Proceeding of the IEEE
International Conference on Robotics and Biomimetics
Dali, China, December 2019

Virtual Calibration Pattern Realization for Hybrid-FOV Multiview Calibration*

Zhong Chen, Qisen Wu, and Xianmin Zhang
Guangdong Provincial Key Laboratory of Precision Equipment and Manufacturing Technology
South China University of Technology
Guangzhou, Guangdong, 510640, China
{*mezhchen, zhangxm@scut.edu.cn*}*@scut.edu.cn*

***Abstract*— Multiview vision systems are commonly applied in different fields at present, including hybrid-field-of-view(hybrid-FOV) multiview systems. The difference between cameras' FOVs results in the difficulty of calibration, such as lack of information for cameras with small FOVs. All the information acquired by the other cameras and the projective geometry between them should be utilized in camera calibration. In this paper, a method of generating redundant constraints with trifocal tensor and the information from the cameras with a large FOV is proposed. The process is described after some reminders of trilinearities. Then, two simulated experiments based on this method have been implemented. The results and analysis show the feasibility of this method.**

***Index Terms*— Multiview geometry, virutal calibration pattern, trifocal tensor.**

I. Introduction

With rapid development of computer vision and image sensing technique, two-view or multiview 3D (Three Dimension) vision technique has been being applied in practical manufacturing process and multidisciplinary scientific researches. Micro-processing and micro/nano robotic manipulation in precision engineering, such as cell injection-force measurement[1], cell micro manipulation [2] and micro injection [3], [4], usually demand 3D geometric information of the manipulated objects based on macro/micro 3D vision in hybrid-FOV (Field of View) multiviews. In order to improve their 3D measurement precision in this kind of configuration, key issues of its calibration should be solved.

Classical calibration for two-view stereo vision system usually is performed by deducing the epipolar geometry from the projection matrices of two cameras after the projection matrix of each camera is estimated using calibration pattern. But this kind of calibration method is only valid for the space near the calibration pattern and the reconstruction precision will becomes worse when the object moves far from the calibration space. Zhang [5] presented a comprehensive calibration method by getting the best use of the rigid geometry relation between two cameras for calibration validness in a wide range in space from the calibration pattern. Xue [6] presented a flexible and low-cost calibration method for a structure-uniform stereovision system including two cameras and a variable projector. Zhao [7] simplified the calibration process for a binocular vision system using 1D calibration pattern through linearization of the colinear equations.

Three-view or multiview vision system can provide more constraints that a two-view vision system, which can stabilize reconstruction and reduce reconstruction deviations [8]. Yang [9] calibrated a infrared (IR) camera with two visible cameras, but did not get the best use of the geometry relation among three cameras. Lu [10] presented a geometric calibration method for IR camera using trinocular vision.

When three-view or multiview vision have different field of view (FOV), the calibration may face the information-loss problem, since pattern geometry can not be obtained entirely by the cameras with FOVs of different sizes simultaneously. This may cause negative effect on the hybrid-FOV multi-view vision calibration. Facing this issue, this paper develops a virtual calibration pattern realization approach for a hybrid-FOV three-view vision system including two camera with a same large FOV and a camera with a small FOV. General calibration patterns can be more applicable through this method. Furthermore, it make it possible to improve the calibration precision for hybrid-FOV multiview vision system.

The following parts of this paper are arranged as follows. First, the basic theory of Multi-view geometry is described in Section II. In Section III, the virtual calibration pattern realization method for a hybrid-FOV tree-view vision is presented. Finally, the part of experimental results are given and discussed in Section IV.

II. Multiview Geometry

A. Epipolar Geometry

Epipolar geometry is a very essential characteristic of binocular vision. Independent of scene structure, it can be considered invariant as the relative pose between two cameras fixed. The fundamental matrix, which is basically a refection of epipolar geometry, can be computed precisely with either matching points from two cameras or two camera projection matrices after calibration. Assuming that a stereo camera is

*The work described in this paper is supported by National Natural Science Foundation of China #51875204 to Z. Chen who is corresponding author.

