

Bundle Adjustment for Stereoscopic 3D

Christian Kurz, Thorsten Thormählen, and Hans-Peter Seidel

Max Planck Institute for Computer Science (MPII)
Campus E1 4, 66123 Saarbrücken, Germany
ckurz@mpi-inf.mpg.de

Abstract. The recent resurgence of stereoscopic 3D films has triggered a high demand for post-processing tools for stereoscopic image sequences. Camera motion estimation, also known as structure-from-motion (SfM) or match-moving, is an essential step in the post-processing pipeline. In order to ensure a high accuracy of the estimated camera parameters, a bundle adjustment algorithm should be employed. We present a new stereo camera model for bundle adjustment. It is designed to be applicable to a wide range of cameras employed in today’s movie productions. In addition, we describe how the model can be integrated efficiently into the sparse bundle adjustment framework, enabling the processing of stereoscopic image sequences with traditional efficiency and improved accuracy. Our camera model is validated by synthetic experiments, on rendered sequences, and on a variety of real-world video sequences.

1 Introduction

In computer vision, stereo image sequences have been employed for a large number of applications over the past decades. However, the largest body of work can be found on robot or autonomous vehicle navigation and motion estimation. Implicated by this predominant area of application, stereo processing pipelines usually have to face restrictive real-time requirements. Furthermore, there are limits on the amount of data the algorithms are allowed to accumulate and process. These requirements influence the types of algorithms employed.

Recently, however, the revival of 3D films using modern stereo 3D (S3D) technology has entailed the creation of an unprecedented amount of high-resolution stereo image data. Today’s movies are often augmented with virtual objects, and sometimes even the major part of the movie is computer generated. In order to composite the virtual objects with a real image sequence, the camera parameters of the real camera have to be estimated to render the virtual object with the corresponding virtual camera. Thus, reliable and accurate camera motion estimation for S3D sequences is a crucial part in movie post-processing and essential for the creation of convincing special effects. Given the amount of computation involved, post-processing is inherently done off-line and does not shy away from computationally expensive algorithms.

Considering the increase in demand, some commercially available match-moving packages already incorporate solvers for stereo cameras. However, the

employed algorithms are not published and an academic paper presenting a solution to high quality camera motion estimation for stereo cameras is (to the best of our knowledge) not yet available.

We present an approach allowing reliable and accurate camera motion estimation for stereo sequences. In contrast to existing real-time approaches, we employ a large number of automatically extracted feature points and optimize the camera parameters with the gold-standard method: bundle adjustment. As known from literature, the naïve implementation of bundle adjustment is computationally expensive beyond feasibility and can be sped up by employing the sparse matrix structure of the Jacobian. The contributions of this paper are:

- An extended camera model for stereo cameras is presented. The model offers great flexibility in terms of its parameters and therefore can be employed for a variety of different cameras, ranging from entry-level consumer 3D camcorders using a 3D conversion lens with a static camera geometry to professional cameras used in movie productions.
- It is shown how the additional constraints introduced by the camera model can be incorporated into the sparse bundle adjustment framework.

The approach is validated on a variety of data sets, from fully synthetic experiments to challenging real-world image sequences.

This paper is organized as follows: Related work will be reviewed in the next section, followed by a brief summary of camera motion estimation in Sec. 3. Sec. 4 introduces our new camera model for stereoscopic bundle adjustment, and the incorporation into bundle adjustment is described in Sec. 5. The results of our new approach are shown in Sec. 6, followed by the conclusion.

2 Related Work

Structure-from-Motion A general introduction to bundle adjustment can be found in [1, 2]. Of late, research has been done towards processing data of multiple independently moving cameras [3], or entire community photo collections [4], demonstrating orthogonal approaches. Multi-camera systems either assume a static and calibrated camera setup on a moving platform [5, 6] or obtain the calibration by averaging parameters of the independent reconstructions [7]. There exist alternative approaches to SfM, but either the stereo rig is assumed to be calibrated and no bundle adjustment is used [8], or the bundle adjustment remains unaffected by the changes to the reconstruction pipeline [9]. To a certain extent, constraints arising from stereo geometry have been included in bundle adjustment [10], but the model is incorporated into the algorithm by simply adding soft constraints and without addressing the sparse structure of the problem.

Self-Calibration The problem of self-calibration for an uncalibrated stereo rig with an unknown motion has been explicitly modelled for two pairs of stereo images [11], even with varying vergence angles [12], but the focus of these papers is rather on obtaining a one-time calibration of these two stereo pairs instead of the optimization over a complete image sequence.

