Calibration of a Vehicle Camera System with Divergent Fields-of-view in an Urban Environment

ALEXANDER HANEL[1] & UWE STILLA[1]

Abstract: Multi camera systems are used in modern cars for environment perception. To combine information extracted from different cameras, their relative pose has to be known. Therefore, a multi camera system can be calibrated. In this contribution, a calibration approach using the structure-from-motion (SfM) method is described. From images of urban structures, 3d points are calculated using SfM. These 3d points are used as reference to estimate the pose of the vehicle cameras. Test images are taken in a car park with a DSLR camera, together with two action cameras mounted in a test vehicle recording the car park. More than 26,000 3d points representing the car park can be obtained. With these points, the poses of the vehicle cameras are estimated successfully. The distance between the vehicle cameras shows the highest deviation from laser reference measurements for images showing a low point density area of the car park.

1 Single and system camera calibration

The road the automotive industry is driving down currently is characterized by autonomously driving cars. To allow advanced driver assistance systems to control a car reliably, they need information about the car's environment. To capture this information, cars can be equipped with several cameras observing the environment. As the relative position of environment objects to the car is important, the positions and orientations (pose) of the vehicle cameras relative to the car have to be known.

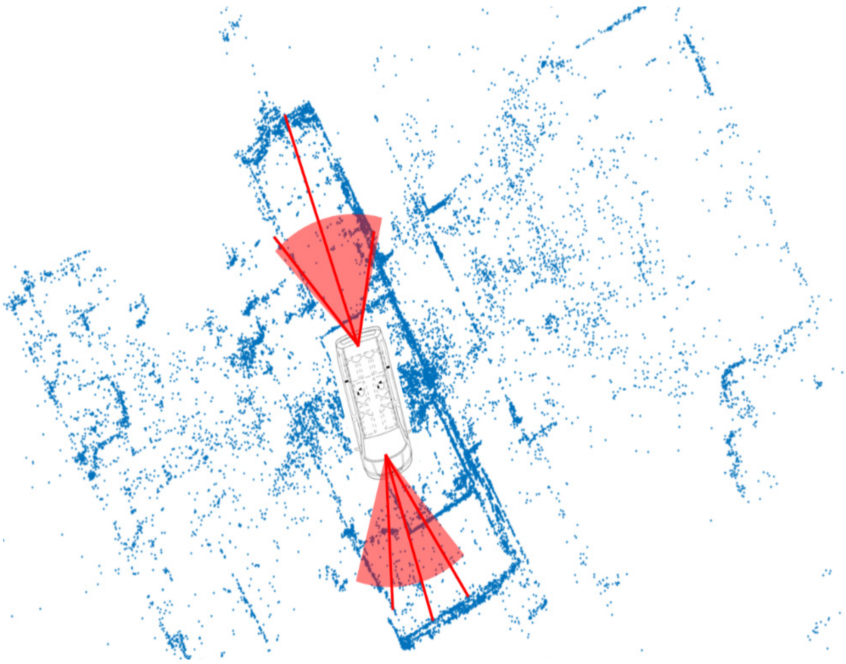

Fig. 1: Top-down view on a car, whose cameras with divergent fields-of-view (in red) can be orientated using 3d points (blue) generated for the calibration environment.

To determine the poses of the vehicle cameras, a system calibration can be performed. To keep costs low, the number of cameras is kept small, too.

[1] Technische Universität München, Photogrammetry & Remote Sensing, Arcisstraße 21, D-80333 München, E-mail: [alexander.hanel, stilla]@tum.de