978-1-7281-6321-5/19//$31.00 ©2019 IEEE

calibrated precisely, the fundamental matrix should be unique and invariant.

In this paper, a trinocular vision is separated into a calibrated binocular vision and an isolated camera. As the projection geometry setting up among these cameras, the epipolar geometry can be a reference to tell whether this setting is correct.

B. Trifocal Tensor and Trilinearity

Similar to the fundamental matrix, the trifocal tensor reflects the intrinsic projection between three views, constituted by a set of three matrices $\{\mathbf{T}_1, \mathbf{T}_2, \mathbf{T}_3\}$ in matrix denotion. It can be derived with three corresponding lines $\mathbf{l}$, $\mathbf{l}'$, $\mathbf{l}''$ as a projection of the same line in object coordinate[8]:

$$l_i = \mathbf{l}'^{\top} \mathrm{T}_i \mathbf{l}'' \tag{1}$$

or

$$\mathbf{l}^{\top} = \mathbf{l}'^{\top} [\mathrm{T}_{1,}, \mathrm{T}_2, \mathrm{T}_3] \mathbf{l}'' \tag{2}$$

where $\mathbf{l}$, $\mathbf{l}'$, $\mathbf{l}''$ are the homogeneous column 3-vectors denoting the lines projected in three different views, i.e. $\mathbf{l}^{\top} = (l_1, l_2, l_3)$. Eq. (1) can be recognized as the foundation of trilinearities. The point-point-point trilinearities, derived by (1) and presented in [12], is described by

$$x^i \left(x'^j \epsilon_{jpr}\right) \left(x''^k \epsilon_{kqs}\right) T_i^{pq} = 0_{rs} \tag{3}$$

where x^i, x'^j, x''^k denote the *i*-th, *j*-th and *k*-th entry of homogeneous column 3-vectors $\mathbf{x}$,$\mathbf{x}'$,$\mathbf{x}''$ indicating the points corresponded in three views. ϵ_{jpr} and ϵ_{kqs} are tensor notations. And T_i^{pq} denotes the q-th entry in p-th row of the i-th matrix of $\{\mathbf{T}_1, \mathbf{T}_2, \mathbf{T}_3\}$. ϵ_{jpr} and ϵ_{kqs} are tensor notation. With (3) and point correspondences, trifocal tensor can be computed by direct linear transformation(DLT). In this paper, the relation between views is basically computed with point correspondences.

As stated above, the trifocal tensor corresponds to the relation among all three cameras. So the fundamental matrices and the epipoles can be computed from the trifocal tensor. Furthermore, camera matrices can be extracted up to a projective ambiguity under the assumption of $\mathbf{P} = [\mathbf{I}|\mathbf{0}]$,$\mathbf{P}' = [\mathbf{A}|\mathbf{a}]$ and $\mathbf{P}'' = [\mathbf{B}|\mathbf{b}]$ [8][14]:

$$\mathbf{P}' = [[\mathbf{T}_1, \mathbf{T}_2, \mathbf{T}_3]\, \mathbf{e}''|\mathbf{e}'] \tag{4}$$

$$\mathbf{P}'' = \left[\left(\mathbf{e}''\mathbf{e}''^{\top} - \mathbf{I}\right) \left[\mathbf{T}_1^{\top}, \mathbf{T}_2^{\top}, \mathbf{T}_3^{\top}\right] \mathbf{e}'|\mathbf{e}''\right] \tag{5}$$

where $\mathbf{P}'$ and $\mathbf{P}''$ are the closed form solution of camera matrices, with $\mathbf{e}'$ and $\mathbf{e}''$ referring to the epipoles on the other two views corresponding to the first view. Responding to what have been described previously, $\mathbf{P}'$ and $\mathbf{P}''$ correspond to the stereo camera in trinocular vision. Supposedly, the epipolar geometry of stereo camera can be a reference to tell the precision of the trifocal tensor.

