

Improving 2D Camera Calibration by LO-RANSAC

Qin Zhang, Shiqian Wu, Wei Wang, and Zhijun Fang

Abstract—The performance of the traditional 2D pattern based camera calibration is affected by sample viewpoints, positions and feature point localization. In this paper, the Locally Optimized RANSAC (Random Sample Consensus) (LO-RANSAC) is employed to remove the unreliable information automatically. To be more particular, a distance between a specific circular point and the underlying image of the absolute conic is adopted, and a local optimization is achieved when the so-far-best model in the RANSAC iterations has been reached. The experiments on artificial and real data demonstrate that the proposed method alleviates the randomness of the RANSAC solution and get more accurate and reliable calibration results than the traditional methods.

Index Terms—2D camera calibration; RANSAC; Lo-RANSAC.

I. INTRODUCTION

Camera calibration refers to finding the parameters of the camera model. The parameters include the intrinsic matrix, which represents the perspective transformation from the scene to the image pixel in the camera coordinates system, and the extrinsic matrix, which transforms the user reference coordinates to the camera coordinates. Usually the camera lens has distortion due to machining error, especially for the wide-angle lens, which need a nonlinear distortion correction model. The performance of camera calibration has a significant influence on the accuracy of 3D reconstruction, vision inspection and robot navigation.

There are many algorithms for camera calibration, which can be classified into two categories: self-calibration and object based calibration. The self-calibration finds corresponds from several images of the static scene to recover the intrinsic and extrinsic parameters without calibration boards. This method is flexible, but the performance is not accurate [1]. Moreover, such calibration doesn't consider the lens distortion [2], which restricts its applications. The object based camera calibration is performed by observing a 3D object with known coordinates. For example, Zhang [3] proposed a flexible 2D camera calibration, which only requires the camera to observe a planar pattern shown in a few (at least two) different orientations. This method has been widely used due to its flexibility and high accuracy. It is noted that the calibration results by Zhang [3] rely on the appropriate views of the

patterns, the precision of the corner detection and so on. In other words, using distinctive subsets of images yields different calibration results. To this end, Lv etc. [2] proposed to use the RANSAC (Random Sample Consensus) [4] to remove the unreliable images. Although the method can get more robust and accurate results than the convention one, the standard RANSAC has some drawbacks: 1) The RANSAC runs much longer (even by an order of magnitude) than theoretically predicted [5]; 2) The assumption in RANSAC, i.e., a model with parameters computed from an outlier-free sample is consistent with all inliers, is not true in practice because there is noise in a real system. Accordingly, the RANSAC method always gets random inliers and leads to distinctive results. To deal with this problem, we present to employ the Locally Optimized RANSAC (LO-RANSAC) proposed in [6]. The key idea behind the solution is to select the optimal subsets of the underlying images to achieve an optimal result. The proposed approach increases the number of inliers found and returns high-quality model.

The rest of this paper is organized as follows: camera model is introduced in Section II. Section III concludes the theory of LO-RANSAC. The method using LO-RANSAC to refine the camera calibration is described in Section IV. The experiment results are shown in Section V. Section VI gives the conclusion.

II. CAMERA MODEL

Camera model includes the linear pinhole model and the nonlinear model of the lens distortion.

A. Linear Model

The camera model is normally denoted in the following equation:

$$m \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = K \begin{bmatrix} r_1 & r_2 & t \end{bmatrix} \begin{bmatrix} X \\ Y \\ 1 \end{bmatrix} \text{ with } K = \begin{bmatrix} f_x & 0 & u_0 \\ 0 & f_y & v_0 \\ 0 & 0 & 1 \end{bmatrix} \quad (1)$$

where K is the intrinsic parameters, in which, f_x, f_y denote the focal length in pixel units; (u_0, v_0) is the center of the image in pixel coordinates; r_1, r_2 are the column vectors of the rotation matrix and t is the translation vector from the world coordinates to the camera coordinates; (X, Y) is the world coordinate of known in the 2D pattern; (u, v) is the projected point in pixel coordinates which can be drawn by the corner detection. Assuming the (X, Y) and the (u, v) to be known, we can easily get the homography H between the pattern and its image as follows:

Manuscript received December 10, 2016; revised May 5, 2017.

Qin Zhang, Shiqian Wu, Wei Wang are with Wuhan University of Science and Technology China (e-mail: 1300027@wust.edu.cn).

