STEREO VISUAL NAVIGATION BASED ON NETWORK ADJUSTMENT OF RELATIVE ORIENTATION PARAMETERS

Kuan-Ying Lin (1), Yi-Hsing Tseng (1), Kai-Wei Chiang (1)

[1] National Cheng-Kung University, No.1, Daxue Road, East District, Tainan, Taiwan
Email: azsxd2014@gmail.com; Tseng@mail.ncku.edu.tw; kwchiang@mail.ncku.edu.tw

KEY WORDS: Stereo image, Visual navigation, Relative orientation, Least-squares, Network adjustment

ABSTRACT: Visual navigation (VN) is an algorithm of estimating the motion of consecutive frames. The motion can be described by the exterior orientation parameters (EOPs) including the attitude and position. However, using a single camera for VN could not recover the real scale. The distance between adjacent images is unknown. Besides, the error would exist in the EOPs due to the error matching. To overcome the above issues, stereo VN is adopted in this research. Stereo VN can recover the real scale according to the pre-calibrated relative orientation of two cameras. However, the error still exists in the solved orientation including the rotation, translation, and scale. Therefore, the network adjustment model based on relative orientation parameters (ROPs) is proposed to optimize the motion of consecutive frames. The proposed network adjustment model is verified by not only simulation data but also the real experiment. Consequently, ROPs are adopted as observations in the model that would update the states of image sequence furthermore. Not only the scale issue could be solved, but also the rotation and translation solution could be improved significantly during the whole operation. In this research, it is worth mentioning that the coordinates of object points are not necessary to be calculated, which means the conventional bundle adjustment is not adopted, but more accurate ROPs still have improved automatically during the process.

1. INTRODUCTION

Visual navigation (VN), or called visual odometry (VO), is a technology applied to track the position and orientation of a moving platform with a camera taking image sequences. In the last two decades, VN has been a hot topic in the field of computer vision and widely applied for robotic navigation. Applying VN for modern mobile mapping technology is also the trend to achieve continuous indoor and outdoor navigation in the field of mobile mapping. Figure 1 shows the workflow of monocular VN (Scaramuzza and Fraundorfer., 2011).

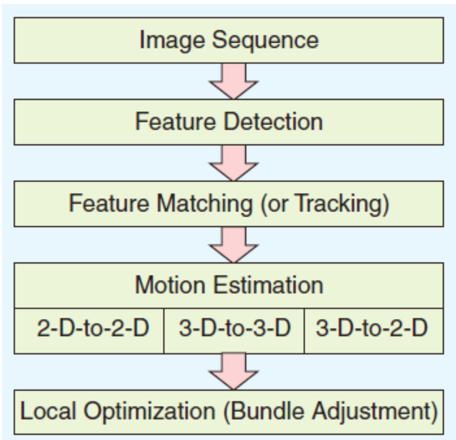

Figure 1. The workflow of monocular VN (Scaramuzza and Fraundorfer., 2011).

However, monocular VN has the problem that the real scale cannot be solved. If only a single camera is used, other measurements about distance are necessary for scale recovering. Recently, it is common that additional sensors like receivers of the global navigation satellite system (GNSS) (Dusha and Mejias, 2012) or inertial navigation system (INS) are adopted and combined with Visual navigation (Qin et al., 2018). They can provide the related scale information for VN. But another problem would exist in the multi-sensor system as well. The observations from GNSS or INS have their own error that would affect the solution of VN. Besides, the error would be accumulated during moving.

To overcome the above issues, stereo VN is adopted in this research. A horizontal bar platform with two fixed cameras instead of a single camera is designed and used. This platform can take stereo images simultaneously and continuously. Based on the pre-calibrated relative orientation of two cameras, stereo VN can recover the real scale without observations of additional sensors. However, the scale would not be very accurate, and the error also exists in the orientation including rotation and translation. Moreover, the network adjustment model based on relative orientation parameters (ROPs) is proposed to optimize the orientations between adjacent images. The related stereo camera theory and mechanism have been developed. Solving ROPs of an image pair is the critical work. Hence, the method of calculating ROPs automatically is also applied. Figure 2 shows the entire workflow. The entire process includes the system calibration of the dual-camera platform, the automatic solution of ROPs between consecutive image pair, the estimation of the moving platform including the trajectory and orientation, and the improvement through computational processing of the least-squares adjustment. The following chapters would explain the stereo VN model and the network adjustment model of ROPs, show the experiment results, and describe the conclusions and future work in the end.