Although (4) and (5) perform unstably because of noise[13], but they can still derive the equation for retrieving fundamental matrix of stereo camera:

$$\mathbf{F}_{32} = \left[\mathbf{e}'' - (\mathbf{e}''\mathbf{e}''^{\top} - \mathbf{I})\mathbf{M}\mathbf{e}''\right]_{\times} \mathbf{M} \tag{6}$$

where the denotation $[\bullet]_{\times}$ indicates a skew-symmetric matrix and

$$\mathbf{M} = [\mathbf{T}_1^T\mathbf{e}', \mathbf{T}_2^T\mathbf{e}', \mathbf{T}_3^T\mathbf{e}'][\mathbf{T}_1\mathbf{e}'', \mathbf{T}_2\mathbf{e}'', \mathbf{T}_3\mathbf{e}'']^{-1}$$

In order to evaluate the computation of trifocal tensor, the accuracy can be quantify by evaluate the fundamental matrix computed by (6). A cost function based on the Sampson approximation for fundamental matrix presented in [15] [8] has been applied as

$$\sum_n \frac{\left(\mathbf{x}_n'\mathbf{F}\mathbf{x}_n\right)^2}{\left(\mathbf{F}\mathbf{x}_n\right)_1^2 + \left(\mathbf{F}\mathbf{x}_n\right)_2^2 + \left(\mathbf{F}^{\top}\mathbf{x}_n'\right)_1^2 + \left(\mathrm{F}^{\top}\mathbf{x}_n'\right)_2^2} \tag{7}$$

where the index n indicates the n-th point correspondence of the information set, and $\left(\mathbf{F}\mathbf{x}_n\right)_i^2$ refers to the square of the i-th entry of vector $\mathbf{F}\mathbf{x}_n$. In this paper, calculation of (7) will be used to figure out some abnormality of a actual program based on this paper.

C. Transferring Points

With the trifocal tensor computed and the trilinearities set up, a image point $\mathbf{x}$ in one view can be extracted if the rest of this correspondence $\mathbf{x}'$,$\mathbf{x}''$ are already acquired[13]. This process, like retrieving epipolar lines with fundamental matrices, can be recognized as point transfer.

There are three ways to preform point transfer. The most complicated one is to solve the formula derived from 3 since it utilizes redundant excessive constraints and cost to find a single image point. On the other hand, referring to eipoplar geometry, $\mathbf{x}$ must be on the epipolar lines corresponding to $\mathbf{x}'$ and $\mathbf{x}''$. So retrieving the fundamental matrices, computing epipolar lines and extract the intersections of these line might be a simpler way, which meets the constraints from epipolar geometry. But unfortunately its performance can be severely influenced by noise and badly-computed fundamental matrices. Last but not least, for the purpose of taking advantage of trilinearities, Ref.[8] presented a method that utilizes point-line-point correspondence:

$$\mathrm{l}'^{\top} \left(\sum_i x^i \mathrm{T}_i\right) [\mathrm{x}'']_{\times} = \mathbf{0}^{\top}. \tag{8}$$

The most fascinating part of this method is that $\mathbf{l}'$ can be any lines $\mathbf{x}'$ lying on but the epipolar lines. So the computation cost can be greatly reduced compared to the first method. And this method is more reasonable than the second one when it comes to accuracy and difficulty to implement.

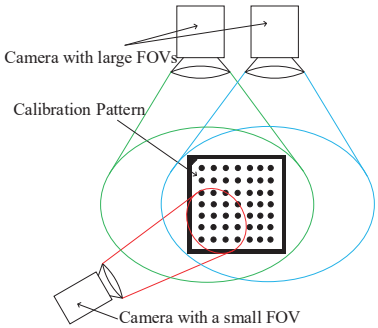


Fig. 1. Schematic of the hybrid-FOV multiview system calibration.

III. Hybrid-FOV Multiview System and Virtual Calibration Pattern

A. Issue of Calibrating Hybrid-FOV System

Hybrid-FOV multiview system usually consist of cameras with different FOVs for satisfying special demand, such as micro manipulation. When the FOVs scales down to 1 centimeter or less, the smaller the FOVs of cameras, the harder to calibrate. The reason is that manipulators are barely compatible with calibration patterns. Furthermore, if hybrid-FOV system is designed to implement 3D reconstruction, the projective geometry among different views should be calibrated. But, one calibration pattern lacks compatibility of FOVs of different sizes. It is a consensus that the FOV should be covered by the calibration pattern in sufficient size in case of noise, deviation and instability of algorithm. If the calibration pattern is not large enough, the calibration will be invalid outside the section that it covers. That means, unless some really special custom products are built for hybrid-FOV system, the mass-produced calibration patterns can't work on all the cameras simultaneously.

