Calibration of Heterogeneous Sensor Systems

Juraj Peršić
University of Zagreb Faculty of Electrical Engineering and Computing,
Department of Control and Computer Engineering,
Laboratory for Autonomous Systems and Mobile Robotics (LAMOR)
Email: juraj.persic@fer.hr

Abstract—Environment perception is a key component of any autonomous system and is often based on a heterogeneous set of sensors and fusion thereof for which sensor sensor calibration plays fundamental role. It can be divided to intrinsic and extrinsic sensor calibration. Former seeks for internal parameters of each individual sensor, while latter provides coordinate frame transformation between sensors. Calibration techniques require correspondence registration in the measurements which is one of the main challenges in the extrinsic calibration of heterogeneous sensors, since generally, each sensor can operate on a different physical principle. Measurement correspondences can originate from a designated calibration target or from features in the environment. Additionally, environment features can be used to estimate motion of individual sensors and the calibration is found by aligning these estimates. Motion-based calibration is the most common approach in the online calibration since it is more practical than the target-based methods, although it can lack in accuracy. Furthermore, online calibration is beneficial for system robustness as it can detect and adjust recalibration of the system in runtime, which can be seen as a prerequisite for long-term autonomy.

I. INTRODUCTION

Robust environment perception is one of the essential tasks which an autonomous mobile robot or vehicle has to accomplish. To achieve this goal, various sensors such as cameras, radars, LiDARs, and inertial navigation units are used and information thereof is often fused. Essential tasks such as simultaneous localization and mapping (SLAM), detection and tracking of moving objects, odometry, etc. are often improved by sensor fusion.

A fundamental step in the fusion process is sensor calibration, both intrinsic and extrinsic. Former provides internal parameters of each sensor (e.g. focal length of a camera, bias in LiDAR range measurements), while latter provides relative transformation from one sensor coordinate frame to the other. The calibration can tackle both parameter groups at the same time or assume that sensors are already intrinsically calibrated and proceed with the extrinsic calibration. Additionally, temporal synchronization of the sensors is sometimes performed within the calibration.

Intrinsic parameters are related to the working principle of the sensor. Therefore, methods for finding intrinsic parameters do not share much similarities for different types of sensors. On the other hand, parametrization of extrinsic calibration, i.e. homogeneous transformation, can always be expressed in the same manner, regardless of the sensors involved in it. Despite that, solving the extrinsic calibration requires finding correspondences in the data acquired by the sensors which can be challenging since different types of sensors measure different physical quantities.

After correspondence registration, optimization steps are performed to estimate the calibration parameters. While some methods require intrinsically calibrated sensors to find the extrinsic calibration, others perform optimization on both parameter groups simultaneously. These methods typically try to satisfy some geometric constraints through minimization of a problem-specific reprojection error. The geometric constraints involve nonlinearities which often cannot be solved analytically. To resolve that problem, estimators use iterative techniques to find the appropriate solution. Due to the non-convexity of the problem caused by the nonlinearities, these methods have a risk of converging to a local minimum. To avoid that risk, some methods divide optimization in initial rough estimates that guarantee near-optimal solutions followed by nonlinear iterative refinement step. The success of the optimization is highly dependant on the provided data. Important step before the data acquisition is to determine minimal requirements on the dataset for which the problem becomes identifiable (or observable in case of dynamical systems).

The calibration approaches can be target-based or targetless. In the case of target-based calibration, correspondences originate from a specially designed target, while targetless methods utilize environment features perceived by both sensors. Former has the advantage of the freedom of design which maximizes the chance of both sensors perceiving the calibration target, but requires the development of such a target and execution of an appropriate offline calibration procedure. The latter has the advantage of using the environment itself as the calibration target and can operate online by registering structural correspondences in the environment, but requires both sensors to be able to extract the same environment features. Registration of structural correspondences can be avoided by motion-based methods, which use the system’s motion estimated by the individual sensors to calibrate them. These methods have two main advantages: (i) they rely less on the sensors operating principles and can be applied to different sensors, assuming that a sensor can estimate its motion, (ii) unlike other methods, they are able to extrinsically calibrate sensors with non-overlapping fields of view.