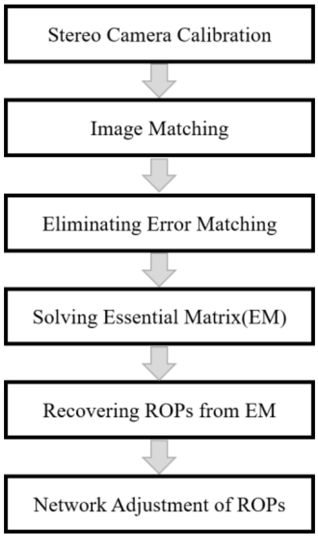

Figure 2. The workflow of proposed stereo VN.

2. METHODOLOGY

2.1 Relative Orientation Parameters (ROPs)

The ROPs is described by the relative rotation and translation of the image pair. Figure 3 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 ($C_1\ frame$), and the other one is s camera coordinate system of camera 2 ($C_2\ frame$). Relative rotation, $R_{C_2}^{C_1}$ means the rotation matrix from $C_2\ frame$ to $C_1\ frame$. There are 9 elements in this rotation matrix. Relative translation, $r_{C_2}^{C_1}$ means the unit vector from the origin of $C_1\ frame$ to the origin of $C_2\ frame$. This unit vector is defined in $C_1\ frame$ and contains 3 elements. The image pair captures the same object point, P. Therefore, a coplanarity condition is formed as figure 3 depicts. 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). EM is composed of relative rotation and translation. 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 $C_1\ frame$, and $r_p^{C_2}$ means the vector of the image point in $C_2\ frame$.

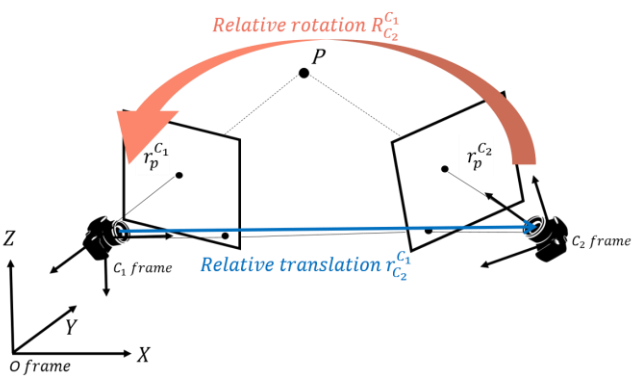

Figure 3. The geometry of the image pair.

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

2.2 The Stereo Camera Model

Figure 4 shows the geometry between consecutive images. Images captured by the left camera is C_1 and C_3 sequentially. Images captured by the right camera is C_2 and C_4 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 scale for relative translation is λ_2^1, λ_4^2, λ_3^4, λ_1^3, λ_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. Therefore, r_{12}^{O} 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^O means the rotation from $C_1\ frame$ to $O\ frame$, and so on. The related equations are listed as the following.

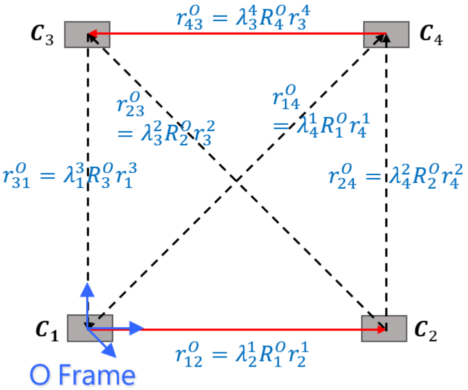

Figure 4. The geometry between consecutive images.

$$r_{12}^{O} = \lambda_2^1 \cdot R_1^O \cdot r_2^1 \tag{2}$$

$$r_{24}^{O} = \lambda_4^2 \cdot R_2^O \cdot r_4^2 \tag{3}$$

$$r_{43}^{O} = \lambda_3^4 \cdot R_4^O \cdot r_3^4 \tag{4}$$

$$r_{31}^{O} = \lambda_1^3 \cdot R_3^O \cdot r_1^3 \tag{5}$$

$$r_{14}^{O} = \lambda_4^1 \cdot R_1^O \cdot r_4^1 \tag{6}$$

$$r_{23}^{O} = \lambda_3^2 \cdot R_2^O \cdot r_3^2 \tag{7}$$

$$R_1^{O} = I \tag{8}$$

$$R_2^{O} = R_1^O \cdot R_2^1 \tag{9}$$

$$R_3^{O} = R_1^O \cdot R_3^1 \tag{10}$$

$$R_4^{O} = R_1^O \cdot R_3^1 \cdot R_4^3 = R_1^O \cdot R_2^1 \cdot R_4^2 \tag{11}$$

Applying the same algorithm of monocular VN, 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.