Stereo Navigation, Ego-motion Estimation, Visual Odometry Stereo rigs used in robot or autonomous vehicle navigation and motion estimation are usually assumed to be calibrated. Due to runtime constraints, the problem of motion estimation is often reduced to estimating the parameters of an inter-frame motion model given two distinct sets of 3D points, and then feeding the results to a Kalman filter to achieve robustness (see [13–16], for example). Optimized feature selection and tracking, especially *multi-frame tracking*, is used in [17] to achieve robustness for tracking features over longer sequences. There are attempts at using bundle adjustment in visual odometry, thereby incorporating the data produced by a calibrated stereo rig directly [18–20], but, in contrast to these approaches, we do not assume the calibration of the stereo rig to be known. A *reduced order bundle adjustment* is used in [21], but the processing and parametrization of the input data are again tailored to meet the real-time requirements of the system. In [22], a correlation-based approach to ego-motion and scene structure estimation from stereo sequences is presented. The approach is different from bundle adjustment and the transformation between left and right frames is assumed to be constant.

Uncalibrated Stereo Various approaches exist to obtain the epipolar geometry of an uncalibrated stereo rig [23–27], but these methods only consider a single pair of images and there is no further optimization. Visual servoing [28–30] and man-machine interaction [31] sometimes rely upon uncalibrated stereo cameras, but the cameras are static and the algorithms avoid explicit 3D reconstructions. For a moving stereo rig, restrictive assumptions on the scene structure have to be made [32]. Quasi-Euclidean epipolar rectification [33] has recently been adapted to work on uncalibrated stereo sequences [34], even with non-linear optimization [35], but the scene representation differs from bundle adjustment.

Optical Flow, Three-Dimensional Scene Flow While camera setups in optical flow applications frequently employ two [36,37] or more cameras [38], research in this area is more geared towards recovering the non-rigid scene motion [39], whereupon the cameras are assumed to be calibrated. Optical flow can be adapted for ego-motion estimation [40], but the method uses rectified input images and makes restrictive assumptions on the scene structure.

Commercial Products Several commercial products feature tools for stereoscopic tracking and stereo solving (PFTrackTM and SynthEyesTM, for example), but the corresponding algorithms have not been published.

3 Structure-from-Motion

Given a sequence of K images I_k , SfM refers to the procedure of deriving a camera matrix \mathbf{A}_k for every image (representing the camera motion), and a set of J 3D object points $\mathbf{P}_j = (P_x, P_y, P_z, 1)^\top$ (representing the static scene structure). The 2D feature point corresponding to \mathbf{P}_j in image I_k is denoted by $\mathbf{p}_{j,k}$.

Traditionally, the SfM pipeline consists of several steps. At first, the 2D feature points are detected and tracked, and outliers are eliminated using geometric constraints (e.g., the fundamental matrix). In the next step, initial camera pa-

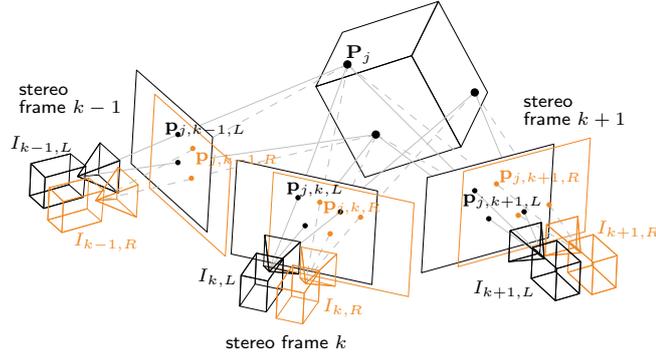


Fig. 1: Each stereo frame consists of a left camera image $I_{k,L}$ and a right camera image $I_{k,R}$. In contrast to monocular SfM, there are now two sets of corresponding 2D feature points $\mathbf{p}_{j,k,L}$ and $\mathbf{p}_{j,k,R}$ for the set of 3D object points \mathbf{P}_j .

rameters and 3D object points are established. To obtain initial values for the intrinsic camera parameters, self-calibration is performed. These steps are not described in this paper; details can be found in the literature [1]. As last step, bundle adjustment is employed, which will be discussed in the following.

The goal of bundle adjustment is to minimize the reprojection error given by the cost function

$$\arg \min_{\mathbf{A}, \mathbf{P}} \sum_{j=1}^J \sum_{k=1}^K d(\mathbf{p}_{j,k}, \mathbf{A}_k \mathbf{P}_j)^2, \quad (1)$$

where $d(\dots)$ denotes the Euclidean distance. Thereby, the error is equally distributed over the whole scene. For numerical optimization of Eq. (1), the sparse Levenberg-Marquardt (LM) algorithm is typically employed [1].

