Evaluation of Direct Plane Fitting for Depth and Parameter Estimation

Nils Einecke, Julian Eggert

2010

2010 Digital Image Computing: Techniques and Applications

Evaluation of Direct Plane Fitting for Depth and Parameter Estimation

Nils Einecke and Julian Eggert
Honda Research Institute Europe
Offenbach, Germany
nils.einecke@honda-ri.de, julian.eggert@honda-ri.de

***Abstract*—Recently, a model-based depth estimation technique has been proposed, which estimates surface model parameters by means of Hooke-Jeeves optimization. Assuming a parametric surface model, the parameters best explaining the perspective changes of the surface between different views are estimated. This constitutes a fitting of models directly into stereo images, which is in contrast to the usual approach of fitting models into pre-processed disparity data. In this paper, we conduct a comparison of the image fitting based on Hooke-Jeeves, an image fitting based on gradient descent and a disparity fitting based on RANSAC. We show that the image fitting based on Hooke-Jeeves as well as the image fitting based on gradient descent are sensitive to occlusion. However, we also propose a simple pre-processing that eliminates this problem. Our experiments revealed that all three approaches have a similar depth accuracy. However, tests under challenging conditions show that the fitting based on Hooke-Jeeves is more robust than RANSAC and gradient descent.**

***Keywords*-plane fitting; stereoscopic depth**

I. Introduction

A lot of approaches in computer vision involve the fitting of planar models. There are two main applications of plane fitting. The first is to improve depth maps generated by a traditional correlation-based stereo processing. The key assumption here is that homogeneous image regions are likely to be planar. As correlation-based stereo is prone to produce errors or holes in homogeneous regions because of the weak texture, the planar fitting of disparity values leads to a robust fill-in for such areas. On the other hand planar fitting is used in many systems to estimate the orientation of planar surfaces.

A naive approach for fitting a surface model into 3D data or a disparity map is a least squared error fitting, but this is seldom done because it is very sensitive to outliers. To account for this problem the random sample consensus (RANSAC) [1] has been developed. The main idea of RANSAC is to find an outlier-free subset of the data points by testing a set of hypotheses. These hypotheses are generated from randomly sampled data points. Because of its simple yet powerful nature, RANSAC has become a common tool for fitting surface models to range data, especially planar models [2], [3].

Fitting models into disparity has one major drawback. It depends on the quality of the pre-processed stereo disparity maps. If the stereo processing fails for a larger area due to strong ambiguities then the model fitting will likely fail. In these challenging situations it is advisable to incorporate the surface model directly into the correspondence search in order to increase the robustness of the depth estimation itself. This could also be interpreted as fitting the surface model directly to the stereo images. State-of-the-art approaches [4], [5] do this by applying a homography [6] mapping. A homography describes the mapping of planar surfaces between different camera views. The parameters of the homography are usually estimated by means of gradient descent.

Unfortunately, approaches based on the homography mapping are restricted to planar surfaces. Moreover, for surface models others than planes it is very difficult to derive the necessary equations for gradient descent. In order to circumvent the need for complex gradient formulas, Einecke et al. proposed a method [7] based on the Hooke-Jeeves optimization [8]. As Hooke-Jeeves does not use gradients for the optimization process it is quite easy to change the formulas for fitting other parameterizable surface models. The only thing that has to be done, is to describe the perspective mapping of a surface model between different camera views. Einecke et al. exemplary demonstrated this for planes, spheres and cylinders and showed that model parameters can be estimated accurately. However, they did not compare their results to other approaches.

In this paper, we compare the fitting of models into stereo images based on Hooke-Jeeves, fitting to images based on gradient descent and model fitting into disparity data based on RANSAC. We limit the comparison to planar surfaces in order to have a fair comparison between all three approaches. We show that image fitting methods are sensitive to occlusion but we present a simple pre-processing that eliminates this problem. Furthermore, our experiments highlight that image fitting techniques are indeed superior to disparity fitting techniques for images with strong ambiguities. Moreover, we show that the fitting based on Hooke-Jeeves optimization is more robust than gradient based approaches.