2.3 Network Adjustment of ROPs

The network adjustment of ROPs is based on least-squares. Observation is relative rotation and translation, not originally image points in bundle adjustment. The network adjustment model proposed is incremental. At the first part, 9 elements in each relative rotation are listed as observation equations sequentially. And then unknown parameters that are rotations belong to EOPs of $C_3\ frame$ and $C_4\ frame$ are calculated during the iteration. At the second part, 3 elements in each relative rotation are listed as observation equations sequentially. And then unknown parameters that are translations and 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 design matrix, and W means the matrix of weight correspondingly.

$$\begin{bmatrix} R_4^2 \\ R_3^4 \\ R_1^3 \\ R_4^1 \\ R_3^2 \\ 6\ contraits\ in\ R_3^O \\ 6\ contraits\ in\ R_4^O \end{bmatrix} + V = A \times \begin{bmatrix} R_3^O \\ R_4^O \end{bmatrix} \sim W \tag{12}$$

$$\begin{bmatrix} \hat{r}_4^2 \\ \hat{r}_3^4 \\ \hat{r}_1^3 \\ \hat{r}_4^1 \\ \hat{r}_3^2 \\ \lambda_3^4 \end{bmatrix} + V = A \times \begin{bmatrix} r_3^O \\ r_4^O \\ \lambda_4^2 \\ \lambda_3^4 \\ \lambda_1^3 \\ \lambda_4^1 \\ \lambda_3^2 \end{bmatrix} \sim W \tag{13}$$

3. EXPERIMENTS AND RESULTS

3.1 Simulation Data

For verifying the performance of the proposed network adjustment model, simulation data is generated as Figure 5 shows. 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. Three testings are implemented both in the Case1 and Case 2. Table 1 shows the solved EOPs and error comparison in Case 1. And Figure 6 shows the differences before and after applying network adjustment of ROPs in Case 1. 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 6 shows the network adjustment of ROPs is feasible and useful so that all ROPs could be optimized. Table 2 shows the solved EOPs and error comparison in Case 2. And Figure 7 shows the differences before and after applying network adjustment of ROPs in Case 2. According to Table 2, the error of EOPs compared to true values become larger due to the larger random bias added. However, Figure 7 shows the network adjustment of ROPs is still feasible and useful so that all ROPs could be optimized significantly.

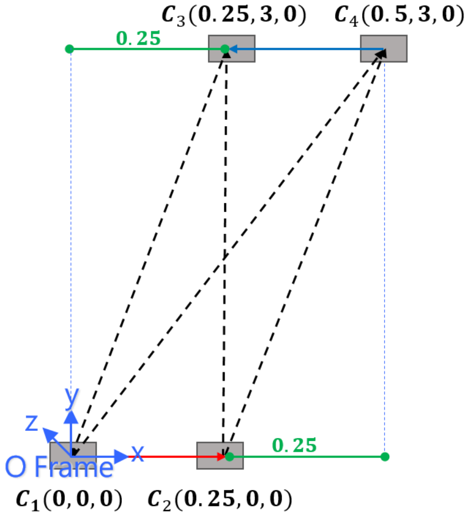

Figure 5. Simulation data of consecutive images.

Table 1. Solved EOPs and error comparison in Case 1.

E.O.	ω	φ	κ	X	Y	Z	$\Delta\omega$	$\Delta\varphi$	$\Delta\kappa$	ΔX	ΔY	ΔZ
Image 3	0.008497	-0.018553	0.009937	0.250	2.997	0.000	0.008497	-0.018553	0.009937	0.000	-0.003	0.000
Image 4	1.007476	1.980988	3.010157	0.500	2.997	0.001	0.007476	-0.019012	0.010157	0.000	-0.003	0.001
Scale	λ_2^1	λ_4^2	λ_3^4	λ_1^3	λ_4^1	λ_3^2	$\Delta\lambda_2^1$	$\Delta\lambda_4^2$	$\Delta\lambda_3^4$	$\Delta\lambda_1^3$	$\Delta\lambda_4^1$	$\Delta\lambda_3^2$
Value	0.250	3.008	0.250	3.007	3.039	2.997	0.000	-0.003	0.000	-0.003	-0.003	-0.003

