2019 IEEE International Conference on Multimedia and Expo (ICME)

REVISIT SURROUND-VIEW CAMERA SYSTEM CALIBRATION

Xuan Shao[1], *Xiao Liu*[1,*], *Lin Zhang*[1,*], *Shengjie Zhao*[1], *Ying Shen*[1], *Yukai Yang*[2]

[1]School of Software Engineering, Tongji University, Shanghai, China
[2]Department of Statistics, Uppsala University, Uppsala, Sweden

ABSTRACT

The surround-view system is an essential component of an advanced driver assistance system especially when the vehicle runs in tight parking space or on a narrow road. To ensure successful maneuvering, a panoramic bird's-eye image with no blind spots is necessarily called for. Hence, a typical surround-view system consists of several cameras mounted around the vehicle capturing images from a top-down viewpoint, and an accurate extrinsic calibration for such system is prerequisite for providing a seamless surround-view image. To achieve this goal, this paper presents a novel extrinsic calibration pipeline which is both easy-to-use and reliable to operate on multiple cameras. Instead of taking the vehicle to a fixed position in a specific calibration site, a single chessboard is the only demand. We adopt a novel refinement procedure that jointly optimizes camera poses in a closed-loop manner. The effectiveness and efficiency of the proposed pipeline to calibrate a surround-view camera system has been corroborated by experiments.

Index Terms— Extrinsic calibration, overlapping field of view, surround-view camera system, bird's-eye image

1. INTRODUCTION

The surround-view system (SVS), as part of most advanced driver assistance systems (ADAS), is now an emerging technology that allows the driver to see a surround-view bird's-eye image around the vehicle [1]. Such a system is particularly essential when the driver maneuvers into a parking slot or on a narrow road. It typically mounts four fisheye cameras around a vehicle capturing images from a top-down viewpoint. However, to seamlessly stitch all captured images, it entails an accurate calibration of the multi-camera system. If calibration parameters are not estimated precisely, a false perception of surroundings will be acquired, which is fatally dangerous in vehicle control. Note that since intrinsic calibration techniques are now quite mature, each camera is delivered to its end-user with precisely accurate intrinsic parameters by quality control. So, instead of intrinsic calibration, the paper draws focus on extrinsic calibration procedure. Besides, normally cameras are extrinsically calibrated by a professional taking vehicle to a specific calibration site in a automobile $4S$ store, which is a challenging and cumbersome task. Hence a reliable and easy-to-use surround-view camera extrinsic calibration pipeline is our main concern here.

*Corresponding author. Email: {1532787, cslinzhang}@tongji.edu.cn

2. RELATED WORK AND OUR CONTRIBUTIONS

2.1. Interest point-based approach

Since extrinsic calibration techniques are inherently based on feature extraction, according to which features are used, the existing methods can be categorized into two approaches: interest point-based approach and pattern-based approach. The interest point-based approach estimates the camera parameters using the interest points extracted from real scene images [2, 3, 4, 5]. Heng *et al.* calibrate the surround-view camera system using tracked interest points and optimize the initial parameters by bundle adjustment [2]. Given intrinsics known in advance, Carerra *et al.* propose a feature-based extrinsic calibration method based on visual SLAM algorithm [6]. However, interest points are hardly tracked to obtain point correspondences between images from adjacent cameras because overlapping areas are severely distorted due to the use of fisheye lenses. So interest point-based approach is not suited for obtaining an accurate calibration parameters.

2.2. Pattern-based approach

A pattern-based approach estimates camera parameters using special patterns including corners, circles, or lines. Since a pattern-based approach uses precisely drawn patterns whose configurations are known, it is possible to accurately estimate the camera parameters, which makes it suitable for accurate calibration for a surround-view camera system. A branch of pattern-based calibration methods is placing calibration pattern in overlapped field of the cameras [7, 8, 9, 10, 11]. Methods in [10, 11] use factorization-based method by placing calibration pattern between adjacent groups of cameras to calibrate the surround-view camera system. However, no further refinement is performed to guarantee high-precision of calibration parameters, which consequently leads to the accumulation errors in calibration results. To incorporate optimization of calibration parameters, Zhang *et al.* propose a

978-1-5386-9552-4/19/$31.00 ©2019 IEEE
DOI 10.1109/ICME.2019.00257

surround-view camera solution designed for embedded systems, based on a photometric alignment to correct brightness and color mismatch synthesis [12]. But a large calibration site is required, which is inconvenient for end-user operation at any time. Apart from above methods, methods in [13, 14, 15] represent non-overlapping pattern-based approaches: they use range photogrammetry and a simple reference plane with a repeated pattern to calibrate such a surround-view system. However, the non-overlapping based method is not stable because noises will influence the result of geometric estimation.