Zhijun Fang is with the College of electrical and electrical engineering, Shanghai university of engineering and technology, Shanghai, China.

$$m \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = H \begin{bmatrix} X \\ Y \\ 1 \end{bmatrix}, \text{ with } H = \begin{bmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & h_{33} \end{bmatrix}$$

B. Nonlinear Model

The lens distortion can be divided into three classes: radial, decentering, and prism, which is totally dominated by the radial components, and especially dominated by the first term of the radial distortion [7]. In this paper, we only consider the radial distortion, because no more elaborated model can improve the accuracy significantly [8]. The model for correcting the radial distortion is given:

$$\begin{aligned} u_d &= u + k_1(u - u_0)r^2 \\ v_d &= v + k_1(v - v_0)r^2 \end{aligned} \quad (2)$$

where $r = (x + y)^2$, (x, y) is the ideal image coordinate, (u_d, v_d) is the distorted image point. The nonlinear model is to remove the distortion of the camera lens to get correct linear model. Then we can use the linear model to recover the parameters of the camera.

III. LOCALLY OPTIMIZED RANSAC

A. Standard RANSAC

In computer vision, RANSAC [2] is widely used for robust estimating a model from a data set contaminated by gross error. First a minimal subset of the data is randomly selected from the input data to estimate the model parameters. Then compute the distance between the data points and the model, the point of which the distance is less than the given threshold is considered as an inlier point. The random sampling iterates until the probability of finding a model which contains more inliers than the current model become low, i.e. the probability of missing a set of inliers of size s within k samples falls under predefined threshold thr .

$$k = \frac{\log(1 - \eta)}{\log(1 - \varepsilon^s)} \quad (3)$$

It is noted that the RANSAC method assumes that the model computed from the inliers is consistent with all the inliers. This is not always true since the inlier points can be contaminated by noise. The standard RANSAC often takes more than k times iteration to find the right model [5]. Moreover, the RANSAC selects the minimal number of samples to compute the model, in which noise can have a significant influence on the computed model. This leads to the fact, i.e., the set of inliers may be smaller than the true inliers, and with a random variety. Our experiments showed that the estimated model by the standard RANSAC method always deviates from the trues with randomness.

B. LO-RANSAC

To cope this problem, the authors [6] proposed the LO-

RANSAC, in which the local optimization of so-far-the-best samples is conducted to increase the number of the inliers found. The local optimization method uses an inner RANSAC with iteration, which is explained in the following two paragraph.

The process of the inner RANSAC is as follows: when the new model of the best-so-far is found in the k th step of the RANSAC, new samples are only selected from the inliers of the best model, hence the number of the samples need not be the minimal. Experiments showed that more number of the samples result in more accuracy of the estimated model [6].

After inner RANSAC operates, to get more reliable inliers, we use an iterative scheme: take all data points within a large threshold ($n * thr$) and use least-square algorithm to compute new model parameters. Then, gradually reduce the threshold, i.e., reduce n and iterate until the threshold is small (thr). The idea behind the iteration is that in the least-square algorithm, a single far-outlying data point will cause a destruction of the estimated model parameters. In the iteration, there are no far-outlying data, as each data point has error below the threshold ($n * thr$) subject to the sampled model.

The inner RANSAC with iteration of least-square can reduce the affection of the noise in the inlier point and the randomness of the RANSAC. Then we can get a more robust and accurate model.

In Zhang's 2D pattern based calibration method, the calibration results are affected by the quality of the image sets, it can be indicated in [3] that the best calibration performance seems to be achieved with an angle around 45 of the model plane to the image plane, and in practice when the angle increases, foreshortening makes the corner detection less precise. Also, the distance between the image and the model plane can influence the accuracy of the corner detection. All these can impact the quality of the sampled images. different image subsets lead to distinctive calibration results for some images become unreliable ones and as the outliers to the camera model. To remove the outliers and get optimal inliers of image sets, we use the LO-RANSAC to remove the unreliable images. As it is analyzed before, the LO-RANSAC can get more inliers than the stand RANSAC does, and it can reduce the randomness of results by the stand RANSAC. Here we try to improve the 2D pattern based calibration method by the LO-RANSAC. In practice, it is easy for us to get many images of the 2D calibration pattern. Then we can use the LO-RANSAC to get the optimal inliers of the image sets automatically, finally we can get an optimal calibration results without much randomness.