If a calibration implemented with normal calibration patterns on hybrid-FOV system, the camera with a small FOV will fail to observe some of the feature points, even the entire patterns in some particular poses. Definitely, the calibration will suffer great lost of accuracy in this way, and it will get even worse when the feature points arrange in some form risking degeneracy. Meanwhile, if the cameras with large FOVs are calibrated precisely, the image, or information outside the a small FOV are still available to the system. Without doubt, the system will perform badly until all the information can be entirely utilized.

B. Process of Generating Virtual Calibration Pattern

Since the cameras with large FOVs can still acquire information that the others with a small FOVs fail to observe, the projective geometry may help to generate information for a valid calibration, approximated generating a calibration pattern. We call this set of information virtual calibration pattern, because the camera can't actually observe them.

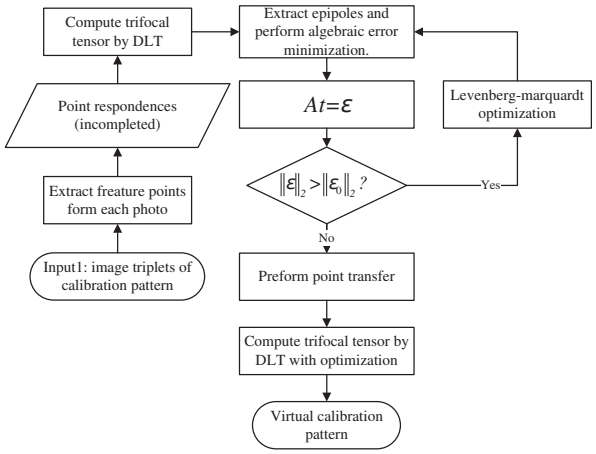


Fig. 2. Flow chart of generating virtual calibration pattern.

Since one image point can be computed from the other two known points or lines with trilinearities, it takes an important role in this process. In fact, through point or line transfer, the information acquired by other view provide a set of reference. If the estimation of trilinearities is precise enough, all information from other cameras will be able to make up the insufficient information for calibration.

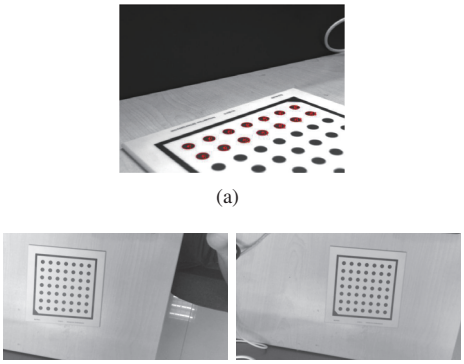
As a point correspondence offering four formulas[13] , 7 point correspondences will be sufficient for computing all 27 entries of the trifocal tensor, slightly larger than 6 for computing a camera matrix. In addition, the computation could be completed with less point correspondences if minimal parameterization was applied[16][17][18]. Therefore, we can still get the trilinearties when camera calibration seriously need redundant constraints, despite the tiny set of information. After generating the virtual calibration pattern, calibration can be carried out in usual ways with them.

There are three phase in this process. First, a minimization algorithm provided in [8] will be carried out with a incomplete set of information for acquiring a good estimation of the trifocal tensor. Then, we perform the point transfer with the rest of the information to generate virtual calibration pattern. After that, the whole set of the information extracted from image and virtual calibration pattern will be applied on another computation for trifocal tensor. The whole process is illustrated in Fig. 2.