The paper is organized as follows. Section I elaborates the calibration methods including calibration targets. Section II explains common approaches in targetless calibration which
uses features in the environment as correspondences in the sensor data. Afterwards, Section V provides details on existing motion-based methods. Lastly, conclusion is given in Section VI with focus on the importance of sensor calibration.

II. TARGET-BASED METHODS

Calibration targets are frequently used due to numerous advantages. They simplify the correspondence registration step since the number and type of correspondences is known in advance which virtually eliminates the problems associated with outliers. Additionally, target-based methods can use a priori knowledge about the target which can enhance the calibration results. Therefore, target-based methods are generally more precise than the targetless. Finally, there are no requirements on the environment which can be uninformative and prevent targetless methods from success. However, it is the least practical method since it requires design and construction of the target and may not always be suitable (e.g. end-user applications like smartphones). Furthermore, it has to be performed offline before any other application for which the calibration is important. Therefore, it is impossible to make any runtime adjustments and the process has to be repeated in case of decalibration.

Properties of a well-designed target are (i) ease of detection and (ii) high localization accuracy for all the sensors in the calibration. The former ensures the success of the correspondence registration, while the latter has great influence on the quality of the results given by the optimization step. Furthermore, if the a priori knowledge about the target is used, construction imprecision may lead to poor calibration results. Perception sensors used in robotics utilize a wide range of physical phenomena to extract information about the environment. Due to different type of data provided by heterogeneous sensors, there exist many diverse target designs. In the sequel we will present some of the designs grouped by the sensor types.

A. Camera

Cameras are passive sensors that utilize the light which goes through the lens and is detected at the optical sensor. They are a rich source of information with an affordable price what makes them commonly used in robotics and other fields. Due to their long presence and frequent usage, intrinsic camera calibration has been given a lot of research attention which resulted in many camera description models and calibration techniques. While cameras with high distortion such as fisheye and omnidirectional cameras require more complex models \cite{Fernandez-Moral2014}, commonly used cameras with slight distortion are usually modelled as pinhole cameras with a previously rectified image. This intrinsic parametrization consists of distortion coefficients (e.g. radial distortion) and camera matrix formed by focal length, pixel scale factors, principal point and skewness between the axis.

Commonly used camera calibration targets are planar checkerboard patterns. They are suitable because they can be easily detected in the image and enable sub-pixel resolution using interpolation based on a known target dimensions. Calibration methods are based on pioneering work by Tsai \cite{Tsai1987} and Zhang \cite{Zhang2000}. Besides checkerboard pattern, a grid of circles is also frequently used \cite{Zhang2000} with comparison study of different patterns given in \cite{Pless2004}. Novel calibration target is presented in \cite{Zhang2000} where authors use a noise-like pattern with many features of varying scales. It is suitable for both intrinsic and extrinsic calibration of multiple cameras with no or little filed of view (FOV) overlap. The only requirement is that the neighbouring cameras observe parts of the target which may not overlap at all. Additionally, it can simultaneously handle both close-range and far-range cameras.

B. LiDAR

LiDARs are active sensors that use light pulses to determine the range of the objects in the environment and provide results in the form of point cloud, i.e., a set of 3D points. They usually consist of one or more rotating light transmitters/receivers. LiDARs are classified as: 1D when they measure a distance on a single ray, 2D when they measure distances in only one plane, and 3D or multi-layer when they measure distances in multiple planes. Considering sensor calibration, 2D and 3D LiDARs have received extensive attention due to the application requirements and possibility to recover structure from the environment.

Intrinsic parameters of interest are range measurement offsets and pose of the individual rays to the common reference frame. Unlike cameras, precision of intrinsic factory calibration parameters is usually considered sufficient. However, for the applications that require higher precision, authors in \cite{Fernandez-Moral2014} propose a method for intrinsic calibration of rotating 3D LiDAR using a box with known dimensions, while the authors in \cite{Zhang2000} use a planar wall as a calibration target.

