

NETWORK ADJUSTMENT OF AUTOMATED RELATIVE ORIENTATION FOR A DUAL-CAMERA SYSTEM

K. Y. Lin ¹*, Y. H. Tseng ¹, K. W. Chiang ¹

¹ Dept. of Geomatics, National Cheng Kung University, No.1, Daxue Road, East District, Tainan, Taiwan - azsxd2014@gmail.com, (tseng, kwchiang)@mail.ncku.edu.tw

KEY WORDS: Visual Odometry, Stereo Image Pair, Relative Orientation Parameters, Least-squares, Network Adjustment

ABSTRACT:

Visual odometry (VO) is a technique applied to track the dynamic positioning and orientation of a moving platform with one or more cameras taking image sequences. The determination relies on the estimation of relative orientation parameters (ROPs) of time adjacent images. The idea of stereo VO to develop a dual-camera system is adopted in this study. By taking advantage of the calibrated stereo camera, this system is able to recover the true scale of relative translation without the need from additional sensors. However, the scale might not be very accurate, and the error also could exist in the orientation including rotation and translation due to environmental factors such as the illumination and texture. Therefore, the primary objective of this study is to find the optimized theory and method of stereo VO. Through the analysis of the geometric relationship of the time adjacent stereo image pairs, locally optimized network adjustment is developed to improve the accuracy of ROPs.

The proposed network adjustment model is verified by the simulation data and experiment data both. ROPs are adopted as observations that would update the states of the image sequence further. Besides, exterior orientation parameters (EOPs) of the dual-camera system could be optimized obviously during the whole operation. In this study, it is worth mentioning that 3D coordinates of object points matched in each image pair are not necessary to be calculated. The conventional bundle adjustment is not adopted, but more accurate EOPs still have been generated automatically during the process.

1. INTRODUCTION

Visual Odometry (VO) is the technique of determining the egomotion including the dynamic positioning and orientation of a platform by using a visual system with one or more cameras (Scaramuzza and Fraundorfer, 2011). A mobile mapping platform could be a land, aerial or underwater vehicle, which could even be an autonomous one (Bonin-Font et al., 2008). The concept and methodology of VO originated by Nister et al. (2004). VO has been developed parallelly and named differently as Visual Navigation or Vision-based Navigation in mobile mapping, navigation and robot fields. VO provides not only moving directions and distances but also three-dimensional trajectory of the sensor as shown in Figure 1. A VO system may involve real-time or post-processing time series image with automated image matching and orientation retrieving computation to generate 3D trajectories of the image sequences.

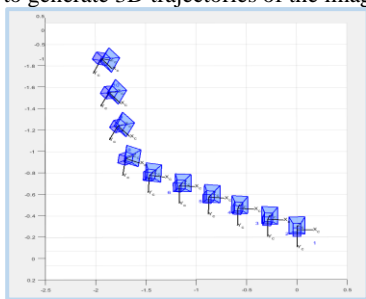


Figure 1. An example of 3D trajectory of a moving sensor generated with visual odometry.

The most famous case of applying VO is the exploration project of Mars. The space project emitted some specially designed

rovers to Mars for collecting geographic and geological information by using VO technique (Cheng et al., 2005). Onboard cameras were used to record the surroundings as well as used for navigation. VO is applied widely in autonomous driving for advanced cars and unmanned aerial vehicles (UAVs) as well. While the signal of positioning satellites is not ideal, VO can show the performance of navigation and obstacle avoidance (Bertozzi et al., 2011; Kelly and Sukhatme, 2007). Besides, there are some applications in agriculture (Ericson and Astrand, 2008) and even underwater archaeology (Foley et al., 2009) nowadays. In a VO system, the determination of platform motion between epochs relies on the estimation of the relative orientation parameters (ROPs) of time adjacent images. However, the true scale of relative translation between images is not solvable for monocular (single camera) VO. Figure 2 shows the difference between unknown and known scale. If the scale is unknown, each relative translation vector is normalized as a unit vector. If the scale is known, each relative translation vector is a real-scale vector. Therefore, how to recover the true scale for applications of monocular VO is the key issue.

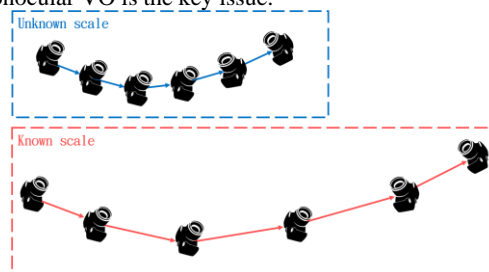


Figure 2. The difference between unknown and known scale.