So, their fields-of-view are typically non-overlapping, making stereo camera calibration with the same planar calibration pattern seen in all cameras impossible.
Due to their wide field-of-view, so called "action cameras" can cover a huge part of the car's environment despite their small number. Furthermore, a removable mount available for such cameras on the car windows with a suction pad can avoid in Germany problems with the car registration caused by additional "built-in" installations in the car.
Therefore, it has to be considered, that the wide field-of-view might cause large image distortions requiring reliable single camera calibration. In addition, the mounting on the windshield might not be rigid over time caused by mechanical movements during a car drive. To check the poses again and again, the system calibration has to be repeated from time to time.
To perform the calibration, a suitable calibration environment has to be selected. This environment has to be large enough to contain cars. Therefore, calibration laboratories are typically not feasible. Urban environments, as for example car parks, are large enough. As it might not be allowed by the owner of the car park and to keep costs and time effort low, the calibration procedure should work without manually placed reference marks. In addition, the number of placed reference marks is typically limited due to the available space. The number of reference marks extracted automatically using image feature descriptors can be higher. In contrast, it is not possible to provide 3d reference information for automatically extracted reference marks in advance.
All three mentioned tasks, single camera calibration, system calibration and generation of 3d reference information can be performed with the structure-from-motion method (Fig. 1). Therefore, in this contribution, a structure-from-motion software toolbox will be used to complete these tasks.

2 Related work

2.1 Camera calibration

Single camera calibration can be divided into photometric calibration (KINGSLAKE 1983; ASADA ET AL. 1996; KRAWCZYK ET AL. 2005), radiometric calibration (HEALEY & KONDEPUDY 1994; SEON & POLLEFEYS 2008; MITSUNAGA & NAYAR 1999) and geometric calibration (BROWN 1971; FRASER 2013; HARTLEY & ZISSERMAN 2003). Geometric camera calibration is used to estimate the interior orientation as well as image distortion parameters.
Geometric camera calibration performed simultaneously for a multi-camera system can be used to obtain the relative pose between the cameras (ZHANG & PLESS 2004). This requires typically a calibration pattern seen by all cameras. To calibrate a vehicle camera system, the car has to be placed in front of a calibration pattern to take the calibration images (DICK AND RICKS 2015). Such expensive calibration patterns are typically only available in calibration laboratories (SAFE CAR NEWS 2016).
There are different toolboxes for camera calibration available. The *camera calibrator app* for MATLAB (2016) bases on the algorithm of ZHANG (2000) and requires a set of images of a planar checkerboard pattern for calibration. Its stereo calibration function allows calibrating a rigid stereo camera system with overlapping fields-of-view. LI ET AL. (2013) provide another Matlab calibration toolbox using a feature descriptor-based calibration pattern, which can handle non-overlapping fields-of-view, as long as each camera looks onto a part of the planar calibration pattern.

To sum up, the mentioned solutions cannot handle the calibration of camera systems, where calibration patterns are not available or where the cameras cannot look onto the same planar calibration pattern.

2.2 Structure from motion

Structure-from-motion is a photogrammetric method to estimate 3d coordinates ('structure') from an image set. Opposite to rigid stereo camera systems (e.g. in HANEL ET AL. 2016), motion between the camera poses at different acquisition times creates the necessary stereo baseline to estimate the 3d coordinates (KOENDERINK & VAN DOORN 1991; STURM & TRIGGS 1996; HÄMING & PETERS 2010).
There are different software toolboxes for structure-from-motion, which use a similar pipeline (SCHÖNBERGER & FRAHM 2016) starting with feature extraction from images and their matching. After that, camera poses are estimated and 3d points triangulated. Finally, camera poses and 3d points are optimized using bundle adjustment. VisualSFM (WU 2007; WU 2011; WU ET AL. 2011) is a SfM toolbox well known in the literature.

3 Vehicle camera calibration in an urban environment

An overview over the method proposed in this contribution is given in **Fehler! Verweisquelle konnte nicht gefunden werden.**. First, 3d points of the calibration environment have to be calculated. Second, these 3d points are used as reference to estimate the pose of the vehicle cameras as well as their interior orientation and distortion parameters.