2.3. Our motivations and contributions

Through the literature survey, we find that in the field of surround-view camera calibration we need to continue to devote efforts in the following two aspects. First, accurate extrinsic calibration is prerequisite to the final seamless bird's-eye image. Unfortunately, the optimization of camera extrinsic calibration parameters is rarely considered in previous methods. Besides, the pipeline for extrinsic calibration of a surround-view camera system needs to be further improved. In fact, previous extrinsic calibration procedures are a mission full of inconvenience. Cumbersome steps make them more formidable than it looks for the driver to operate. Drivers must take the vehicle to a specific calibration site in an authorized dealer and stop the vehicle at a fixed point to calibrate. When the car needs to be re-calibrated due to vibrations of mounted cameras, such steps are really inconvenient to the end-user. By far, the latest work that is closely related to our work is one established in [16], but no further refinement is performed and the operation pipeline is cumbersome for end-user operation. In this work, we attempt to fill the aforementioned research gaps to some extent. Our pipeline is easy-to-use and relies simply on chessboard placed in the overlapped field between adjacent cameras. Our major contributions are three-fold:

(1) Our pipeline is easy-to-use and spares lots of trouble in finding a specific broad calibration site. The set-up configuration is just a chessboard; and no prior information of the relative position between the vehicle and the chessboard is necessary.

(2) We adopt a novel refinement approach that jointly optimizes camera poses in a closed-loop manner.

(3) In order to get bird's-eye image, rather than choosing pixels in undistorted image, no information is lost by direct mapping from fisheye image to bird's-eye image.

3. OUR PROPOSED PIPELINE

As illustrated in Fig. 1, four fisheye cameras are mounted on a vehicle in four directions: in front, under two side mirrors and on the back side. Each camera has over $180°$ field of view. Adjacent cameras share some overlapped areas and four fisheye cameras together cover the whole surrounding area of the vehicle. The flowchart of the proposed pipeline is depicted in Fig. 2. It mainly consists of three parts: (1) pose initialization and joint optimization of surround-view camera system (2) calculation of camera-ground relation and (3) mapping table formulation from the bird's-eye image to the original fisheye image. The details are given as follows.

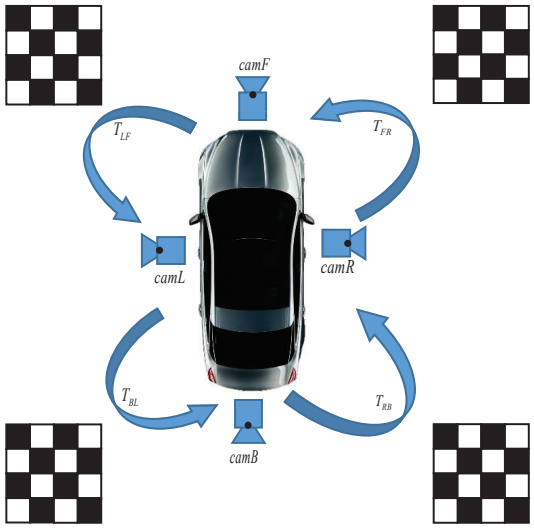

Fig. 1. Configuration set-up for surround-view camera system.

3.1. Pose initialization and joint optimization

3.1.1. Pose initialization for each camera

Relative motion between front and left cameras T_{LF} can be estimated by epipolar constraint using a chessboard placed in the common field of view [17]. Similarly, we can get the relative motion T_{BL}, T_{RB} and T_{FR}. By setting front camera coordinates as reference coordinates, we can get four camera poses in front camera coordinates (Table 1).

