

NETWORK ADJUSTMENT OF AUTOMATIC RELATIVE ORIENTATION FROM IMAGE SEQUENCES

C. C. Lin¹, Y. H. Tseng¹, K. Y. Lin¹, K. W. Chiang¹

¹ Dept. of Geomatics, National Cheng-Kung University, No. 1, Daxue Road, East District, Tainan, Taiwan - azsxd2014@gmail.com

Commission I, WG I/7

KEY WORDS: Visual Odometer, Monocular Camera, Photogrammetry, Network Adjustment, Relative Orientation

ABSTRACT:

Recently, Visual Odometry (VO) using cameras for navigation is known as an alternative solution in GNSS-hostile environments. VO is a process of estimating the egomotion based on consecutive frames captured by the camera. 3D Motion including the attitude and position can be described as the exterior orientation parameters (EOPs) in photogrammetry. The advantage of VO compared with wheel odometry is that VO is not affected by wheel slip in uneven terrain or other adverse conditions. Since VO computes the camera path incrementally, the errors are accumulated as well according to the motion of each new frame-to-frame over time. That would cause the drift in the estimated trajectory compared to the real path. To solve this issue, this research proposes the network adjustment model based on relative orientation parameters (ROPs) for monocular VO. The fundamental idea originates from the traverse in the field of surveying. A traverse is a series of consecutive lines whose ends have been marked in the field and whose lengths and angles have been determined from observations. Consequently, ROPs are adopted as observations in the model that would update the states of image sequence furthermore. In this research, it is worth mentioning that the coordinates of object points are not necessary to be calculated, and more accurate ROPs are improved automatically during the process. In the future, VO with this proposed method could be integrated with GNSS/INS to a navigation system.

1. INTRODUCTION

VO is very popular and discussed actively because rich visual information can be captured in images by only using affordable and simple camera. VO is defined as determining location and orientation of the robotic by analyzing the consecutive images derived from the cameras (Nistér et al., 2004). VO can be divided into monocular (Davison, 2003) and binocular method (Fraundorfer and Scaramuzza, 2012). Monocular VO is using single camera to capture consecutive images over time, and adjacent images can form a stereo image pair to estimate camera path. On the other hands, binocular VO is known as stereo VO. In stereo VO, two cameras are attached on a platform (baseline is fixed) to capture consecutive image pairs simultaneously to estimate the path. In general, binocular is better than monocular because the binocular can provide the true scale of the translation of the cameras, but monocular cannot. However, in cases where the distance between object and binocular is too far, the stereo case degenerates to the monocular case (Lemaire et al., 2007), especially for small robotic (baseline of binocular is too short). Moreover, the requirement of UAV in robotic is small and light currently. Consequently, monocular VO is also attractive in VO technique.

Actually, estimating camera path and mapping can also be achieved in photogrammetry. However, there are some difference between the real-time navigation in computer vision and offline mapping in photogrammetry. Simply to say, the former one is automatic method, and the latter one is semi-automatic method. It is mainly caused by the methodology in two fields. In computer vision, the feature-based methods have been widely used to search the correspondences of an image pair automatically. For arbitrary two images, if they are overlapped in visualization, the feature-based methods may generate an acceptable matching result. On the other hands, area-based

matching in photogrammetry merge the feature detection step with the matching part. Instead of detecting salient features, window of predefined size is used for the estimation of correspondences. There are certain limitations in area-based matching (Joglekar and Gedam, 2012). Firstly, if images are deformed by complex transformation, rectangular window cannot able to cover same part of scene, especially for two images with large intersection angle in close-range photogrammetry. Secondly, reasonable search window is usually predefined necessarily with manual input to obtain a reliable matching result. Therefore, it is hard to achieve automatic processing for area-based matching.

