2018 15th International Conference on
Control, Automation, Robotics and Vision (ICARCV)
Singapore, November 18-21, 2018

A Comparative Study on Camera-Radar Calibration Methods

Jiyong Oh, Ki-Seok Kim, Miryong Park, and Sungho Kim

Abstract— **Camera-radar fusion has been applied in obstacle detection or moving object tracking for autonomous vehicles and advanced driver assistance systems. When utilizing multiple sensors, their calibration is not only essential but also important because it gives great impacts on subsequent procedures. Nonetheless, camera-radar calibration methods have not been compared in the literature qualitatively or quantitatively. In this paper, we compare three types of the calibration methods presented in previous studies on fusing camera and radar in terms of calibration accuracy. Especially, the comparison is conducted in the situation of varying the number of radar-image data pairs used in their calibration. Experimental results show that one type of the methods is not appropriated to the camera-radar calibration, and the methods belonging to the other types provide quite similar accuracy.**

I. Introduction

After three competitions of DARPA Grand Challenges in 2004, 2005, and 2007, researches on autonomous vehicles or self-driving cars have dramatically increased in both academic and industrial fields. The related techniques have been rapidly developed so that advanced driver assistance systems (ADAS) such as adaptive cruise control and automatic emergency braking have not already been surprised any more in these days. Autonomous vehicle or ADAS should be aware of surrounding obstacles and environment to appropriately react to various situations for safety, and it is necessary to utilize various sensors for those perception. Among the sensors, lidar and radar are known as the most popular sensors together with camera [1], [2]. Although lidar is better than radar in terms of accuracy and resolution, radar is more robust to weather condition and surface occlusion by foreign substance such as water drop than lidar. Thus, the fusion of camera and radar has been studied steadily [3]–[12]. Most of the studies on the camera-radar fusion addressed obstacle detection or moving object detection and tracking, but it was also applied in automatic traffic surveillance [7] and three dimensional reconstruction [11], [12].

In camera-radar fusion techniques, the calibration between the two sensors with different types is not only essential but also important because it gives great impacts on subsequent procedures. In this study, we deal with the calibration to estimate an transformation mapping a real location of an object provided by a radar into the coordinate on image provided by a camera. The camera-radar calibration presented in the literature can be divided into three categories. The methods in the first category are based on affine transformation, which is obtained by the least squares using pseudo inverse [8]. On the other hand, the methods in the second category are based on projective transformation. In these methods, the transformation can be regard as a two dimensional homography [13] from radar plane to image plane, and it is obtained using the least squares using singular value decomposition. Many of the studies based on the camera-radar fusion used one of those calibration methods [3], [7], [9], [10]. The methods in the last category are based on the transformation that maps a point in a three dimensional space to a point on image plane of camera. Those methods are similar to the extrinsic calibration for camera pose estimation. Related to the methods, it was mentioned in [5], [6] that a coordinate with respect to radar was projected to the corresponding image coordinate using intrinsic and extrinsic parameters. In [11] and [12], a similar method based on the same transformation was also proposed as an intermediate procedure of reconstructing three dimensional information based on the camera-radar fusion. However, to our best knowledge, any study did not compare the camera-radar calibration methods quantitatively or qualitatively. This paper focuses on comparing the camera-radar methods presented in the literature, and we describe the methods in detail as well so that this paper provides a guide of the camera-radar calibration to readers.

This paper is organized as follows. In the next section, the calibration methods are described, and they are compared in terms of accuracy in Sec. III. The last section concludes this paper with future work.

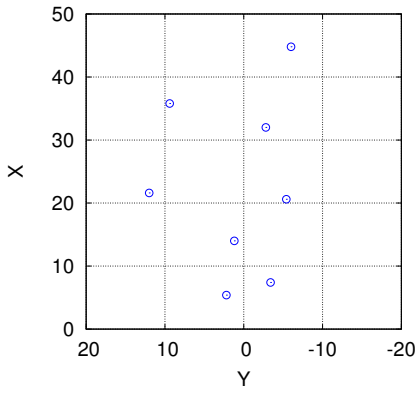
II. Calibration methods

Let us consider N radar targets with their coordinates of $\{\mathbf{p}_i = [x_i, y_i]^T\}_{i=1}^N$ and their corresponding image coordinates $\{\mathbf{q}_i = [u_i, v_i]^T\}_{i=1}^N$, which are captured in the common field of view of the radar and the camera. For example, eight targets detected by a radar are shown in Fig. 1a where the targets are denoted as blue circles, and they are also captured by a camera as in Fig. 1b where their coordinates are denoted as red crosses. Camera-radar calibration in this study is to estimate a transformation mapping the radar targets of Fig. 1a on the image coordinates of 1b given a pairs of $\{\mathbf{p}_i\}_{i=1}^N$ and $\{\mathbf{q}_i\}_{i=1}^N$ and the one-to-one correspondence between