E.O.	ω	φ	κ	X	Y	Z	$\Delta\omega$	$\Delta\varphi$	$\Delta\kappa$	ΔX	ΔY	ΔZ
Image 3	-0.005387	0.009949	-0.005071	0.252	3.000	-0.005	-0.005387	0.009949	-0.005071	0.002	0.000	-0.005
Image 4	0.995161	2.010240	2.994783	0.502	3.000	-0.005	-0.004839	0.010240	-0.005217	0.002	0.000	-0.005
Scale	λ_2^1	λ_4^2	λ_3^4	λ_1^3	λ_4^1	λ_3^2	$\Delta\lambda_2^1$	$\Delta\lambda_4^2$	$\Delta\lambda_3^4$	$\Delta\lambda_1^3$	$\Delta\lambda_4^1$	$\Delta\lambda_3^2$
Value	0.250	3.010	0.250	3.010	3.041	3.000	0.000	0.000	0.000	0.000	0.000	0.000

E.O.	ω	φ	κ	X	Y	Z	$\Delta\omega$	$\Delta\varphi$	$\Delta\kappa$	ΔX	ΔY	ΔZ
Image 3	0.006471	-0.013764	0.007305	0.247	3.002	-0.003	0.006471	-0.013764	0.007305	-0.003	0.002	-0.003
Image 4	1.005708	1.985886	3.007473	0.497	3.002	-0.002	0.005708	-0.014114	0.007473	-0.003	0.002	-0.002
Scale	λ_2^1	λ_4^2	λ_3^4	λ_1^3	λ_4^1	λ_3^2	$\Delta\lambda_2^1$	$\Delta\lambda_4^2$	$\Delta\lambda_3^4$	$\Delta\lambda_1^3$	$\Delta\lambda_4^1$	$\Delta\lambda_3^2$
Value	0.250	3.012	0.250	3.012	3.043	3.002	0.000	0.002	0.000	0.002	0.002	0.002

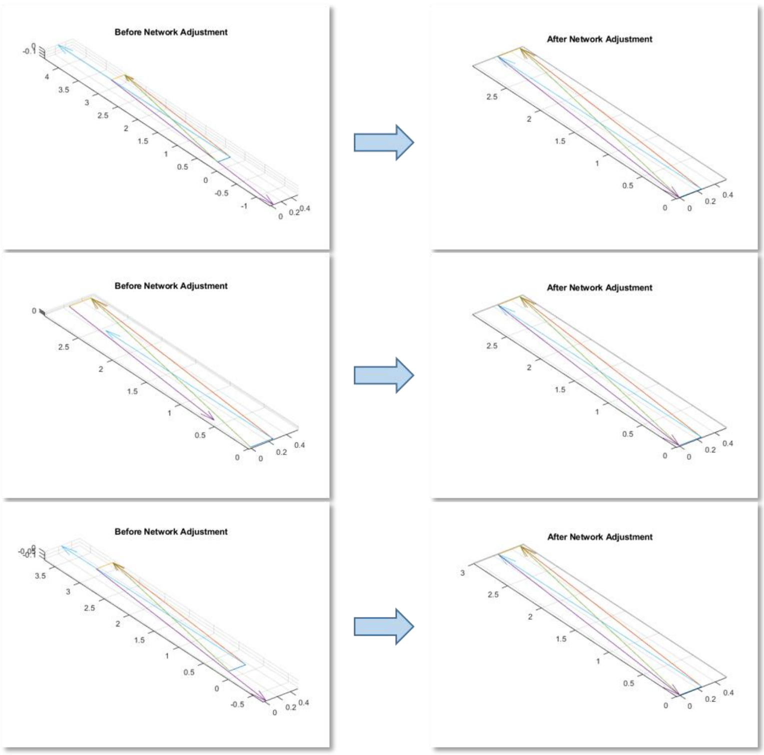

Figure 6. The differences before and after applying network adjustment of ROPs in Case 1.

Table 2. Solved EOPs and error comparison in Case 2.

