On-line Stereo Self-Calibration through Minimization of Matching Costs

Robert Spangenberg, Tobias Langner and Raúl Rojas

Original Publication:

Spangenberg, R., Langner, T., Rojas, R.: On-line stereo self-calibration through minimization of matching costs. In Kämäräinen, J.K., Koskela, M., eds.: Image Analysis. Volume 7944 of Lecture Notes in Computer Science. Springer Berlin Heidelberg (2013), 545-554

The final publication is available at link.springer.com:

http://link.springer.com/chapter/10.1007/978-3-642-38886-6_51

On-line Stereo Self-Calibration through Minimization of Matching Costs

Robert Spangenberg, Tobias Langner, and Raúl Rojas

Freie Universität Berlin, Institut für Informatik,
Arnimallee 7, 14195 Berlin, Germany
robert.spangenberg@fu-berlin.de,
http://www.autonomos.inf.fu-berlin.de

Abstract. This paper presents an approach to the problem of on-line stereo self-calibration. After a short introduction of the general method, we propose a new one, based on the minimization of matching costs. We furthermore show that the number of matched pixels can be used as a quality measure. A Metropolis algorithm based Monte-Carlo scheme is employed to reliably minimize the costs. We present experimental results in the context of automotive stereo with different matching algorithms. These show the effectiveness for the calibration of roll and pitch angle offsets.

Keywords: self-calibration, stereo vision, matching costs

1 Introduction

As a consequence of their low cost, vision systems are being used more and more in the automotive context to implement various driver assistance systems. Stereo-vision adds dense depth information to the intensity/texture information a monocular camera can provide. To enable a fast estimation of depth out of the stereo camera images, most of the systems rely on perfectly aligned cameras or sufficient information about their relative position and orientation in order to align them mathematically. This enables to work on a rectified images, where each 3D point is projected on the same line in both virtual camera images. Real-time stereo processing relies heavily on this fact. The calibration process has to estimate the camera orientation with an accuracy of about 10^{-2} degrees [1]. This means sub-pixel accuracy. The usage of a suitable stereo matching algorithm might mitigate the errors created by mis-calibration [5], but is not sufficient in general to reach the desired accuracies.

The assumption of an "ideal" rigid stereo rig, whose cameras do not change their orientations and positions, leads to the sole usage of a traditional camera calibration method using special reference patterns [12]. In an industrial or automotive context, mechanical vibrations, large temperature variations and material fatigue cause drifting of the camera parameters. Only self-calibration can compensate for this and enable a long-term operation of the stereo vision system without any need of manual intervention. In addition, the often costly initial off-line calibration step could be saved, enabling a faster and more economical production.

We are working on an integrated system in an automotive urban context, where the vision system recalibrates itself periodically. This calibration is only a support task. It does not have to work in real-time, but should be functional in a wide range of scenes that can be encountered.

2 Related Work

A comprehensive approach to on-line stereo-calibration in the automotive context is presented in [1]. In this article, constraints arising out of recursive bundle adjustment are used, furthermore the epipolar constraint between a pair of stereo images and the trifocal constraint. Feature points are detected using the SIFT feature detector [9]. In combination with an Iterated Extended Kalman Filter they achieve a robust framework even in the context of active vision. The work of [8] focuses on the epipolar constraint and follows the common steps in the self-calibration procedure, which comprises of a feature point detector, a matching algorithm and the computation of the associated Fundamental matrix. The ”Minimum Eigen Value” in the classical structure tensor are used as input feature. They are not sub-pixel refined. Matching is modified by an additional correlation based filtering step, using the correlation score to measure the ambiguity of the features.

In both approaches, the calculation of the depth image through stereo matching and the self-calibration procedure are distinct tasks. Reuse of algorithms for stereo matching is low and especially the approach in [1] is rather complicated. Our approach is somewhat different and simpler, as it aims to use the outputs of the stereo matching algorithm itself as a tool to improve the calibration.

In contrast to the standard approach, we modify the calibration and measure the accuracy of the calibration through the output of the stereo matching. Therefore we first present measures that are readily available as outcomes of the stereo matching - matching costs and the percentage of unmatched pixels. Secondly, we propose a scheme to efficiently guide the search process using a Markov-chain Monte Carlo method. Finally, we present some experimental results for different matching algorithms.