This work was not supported by Electronics and Telecommunications Research Institute (ETRI) grant funded by the Korean government [18ZD1120, Development of ICT Convergence Technology for Daegu-GyeongBuk Regional Industry

J. Oh, K.-S. Kim, and M. Park are with the Smart Vehicle Section, Daegu-Gyeongbuk Research Center, Electronics and Telecommunications Research Institute (ETRI), Daegu, Korea. e-mail: {jiyongoh, kskim, mrpark}@etri.re.kr

S. Kim is with the Department of Electronic Engineering, Yeungnam University, Gyeongsan-si, Gyeongangbuk-do, Korea. email: sunghokim@ynu.ac.kr



(a) Radar targets

(b) An image

Fig. 1: A pair of radar targets and image captured at the same time.

them. In this study, we consider three types of the calibration methods and describe them in this section.

• *Pseudo inverse* (PI) [8]: In this method, the coordinate system of the radar is defined as in Fig. 1a and the coordinate system of the camera is conventionally defined: the column of the image is associated with $\mathbf{U}$ and the row of the image is associated with $\mathbf{V}$, then the transformation is modeled as follows.

$$\widetilde{\mathbf{q}} = \mathbf{H}\widetilde{\mathbf{p}},$$

where $\widetilde{\mathbf{p}}$ and $\widetilde{\mathbf{q}}$ are the homogeneous coordinates of $\mathbf{p}$ and $\mathbf{q}$, respectively. Here, $\mathbf{H}$ is assumed to be a two dimensional affine transformation, i.e.,

$$\mathbf{H} = \begin{bmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ 0 & 0 & 1 \end{bmatrix} \quad (6 \text{ Dof}).$$

Since this problem has six unknown variables, we need $N \geq 4$ pairs of $\{\mathbf{p}_i\}_{i=1}^N$ and $\{\mathbf{q}_i\}_{i=1}^N$ in order to compute $\mathbf{H}$ [8]. Using a given N pairs of $\{\mathbf{p}_i, \mathbf{q}_i\}_{i=1}^N$, an optimization problem can be formulated to minimize the projection error on the image plane as follows.

$$\arg\min_{\mathbf{H}} \sum_{i=1}^{N} d(\mathbf{H}\widetilde{\mathbf{p}}_i, \widetilde{\mathbf{q}}_i)^2,$$

where $d(\mathbf{p}, \mathbf{q})$ is the Euclidean distance between $\mathbf{p}$ and $\mathbf{q}$ on the image plane. This problem can be more compact as the following matrix form:

$$\arg\min_{\mathbf{H}} ||\mathbf{Q} - \mathbf{H}\mathbf{P}||_2^2,$$

where

$$\mathbf{P} = \begin{bmatrix} x_1 & \cdots & x_N \\ y_1 & \cdots & y_N \\ 1 & \cdots & 1 \end{bmatrix}, \quad \mathbf{Q} = \begin{bmatrix} u_1 & \cdots & u_N \\ v_1 & \cdots & v_N \\ 1 & \cdots & 1 \end{bmatrix}.$$

Here, $||\mathbf{X}||_2$ is the Frobenius norm [14] of a matrix $\mathbf{X}$. This is a least squares problem, and the transformation matrix can be easily computed as follows.

$$\widehat{\mathbf{H}} = \mathbf{Q}\mathbf{P}\left(\mathbf{P}\mathbf{P}^T\right)^{-1} = \mathbf{Q}\mathbf{P}^{\dagger},$$

where $\mathbf{P}^{\dagger}$ is known as the pseudo inverse of $\mathbf{P}$. After obtaining $\widehat{\mathbf{H}}$, an arbitrary radar target $\mathbf{z}$ is transformed to $\widehat{\mathbf{H}}\widetilde{\mathbf{z}}$ where $\widetilde{\mathbf{z}}$ is the homogeneous coordinate of $\mathbf{z}$.