II. Plane Fitting Methods

In this paper, we compare three different methods for planar fitting: fitting to stereo disparity data using random

978-0-7695-4271-3/10 $26.00 © 2010 IEEE
DOI 10.1109/DICTA.2010.89

IEEE computer society

sample consensus (RANSAC), fitting to images by gradient descent and fitting to images by Hooke-Jeeves optimization.

A. RANSAC

The most straightforward way of fitting a plane into stereo disparity data is to apply a least squared error (LSE) fitting. It is well known, however, that LSE fitting is very sensitive to outliers. Hence, these outliers have to be removed. A common approach for this is the RANSAC [1] method, which comprises three basic steps:

1) Randomly select just enough data points for the model parameter calculation, i.e. three data points for a planar model.
2) Analytically determine the model parameters from the selected data points.
3) Calculate the size of the census set. This set consists of all data points whose distance from the estimated model is below a threshold δ.

These steps are repeated for a certain number of times. Afterwards the largest census set is used for fitting the model, e.g. by means of LSE fitting.

B. Gradient Descent

As explained above, stereo estimation itself can only be improved by incorporating the surface models directly into the correspondence search. The reason is that this allows for larger patch sizes to be matched. A common way of doing so is to incorporate a homography H which describes the transformation of a planar surface between two camera views. The goal is to estimate the homography parameters that best describe the observed mapping of an image region S. For two stereo images I^L (left) and I^R (right), this estimation is described by the minimization

$$\min_{\mathbf{p}} \sum_{\mathbf{x} \in S} \left(I^L(\mathbf{x}) - I^R(H(\mathbf{x}, \mathbf{p})) \right)^2 \ , \qquad (1)$$

where $H(\mathbf{x}, \mathbf{p})$ is the homography mapping of the image coordinates $\mathbf{x}$ with the parameters $\mathbf{p}$. Deriving (1) for the homography parameters $\mathbf{p}$ leads to an iterative gradient descent which is very similar to the image registration proposed by Lucas and Kanade [9]. For example Habbecke and Kobbelt [4] elaborated on this idea using a Gauss-Newton style matching and approximated partial image derivatives. In this paper, we use the traditional approach of Lucas and Kanade for comparison. This becomes possible as the perspective changes of a plane reduce to an affine warping for rectified stereo images with horizontal epipolar lines

$$\min_{\mathbf{A}, \mathbf{d}_a} \sum_{\mathbf{x} \in S} \left(I^L(\mathbf{x}) - I^R(\mathbf{A}\mathbf{x} + \mathbf{d}_a) \right)^2 \ . \qquad (2)$$

Here $\mathbf{A}$ describes the scaling and the shear and $\mathbf{d}_a$ describes the translation. For the planar estimation in a parallel camera setting only three parameters are of interest $\mathbf{p} = (p_1, p_2, p_3)$,

$$\mathbf{A} = \begin{pmatrix} p_1 & p_2 \\ 0 & 1 \end{pmatrix} \ , \ \mathbf{d}_a = (p_3, 0)^T \ . \qquad (3)$$

C. Hooke-Jeeves Optimization

A major drawback of the homography estimation is its restriction to a planar model. Recently, Einecke et al. proposed a surface fitting [7] based on Hooke-Jeeves optimization that overcomes this limitation. The starting point for this approach is also the minimization (1). However, instead of the homography, they use the basic stereo mapping equation

$$\mathbf{u}_R = \mathbf{u}_L - b\frac{f}{z} \begin{pmatrix} 1 \\ 0 \end{pmatrix} \ , \qquad (4)$$

where b and f are the baseline and the focal length of the stereo camera system. By means of the above equation a pixel $\mathbf{u}_L$ from the left camera can be mapped to a pixel $\mathbf{u}_R$ in the right camera using the depth z. In order to map a parametric surface, z has to be described in terms of the surface's parametric description. This means that the 3-D formulation of a surface has to be rearranged for z. For a planar model this leads to

$$z = f \frac{x_a \sin\alpha_y - y_a \tan\alpha_x + z_a \cos\alpha_y}{u_{Lx} \sin\alpha_y - u_{Ly} \tan\alpha_x + f \cos\alpha_y} \ . \qquad (5)$$

In this equation (x_a, y_a) is the 2-D anchor point of the plane which can be chosen freely. The parameters to estimate are the depth z_a of the anchor point and the orientation (α_x, α_y) of the plane. Mapping formulas for spheres and cylinders are described in [7].