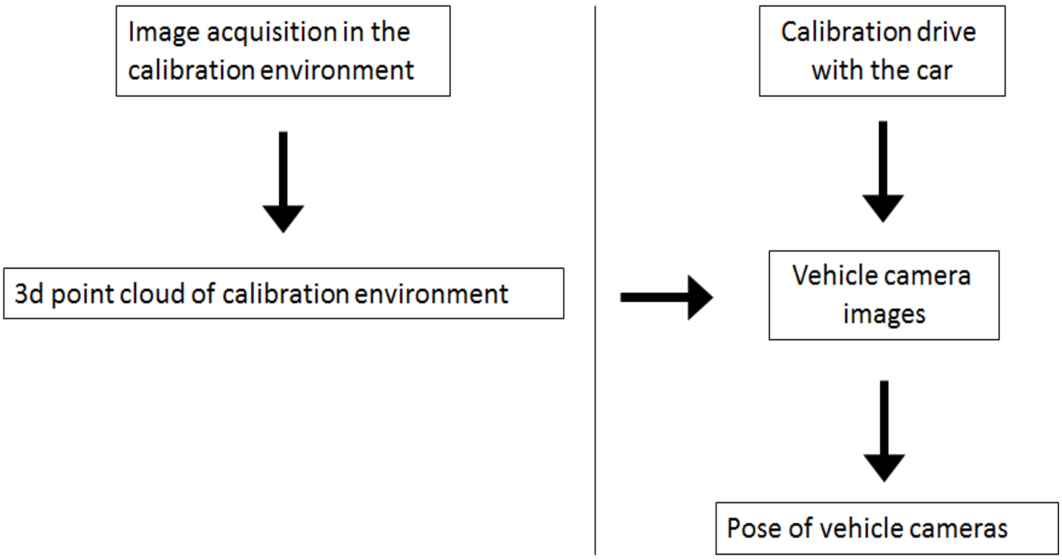

Fig. 2: Method for calibration of a vehicle camera system with divergent fields-of-view in an urban environment.

3.1 Creating a 3d point cloud of the calibration environment

In the first part, images of the urban calibration environment have to be acquired to obtain 3d points of the environment to use them as reference for the vehicle camera calibration.

To select a suitable calibration environment, the following aspects have to be considered. The objects within the environment have to have textured surfaces to allow a high number of 3d points and to avoid misassignments between different parts of the environment. Placing photogrammetric reference marks manually is therefore not necessary. The environment has to be large enough to allow a car to turn around, so that the vehicle cameras look at different reference points within the environment. For example, a car park can be used as calibration environment.

A professional camera with a fixed focal length lens is recommended to be used to ensure high quality images with minimal distortions. While the focus should be set to infinity to achieve a consistent interior orientation, aperture should have a medium value to get a wide depth of sharpness, but to avoid diffraction blur.

SIFT features are extracted from each image. Each image pair is checked for matching SIFT features. The pose of the used images is estimated as well as the 3d point coordinates are calculated for the matched image points using SfM. This results in a 3d point cloud of the calibration environment.

3.2 Estimating the pose of vehicle cameras using the point cloud

Having generated the 3d reference, the calibration of the vehicle cameras can be performed. Therefore, a car equipped with cameras looking at the area around the vehicle has to perform a calibration drive within the urban calibration environment. Images or videos of the vehicle cameras have to be recorded, and if necessary images extracted from the videos.

Temporarily corresponding images of all vehicle cameras are fed into the processing chain. SIFT features are extracted from the vehicle camera images and used to match them to the environment images (section 3.1).

A part of the matches between the vehicle camera images and the environment images relies on SIFT features, which are already used to calculate the 3d point cloud. These 3d points are now used as reference to estimate the pose of the vehicle camera images, as well as their interior orientation and distortion parameters. From a set of temporarily corresponding vehicle camera images, the mean relative pose between the vehicle cameras can be calculated.

4 Experiments

The calibration environment (Fig. 3) contains one parking lot on one side of a driveway and two parking lots on the other side. The environment consists of floor, ceiling and walls made of concrete. The walls and parts of the ceiling are painted white. Some signs are painted onto the walls. The experiments are conducted in the late evening, wherefore no cars are in the calibration environment at that time. The only illumination source in the car park is an artificial lamp, illuminating the environment quite darker than natural daylight.

a) b)

Fig. 3: a) Sample image of the calibration environment taken with a Nikon D3. Two parking lots can be seen, b) Sample image of the front camera of the car. The sharp shadow on the left is caused by powerful external LED lights, which are not available in a). The reference marks at the wall are not used in this contribution.