The common approach is to integrate with other sensors like a wheel odometry or a GNSS to provide observations of moving

* Corresponding author

distances (Dusha and Mejias, 2012). These sensors can update the real translation to VO. Moreover, combining Inertial Navigation System (INS) is also a popular alternative (Jones and Soatto, 2011). Especially, VO integrated with INS applying to indoor navigation is more suitable and useful (Kneip et al., 2011) when GNSS signal is blocked. However, a multi-sensor system tends to have system calibration problems may affect the solution of VO. Furthermore, observation errors of sensors would be propagated and accumulated during moving as well.

Stereo VO is a dual-camera system normally installed on a horizontal bar platform. This system can take stereo image pairs simultaneously and continuously. By taking the advantage of a pair of calibrated stereo cameras, this system is able to recover the true scale of relative translation without the need of additional sensors. Hence, the error propagation from other sensors would be avoided. Figure 3 shows the geometric relationship of adjacent image pairs in stereo VO. The true scale can be recovered based on the known baseline calibrated previously between the dual camera and matching the same feature points to calculate 3D coordinates of object points. However, the scale might not be accurate, and the error also could exist in the orientation including rotation and translation according to the quality of calibration and condition of illumination and texture in the environment. Therefore, how to improve the estimated rotation and translation becomes another important issue.

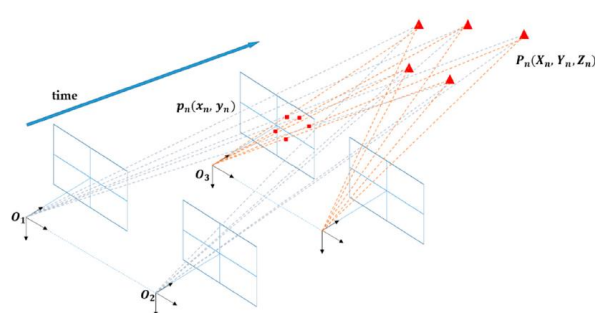


Figure 3. The geometric relationship of adjacent image pairs in stereo VO (Yoon and Kim, 2019).

Considering the aforementioned issues, this study aims to develop a dual-camera system to implement stereo VO. The primary objective is to develop a theory and robust computation algorithm for stereo VO to obtain complete navigation information without additional assistance from other sensors.

In general, same as an INS, the navigation of VO is a kind of dead reckoning process. The errors of positioning and orientation on each epoch will be accumulated, which continuously enlarges the drift errors of trajectory. Therefore, developing an optimization approach for decreasing drift errors is necessary. There are two major categories of local optimization methods. The first one applies bundle adjustment. Assuming the object point, image point and perspective center is colinear, the image points are taken as observations. The coordinates of object points and image orientations are optimized by least-squares method (Triggs et al, 1999). The second one applies pose-graph optimization like loop closure in the field of simultaneous localization and mapping (SLAM) (Grisetti et al., 2010). When the platform moves and detects the area as same as the previous visited, the trajectory could become a close graph, which is considered as constraints to optimize the orientation of related images in this scene. Figure 4 shows the illustration of these two methods. However, while the number of image points increases, the calculation of bundle adjustment enlarge as well that needs much more computer resource to implement. For pose-graph optimization, there is also a limitation that the platform must revisit the place to form a

closure. Passing by the same position is not necessary for the navigation purpose in reality.

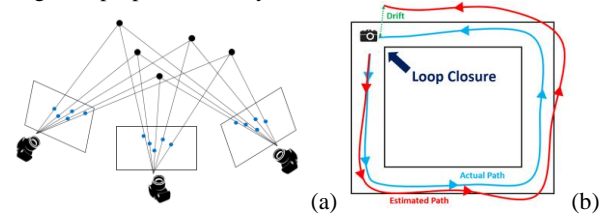


Figure 4. Local optimization (a) Bundle adjustment (b) Pose-graph optimization.

Moreover, the geometric constraint of multiple images can be used to optimize locally as well. Three images can form the geometric constraints based on conjugate points and lines that are called Trifocal tensor (Kitt and Lategahn, 2010). Four images also can form the geometric constraints called Quadrfocal tensor (Comport et al, 2007). Figure 5 shows the illustration of them that object points captured must be projected into a line on relative images. And these lines will be intersected into a line as well. Hence, the calculation would be more complicated and the texture in the surroundings has to be enough.