Please note that all three methods (RANSAC, gradient descent and Hooke-Jeeves) are comparable with respect to the complexity of the estimation. All three methods have to estimate three model parameters. Of course these parameters can easily be transformed from one algorithm's domain to the domain of the other algorithms, which is necessary in order to compare the different methods.

III. Experiments

In this paper, we use two different sets of stereo images for comparing and evaluating the three plane fitting methods. Firstly, we use the Venus, Bull, Sawtooth and Poster scenes (see Fig. 1) from the Middlebury stereo evaluation data sets [10]. These scenes consist solely of planar surfaces which makes them well suited for the analysis of planar fitting. In order to test the fitting performance of the different approaches under more realistic conditions, we refrained from using a hand segmentation of the scene. Instead, we applied a simple region growing to the left camera images. From these regions we selected the ones with more than 100 pixels (see Fig. 2). The other regions are ignored in the following evaluations. This first setup assesses the general performance of the three methods and how well they can cope with partial occlusions. Note that we compare the methods on the basis of the disparity error and not the planar parameter error because this is more expressive for the actual performance.

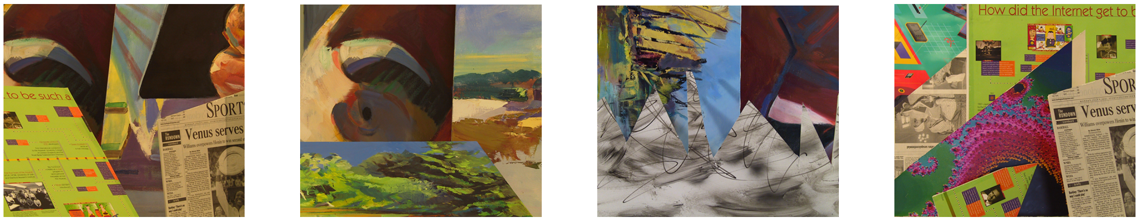
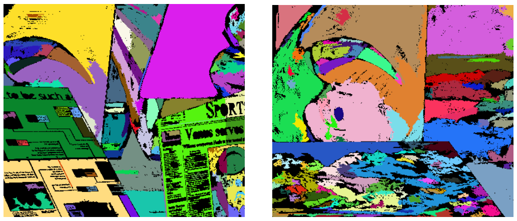


Figure 1. Left images of the Venus, Bull, Sawtooth and Poster scene.



(a) Venus

(b) Bull

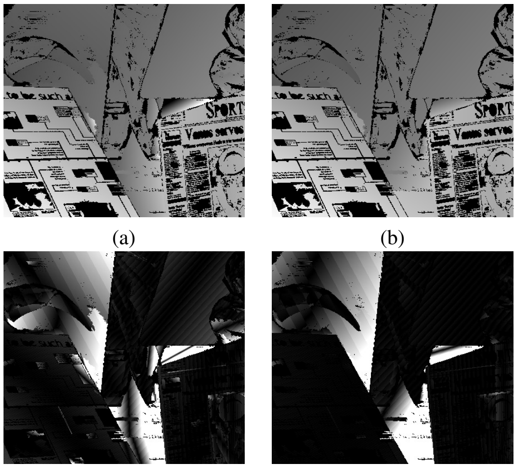
Figure 2. Segmentation of the Venus and the Bulls scene into homogeneous regions by means of region growing. The regions are illustrated in pseudo-colors. Regions smaller than 100 pixels are discarded and shown in black.

The second setup consists of some stereo camera images taken from within a car while driving. The task for the three plane estimation methods is to estimate the position and orientation of the street. We use the challenging images of a wet street which exhibits many distracting reflections.

A. General Performance and Occlusions

When we started our evaluation, we quickly found out that approaches that fit models into images are prone to fail for regions that are partially occluded in the other image. For example Fig. 3 shows the resulting disparity and disparity error maps for the Venus scene for gradient descent and Hooke-Jeeves optimization. It can be observed that both methods tend to wrongly estimate the orientations of background regions near occluding objects. This effect is most striking for the region above the newspaper (brownish, triangular region in Fig. 2a).