• *Direct linear transformation* (DLT) [3], [9], [10]: This method corresponds to the two dimensional homography estimation [13] between radar plane and image plane. Whereas the same coordinate systems of radar and camera data as mentioned above are used in this method, the transformation is modeled as the two dimensional projective transformation instead of the affine transformation:

$$\widetilde{\mathbf{q}} = \mathbf{H}\widetilde{\mathbf{p}},$$

where $\widetilde{\mathbf{p}}$ and $\widetilde{\mathbf{q}}$ are the homogeneous coordinates of $\mathbf{p}$ and $\mathbf{q}$, respectively, and

$$\mathbf{H} = \begin{bmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & h_{33} \end{bmatrix} \quad (8 \text{ Dof}).$$

With this transformation, an arbitrary coordinate $[x_i, y_i]^T$ is mapped to

$$\begin{bmatrix} u_i \\ v_i \end{bmatrix} = \begin{bmatrix} \frac{h_{11}x_i + h_{12}y_i + h_{13}}{h_{31}x_i + h_{32}y_i + h_{33}} \\ \frac{h_{21}x_i + h_{22}y_i + h_{23}}{h_{31}x_i + h_{32}y_i + h_{33}} \end{bmatrix}. \tag{1}$$

In terms of $\mathbf{h} = [h_{11}, h_{12}, h_{13}, h_{21}, h_{22}, h_{23}, h_{31}, h_{32}, h_{33}]^T$, (1) can be differently represented as follows.

$$\mathbf{A}_i\mathbf{h} = 0,$$

where

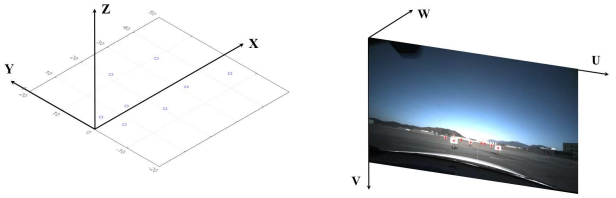
$$\mathbf{A}_i = \begin{bmatrix} -x_i & -y_i & -1 & 0 & 0 & 0 & u_i x_i & u_i y_i & u_i \\ 0 & 0 & 0 & -x_i & -y_i & -1 & v_i x_i & v_i y_i & v_i \end{bmatrix}.$$

Since we have N data pairs, the optimization problem can be formulated as follows.

$$\arg\min_{||\mathbf{h}||_2=1} ||\mathbf{A}\mathbf{h}||_2^2,$$

where

$$\mathbf{A} = \left[\mathbf{A}_1^T, \cdots, \mathbf{A}_N^T\right]^T.$$



(a) Radar (b) Camera

Fig. 2: A three dimensional coordinate systems of two sensors.

Here, $||\mathbf{x}||_2$ is the Euclidean norm of a vector $\mathbf{x}$. If $N \geq 4$, the solution of this optimization problem is the singular vector with respect to the smallest singular value of $\mathbf{A}$ [13].

According to [13], the solution computed by DLT depends on the coordinate system, and a normalization, which is called as the *pre-conditioning*, is recommended to be conducted together with DLT. It is an isotropic scaling that $\{\mathbf{p}_i\}_{i=1}^N$ have zero mean and their average distance from the mean is $\sqrt{2}$ after the normalization, which can be performed by the following similarity transformation.

$$\overline{\mathbf{p}}_i = \mathbf{T}_p \widetilde{\mathbf{p}}_i = \begin{bmatrix} c & 0 & -cm_1 \\ 0 & c & -cm_2 \\ 0 & 0 & 1 \end{bmatrix} \widetilde{\mathbf{p}}_i,$$

where m_1 and m_2 are the elements of the mean vector $\mathbf{m} = \frac{\sum_i \mathbf{p}_i}{N}$ and $c = \frac{\sqrt{2}}{\overline{d}}$ where $\overline{d}$ is the average of the norms of $\{\mathbf{p}_i - \mathbf{m}\}_{i=1}^N$. Similarly, we can also perform another normalization associated with $\mathbf{T}_q$ transforming $\{\mathbf{q}_i\}_{i=1}^N$ to $\{\overline{\mathbf{q}}_i\}_{i=1}^N$. After performing normalization using $\mathbf{T}_p$ and $\mathbf{T}_q$, DLT is applied to the normalized radar data $\{\overline{\mathbf{p}}_i\}_{i=1}^N$ and $\{\overline{\mathbf{q}}_i\}_{i=1}^N$ to obtain $\widetilde{\mathbf{H}}$. Then, an arbitrary radar target $[x,\ y]^T$ can be mapped to a point on the input image with the coordinates of