IV. CAMERA CALIBRATION WITH LO-RANSAC

In Zhang's camera calibration, two constrains are put in solving (1), i.e.,

$$\begin{aligned} h_1^T K^{-T} K^{-1} h_2 &= 0 \\ h_1^T K^{-T} K^{-1} h_1 &= h_2^T K^{-T} K^{-1} h_2 \end{aligned} \quad (4)$$

where h_1, h_2 are the column of the homography matrix H . Hence, at least two equations from (4) are needed to compute $K^{-T}K^{-1}$, in [9], which is called the image of absolute conic (IAC) in [9]. Here we don't need to know the meaning behind it, we just need to know it represent the matrix $K^{-T}K^{-1}$. The closed form solution of (1) can be obtained after that. Finally, the closed form solution is used as the initial value to refine the whole parameters including the lens distortion by Levenberg-Marquardt algorithm. Hence the IAC is crucial for the nonlinear optimization to avoid local optimum. To remove unreliable IAC, the authors [2] defined a distance between the homography and the IAC as follows.

$$d = \frac{(h_1^T B h_2)^2}{B h_1(1)^2 + B h_1(2)^2 + B h_2(1)^2 + B h_2(2)^2} + \frac{(h_3^T B h_4)^2}{B h_3(1)^2 + B h_3(2)^2 + B h_4(1)^2 + B h_4(2)^2} \quad (5)$$

It is noted that h is the column of the homography and $Bh(j)(j=1,2)$ represents the elements of the vector Bh . The threshold is given by computer simulation (refer to section V). In our work, the thr is set to be $2*10^{-5}$. The proposed algorithm is summarized below:

1. Take abundant images of the model plane under different orientations by moving the model plane;
2. Extract the feature points in the images and compute the homography matrix for each image;
3. Set parameters for RANSAC method: $s = 2$, $\varepsilon = 0$, $k = \infty$, $thr = 2*10^{-5}$, $i = 1$;
4. Randomly select two homography matrixes and estimate the IAC according to (4);
5. Identify a set of inliers consistent with the evaluated IAC according to the threshold value thr ;
6. Update the sample count value by (3), if a larger consistent is found; local optimization.
7. If $i > k$, go to step 8; else $i = i + 1$, go to step 4;
8. Once the largest consensus is obtained, the subset of calibrating images is determined. Then the camera parameters are calculated by Zhang's method.

V. EXPERIMENT RESULT

A. Results on Synthetic Data

If we synthesize camera parameters randomly, it will not be robust to larger noise, thus when we test the algorithm using the images within larger random noise, it often caught to get unreal solution. Here we derive the synthetic camera parameters from the reality, the calibration results in [10]. we get the parameters of following. The intrinsic parameters, $f_x = 657.384416175761$; $f_y = 658.058046335663$; $u_0 = 303.625818604402$; $v_0 = 244.843359357986$. The extrinsic parameters are the rotation vector r and the translation vector t , in which the rotation vector r is represented by the Rodrigues notation of rotation matrix from world coordinate system to camera coordinate system.

The 20 different kinds of rotation and translation vectors are shown in Table I and Table II respectively. We set the lens distortion parameter as 0. The synthesized model plane is a chessboard pattern containing $12 \times 13 = 144$ corner points. The size of pattern is 30 mm \times 30 mm. using these parameters, we can generate the image coordinates of the corner points in the 2D model pattern. Finally, we use the synthetic image coordinates of the corner and the known world coordinates of the corner to simulate the calibration algorithms.