An analysis of the problematic regions revealed that the main reason for the bad performance are the usually large intensity differences between a background region and its occluder. As parts of an occluded region are only visible in one camera image, a correct warping from one camera image into the view of the other camera image will lead to an overlay of completely different intensity values. This in turn will lead to large errors in the squared error distance of the minimization equation (1). As a first solution, we tried to replace the squared error by other error measures. The best results were achieved with truncated error measures. However, truncated measures are unsatisfactory because the threshold for truncation is not stable over different scenes. In search of a better solution, we came up with an old idea that is usually used to make correspondence search between stereo images invariant to illumination changes.



(a)

(b)

(c)

(d)

Figure 3. The first row shows the disparity maps of the Venus scene for planar image fitting using a) gradient descent and b) Hooke-Jeeves optimization. The planar fitting was applied to each region shown in Fig. 2a. For better visualization c) and d) show the disparity error maps (pixel-wise absolute difference to ground truth). The error is encoded by intensity, large errors (more than 2 pixels disparity) are shown in white and small errors in black. These results were achieved using the plain gray images. They highlight the usual tendency of image fitting methods to fail for partially occluded regions.

We found out that normalizing the stereo images reduces the intensity difference to a level were it does not disturb the surface matching. In particular, we normalize the images by:

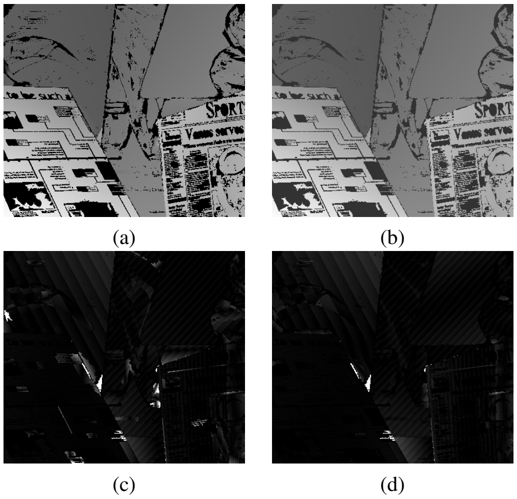
$$I_x^{\text{norm}} = \frac{I_x - \mu_x}{\sigma_x} \ , \tag{6}$$

where

$$\mu_x = \frac{1}{|N(x)|} \sum_{x' \in N(x)} I_{x'} \ , \tag{7}$$

$$\sigma_x = \sqrt{\frac{1}{|N(x)|} \sum_{x' \in N(x)} (I_{x'} - \mu_x)^2} \ . \tag{8}$$

This means that for each pixel x, we calculate the mean intensity and the intensity variance in its neighborhood $N(x)$. By subtracting the mean and dividing by the standard



(a) (b)

(c) (d)

Figure 4. Similar to Fig. 3 the disparity and disparity error maps for Hooke-Jeeves and gradient descent planar image fitting are shown. In contrast to Fig. 3, mean and variance normalized images were used for matching. It demonstrates that image normalization is able to dramatically reduce the occlusion problem for image fitting methods.

deviation the pixel is normalized. We found that the normalization is most effective for small neighborhoods. Here we use a neighborhood of 3x3 pixels.

The significant improvement for both gradient descent and Hooke-Jeeves using the normalized images is shown in Fig. 4. Comparing the error maps of gradient descent and Hooke-Jeeves without image normalization (Fig. 3c and 3d) to the resulting error maps with image normalization (Fig. 4c and 4d) shows a clear improvement. These results demonstrate that image normalization dramatically reduces the occlusion problem for image fitting methods.

It is important to note, that the improvement cannot be explained by a reduced difference of corresponding pixels. Firstly, the Venus scene was taken under ideal conditions, i.e. corresponding pixels have almost the same intensity. Secondly, the improvement vanishes if larger neighborhoods are used for mean and variance calculation. Last but not least, wrongly estimated planar fittings would not populate around depth discontinuities but rather distribute equally all over the image. The actual reason for the improvement is that the normalization makes pixels within one image more equal. Thus, the influence of strong contrasting structures, like the strong contrast edge between the bright newspaper and the dark background, is reduced.