$$\widehat{u} = \frac{\widehat{h}_{11}x + \widehat{h}_{12}y + \widehat{h}_{13}}{\widehat{h}_{31}x + \widehat{h}_{32}y + \widehat{h}_{33}},$$
$$\widehat{v} = \frac{\widehat{h}_{21}x + \widehat{h}_{22}y + \widehat{h}_{23}}{\widehat{h}_{31}x + \widehat{h}_{32}y + \widehat{h}_{33}},$$

where $\widehat{h}_{ij}$ for $i = 1, \ldots, 3$ and $j = 1, \ldots, 3$ are the elements of $\widehat{\mathbf{H}} = \mathbf{T}_q^{-1}\widetilde{\mathbf{H}}\mathbf{T}_p$. Furthermore, it is known that the transformation $\widehat{\mathbf{H}}$ can be more refined by additionally performing a nonlinear optimization method such as Levenberg-Marquardt (LM) algorithm [15] using $\widehat{\mathbf{H}}$ as its initial point. One of the popular objective functions used in the nonlinear optimization is the symmetric transfer error as follows.

$$\sum_{i=1}^{N} d(\mathbf{H}\mathbf{p}_i, \mathbf{q}_i)^2 + d\big(\mathbf{p}_i, \mathbf{H}^{-1}\mathbf{q}_i\big)^2.$$

• *Extrinsic calibration* (EC): In [5] and [6], it was mentioned that they calibrated radar and camera using intrinsic and extrinsic parameters of the camera, but the procedure was not provided in detail. Different from the previous two calibration methods, the coordinate systems of radar and camera are defined as in Fig. 2, and this method models the camera-radar calibration as the following:

$$\widetilde{\mathbf{q}} = \mathbf{K}\,[\mathbf{R}\ \mathbf{t}]\,\widetilde{\mathbf{p}}, \tag{2}$$

where $\widetilde{\mathbf{p}} = [x,\ y,\ 0,\ 1]^T$ and $\widetilde{\mathbf{q}} = [u,\ v,\ 1]^T$ are the homogeneous coordinates of $\mathbf{p}$ and $\mathbf{q}$, respectively. Note that the third element of $\widetilde{\mathbf{p}}$ is set to 0 since radar provides the target location projected on xy-plane in Fig. 2a, in other words, the radar targets are regarded as three dimensional points with the same value of height ($z = 0$). In (2), $\mathbf{K}$ is the camera intrinsic matrix as follows.

$$\begin{bmatrix} -k_u f_x & s & u_0 \\ 0 & k_v f_y & v_0 \\ 0 & 0 & 1 \end{bmatrix},$$

where s is set to zero as in [12] and this matrix can be computed by [16]. Also, the matrix $[\mathbf{R}\ \mathbf{t}]$ in (2) is called the extrinsic matrix, and consists of the rotation matrix as

$$\mathbf{R} = \mathbf{R}_z\mathbf{R}_y\mathbf{R}_x,$$
$$\mathbf{R}_x = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\theta_x & \sin\theta_x \\ 0 & -\sin\theta_x & \cos\theta_x \end{bmatrix},\ \mathbf{R}_y = \begin{bmatrix} \cos\theta_y & 0 & -\sin\theta_y \\ 0 & 1 & 0 \\ \sin\theta_y & 0 & \cos\theta_y \end{bmatrix},$$
$$\mathbf{R}_z = \begin{bmatrix} \cos\theta_z & \sin\theta_z & 0 \\ -\sin\theta_z & \cos\theta_z & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

and the translation matrix as

$$\mathbf{t} = [t_x,\ t_y,\ t_z]^T.$$

The extrinsic calibration is to estimate the rigid transformation between the coordinate systems of two sensors, and the transformation in (2) is defined by the six variables as follows.

$$\mathbf{w} = [\theta_x,\ \theta_y,\ \theta_z,\ t_x,\ t_y,\ t_z] \quad (6\ \text{Dof}).$$

In the literature, the extrinsic calibration is performed by solving a nonlinear optimization problem with the following objective function:

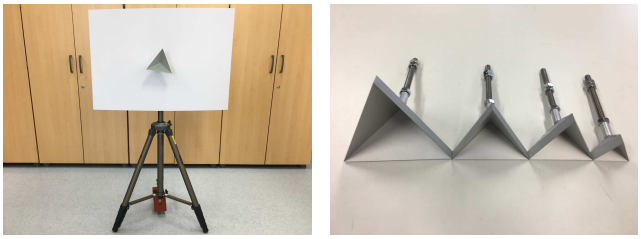
$$\arg\min_{\mathbf{w}} \sum_{i=1}^{N} ||\widetilde{\mathbf{q}}_i - \mathbf{K}\,[\mathbf{R}\ \mathbf{t}]\,\widetilde{\mathbf{p}}_i||_2^2.$$