Extrinsic calibration between multiple LiDARs is mostly done by motion-based methods which are applicable when there is small or none FOV overlap. For sensor configurations in which multiple 2D LiDARs share the same parts of FOV, Fernandez-Moral et al. \cite{Fernandez-Moral2014} presented a solution which uses corner structures to perform extrinsic calibration. Additionally, using the rank of Fisher Information Matrix (FIM) they show that problem becomes identifiable when at least three perpendicular planes are observed.

C. LiDAR–Camera

Point clouds from 2D/3D LiDARs are often fused with camera images. Both are rich information sources and precise extrinsic calibration is crucial for tasks such as 3D reconstruction what led to development of many calibration methods. A common approach in target-based LiDAR–camera calibration is using planar targets which are easily detected and localized in the point cloud covered by a pattern (e.g. checkerboard) which allows estimation of the plane position and orientation in the image.

Widely adopted and extended method presented by Zhang and Pless \cite{Zhang2000} introduced point-plane geometric constraint initially designed for 2D LiDAR–camera calibration. LiDAR points originating from the target plane are transformed into
the camera frame and the method tries to minimize point to plane distances based on the estimated plane parameters in the image. Pandey et al. [11] showed that the method is also applicable in case of 3D LiDAR—camera calibration. Zhou and Deng [12] improved the method by introduction of additional constraints which decoupled rotation from translation. They achieved better results than other methods because their method is less affected by errors in the plane parameters estimation in the checkerboard image. Additionally, they showed that for a 2D LiDAR, at least five correspondences should be made with different target orientations, while a 3D LiDAR required minimum of 3 different views. Geiger et al. [13] tried to reduce time of the calibration procedure by extending the method with global correspondence registration which allows for multiple plane observations in a single shot. The same constraint was used by Mirzaei et al. in [14] where instead of checkerboard pattern, ApriTag fiducial markers were used [15]. Additionally, they extended the extrinsic calibration with estimation of intrinsic LiDAR parameters. ApriTag markers and the same geometric constraint were also used in [16] as a part of multi-sensor graph based calibration.

Besides commonly used point-plane constraint, 3D LiDAR-camera pair was calibrated based on the point-point correspondences. Velas et al. [17] used a target with circular holes which allows a single-shot calibration and does not require observation of the plane in multiple orientations. Similar geometric constraints where used by Kwak et al. [18] for 2D LiDAR-camera calibration. Improvements were made by extracting centreline and edge features of a V-shaped planar target. Furthermore, an interesting target adaptation to the working principle of different sensors was presented by Bormann et al. [19], where the authors proposed a method for extrinsic calibration of a 3D LiDAR and a thermal camera by expanding a planar checkerboard surface with a grid consisting of light bulbs.

D. Radar—Camera/Lidar

Radars are active sensors which, similarly to the LiDAR, emit an electromagnetic signal and determine the range of objects in the vicinity based on the returned echo. Although being frequently used in automotive applications due to their low price and robustness, extrinsic radar calibration has not gained much research attention. The existing methods are all target-based since, for all practical means and purposes, the targetless methods are hardly feasible due to limited resolution of current automotive radar systems, as the radar is virtually unable to infer the structure of the detected object and extract features such as lines or corners. Current radars have no elevation resolution while the information about the detected objects they provide contains range, azimuth angle, radar cross section (RCS) and range-rate based on the Doppler effect. Although having no elevation resolution, radars have substantial elevation FOV which makes the extrinsic calibration challenging due to the uncertainty of the measurements.