The coplanarity condition and collinearity equation are usually applied to solve the relative orientation in photogrammetry. Both of them are nonlinear equation, so the reasonable initial approximation of unknown parameters is necessary. However, reasonable approximation in close-range photogrammetry will be a challenge. That is the major reason that photogrammetry is always offline work, and real time navigation is hardly achieved. However, the knowledge in the geometry of images is perfect in photogrammetry (Förstner, 2002). For example, the concept of local bundle adjustment in computer vision is developed from photogrammetry. Consequently, it would be a practical way to merge methodology in these two fields. In this study, reconstructing the relative orientation of image sequences automatically based on computer vision and solving the geometry of multiple cameras based on photogrammetry are combined preliminarily.

2. METHODOLOGY

The overall workflow in the study is summarized in figure 1. Before capturing images, the interior parameters of the camera need to be calibrated in advance that can rectify the lens

distortion. The checkerboard (Zhang, 1999) is used in the camera calibration. And then image sequences can be obtained and rectified based on the previous calibration results.

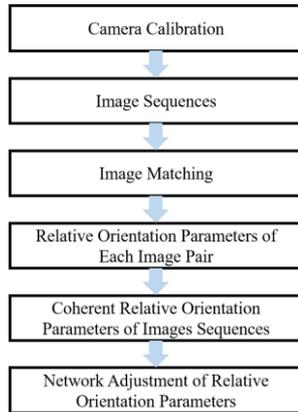


Figure 1. The workflow in this study

Following steps can be separated into two major parts. The first one containing image matching and relative orientation parameters of each image pair belongs to automatic relative orientation in computer vision. The second one containing coherent relative orientation parameters of image sequences and network adjustment of relative orientation parameters belongs to the geometry of multiple cameras in photogrammetry. The following sections would describe the more details.

2.1 Image Matching

2.1.1 Feature-based method

An image feature is an interesting or significant part of the image, such as edges, corner, blobs, ridges, and so on. Image features are often extracted as the starting points for many algorithms applied in computer vision. In this study, feature-based methods are used to estimate the relative orientation parameters (ROPs) of an image pair. SURF (Speeded up robust features) (Bay et al., 2006) and SIFT (Scale-invariant feature translation) (Lowe, 2004) are two most popular feature-based methods. In general, SURF is several times faster, and more robust against different image transformation than SIFT, but generates less correspondences than SIFT (Juan and Gwun, 2009). Considering the processing efficiency, SURF is adopted in this study.

2.1.2 Random Sample Consensus (RANSAC)

The matching result would exist some mismatches. These mismatches are defined as outliers here, and correct point correspondences are defined as inliers. If inliers are not sufficient or outliers are more than the inliers in the matching result, ROPs cannot be estimated accurately. In order to reduce outliers, the random sample consensus (RANSAC) (Fischler and Bolles, 1981) strategy is applied to improve the matching results.

2.2 Motion estimation

2.2.1 Relative Orientation

The relative orientation of an image pair can be expressed as the exterior orientation parameters (EOPs). Consequently, the relative orientation of the image pair can be described as the parameters of relative translations $(\Delta X, \Delta Y, \Delta Z)$ and relative rotations $(\Delta\omega, \Delta\phi, \Delta\kappa)$. Therefore, they are named as relative orientation parameters (ROPs) in this study. If the amount of correspondences in each image pair is sufficient, ROPs can be estimated. Figure 2 illustrates the relationship of an image pair and an object point. Eq. (1) and Eq. (2)

express the rotation matrix of first and second camera respectively, and Eq. (3) and Eq. (4) express the vector of the first and second camera in object frame (O frame). Consequently, the definition of relative orientation and translation of an image pair can be explained as the Eq. (5) and Eq. (6). In this study, both definitions of the relative translation and rotation of an image pair are transforming the first image into the second image.

$$R_O^{C_1} = R_{\kappa_1} R_{\phi_1} R_{\omega_1} \quad (1)$$

$$R_O^{C_2} = R_{\kappa_2} R_{\phi_2} R_{\omega_2} \quad (2)$$

$$r_{C_1}^O = \begin{bmatrix} X_{C_1} \\ Y_{C_1} \\ Z_{C_1} \end{bmatrix} \quad (3)$$

$$r_{C_2}^O = \begin{bmatrix} X_{C_2} \\ Y_{C_2} \\ Z_{C_2} \end{bmatrix} \quad (4)$$

$$R_{C_1}^{C_2} = R_O^{C_2} (R_O^{C_1})^T = R_{\Delta\omega} R_{\Delta\phi} R_{\Delta\kappa} \quad (5)$$

$$r_{C_1 \rightarrow C_2}^O = r_{C_2}^O - r_{C_1}^O = \begin{bmatrix} \Delta X \\ \Delta Y \\ \Delta Z \end{bmatrix} \quad (6)$$

