Accurate Sensor Calibration for Geometric Computer Vision

Zoltán Pusztai

Theses of PH.D. Dissertation

Supervisor:
Levente Hajder, PhD (ELTE IK)

Department of Algorithms and Their Applications,
Faculty of Informatics
Eötvös Loránd University

Budapest, 2019

Motivation

Sensor calibration aims to find the internal or external parameters of a selected mathematical model for a specific device. This dissertation focuses on the calibration of widely used modalities in computer vision, e.g. structured-light scanners, cameras, and LiDAR (Light Detection and Ranging) sensors.

Calibration is a crucial step for many applications in computer vision. The a priori knowledge of parameters (e.g. the focal length or camera poses) makes the processing of other algorithms faster and more accurate. Some of these parameters can be estimated online, during the usage of the device. However, online algorithms usually achieve lower accuracy, and used only when the device is unreachable by a human and no manual calibration is available. This dissertation focuses on accuracy rather than automatic or object-free calibration. The proposed methods are based on the known geometrical properties of the observed objects and correspondences between different views.

Sometimes the calibration of unusual devices is needed. While the literature of camera, LiDAR, GPS, IMU calibrations are wide, there exist only a few calibration methods for a turntable calibration. Turntables are widely used in a 3D scanning, however, most of the existing calibration pipelines do not deal with the calibration of such a device. The calibration, in case of a turntable, means that its centerline has to be determined in the camera coordinate system. Using a turntable makes it possible to obtain high-resolution scans of the objects in 360 degrees. The axis of the turntable determines how the scans from the individual viewpoints can be merged.

The accurate calibration of a 3D scanner makes it possible to generate GT (Ground Truth) data for feature matching and the quantitative comparison of affine feature detectors. Most of the evaluation databases use homography-based image pairs for the comparison. However, the accurate depth and sensor parameters acquired using the accurate calibration of a structured-light scanner can be used to calculate the movement of the objects and to generate GT. This makes it possible to quantitatively compare the feature detectors and matchers on real-world rotating objects, which has not been introduced in literature before. Moreover, the database is extended with outdoor images taken by quadcopter-mounted camera. The GT data are calculated from the restricted flight-path of the quadcopter.

Calibration is especially challenging when modalities of different kinds are considered, e.g. cameras and LiDAR sensors. The former takes high-resolution color images, while the latter acquires sparse point clouds. Even the dimensions and density properties of these outputs are significantly distinct; thus, the calibration needs to take place so that accurate correspondences are easily acquired from the devices.

Multi-device systems consisting of cameras and LiDAR sensors are widely used these days in autonomous driving. The external calibration determines the relative pose of these. Due to accurate calibration and pose estimation, the surroundings of the autonomous vehicle can be mapped, even if just some of the sensors share a common field of view.

Besides the theoretical background, this dissertation presents easy to use, novel methods for the calibration of different modalities. Each of the presented calibration methods works well in real-life and some of them even attract the attention of industry purposes as well.

Previous Works

The thesis contributes to three topics to the state-of-the-art. Here, the previous works on individual topics are reviewed.

Turntable-aided structured-light scanning:

Structured-light scanning has been a long-studied subject of computer vision. These scanners are equipped with a camera and a projector, and use an image sequence, called structured-light, to reconstruct the shape of the objects. Structured-light encodes each projector pixel individually. Then the shape of the object is obtained by triangulation. A pre-calibration step is required to determine the intrinsic and extrinsic parameters of the sensors. The calibration is usually done using an object with known geometry, for example, a chessboard. The spatial chessboard corners are known and the projections of the corners can be found in the images. Then the intrinsic parameters of the camera and the spatial location of the chessboard can be estimated. The same procedure is used for the projector as well; however, the projections of the corners must be calculated by decoding the structured-light. Previous calibration methods use only the raw decoded projector values from the camera to establish camera-projector pixel correspondences. However, these methods achieve only pixel precision and are not robust to decoding errors. Thus, the first thesis contributes to this calibration step by applying homographies. A homography describes the transformation between two projections of a plane. It eliminates the noisy decoded projector pixels and achieves sub-pixel precision, making the camera-projector calibration more accurate.

Our particular scanner is also equipped with a turntable to make 360° reconstruction possible. The scanner obtains scans from different viewpoints by rotating the object. To merge the point clouds, the axis of the turntable needs to be determined relative to the camera. Previous methods on the topic mostly use ICP (Iterative-Closest-Points) algorithm, which uses only the point clouds for merging. Other methods use images of a rotating chessboard, calculate the extrinsic locations of the camera, and fit a circle to determine the turntable axis. Neither result of the mentioned family of methods is satisfying. The aligning error of different scans is visible in the final point cloud of the object. Thus, a novel method is proposed for the turntable axis calibration. Determining the centerline is especially challenging since only a small error in the axis estimation produces visible aligning errors in the merged point cloud. The proposed method not only calculates the centerline of the turntable accurately, but also refines the extrinsic parameters of the camera and the projector. Thus, the calibration system is extremely accurate.