3 Calibration Parameters of Interest

We assume the internal calibration parameters of both cameras to be stable and known. Furthermore, an initial guess of the external calibration is assumed to be available as well. Regarding the external parameters, we have a relative rotation and translation which are of interest. A thorough sensitivity analysis in [1] and practical experience show strong correlations between small distortions of principal point coordinates and extrinsic orientation parameters. The base length cannot be retrieved from image observations alone. So, for a fixed camera set-up, the calculation of offsets for yaw, pitch and roll angle is sufficient.

4 Measuring Calibration Accuracy

In contrast to the typical task of stereo matching evaluation, no ground truth is available. We have to rely on the output of the algorithm. As stereo matching is typically seen as a minimization problem, we can use the associated costs as an evaluation criterion.

We can approach the process of stereo matching as a labelling problem. Let the set of all pixels in the image be P and D be a finite set of disparities. Then f is the result of a stereo matching and assigns every pixel $p \in P$ a disparity $f_p \in D$. The quality of this labelling can be evaluated by the following energy function:

$$E(f, P) = \sum_{p \in P} \left(D_p(f_p) + \sum_{q \in N(p)} W(f_p, f_q) \right) \tag{1}$$

being $N(p)$ the 4-way neighborhood of p, $D_p(f_p)$ the data costs of the assignment of f_p to p and $W(f_p, f_q)$ a cost measure between f_p and its neighbor pixels f_q, the smoothness costs. Common stereo matching algorithms try to minimize these cost either locally or globally. Simple algorithms neglect the $W(f_p, f_q)$ part and solely rely on the data cost part. Let Q_f be the set of pixels with a valid disparity value in the assignment f, then

$$\bar{E}(Q_f) = \frac{E(f, Q_f)}{|Q_f|} \tag{2}$$

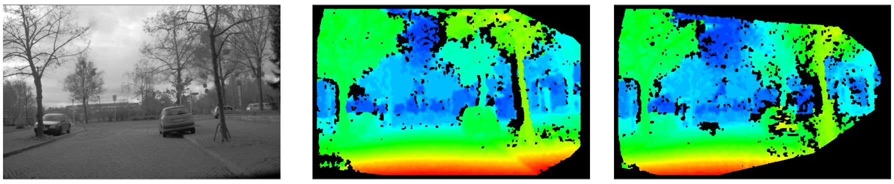
is the average matching cost per valid pixel. This cost should be higher, if the calibration is incorrect.

For most algorithms it is rather easy to calculate the cost of the chosen assignment. In some cases, the matching cost might be not readily available or induce a performance penalty, e.g. in a fixed FPGA implementation. In these cases, the percentage of pixels with an valid disparity could be a measure for calibration accuracy

$$Val(f) = \frac{|Q_f|}{|P|}. \tag{3}$$

Again $Val(f)$ should be higher compared to $Val(f')$ of the assignment f' with the the correct calibration. An example backing this notion can be seen in figure 1. A mis-calibration leads to a significant reduction of valid pixels compared to the result for the optimal calibration. We use the method proposed in [3] as matching algorithm (ELAS), adapted for the use with 12 bit images. The matching costs calculated for this method are simply the data costs, based on the differences of Sobel operator signatures.

A systematic variation of the pitch, roll and yaw angle is depicted in figure 2. It shows clear minima in the proposed costs for pitch and roll. The matching cost shows an advantage in the extent and shape of its minima, but the invalid pixel ratio is usable as well. Nevertheless, a simple gradient descent method is prone to fail in case of bigger changes in calibration even for pitch and roll angle. The yaw angle plot shows no clear minimum and it seems not feasible to estimate yaw angle offsets out of a single image.



(a) Scene 1 (b) Optimal calibration (c) Pitch offset −0.1° and roll offset 0.1°

Fig. 1: Influence of mis-calibration on stereo matching - same scene with different calibrations, disparity hue-coded, black: no valid disparity available (image size 768x480 with 12 bit intensity values, 48° field of view)

5 Skilling's Method

To robustly find the minima of the objective functions, Markov-chain Monte Carlo methods are an option. As an extension to the standard Metropolis algorithm, [10] describes the Leapfrog-method by Skilling. We shortly recall the Metropolis algorithm. It simulates a Markov-chain, with the probability density $P(\mathbf{x})$ using a symmetric proposal or jumping density $Q(\mathbf{x}'|\mathbf{x}_t)$, which is simple to evaluate. The result is a set of samples S=$\{\mathbf{x}_t\}_{t=0...N}$.