In the case of a stereo camera setup, the input consists of K stereo frames. For convenience, the individual images are now denoted as $I_{k,L}$ for the image of the left camera, and $I_{k,R}$ for the image of the right camera. Analogical, we get separate projection matrices $\mathbf{A}_{k,L}$ and $\mathbf{A}_{k,R}$, and we have to distinguish between 2D feature points $\mathbf{p}_{j,k,L}$ and $\mathbf{p}_{j,k,R}$, respectively (see Fig.1).

Introducing $x \in \{L, R\}$, the cost function from Eq. (1) translates to

$$\arg \min_{\mathbf{A}, \mathbf{P}} \sum_{j=1}^J \sum_{k=1}^K \sum_x d(\mathbf{p}_{j,k,x}, \mathbf{A}_{k,x} \mathbf{P}_j)^2. \quad (2)$$

4 Camera Model

In this section, we first describe the camera model for our stereo bundle adjustment for a metric camera. Bundle adjustment for monocular sequences is often also performed with a projective camera model [1]. However, the representation

of the geometric constraints between the left and the right camera is not possible in the projective framework, because transformations in the local camera coordinate system including rotations and translations cannot be parametrized independently from the current projective camera matrix. Thus, we propose to enforce the constraints introduced by our metric stereo camera model after an update from projective to metric space has been performed (cp. [1, 41]).

The 3×4 projection matrix \mathbf{A} of a metric camera can be decomposed as

$$\mathbf{A} = \mathbf{K} [\mathbf{I} | \mathbf{0}] \begin{bmatrix} \mathbf{R} & -\mathbf{R}\mathbf{C} \\ \mathbf{0} & 1 \end{bmatrix}, \quad (3)$$

where \mathbf{C} is the position of the camera center in world coordinate frame, \mathbf{R} is a rotation matrix representing the camera orientation, and \mathbf{K} is a calibration matrix comprising the intrinsic camera parameters, such as focal length. The index k assigning a projection matrix to the corresponding image is omitted throughout this chapter for the sake of readability.

Considering a standard stereo camera setup as employed in movie productions, our first observation is that the two cameras of the stereo system undergo only dependent motion – if the left camera translates to the right, the right camera will inherently have to follow that same translation. Now, in order to improve over the conventional bundle adjustment algorithm, we exploit this dependency: Instead of treating the left and the right camera as separate entities, we consider them as instances of the same camera system. A change of parameters introduced by the left camera will therefore influence the position and orientation of the right camera, and vice versa.

Secondly, to benefit from the combined camera model, the total number of parameters representing the camera over the whole image sequence has to be reduced. Since modern stereo camera systems allow the point of convergence of the two cameras to change during acquisition, the relative rotation between the cameras can not always be assumed to be constant over the sequence. Therefore, this constraint, which would reduce the number of parameters significantly, is only optionally enforced (however, all our results enforce this constraint).

Assuming the relative position offset of the two camera centers to be unknown but constant is a constraint we always enforce, because the baseline between the cameras is usually not changed. As a matter of principle, there is some freedom in the choice of the stereo system base position. We chose it to coincide with the center of the left camera. The result are two different decompositions for the left and the right camera that can be expressed as

$$\mathbf{A}_L = \mathbf{K}_L \quad [\mathbf{R}_L | \mathbf{0}] \quad \begin{bmatrix} \mathbf{R} & -\mathbf{R}\mathbf{C} \\ \mathbf{0} & 1 \end{bmatrix}, \quad (4)$$

$$\mathbf{A}_R = \mathbf{K}_R [\mathbf{R}_R | -\mathbf{R}_R\mathbf{C}_R] \begin{bmatrix} \mathbf{R} & -\mathbf{R}\mathbf{C} \\ \mathbf{0} & 1 \end{bmatrix}, \quad (5)$$

where subscripts L and R denote parameters that are exclusive to the left and right camera respectively.

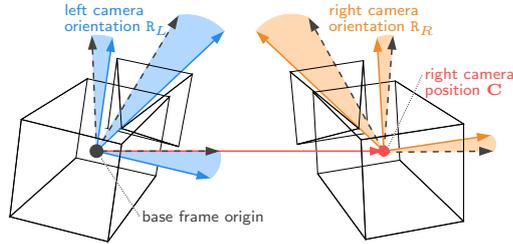


Fig. 2: Our novel camera model for bundle adjustment. The camera geometry of every stereo frame is given by a base frame (dashed lines), whose origin is aligned with the center of the left camera. The orientation \mathbf{R}_L of the left camera is encoded independent from the orientation of the base frame, allowing the position of the right camera to be specified by a single parameter \mathbf{C} (red arrow) for the whole sequence.