Table 1. Poses in front camera coordinates

camera	pose
$camF$	$T_{FF} = I$
$camL$	$T_{LF} = T_{LF}T_{FF}$
$camB$	$T_{BF} = T_{BL}T_{LF}T_{FF}$
$camR$	$T_{RF} = T_{RB}T_{BL}T_{LF}T_{FF}$
$camF^{'}$	$T^{'}_{FF} = T_{FR}T_{RB}T_{BL}T_{LF}T_{FF}$

Note that by multiplying T_{FR} with T_{RF}, we define a new front camera pose after a closed-loop transformation $T^{'}_{FF} = T_{FR}T_{RB}T_{BL}T_{LF}T_{FF}$, which, under ideal conditions, should be equal to T_{FF}. However, due to accumulation error, $T^{'}_{FF} \neq T_{FF}$ which can be used to improve the initial estimation of camera poses.

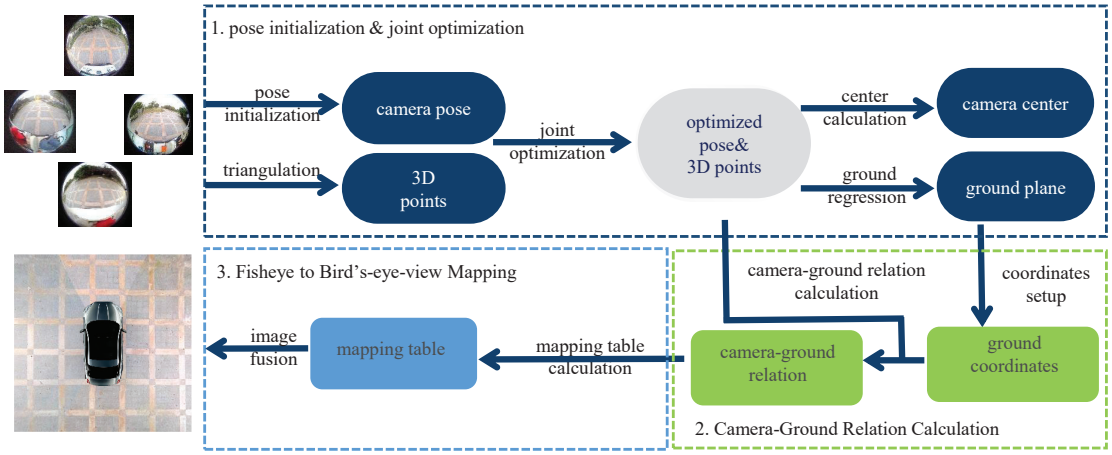

Fig. 2. Pipeline of calibration for surround-view camera system.

3.1.2. *Joint optimization of surround-view cameras*

A graph based [18] joint optimization method is used to optimize the pose of each camera. As denoted in Fig. 3, we triangulate all chessboard corners to get $3D$ points from four directions: P_{FF}, P_{LF}, P_{BF} and P_{RF} respectively [19, 20]. Each edge denotes the reprojection error between the original $3D$ point and the reprojected point in the camera. Apart from T_{FF}, T_{LF}, T_{BF} and T_{RF}, T'_{FF} needs to be simultaneously optimized. For each point $P_j(X_j, Y_j, Z_j)$, its image $\boldsymbol{u_{ij}}(u_{ij}, v_{ij})$ and the ith camera's pose ξ_i, we sum all reprojection errors up and build up a least squares minimization problem:

$$P^*, \xi^* = \arg\min_{P,\xi} \frac{1}{2} \sum_{i=1}^{5} \sum_{j=1}^{N_i} (\boldsymbol{u_{ij}} - \frac{1}{s_{ij}} K_i \exp(\xi_i) P_j) \quad (1)$$

K denotes camera intrinsics. During optimization, it will adjust poses of the four cameras in order to make $T'_{FF} = T_{FF}$, and solve for the optimal camera pose: T^*_{FF}, T^*_{LF}, T^*_{BF} and T^*_{RF}. $3D$ points are also optimized simultaneously.

3.2. Camera-ground relation calculation

After getting the pose of each camera in reference coordinates (front camera coordinates), as long as we know the relation between the front camera and the ground, we can obtain the relation between the whole surround-view camera system and the ground. By setting up a ground coordinates, the relation between the ground and front camera can be calculated by the coordinates of four points in both coordinate systems, respectively (seen in Fig. 4).