Concerning automotive radars, common operating frequencies (24 GHz and 77 GHz) result with reliable detections of conductive objects, such as plates, cylinders and corner reflectors, which are then used in intrinsic and extrinsic calibration methods [20]. Wang et al. [21] used a metal panel as the target for radar—camera calibration. They assume that all radar measurements originate from a single ground plane, thereby neglecting the 3D nature of the problem. The calibration is found by optimizing homography transformation between the ground and image plane. Contrary to [21], Sugimoto et al. [22] take into account the 3D nature of the problem. Therein, they manually search for detection intensity maximums by moving a corner reflector within the FOV. They assume that detections lie on the radar plane (zero elevation plane in the radar coordinate frame). Using these points a homography transformation is optimized between the radar and the camera. The drawback of this method is that the maximum intensity search is prone to errors, since the return intensity depends on a number of factors, e.g., target orientation and radar antenna radiation pattern which is usually designed to be as constant as possible in the FOV.

While the above described radar calibration methods provide sufficiently good results for the targeted applications, they lack the possibility to fully assess the placement of the radar with respect to other sensors. Research on 3D LiDAR—radar calibration was conducted by Peršić et al. [23]. They propose a method which estimates a 6 degrees of freedom (DOF) extrinsic calibration of a 3D LiDAR—radar pair. The method includes a target design suitable both for the LiDAR and the radar shown in Fig. [14]. It is inspired by a target constructed by Stanislas and Peynot [24] where radar performance is evaluated using a 2D LiDAR as a ground truth with a target composed of radar tube reflector and a square cardboard. Target for 3D LiDAR—radar calibration consists of a styrofoam triangle which is invisible to the radar while it has good properties for detection and localization in the LiDAR point cloud. Radar receives the echo from the trihedral corner reflector shown in Fig. [18] which has high RCS and low orientation sensitivity. In the end, extrinsic calibration parameters are found by two-step optimization. The first step is based on the reprojection error minimization while the second uses space distribution of RCS, measure of the detection intensity, to estimate variables which are not observable from the reprojection error due to the lack of radar’s vertical resolution.

III. Targetless methods

In order to maintain reliability of a perception system, sensor calibration has to be performed occasionally. Sensors displacement due to mechanical vibrations, changes of intrinsic parameters due to the variable environment conditions such as temperature and pressure, are some of the effects that can cause sensor decalibration. In such cases target-based methods are impractical and can restrict usability of the system which led to development of the targetless methods. They eliminate the need for artificial targets by using environment features to match correspondences in the sensor data. This problem is especially challenging in the heterogeneous sensor systems. It
is feasible when sensors provide enough information about the environment to extract its structure. Therefore, this techniques, described in the sequel, are mainly used in camera and LiDAR calibration.

A. Camera

Barazzetti et al. [25] proposed an approach for intrinsic camera calibration using only natural scenes. Their method uses feature extraction methods and robust estimation techniques to create correspondences between different views of the same scene. It is suitable for scenes with many features that can be uniquely described. However, repetitive textures (e.g. building facades, tiles) result in image features with similar descriptors which can be easily mismatched and thus compromise the calibration results. Although showing valuable results, authors conclude that high precision and industry applications still require target-based methods for desired calibration accuracy. Similar approach was adopted by Fraser and Stamatopoulos [26] where they showed comparable results to the target-based methods.

In order to retrieve depth information about the environment, two cameras are often rigidly connected to form a stereo vision system. Besides the intrinsic calibration of individual cameras, high precision of extrinsic calibration between the cameras is crucial for successful stereo reconstruction. Common methods for intrinsic target-based can be used to obtain the extrinsic calibration. Online targetless calibration is a greater challenge [27], [28], [29] and Ling and Shen [30] have presented an online targetless approach. It minimize epipolar errors between the image pairs based on the sparse natural features to obtain 5 degrees-of-freedom (DOF) transformation between the cameras. Similarly as in monocular vision, scale of the translation vector is unobservable. They show comparable results to the target-based methods.

B. Camera–LiDAR

Informativeness of these sensors enables inference on the structure of the environment that can be used in generating correspondences. For example, Levinson and Thrun [31] based their calibration of a 3D–LiDAR and a camera on line features detected as intensity edges in the image and depth discontinuities in the point cloud. Their method is able to detect decalibration on-the-fly and track the gradual drift of the sensor pose over time. Similar approach was adopted by Moghdam et al. [32] where they increased the robustness of their method by handling one-to-many correspondence registration by re-weighting the error metric. Gong et al. [33] proposed an approach in which they use arbitrary trihedrons commonly found in urban and indoor environments (e.g. corners of the buildings).