- where
- $R_O^{C_1}$: The rotation matrix from O frame to C_1 frame
 - $R_O^{C_2}$: The rotation matrix from O frame to C_2 frame
 - $r_{C_1}^O$: The vector of C_1 in O frame
 - $r_{C_2}^O$: The vector of C_2 in O frame
 - $R_{C_1}^{C_2}$: The rotation matrix from C_1 frame to C_2 frame
 - $r_{C_1 \rightarrow C_2}^O$: The translation from C_1 to C_2 in O frame

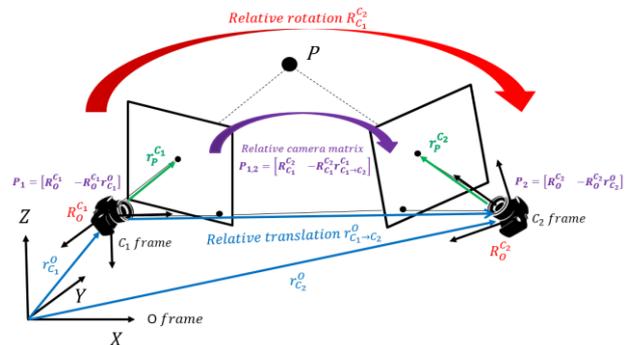


Figure 2. Relationship of an image pair and an object point

2.2.2 Coherent Relative Orientation

In order to establish the relationship of image sequences incrementally, Individual ROPs of consecutive image pairs should be sequentially coherent. Since the localization technique in VO is incremental, the initialization should be done to determine the starting points of the path. The first camera frame here is the reference frame as O frame. Based on the initialized camera, EOPs of each image can be determined incrementally based on ROPs of consecutive image pairs. However, EOPs here are not defined in the mapping Frame (M frame) that cannot describe absolute orientation known in general. Therefore, a new term, coherent relative orientation parameters (CROPs) which substitutes the absolute orientation to describe EOPs is proposed. Figure 3 shows CROPs and ROPs of image sequences. Eventually, each CROPs of image sequences can be determined incrementally by using ROPs of consecutive image pairs.

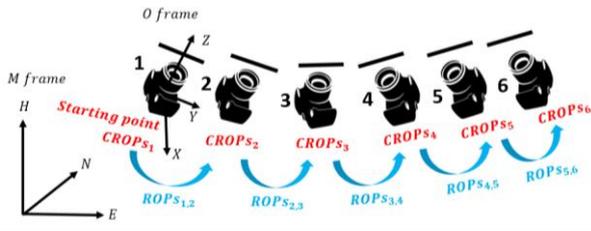


Figure 3. CROPs and ROPs of image sequences

2.3 Network Adjustment

In order to suppress the accumulated error over time, there are several methods proposed such as bundle adjustment (Triggs et al., 1999) in photogrammetry and pose-graph optimization (Grisetti et al., 2010) in computer vision. However, only ROPs estimated is unable to refine EOPs so far no matter which methods are used. Therefore, considering all possible ROPs of arbitrary image pair in images sequences become the strategy to overcome the above mentioned issue. When a camera is focus on an object to take images sequentially from different location, a consecutive image pairs could be obtained. For arbitrary image pair that is not necessarily adjacent, there is a relationship between them which can be described by ROPs. Consequently, all possible ROPs of arbitrary image pairs can form the network image pairs. Figure 4 illustrates an example of initial network image pairs. There are 15 set of possible ROPs of arbitrary image pair.