3.2.1. *Ground coordinates set-up*

In order to set up ground coordinates, three concurrent lines orthogonal with one another are required. Thus four points

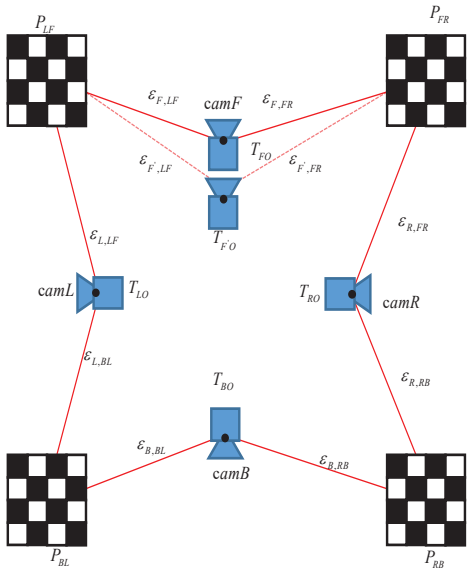

Fig. 3. Graph based joint optimization of surround-view camera system.

are needed:

(1) P_O and P_Y: We set each camera center in its own coordinates as: $P = (0, 0, 0, 1)$. Given camera poses, we get the camera center for each camera and the whole system as shown in Table 2. Using all points calculated in Sec. 3.1.2, a ground plane G_F in front camera coordinates can be regressed:

$$\boldsymbol{G_F} : ax + by + cz + d = 0 \quad (2)$$

By calculating projection points of P_{center} and P_{camF} to the ground plane : P_O and P_Y. We then define Y-axis pointing from the origin P_O to the point P_Y in ground coordinates (red color in Fig. 4) .

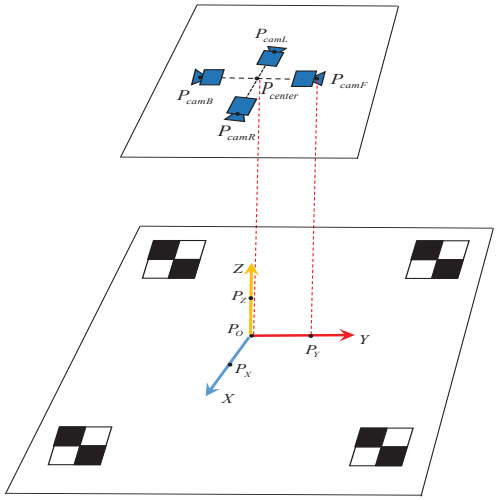

Fig. 4. Ground coordinates set-up.

Table 2. Camera center and surround-view system center

camera	center
$camF$	$T_{FF}^{-1}(0,0,0,1)^T$
$camL$	$T_{LF}^{-1}(0,0,0,1)^T$
$camB$	$T_{BF}^{-1}(0,0,0,1)^T$
$camR$	$T_{RF}^{-1}(0,0,0,1)^T$
$center$	$\frac{1}{4}(camF + camL + camB + camR)$

(2) P_X: Similarly, X-axis is defined by vector pointing from P_O to P_X that is orthogonal to Y-axis (blue color in Fig. 4). P_X must be in the ground plane $\boldsymbol{G}$. Hence, we can get P_X by solving the following equations:

$$\begin{cases} (P_X - P_O)^T (P_Y - P_O) = 0 \\ \begin{bmatrix} a & b & c & d \end{bmatrix} P_X = 0 \end{cases} \tag{3}$$

(3) P_Z: Since X, Y and Z axes are orthogonal with one another, a point P_Z in Z-axis can be calculated by the following equation:

$$\begin{aligned} &(P_Y - P_O) \times (P_X - P_O) = (P_Z - P_O) \\ &\Rightarrow P_Z = (P_Y - P_O)(P_X - P_O) + P_O \end{aligned} \tag{4}$$

3.2.2. Camera-ground relation

For points in front camera coordinates P_O, P_X, P_Y and P_Z, we have their coordinates in ground coordinates: P_{GO}, P_{GX}, P_{GY}, P_{GZ} by setting the length of corresponding line segment. Four corresponding points in ground and front camera coordinates are shown in Table 3.