The rotation matrix of the left camera \mathbf{R}_L could be omitted for a static stereo setup. However, if the point of camera convergence changes in a dynamic setup, it is necessary to encode the orientation of the left camera separately from the orientation of the stereo system. This is due to the fact that a rotation of the left camera would otherwise inherently lead to a rotation of the coordinate frame in which the relative translation of the right camera takes place (see Fig. 2).

Depending on the actual acquisition system in operation, parameters can be chosen to be estimated for every frame, for a subset of frames, or for the whole sequence. Furthermore, the intrinsic camera parameters can of course be treated as shared between the two cameras, if this was the case at the time of recording.

5 Bundle Adjustment

To optimize Eq. (2), we extend the sparse LM algorithm [1].

First, we assemble a parameter vector $\mathbf{q} = (\mathbf{b}^\top, \mathbf{c}^\top, \mathbf{d}^\top, \mathbf{e}^\top, \mathbf{f}^\top, \mathbf{g}^\top)^\top$. The designation of the corresponding subvector for all parameters of our camera model can be found in Tab. 1, along with a listing of the number of parameters and the number of the respective vector entries.

Most parameters can either be assumed to be variable for each frame or joined (i.e., estimated conjointly) over the whole sequence. The intrinsic parameters can also be shared for both cameras.

It is also possible to restrict \mathbf{R}_L and \mathbf{R}_R in a way that makes them depend on the vergence angle only. Dependent on the degrees of freedom for the convergence point, this results in 1 or 2 degrees of freedom for the rotation matrices \mathbf{R}_L and \mathbf{R}_R (cp. Tab.1).

For the sake of simplicity, we will assume a static stereo setup with joined and shared intrinsic parameters henceforth, resulting in two single rotation matrices \mathbf{R}_L and \mathbf{R}_R over the whole sequence, and a single calibration matrix \mathbf{K} . This would be the case in a stereo setup with a fixed convergence point, e.g., a camcorder with a 3D conversion lens.

Model parameters	# of parameters	# of vector elements	designation
base frame \mathbf{C}, \mathbf{R}	6	K	\mathbf{b}
left orientation \mathbf{R}_L	1-3	$K, 1$ (joined)	\mathbf{c}
right position \mathbf{C}_R	3	1	\mathbf{d}
right orientation \mathbf{R}_R	1-3	$K, 1$ (joined), 0 (shared)	\mathbf{e}
left intrinsics \mathbf{K}_L	3	$K, 1$ (joined)	\mathbf{f}
right intrinsics \mathbf{K}_R	3	$K, 1$ (joined), 0 (shared)	\mathbf{f}
3D object points \mathbf{P}_j	3	J	\mathbf{g}

Table 1: Stereo model parameters with their typical parameter count, the number of elements in the associated vector, and the designation of the corresponding vector. Example: For a sequence of $K = 10$ images, \mathbf{b} contains 10 elements with 6 parameters each, i.e., 60 entries in total. 'Joined' indicates that the parameters are constant and are jointly estimated over the whole sequence. 'Shared' indicates that the respective parameters of the right camera are estimated in combination with the corresponding parameters of the left camera, so that there are no separate entries for these parameters in the matrix $\mathbf{J}^\top \mathbf{J}$.

The least squares problem that is the core of bundle adjustment is tackled by the sparse LM algorithm that solves the linear equation system

$$\mathbf{J}\boldsymbol{\delta} = \boldsymbol{\epsilon} \quad (6)$$

with the Jacobian matrix $\mathbf{J} = \partial\mathbf{p}/\partial\mathbf{q}$, the residual vector $\boldsymbol{\epsilon}$, and the update vector $\boldsymbol{\delta}$. The Jacobian matrix \mathbf{J} has the block structure $\mathbf{J} = [\mathbf{B}\mathbf{C}\mathbf{D}\mathbf{E}\mathbf{F}\mathbf{G}]$, where $\mathbf{B} = \partial\mathbf{p}/\partial\mathbf{b}$, $\mathbf{C} = \partial\mathbf{p}/\partial\mathbf{c}$, et cetera. In the case of a conventional bundle adjustment that allows to enforce joined intrinsic parameters over the sequence, the Jacobian \mathbf{J} only comprises the matrices \mathbf{B} , \mathbf{F} , and \mathbf{G} . Depending on the parameter interdependencies, \mathbf{J} usually has a lot of zero entries (cp. Fig. 3). The measurement vector \mathbf{p} is constructed by placing all the 2D feature points from all camera images in a single column vector. For the purpose of illustration, we assume them to be sorted by their affiliation to the left or right camera, then their image index k , and finally their corresponding 3D object point index j .