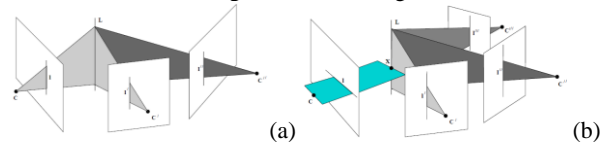


Figure 5. The geometric constraint of multiple (a) Trifocal tensor (b) Quadrfocal tensor (Hartley and Zisserman, 2003).

Consequently, the optimized theory applied in stereo VO algorithm would be proposed as well in this study. Through the analysis of the geometric relationship of the time adjacent stereo image pairs, a locally optimized approach is built for improving the position and orientation accuracy of stereo VO. The following chapters would explain the related details.

2. METHODOLOGY

The entire workflow is shown as Figure 6. The procedures include the system calibration of a dual-camera system, the solution of ROPs between consecutive stereo image pairs, the motion estimation including the position and orientation, and local optimization through computational processing of network adjustment.

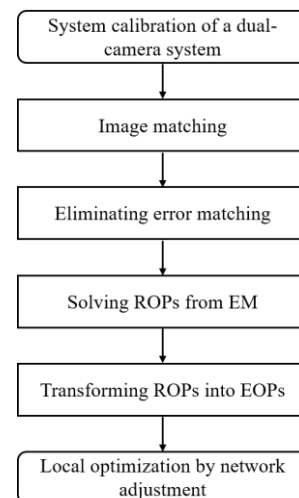


Figure 6. The workflow of proposed stereo VO.

The system calibration includes the calibration of IOPs and ROPs. Both must be calibrated in advance and then adopted in the stereo VO algorithm. IOPs mean interior orientation parameters of each camera. IOPs would not be the absolutely same in the dual-camera system. ROPs mean the relative orientation parameters between two lenses of a dual camera. ROPs play an important role in stereo VO for retrieving the true scale.

2.1 Automated Calculation of ROPs

The steps of image matching and eliminating error matching are contained in this part. The geometry between adjacent image pairs is built by ROPs. During the process of image matching, feature points are selected and filtered. Then EM is estimated and decomposed into the relative rotation and relative translation called ROPs. Furthermore, by defining the orientation of the first camera as the origin of the local coordinate system, the following ROPs of continuous image pairs are transformed into exterior orientation parameters (EOPs) in this local coordinate system. Figure 7 depicts the geometry of the image pair. There are three coordinates systems. The first one is object coordinate system (*O frame*), the second one is camera coordinate system of camera 1 (*C₁ frame*), and the other one is camera coordinate system of camera 2 (*C₂ frame*). Relative rotation, $R_{C_2}^{C_1}$ means the rotation matrix from *C₂ frame* to *C₁ frame*. There are 9 elements in this rotation matrix. Relative translation, $r_{C_2}^{C_1}$ means the unit vector from the origin of *C₁ frame* to the origin of *C₂ frame*. This unit vector is defined in *C₁ frame* and contains 3 elements. The image pair captures the same object point, P. Therefore, a coplanarity condition is formed. This condition is also called as the epipolar geometry. The algebraic representation of epipolar geometry can be expressed as a 3×3 matrix, which is named Essential Matrix (EM) (Longuet-Higgins, 1981). Every point correspondence, $r_p^{C_1}$ and $r_p^{C_2}$ should be satisfied with epipolar constraint described in the Equation 1. $r_p^{C_1}$ means the vector of the image point in *C₁ frame*, and $r_p^{C_2}$ means the vector of the image point in *C₂ frame*.

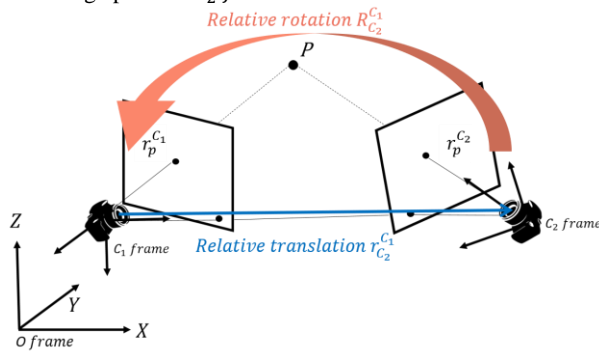


Figure 7. The geometry of the image pair.

$$(r_p^{C_2})^T E r_p^{C_1} = 0 \quad (1)$$

In this study, SURF algorithm (Bay et al., 2008) is used for image matching and Nister's five-point algorithm (Nistér, 2004) is adopted for estimating EM. It has set the inner constraint in EM, but the feature points are selected randomly without considering the distribution. The selected feature points do not distribute averagely on a whole image that would affect the reliability of ROPs. Therefore, the geometry constraint needs to be considered to eliminate the matching error as well. The solution in this study is to form a convex hull based on selected feature points. The area threshold of the convex hull is applied to eliminate error

matching. The distribution of matching feature points would become more average which would improve the estimation of ROPs.