In addition to range measurements, LiDARs also provide information about returned signal’s intensity. Pandey et al. [34] find extrinsic calibration by maximizing the mutual information between the cameras grayscale pixel intensities and projected surface reflectivity values measured by the LiDAR. For success of their method it is important to first perform intrinsic inter-beam calibration of the surface reflectivity values. The concept of mutual information was also used by Taylor and Nieto [35]. Instead of using dense information from the point cloud, they only project selected features into the 2D LiDAR image. Additionally, they complement returned intensity information with estimated surface normal as there exist strong statistical dependence between these quantities. They show that the method is applicable to variety of LiDARs. Furthermore, mutual information between camera image and LiDAR generated reflectance image was used by Napier et al. [36] to calibrate a push-broom 2D LiDAR with camera in natural scenes. The method allows calibration of sensors without overlapping FOVs, but it requires ego-motion information.

Generation of 2D image from the LiDAR’s point cloud was also done by Scaramuzza et al. [37]. Instead of intensity, they introduced bearing angle images which are constructed from angle difference of the surface normals in the point cloud. This metric highlights environment plane intersections arising from wall corners and other similar discontinuities. However, their method requires manual registration of the correspondences. Lastly, extracting features from the environment can lead to a high number of correspondences. Scott et al. [38] claimed that not all correspondences are equally informative and that appropriate choice of scenes can improve calibration. They use normalised information distance as criteria for scene selection scheme which provided more effective and precise calibration results using a few scenes.

IV. MOTION-BASED METHODS

Motion-based calibration techniques compare ego-motion estimates from individual sensors to perform calibration. These methods can be classified as targetless methods because they also use environment features, but since they are only used to estimate ego-motion, developed methods are applicable to a wider range of sensor configurations. Only requirement is that the sensor can estimate its motion. Moreover, for the sensors such as IMU or encoder odometry, motion-based methods are the only viable solution. Additionally, they are virtually the only option for calibration of sensors whose FOVs do not overlap. Many of these methods are agnostic in terms of the sensor choice. In the sequel, some of the general methods will
be addressed, followed by methods which have some special contribution in specific sensor configurations.

A. General Methods

In their work [49], Schneider et al. have proposed a solution for extrinsic calibration of sensors based on Unscented Kalman Filter. The method is generic and can be used with sensors which provide both 3DOF and 6DOF motion estimates. However, they require time-synchronized delta poses. Furgale et al. [40], [41] have proposed a method which relaxes that constraint by using continuous-time batch estimation while simultaneously estimating both spatial and temporal calibration parameters. Brookshire and Teller [42] proposed an approach in which they explicitly model the noise via the Lie algebra yielding a constrained FIM from which they analyse motion degeneracy and proceed to singularity-free optimization procedure. FIM was also used by Maye et al. [43] to detect unobservable directions in parameter space from the available data. Additionally, they used information gain as a measure upon which they accept new measurements into the batch optimization. By reducing the total number of correspondences they created a framework for feasible online calibration.

Furthermore, Huang and Stachniss [44] have addressed the problem of high measurement noise which compromises the results of the commonly used least square optimization techniques. They improved the calibration results by adopting Gauss-Helmert optimization paradigm which jointly optimizes calibration parameters and pose observation errors.

B. Camera - IMU

Visual-inertial odometry is able to accurately estimate the 6DOF motion and it is well suited for many robotic tasks. However, it requires precise extrinsic calibration. Mirzaei and Roumeliotis [45] proposed a Kalman filter based approach for IMU – camera calibration. They based their method on estimating the camera motion using checkerboard pattern. Through the observability analysis based on the Lie derivatives rank criterion they showed that it is necessary to excite at least two rotational axis of the system to make the calibration parameters observable. Kelly and Sukhatme [46] have continued on the previous research by discarding the checkerboard and using environment features for the visual odometry. They showed that additional two translational axes need to be excited in order to resolve camera scale issue and make the calibration parameters observable. Keivan and Sibley [47] proposed a SLAM solution which is able to detect system decalibration and perform calibration online. Compared to the previously described methods, they are additionally able to recalibrate camera’s intrinsic parameters. They evaluated the method on an experiment where they doubled the focal length on-the-fly.