The solution to Eq. (6) is obtained by multiplication with \mathbf{J}^\top , thereby directly evaluating $\mathbf{J}^\top \mathbf{J}$ and $\mathbf{J}^\top \boldsymbol{\epsilon}$, leaving the explicit construction of \mathbf{J} unnecessary.

A comparison of the structure of $\mathbf{J}^\top \mathbf{J}$ taken from our stereo bundle adjustment and from a conventional bundle adjustment can be found in Fig. 4. As becomes evident, we only introduce changes to one block in the structure, which is the top left one. Although the structure in the block is no longer sparse, this does not have any influence on the matrix inversion $(\mathbf{J}^\top \mathbf{J})^{-1}$, since other elements added on top during the sparse matrix inversion cause the sparse structure of this block to break down anyway (cp. [1]). Furthermore, the size of this block is significantly reduced due to the reduced number of parameters when using stereo bundle adjustment with constant convergence point, leading to better computational performance.

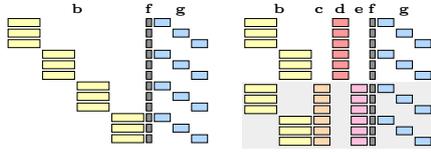


Fig. 3: Block structure of the Jacobian matrix J for a conventional bundle adjustment with joined intrinsic parameters (left), and for our stereo bundle adjustment (right). The individual block matrices are set apart by different coloring. The gray background on the right indicates derivatives contributed by the right camera.

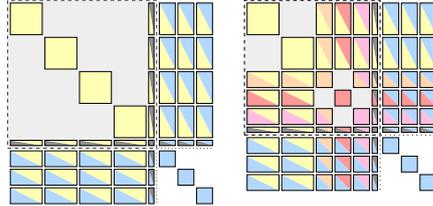


Fig. 4: Structure of the matrix $J^T J$ used in the solution of Eq. (6) for a conventional bundle adjustment with joined intrinsic parameters (left), and for our stereo bundle adjustment (right). The color indicates the contribution of the individual elements in the matrix multiplication. The dashed square indicates the relevant block for matrix inversion.

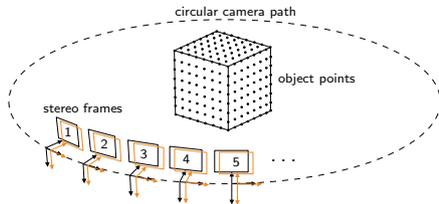


Fig. 5: The setup used in the synthetic experiments for the generation of the ground truth camera and 3D object point parameters.

RMSE	unconst.	joined	stereo
translation	1.7274 mm	0.6459 mm	0.5964 mm
rotation	0.0112 deg	0.0026 deg	0.0024 deg
focal length	1.3609 mm	0.0975 mm	0.0600 mm
Avg. time	719 ms	860 ms	733 ms

Table 2: Average translation, rotation, and focal length error, and average time per iteration for the rendered sequence for an unconstrained bundle adjustment, a bundle adjustment with joined focal length, and our stereo bundle adjustment.

6 Results

In this section we present the evaluation of our stereo bundle adjustment with purely synthetic data, rendered sequences and real-world sequences. The latter can also be found in the video accompanying this paper, which can be downloaded from http://www.mpi-inf.mpg.de/users/ckurz/papers/Kurz_MIRAGE2011.mov.

Our setup for the synthetic experiments is sketched in Fig. 5. It consists of a virtual stereo configuration composed of two cameras. The cameras execute a circular motion around a set of 296 3D object points arranged in a regular grid on the surface of a cube. The cube has an edge length of 100 mm, the radius of the camera path is 300 mm, and the opening angle of the cameras is 30 degrees.

We generate a total of 40 stereo pairs per trial, providing 80 images per sequence. All the ground truth measurements for the 2D feature points contained in these images are calculated from the known ground truth camera and 3D object points parameters. In a last step before the reconstruction process, Gaussian noise with a standard deviation σ_{syn} is applied to the measurements.