The LM algorithm is one of the popular algorithms to solve this optimization problem when $N \geq 6$ as in [11], [12]. After obtaining $\widehat{\mathbf{w}}$, an arbitrary radar target with the coordinate of $[x, y]^T$ can be transformed onto $[\widehat{u}, \widehat{v}]^T$

$$\widehat{u} = \frac{\widehat{q}_1}{\widehat{q}_3}, \quad \widehat{v} = \frac{\widehat{q}_2}{\widehat{q}_3},$$
$$[\widehat{q}_1,\ \widehat{q}_2,\ \widehat{q}_3]^T = \mathbf{K}\left[\widehat{\mathbf{R}}\ \widehat{\mathbf{t}}\right][x,\ y,\ 0,\ 1]^T.$$

In the next section, we will present comparisons among the calibration methods described in this section.

Fig. 3: camera and radar sensors and their locations installed in a vehicle platform.



(a) A corner reflector on tripod (b) Four types of corner reflectors

Fig. 4: Corner reflectors used in our experiments.

III. Experiments

In this study, we utilized a radar sensor (Continental ARS 408-21) and a camera sensor (FLIR Chameleon3 with the lens of Theia SL183M). The radar and the camera were installed behind the frontal grill and behind the windshield of our automotive platform (Hyundai IONIQ Electric), respectively, as in Fig. 3. With the two sensors, we collected radar and image data using corner reflectors shown in Fig. 4. The resolution of the image was fixed to 1280×720 pixels. Since the absolute sizes of the corner reflectors are not large enough, they are invisible if they are placed at a distance from the camera. To address this problem, we used a piece of A_0 hard board together with each corner reflector as Fig. 4a. After putting the eight corner reflectors into the common field of view of the two sensors, their data was captured simultaneously. Fig. 1 shows an example of the data captured by our radar and camera. The radar coordinates of the corner reflectors ($\mathbf{p}$) were provided by our radar, and their image coordinates ($\mathbf{q}$) were manually determined. Nonetheless, it was not easy to manually decide the image coordinates of the corner reflectors with the distances above about 40 meters. In addition to the image coordinates, we also manually associated the radar and image coordinates.

By these ways, we constructed twelve sets to be used in the camera-radar calibration as Fig. 1 for our experiments. The twelve sets were divided into *training data* and *test data*: the data pairs in the training data were used to estimate transformations by the calibration methods, the data pairs in the test data were used to evaluate them. In order to establish a quantitative comparison of the calibration methods, we used the *calibration error* for given N pairs of $\{\mathbf{p}_i\}$ and $\{\mathbf{q}_i\}$ as the following:

$$\frac{1}{N}\sum_{i=1}^{N} d(\widehat{\mathbf{q}}_i, \mathbf{q}_i)\,,$$

where $\widehat{\mathbf{q}}_i$ is an image coordinate mapped from $\mathbf{p}_i$ by an estimated transformation. We implemented all of the calibration methods mentioned in our experiments using Python. In the case of EC, it needs the camera intrinsic matrix, and the camera calibration was conducted using OpenCV to obtain the following camera intrinsic matrix:

$$\begin{bmatrix} 375.40 & 0 & 630.97 \\ 0 & 374.23 & 491.74 \\ 0 & 0 & 1 \end{bmatrix},$$

At first, in order to investigate the relation between the number pairs of $\{\mathbf{p}, \mathbf{q}\}$ and the calibration accuracy, we compared the calibration methods in the situation of varying the number of the radar image data pairs N_r. For $4 \leq N_r \leq 8$, we randomly selected N_r pairs from one set to perform the calibrations (training data), and all of the pairs in the other eleven sets were used to measure the calibration error (test data). In this case, N_t was set to $88\,(= 11 \times 8)$. To avoid the bias by randomness, we repeated the selection from one set 10 times so that the training and testing for a calibration method was performed 120 times, 96 times and 12 times for $4 \leq N_r \leq 6$, $N_r = 7$, and $N_r = 8$, respectively. For $N_r = 16, 24, \ldots, 54$, we randomly selected $N_r/8$ sets from the twelve sets, and all of the N_r pairs in the selected sets were used as training data and all of the pairs in the remaining sets were used as test data. In this case, N_t was set to $12 \times 8 - N_r$. For the same purpose as above, we repeated the random selection of $N_r/8$ sets 5 times so that the training and testing of each calibration method was conducted 60 times for $N_r = 16, 24, \ldots, 54$. We averaged all of the calibration errors for a specific value of N_r, and Table I shows the average calibration errors for various N_r. In the table, NDLT is DLT along with the normalization described in the previous section, and LM is the nonlinear refinement using LM algorithm. From the table, we can see that, for all the methods, their calibration errors decrease as N_r increase, and the decrease was seemed to be stopped around $N = 50$ except PI. Interestingly, PI showed the best performance when $N_r = 4$, but it provided very large calibration errors for all the cases of $N_r \geq 5$ compared to the other methods. This result agrees with the fact that the parallelism of two lines is not preserved by the transformation from radar plane to image plane, and we can figure out that PI based on the affine transformation preserving the parallelism is not appropriate to the camera-radar calibration. On the other hand, in total, DLT based methods yielded almost same level in the calibration error