2.2 Local Optimization

The steps of transforming ROPs into EOPs and local optimization by network adjustment are contained in this part. The dual-camera system captures two-time adjacent image pairs that contain four images. Currently, there are two major methods to solve the relative orientation between them. The first one is the algorithm with 3D-to-3D correspondences. The conjugate points (*i*) of four images are matched first. Then objective points corresponding to these conjugate points can be estimated by forward intersection. There are two sets of 3D coordinates defined in the different image pair. The 3D coordinates of object points (\tilde{X}_{k-1}^i) at the previous time (*k-1*) are transformed into the 3D coordinates of object points ($T_k \tilde{X}_{k-1}^i$) at the current time (*k*). The 3D coordinates estimated by current image pairs are \tilde{X}_k^i . Based on minimizing the difference between \tilde{X}_k^i and $T_k \tilde{X}_{k-1}^i$, ROPs can be solved in Equation 2. $T_{k,k-1}$ in Equation 3 means the transformation matrix between *k* and *k-1* that are formed by ROPs. Equation 3 shows the elements of T_k . $R_{k,k-1}$ means the relative rotation. $t_{k,k-1}$ means the relative translation.

$$\underset{T_k}{\operatorname{argmin}} \sum_i \|\tilde{X}_k^i - T_k \tilde{X}_{k-1}^i\| \quad (2)$$

$$T_k = \begin{bmatrix} R_{k,k-1} & t_{k,k-1} \\ 0 & 1 \end{bmatrix} \quad (3)$$

The second one is the algorithm with 3D-to-2D correspondences. The conjugate points (*i*) of four images are matched. First, 3D coordinates (\tilde{X}_{k-1}^i) of objective points corresponding to these conjugate points are estimated by forward intersection at the previous time (*k-1*). Then these objective points (\tilde{X}_{k-1}^i) are transformed and projected onto the image at the current time (*k*). Corresponding image points are \hat{p}_{k-1}^i formed according to \tilde{X}_{k-1}^i and T_k in Equation 4. The current image points are p_k^i . Based on minimizing the difference between p_k^i and \hat{p}_{k-1}^i , ROPs can be solved in Equation 5.

$$\hat{p}_{k-1}^i = F(T_k, \tilde{X}_{k-1}^i) \quad (4)$$

$$\underset{T_k}{\operatorname{argmin}} \sum_i \|p_k^i - \hat{p}_{k-1}^i\| \quad (5)$$

However, no matter adopting the algorithm with 3D-to-3D or 3D-to-2D correspondences, the 3D coordinates of object points must be computed first. Then ROPs are solved based on the minimization principle. The matching error would be propagated into object points and then ROPs. Therefore, local optimization by bundle adjustment is necessary. The EOPs of images and 3D coordinates of object points are improved during the least-squares process. More image points generate more observations, and the calculation becomes more complicated that cost much more time and computer resource. With 2D-to-2D correspondences, ROPs still can be solved from EM as monocular VO workflow. The 3D coordinates of object points do not have to be computed, but the true scale is unknown.

In this study, a novel local optimization is proposed based on 2D-to-2D correspondences for above issues. Two-time adjacent image pairs can generate six sets of ROPs totally. Their geometric relationship is depicted in Figure 8. Images captured by the left camera is *C₁* and *C₃* sequentially. Images captured by the right camera is *C₂* and *C₄* sequentially. There are totally six

combinations of ROPs. Relative rotation is $R_2^1, R_4^2, R_3^4, R_1^3, R_1^4$, and R_3^2 correspondingly. Relative translation is $r_2^1, r_4^2, r_3^4, r_1^3, r_1^4$, and r_3^2 correspondingly. Each true scale for relative translation is $\lambda_2^1, \lambda_4^2, \lambda_3^4, \lambda_1^3, \lambda_1^4$, and λ_3^2 correspondingly. Assuming C_1 frame is O frame, all rotation and translation in each camera frame are transformed into O frame. So that EOPs of four images could be obtained. r_{12}^0 means the vector defined in O frame from the origin of C_1 frame to the origin of C_2 frame, and so on. R_1^0 means the rotation from C_1 frame to O frame, and so on. The related equations are listed as the following.