TABLE I: ROTATION VECTORS R IN RODRIGUES NOTATION

| Rotation Vector(r) | r(1) (rad) | r(2) (rad) | r(3) (rad) |
|--------------------|------------|------------|------------|
| r1 | -0.88368 | -0.12948 | -0.00142 |
| r2 | -0.53129 | -0.11972 | 0.026916 |
| r3 | -0.5286 | -0.01454 | 0.170982 |
| r4 | -0.62672 | 0.43339 | 0.139589 |
| r5 | -0.81044 | -0.50949 | 0.504976 |
| r6 | 0.740578 | -0.09123 | 0.119276 |
| r7 | 0.574271 | -0.67371 | -0.03175 |
| r8 | 0.5394 | -0.76097 | -0.06779 |
| r9 | 0.296679 | 0.497443 | 0.307781 |
| r10 | 0.504538 | 0.016748 | 0.137809 |
| r11 | 0.48148 | -0.03432 | 0.123936 |
| r12 | 0.450067 | -0.12876 | 0.095866 |
| r13 | 0.423694 | -0.14126 | 0.081879 |
| r14 | 0.237357 | -0.21432 | 0.064762 |
| r15 | 0.605123 | 0.228369 | 0.183932 |
| r16 | 0.863768 | 0.505338 | 0.211123 |
| r17 | 0.787335 | -0.61574 | -0.04347 |
| r18 | 0.563541 | -0.94827 | -0.15624 |
| r19 | -0.17426 | -0.01153 | 0.211251 |
| r20 | 0.332693 | 0.690128 | 0.347512 |

TABLE II: TRANSLATION VECTORS T

| Translation Vector (t) | t(1) | t(2) | t(3) |
|------------------------|----------|----------|----------|
| t1 | -177.038 | -85.5073 | 852.5449 |
| t2 | -154.489 | -160.938 | 756.8986 |
| t3 | -124.752 | -176.175 | 774.7837 |
| t4 | -63.98 | -156.488 | 778.3091 |
| t5 | -91.8725 | -230.786 | 735.7576 |
| t6 | -148.555 | -80.589 | 444.3836 |
| t7 | -82.6357 | -78.6925 | 439.5382 |
| t8 | -169.658 | -104.548 | 461.2175 |
| t9 | -28.9742 | -302.186 | 858.953 |
| t10 | -150.491 | -236.925 | 703.8096 |
| t11 | -133.023 | -178.56 | 604.1523 |
| t12 | -132.299 | -144.759 | 544.1578 |
| t13 | -123.219 | -138.183 | 490.157 |
| t14 | -198.764 | -135.526 | 474.2777 |
| t15 | -134.843 | -139.949 | 489.3394 |
| t16 | -185.005 | -158.703 | 440.5264 |
| t17 | -106.28 | -80.2875 | 333.4269 |
| t18 | -143.429 | -88.845 | 395.2667 |
| t19 | -15.4968 | -171.741 | 694.9975 |
| t20 | -1.64485 | -226.691 | 727.0604 |

The image coordinates of the corner points are added with mean 0 and standard deviation Gaussian noise. First the standard deviation is varied from 0 to 1.0 pixel with the step of 0.1 pixel. The distance of the homography to the true IAC is computed, the results are shown in Fig. 1. The noise is about 0.2 pixel in reality, so we set the thr to be $2*10^{-5}$. Next, we change the standard deviation from 0 to 4 pixels with the step length of 0.2 pixel. In each noise level, the RANSAC and the LO-RANSAC are executed for 100 times. The calibration results are shown in Fig. 2 and Fig. 3. In Fig. 2, a) and b) represent the mean relative error of the calibrated f_x and f_y with respect to the ground truth in 100

iterations; c) and d) represent the absolute error of u_0 and v_0 with respect to the ground truth. In each figure, the red one represents the RANSAC method; the green one represents the LO-RANSAC method. We can see that the mean error by LO-RANSAC are smaller than the one by RANSAC, especially for u_0, v_0 . This indicates that our proposed method can get more accurate results than the

RANSAC method. In Fig. 3, a), b), c), d) represent the standard deviation of the error in Fig. 2 also the red one represents RANSAC and green one represents LO-RANSAC. We can see that the standard deviation by the LO-RANSAC is smaller than the one by RANSAC, especially for the u_0, v_0 , which shows that our method is more stable than the RANSAC.