A Volkswagen Transporter 5 van is used as test car. Two vehicle cameras are placed. One camera is mounted at the front of the car looking through the windshield to the area in front of the car (**Fehler! Verweisquelle konnte nicht gefunden werden.**). The other camera is mounted at the rear of the car looking through the rear window to the area behind the car.

143 calibration environment images (Fig. 3a) are taken with a Nikon D3 camera with a 24 mm fixed focal length lens, acquiring 13 megapixels images. A tripod is used to keep ISO low, focus is set to infinite and the aperture to f/8. The lengths and height of the walls in Fig. 3a are measured with a laser distance meter to calculate ground control points for an Euclidean 3d reconstruction. Four ground control points are marked manually in seven images showing all four points together. The environment images are processed in VisualSfM. SIFT features are extracted and used to match the images. 3d point coordinates of parts within the TUM car park are calculated for the matching points.

To calibrate the vehicle cameras, they record videos with 4 megapixels resolution and 30 frames per second, while the car is driving slowly (< 5 km/h) through the calibration environment. 248 pairs of vehicle images (Fig. 3b) are extracted from the videos, one image pair per second. Videos instead of images are recorded, as the street scenes for later research will be recorded with videos, too. Garmin VIRB Ultra 30 action cameras with 2.73 mm focal length are used as vehicle cameras. Acquisition parameters are set automatically by the cameras. The recording is started and stopped with the Garmin VIRB app, allowing synchronous control of all cameras. During video recording,

the area near the vehicle path is illuminated with powerful permanent LED light used in photographic studios.
The vehicle camera images are processed in VisualSfM, too. They are linked to the environment images with common SIFT features, so that the 3d point coordinates can be used to estimate the pose of the vehicle cameras.

a) b)

Fig. 4: a) Position (red rectangle) of the front camera, looking through the windshield to the area in front of the car, b) Position of the rear camera, looking through the rear window to the area behind the car. Image brightness strongly increased for better visualization.

5 Results

The 3d point cloud of the TUM car park is created with around 26,000 3d points (Fig. 5). All of the 143 calibration environment images are used and more than 600,000 SIFT features are extracted from them, in average around 4,800 SIFT features per image. More than 126 thousand combined matches (i.e., independent whether a SIFT feature is found in two or more images, it is always counted as one match) are calculated for all image pairs. That means, 3d points can be calculated for around 4 % of all matches.
In the following, the appearance of the point cloud is discussed. According to Fig. 5, the density of the point cloud in highest in the central part of the figure. The closer the 3d points to the camera positions are, the higher their density is. In areas further away from camera positions, the point density is lower. This observation correlates with few aspects. First, in closer areas, the building walls are shown with more details (e.g. wall signs), which can be seen as intensity differences in images. These differences are needed by the SIFT algorithm to extract features. Second, closer areas are pictured in a higher number of images than areas further away, leading to a higher number of matches in closer areas. Third, as the illumination in the car park is quite low, remote areas tend to appear dark in the images, which reduces the intensity differences needed by SIFT.

Accuracy information about the 3d points are not given by VisualSfM. Furthermore, as the only parameter of the interior orientation, the focal length is estimated with $f = 18.23\,mm$. For comparison purpose, the focal length calibrated in a laboratory environment is f = 24.29 mm. The difference of around 6 mm might be interpreted, that VisualSfM might require calibration parameters as well as undistorted images as a-priori data.
In sum, as the area with the highest point density is the area used to record vehicle camera videos, the different point densities do not have negative consequences on the estimation of the vehicle camera poses.