IV. Experiments and Analysis

A. Virtual Calibration Pattern Evaluation with Single Triplet

To evaluate the feasibility of generating a virtual calibration pattern, an experiment was implemented with a single image triplet, shown in Fig. 3. The trinocular system adopted



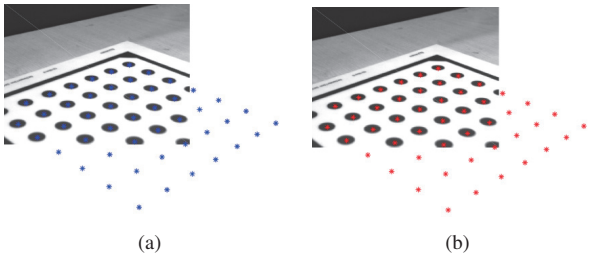
(a) (b) (c)

Fig. 3. A image triplet used in this experiment, which are: (a) The image captured by camera with small FOV; (b)(c)Two images captured by the stereo camera.

in this experiment is consist of a Lena HNY-CV-002 stereo camera (Both are: Resolution:1280×720; Sensor Size:$1/4''$; Cell Size: 3.0μm) with two lenses (Focal length:3.1mm) and a MindVision MV-UB130GM-T camera(Resolution: full resolution was 1280×1024, set to 640×480, Sensor Size:1/3″, Cell Size: 5.2μm) with Computar M0814-MP2 lens (Focal length: 8mm; Max aperture ratio: 1:14). The planar calibration pattern used in the experiment is Caliboptics CO-50-H-15 grid plate (Dot diameter: 7.5mm; Dot center spacing 15mm). A part of the visible feature points in the small view were not applied on the initial computation. Instead, these points were considered as references or the ideal position of transferred points.

There were 49 points extracted from both views of the stereo cameras, and 28 from the stand-alone camera. In the first computation phase for the trifocal-tensor, 14 dot points was used, marked in Fig. 3(a). After that, the point transfer was performed using all 49 point correspondences. As a comparison, it was performed simultaneously in two different ways to find out a better solution. With result of point transfer, all the point correspondences were utilized to compute trifocal tensor again.

The result of the first point transfer is shown in Fig. 4(a).

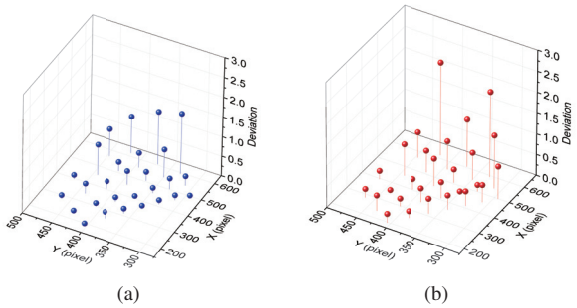


(a) (b)

Fig. 4. Results from point transferring are indicated as (a) After the first computation of the trifocal tensor ; (b) After the second computation of teh trifocal tensor using the whole set of point correspondences.

This seems reasonable, because the transferred points nearly kept their colinear relationship.

Deviation between the original set of points extracted from images directly and the point-transfer result was shown in Fig. 5(a). It is obvious that the deviation of each correspondence is less than 1 pixel. The further the points are from the part for computing trifocal tensor, the greater their deviations. The most distant visible points suffer the greatest, but the deviations are less than 2 pixels, slightly larger than 1 pixel. In total, 2 of the 28 points deviated from their ideal position larger than 1 pixel. Apart from what was stated, we also performed point transfer by utilizing point-line-point correspondences, and the deviations were shown in Fig. 6. We can observe in Fig. 6 that severe deviations would be caused with the point-line-point method compared to the one we performed previously. The deviation of trifocal tensor and computer algorithm might result in it.



(a) (b)

Fig. 5. Deviations of point transfer, which are: (a) After the first computation of trifocal tensor ; (b) After the second computation of trifocal tensor with the whole set of point correspondences.

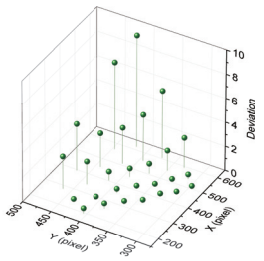


Fig. 6. Deviations of point transfer with point-line-point correspondence.