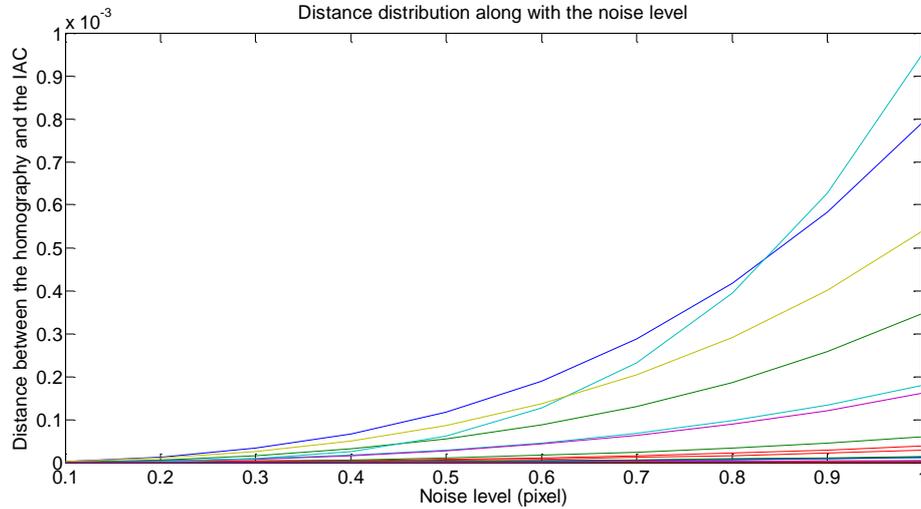
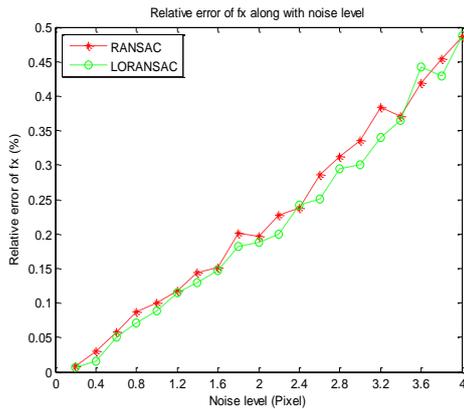
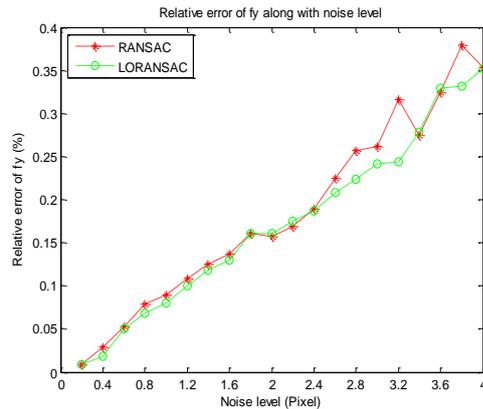


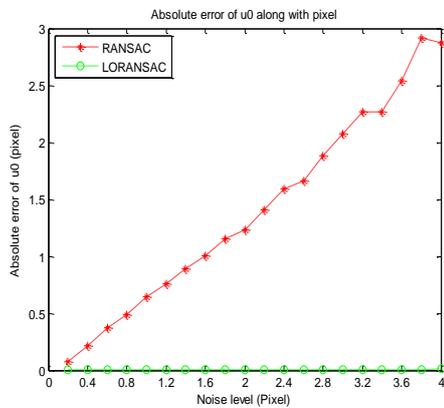
Fig. 1. Distance between the homography and the IAC along with the noise level in 20 different homographies.



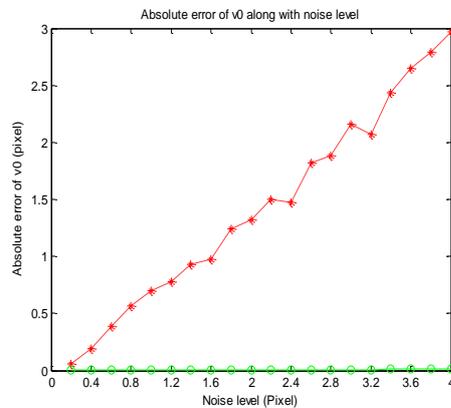
a) Relative error of f_x



b) Relative error of f_y



c) Absolute error of u_0



d) Absolute error of v_0

Fig. 2. Mean error of the calibration results, a) relative error of f_x , b) relative error of f_y , c) absolute error of u_0 , d) absolute error of v_0 .