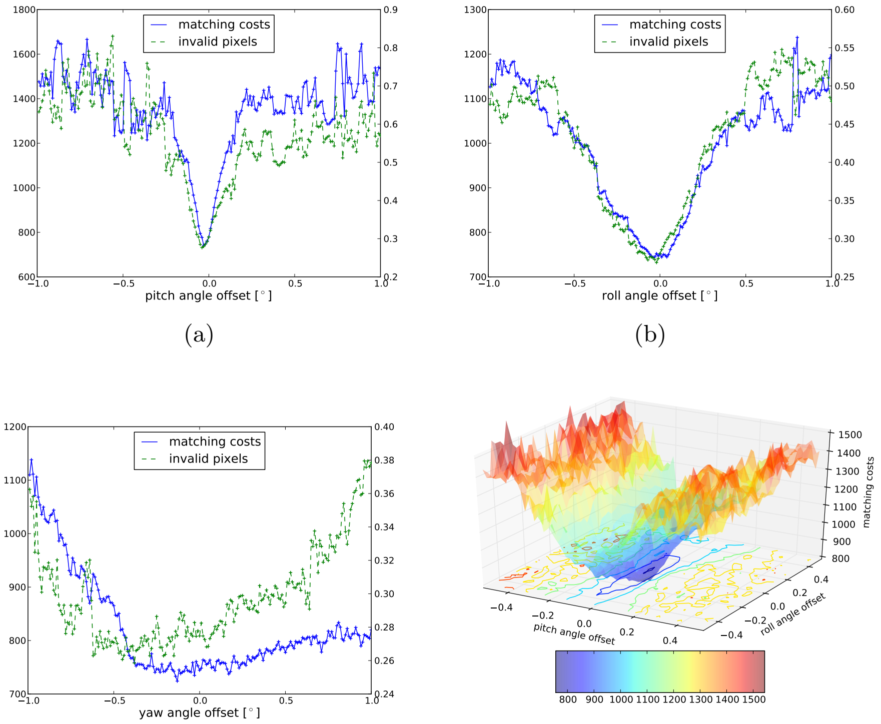
Metropolis-Algorithm

1. $\mathbf{x}_0$ is chosen by chance as first sample
2. Given $\mathbf{x}_t$ one creates a $\mathbf{x}_{t+1}$ through:
   - $\mathbf{x}'$ is created by sampling $Q(\mathbf{x}'|\mathbf{x}_t)$.
   - Calculate acceptance rate $\alpha = \frac{P(\mathbf{x}')}{P(\mathbf{x}_t)}$.
   - Create a uniform random number $r \in [0, 1]$
   - $\mathbf{x}_{t+1} = \begin{cases} \mathbf{x}', & \text{if } \alpha \geq r \\ \mathbf{x}_t, & \text{else} \end{cases}$

The expectation is that the simulated Markov-chain will visit the maxima in the state space. As the Metropolis algorithm is highly inefficient, especially for densities shaped like the one arising of our optimization goal (figure 2), we employ the Leapfrog-method by Skilling, which is a variation of the above. Instead of one state vector $\mathbf{x}$ we employ a small number, say 6 or 12 of state vectors $\mathbf{x}^{(s)}$ simultaneously. A new state vector $\mathbf{x}^{(s)}$' is created with another one $\mathbf{x}^{(t)}$ by:

$$\mathbf{x}'^{(s)} = \mathbf{x}^{(t)} + (\mathbf{x}^{(t)} - \mathbf{x}^{(s)}) = 2\mathbf{x}^{(t)} - \mathbf{x}^{(s)} \tag{4}$$

The partner state $\mathbf{x}^{(t)}$ is chosen either at random or weighted by a distance function. In the latter case the detailed balance condition has to be fulfilled. This algorithm has the advantage to adapt better to the shape of the estimated density, as can be seen in figure 3. This leads to a more efficient way through



(a)

(b)

(c)

(d) Pitch and roll angle

Fig. 2: Systematic variation of relative angles in the stereo rig through additional offsets.

