LAWRENCE
LIVERMORE
NATIONAL
LABORATORY

LLNL-PROC-417469

Factorization of correspondence and camera error for unconstrained dense correspondence applications

D. Knoblauch, M. Hess-Flores, M. Duchaineau, F. Kuester

October 2, 2009

5th International Symposium on Visual Computing (ISVC 2009)
Las Vegas, NV, United States
November 30, 2009 through December 2, 2009

Disclaimer

This document was prepared as an account of work sponsored by an agency of the United States government. Neither the United States government nor Lawrence Livermore National Security, LLC, nor any of their employees makes any warranty, expressed or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States government or Lawrence Livermore National Security, LLC. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States government or Lawrence Livermore National Security, LLC, and shall not be used for advertising or product endorsement purposes.

Factorization of correspondence and camera error for unconstrained dense correspondence applications

Daniel Knoblauch[1], Mauricio Hess-Flores[2], Mark Duchaineau[3], and Falko Kuester[1]

[1] University of California, San Diego
[2] University of California, Davis
[3] Lawrence Livermore National Laboratory

Abstract. A correspondence and camera error analysis for dense correspondence applications such as structure from motion is introduced. This provides error introspection, opening up the possibility of adaptively and progressively applying more expensive correspondence and camera parameter estimation methods to reduce these errors. The presented algorithm evaluates the given correspondences and camera parameters based on an error generated through simple triangulation. This triangulation is based on the given dense, non-epipolar constraint, correspondences and estimated camera parameters. This provides an error map without requiring any information about the perfect solution or making assumptions about the scene. The resulting error is a combination of correspondence and camera parameter errors. An simple, fast low/high pass filter error factorization is introduced, allowing for the separation of correspondence error and camera error. Further analysis of the resulting error maps is applied to allow efficient iterative improvement of correspondences and cameras.

1 Introduction

The main challenges in tracking, structure from motion and other applications that make use of dense correspondences are attributable to faulty correspondences and the estimated camera parameters. These challenges result from different lighting conditions, occlusions, and moving objects within the scene, which introduce uncertainty to the correspondence algorithm. This makes it desirable to be able to iteratively improve these correspondences based on an error metric. To the knowledge of the authors there has been no work evaluating correspondences and camera pose without knowledge of the ground truth. This paper introduces a novel, simple error evaluation based on the triangulation error, without ground truth knowledge.

The usual approach for epipolar-constrained applications is based on the following steps: (1) Find a small number of reliable correspondences between the two images. (2) Estimate camera poses with calculated correspondences. (3) Calculate dense correspondences and scene structure with the help of epipolar constraints.

This project does it by calculating general dense, non epipolar-constrained correspondences and the camera pose estimation from a subset of these correspondences. Based on these two steps, this paper introduces an error metric based on a 3D geometric error. The main contribution of this paper is the factorization of the error into the two main error sources, the camera parameter error and the correspondence error. This error metric

opens up the possibility of automatically performing feedback on both correspondence and camera parameter calculation given a general, non epipolar-constrained, dense correspondence algorithm. While this is not the paper that introduces such a feedback loop, it lays the fundamentals for it.
Further analysis of the extracted errors is performed to allow a quantitative error evaluation. This analysis allows a more systematic decision in which previous steps, correspondence calculation or camera parameter estimation, need further exploration.
The approach presented in this project was chosen because of improvements in hierachical dense correspondence algorithms, which allows efficient general correspondence calculation without knowledge of epipolar geometry. The main reason for this approach is the potential to incrementally improve the correspondences and camera poses by the proposed feedback loop. This is possible because correspondence errors for non-epipolar constraint dense correspondences are independent of the epipolar mapping. The freedom in unconstrained correspondence calculation allows the introduction of geometrical error extraction. By the knowledge of the authors this is the first time that an error factorization for correspondence and camera error in dense correspondence algorithms is possible.