Quantitative Evaluation of Affine Covariant Feature Matching:

Point features are frequently used as input data for many computer vision problems. Recently, affine transformations are also considered. An affine transformation describes the local image warp around the point features. Thus, the problems of epipolar geometry, that can be solved by point correspondences, can also be solved by affine correspondences using less number of features than using only point correspondences. Despite that many comparison datasets are available with accurate GT point features, the literature focusing on quantitative affine covariant feature matching is very limited. Most of the well-known quantitative comparisons are based on homography. The main problem with these comparisons is twofold: feature matching is usually not involved, and only planar objects can be observed or the camera has to be fixed.

The used measure, called repeatability, is also problematic. It is calculated as the ratio between the number of repeated features over the minimum of the number of detected features in an image pair. A feature pair is considered repeated if their overlapping region is higher than a pre-selected threshold. However, affine feature regions with larger regions have a better chance to reach the overlapping threshold. Moreover, the minimum of the number of features in the denominator makes the measure error prone, if only a low number of features are found in one of the images.

To overcome the shortcomings of the above-mentioned comparisons, novel GT generation, and quantitative evaluations are introduced. In the first scenario, the GT data are generated by the accurate structured-light scanner, introduced in Thesis 1. The scanned objects are making rotation movement between the images. In the second scenario, the images are captured by a quadcopter mounted camera, and the GT is obtained from the restricted flight-path of the quadcopter. Instead of the repeatability, the Frobenius norm of the affine transformations and the Euclidean distance of the affine correspondences are evaluated.

Accurate Calibration of Multi-LiDAR-Multi-Camera Systems:

Autonomous driving is currently in the focus of several research communities and industrial partners. Several sensors have to be used together to map the surroundings of an autonomous vehicle. In order to efficiently work these sensors together, their relative position and orientation need to be known a priori; thus, extrinsic calibration is required.

The joint application of camera and LiDAR sensors compensate for the shortcomings of the other; thus, they can be effectively used jointly in many scenarios. The extrinsic calibration of such a sensor pair is an active research topic nowadays. Online and offline calibrations can be distinguished. Online calibration is applied while the sensors are in use and usually no calibration object is needed. Offline calibrations methods need one or more calibration objects and usually human intervention to estimate the extrinsic parameters. In general, offline calibration methods achieve better accuracy than online methods.

Most of the previously proposed calibration use chessboards or other planar objects. The motivation behind the chessboards is that they can be used for camera calibration automatically and achieve high accuracy. However, all of these methods struggle to find the chessboard edges in the LiDAR point cloud. Several solutions are published to determine these edges accurately; however, none of them proved to be accurate enough. Moreover, the chessboard pattern makes the LiDAR point clouds noisier, which affects the accuracy of the calibration.

Other types of methods reconstruct a part of the scene from multiple camera images and try to merge the LiDAR point cloud with the reconstructed one. Some of these use ICP to estimate the rigid transformation between the devices; however, ICP needs good initial parameters to converge and its accuracy is not satisfying. More sophisticated methods try to establish correspondences between the point clouds by the detection of common planes. However, establishing plane correspondences is also problematic.

The proposed method uses cardboard boxes for the calibration. Cardboard boxes have 3 perpendicular planes; thus, the intersections of these planes can be accurately calculated. From the edges of the box, the corners are determined in the LiDAR point clouds. These corners can be found in the camera images with high accuracy. From the spatial and image locations, 3D-2D point correspondences are obtained. Thus, the extrinsic calibration is reduced to the PnP (Perspective-n-Point) problem, which can be solved effectively. Moreover, the method can also be used for LiDAR-LiDAR calibration. In this case, the calibration is equivalent to the registration of spatial points. Furthermore, the method can be easily extended to handle multiple LiDAR and multiple camera devices.

Main Contributions

The main contributions are summarized in the following three theses.

Thesis 1 – Calibration of Turntable-Aided Structured-Light Scanners [4;7;8]

Considering the problem of turntable-aided structured-light scanner calibration, a new method is proposed to achieve sub-pixel accurate correspondences between the camera and projector pixels. A homography is robustly estimated between the chessboard regions visible from both the camera and the projector. Then this homography is used to eliminate the false matches obtained from the wrong decodings of the structured-light and to refine the correspondences to the sub-pixel level. For turntable calibration, a novel method is proposed to calculate its centerline in the camera coordinate system. The method iteratively refines the extrinsic parameters of the camera and the projector. A visual comparison of calibration methods indicates that the proposed method outperforms the others.

Thesis 2 – GT Data Generation for Affine Covariant Feature Detectors and Their Quantitative Evaluation [1;3;6]

A GT generation method is proposed for affine covariant feature detectors based-on the high-precision structured-light scanner. The GT data consist of not only point correspondences between the images, but also affine transformations. The comparison database improves on the previously proposed comparisons, in the sense that real-world, rotating 3D objects are used, instead of just static scenes. The database is extended with outdoor images taken by a quadcopter mounted camera. The GT data generation methods for the several types of restricted movement of the quadcopter are also proposed.