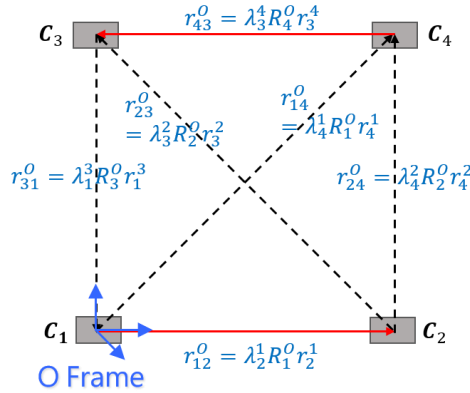


Figure 8. The geometry between consecutive images.

$$r_{12}^0 = \lambda_2^1 \cdot R_1^0 \cdot r_2^1 \quad (6)$$

$$r_{24}^0 = \lambda_4^2 \cdot R_2^0 \cdot r_4^2 \quad (7)$$

$$r_{43}^0 = \lambda_3^4 \cdot R_4^0 \cdot r_3^4 \quad (8)$$

$$r_{31}^0 = \lambda_1^3 \cdot R_3^0 \cdot r_1^3 \quad (9)$$

$$r_{14}^0 = \lambda_4^1 \cdot R_1^0 \cdot r_4^1 \quad (10)$$

$$r_{23}^0 = \lambda_3^2 \cdot R_2^0 \cdot r_3^2 \quad (11)$$

$$R_1^0 = I \quad (12)$$

$$R_2^0 = R_1^0 \cdot R_2^1 \quad (13)$$

$$R_3^0 = R_1^0 \cdot R_3^1 \quad (14)$$

$$R_4^0 = R_1^0 \cdot R_3^1 \cdot R_4^3 = R_1^0 \cdot R_2^1 \cdot R_4^2 \quad (15)$$

Applying the same algorithm of monocular VO, ROPs of each image pairs could be solved. Besides, the scale of the origin of C_1 frame to the origin of C_2 frame, and the origin of C_4 frame to the origin of C_3 frame is known based on the previous calibration. For the other 4 scales, the approximation could be estimated based on the principles of the triangle including inner product and sine rule.

The network adjustment of ROPs is based on least-squares. Observations are relative rotations and translations, not originally image points in bundle adjustment. The process is incremental that is designed into two parts. In the first part, 9 elements in each relative rotation and six inner constraints in each rotation matrix are listed as observation equations sequentially. Unknown parameters that are rotations belong to EOPs of C_3 frame and C_4 frame are calculated during the iteration. In the second part, 3 elements in each relative translation and one baseline of dual-camera system as the known true scale (λ_3^4) are listed as

observation equations sequentially. Unknown parameters that are translations and other true scales belong to EOPs of C_3 frame and C_4 frame are calculated during the iteration. The related observation equations are listed as the following. V means the matrix of residual, A means the design matrix, and W means the matrix of weight correspondingly.

$$\begin{bmatrix} R_4^2 \\ R_3^4 \\ R_1^3 \\ R_4^1 \\ R_3^2 \\ 6 \text{ constraints in } R_3^0 \\ 6 \text{ constraints in } R_4^0 \end{bmatrix} + V = A \times \begin{bmatrix} R_3^0 \\ R_4^0 \end{bmatrix} \sim W \quad (16)$$

$$\begin{bmatrix} \hat{r}_4^2 \\ \hat{r}_3^4 \\ \hat{r}_1^3 \\ \hat{r}_4^1 \\ \hat{r}_3^2 \\ \lambda_3^4 \end{bmatrix} + V = A \times \begin{bmatrix} r_3^0 \\ r_4^0 \\ \lambda_2^1 \\ \lambda_4^2 \\ \lambda_3^4 \\ \lambda_1^3 \\ \lambda_4^1 \\ \lambda_3^2 \end{bmatrix} \sim W \quad (17)$$

3. EXPERIMENTS AND RESULTS

3.1 Simulation Data

For verifying the performance of the proposed network adjustment model, simulation data is generated as shown in Figure 9. EOPs of consecutive images are known. Therefore, all ROPs of each image pairs are also known. The random bias is added in all ROPs. There are two cases are designed. Random bias in ROPs is set as 1 degree and 0.01meter in Case 1. Random bias in ROPs is set as 10 degrees and 0.1 meter in Case 2. Testing is implemented three times both in Case1 and Case 2.