2 Previous Work

Image registration, establishing the correspondence between two images, is a major step towards extracting model geometry. There are different approaches to evaluate pixels in two images representing the same object. Harris and Stephens [1] introduced a motion analysis algorithm based on corners and edges. This approach is only suitable for image motion analysis where objects of interest have to be tracked. Other approaches base the correspondence search on epipolar constraints as shown in [2]. To exploit the epipolar constraints the camera poses have to be known in advance or have to be calculated with a subset of reliable correspondences. The algorithm used in this project is based on dense, non-epipolar constraint correspondences. Due to this constraint a direct method solving correspondences coarse-to-fine on 4-8 mesh image pyramids, with a 5x5 local affine motion model as outlined by Duchaineau et al. [3] has been introduced. The algorithm guarantees that every destination pixel is used only once and, if possible, every pixel gets a correspondence pixel in the destination frame. All the correspondences are calculated without any knowlegde of the camera pose or epipolar constraints. This leads to a very flexible but still reliable correspondence calculation, which fits our newly proposed iterative correspondence calculation.
The camera pose estimation from two corresponding images has been extensively studied in Computer Vision. Hartley [4] introduced the eight-point algorithm that requires at least eight correspondences to evaluate the relative poses of the cameras. Nistér introduced a five-point algorithm in [5]. According to the literature, this algorithm is considered to be more robust than the eight-point algorithm. The five-point algorithm, embeded in RANSAC [6] is used in this project.
The fundamental question is how to quantify the quality of correspondences and calculated camera pose. There has been a vast amount of work in the stereo vision community in error and quality analysis for correspondence and camera pose estimation algorithms.

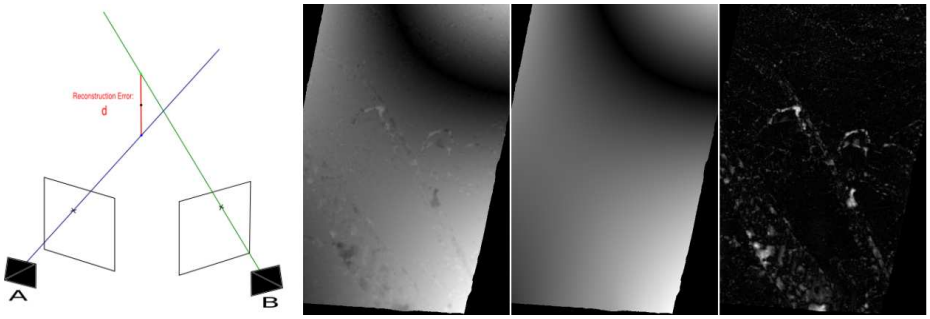

Fig. 1. Triangulation based on camera pose and correspondences. Error metric is defined by d. Resulting Error Maps: Total Error (left), Camera Error (middle), Correspondence Error (right). All error maps are normalized.

Rodehorst et al. [7] introduced an approach to evaluate camera pose estimation based on ground truth data. For correspondence evaluation Seitz et al. [8] introduced a comparison and evaluation platform for reconstructions from stereo, termed the Middlebury Stereo Evaluation. This approach is based on reconstructing scenes that are known exactly and comparing the reconstructions against the ground truth data. Mayoral et al. [9] introduced an approach to evaluate the best matching algorithm by introducing a disparity space image based on matching errors. Xiong and Matthies [10] analyse and correct major error sources, based on matching errors, for a certain scene type, in this case a cross country navigation of an autonomous vehicle. All these approaches are based on matching errors in epipolar-constrained correspondence algorithms. To the knowledge of the authors there is no work covering error analysis for correspondences and camera pose at the same time. This paper on the other hand presents a novel technique based on non-epipolar constrained correspondences and a geometric error extraction to evaluate correspondence and camera errors on the fly, without the prerequisite of ground truth data or assumptions about the scene, and lays the fundamentals for an iterative correspondence and camera pose improvement.

3 Factorization of Correspondence and Camera Pose Errors

The error factorization is based on two preceding steps not further covered in this paper, the general dense correspondence calculation and the camera pose estimation. The main achievement of this paper is the introduction of a measure for correspondence and camera quality without knowledge of the perfect solution, any information, or assumption of the scene. This opens the way for automatic iterative correspondence and camera pose calculation.

The error metric is based on triangulation. The basic idea is to intersect rays coming from both cameras which go through corresponding pixels. In order to calculate the direction of these rays we have to take the extrinsic and intrinsic parameters of the cameras into consideration. The camera pose is defined by R, the camera rotation, and T,

