An Online Intelligent Method to Calibrate Radar and Camera Sensors for Data Fusing

Ming Liu, Dongdong Li, Qiankun Li, Wei Lu and Jun Yin
Zhejiang Dahua Technology Co., Ltd. 1199 Binan Road, Binjiang District, Hangzhou, Zhejiang Province, China
Email: aiguzzhefei@126.com

Abstract. This paper introduces an intelligent method to calibrate radar and camera sensors for data fusing. Camera can recognize pedestrians, non-motor vehicles, motor vehicles high accuracy by advanced deep learning neural networks. But it's hard to acquire the target’s spatial position and speed information simply relying on image. For the radar sensor, it’s easy to acquire the target’s spatial position and speed information by itself, but it cannot recognize target. In other words, the advantage of radar is the disadvantage of camera, and the advantage of camera is the disadvantage of radar. So combining these two complementary sensors is valid and necessary for solving the problem: Who is doing what. Then the first and foremost thing is to calibrate the radar and camera sensors. Calibration includes time synchronization and space calibration, which will seriously affect the sensors data fusing performance. This paper emphasizes on spatial calibration. Traditional methods (based on the principle of geometric projection or four-point calibration method) cannot acquire enough high calibration accuracy in the general scenarios with low time cost or easy operation. So this paper introduces an online intelligent method to calibrate radar and camera sensor without human assistance operation in the general scenarios. Firstly, we introduce what is spatial calibration. Secondly, we introduce two conventional methods and their defects. And then we introduce the online intelligent calibration method. At last we compare three calibration methods with the actual data to further explain the advance of our new method.

1. Introduction
With the development of Artificial Intelligence [1, 2], the intelligent surveillance system has been widely used in various aspects of people’s lives. Target detection, tracking, behaviour recognition and detection are the key technology of intelligent surveillance system, and they have become one of the hot areas studied by the specialists and scholars [3, 4]. For example, in the intelligent transportation field, instead of human, more and more cameras or radars detect the dangerous behaviours or acts in violation of regulations. The intelligent devices not only improve work efficiency, but also reduce hassles and costs.

As for the intelligent transportation field, there is an import issue: Who is doing what? That is to say, we should save the evidence that can determine the identification and behaviours [5-7]. The intelligent camera can solve the issue “Who” by the deep learning neural network. But there are some difficulties to recognize abnormal behaviours, such as over speeding, pressing line, S-curve driving. The Radar can solve the issue “Doing what” by itself attribution. Radar device can acquire the spatial position, speed, RCS (Radar Cross Section) information of targets real timely. So it is easy to determine whether one target is over speeding, S-curve driving, or other abnormal behaviours. But
radar doesn’t know that target is who. With above analysis, the advantage of radar is the disadvantage of camera, and the advantage of camera is the disadvantage of radar. In theory, a system including radar and camera data fusing can solve the issue: Who is doing what.

In practice, there are some problems to realize radar and camera data fusing, such as time synchronization and spatial calibration, identity fusion, state estimation, and so on. One of the first and foremost things is spatial calibration. The spatial calibration means that you can get measurements in an image basing on the known radar measurements, or get radar measurements using real spatial units (cm, m, etc.), basing on the known image measurements. That is to say, we construct a mapping function between radar coordinate system and camera coordinate system (R-C spatial calibration in short).

No matter what R-C spatial calibration methods, currently, the most common studies are the solutions how to more efficiently calibrate the two-coordinate system offline or correct manual. However few studies tackle the problems that calibrate or automatically correct online, which is the main focus of this paper. In the multi-sensors calibration field, which provides the motivation for our method, R-C spatial calibration is related to device structure, camera intrinsic parameters and external parameters. That’s to say, it takes lots of effort and time to solve many parameter values, and if the device has been moved, we should recalibrate again. Therefore, the online automatic and intelligent calibration method is significant for reducing development and maintenance costs.

The structure of this paper is shown as follow. Section II gives a brief description of R-C spatial calibration, including principle and key points. Section III introduces two conventional calibration methods and their defects. We propose our new method in the section IV. The simulation and analyses in Section V show a notable improvement in the performance of our new method comparing with the two conventional calibration methods. Finally, Section VI summarizes the main conclusions and results of the paper.

2. Introduction for R-C Spatial Calibration
The radar-camera device is shown in figure 1. The radar sensor and camera sensor are installed together to detect and tracking targets, as shown in figure 2. The left figure is the camera monitor result, and the right figure is the radar monitor result. In the left figure, it shows ID of every vehicle as same as in the right figure. But we should concern their differences specially: 1) a target in the left figure is rectangular area but the target in the right figure is one point. 2) The left figure is perspective, meaning that near looks large and far looks small. The scale of right figure is equal. 3) In the left figure, we can acquire target’s category and position with pixels, but cannot acquire the position and velocity in the camera coordinate system. In the right figure, we can acquire target position and velocity in the radar coordinate system, but cannot acquire target’s category.