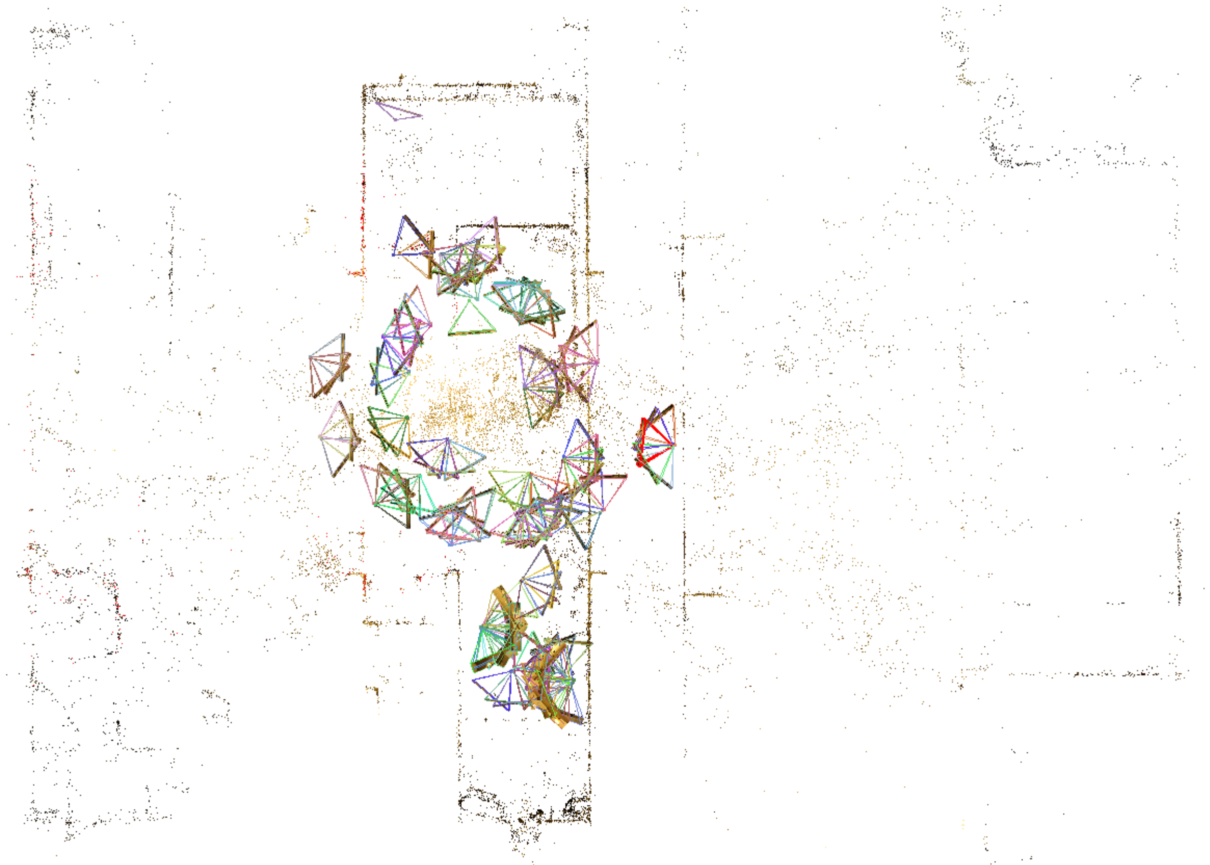

Fig. 5: Top-down view on the 3d point cloud of the TUM car park. Most points are along walls. The drive way goes horizontally from left to right in the middle of the figure, parking lots are orthogonal to it. The area close to the camera positions contains the most 3d points and is used to record the vehicle camera videos.

All 248 image pairs of the front and rear vehicle camera are matched to the environment images, so in total 639 (143 + 248 + 248) images are processed together. 378,867 combined matches are found. Around 10,000 3d points are calculated after adding the front and rear camera images. Previous tests of image pairs of the vehicle cameras have shown a maximum time gap between the two cameras of 0.1 second. As the car is standing or driving very slowly during recording, the image pairs are approximated as temporarily synchronized.