For the results in Case 1, Table 1 indicates the error comparison, and Figure 10 shows the differences of ROPs before and after applying network adjustment. Six vectors mean six sets of ROPs from two-time adjacent image pairs. According to Table 1, the error of EOPs compared to true values is very small. Most rotation differences are less than 0.01 degree, and most translation differences are less than 0.003 meter. Besides, Figure 10 shows the network adjustment of ROPs is feasible and useful so that all ROPs could be optimized.

For the results in Case 2, Table 2 indicates the error comparison and Figure 11 shows the differences of ROPs before and after applying network adjustment. Six vectors mean six sets of ROPs from two-time adjacent image pairs. According to Table 2, the error of EOPs compared to true values become larger due to the larger random bias added. However, Figure 11 shows the network adjustment of ROPs is still feasible and useful so that all ROPs could be optimized significantly.

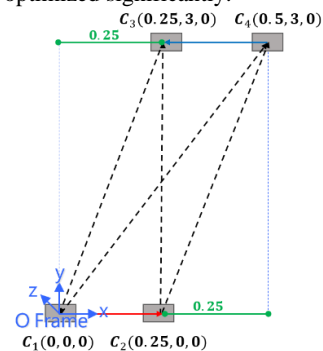


Figure 9. Simulation data of consecutive images.

	$\Delta\omega$	$\Delta\varphi$	$\Delta\kappa$	ΔX	ΔY	ΔZ
1 st	0.0085	-0.0186	0.0099	0.000	-0.003	0.000
	0.0075	-0.0190	0.0102	0.000	-0.003	0.001
2 nd	$\Delta\omega$	$\Delta\varphi$	$\Delta\kappa$	ΔX	ΔY	ΔZ
	-0.0054	0.0099	-0.0051	0.002	0	-0.005
3 rd	-0.0048	0.0102	-0.0052	0.002	0	-0.005
	$\Delta\omega$	$\Delta\varphi$	$\Delta\kappa$	ΔX	ΔY	ΔZ
	0.0065	-0.0138	0.0073	-0.003	0.002	-0.003
	0.0057	-0.0141	0.0075	-0.003	0.002	-0.002

Table 1. Error comparison in Case 1.

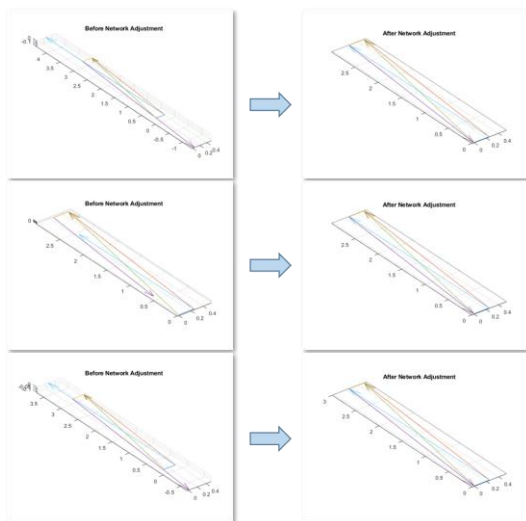


Figure 10. The differences of ROPs before and after applying network adjustment in Case 1.

	$\Delta\omega(^{\circ})$	$\Delta\varphi(^{\circ})$	$\Delta\kappa(^{\circ})$	$\Delta X(m)$	$\Delta Y(m)$	$\Delta Z(m)$
1 st	-0.0023	0.0044	-0.0023	0.077	-0.005	-0.063
	-0.0021	0.0045	-0.0023	0.077	-0.005	-0.063
2 nd	$\Delta\omega(^{\circ})$	$\Delta\varphi(^{\circ})$	$\Delta\kappa(^{\circ})$	$\Delta X(m)$	$\Delta Y(m)$	$\Delta Z(m)$
	0.0186	-0.0490	0.0274	-0.107	-0.104	0.064
3 rd	0.0159	-0.0500	0.0279	-0.107	-0.104	0.064
	$\Delta\omega(^{\circ})$	$\Delta\varphi(^{\circ})$	$\Delta\kappa(^{\circ})$	$\Delta X(m)$	$\Delta Y(m)$	$\Delta Z(m)$
	-0.0893	0.0995	-0.0396	-0.012	0.003	-0.004
	-0.0838	0.1043	-0.0424	-0.012	0.003	-0.004

Table 2. Error comparison in Case 2.