Figure 3 is the sketch map of figure 2. Left figure is camera image, and right figure is radar image. Every sensor works independently. As for the camera sensor, it can get below target’s measurements: category, and target box. As for the right sensor, it can get \([x, vx, y, vy]\) in the radar coordinate system. If Target 1 is associated with target A, it means one vehicle is coming, its speed is 10m/s and position is \([x = 3\text{m}, y = 150\text{m}]\). Target 1(A) observes traffic regulations base on that its trajectory is straight line and it stays driving inside in one lane. But if target 1 is associated with target D, the vehicle is dangerous, because its trajectory bends significantly and occupies two lanes. If necessary, we should record the plates of target 1(D), and warn the driver.

With above examples, we have understood the R-C calibration meaning: providing the mapping function that can correctly associate camera targets with radar targets. The mapping function is shown below.

A 2D target point in radar coordinate system is denoted by \(X = [x, y]^T\) and a 2D target point in camera coordinate system is denoted by \(U = [u, v]^T\). The relationship between a point in radar coordinate and a point in camera coordinate is given by

\[
U = F(X)
\]
where $F$ and $G$ is multi-input-multi-output functions. The task of spatial calibration is finding or solving $F$ and $G$.

$$X = G(U)$$  \hspace{1cm} (2)

$$F = G^{-1}$$  \hspace{1cm} (3)

$$G = F^{-1}$$  \hspace{1cm} (4)

Figure 1. The radar-camera device. Camera sensor and radar sensor are installed together.

Figure 2. The monitor scenario.

As for figure 2, left figure is the image of camera sensor, and the right figure is the image of radar. In the left camera figure, vehicles have been marked by box, meaning that vehicles have been recognized by AI algorithm. In the right radar figure, the track of every target has been shown with current position, and if we want, we can show the position and velocity at each moment.
As for figure 3, left figure is camera image, and right figure is radar image. There are three targets in left figure: target 1 is vehicle, target 2 is human, and target 3 is non-motor vehicle. There are four targets in right figure; we can capture there are four targets, but do not known what they are. Maybe Target A of right figure is target 1 of left figure or Target A is target 3.

3. Conventional Methods and Their Defects

Many papers have researched the spatial calibration [8, 9], and the representative is modified Zhang calibration method and modified four points calibration method.

The modified Zhang calibration method translates a point of radar coordinate system to a point of world coordinate system, and then translates the point of world coordinate system to a pixel point of camera coordinate system, as shown in figure 4.

As for the step 1, we should get the $x, y, z, \theta, \phi, \psi$ value of radar sensor in the world coordinate, and as for the step 2, we also get the intrinsic and external parameters of a camera. The intrinsic parameters include principal point, focus length, and image size. The external parameters include spatial positions of camera sensor in the world coordinate system and three rotation angles. Without loss of generality, after the world coordinate system has assumed, we have

$$[R, T] = \begin{bmatrix} R & T \\ 0 & 1 \end{bmatrix}$$  \hspace{1cm} (5)$$

$$R = \begin{bmatrix} \cos \theta & -\sin \theta & 0 \\ \sin \theta & \cos \theta & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos \phi & \sin \phi \\ 0 & -\sin \phi & \cos \phi \end{bmatrix} \begin{bmatrix} \cos \psi & 0 & -\sin \psi \\ 0 & 1 & 0 \\ \sin \psi & 0 & \cos \psi \end{bmatrix}$$  \hspace{1cm} (6)$$

$$T = [x, y, z]^T$$  \hspace{1cm} (7)
\[ A = \begin{bmatrix} f_x & 0 & u_0 & 0 \\ 0 & f_y & v_0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \]  
\[ F = A^* R, T \]  
\[ F = A^* R, T, z \]  

where \( R \) is rotate matrix, \( T \) is translation vector, and \( A \) is camera coordinate system. Equation (9) is the mapping function. Its advantage is that every parameter has definite physical meaning. But it is too hard to get the correct value of every parameter, and we should calibrate and recalibrate every device, which seriously affect the popularizing of market.

The modified four points calibration method constructs the mapping function from the radar coordinate system to cameral coordinate system. It needs at least four-pair points that are corresponding in the radar and camera coordinate system. Then it is easy to get the mapping function, which is not necessary to show here again. But as you know, it is difficult to find perfect four-pair points: (1) the detect range of radar is larger than 100m generally, so it requires more points to solve the mapping function. (2) Any significant target in the camera image should occupy some area, so select one point in the area will create error, and in order to reducing the error, we need more paired points; (3) If somebody has moved devices, we should recalibrate the devices.