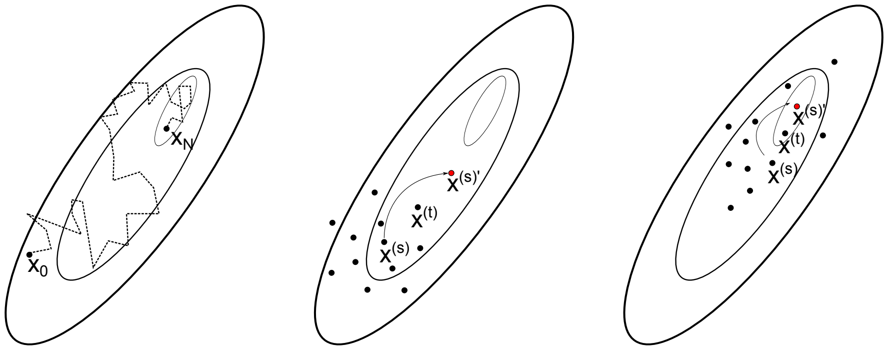
state-space. Furthermore it needs no proposal density, which is sometimes hard to specify.

The simulated chain of the leapfrog method is filtered for the cost minimum. We finally perform a local deterministic search around this point in state space to gain additional accuracy.

6 Experimental Results and Discussion

As a first test, we used the algorithm on a scene directly after an traditional off-line stereo calibration. The calculated offsets were zero for pitch and roll angle.

To evaluate the robustness of the approach we took some sample scenes and recorded the results of the algorithm with a spacing of around one second for each used frame. Scene 1 exposes a nearly optimal calibration. Scene 2 has a rather strong de-calibration of the roll angle. Both were recorded on the university



(a) Metropolis walk (b) Skilling: initial step (c) Skilling: after a few iterations

Fig. 3: Markov-Chain Monte-Carlo

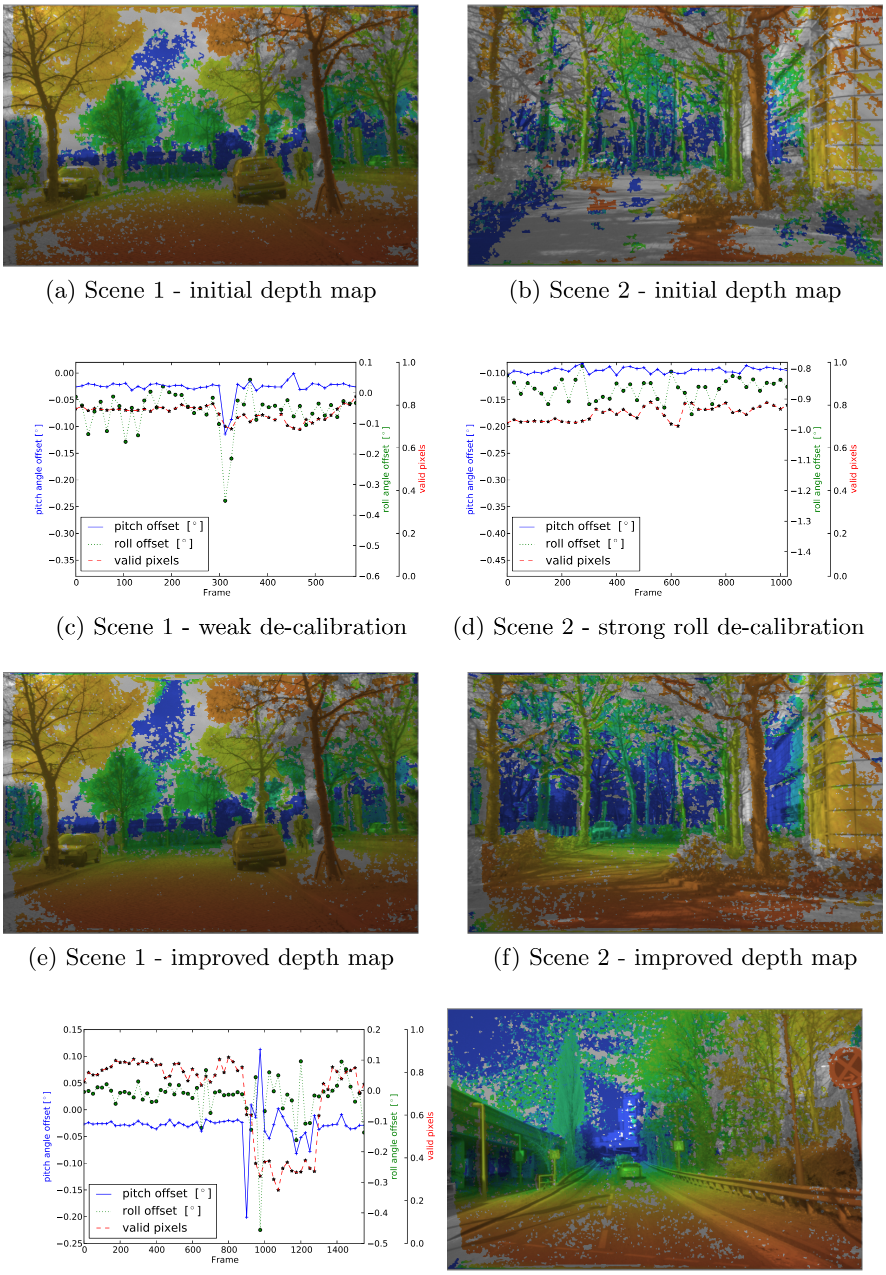
campus, that consists of rather narrow streets. Scene 3 contains bigger streets, tunnels and inner city highways. Starting calibration is nearly optimal here.