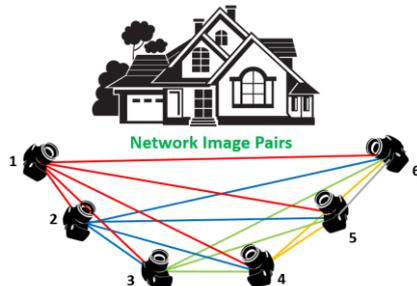


Figure 4. An example of initial network image pairs

However, the matching results may not be reliable since SURF does not take the geometry into account. Therefore, related examinations to remove the mismatches are necessary. After implementing the examinations, some image pairs may be discarded. However, the final ROPs would be more reliable. Figure 5 shows an example of the final network image pairs after the examinations.

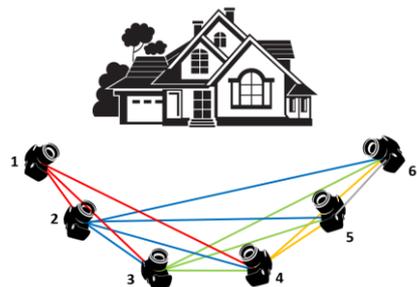


Figure 5. An example of the final network image pairs after the examinations

Figure 6 (a) shows an example of the open path and (b) shows this path is divided in to 4 groups. Here are two properties of these consecutive images. The first one is the current image is far from the initial image; therefore, the loop closure cannot be

achieved. The second one is the adjacent images are overlapped in a period of traveling. According to the overlapping in a part of traveling, consecutive images could be divided into 4 groups. Actually, it is an alternative way to achieve loop closure for a group. If optimal solution of CROPs in each group is obtained, the overall accuracy will be improved as well. Consequently, this study focuses on network adjustment of ROPs for a small group of consecutive images. Take the first group in figure 6 (b) as an example. There are 4 consecutive images covering one building and 6 possible image pairs should be tested to estimate reliable ROPs of consecutive image pairs.

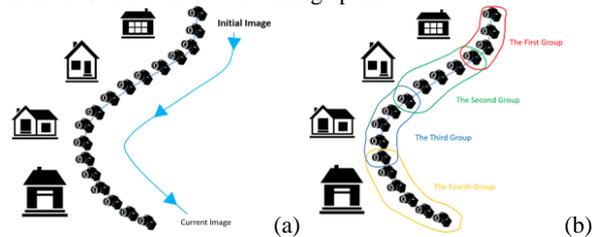


Figure 6. An example of the open path divided in to 4 groups

Network adjustment can be separated into the angular adjustment and translation adjustment. The angular adjustment is implemented at first. In angular adjustment, the observation equation of rotation matrix can be described as Eq. (7). However, it is a nonlinear problem. The linearization is not implemented easily, since three elements, ω , φ and κ , in rotation matrix are not independent. In this study, the hypothesis is built: three rotation angles, ω , φ and κ , are independent of each other. Consequently, the Eq. (7) can be divided into three parts, which are described in the Eq. (8) to Eq. (10).

$$R_{C_i}^{C_j} + v_{R_{C_i}^{C_j}} = R_O^{C_j} \cdot R_{C_i}^O \quad (7)$$

$$\omega_{i,j} + v_{\omega_{i,j}} = \omega_j - \omega_i \quad (8)$$

$$\varphi_{i,j} + v_{\varphi_{i,j}} = \varphi_j - \varphi_i \quad (9)$$

$$\kappa_{i,j} + v_{\kappa_{i,j}} = \kappa_j - \kappa_i \quad (10)$$

where $R_{C_i}^{C_j}$: the rotation from C_i frame to C_j frame
 i, j : the order of the first and second image

Next, the translation adjustment is implemented after the angular adjustment. The Translation is composed of direction and scale of vector. Therefore, both of them should be solved in translation adjustment together. The observation equation of translation vector can be written as Eq. (11). However, only observation of direction is not enough to obtain the reliable solution of scale factors. Consequently, additional pseudo observation of scales should be added into network adjustment. It is known as additional conditions of adjustment. Eq. (12) expresses the pseudo observation equation of scale factor. These pseudo observations of scale factors are derived from measurement of the translation between two cameras. In order to recover all scale factors of consecutive image pairs, one real scale of image pair should be fixed. In general, the measured scale between the first image and second image, $S_{1,2}$, is fixed.