Thesis 3 – Extrinsic Calibration of Camera-Lidar Systems
[2;5]

A novel method is proposed for camera-LiDAR extrinsic calibration. Compared to the current state-of-the-art, this method uses a 3D calibration objects, instead of planar ones. The 3D corners of the cardboard boxes are accurately computed in the LiDAR points. Then the extrinsic camera-LiDAR calibration is reduced to the PnP (Perspective-n-Point) problem by finding 3D-2D correspondences between the camera image and the LiDAR point cloud. LiDAR-LiDAR calibration is possible by finding the same spatial corners in the point cloud and by point registration. The proposed method is evaluated in both synthetic and real-world tests, outperforming the current state-of-the-art techniques. Moreover, the method can be easily extended for LiDAR-camera system calibration, containing multiple sensors.

Acknowledgements

I gratefully acknowledge the support of the following projects:

EMBERI ERŐFORRÁSOK MINISZTÉRIUMA

Supported BY the ÚNKP-18-3 New National Excellence Program of the Ministry of Human Capacities.

EFOP-3.6.3-VEKOP-16-2017-00001: Talent Management in Autonomous Vehicle Control Technologies – The Project is supported by the Hungarian Government and co-financed by the European Social Fund.

List of the author's publications:

[1] **Pusztai, Z;** Hajder, L. *Quantitative Affine Feature Detector Comparison based on Real-World Images Taken by a Quadcopter*. In: Proceedings of the 14th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications - Volume 5: VISAPP, 2019, ISBN:978-989-758-354-4, pages 704-715. DOI:10.5220/0007372907040715

[2] **Pusztai, Z**; Eichhardt I; Hajder L. *Accurate Calibration of Multi-LiDAR-Multi-Camera Systems*. In: SENSORS (Basel). 2018; 18(7):2139. Published 2018 Jul 3. doi:10.3390/s18072139

[3] **Pusztai, Z**; Hajder, L: *Affine Invariant Feature Tracker Evaluation based on Ground-Truth Data*, In: Szirmay-Kalos, László; Renner, Gábor, IX. magyar számítógépes grafika és geometria konferencia, GRAFGEO 2018, Budapest, Magyarország : Neumann János Számítógép-tudományi Társaság (NJSZT), (2018) pp. 164-173. , 10 p. ISBN : 978-963-313-282-1

[4] **Pusztai, Z**; Hajder, L: *Ground-truth tracking data generation using rotating real-world objects* In: Braz, J; [et, al.] Computer Vision, Imaging and Computer Graphics Theory and Applications, 11th International Joint Conference, VISIGRAPP 2016, Revised Selected Papers Cham (Németország), Németország : Springer, (2017), pp. 395-417. ISBN: 978-3-319-64870-5

[5] **Pusztai, Z**; Hajder, L: *Accurate Calibration of LiDAR-Camera Systems using Ordinary Boxes.* In: CVF 2017 IEEE International Conference on Computer Vision Workshops (ICCVW), (2017) pp. 394-402. , 9 p. ISSN: 2473-9944, DOI: 10.1109/ICCVW.2017.53

[6] **Pusztai, Z**; Hajder, L: *Quantitative Comparison of Affine Invariant Feature Matching* In: Imai, F; Tremeau, A; Braz, J: VISAPP 2017. Proceedings of

the 12th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications. Vol. 6, Setúbal, Portugália : SciTePress, (2017), pp. 515-522. , 8 p., DOI: 10.5220/0006263005150522

[7] **Pusztai, Z**; Hajder, L: *Quantitative Comparison of Feature Matchers Implemented in OpenCV3* In: Čehovin, L; Mandeljc, R; Štruc, V (szerk.), Proceedings of the 21st Computer Vision Winter Workshop. CVWW 2016, Ljubljana, Szlovénia : SDRV, (2016), pp. 1-9. ISBN 978-961-90901-7-6

[8] **Pusztai, Z**; Hajder, L: *A Turntable-based Approach for Ground Truth Tracking Data Generation* In: Magnenat-Thalmann, N et al. (szerk.), VISIGRAPP 2016. Proceedings of the 11th Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications. Vol. 3. VISAPP 2016, Setúbal, Portugália : SciTePress, (2016), pp. 500-511., ISBN: 978-989-758-175-5, DOI:10.5220/0005719404980509

List of the author's other publications:

[9] Gábor, Rácz; **Zoltán, Pusztai**; Balázs, Kósa; Attila, Kiss: *An improved Community-based Greedy algorithm for solving the influence maximization problem in social networks*, ANNALES MATHEMATICAE ET INFORMATICAE 44 pp. 141-150. , 10 p. (2015), ISSN 1787-5021

[10] Balázs, Kósa; Márton, Balassi; Péter, Englert; Gábor, Rácz; **Zoltán, Pusztai**; Attila, Kiss: *A Basic Network Analytic Package for RapidMiner* In: Simon, Fischer; Ingo, Mierswa (szerk.) Proceedings of the 5th Rapidminer World 2014, Aachen, Németország : Shaker Verlag, (2014), pp. 47-59. , 13 p., ISBN-13: 978-3844029468