The results for each evaluated frame can be seen in figure 4 and the associated statistics in table 1.

Table 1: Optimal angle offsets statistics for scenes 1 to 3 - Results 1* and 3* are the statistics for scene 1 and 3 respectively with an additional valid frame filter, based on the valid pixel ratio.

scene	1	1*	2	3	3*
minimum pitch [°]	-0.114	-0.033	-0.105	-0.201	-0.036
maximum pitch [°]	-0.001	-0.014	-0.082	0.113	-0.009
mean pitch [°]	-0.027	-0.024	-0.095	-0.029	-0.026
stddev. pitch [°]	0.017	0.004	0.005	0.032	0.005
minimum roll [°]	-0.353	-0.160	-0.950	-0.456	-0.073
maximum roll [°]	0.043	0.043	-0.792	0.096	0.095
mean roll [°]	-0.059	-0.049	-0.866	-0.021	-0.001
stddev. roll [°]	0.065	0.042	0.034	0.075	0.030

As no ground truth data is available for these scenes, we can only assess them by their consistency and visual inspection. In scene 2 we see a very consistent algorithm performance. Sub-pixel accuracy for pitch and roll angle is reached. Scene 1 shows a couple of frames with stronger deviations from the estimated mean. This is due to insufficient texture on the tarmac inhibiting a reliable matching of significant fraction of the image (figure 5a and 5b). These frames can be filtered out using the percentage of valid pixels as an indicator (figure 4c). With that adaptation, sub-pixel accuracy is reached here as well. In scene