$$\Delta X_{i,j}^{(0)} + v_{\Delta X_{i,j}^{(0)}} = \frac{X_j - X_i}{S_{i,j}} \quad (11)$$

$$S_{i,j}^{(0)} + v_{S_{i,j}^{(0)}} = S_{i,j} \quad (12)$$

where i, j : The order of the first and second image
 $\Delta X_{i,j}^{(0)}$: Observation of translation vector
 X_i, X_j : Location of the first and second camera

$S_{i,j}$: Real scale of translation vector
 $S_{i,j}^{(O)}$: Observation of scale factor

3. EXPERIMENT

PointGrey Camera is used to capture the image sequence in this experiment. About 20 images are captured in different angle and used by Matlab Camera Calibration APP to calibrate the camera in advance. Figure 7 (a) shows the test field and (b) shows the path of camera and image sequence. 10 stations are set up to capture 9 consecutive image pairs. The approximate length of baseline of each adjacent station is measured which is close to 1 meter.

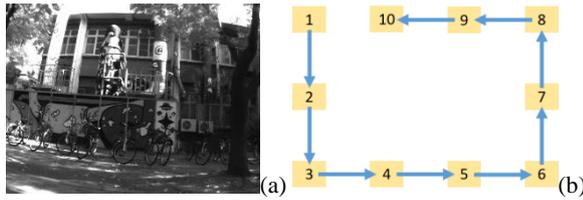


Figure 7. (a) The test field (b) The path of the camera and image sequence

4. RESULTS AND DISCUSSIONS

Figure 8 shows the SURF matching result of one image pair that the outliers have been eliminated by modified RANSAC. It can be observed that the correspondences are nearly correct.



Figure 8. The SURF matching result of one image pair after RANSA

There are totally 44 ROPs are estimated. However, only 11 ROPs could pass the examinations in the end. These ROPs are used as Observations in network adjustment. Figure 9 shows the estimated pose and location of each camera. Figure 10 shows the comparison of related trajectories. Red one means the estimated path before adjustment. Blue ne means the estimated path after adjustment. The difference between them can be distinguished clearly.

Because there is no actual path to directly validate the proposed methodology. Several random multiple-conjugate points are searched to compute the deviation of the object points as the evaluation. In the experiment, 10 random triple-conjugate points are selected from image NO. 1, image NO. 2 and image NO. 3 to reconstruct the object points. Table 1 lists deviations of 10 random triple-conjugate points and the comparison. The first two columns show the deviation of object points before and after adjustment. The last column shows the change of the deviation. According to table 1, the change of deviation of object points almost decrease after adjustment. Consequently, network adjustment is effective to solve the better CROPs.

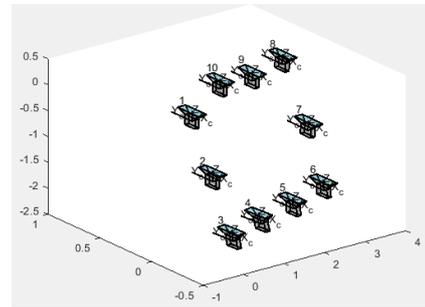


Figure 9. The estimated pose and location of each camera after adjustment

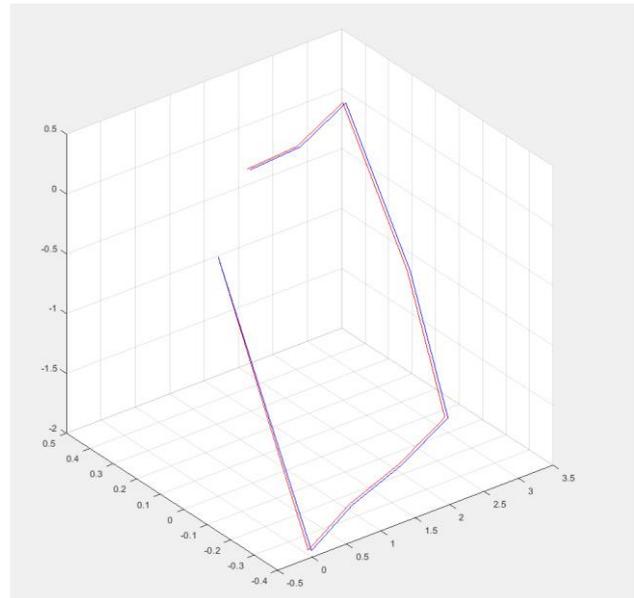


Figure 10. Comparison of estimated camera path before adjustment and estimated camera path after adjustment