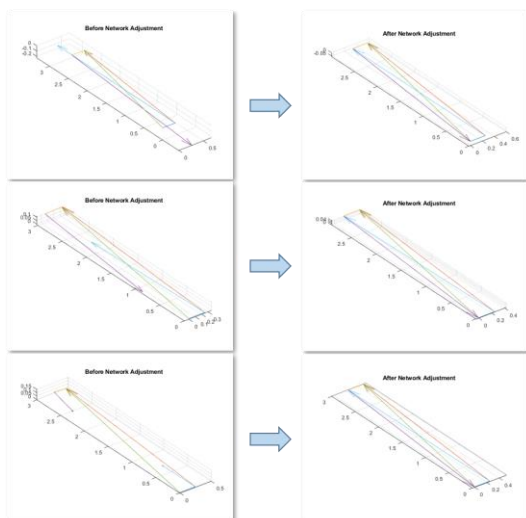


Figure 11. The differences of ROPs before and after applying network adjustment in Case 2.

3.2 Experiment Data

The experimental setting of the dual-camera system is shown as Figure 12. Images are taken in front of the department building at a different time. Table 3 indicates the solved EOPs and true scales in the experiment. Figure 13 shows the difference of ROPs before and after applying network adjustment in the experiment. Six vectors mean six sets of ROPs from two-time adjacent image pairs. Figure 14 shows the position and orientation of adjacent image pairs in stereo VO. The results represent the proposed stereo VO algorithm is feasible. The unknown orientation of images can be recovered in the local frame. Especially, network adjustment of ROPs could optimize the solved EOPs without the observations of image points.

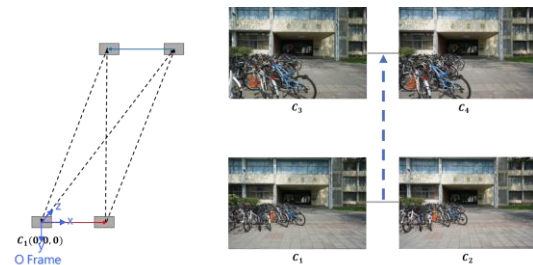


Figure 12. The experimental setting of the dual-camera system.

EOPs	$\omega(^{\circ})$	$\varphi(^{\circ})$	$\kappa(^{\circ})$	$X(m)$	$Y(m)$	$Z(m)$
C_3	0.6952	-1.2449	-0.6450	-0.052	-0.004	2.015
C_4	0.6032	0.3745	-0.0733	0.22	-0.009	1.993
Scale	λ_2^1	λ_4^2	λ_3^4	λ_1^3	λ_4^1	λ_3^2
Value	0.273	1.997	0.273	2.016	2.005	2.044

Table 3. Solved EOPs in the experiment.

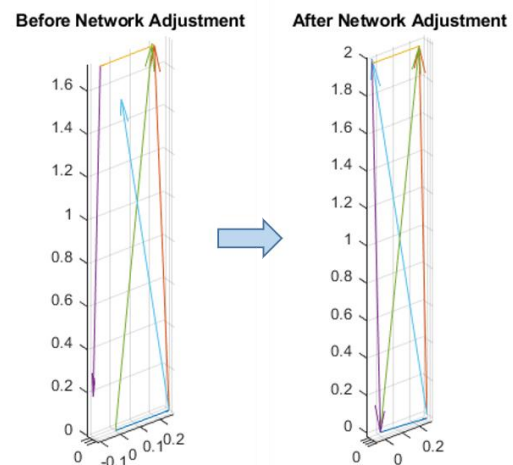


Figure 13. The difference of ROPs before and after applying network adjustment in the experiment.

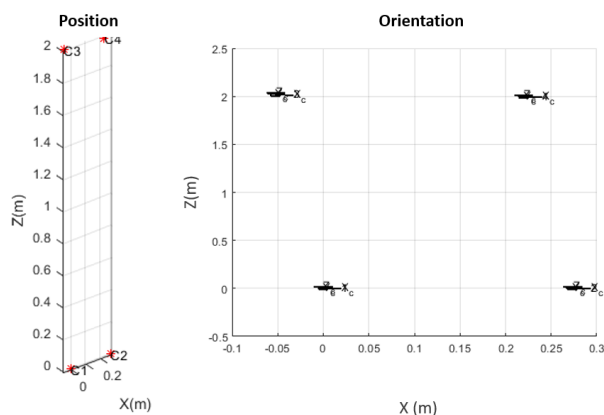


Figure 14. The position and orientation of adjacent image pairs in stereo VO.