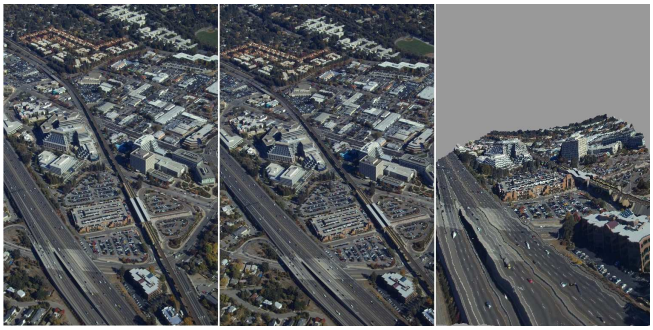

Fig. 2. Aerial source images (left, middle) and the resulting triangulation from a different view point(right).

the camera translation, which are given through camera pose estimation. The intrinsic parameters K are given through a one time calibration of the cameras. With these parameters and the correspondences the ray directions D_A and D_B can be calculated; x and y are defined as the pixel coordinates in the base image or the corresponding pixel in the destination image.

$$D_i = R_i * K_i^{-1} * \begin{pmatrix} x \\ y \\ 1 \end{pmatrix} \tag{1}$$

Having the directions of the rays calculated and knowing the start points T_A and T_B, which are based on camera pose estimation, the shortest distance between the two rays can be calculated. The points P_A and P_B on the rays that correspond to the nearest distance points on the rays are defined in the following equations, where t_A and t_B define how far in the given direction the points P_A and P_B are from the camera locations T_A and T_B.

$$P_A = T_A + t_A * D_A \tag{2}$$

$$P_B = T_B + t_B * D_B \tag{3}$$

Knowing P_A and P_B the length of the shortest distance between the two rays is defined by d. In general these rays will not intersect because of noise and errors in the camera pose and the correspondences. Based on this knowledge the length of the nearest distance between the two rays is introduced as the error metric d. Figure 1 illustrates the error calculation. The error is considered to be directional for further calculations. In order to get a direction, the cross product of the two direction rays is evaluated and the resulting vector direction is considered to be the positive direction.

This error calculation is performed for every given correspondence pair across the pair of images. Figure 1 shows the resulting error map on the left. This map consists of a smooth global error superimposed by high frequency errors. Knowing that the main error sources are the camera parameters and the correspondences it can be said that the

camera parameters would have to introduce a smooth overall error and the correspondences have a local high frequency error. A closer look at error d results in equations (4) and (5) respectively for P_A and P_B. For this analysis the calculations are done in the coordinate system of camera A. This means that the rotation and translation between camera A and B are relative.

$$P_A = t_A K_{E_A}^{-1} \begin{pmatrix} x + x_{E_d} \\ y + y_{E_d} \\ 1 \end{pmatrix} \tag{4}$$

$$P_B = R_E t_B K_{E_B}^{-1} \begin{pmatrix} x_c + x_{E_d} + x_{E_c} \\ y_c + y_{E_d} + y_{E_c} \\ 1 \end{pmatrix} + T_E \tag{5}$$

Camera parameter errors consist of error in relative rotation R_E, relative translation T_E, the intrinsic errors K_{E_A} and K_{E_B}, and the radial distortion errors x_{E_d} and y_{E_d}. The errors introduced by the inaccuracies in correspondence calculation are represented by x_{E_c} and y_{E_c}. All errors introduced by the camera parameters are global and influence the reconstruction in a smooth manner, resulting in the smooth parts of the total error map. The correspondence errors on the other side are local and therefore result in high frequency errors in the error map. To separate the two error sources from each other the camera error is first estimated. The error map shown in Figure 1 shows the absolute values of the error. The direction of the error is taken into account as it is possible that the crossing rays change their spatial order. This can be seen in the upper right corner of the total error map in Figure 1. The black circle corresponds to an area where the sign of the error changes.

To extract the camera error, a least-square B-spline approximation to the total error height field is introduced. This approximation consists of a 5x5 support point grid and is a special case called Bézier Curve. The goal is to filter the smooth camera error out. The correspondence error is defined as the difference between the total error and the camera error. The resulting smooth camera error can be seen in Figure 1 in the middle. The correspondence error can now be calculated by subtracting the camera error from the total error in each pixel. Figure 1 shows the correspondence error (on the right) based on the image pair in Figure 2. The latter figure shows the triangluation based on the given correspondences and the calculated camera pose. In areas of high error in the correspondence error map the resulting reconstruction shows artifacts. Also the camera error map represents the smooth reconstruction displacements seen in the upper right corner of the reconstruction. This error is produced by internal and external errors of the camera.

