University of Cape Town

Spacial and Temporal Calibration of Multi-Sensor Systems

Michael Katsoulis

KTSMIC005

24 August 2020

Plagerism Declaration

1. I know that plagiarism is wrong. Plagiarism is to use another's work and pretend that it is one's own.

2. I have used the IEEE convention for citation and referencing. Each contribution to, and quotation in, this final year project report from the work(s) of other people, has been attributed and has been cited and referenced.

3. This project report is my own work.

4. I have not allowed, and will not allow, anyone to copy my work with the intention of passing it off as their own work or part thereof.

<div align="right">

Michael Katsoulis
24 August 2020

</div>

1 Literature Review

1.1 Introduction to the problem

In the modern world, the need for autonomous vehicles is growing daily. If common vehicles could become autonomous, it would improve the quality of human life, provide advances in science and open new opportunities. One of the problems with trying to make vehicles autonomous is that they are normally different shapes and sizes. Vehicles are sometimes modified or have accidents which would deform them in some way. This leads to the sensors on each vehicle being in different locations.

The goal of this project is to investigate the current methods of determining the spacial and temporal offsets between sensors on a vehicle automatically. This would allow a decrease in calibration time, ease of manufacture as well as allow vehicles to continue to operate in the case of an unforeseen incident which moves a sensor.

This would save costs as the same algorithm on multiple models of vehicle. Drones could be built which can reconfigure their geometry mid-flight or ,if damaged on a mission, recalibrate and still get home safely.

1.2 How do we get each form of error and why is it important to calibrate a system properly?

Spatial errors occur when a sensor is not where the vehicle thinks it is. When a system has not been spatially calibrated properly, the readings for each sensor cannot be transformed and fitted to one another accurately. This spatial calibration error could be because the sensors have been moved from the location where they were calibrated in or it could be that the positions of the sensors are simply unknown. This results in misalignment of the sensor readings and therefore, the vehicle's world view is distorted and inaccurate.

Temporal errors are caused by a wide variety of factors that influence the time at which the readings are taken from all the sensors. It is common for cameras and lidar sensors to have their own onboard clocks which may not be perfectly synchronized. There may be filters on the readings, different lengths of communication channels and other factors too. The reason it is important to have a temporally calibrated system is so that there is no changes in the data between readings. This way, the vehicle knows that the data that is being used to make decisions reflects what was happening at the same instant in time for all the sensors. There have been attempts to fix the temporal calibration problem through clock synchronization and time stamping sensor readings but this is not a long term solution in case the clocks go out of sync while the vehicle is in the field. It is also an expensive way of solving a problem which could be solved with an algorithm and some code.

1.3 Existing Literature

From the literature that has been reviewed, there are two main methods for obtaining targets in each sensor to use as references during the calibration and evaluation stages. The first method uses markers that are of known shape, texture and that are placed into what is often a known, simple environment. This is a more hands-on, involved approach but it is often simpler to find methods of calibration. The markers are normally planar with patterns, holes and other features on them to help make them easy to distinguish from the environment. The other method is markerless calibration. In these cases, the goal is to use the information in the natural environment around the vehicle in order to calibrate the sensors. This often involves identifying and removing targets which are moving relative to the global frame(dynamic targets) and trying to isolate those which are static. These are then used to determine the calibration parameters.
Markerless calibration is the newest form and is the focus of most research in an attempt to find an elegant solution.

The most common sensors used in autonomous vehicles are lidar and stereo cameras. The calibration of these sensors poses two problems as stated by Fu,B et al [1]:

- The first is that there is a common field of view constraint. Both the lidar and the camera need to be able to see the same objects.

- The second is that two optimizations need to happen simultaneously in both spacial and temporal calibration. This makes the optimization more challenging.

In order to simplify the problems described above, the solution in [1] was to remain stationary while capturing a 3D image of the surrounding enviroment and comparing that to the lidar point map. There is not a temporal calibration that needs to occur as the scene is stationary. This meant that the researchers could simply solve the spatial calibration on it's own. By having a 360° lidar and camera image, they also removed the common field of view constraint as both images were showing the same frames. A limitation to this research was that they used a target based calibration with checkerboards which had to be set up in a room surrounding the vehicle. They also did not remove dynamic objects in the field of view by taking more than one 360° image and removing the items that change. This leaves two areas for improvement in future research.

In an attempt to make marker based lidar - camera calibration easier to implement, Z. Pusztai and L. Hajder. [3] developed a way to use ordinary cardboard boxes as the calibration targets. This meant that calibration of the system spatially became a simpler, less speciallised task. Boxes of known size were placed in a room in arbitrary configurations. Then the lidar point maps were used to find planes in the environment and from there, isolate the potential cardboard boxes. They also developed a way of predicting where the corners of the boxes were if they were between two points. This became particularly useful when the targets were further from the vehicle and the lidar point map was starting to become spread out with distance. All that was left was to find the boxes in the camera images and map the images from the lidar onto the camera images. They used the dimensions of the boxes to then finish calibrating the system spatially. Similarly to the method proposed in [1], the problem with this technique is that it requires supervision and cannot be done on the fly in an unknown environment. But it provides some good ideas for showing there is no need for complex planar targets by achieving the same accuracy with simple boxes.