Table 1. Deviations of 10 random triple-conjugate points and the comparison.

Random object points	Deviations before adjustment	Deviations after adjustment	Comparison
	$\begin{bmatrix} X_{13} - X_{12} \\ Y_{13} - Y_{12} \\ Z_{13} - Z_{12} \end{bmatrix}$	$\begin{bmatrix} \hat{X}_{13} - \hat{X}_{12} \\ \hat{Y}_{13} - \hat{Y}_{12} \\ \hat{Z}_{13} - \hat{Z}_{12} \end{bmatrix}$	$\left\ \begin{bmatrix} \hat{X}_{13} - \hat{X}_{12} \\ \hat{Y}_{13} - \hat{Y}_{12} \\ \hat{Z}_{13} - \hat{Z}_{12} \end{bmatrix} - \begin{bmatrix} X_{13} - X_{12} \\ Y_{13} - Y_{12} \\ Z_{13} - Z_{12} \end{bmatrix} \right\ $
1	$\begin{bmatrix} 0.010 \\ -0.004 \\ 0.050 \end{bmatrix}$	$\begin{bmatrix} -0.005 \\ 0.001 \\ -0.022 \end{bmatrix}$	$\begin{bmatrix} -0.005 \\ -0.003 \\ -0.028 \end{bmatrix}$
2	$\begin{bmatrix} 0.017 \\ -0.004 \\ 0.057 \end{bmatrix}$	$\begin{bmatrix} 0.000 \\ 0.004 \\ -0.013 \end{bmatrix}$	$\begin{bmatrix} -0.017 \\ -0.001 \\ -0.044 \end{bmatrix}$
3	$\begin{bmatrix} 0.032 \\ -0.052 \\ -0.157 \end{bmatrix}$	$\begin{bmatrix} 0.021 \\ -0.037 \\ -0.110 \end{bmatrix}$	$\begin{bmatrix} -0.011 \\ -0.015 \\ -0.047 \end{bmatrix}$
4	$\begin{bmatrix} 0.019 \\ -0.008 \\ -0.044 \end{bmatrix}$	$\begin{bmatrix} -0.005 \\ 0.001 \\ -0.010 \end{bmatrix}$	$\begin{bmatrix} -0.015 \\ -0.007 \\ -0.034 \end{bmatrix}$
5	$\begin{bmatrix} 0.028 \\ 0.008 \\ 0.075 \end{bmatrix}$	$\begin{bmatrix} 0.006 \\ 0.004 \\ 0.015 \end{bmatrix}$	$\begin{bmatrix} -0.022 \\ -0.005 \\ -0.060 \end{bmatrix}$
6	$\begin{bmatrix} 0.003 \\ -0.045 \\ 0.171 \end{bmatrix}$	$\begin{bmatrix} -0.004 \\ -0.027 \\ 0.102 \end{bmatrix}$	$\begin{bmatrix} 0.002 \\ -0.018 \\ -0.069 \end{bmatrix}$
7	$\begin{bmatrix} 0.006 \\ 0.012 \\ 0.110 \end{bmatrix}$	$\begin{bmatrix} -0.005 \\ -0.012 \\ -0.103 \end{bmatrix}$	$\begin{bmatrix} -0.001 \\ 0.000 \\ -0.007 \end{bmatrix}$
8	$\begin{bmatrix} 0.030 \\ 0.011 \\ 0.095 \end{bmatrix}$	$\begin{bmatrix} 0.007 \\ 0.005 \\ 0.020 \end{bmatrix}$	$\begin{bmatrix} -0.022 \\ -0.006 \\ -0.066 \end{bmatrix}$
9	$\begin{bmatrix} 0.011 \\ -0.018 \\ 0.113 \end{bmatrix}$	$\begin{bmatrix} 0.000 \\ -0.008 \\ 0.040 \end{bmatrix}$	$\begin{bmatrix} -0.011 \\ -0.010 \\ -0.073 \end{bmatrix}$
10	$\begin{bmatrix} -0.007 \\ -0.018 \\ 0.057 \end{bmatrix}$	$\begin{bmatrix} -0.013 \\ -0.003 \\ 0.012 \end{bmatrix}$	$\begin{bmatrix} 0.006 \\ -0.015 \\ -0.044 \end{bmatrix}$