Then we can calculate transformation matrix from the ground coordinates to the front camera coordinates T_{FG} by stacking four points into a matrix:

Table 3. Correspondent points between the ground coordinates and front camera coordinates

axis	front camera	ground coordinates
O	P_O	P_{GO}: (0,0,0,1)
X	P_X	P_{GX} : ($\|P_X - P_O\|_2$,0,0,1)
Y	P_Y	P_{GY}: (0,$\|P_Y - P_O\|_2$,0,1)
Z	P_Z	P_{GZ}: (0,0,$\|P_Z - P_O\|_2$,1)

$$P_c = T_{FG} P_g \Rightarrow T_{FG} = P_c P_g^{-1} \tag{5}$$

Here, $P_c = [P_O \quad P_X \quad P_Y \quad P_Z]$ and $P_g = [P_{GO} \quad P_{GX} \quad P_{GY} \quad P_{GZ}]$. According to each camera pose in front camera coordinates, we can obtain camera-ground relation for the whole surround-view camera system: T_{FG}, T_{LG}, T_{BG} and T_{RG}.

3.3. Fisheye image to bird's-eye image

3.3.1. Bird's-eye image to ground coordinates

We define a bird's-eye image of size $W \times H$ and each pixel has the size of $dx \times dy$. The origin of the ground coordinates is set in the image center and the origin of the bird's-eye image is set in the upper-left of the image (seen in Fig. 5). Thus the relation between each pixel in bird's-eye image $\boldsymbol{u_G}(u_G, v_G)$ and its ground coordinate $P_G\ (x, y)$ can be expressed as:

$$\begin{cases} u_G = \dfrac{x}{dx} + \dfrac{W}{2dx} \\ v_G = \dfrac{-y}{dy} + \dfrac{H}{2dy}, \end{cases} \tag{6}$$

More concisely, we have the following equation:

$$\begin{bmatrix} u_G \\ v_G \\ 1 \end{bmatrix} = \begin{bmatrix} 1/d_x & 0 & W/2d_x \\ 0 & -1/d_y & H/2d_y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \tag{7}$$

In the above equation, we omit Z-axis, since all points in the ground share the same value $z = 0.2$, which is the depth of our chessboard. More concisely, denoting the relation as K_G, we have:

$$u_G = K_G P_G \tag{8}$$

3.3.2. Fisheye image to bird's-eye image

Since K_G and camera-ground relation T_{camG}, $camG \in \{FG, LG, BG, RG\}$ are known, we transform each pixel in bird's-eye image u_G into a point in each camera coordinate system P_{cam}:

$$P_{cam} = T_{camG} P_G = T_{camG} K_G^{-1} u_G \tag{9}$$

Thus a mapping table from bird's-eye image u_G to fisheye image u_{cam} can be straightly acquired by pinhole camera

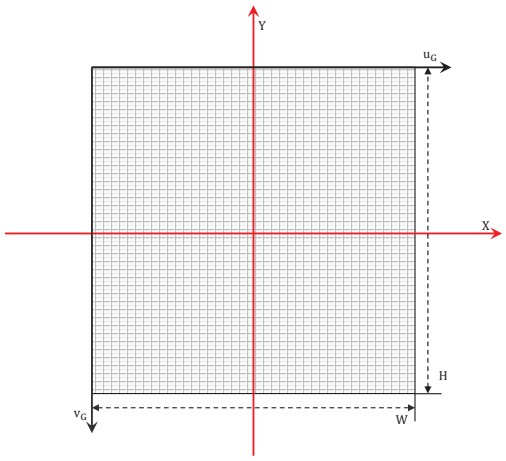

Fig. 5. Bird's-eye image and ground coordinates.

model for each camera:

$$u_{cam} = K_{cam} f_D^{cam}(T_{camG} K_G^{-1} u_G) \tag{10}$$

Where f_D^{cam} is distortion function. By the above equation, we can transform from a pixel u_G in bird's-eye image straightly to a pixel u_{cam} in original fisheye-image. No information is lost here compared with methods that select pixels from undistorted image. By different blending techniques, we can stitch the images from four directions in order to get a surround-view bird's-eye image.

4. RESULTS

4.1. Experimental set-up

In our approach, a single chessboard is used. A larger pattern improves precision of corner detection further the precision of camera extrinsic calibration. However, a larger one is cumbersome for the user to carry around. To balance off, we use the chessboard with 9×6 squares and each square is 10 centimeters in length. Our system consists of four fisheye cameras mounted at the center of front and rear car windows, and under two side view mirrors. The resolution, field-of-view, and acquisition frequency of the fisheye camera are 1280×1080, 190 degrees and 30 frames per second, respectively.