With above analysis, for the conventional calibration methods, their defects have seriously affect volume production, popularization and widespread use. So we need one more intelligent method to calibrate the radar-camera device.

4. Online Intelligent Calibration Method

As for figure 2, we provide another perspective to analyse the mapping relationship between radar and camera coordinate system, as shown in figure 5. Area 1 is the blind spot of device, meaning that both radar and camera cannot detect the targets in area 1. Area 2 is detected only by the camera sensor, area 3 is detected by both camera and radar sensor, and area 4 is detected only by the radar sensor.

For any target passing the radar or camera monitor area, we have

\[ R^x = [X_1, X_2, \ldots, X_n] \times X = [x, y]^T \]  
\[ P^c = [U_{i1}, U_{i2}, \ldots, U_{in}] \times U = [u, v]^T \]  
\[ S^{RC} = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{bmatrix} \begin{bmatrix} U_{i1} \\ U_{i2} \\ \vdots \\ U_{in} \end{bmatrix} \]  

where, \( P^R \) is the spatial position sequence of radar sensor when a target is passing the area 2 and 3, and \( P^C \) is the pixel sequence of camera sensor when a target is passing the area 1 and 2. Both \( t_i \) (1 ≤ \( i \) ≤ \( n \)) and \( k_j \) (1 ≤ \( j \) ≤ \( m \)) are the time stamp. If the time synchronization has been solved, then we can get paired measurements \( S^{RC} \) from radar and camera sensors when a target passes the area 2, which provides possibilities for online intelligent R-C calibration. In theory, we have solved the issue of collecting data online.

\[ \text{Figure 5. The plan of radar-camera monitor scenario.} \]
The next key step is how to design the mapping function. As we all know, the deep neural network can fit any multi-input-multi-output nonlinear function [10]. Based on this theory, the neural network is used to construct the mapping function, as shown in figure 6. Both the input and output layer of neural network include two nodes. It has five hidden layers. The architecture of whole network is $2 \rightarrow 5 \rightarrow 10 \rightarrow 20 \rightarrow 10 \rightarrow 5 \rightarrow 2$. The activation function is ReLU, and the loss function is MSE. Optimizer is stochastic gradient descent with a batch size of 128 examples, learning rate of 0.01, momentum of 0.9, and weight decay of 0.0006.

The third key step is how to train and test the neural network with collected data. Special weight is given to that the data is collected online and the neural network is trained and tested online. So large memory device is necessary because all collected data should be remained, and the network is trained and tested repeatedly in order to obtaining stable network. We train the network for roughly 100 cycles by the training data set of 10000 pairs, and test the network by the test data set of 2000 pairs.

We summarize the major steps of the online intelligent calibration method, whose flow diagram is shown in figure 7.

![Figure 6. The neural network of intelligent calibration method.](image)

Step 1: Initializing. We should initialize hyper-parameters of the network, such as batch size, learning ratio, momentum, and weight decay. Training cycles, the size of train data set and test data set also are initialized at this step. Skip to step 2.

Step 2: Recognize environment. The purpose of this step is to judge when we begin to collect data. That is to say, ideal environment makes unambiguous pairs to ensure the correctness of collected data set, as shown figure 8. Both radar and camera sensor only detect one target, so the target 1 must be associated with target A. At this step, the target tracking algorithm of radar sensor can provide the target trajectory information real-time (for example, target A), and the artificial intelligent target detecting algorithm of camera sensor can provide the target information real-time (for example, target 1). Skip to step 3.

Step 3: Is ideal environment. If current environment is ideal environment, Skip to step 4 to collect data. Otherwise, skip to step 2, wait for next scan, and recognize environment. With the actual experience and business requirements, most of environment is not ideal environment, but we have enough time to collect data.

Step 4: Collect data. Although it is ideal environment, there is still plenty of work to be done: attaching time stamp for radar and camera data and selecting one point in the target area of the camera image. Skip to step 5.
Figure 7. Flow diagram.

Step 5: Time synchronizing. Based on the time stamp, we get the paired radar and camera data by time synchronizing process. Skip to step 6.

Step 6: Save data. The paired data are saved to the non-volatile storage medium. Skip to step 7.
Step 7: Quality and quantity judge. We should judge the quality and quantity of data set in order to get better training result. Quantity requirement means that we should collect enough data. Quality requirement means that collected data distribute balanced and uniformly in area 3 of figure 5. Skip to step 8.

Step 8: Is satisfying. If quality and quantity of collected data have been satisfied, skip to step 9. Otherwise, skip to step 12.