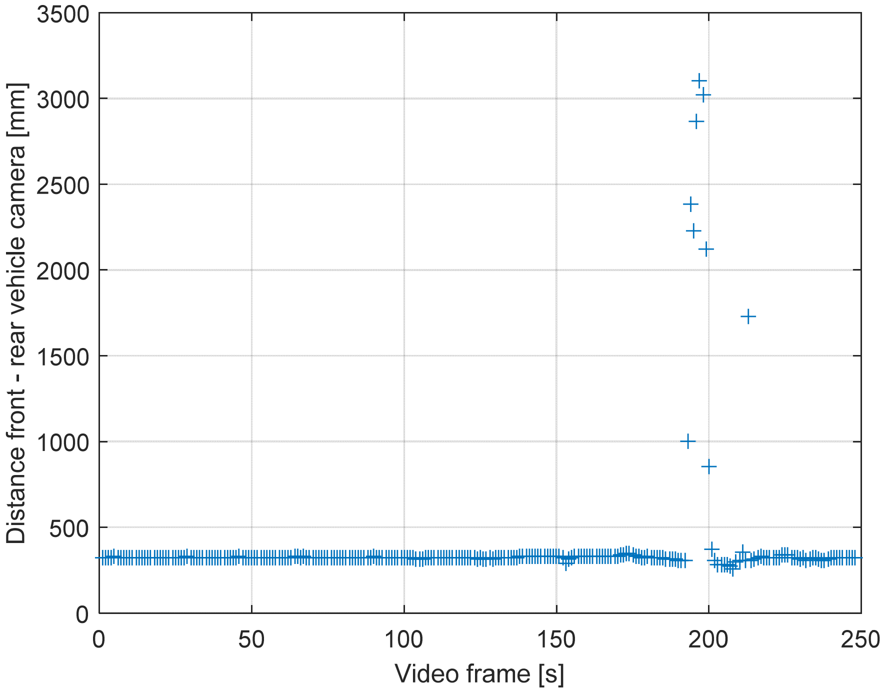

Fig. 6: Distance between the front and rear vehicle camera for the video recorded during the calibration drive. At most times, there is a deviation of the distance over time of smaller than 10 cm, but around frame 200, there is a large deviation of few meters. At that time, the vehicle cameras are looking at low point density areas.

The distance between the front and rear camera calculated for each of the 248 image pairs is used as evaluation metric for the estimated pose (**Fehler! Verweisquelle konnte nicht gefunden werden.**). The distance varies over time within an interval of 10 cm. Only around second 200, there is a way larger deviation of few meters compared to other points of time. At that time, the vehicle cameras are looking to an area (cf. Fig. 5: cameras looking at second 200 to the left and to right part of figure, respectively) of the calibration environment with low 3d point density. That might explain also, why the deviation is increasing to the maximum step by step and decreasing also step by step, as the vehicle moves slowly to and from the area with low point density.

The median distance is 3.26 m, while laser distance measurements show 3.22 m. These values vary only by a few centimetres, which might lead to the conclusion, that the estimated poses are reliable. Again, the missing accuracy information from VisualSfM prevents such conclusion.

To sum up, a car park built of concrete with relatively low structured surface allows to obtain a high number of several thousand 3d points. As these points are spread within the calibration area, they allow to calibrate the vehicle cameras even without overlapping fields-of-view. Neither standard planar calibration patterns, nor patterns designed specifically for the calibration environment are necessary.

As the generation of the 3d reference information and the vehicle camera calibration are two independent steps, the calibration can be repeated every time the car is driving through the part of the car park with the 3d information available.

Processing the images in VisualSfM is easy to use. But there are limitations in the estimated camera parameters (only interior, distortion) as well as there is no accuracy information about the obtained results.

6 Conclusion

In this paper, an approach to calibrate a vehicle camera system with divergent fields-of-view using urban structures has been proposed. It is shown, that several thousand 3d points can be obtained from images of urban structures using SfM. For images of vehicle cameras, taken during a drive through the urban structure, these 3d points can be used as reference to estimate their poses. The distance between the front and rear camera calculated from a set of image pairs, deviates in median only few centimetres from laser distance measurements. But it has been shown, that a high deviation correlates with a low number of 3d points.

For further investigation, the accuracy of the obtained 3d points as well as of the vehicle camera poses should be analysed. Furthermore, pre-processing steps for the images before using them for 3d point cloud generation should be considered to ensure a high accuracy pose estimation. This might involve separate single camera calibration, as well as the use of highly reliable ground control points, like for example obtained from tachymeter measurements.

7 References

ASADA, N., AMANO, A. & BABA, M., 1996: Photometric calibration of zoom lens systems. Proceedings of the 13th International Conference on Pattern Recognition **1**, 186-190.