4.2. Quantitative experiment

4.2.1. Reprojection errors before and after refinement of the initial estimation

Table 4 shows the average reprojection errors of each camera using initial estimated poses and poses after joint optimization. From this table, we can see that the average reprojection error of all cameras decreases from 0.3392 to 0.3009. Errors are distributed more evenly among cameras after joint optimization, which account for the small increase in left and back cameras' reprojection errors. Reprojection error of right camera is the largest among all cameras. After refinement, it declines by a large margin, which demonstrates the effectiveness of the proposed joint optimization approach.

Table 4. Reprojection errors before and after Refinement

errors	front	left	back	right	average
before	0.5804	0.0952	0.1135	0.5667	0.3392
after	0.3757	0.3040	0.3667	0.1570	0.3009

4.2.2. Comparison with previous methods

As aforementioned in Sect. 2, previous methods for calibrating surround-view camera system can be categorized into pattern-based approach and interest point-based approach. Table 5 shows comparison of eight previous methods and our proposed one from the viewpoint of three aspects: prior information of pattern position, accuracy and information loss during bird's-eye image generation. From the following table, we could see that our pipeline is not only easy-to-use, but also more accurate. In addition, no information is lost during the whole process (seen in supplementary material).

Table 5. Comparison with other methods

method	category	accuracy	prior info	loss info
Heng [3]	pattern	good	×	-
Choi [4]	interest point	good	√	-
Liu [5]	interest point	fair	-	-
Hedi [7]	pattern	fair	√	√
Natro [8]	pattern	fair	×	×
Liu [9]	pattern	good	×	√
Zhang [12]	pattern	good	-	√
Zhang [16]	pattern	fair	√	√
ours	pattern	good	×	×

4.3. Qualitative experiment

Fig. 6 shows poses of the surround-view cameras before and after joint optimization. In (a), the new front camera (colored in red) defined after a closed-loop transformation deviates from the original front camera (colored in blue) by a large margin, whereas both cameras in (b) almost coincide. The output of the surround-view bird's-eye image using our proposed approach is shown in Fig. 7. Four images collected from outdoor environment are stitched correctly, which further proves that the estimated calibration parameters are precisely accurate. More results are shown in our supplementary material.

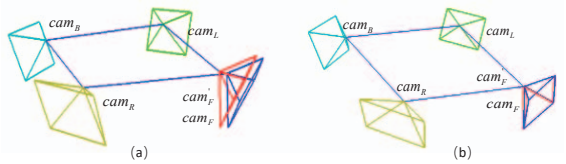

Fig. 6. Relationship among cameras before/after refinement.

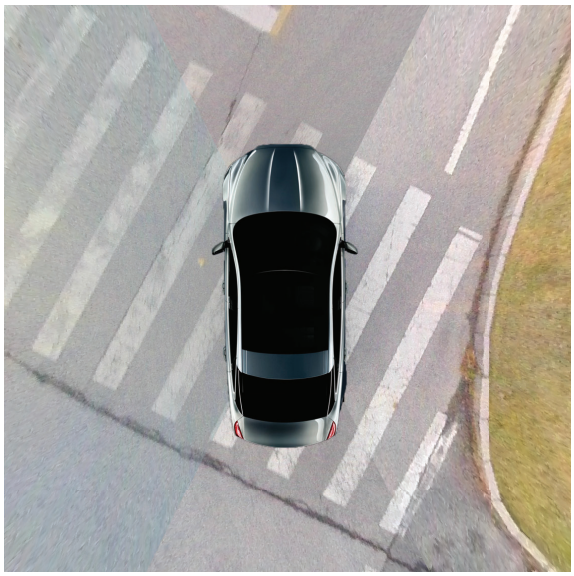

Fig. 7. Result of surround-view bird's-eye image synthesis.

5. CONCLUSIONS

This paper proposes a novel and practical pipeline that calibrates a multi-camera system in a fully easy-to-use manner. This pipeline is expected to improve the drivers' experience by simply using a single chessboard and does not rely on any prior information of the relative position between the vehicle and the calibration board. During the whole process, no information is lost by direct mapping from the fisheye image to the bird's-eye image.