Now that the occlusion problem is resolved a reasonable comparison of the image fitting methods to disparity fitting based on RANSAC is feasible. Table I compares the performance of the three approaches for the four Middlebury test scenes. Here we calculated the percentage of bad pixels [10] for the estimated planar surfaces for an error threshold of 0.5. These results show that all three approaches have a similar performance.

Table I

PERFORMANCE OF THE THREE PLANE FITTING METHODS RANSAC, GRADIENT DESCENT AND HOOKE-JEEVES. THE PERFORMANCE IS MEASURED BY THE PERCENTAGE OF BAD PIXELS, I.E. PIXELS WHOSE DISPARITY DIFFERS MORE THAN 0.5 FROM GROUND TRUTH.

scene	RANSAC	gradient descent	Hooke-Jeeves
Venus	1.42%	1.65%	**0.37**%
Bull	**1.37**%	1.68%	1.68%
Sawtooth	8.24%	7.55%	**7.05**%
Poster	**5.20**%	8.49%	5.44%

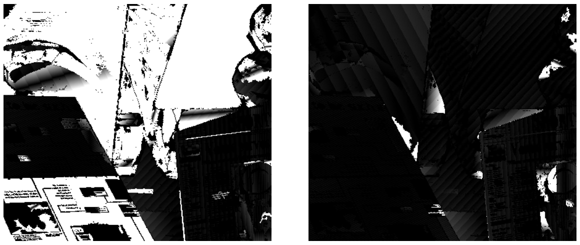


Figure 5. This figure shows the error maps of gradient descent fitting (left) and Hooke-Jeeves fitting (right) with fixed initialization at 10 disparity. It is clearly observable that the image fitting based on Hooke-Jeeves is much less sensitive to initialization as compared to gradient descent fitting.

In contrast to RANSAC, image fitting approaches based on gradient descent or Hooke-Jeeves need some initial guess of the parameters. Thus, an important criterion for image fitting approaches is the robustness against the initialization. So far we initialized the fitting with a fronto-parallel assumption and the starting disparity was acquired by standard patch matching. In order to test how well gradient descent and Hooke-Jeeves can cope with errors of the initialization, we applied them again on the Venus scene but this time the initial disparity was fixed to 10 pixels disparity. The resulting error maps in Fig. 5 demonstrate that Hooke-Jeeves can cope very well with such a bad initialization. On the other hand, gradient descent gets stuck in a local minimum for many regions. The sensitivity of gradient descent image fitting could be reduced with a resolution pyramid. However, the results show that in most cases such a costly processing is not necessary for the Hooke-Jeeves optimization.

B. Performance under Heavy Distortions

The general claim of image fitting approaches is that by integrating over larger image areas for fitting, the robustness is improved and the aperture problem reduced. In order to test this hypothesis, we recorded a short stream (320 frames) with a camera system mounted on a car. The stream was recorded under rainy weather condition. The goal for all three approaches is to estimate the planar parameters of the street in front of the car.

Fig. 6a shows exemplarily one frame of this rain stream. As can be seen in Fig. 6b the traditional correlation-based stereo processing struggles to find correct correspondences. The reason for this is the superposition of the street structure