E.O.	ω	φ	κ	X	Y	Z	$\Delta\omega$	$\Delta\varphi$	$\Delta\kappa$	ΔX	ΔY	ΔZ
Image 3	-0.002316	0.004412	-0.002273	0.327	2.995	-0.063	-0.002316	0.004412	-0.002273	0.077	-0.005	-0.063
Image 4	0.997926	2.004537	2.997664	0.577	2.995	-0.063	-0.002074	0.004537	-0.002336	0.077	-0.005	-0.063
Scale	λ_2^1	λ_4^2	λ_3^4	λ_1^3	λ_4^1	λ_3^2	$\Delta\lambda_2^1$	$\Delta\lambda_4^2$	$\Delta\lambda_3^4$	$\Delta\lambda_1^3$	$\Delta\lambda_4^1$	$\Delta\lambda_3^2$
Value	0.250	3.015	0.250	3.029	3.053	2.999	0.000	0.005	0.000	0.018	0.011	-0.001

E.O.	ω	φ	κ	X	Y	Z	$\Delta\omega$	$\Delta\varphi$	$\Delta\kappa$	ΔX	ΔY	ΔZ
Image 3	0.018553	-0.048965	0.027428	0.143	2.896	0.064	0.018553	-0.048965	0.027428	-0.107	-0.104	0.064
Image 4	1.015862	1.950032	3.027874	0.393	2.896	0.064	0.015862	-0.049968	0.027874	-0.107	-0.104	0.064
Scale	λ_2^1	λ_4^2	λ_3^4	λ_1^3	λ_4^1	λ_3^2	$\Delta\lambda_2^1$	$\Delta\lambda_4^2$	$\Delta\lambda_3^4$	$\Delta\lambda_1^3$	$\Delta\lambda_4^1$	$\Delta\lambda_3^2$
Value	0.250	2.902	0.250	2.918	2.925	2.901	0.000	-0.108	0.000	-0.093	-0.116	-0.099

E.O.	ω	φ	κ	X	Y	Z	$\Delta\omega$	$\Delta\varphi$	$\Delta\kappa$	ΔX	ΔY	ΔZ
Image 3	-0.089291	0.099454	-0.039645	0.238	3.003	-0.004	-0.089291	0.099454	-0.039645	-0.012	0.003	-0.004
Image 4	0.916206	2.104263	2.957648	0.488	3.003	-0.004	-0.083794	0.104263	-0.042352	-0.012	0.003	-0.004
Scale	λ_2^1	λ_4^2	λ_3^4	λ_1^3	λ_4^1	λ_3^2	$\Delta\lambda_2^1$	$\Delta\lambda_4^2$	$\Delta\lambda_3^4$	$\Delta\lambda_1^3$	$\Delta\lambda_4^1$	$\Delta\lambda_3^2$
Value	0.250	3.012	0.250	3.013	3.042	3.003	0.000	0.002	0.000	0.002	0.001	0.003

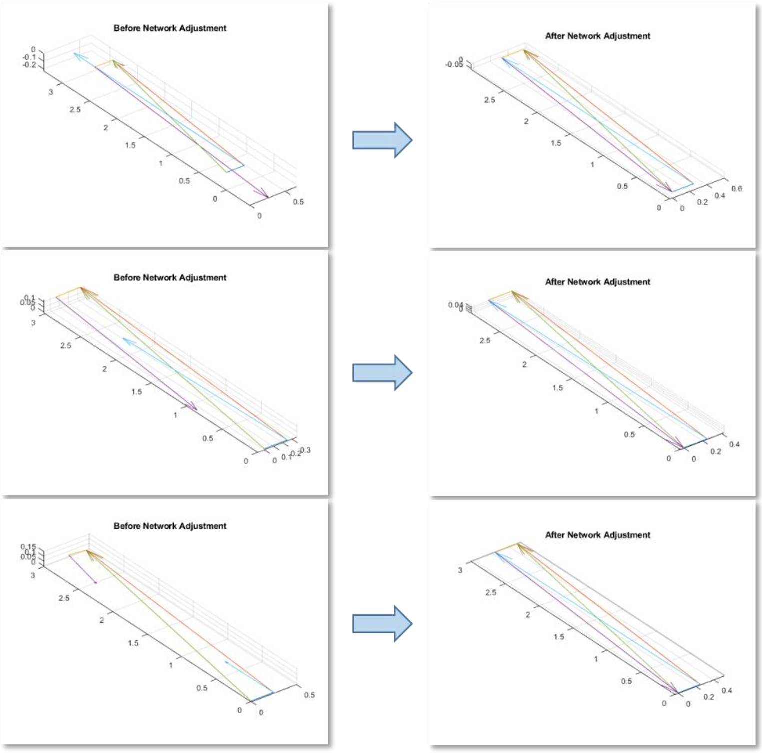

Figure 7. The differences before and after applying network adjustment of ROPs in Case 2.