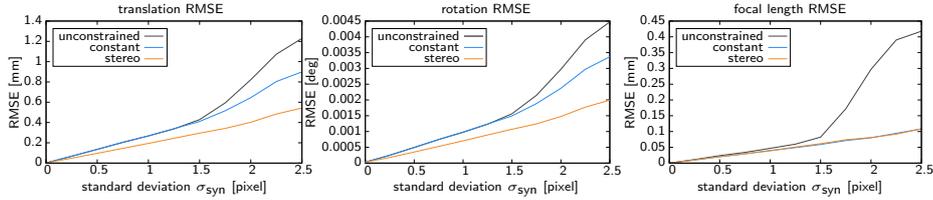


Fig. 6: Average translation, rotation, and focal length error for a given Gaussian error σ_{syn} of the 2D feature points over 1000 trials. The setup sketched in Fig. 5 was used for the generation of the ground truth parameters.

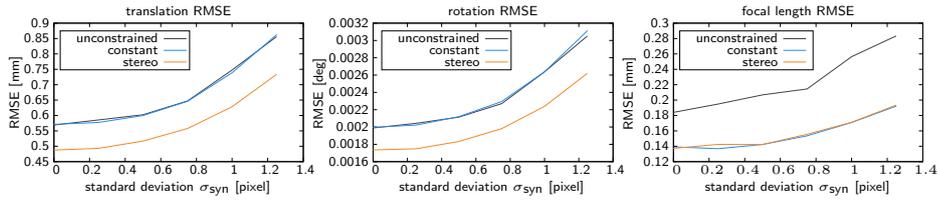


Fig. 7: Average translation, rotation and focal length error for a given Gaussian error σ_{syn} of the 2D feature points over 1000 trials, while 20 percent of the feature points were additionally disturbed by a large offset. The setup sketched in Fig. 5 was used for the generation of the ground truth parameters.

For each value of σ_{syn} , we perform a total of 1000 trials for a conventional bundle adjustment, a conventional bundle adjustment with joined focal length over the sequence, and our novel stereo bundle adjustment, whereas a different random disturbance is introduced in the measurements each time. For each reconstruction, a similarity transformation is estimated to register it to the ground truth, and then the average absolute position and orientation error is calculated. The results can be found in Fig. 6. Our stereo bundle adjustment clearly outperforms the conventional methods in terms of the translation and rotation error, while being on par with the conventional bundle adjustment with joined focal length for the error in the estimated focal length.

Furthermore, to simulate outliers, another test series was conducted. In this series, 20 percent of the measurements were disturbed by an offset of up to 12 pixel in addition to the Gaussian noise. Since not all outliers can be removed in the outlier elimination step, the results, which can be found in Fig. 7, differ. Our stereo bundle adjustment clearly outperforms both competitors again.

The second step in the evaluation was to process a rendered sequence with known ground truth parameters. Again, results were generated for a conventional bundle adjustment, a conventional bundle adjustment with joined focal length, and our stereo bundle adjustment (see Tab. 2). Our algorithm achieves the best results. In addition, Fig. 8 shows two sample stereo frames from the rendered sequence with a wireframe overlay using the estimated camera parameters. As can also be seen in the supplemental video, the wireframe fits the true scene geometry almost perfectly.



Fig. 8: This figure shows three example stereo frames from a rendered indoor sequence. The left images show the actual frames, whereas the right images the same images augmented with the wire-frame model of the scene placed using the estimated camera parameters. The results can also be found in the video accompanying this paper.

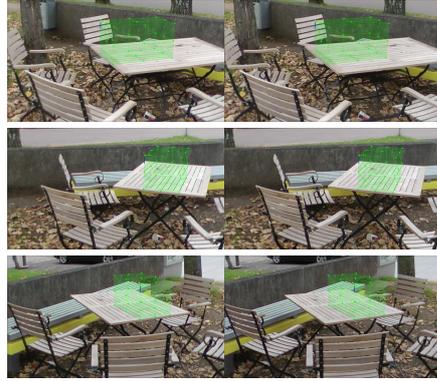


Fig. 9: Real-world sequence shot with a HD camcorder with a 3D conversion lens. The scene has been augmented by a green cuboid to demonstrate the quality of the estimated camera parameters.

The first real-world sequence (see Fig. 9) was captured with a Panasonic HDC-SDT750 camcorder with a 3D conversion lens and depicts some pieces of garden furniture. As can be seen by the overlay geometry, our stereo bundle adjustment was able to obtain excellent results for the camera parameters. The second sequence (see Fig. 10) depicts a flyover over Ehrenbreitstein Fortress in the Upper Rhine valley from the documentary *UNESCO World Heritage - Upper Middle Rhine Valley* (courtesy of *cinovent entertainment*). In the third sequence (see Fig. 11), a scene at a train station from *Grand Canyon Adventure 3D* (courtesy of *MacGillivray Freeman Films*) is shown.