(a) Scene 1 - initial depth map

(b) Scene 2 - initial depth map

(c) Scene 1 - weak de-calibration

(d) Scene 2 - strong roll de-calibration

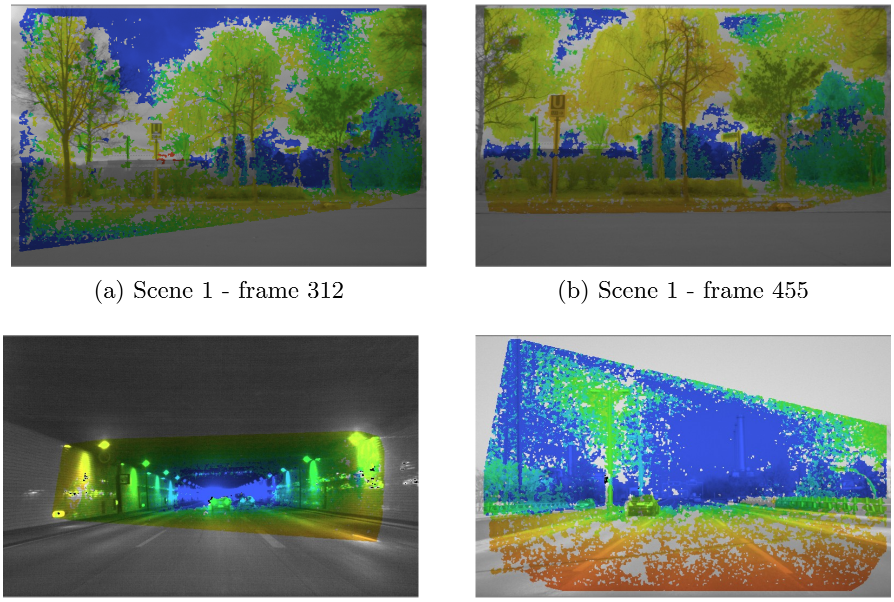
(e) Scene 1 - improved depth map

(f) Scene 2 - improved depth map

(g) Scene 3

(h) Scene 3 - improved depth map

Fig. 4: Results for roll and pitch angle offset estimation for different scenes. ELAS based matching cost measure, improved depth map created using mean estimated offsets for pitch and roll angle. Depth maps are an overlay of the used base image and the color coded disparities.



(a) Scene 1 - frame 312

(b) Scene 1 - frame 455

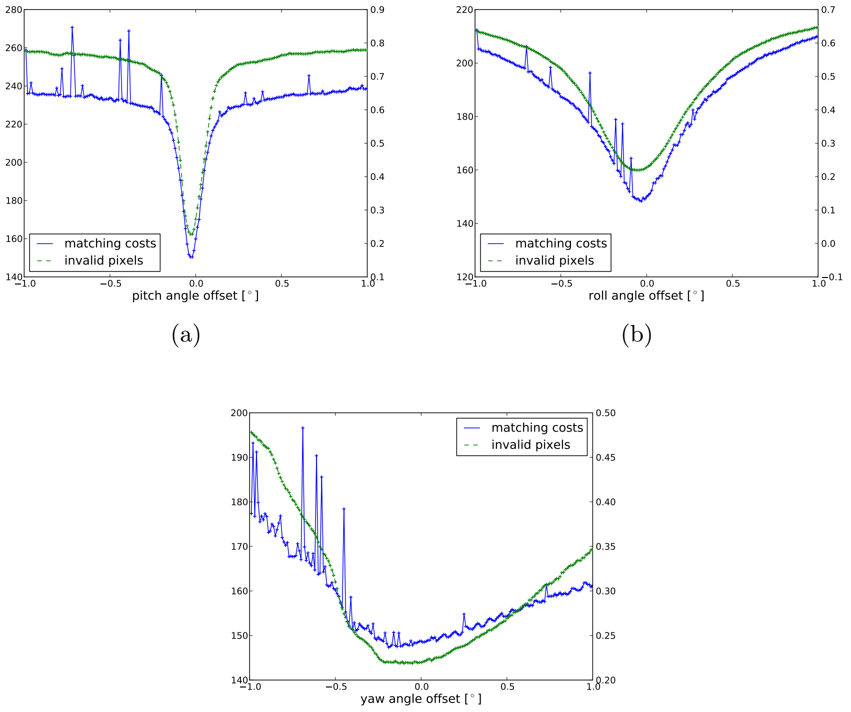
(c) Scene 3 - frame 975

(d) Scene 3 - frame 1525

Fig. 5: Example failure frames - Due to low contrast a significant portion of the image cannot be matched reliably.

3 this validity indicator works as well. Tunnel scenes as in figure 5c permit no matching of a large fraction of the image due to bad lighting and motion blur effects. Some frames are incorrectly classified as invalid, as figure 5d, though overall availability is sufficient. Dynamic thresholding based on the matching costs could be a remedy here.