4. CONCLUSIONS AND FUTURE WORKS

The proposed stereo VO is feasible and has been implemented. The true scale of translation can be recovered. Network adjustment of ROPs is validated by both simulation and experiment data. No matter the random bias in relative rotation is 1 or 10 degrees or the random bias in relative translation is 0.01 or 0.1 meter, ROPs could be optimized significantly. Therefore, the position and orientation of images can be estimated better. The experiment data also represent the same effect. However, the combination of the baseline and intersection geometry needs to be analyzed and tested in more experiments. The trajectory in the experiment also needs to be larger to estimate the accumulated error. Besides, the reference solution can be set to compare the precision of the solved position and orientation of images frame by frame.

ACKNOWLEDGEMENTS

The authors would acknowledge the supports provided by the Ministry of the Interior.

REFERENCES

- Bay, H., Ess, A., Tuytelaars, T., & Van Gool, L., 2008. Speeded-up robust features (SURF). *Computer vision and image understanding*, 110(3), 346-359.
- Bertozzi, M., Broggi, A., Cardarelli, E., Fedriga, R. I., Mazzei, L., & Porta, P. P., 2011. Viac expedition toward autonomous mobility [from the field]. *IEEE Robotics & Automation Magazine*, 18(3), 120-124.
- Bonin-Font, F., Ortiz, A., & Oliver, G., 2008. Visual navigation for mobile robots: A survey. *Journal of intelligent and robotic systems*, 53(3), 263.
- Cheng, Y., Maimone, M., & Matthies, L., 2005. Visual odometry on the Mars exploration rovers. *IEEE International Conference on Systems, Man and Cybernetics*, Vol. 1, pp. 903-910.
- Comport, A.I., Malis, E., & Rives, P., 2007. Accurate quadrifocal tracking for robust 3D visual odometry. *IEEE International Conference on Robotics and Automation*, pp. 40-45.
- Dusha, D., & Mejias, L., 2012. Error analysis and attitude observability of a monocular GPS/visual odometry integrated navigation filter. *The International Journal of Robotics Research*, 31(6), 714-737.
- Ericson, E., & Astrand, B., 2008. Visual odometry system for agricultural field robots. In *Anonymous Proceedings of the World Congress on engineering and computer science*.
- Foley, B.P., Dellaporta, K., Sakellariou, D., Bingham, B.S., Camilli, R., Eustice, R.M., ... & Mallios, A., 2009. The 2005 Chios ancient shipwreck survey: new methods for underwater archaeology. *Hesperia*, 269-305.
- Grisetti, G., Kummerle, R., Stachniss, C., & Burgard, W., 2010. A tutorial on graph-based SLAM. *IEEE Intelligent Transportation Systems Magazine*, 2(4), 31-43.
- Hartley, R., & Zisserman, A., 2003. *Multiple view geometry in computer vision*. Cambridge university press.
- Jones, E. S., & Soatto, S., 2011. Visual-inertial navigation, mapping and localization: A scalable real-time causal approach. *The International Journal of Robotics Research*, 30(4), 407-430.
- Kelly, J., & Sukhatme, G. S., 2007. An experimental study of aerial stereo visual odometry. *IFAC Proceedings Volumes*, 40(15), 197-202.
- Kitt, B., Geiger, A., & Lategahn, H., 2010. Visual odometry based on stereo image sequences with ransac-based outlier rejection scheme. In *2010 IEEE intelligent vehicles symposium*, pp. 486-492.
- Kneip, L., Chli, M., & Siegwart, R.Y., 2011. Robust real-time visual odometry with a single camera and an IMU. In *Proceedings of the British Machine Vision Conference 2011*. British Machine Vision Association.
- Longuet-Higgins, H. C., 1981. A computer algorithm for reconstructing a scene from two projections. *Nature*, 293(5828), 133-135.
- Nistér, D., 2004. An efficient solution to the five-point relative pose problem. *IEEE transactions on pattern analysis and machine intelligence*, 26(6), 756-770.
- Nistér, D., Naroditsky, O., & Bergen, J., 2004. Visual odometry. In *Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2004. CVPR 2004. (Vol. 1, pp. I-I). Ieee.
- Scaramuzza, D., & Fraundorfer, F., 2011. Visual odometry (tutorial). *IEEE robotics & automation magazine*, 18(4), 80-92.
- Triggs, B., McLauchlan, P.F., Hartley, R.I., & Fitzgibbon, A.W., 1999. Bundle adjustment—a modern synthesis. In *International workshop on vision algorithms*, pp. 298-372. Springer, Berlin, Heidelberg.
- Yoon, S. J., & Kim, T., 2019. Development of Stereo Visual Odometry Based on Photogrammetric Feature Optimization. *Remote Sensing*, 11(1), 67.

Revised May 2020