7 Conclusion and Future Work

We have presented a novel camera model for stereo cameras for use in bundle adjustment. The model has the generality to accommodate a wide range of the stereo cameras used in today's movie productions, and can be incorporated efficiently into the conventional sparse bundle adjustment algorithms. A multitude of tests has been conducted, validating our model.

For future work, we will update the other stages of the SfM pipeline to make full use of the additional information provided by stereoscopic image sequences.

References

1. Hartley, R.I., Zisserman, A.: Multiple View Geometry in Computer Vision, 2nd Edition. Cambridge University Press (2003)
2. Triggs, B., McLauchlan, P.F., Hartley, R.I., Fitzgibbon, A.W.: Bundle adjustment – a modern synthesis. Lecture Notes in Computer Science **1883** (2000) 298ff

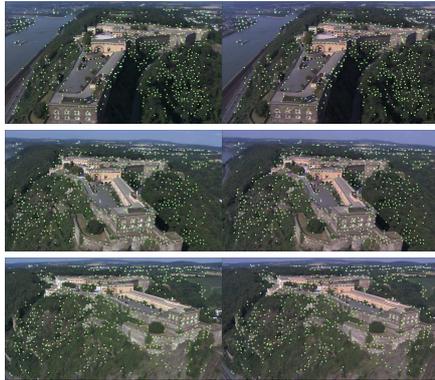


Fig. 10: Stereo frames from the Ehrenbreitstein Fortress sequence. The green dots signify the reprojections of reconstructed 3D points, showing that no drift in the parameters has occurred.



Fig. 11: Stereo frames from the train station sequence. The scene has been augmented by a yellow cuboid to demonstrate the quality of the estimated camera parameters.

3. Hasler, N., Rosenhahn, B., Thormählen, T., Wand, M., Gall, J., Seidel, H.P.: Markerless motion capture with unsynchronized moving cameras. In: CVPR. (2009)
4. Goesele, M., Snavely, N., Curless, B., Hoppe, H., Seitz, S.M.: Multi-view stereo for community photo collections. In: Intl. Conference on Computer Vision. (2007)
5. Kim, J.H., Li, H., Hartley, R.: Motion estimation for multi-camera systems using global optimization. In: Computer Vision and Pattern Recognition. (2008)
6. Stewénius, H., Åström, K.: Structure and motion problems for multiple rigidly moving cameras. In: European Conference on Computer Vision. (2004) 238ff
7. Frahm, J.M., Köser, K., Koch, R.: Pose estimation for multi-camera systems. In: DAGM Symposium, Tübingen, Germany (2004) 27–35
8. Chandraker, M., Lim, J., Kriegman, D.J.: Moving in stereo: Efficient structure and motion using lines. In: International Conference on Computer Vision. (2009)
9. Hirschmüller, H., Innocent, P.R., Garibaldi, J.M.: Fast, unconstrained camera motion estimation from stereo without tracking and robust statistics. In: International Conference on Control, Automation, Robotics and Vision. (2002)
10. Di, K., Xu, F., Li, R.: Constrained bundle adjustment of panoramic stereo images for mars landing site mapping. In: Mobile Mapping Technology. (2004)
11. Zhang, Z., Luong, Q.T., Faugeras, O.: Motion of an uncalibrated stereo rig: Self-calibration and metric reconstruction. *TRA* **12** (1996) 103–113
12. Brooks, M.J., de Agapito, L., Huynh, D.Q., Baumela, L.: Towards robust metric reconstruction via a dynamic uncalibrated stereo head. *Image and Vision Computing* **16** (1998) 989–1002
13. Matthies, L., Shafer, S.A.: Error modeling in stereo navigation. *IEEE Journal of Robotics and Automation* **3** (1987) 239–250
14. Molton, N., Brady, M.: Practical structure and motion from stereo when motion is unconstrained. *International Journal of Computer Vision* **39** (2000) 5–23
15. Saeedi, P., Lawrence, P.D., Lowe, D.G.: 3d motion tracking of a mobile robot in a natural environment. In: ICRA. (2000) 1682–1687
16. Weng, J., Cohen, P., Rebibo, N.: Motion and structure estimation from stereo image sequences. *Transactions on Robotics and Automation* **8** (1992) 362–382