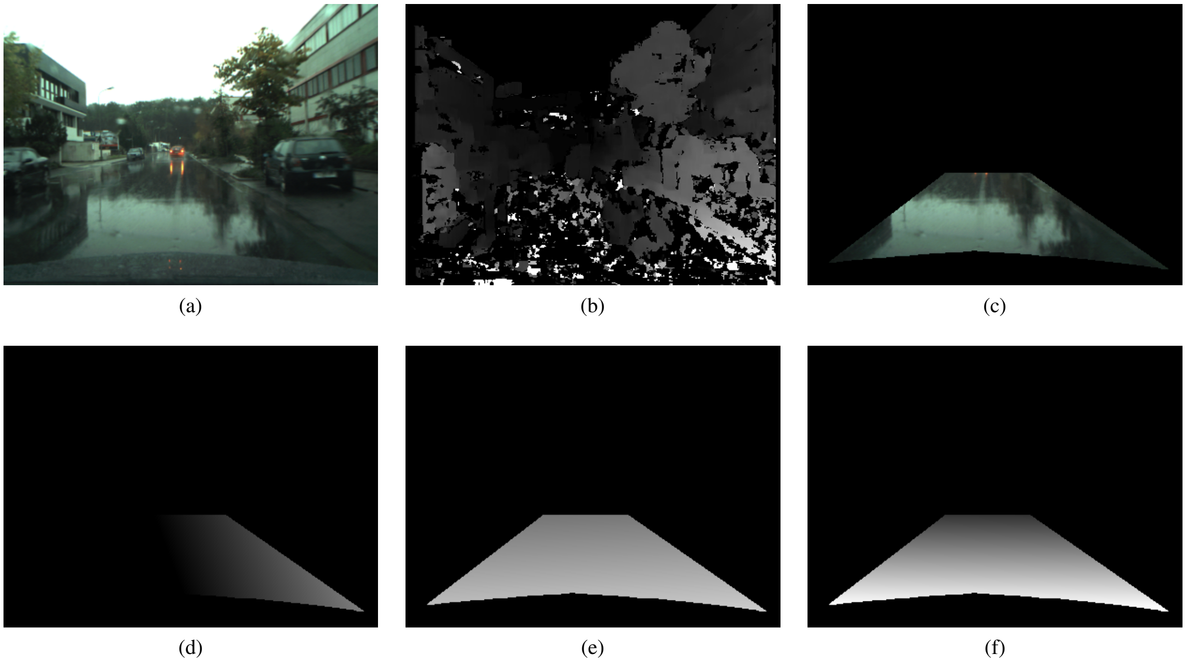
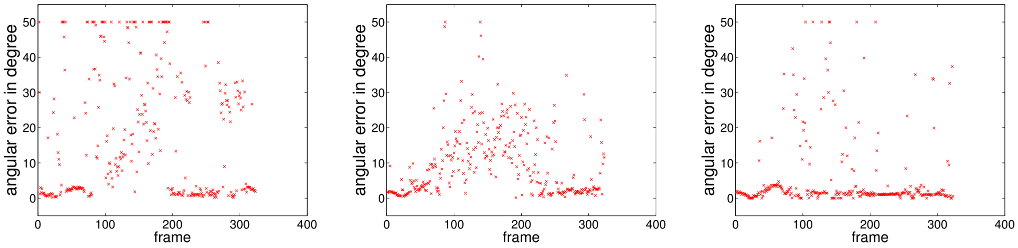


Figure 6. a) Wet street with a mirror-like state and b) disparity map of traditional stereo. c) Image region for which the planar fitting is done. Resulting disparity maps: d) RANSAC, e) gradient descent f) the method based on Hooke-Jeeves.



(a) RANSAC (b) Gradient Descent (c) Hooke-Jeeves

Figure 7. These plots show the angular error of the estimation of the pitch angle of a camera mounted in a car. The stereo images were taken under rainy weather condition (see also Fig. 6).

with the image of the surrounding because of the mirror-like state of the street. Due to the bad performance of the correlation-based stereo, the RANSAC fitting is prone to fail to find a reasonable estimation of the street. Fig. 6d shows that the fitted plane is very skewed. In contrast to this the result of gradient descent (see Fig. 6e) and Hooke-Jeeves (see Fig. 6f) are much better. However, in this frame gradient descent was not able to find the correct inclination of the street.

We applied all three methods to the 320 frames (32 seconds) of the rain stream. In order to assess the robustness of the plane fitting under these challenging conditions, we evaluate the so-called pitch angle. The pitch angle is used in many intelligent vehicle approaches, for example as an additional input for obstacle segmentation in order to account for camera rotations relative to the street. This angle changes when the car accelerates, decelerates or due to small irregularities on the street's surface. Here, we have no ground truth of the pitch angle, but we know that it closely varies around $90°$ because the angle between camera and street is roughly $90°$ when the car is standing still. By means of this knowledge, we can calculate an angular error that should be close to zero but is allowed to have small deviations from this as long as it is a smooth change. Fig. 7 shows the angular error of the pitch angle for all three methods. For better contrast we clipped angular errors larger than $50°$.