TABLE I: Average calibration errors for various N_r

Method	$N_r = 4$ (120)	$N_r = 5$ (120)	$N_r = 6$ (120)	$N_r = 7$ (96)	$N_r = 8$ (12)	$N_r = 16$ (60)	$N_r = 24$ (60)	$N_r = 32$ (60)	$N_r = 40$ (60)	$N_r = 48$ (60)	$N_r = 54$ (60)
PI	84.439	73.092	61.203	56.499	53.302	50.339	49.577	49.170	48.711	48.339	48.294
DLT	142.899	16.724	4.605	3.260	2.953	2.688	2.593	2.530	2.478	2.404	2.487
NDLT	1172.922	11.468	4.074	3.174	2.896	2.573	2.467	2.430	2.375	2.318	2.340
DLT+LM	202.319	17.105	4.694	3.348	3.060	2.850	2.703	2.743	2.618	2.563	2.592
NDLT+LM	578.322	7.819	4.149	3.269	2.971	2.849	2.702	2.740	2.617	2.563	2.591
EC	-	-	3.370	3.263	3.194	2.942	2.824	2.880	2.785	2.739	2.7415

for the same value of $N_r \geq 6$, and they were the most competitive when $N_r \geq 7$. It is worthy to note that the errors of NDLT and NDLT+LM were slightly lower than DLT and DLT+LM, respectively, for $N_r \geq 5$, and the improvement by the normalization was obviously shown for $N_r = 5$. This means that the pre-conditioning normalization is effective in being applied with DLT. On the contrary, the errors of DLT+LM and NLDT+LM were slightly larger than DLT and NDLT, respectively, except $N_r = 4$, i.e., the nonlinear refinement by LM was seemed to be not effective different from the pre-conditioning normalization. In the case of EC, it was the best when $N_r = 6$, but it gave slightly larger errors than the DLT based methods for $N_r \geq 8$. We also found in our experiments that the calibration error of EC is highly dependent on the camera intrinsic matrix. Thus, the precise camera calibration is one of the important preliminaries of the camera-radar calibration using the method of EC.

In addition to comparing the calibration errors, we compared the locations on the image where the radar data are projected by the transformations obtained from the different calibration methods. To do this, we performed estimating the transformations using all the calibration methods and measuring their calibration errors for each combination to select two sets from the twelve sets. In this case, both N_r and N_t were set to 8. The average errors for all the combinations are shown as Fig. 5. Among the twelve sets, the eleventh and the ninth sets gave the minimum and maximum calibration errors on average when they used for testing. We plotted their radar targets projected onto the images by the transformations computed using the training sets which provided the minimum and maximum errors. Fig. 6 shows the radar targets projected on the images for four combinations of the training and test data sets. Note that each radar target is projected onto almost same locations by the transformations obtained from all the methods except PI regardless of the combination of the training and test data sets.

In summary, PI was not suitable for the camera-radar calibration, and the differences of the other calibration methods were insignificant quantitatively and qualitatively. In a situation that the calibration speed is critical together with its accuracy, NDLT may be a better choice than the other methods because DLT is based on the optimization problem having a closed form solution whereas LM is an iterative