C. Hand – Eye calibration

Many robotic applications involve manipulator equipped with a wrist-mounted sensor such as camera. Calibration between the end of the manipulator and the perceptive sensor is crucial in these applications. This problem is often referred to as an \( AX = XB \) problem due to the emerging equation that needs to be solved (\( A \) and \( B \) represent manipulator and sensor movement, respectively, while \( X \) represents the extrinsic calibration). It has been studied for more than three decades [48] and many solutions exist.

Some of the recent advances in the field have dealt with the problem of unknown correspondences caused by asynchronous sensors or missing detections [49], [50]. Furthermore, a general solution for motion-based extrinsic and temporal calibration was given by Taylor and Nieto [51]. The solution is based upon the framework of \( AX =XA \) problem which is further enhanced by targetless methods if the sensor types and overlaps allow such refinements. It was evaluated through calibration of several vehicle-mounted sensor configurations.

V. Conclusion

Robotic systems are increasingly adopting heterogeneous multi-sensor approaches for which correct sensors calibration is a necessity, both intrinsic and extrinsic. Nowadays, robotic applications use a wide spectrum of sensors for which many calibration methods have been developed. While some solve specific sensor configurations, others aim at more general solutions. In this overview, some of the representative methods were presented for commonly used sensors such as cameras, LiDARs, IMUs and radars. Offline calibration has been studied for a long time, while the field of online calibration has been given a special focus only as of lately. Online detection and correction of system decalibration is crucial for robustness of any autonomous system which makes it one of the prerequisites for reliable long-term autonomy.

REFERENCES

[1] D. Scaramuzza, “Omnidirectional vision: from calibration to robot motion estimation,” ETH Zurich, PhD Thesis, no. 17635, p. 189, 2008.
[2] R. Y. Tsai, “A Versatile Camera Calibration Technique for High-Accuracy 3D Machine Vision Metrology Using Off-the-Shelf TV Cameras and Lenses,” IEEE Journal on Robotics and Automation, vol. 3, no. 4, pp. 323–344, 1987.
[3] Z. Zhang, “A Flexible New Technique for Camera Calibration (Technical Report),” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 22, no. 11, pp. 1330–1343, 2002.
[4] J. Heikkila and O. Silven, “A four-step camera calibration procedure with implicit image correction,” IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1106–1112, 1997.
[5] J. Mallon and P. F. Whelan, “Which pattern? Biasing aspects of planar calibration patterns and detection methods,” Pattern Recognition Letters, vol. 28, no. 8, pp. 921–930, 2007.
[6] B. Li, L. Heng, K. Koser, and M. Pollefeys, “A multiple-camera system calibration toolbox using a feature descriptor-based calibration pattern,” IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pp. 1301–1307, 2013.
[7] G. Atanacio-Jiménez, J. -J. -J. González-Barbosa, J. B. -J. Hurtado-Ramos, F. -J. J. Ornelas-Rodríguez, H. -J. Jiménez-Hernández, T. -J. García-Ramirez, and R. -J. González-Barbosa, “LIDAR Velodyne HDL-64E Calibration Using Pattern Planes,” International Journal of Advanced Robotic Systems, vol. 8, no. 5, pp. 70–82, 2011.
[8] N. Muhammad and S. Lacroix, “Calibration of a rotating multi-beam Lidar,” IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pp. 5648–5653, 2010.
[9] E. Fernández-Moral, J. González-Jiménez, and V. Arevalo, “Extrinsic calibration of 2D laser range finders from perpendicular plane observations,” International Journal of Robotics Research, vol. 34, no. 11, pp. 1401–1417, 2015.