Step 9: Training online. Firstly, collected data are divided into train data set and test data set. Secondly, the train data are input into neural network model for training. Skip to step 10.

Step 10: Testing online. After training, we select the best trained network based on the loss function by testing on test data. Skip to step 11.

Step 11: Pass. If testing accuracy is higher than threshold, it means we have obtained efficient network, and skip to step 14. Otherwise, skip to step 13.

Step 12: Delete redundant data. In view of this situation, generally, it is triggered by unbalanced and ununiformed data. So we should delete some data randomly. Skip to step 2.

Step 13: Modifying data. In view of this situation, generally it is trigged by a lack of quantitative data or wrong paired data. We can add more paired data or recollect paired data. Skip to step 2.

Step 14: Save network. We have obtained efficient network. Then the network should be saved and used to data fusing. Skip to step 15.

Step 15: End.

5. Simulation and Analyses
In this section, some simulations and analyses will be made to verify the performance of our new online intelligent R-C calibration method. The evaluation index is the MSE

\[
MSE = \frac{1}{m} \sum_{i=1}^{m} [(x_i^r - x_i^p)^2 + (y_i^r - y_i^p)^2]
\]  

(13)

where \(x_i^r\) and \(y_i^r\) is the real measurement of a pedestrian at time \(i\), \(x_i^p\) and \(y_i^p\) is the predict value of the responding calibration method. Here there are three methods: modified Zhang calibration method, modified four points calibration method and online intelligent calibration method. The result is shown figure 9.

Figure 9. Simulation result.
The red, green, blue line represents the error of modified Zhang, modified four points, the new online intelligent calibration method, respectively.

In the figure 9a, if we set the right x, y, z, heading, pitch, roll, intrinsic parameters, and position of the four points, the two conventional methods can get wonderful effect as same as the intelligent calibration method. In figure 9b, the measurement of electronic gyroscope and altimeter is added by error, so compared with figure 9a, the error has increased for the two conventional methods, but the performance of the new method doesn’t decrease. In the figure 9c, during the testing, we have moved device at the third stage, then the error of three methods all increases. But some time later, the new method has self-corrected. The other two conventional methods don’t self-correct. In figure 9d, if the distorted camera is used to test, the result indicates that the new method perfects better than the two conventional methods, because the neural network can fit nonlinear functions.

6. Conclusions
Spatial calibration is significant for data fusing of multi-sensors. As for the radar-camera spatial calibration, the conventional methods have obvious defects: depending on other sensors (gyroscope and altimeter), low anti-interference and anti-distortion ability, and lack of self-correction ability. Then we propose a new online intelligent calibration method. It improves the calibration performance, and self-correct if the environment has changed, meanwhile reduces the dependence on manual operation. Its performance is illustrated by the simulations and analyses with actual data. But there are some problems for our new method: it cost lots of computing resources and storage. So in the future, research focuses on how to decrease computing resources and storage.

References
[1] Krizhevsky A, Sutskever I and Hinton G 2012 Imagenet classification with deep convolutional neural networks Proceedings of the Advances in Neural Information Processing Systems 1097-1105.
[2] Wanli O, Xingyu Z and Xiaogang W 2015 Single-pedestrian detection aided by two-pedestrian detection [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence 37 (9) 1875-1889.
[3] Ricardo C and Aycard O 2016 Multiple sensor fusion and classification for moving object detection and tracking IEEE Transactions on Intelligent Transportation Systems 17 (2) 525-534.
[4] Shaoqing R, Kaiming H, Girshick R and Jian S 2017 Faster R-CNN: Towards real-time object detection with region proposal networks IEEE Transactions on Pattern Analysis and Machine Intelligence 39 (6) 1137-1149.
[5] Xinchen L, Wu L and Mei T 2018 Progressive and multimodal vehicle re-identification for large-scale urban surveillance IEEE Transactions on Multimedia 20 (3) 645-658.
[6] Sun C, Arr S and Ramachandran P 2004 Vehicle reidentification using multidetector fusion IEEE Transactions on Intelligent Transportation Systems 5 (3) 155-164.
[7] Shuiwang J, Wei X and Ming Y 2013 3D convolutional neural networks for human action recognition IEEE Transactions on Pattern Analysis and Machine Intelligence 35 (1) 221-231.
[8] Dubská M, Herout A and Sochor J 2014 Automatic camera calibration for traffic understanding British Machine Vision Conference p 8.
[9] Dubská M, Herout A and Jurá R 2015 Fully automatic roadside camera calibration for traffic surveillance IEEE Transactions on Intelligent Transportation System 16 (3) 1162-1171.
[10] Rupesh S, Greff K and Jurgen S 2017 Training very deep networks Proceedings of the Conference and Workshop on Neural Information Processing Systems pp 2377-2385.