BROWN, D. C., 1971: Close-range camera calibration. Photogrammetric Engineering **78**(8), 855-866.

DICK AND RICKS AUTO INTERIORS INC., 2015: Windshield Replacement. http://www.dickandricks.com/blog/windshieldreplacement-not-the-same-glass-repair-as-yr-last-vehicle, last access on 2015-12-08.

FRASER, C., 2013: Automatic Camera Calibration in Close Range Photogrammetry. Photogrammetric Engineering and Remote Sensing **79**(4), 381-388.

HANEL, A., HOEGNER, L. & STILLA, U., 2016: Towards the influence of a car windshield on depth calculation with a stereo camera system. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences **41**(B5), 461-468.

HARTLEY, R. & ZISSERMAN, A., 2003: Multiple View Geometry in Computer Vision. 2nd edition. Cambridge University Press, New York, USA.

HEALEY, G. E. & KONDEPUDY, R., 1994: Radiometric CCD camera calibration and noise estimation. IEEE Transactions on Pattern Analysis and Machine Intelligence **16**(3), 267-276.

HÄMING, K. & PETERS, G., 2010: The structure-from-motion reconstruction pipeline – A survey with focus on short image sequences. Kybernetika **46**(5), 926-937.

KINGSLAKE, R., 1983: Optical system design. Academic Press, Waltham, USA.

KOENDERINK, J. J. & VAN DOORN, A. J., 1991: Affine structure from motion. Journal of the Optical Society of America A **8**(2), 377-385.

KRAWCZYK, G., GOESELE, M., SEIDEL, H.-P., 2005: Photometric Calibration of High Dynamic Range Cameras. Research Report. Max-Planck Institute for Informatics, Saarbrücken, Germany.

LI, B., HENG, L., KÖSER, K. & POLLEFEYS, M., 2013: A Multiple-Camera System Calibration Toolbox Using a Feature Descriptor-Based Calibration Pattern. 2013 IEEE/RSJ International Conference on Intelligent Robots and Systems, 1301-1307.

MATLAB, 2016: Single Camera Calibration App – MATLAB & Simulink. https://de.mathworks.com/help/vision/ug/single-camera-calibrator-app.html?requestedDomain=www.mathworks.com, last access on 2017-01-10.

MITSUNAGA, T. & NAYAR, S. K., 1999: Radiometric self calibration. IEEE Computer Society Conference on Computer Vision and Pattern Recognition **1**, 374-380.

SAFE CAR NEWS, 2016: National Windscreen launches UK-wide ADAS calibration service, Website, http://safecarnews.com/national-windscreen-launches-uk-wide-adas-calibration-service-ma793/, last access on 2017-01-10.

SCHÖNBERGER, J. L. & FRAM, J.-M., 2016: Structure-from-Motion Revisited, IEEE Conference on Computer Vision and Pattern Recognition, 4104-4113.

SEON, J. K. & POLLEFEYS, M., 2008: Robust Radiometric Calibration and Vignetting Correction. IEEE Transactions on Pattern Analysis and Machine Intelligence **30**(4), 562-576.

STURM, P. & TRIGGS, B., 1996: A factorization based algorithm for multi-image projective structure and motion. European Conference on Computer Vision, 709-720.

WU, C., 2007: SiftGPU: A GPU implementation of Scale Invaraint Feature Transform (SIFT), http://cs.unc.edu/~ccwu/siftgpu, 2017-01-10.

WU, C., 2011: VisualSFM: A Visual Structure from Motion System, http://ccwu.me/vsfm/, last access on 2017-01-10.

WU, C., AGARWAL, S., CURLESS, B. & SEITZ, S. M., 2011: Multicore Bundle Adjustment, CVPR 2011, 3057-3064.

ZHANG, Q. & PLESS, R., 2004: Extrinsic Calibration of a Camera and Laser Range Finder (improves camera calibration). IEEE/RSJ International Conference on Intelligent Robots and Systems, Proceedings **3**, 2301-2306.

ZHANG, Z., 2000: A Flexible New Technique for Camera Calibration. IEEE Transactions on Pattern Analysis and Machine Intelligence **24**(11), 1330-1334.