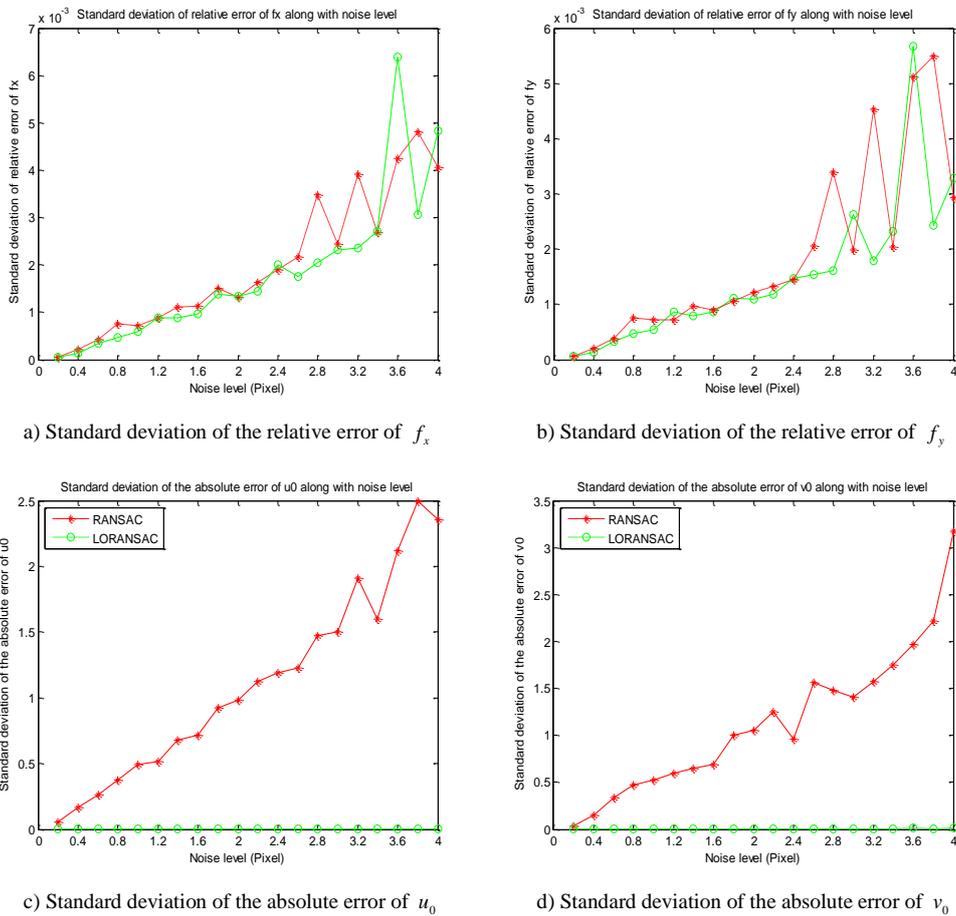


Fig. 3. Standard deviation of error of the calibration results, a) standard deviation of the relative error of f_x , b) standard deviation of the relative error of f_y , c) standard deviation of absolute error of u_0 , d) standard deviation of the absolute error of v_0 .

B. Results on Real Data

In this part, we use 20 images pattern from the online sources [10] to do the calibration. As there is no information about the true parameters, we use the mean re-projection error (error between the detected image coordinates and the re-projected image coordinates by using the parameters get by the calibration method and the known world coordinates of the pattern) to evaluate the quality of the different calibration methods. Both the RANSAC and the LO-

RANSAC execute for 100 times. The mean and standard deviation of the re-projection error are shown in Table III. As we can see from the Table III, both the mean and standard deviation of the re-projection error in RANSAC are bigger than the one in LO-RANSAC. The standard deviation of the LO-RANSAC is near zero, which implies that using the LO-RANSAC can get a more accurate and robust calibration result than using the RANSAC method.

TABLE III: CALIBRATION WITH REAL DATA

| Calibration method | Mean reprojection error | Standard deviation of the error |
|--------------------|--------------------------|---|
| RANSAC | 0.159679252519579 | 0.004770649100921 |
| LO-RANSAC | 0.156598456968103 | 1.952678215459548 * 10⁻¹⁶ |

VI. CONCLUSION

Camera calibration is a fundamental issue in computer vision, especially in visual measurement. The measurement performance is significantly dependent on calibration accuracy, which is affected by experimental conditions as well as calibration method. In this paper, the Locally Optimized RANSAC (LO-RANSAC) is adopted to remove unreliable images automatically. The LO-RANSAC determine the optimal image sets as the inlier to the camera model by define a distance between a specific circular point

and the underlying image of the absolute conic, where the LO-RANSAC is achieved by execute a local optimization when the so-far-best model in the RANSAC iterations has been reached. The local optimization step can get rid of the noise which can influence the performance of the RANSAC in the sampled inliers and get optimal image sets as the inlier to the image model. Hence get the optimal calibration results automatically. Experiments on computer simulation and the real data demonstrate that our method can get more accurate and robust calibration results than the traditional one.