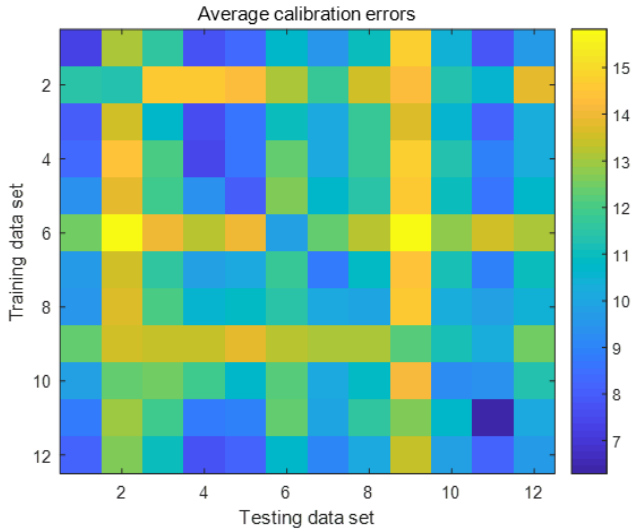


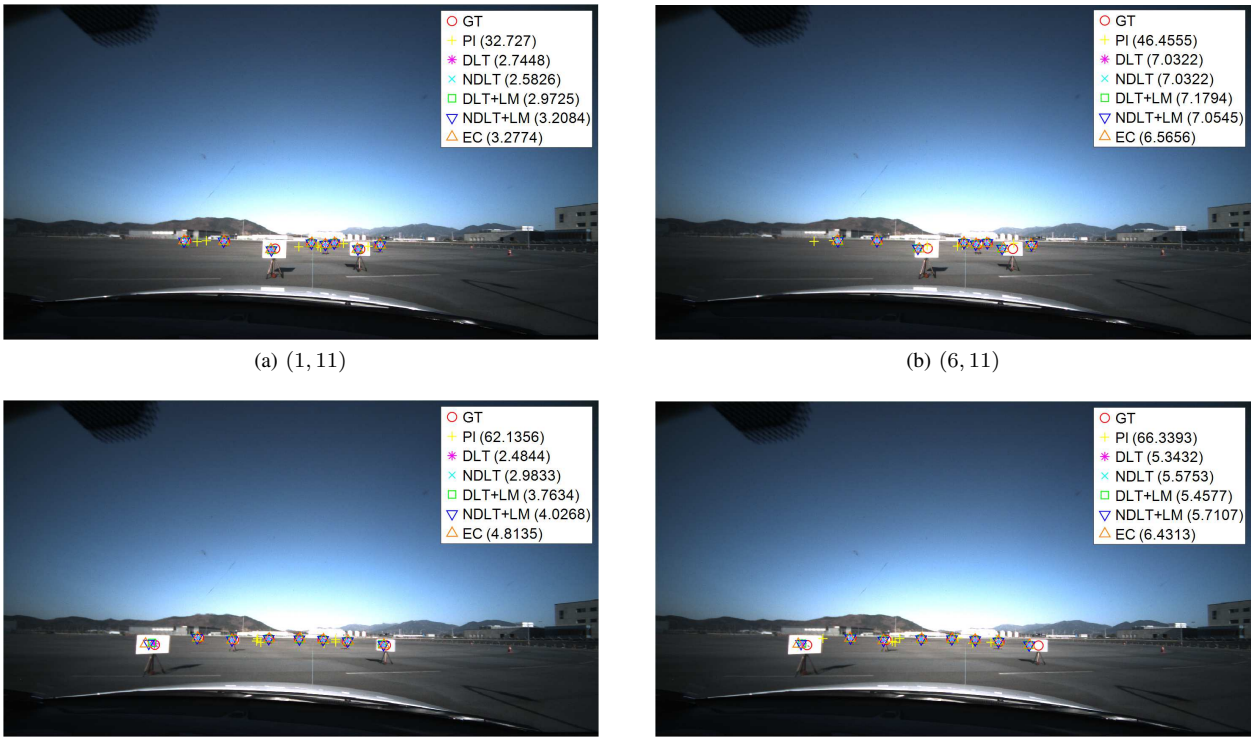
Fig. 5: Average calibration errors for each combination to select two sets for training and testing.

algorithm. Lastly, the accurate calibration can be obtained from about 20 radar-image data pairs.

IV. Conclusion

In this paper, we described the camera-radar calibration methods presented in the literature, and compared them in terms of the calibration error using the data collected by the camera and radar installed in a vehicle platform. Three types of the calibration methods (PI, DLT, and EC) were considered in this study, and the normalization based on the pre-conditioning and the nonlinear refinement were applied along with DLT. Our experiments showed that PI was not appropriate to the camera-radar calibration and DLT based methods provided slightly better than EC in most cases we considered. Also, the pre-conditioning normalization is slightly more effective than the nonlinear refinement under the assumption that target objects have almost same height.