The first thing that strikes is that RANSAC and gradient decent fail to estimate the pitch angle between frame 80 and 200. This is the most difficult part of the stream with heavy reflections as seen in Fig. 6a. In contrast the estimation with Hooke-Jeeves produces good results with only a few outliers for the whole stream. For the frames before 80 and after 200 the three approaches have roughly the same performance. This shows that in principle all three approaches have a similar performance but under difficult conditions the estimation based on the Hooke-Jeeves optimization is more robust than RANSAC or gradient descent estimation. However, the results have to be seen with same caution. The performance of the gradient descent and Hooke-Jeeves optimization depend on the initialization. In order to be fair for RANSAC we initialized the search for gradient descent and Hooke-Jeeves at 70°, i.e. with an initial error of approximately 20°. This is in general close enough for gradient descent but not so close that the estimation becomes too easy with respect to the RANSAC estimation. Nevertheless, a more thorough analysis has to be done in future work in order to investigate the degradation of the image fitting approaches with respect to a decreasing accuracy of the initialization. On the other hand a processing on image streams allows for a temporal integration or tracking [11] over time which limits the necessity of a good initialization to the first frames.

IV. Summary

In this paper, we compared a recently proposed method based on Hooke-Jeeves optimization for fitting planes directly to stereo images to a gradient descent approach for fitting planes into images and to RANSAC for fitting planar models into pre-processed disparity maps. We showed that the image fitting approaches tend to give wrong estimates for partially occluded surface regions. However, we also showed that this problem can be eliminated by a simple normalization of the stereo images. Our experiments also revealed that Hooke-Jeeves is much less sensitive to initialization as compared to gradient descent. This makes a costly processing on multiple scales superfluous.

Furthermore, we compared the overall performance of the three approaches. RANSAC, gradient decent and Hooke-Jeeves perform quite good for most situations. However, in challenging scenes RANSAC occasionally fails due to the bad performance of the stereo estimation step. In these cases the image fitting methods should be advantageous as they base their correspondence search on larger image patches. Surprisingly, we could only observe this improvement for the image fitting based on Hooke-Jeeves. The gradient based approach failed. This could be related to the sensitivity of gradient descent against the initialization and should be analyzed more in detail in future work. Altogether, we conclude that RANSAC, gradient descent and Hooke-Jeeves show a similar performance for most situations but the image fitting based on Hooke-Jeeves produces more reliable results than RANSAC for heavily distorted images and is less sensitive to initialization compared to gradient descent.

References

[1] M. A. Fischler and R. C. Bolles, "Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography," *Commun. ACM*, vol. 24, no. 6, pp. 381–395, 1981.

[2] M. Heracles, B. Bolder, and C. Goerick, "Fast detection of arbitrary planar surfaces from unreliable 3D data," in *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2009, pp. 5717–5724.

[3] Q. Yang, C. Engels, and A. Akbarzadeh, "Near real-time stereo for weakly-textured scenes," in *BMVC08*, 2008.

[4] M. Habbecke and L. Kobbelt, "Iterative multi-view plane fitting," in *Vision, Modeling, Visualization VMV'06*, 2005, pp. 73–80.

[5] 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*, 2002, pp. 3321–3327.

[6] R. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*, 2nd ed. Cambridge University Press, 2004.

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

[8] R. Hooke and T. A. Jeeves, ""Direct Search" Solution of numerical and statistical problems," *Journal of the Association for Computing Machinery*, vol. 8, no. 2, pp. 212–229, 1961.

[9] B. D. Lucas and T. Kanade, "An iterative image registration technique with an application to stereo vision," in *International Joint Conference on Artificial Intelligence*, 1981, pp. 674–679.

[10] D. Scharstein and R. Szeliski, "A taxonomy and evaluation of dense two-frame stereo correspondence algorithms," *International Journal of Computer Vision*, vol. 47, no. 1-3, pp. 7–42, April-June 2002.

[11] J. J. Corso, D. Burschka, and G. D. Hager, "Direct plane tracking in stereo images for mobile navigation," in *International Conference on Robotics and Automation*, 2003, pp. 875–880.