ACKNOWLEDGMENT

This work was supported in part by the National Natural Science Foundation of China under Grant 61371190. The authors wish to acknowledge the anonymous reviewers' insightful and inspirational comments that have helped to improve the technical contents and readability of this paper.

REFERENCES

- [1] S. Bougnoux, "From projective to euclidean space under any practical situation, a criticism of self-calibration," in *Proc. IEEE International Conference on Computer Vision (ICCV 98)*, IEEE, CS Press, pp. 790–796, Dec. 1998.
- [2] Y. Lv, J. Feng, Z. Li, W. Li, and J. Cao. "A new robust 2D Camera calibration method using RANSAC," *Optik -International Journal for Light and Electron Optics*, vol. 126, pp. 4910–4915, Dec. 2015.
- [3] Z. Zhang, "A flexible new technique for camera calibration," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, pp. 1330–1334, Dec. 2000.
- [4] M. A. Fischler and R. C. Bolles, "Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography," *Communications of the ACM*, vol. 24, pp. 381–395, Dec. 1981.
- [5] B. Tordoff and D. W. Murray, "Guided sampling and consensus for motion estimation," in *Proc. European Conference on Computer Vision (ECCV 02)*, Springer-Verlag Press, pp. 82–96, Dec. 2002
- [6] O. Chum, J. Matas, and J. Kittler, "Locally optimized RANSAC," *Lecture Notes in Computer Science*, vol. 2781, pp. 236–243, Dec. 2003.
- [7] D. C. Brown, "Close-range Camera Calibration," *Photogramm Eng*, vol. 37, pp. 855–866, Dec. 2002.
- [8] J. Salvi, X. Armangué, and J. Batlle, "A Comparative review of camera calibrating methods with accuracy evaluation," *Pattern Recognition*, vol. 35, pp. 1617–1635, Dec. 2002.
- [9] R. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*, 2nd ed., vol. 30, Cambridge University Press, 2003, pp. 1865–1872.
- [10] Camera Calibration Toolbox for Matlab. [Online]. Available: http://www.vision.caltech.edu/bouguetj/calib_doc/



Qin Zhang was born in Hubei, China on January 13, 1992. She received the bachelor of engineering degree in mechanical engineering and automation from Wuhan University of Science and Technology in 2015. She is currently working toward the Ph.D. degree in mechanical engineering at the same university.

Her research interests include the image processing and computer vision. This is the first paper she has published. At present, she is working in the Key Laboratory

of Metallurgical Equipment and Control Technology, Ministry of Education, Wuhan University of Science and Technology, Wuhan, China as a full-time student.



Shiqian Wu received the B.Eng. and M.Eng. degrees from the Huazhong University of Science and Technology (HUST), Wuhan, China, in 1985 and 1988, respectively, and the Ph.D. degree from Nanyang Technological University, Singapore, in 2001. He is currently a Professor with the School of Machinery and Automation, Wuhan University of Science and Technology, Wuhan.

He was an Assistant Professor, a Lecturer, and an Associate Professor at HUST, from 1988 to 1997. From 2000 to 2014, he was a Research Fellow or Research Scientist with the Agency for Science, Technology and Research, Singapore. He has co-authored the book entitled *Dynamic Fuzzy Neural Networks* (Singapore: McGraw-Hill, 2003).

He has authored or co-authored over 160 scientific publications (book chapters and journal/conference papers). His current research interests include image processing, pattern recognition, machine vision, fuzzy systems, and neural networks.



Wei Wang was born in Wuhan, Hubei on 6th September 1990. She received the bachelor degree and Meng degree from Wuhan University of Science and Technology and University of Ottawa in 2012 and 2014 respectively. At present, she is pursuing her Ph.D. degree in mechanical engineering in Wuhan University of Science and Technology. Her research interests include high-dynamic-range imaging and image calibration.



Zhijun Fang was born in Jiangxi, China in September, 1971. He received his Ph.D. degree in Shanghai Jiao Tong University. He is currently a professor in College of electrical and electrical engineering, Shanghai university of engineering and technology, Shanghai. He has presided 5 national or equivalent level research projects including 2 national science foundation of China projects. His research interest includes video image analysis, machine vision and intelligent data analysis.