4 Error Metric Analysis

4.1 Signal-to-noise ratio (SNR) analysis

As discussed in the previous section, the camera error is modeled as a deterministic function. Correspondence error, on the other hand, appears to be more like ‘noise’, due to its high-frequency and non-deterministic nature. Thus, a convenient way to examine

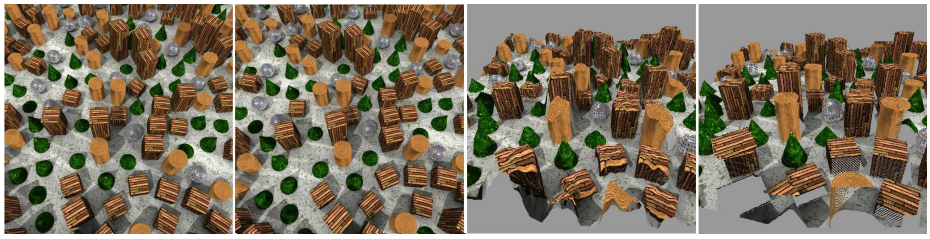

Fig. 3. Source images (left pair), 3D reconstruction with calculated camera pose and correspondences (middle) / perfect camera pose and perfect correspondences (right).

their relationship is by applying the signal-to-noise ratio (SNR) concept, which is commonly used in image and signal processing. It must be mentioned that the two errors are assumed to be independent, since only an inlier subset of the correspondences chosen by RANSAC are used to compute the camera pose, which are not necessarily representative of the entire set of correspondences. It is also important to take into account that correspondence error is a signal in itself, despite its treatment as noise here for our purposes. Thus, a range of low SNR values, uncommon in normal signal and processing applications, is permissive here, whenever the influence of camera error is less than that of the correspondence error. Our formulation fo the SNR is given by eq. 6, where μ_s and μ_n are respectively the average camera and correspondence errors, while σ_n is the standard deviation of the correspondence error.

$$SNR = \frac{\mu_s - \mu_n}{\sigma_n} \tag{6}$$

A high SNR indicates numerically that the camera error is dominant, and that some algorithm should be applied to overcome this deficiency. On the other hand a SNR smaller than one suggests that the correspondences are the main error source and that the main focus should be on global or local correspondence improvement.

5 Results

In this section the results of the presented approach are discussed and evaluated by using different data sets to show the flexibility of the approach. These tests have been conducted on a machine with Quad Core CPU @2.66 Mhz and 4 GB of RAM. All results were achieved in a few seconds depending on the size of the input images.

5.1 Aerial Imagery

The first data set consists of aerial images taken from different viewing angles of a downtown district. Figure 2 shows the image pair and resulting reconstruction. Figure 1 displays the error maps resulting from the introduced approach. It can be concluded that the largest correspondence errors appear in occlusion areas and in areas where

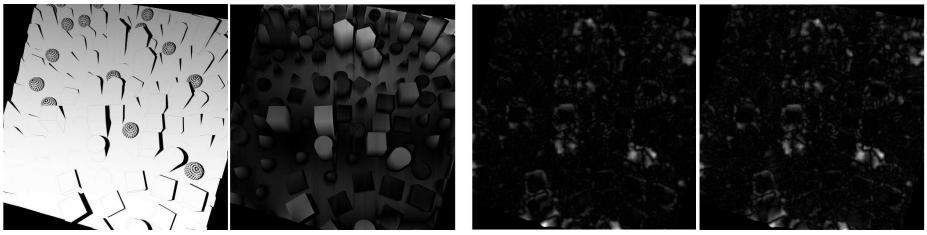

Fig. 4. Left pair: Calculated camera pose/perfect correspondences: camera error (left)/correspondence error (right). Right pair: Perfect camera pose/calculated correspondences: total error (left)/correspondence error (right).

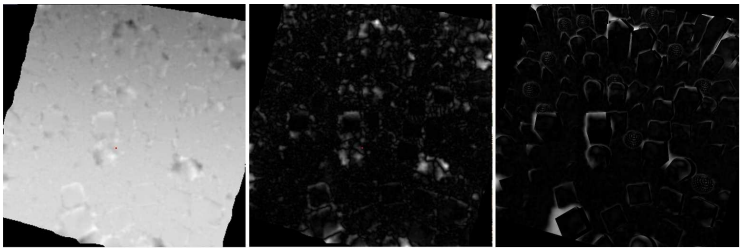

Fig. 5. Estimated camera pose and calculated correspondences. Total error map (left), correspondence error map (middle) and ground truth correspondence error map (right).

there is not enough texture for the correspondence algorithm to lock down the best correspondences. There are also high errors on the reconstruction of the static scene, such as the streets, where movers appear. The problem is that these objects move from one frame to the other and therefore the correspondences are incorrect. These results demonstrate that problem areas are found by the introduced correspondence error map.