17. Olson, C.F., Matthies, L.H., Schoppers, M., Maimone, M.W.: Rover navigation using stereo ego-motion. *Robotics and Autonomous Systems* **43** (2003) 215–229
18. Sünderhauf, N., Konolige, K., Lacroix, S., Protzel, P.: Visual odometry using sparse bundle adjustment on an autonomous outdoor vehicle. In: AMS. (2005)
19. Sünderhauf, N., Protzel, P.: Towards using sparse bundle adjustment for robust stereo odometry in outdoor terrain. In: TAROS. (2006) 206–213
20. Nister, D., Naroditsky, O., Bergen, J.: Visual odometry. In: CVPR. (2004)
21. Dang, T., Hoffmann, C., Stiller, C.: Continuous stereo self-calibration by camera parameter tracking. *Transactions on Image Processing* **18** (2009) 1536–1550
22. Mandelbaum, R., Salgian, G., Sawhney, H.: Correlation-based estimation of ego-motion and structure from motion and stereo. In: ICCV. Volume 1. (1999) 544–550
23. Akhloufi, M., Polotski, V., Cohen, P.: Virtual view synthesis from uncalibrated stereo cameras. In: *Multimedia Computing and Systems*. (1999) 672–677
24. Hartley, R., Gupta, R., Chang, T.: Stereo from uncalibrated cameras. In: *IEEE Conference on Computer Vision and Pattern Recognition*. (1992)
25. Ko, J.H., Park, C.J., Kim, E.S.: A new rectification scheme for uncalibrated stereo image pairs and its application to intermediate view reconstruction. *Optical Information Systems II, Proceedings of SPIE* **5557** (2004) 98–109
26. Yin, X., Xie, M.: Estimation of the fundamental matrix from uncalibrated stereo hand images for 3d hand gesture recognition. *PR* **36** (2003) 567–584
27. Zhang, Z., Xu, G.: A unified theory of uncalibrated stereo for both perspective and affine cameras. *Journal of Mathematical Imaging and Vision* **9** (1998) 213–229
28. Hodges, S., Richards, R.: Uncalibrated stereo vision for pcb drilling. In: *IEEE Colloquium on Application of Machine Vision*. (1995)
29. Park, J.S., Chung, M.J.: Path planning with uncalibrated stereo rig for image-based visual servoing under large pose discrepancy. *TRA* **19** (2003) 250–258
30. Shimizu, Y., Sato, J.: Visual navigation of uncalibrated mobile robots from uncalibrated stereo pointers. In: *Intl. Conf. on Pattern Recognition*. (2000) 346–349
31. Cipolla, R., Hadfield, P.A., Hollinghurst, N.J.: Uncalibrated stereo vision with pointing for a man-machine interface. In: MVA. (1994) 163–166
32. Simond, N., Rives, P.: Trajectory of an uncalibrated stereo rig in urban environments. In: *Intelligent Robots and Systems*. Volume 4. (2004) 3381–3386
33. Fusiello, A., Irsara, L.: Quasi-euclidean uncalibrated epipolar rectification. In: *IEEE International Conference on Pattern Recognition*. (2008)
34. Bleyer, M., Gelautz, M.: Temporally consistent disparity maps from uncalibrated stereo videos. In: *Image and Signal Processing and Analysis*. (2009)
35. Cheng, C.M., Lai, S.H., Su, S.H.: Self image rectification for uncalibrated stereo video with varying camera motions and zooming effects. In: MVA. (2009) 21–24
36. Huguet, F., Devernay, F.: A variational method for scene flow estimation from stereo sequences. In: *International Conference on Computer Vision*. (2007)
37. Min, D., Sohn, K.: Edge-preserving simultaneous joint motion-disparity estimation. In: *IEEE International Conference on Pattern Recognition*. (2006) 74–77
38. Zhang, Y., Kambhampettu, C.: On 3-D scene flow and structure recovery from multiview image sequences. *Systems, Man, and Cybernetics* **33** (2003) 592–606
39. Vedula, S., Baker, S., Rander, P., Collins, R.T., Kanade, T.: Three-dimensional scene flow. In: *International Conference on Computer Vision*. (1999) 722–729
40. Trinh, H., McAllester, D.: Structure and motion from road-driving stereo sequences. In: *3D Information Extraction for Video Analysis and Mining*. (2009)
41. Pollefeys, M., Gool, L.V., Vergauwen, M., Cornelis, K., Verbiest, F., Tops, J.: Video-to-3D. In: *ISPRS Commission V Symposium*. (2002)