We tested the sensitivity of the approach to the chosen matching algorithm by running the Semi-Global matching (SGM) algorithm [4] on the same frame as in figure 2. Instead of mutual information we used a 5x5 Census similarity criterion [11] and fixed penalties $P_1 = 7$ and $P_2 = 20$. Census is reported to be comparable to mutual information in case of automotive stereo and even superior in some cases [2],[6]. The results are shown in figure 6 and it shows much smoother plots than figure 2. This might be caused by the inclusion of smoothness costs during the accumulation process in contrast to the ELAS algorithm. A better detection of invalid assignments could contribute to this as well. With SGM the Monte-Carlo-Approach for minimization might not be necessary and one could rely on standard gradient descent methods. This might speed up the calibration process, despite SGM being more costly.



(a)

(b)

(c)

Fig. 6: Systematic variation of relative angles in the stereo rig through additional offsets, SGM algorithm.

7 Conclusion

In this paper we have shown that matching costs are a valuable source of information to improve relative stereo calibration. They can be used for offset estimation of relative roll and pitch angle in a stereo rig. The chosen Monte-Carlo algorithm makes the procedure reliable. Using the fraction of matched pixels as a quality measure, we can filter out situations not fitting the assumptions of the algorithm. In contrast to the general approach of on-line stereo calibration, the approach is rather simple and does not need much additional code. It furthermore seems favorable to perform matching and self-calibration within the same framework. The experiments show the effectiveness of the approach in an automotive context. Similar results can be expected in applications of stereo vision like field robotics or active vision.

Whether it is feasible to calibrate the yaw angle using only matching costs is still an open point in the presented data. One could think about integrating the matching costs from several frames or scenes to get a reliable minimum.

Enhanced analysis of the matching costs arising of each frame might be needed as well. Furthermore using more robust estimators than the mean seems worth additional research.

Acknowledgments. We thank the Autonomos Labs team for their support and constant effort. Thanks to the creators and contributors of Matplotlib for their versatile plotting environment [7].

References

1. Dang, T., Hoffmann, C., Stiller, C.: Continuous stereo self-calibration by camera parameter tracking. Trans. Img. Proc. 18(7), 1536–1550 (Jul 2009), http://dx.doi.org/10.1109/TIP.2009.2017824
2. Gehrig, S.K., Rabe, C.: Real-Time Semi-Global Matching on the CPU. In: Proceedings of the IEEE Computer Vision and Pattern Recognition Workshops. pp. 85–92. San Francisco, CA, USA (June 2010)
3. Geiger, A., Roser, M., Urtasun, R.: Efficient large-scale stereo matching. In: Asian Conference on Computer Vision. Queenstown, New Zealand (November 2010)
4. Hirschmüller, H.: Stereo processing by semiglobal matching and mutual information. IEEE Trans. Pattern Anal. Mach. Intell. 30(2), 328–341 (2008)
5. Hirschmüller, H., Gehrig, S.K.: Stereo matching in the presence of sub-pixel calibration errors. In: CVPR. pp. 437–444. IEEE (2009)
6. Hirschmüller, H., Scharstein, D.: Evaluation of stereo matching costs on images with radiometric differences. IEEE Trans. Pattern Anal. Mach. Intell. 31(9), 1582–1599 (2009)
7. Hunter, J.D.: Matplotlib: A 2d graphics environment. Computing In Science & Engineering 9(3), 90–95 (2007)
8. Kramm, S., Miche, P., Bensrhair, A.: Self calibration of a road stereo vision system through correlation criterions. In: Intelligent Vehicles Symposium, 2006 IEEE. pp. 36 – 41
9. Lowe, D.G.: Distinctive image features from scale-invariant keypoints. International Journal of Computer Vision 60, 91–110 (2004)
10. MacKay, D.J.C.: Information Theory, Inference, and Learning Algorithms. Cambridge University Press (2003), http://www.cambridge.org/0521642981, available from http://www.inference.phy.cam.ac.uk/mackay/itila/
11. Zabih, R., Woodfill, J.: Non-parametric local transforms for computing visual correspondence. In: Proceedings of the third European conference on Computer Vision (Vol. II). pp. 151–158. ECCV '94, Springer-Verlag New York, Inc., Secaucus, NJ, USA (1994), http://dl.acm.org/citation.cfm?id=200241.200258
12. Zhang, Z.: A flexible new technique for camera calibration. IEEE Trans. Pattern Anal. Mach. Intell. 22(11), 1330–1334 (2000)