5.2 Artificial Data

Further tests have been completed with artificial data, where the perfect camera positions and correspondences are known. Figure 3 shows the used camera views and the resulting triangulation with calculated and perfect correspondences. The goal of this test is to prove that the assumption of the smooth camera error is correct and that the extraction of correspondence errors results in a reliable error map. The algorithm was run with the perfect camera pose and the calculated correspondences. The resulting correspondence error map can be seen in Figure 4 on the right. This error map illustrates that the main errors in the correspondences are around occlusions and repetitive textures on the cylinders. Considering that we have perfect camera poses the camera error is very small overall, which is supported by the resulting SNR of -0.77. This explains the similarity of the total error map and the correspondence error map. Figure 5 shows

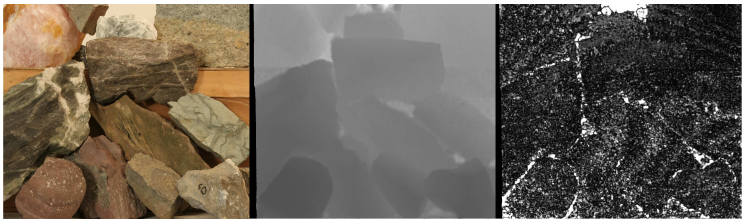

Fig. 6. Middlebury data set. Base camera view (left), depth map extracted after reconstruction (middle) and Correspondence Error Map (right).

the extracted error maps with calculated camera poses and calculated correspondences. This demonstrates that the resulting correspondence error map (middle) is up to normalization just like the one with the perfect camera poses. This shows that the assumed interaction of camera error and correspondence error is correct. In this case the SNR result is 28.31 which implies that the camera error is dominant.
The next test has been conducted to prove that in case of perfect correspondences the total error corresponds to the camera error. Figure 4 shows the results of this test (left). It can be seen that the camera error represents the entire error. The SNR of value 292.96 implies that all the error is in the camera and correspondence errors are negligable. This fulfills the assumed error relation.
A ground truth correspondence error map is introduced to support the extracted correspondence error map. The ground truth correspondence error is given by the distance of the two 3D points based on the perfect camera pose and the perfect correspondences or calculated correspondences respectively. In areas of occlusions no ground truth data can be produced as no perfect correspondences exist. Figure 5 shows that the error areas in the ground truth (right) and the calculated correspondence error (middle) maps are similar up to scale. By taking into account that the error estimation is done without knowledge about the scene it can be said that the results are conclusive.
The introduced approach has additionally been tested with the 'Rocks2' data set from the Middlebury Stereo Evaluation data sets [11]. The resulting correspondence error map can be seen in Figure 6. The comparison of extracted depth map, input image and the correpondence error map shows that problem areas for the correspondences are detected.

5.3 Signal-to-noise ratio

The SNR based on the error maps gives information about the relative importance relation between camera and correspondence errors. To underline the benefit of this analysis, tests with the artifical data were conducted. The results in table 1 show the changes in SNR, based on different camera poses and correspondences. The correspondences are reduced in quality from left to right. In the calculated correspondences the number of iterations the algorithm runs are restricted to get different correspondence qualities. For testing purposes the camera poses are optimized based on bundle adjustment (BA) [12].

SNR	Perfect	Calculated (5x)	Calculated (1x)
Perfect	6.69695	-0.789583	-0.770614
Calculated (BA)	278.659	28.5515	27.3323
Calculated	292.969	30.4882	28.3113

Table 1. SNR values based on different camera poses (vertical) and different correspondences (horizontal).