6. ACKNOWLEDGMENT

This research was funded in part by the Natural Science Foundation of China under Grant No. 61672380, in part by the National Key Research and Development Project under Grant No. 2017YFE0119300, and in part by the Fundamental Research Funds for the Central Universities under Grant No. 2100219068.

7. REFERENCES

[1] S. Hecker, D. Dai, and L. Gool, "Learning driving models with a surround-view camera system and a route planner.," *CVPR*, 2018.

[2] L. Heng, B. Li, and M. Pollefeys, "Camodocal: Automatic intrinsic and extrinsic calibration of a rig with multiple generic cameras and odometry," in *IROS*, 2013, pp. 1793–1800.

[3] L. Heng, M. Brki, G. H. Lee, P. Furgale, R. Siegwart, and M. Pollefeys, "Infrastructure-based calibration of a multi-camera rig," in *ICRA*, 2014, pp. 4912–4919.

[4] K. Choi, H. G. Jung, and J. Suhr, "Automatic calibration of an around view monitor system exploiting lane markings," *Sensors*, vol. 18, no. 9, pp. 2956, 2018.

[5] Y. C. Liu, K. Lin, and Y. Chen, "Birds-eye view vision system for vehicle surrounding monitoring," in *ICRV*, 2008, pp. 207–218.

[6] G. Carrera, A. Angeli, and A. Davison, "Slam-based automatic extrinsic calibration of a multi-camera rig," in *ICRA*, 2011, pp. 2652–2659.

[7] A. Hedi and S. Lonari, "A system for vehicle surround view," *IFAC Proceedings Volumes*, vol. 45, no. 22, pp. 120–125, 2012.

[8] K. Natroshvili and K. Scholl, "Automatic extrinsic calibration methods for surround view systems," in *IV*, 2017, pp. 82–88.

[9] Y. Liu and B. Zhang, "Photometric alignment for surround view camera system," in *ICIP*, 2014, pp. 1827–1831.

[10] T. Ueshiba and F. Tomita, "Calibration of multi-camera systems using planar patterns," *Sensors*, vol. 8, pp. 4, 2002.

[11] L. Zhang, B. Li, and Y. Jia, "A practical calibration method for multiple cameras," in *ICIG*, 2007, pp. 45–50.

[12] B. Zhang, V. Appia, I. Pekkucuksen, Y. Liu, A. Batur, P. Shastry, S. Liu, S. Sivasankaran, and K. Chitnis, "A surround view camera solution for embedded systems," in *CVPRW*, 2014, pp. 676–681.

[13] S. Dong, X. Shao, X. Kang, F. Yang, and X. He, "Extrinsic calibration of a non-overlapping camera network based on close-range photogrammetry," *Applied Optics*, vol. 55, no. 23, pp. 6363–6370, 2016.

[14] S. Kawabata and Y. Kawai, "Plane based multi camera calibration under unknown correspondence using icp-like approach," in *ICPR*, 2012, pp. 3700–3703.

[15] P. Lebraly, E. Royer, O. Ait-Aider, C. Deymier, and M. Dhome, "Fast calibration of embedded non-overlapping cameras," in *ICRA*, 2011, pp. 221–227.

[16] L. Zhang, J. Huang, X. Li, and L. Xiong, "Vision-based parking-slot detection: A dcnn-based approach and a large-scale benchmark dataset," *IEEE Trans. IP*, vol. 27, no. 11, pp. 5350–5364, 2018.

[17] H. Li and R. Hartley, "Five-point motion estimation made easy," *ICPR*, pp. 630–633, 2006.

[18] R. Kummerle, G. Grisetti, H. Strasdat, K. Konolige, and W. Burgard, "G2o: A general framework for graph optimization," *ICRA*, pp. 3607–3613, 2011.

[19] G. Vogiatzis and C. Hernandez, "Editor's choice article: Video-based, real-time multi-view stereo," *Image and Vision Computing*, vol. 29, no. 7, pp. 434–441, 2011.

[20] C. Forster, M. Pizzoli, and D. Scaramuzza, "SVO: Fast semi-direct monocular visual odometry," *ICRA*, pp. 15–22, 2014.