Utilizing all the correspondence, we recomputed the trifocal tensor, and performed point transfer. The result is shown in Fig. 4(b), which looks very similar to the previous one shown in Fig. 4(a). The difference is, the arrangement of point-transfer result looks even closer to a actual projected calibration pattern compared to the previous set. Still, the deviated distance of each visible point correspondence was computed, shown in Fig. 5(b). Although the deviation increased, most of the point-transfer result kept the deviation less than 1 pixel. 2 of them deviated from their ideal

position more than 2 pixels, but these deviations would be minimized theoretically since their deviations in x and y direction were (2.0499, 0.1052) and (1.9807, 1.0448), which were reasonable. If the trifocal tensor is computed well, the deviation can be of subpixel scale. So, with trifocal tensor and point transfer we can generate a virtual calibration patterns, consisting a valid set of redundant information for camera calibration with a small FOV in hybrid-FOV vision system without any custom made calibration patterns.

B. Experiment with Multi Triplets

For more evaluation, we implemented an experiment with multi triplets, which consist of 20 triplet images of the calibration pattern under different poses. One of these triplets was presented in Fig. 7.The maximum amount of points in each view was 6, which was closed to the actual conditions. A half of the image set was applied on the calculation of the trifocal tensor, and the rest were applied on the comparison between the point-transfer result and the extracted from image.

The deviation of the second half of the set shown presented in Fig. 8. The deviation in y direction was roughly maintain from -1 pixel to 1.5 pixel, but the outcome in x direction was disappointing. This might be caused by the inaccurate point extraction and the deviation of the trifocal tensor due to poor illumination.

Apart from the deviation cause by poor illumination this process may be valid, because the point-transfer result basically stayed closed to the correct location despite the obstacle of awful point distribution.

C. Discussion

Because most of the matrices can be derived by epipoles in a trinocular vision system, we can estimate the parameters of a system with low computation cost if the epipoles are extracted precisely. But, the extraction can be influenced by

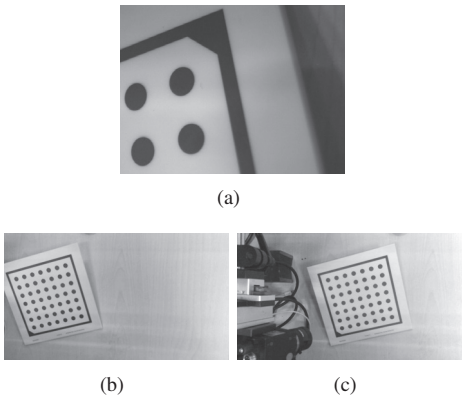


Fig. 7. One of the triplets used in this experiment, which are: (a) The image captured by camera with small FOV; (b)(c)Two images captured by the stereo cameras.

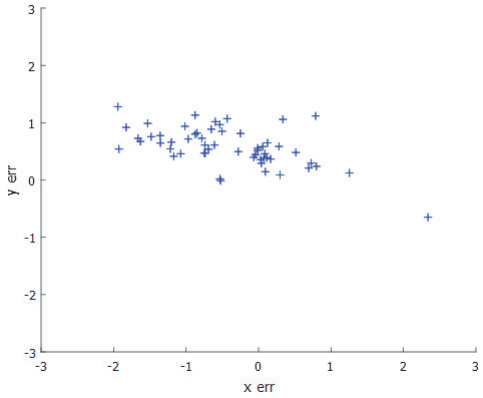


Fig. 8. Deviations of point transfer using the data set which was not applied on calculation of the trifocal tensor.

lack of pose variations severely. To ensure the accuracy of the process, the rotation of calibration pattern is essential. Making (7) the cost function of optimization for trifocal tensor may be a rational choice to ensure the accuracy of a trifocal tensor and epipoles.

V. Conclusion

To meet the demand of hybrid-FOV multiview system on calibration, a process of utilizing redundant information from the cameras with large FOVs by generating virtual calibration patterns is presented in this paper. With the virtual calibration pattern, redundant constraints can be provided to the camera with a small FOV for calibration by the rest of system, despite the lack of compatibility of usual calibration pattern. Two simulated experiments of generating were implemented to evaluate the method, and the results showed its capacity to offer redundant information while keeping accuracy. Applying this method on micromanipulation would be a rational way to decrease the cost of calibration pattern, and make it easier to implement calibration in such a small scale.