In the paper by B. Nagy and C. Benedek [2], the researchers used the motion of a car to synthesize a 3D image of the surrounding enviroment from a series of consecutive images. This 3D, structure from motion image was then used to synthesize a dense point cloud which could be used to map to the lidar's point cloud. To find the transformation between the two point clouds, the researchers then found potential objects in the image and mapped these to the lidar point cloud. A finer alignment was then done using a Non-Uniform Rational B-Spline which allowed for flexibility in the shape of the point clouds. This allowed them to account for elipsoid-shape and other types of distortion due to the motion of the vehicle. Using movement to calibrate the camera-lidar system removes the dependency on high fidelity GPS and inertial measurement units. Another benefit to this technique of generating synthetic point clouds and mapping those to the lidar point clouds is that there was no need for hardware time synchronization. A downside to this method is that there needs to be a fairly powerful GPU on board if this were to be done on the vehicle. The researchers used a Nvidia GTX1080Ti GPU with 11 GB RAM and it took close to 3 minutes to complete the whole calibration from start to finish. This is a decent time but considering the hardware used, it is unlikely that embedded systems could perform the same steps in similar times.

Another method that has been used is a spin on the hand-eye calibration used when calibrating robot arms. In hand-eye calibration, the spatial transformation from the robot "hand" to the camera is not known. The robot then undergoes a series of known transformations which we will call A and B. Then at each pose, the position of the hand-eye can be given by the combination of the known transformation and the spatial transformation between the hand and eye (X).
So using the relationship $AX = XB$, we can determine X. In their paper, Motion-Based Calibration of Multimodal Sensor Extrinsic and Timing Offset Estimation [4], the authors Z. Taylor and J. Nieto. showed that a similar algorithm could be used to determine the spatial offset between a camera, lidar and navigation sensor. This meant that the authors could get a coarse calibration by estimating the motion of

each of the sensors from what it has been recording before then using a similar approach to the hand-eye problem to determine the coarse calibration parameters.

The researchers then used appearance based metrics to extract objects from the environment where the sensors had an overlapping field of view. They used these objects to fine tune the alignment of the sensors and to confirm that the coarse alignment was correct.

The methods used resulted in no need for calibration targets while also simplifying the coarse calibration step. A problem with it however is that some form of a navigational sensor is needed. It did however provide a very easy way to use the same technique for different vehicles which the authors demonstrated.They also showed how the length of the data used during calibration played a role in the final accuracy. It is an algorithm that seems to be of very good practical use and from the report also seems to be the easiest to implement.

1.4 Reflection on the related work

The literature has influenced the ideas for this thesis by providing a clearer description of the problem. It has also spurred several ideas on how to tackle the challenge. Markerless calibration is the challenge for the future although it is a significantly more challenging task. In all markerless calibration methods, there is a step where objects are identified and those that are static objects are removed. This seems to be the biggest challenge as finding ways to correlate lidar and camera images is not always straightforward, especially when the target objects have a visual texture which makes it harder to determine the edges of the shape.

The Structure from Motion approach to synthesize a point cloud from images seems like a good way to deal with the texturing problem while also seperating the spatial and temporal aspects of the calibration into two seperate challenges. The hand-eye technique however seems to be much lighter on processing power but it requires a high fedelity IMU or navigational sensor. To implement it in the real world, there would need to be some way of identifying a target but even then, a calibration routine would need to be setup on the vehicle such that known transformations can be choreographed.

It seems that the biggest challenge for moving towards targetless calibration is still the identifying of static targets in both camera and lidar images while filtering out other objects which cannot be used for calibration. It would be interesting to see if simpler, less hardware demanding calibration techniques could be implemented once this step is done as well as investigating whether it is a task that would be accomplished more easily using OpenCV and algorithms or if it would be more efficient when it is implemented by a convolutional neural network.

References

[1] B. Fu, Y. Wang, X. Ding, Y. Jiao, L. Tang, and R. Xiong. LiDAR-Camera Calibration Under Arbitrary Configurations: Observability and Methods. *IEEE Transactions on Instrumentation and Measurement*, 69(6):3089–3102, June 2020.

[2] B. Nagy and C. Benedek. On-the-Fly Camera and Lidar Calibration. *Remote Sensing*, 12(7):1137, Apr. 2020.

[3] Z. Pusztai and L. Hajder. Accurate Calibration of LiDAR-Camera Systems Using Ordinary Boxes. In *2017 IEEE International Conference on Computer Vision Workshops (ICCVW)*, pages 394–402, Venice, Italy, Oct. 2017. IEEE.

[4] Z. Taylor and J. Nieto. Motion-Based Calibration of Multimodal Sensor Extrinsics and Timing Offset Estimation. *IEEE Transactions on Robotics*, 32(5):1215–1229, Oct. 2016.