In the future work, it will be included to efficiently increase the number of the radar-image data pairs because it was demonstrated that the calibration accuracy can be improved together with the number of the pairs, but the number of simulated targets such as corner reflectors is generally limited. It can also be one of the important progresses in the camera-radar calibration to automatically determine the



(a) $(1, 11)$ (b) $(6, 11)$

(c) $(12, 9)$ (d) $(6, 9)$

Fig. 6: The radar targets projected on the corresponding image for four combinations of $(\text{training set}, \text{test set})$. The numbers in the parentheses of legends are the calibration errors

image coordinates of those targets and to associate the radar coordinates and the corresponding image coordinates without any manual process. Moreover, it is also meaningful to apply the comparisons performed in this study to real data such as cars, cyclists, and pedestrians used in obstacle detection or moving object tracking.

REFERENCES

[1] H. Cho, Y. W. Seo, B. V. K. V. Kumar, and R. R. Rajkumar, "A Multi-Sensor Fusion System for Moving Object Detection and Tracking in Urban Driving Environments," in *Proc. 2014 IEEE International Conference on Robotics and Automation (ICRA)*, May 2014, pp. 1836–1843.

[2] R. O. Chavez-Garcia and O. Aycard, "Multiple Sensor Fusion and Classification for Moving Object Detection and Tracking," *IEEE Transactions on Intelligent Transportation Systems*, vol. 17, no. 2, pp. 525–534, Feb 2016.

[3] S. Sugimoto, H. Tateda, H. Takahashi, and M. Okutomi, "Obstacle Detection Using Millimeter-wave Radar and Its Visualization on Image Sequence," in *Proc. 17th International Conference on Pattern Recognition, 2004. ICPR 2004.*, Aug. 2004.

[4] M. Bertozzi, L. Bombini, P. Cerri, P. Medici, P. C. Antonello, and M. Miglietta, "Obstacle Detection and Classification fusing Radar and Vision," in *Proc. 2008 IEEE Intelligent Vehicles Symposium*, June 2008, pp. 608–613.

[5] Z. Ji and D. Prokhorov, "Radar-Vision Fusion for Object Classification," in *Proc. 2008 11th International Conference on Information Fusion*, June 2008, pp. 1–7.

[6] Z. Ji, M. Luciw, J. Weng, and S. Zeng, "Incremental Online Object Learning in a Vehicular Radar-Vision Fusion Framework," *IEEE Transactions on Intelligent Transportation Systems*, vol. 12, no. 2, pp. 402–411, June 2011.

[7] A. Roy, N. Gale, and L. Hong, "Fusion of Doppler Radar and Video Information for Automated Traffic Surveillance," in *Proc. 2009 12th International Conference on Information Fusion*, July 2009, pp. 1989–1996.

[8] T. Wang, N. Zheng, J. Xin, and Z. Ma, "Integrating Millimeter Wave Radar with a Monocular Vision Sensor for On-Road Obstacle Detection Applications," *Sensors*, vol. 11, pp. 8992–9008, 2011. [Online]. Available: http://www.mdpi.com/1424-8220/11/9/8992

[9] D. Y. Kim and M. Jeon, "Data fusion of radar and image measurements for multi-object tracking via Kalman filtering," *Information Science*, vol. 278, pp. 641–652, Sept. 2014.

[10] X. Wang, L. Xu, H. Sun, J. Xin, and N. Zheng, "Bionic Vision Inspired On-Road Obstacle Detection and Tracking using Radar and Visual Information," in *Proc. 17th International IEEE Conference on Intelligent Transportation Systems (ITSC)*, Oct. 2014, pp. 39–44.

[11] G. E. Natour, O. A. Aider, R. Rouveure, F. Berry, and P. Faure, "Radar and vision sensors calibration for outdoor 3D reconstruction," in *Proc. 2015 IEEE International Conference on Robotics and Automation (ICRA)*, May 2015, pp. 2084–2089.

[12] G. E. Natour, O. Ait-Aider, R. Rouveure, F. Berry, and P. Faure, "Toward 3D Reconstruction of Outdoor Scenes Using an MMW Radar and a Monocular Vision Sensor," *Sensors*, vol. 15, pp. 25 937–25 967, 2015. [Online]. Available: http://www.mdpi.com/1424-8220/15/10/25937

[13] R. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*, 2nd ed. New York, NY, USA: Cambridge University Press, 2003.

[14] S. Boyd and L. Vandenberghe, *Convex Optimization*. New York, NY, USA: Cambridge University Press, 2004.

[15] J. Nocedal and S. J. Wright, *Numerical Optimization*, 2nd ed. New York, NY, USA: Springer, 2006.

[16] Z. Zhang, "A Flexible New Technique for Camera Calibration," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, pp. 1330–1334, Dec. 2000.