It can be seen that the better the camera poses get, the lower the SNR is. The worse the correspondences are, the lower the SNR is too. The SNR gives us a measure to estimate which error sources are relatively more dominant. A SNR of approximately one implies that both errors have about the same influence. The SNR value for perfect camera pose and perfect correspondences results because the intrinsic parameters in this data set are not perfect, which shows in the camera error map. These tests were run with all the used data sets and the results comply.
At this point a simple feedback loop can be introduced. An initial camera pose based on dense correspondence calculation is calculated and the resulting SNR is 28.31. This implies that the camera pose is the dominant error source. By using BA the extrinsic parameters can be improved, which shows in the smaller SNR value 27.33. Despite the correction the camera error is still dominant, which implies that most of the remaining camera error is based on internal camera parameters and distortion. To improve this, further parameters could be added to the camera refinement. On the other side a simple global correspondence improvement would be to run the correspondence algorithm with more iterations. This leads to a higher SNR, which implies that the correspondences get globally relatively better. A more advanced way to improve correspondences would be to use the correspondence error map to locally improve bad correspondences with more expensive fitting algorithms. This discussion shows the proof of concept for an iterative correspondence and camera pose estimation algorithm, based on error analysis and separation, though this is out of the scope of this particular paper.

5.4 Zero crossings

A closer look at the correspondence error map reveals thin lines of ‘no-error’ inbetween high error regions. The same regions in the total error map reveal that this occurs where total error changes from being smaller to being bigger than the estimated camera error. This is a result of the taken assumptions as we try to calculate the error without comparison to the perfect solution. This artifact is acceptable as it is only a small part of the error and only introduces false positives. Used in an iterative correspondence calculation algorithm these areas will show up as errors in the next iterations.

6 Conclusion

This paper introduces an automated correspondence and camera error metric based on triangulation for general, non-epipolar constrained, dense correspondence applications.

The goal of this error metric is to find faulty camera poses and correspondences and lay foundation for feedback to allow updates with more sophisticated and expensive algorithms. To solve for this error metric a triangulation based on correspondences and camera poses is executed. The length of the nearest distance between the resulting triangulation rays is used as the error for this approach. Based on the assumption that the error introduced by the camera is smooth over an image the camera error can be extracted with a least squares B-spline approximation of the total error. The correspondence error, which is considered to be local and represented by high frequency errors is the difference between camera error and total error. Further SNR analysis of the errors reveals if correspondence or camera parameter errors are dominant and helps to iteratively improve the weaker link. An overall test demonstrates the usefulness of the SNR value towards identifying which source of error is relatively dominant, and in case it is the correspondence error the problem areas identified are consistent with ground-truth error maps, such that posterior local corrections can be applied in only these regions.

References

1. Harris, C., Stephens, M.: A combined corner and edge detector. In: Alvey Vision Conference. Volume 15. (1988) 50 2
2. Pollefeys, M., Van Gool, L., Vergauwen, M., Verbiest, F., Cornelis, K., Tops, J., Koch, R.: Visual Modeling with a Hand-Held Camera. International Journal of Computer Vision **59** (2004) 207–232 2
3. Duchaineau, M., Cohen, J., Vaidya, S.: Toward Fast Computation of Dense Image Correspondence on the GPU. In: Proceedings of HPEC 2007, High Performance Embedded Computing, Eleventh Annual Workshop, Lincoln Laboratory, Massachusetts Institute of Technology. (2007) 91–92 2
4. Hartley, R.: In defense of the eight-point algorithm. Pattern Analysis and Machine Intelligence, IEEE Transactions on **19** (1997) 580–593 2
5. Nistér, D.: An Efficient Solution to the Five-Point Relative Pose Problem. IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE (2004) 756–777 2
6. Fischler, M., Bolles, R.: Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. Communications of the ACM **24** (1981) 381–395 2
7. Rodehorst, V., Heinrichs, M., Hellwich, O.: Evaluation of relative pose estimation methods for multi-camera setups. In: ISPRS08. (2008) B3b: 135 ff 3
8. Seitz, S., Curless, B., Diebel, J., Scharstein, D., Szeliski, R.: A comparison and evaluation of multi-view stereo reconstruction algorithms. In: Int. Conf. on Computer Vision and Pattern Recognition. (2006) 519–528 3
9. Mayoral, R., Lera, G., Perez Ilzarbe, M.: Evaluation of correspondence errors for stereo. IVC **24** (2006) 1288–1300 3
10. Xiong, Y., Matthies, L.: Error analysis of a real time stereo system. In: CVPR97. (1997) 1087–1093 3
11. Hirschmuller, H., Scharstein, D.: Evaluation of cost functions for stereo matching. In: IEEE CVPR. (2007) 1–8 8
12. Triggs, B., McLauchlan, P., Hartley, R., Fitzgibbon, A.: Bundle adjustment-a modern synthesis. Lecture Notes in Computer Science, vol. 1883 (1999) 298–372 8

This work performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under Contract DE-AC52-07NA27344.