3.2 Experiment Data

Figure 8 shows the experimental setting of the stereo camera. The stereo camera is used to take photos in front of the department building at a different time. Table 3 shows the solved EOPs in the experiment. And Figure 9 shows the difference before and after applying network adjustment of ROPs in the experiment. The results represent the proposed stereo VN algorithm is feasible. The unknown orientation of images can be recovered. Especially, network adjustment of ROPs could optimize the solved EOPs without the observations of image points.

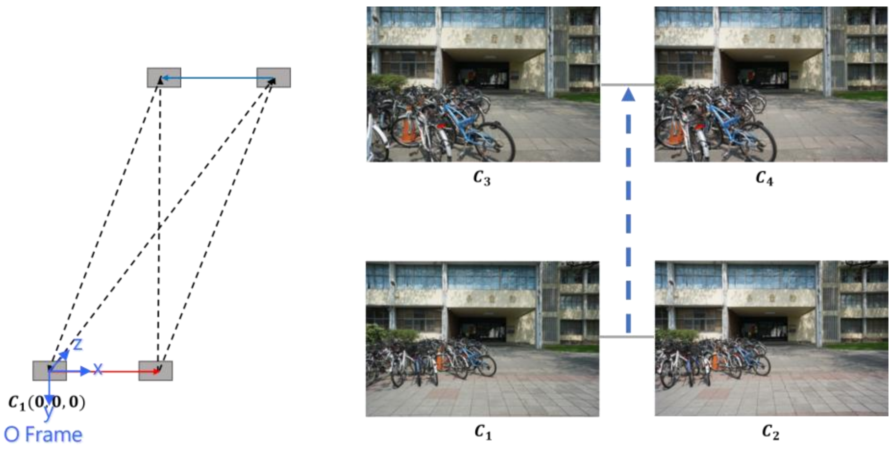

Figure 8. The experimental setting of the stereo camera.

Table 3. Solved EOPs in the experiment.

E.O.	ω	φ	κ	X	Y	Z
Image 3	0.695187	-1.244894	-0.645041	-0.052	-0.004	2.015
Image 4	0.603215	0.374489	-0.073307	0.220	-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

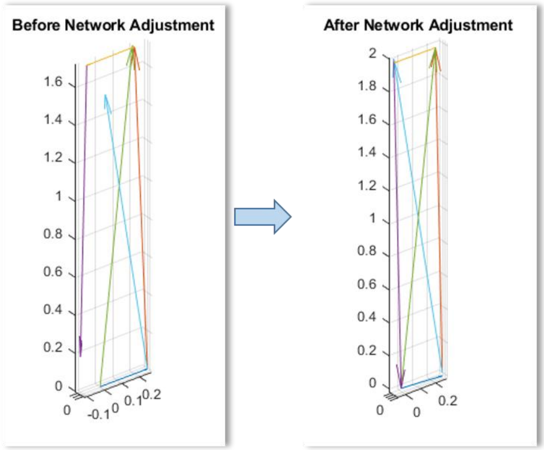

Figure 9. The difference before and after applying network adjustment of ROPs in the experiment.

4. CONCLUSIONS AND FUTURE WORK

The proposed stereo VN is feasible and has been implemented. The 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, the EOPs could be optimized significantly. The error in ROPs can be corrected. Moreover, the orientation of images can be solved and improved. The experiment data also represent the same effect. However, more experiments including the baseline and intersection geometry need to be implemented and analyzed. 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 solved orientation of images.

REFERENCES

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), pp.714-737.

Longuet-Higgins, H. C. ,1981. A computer algorithm for reconstructing a scene from two projections. Nature, 293(5828), pp.133.

Qin, T., Li, P., & Shen, S., 2018. Vins-mono: A robust and versatile monocular visual-inertial state estimator. IEEE Transactions on Robotics, 34(4), pp.1004-1020.

Scaramuzza, D., & Fraundorfer, F. ,2011. Visual odometry [tutorial]. IEEE robotics & automation magazine, 18(4), pp. 80-92.