5. CONCLUSIONS AND SUGGESTIONS

This study proposed a strategy to automatically solve the relative orientation of image sequences. Moreover, instead of

current optimization methods including bundle adjustment and pose-graph optimization, network adjustment based on ROPs is applied to solve the better CROPs. In other words, the accumulated errors over time could be reduced. According to deviations of object points derived from different image pairs before and after network adjustment, the network adjustment is feasible and effective to suppress the drift in the navigation.

However, the proposed network adjustment using ROPs is still a preliminary study due to two hypotheses which are not entirely correct and should be discarded. In addition, real-time navigation cannot be achieved by manually measuring translation as the scale. Consequently, alternative approaches could be tried, such as the aid of other sensors or the reference of road plane and camera height.

ACKNOWLEDGEMENTS

The author would acknowledge the financial supports provided by the Ministry of Science and Technology.

REFERENCES

- Bay, H., Tuytelaars, T., Van Gool, L., 2006. Surf: Speeded up robust features, European conference on computer vision. Springer, pp. 404-417.
- Davison, A.J., 2003. Real-time simultaneous localisation and mapping with a single camera. IEEE International Conference on Computer Vision, p. 1403.
- Fischler, M.A., Bolles, R.C., 1981. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM* 24, 381-395.
- Fraundorfer, F., Scaramuzza, D., 2012. Visual odometry: Part ii: Matching, robustness, optimization, and applications. *IEEE Robotics & Automation Magazine* 19, 78-90.
- Förstner, W., 2002. Computer vision and photogrammetry—mutual questions: geometry, statistics and cognition. *Bildtechnik/Image Science, Swedish Society for Photogrammetry and Remote Sensing*, pp 151-164.
- Grisetti, G., Kummerle, R., Stachniss, C., Burgard, W., 2010. A tutorial on graph-based SLAM. *IEEE Intelligent Transportation Systems Magazine* 2, 31-43.
- Joglekar, J., Gedam, S.S., 2012. Area based image matching methods—A survey. *Int. J. Emerg. Technol. Adv. Eng* 2, 130-136.
- Juan, L., Gwon, O., 2009. A comparison of sift, pca-sift and surf. *International Journal of Image Processing (IJIP)* 3, 143-152.
- Lemaire, T., Berger, C., Jung, I.-K., Lacroix, S., 2007. Vision-based slam: Stereo and monocular approaches. *International Journal of Computer Vision* 74, 343-364.
- Lowe, D.G., 2004. Distinctive image features from scale-invariant keypoints. *International journal of computer vision* 60, 91-110.
- Nistér, D., Naroditsky, O., Bergen, J., 2004. Visual odometry, *Computer Vision and Pattern Recognition, 2004. CVPR 2004. Proceedings of the 2004 IEEE Computer Society Conference on*. Ieee, pp. I-I.
- Triggs, B., McLauchlan, P.F., Hartley, R.I., Fitzgibbon, A.W., 1999. Bundle adjustment—a modern synthesis, *International workshop on vision algorithms*. Springer, pp. 298-372.
- Zhang, Z., 1999. Flexible camera calibration by viewing a plane from unknown orientations, *Computer Vision, 1999. The Proceedings of the Seventh IEEE International Conference on*. Ieee, pp. 666-673.