References

[1] F. Karimirad, S. Chauhan, and B. Shirinzadeh, "Vision-based force measurement using neural networks for biological cell microinjection," Journal of Biomechanics, vol. 47, no. 5, pp. 1157-1163, 2014.

[2] H. Yang, X. Li, Y. Wang, G. Feng, and D. Sun, "Control of Single-Cell Migration Using a Robot-Aided Stimulus-Induced Manipulation System," IEEE/ASME Transactions on Mechatronics, vol. 22, no. 2, pp. 815-825 , 2017.

[3] Y. Zhao, H. Sun, X. Sha, L. Gu, Z. Zhan, and W. J. Li, "A Review of Automated Microinjection of Zebrafish Embryos," Micromachines, vol. 10, no. 1, p. 7, 2019.; https://doi.org/10.3390/mi10010007.

[4] S. Permana, E. Grant, G. M. Walker, and J. A. Yoder, "A Review of Automated Microinjection Systems for Single Cells in the Embryogenesis Stage," IEEE/ASME Transactions on Mechatronics, vol. 21, no. 5, pp. 2391-2404, 2016.

[5] Z. Zhang, O. Faugeras, and R. Deriche, "An Effective Technique for Calibrating a Binocular Stereo Through and the Environment," *Journal of Computer Vision Research*, vol. 1, no. 1, pp. 58-68, 1997.

[6] T. Xue, B. W, J.G. Zhu, and S.H. Ye, "Complete calibration of a structure-uniform stereovision sensor with free-position planar pattern," *Sensors and Actuators A: Physical*, vol. 135, no. 1, pp. 185-191, 2007.
[7] Y. Zhao, X. Li, and W. Li, "Binocular Vision System Calibartion Based on a One-dimensional Target," *Applied Optics*, vol. 51, no. 16, pp. 3338-3345.
[8] R. Hartley, and A. Zisserman, *Multiple View Geometry in Computer Vision, Second Edition,*. Cambridge: Cambridge University Press, 2003.
[9] R. Yang, W. Yang, Y. Chen, and W. Wu, "Geometric Calibration of IR Camera Using Trinocular Vision," *Journal of Lightwave technology*, vol. 29, no. 24, pp. 3797-3803, 2011.
[10] R. Lu, and M. Shao, "Sphere-based Calibration Method for Trinocular Vision Sensor," *Optics and Lasers in Engineering*, vol. 90, pp. 119-127, 2017.
[11] A. Shashua, "Algebraic Functions For Recognition," *IEEE Transaction on Pattern Analysis and Machine Intelligence*, vol. 17, pp. 779-789, 1995.
[12] A. Shashua, and M. Werman, "Trilinearity of three perspective views and its associated tensor," presented at 1995 Proceedings of IEEE International Conference on Computer Vision, Cambridge, MA, USA.
[13] R. Hartley, "Lines and Points in Three Views and the Trifocal Tensor," *International Journal of Computer Vision*, vol. 22, no. 2, pp. 125-140, 1997.
[14] R. Hartley, "Projective reconstruction from line correspondences," in *Proceedings 1994 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 1994, pp. 903-097.
[15] P. D. Sampson, "Fitting Conic Sections to "Very Scattered" Data: An iterative Refinement of the Bookstein Algorithm," *Computer graphics and image processing*, vol. 18, no. 1, pp. 97-108, 1982.
[16] L. F. Julià and P. Monasse, "A Critical Review of the Trifocal Tensor Estimation," in *Pacific-Rim Symposium on Image and Video Technology*, 2017, pp. 337-349.
[17] K. Nordberg, "A minimal parameterization of the trifocal tensor," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2009, pp. 1224-1230.
[18] C. Ressl, "A minimal set of constraints and a minimal parameterization for the trifocal tensor," in *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 2002, Part 3A, vol. XXXIV.