

Realtime Image Matching for Vision Based Car Navigation with Built-in Sensory Data

K. Choi, S. Tanathong, H. Kim, I. Lee*

Dept. of Geoinformatics, The University of Seoul, 163 Seoulsiripdaero, Dongdaemun-gu, Seoul 130-743, Korea
– shale@uos.ac.kr, littlebearproject@gmail.com, zzimss@uos.ac.kr, iplee@uos.ac.kr

Commission III, WG III/3

KEY WORDS: Position, Attitude, Navigation, Image, Velocity, Angular rate, Front Camera, Built-in Sensors

ABSTRACT:

Recently, a car employs various built-in sensors such as a speedometer, odometer, accelerometer and angular rate sensor for safety and maintenance. These sensory data can be provided in real time through a CAN (Controller Area Network) bus. In addition, image sequences can be provided from various cameras mounted to the car, such as built-in front and around view monitoring cameras. We thus propose an image based car navigation framework to determine car position and attitude using the built-in sensory data such as a speed, angular rate and images from a front view camera. First, we determine the two-dimensional position and attitude of a car using the velocity and angular rate provided in real-time through the CAN bus. We then estimate the three-dimensional position and attitude by conducting sequential bundle block adjustment using the two-dimensional position and attitude and tie points between image sequences. The sequential bundle adjustment can produce accurate results comparable to those from the conventional simultaneous bundle adjustment in real time. As the input to this process, it needs reliable tie points between adjacent images acquired from a real-time image matching process. Hence, we develop an image matching process based on the enhanced KLT algorithm using preliminary exterior orientation parameters. We also construct a test system that can acquire and store built-in sensory data and front camera images at the same time, and conduct experiments with the real data acquired by the system. The experimental results show that the proposed image matching process can generate accurate tie-points with about 0.2 second in average at each epoch. It can successfully meet the requirements from real-time bundle adjustment for image based car navigation.

1. INTRODUCTION

A car navigation system typically determines the car position on a road of a map using a satellite navigation system and map-matching techniques. Thanks to the advances of navigation sensors and map making technology, the navigation systems provide better performance with lower prices and thus are popularly used. Furthermore, as the use of mobile devices such as smart phones increases, car navigation based on such a device becomes more prevalent. However, the positional accuracy of such systems is still not fully satisfactory for more sophisticated guidance with high resolution maps, which requires less than 1 m accuracy. It is mainly because most navigation systems highly depend on the GPS performance that can be greatly affected by the operating environment (Bayoud, 2005; Dissanayake et al., 2001; Hasan et al., 2010; El-Sheimy et al., 2006; Limsuonthrakul et al., 2009). If a car is in the areas where GPS signals are weak or in outage, for example, a road surrounded by high buildings or a tunnel, the navigation system cannot define the accurate position and attitude of the car. In addition, when a car is in an interchange, it has difficulty to decide whether the car is on a ramp or not.

To overcome these problems, some studies have attempted to combine GPS with other sensors such as INS and cameras (Langel Steven et al., 2010; Andrey and Donald, 2010; Kim et al., 2011; Yoo et al., 2005). Especially, hybrid approaches with additional cameras seem to be promising since many cars tend to have built-in front and around view monitoring cameras (Andrey and Donald, 2010; Kim et al., 2011; Yoo et al., 2005; Madison et al., 2007; Nettleton and Sukkarieh, 2004; Goldshtein et al., 2007; Kim et al., 2012). For example, Kim et al. (2012) extracted lanes from images and estimated displacement to correct GPS estimation, but this approach has limitation under the situation where the number of visible satellites is not enough for a long time or the lane detection are

difficult. In addition, Kim et al. (2011) employed omnidirectional cameras to solve a problem with narrow field of view. They used not only vision sensors but also navigational sensors such as INS and an odometer. In this approach, they extracted and tracked vanishing points from the omnidirectional images. The tracking results were utilized to estimate vehicle attitude, which then complemented the INS quality. Although the accuracy of this approach is better than conventional methods, it is difficult to commercialize because the omnidirectional cameras are still expensive.

As inexpensive solutions, some researchers proposed smartphone based methods since a smartphone inherently include many sensors such as GPS, gyro, compass, and cameras (Retsher et al., 2013; Chang et al., 2013). The combination of these sensory data is useful for the estimation of the position and attitude of the smart phone and so the car where it is inside. Furthermore, other researchers focused on using only in-vehicle built-in sensors (Jo et al., 2011; Jung et al., 2012). Most cars already have many built-in sensors providing velocity, yaw-rate, longitudinal/lateral acceleration, etc. These data can be combined to improve the navigational accuracy.

In this study, we thus propose a vision based car navigation framework using not only the built-in sensory data but also image sequences. This framework consists of two stages, (1) preliminary determination using built-in sensory data, and (2) enhancement using image sequences. First, we determine the two-dimensional position and attitude using the velocity and angular rate. We then estimate the three-dimensional position and attitude by conducting sequential bundle block adjustment using the two-dimensional position and attitude and tie points between image sequences. As the input to this process, it needs reliable tie points between adjacent images acquired from a real-time image matching process.

* Corresponding author

There exist a number of studies that present image matching techniques to obtain reliable tie-points (Honkavaara and Hoghollen, 1996; Barazzetti et al., 2010; Jin et al., 2001; Hwangbo et al., 2011; Tanathong and Lee, 2011). Most of these techniques deal with images captured over static scene. The car navigation problem, however, involves extensively with traffic scene. This is a challenging task for image matching as the captured images usually include moving objects, while the adjustment process requires that tie-points must only be associated with stationary objects.

In order to detect moving objects, many research studies consider optical flow constraints such as presented in (Ebine and Hamada, 2002; Shibata et al., 2008). Jung and Sukhatme (2004) used the obtained tie points to determine image transformation parameters which can be affine, bilinear or pseudo-perspective models. Tie-points poorly fitting the selected transformation model are classified as mismatches. The existing works, however, only apply to the case when there are a few moving objects. They are not applicable for car navigation problem, especially under urban traffic environment.

Hence, in addition to propose the framework, we also develop an image matching technique for the sequential bundle adjustment. The technique is based on the enhanced KLT algorithm using the preliminary position and attitudes, being equipped with novel post-processing to remove the points from moving objects. Following this introduction, we introduce the image based navigation framework, describe our image matching technique, provide some experimental results, and present some conclusions and future research.

2. IMAGE BASED NAVIGATION FRAMEWORK

We propose a framework to determine the position and attitude more precisely using in-vehicle built-in sensors and image sequences. As shown in Figure 1, the framework consists of two stages: (1) determining preliminary position and attitude using built-in sensory data such as velocity and angular rates, and (2) refining the position and attitude through sequential bundle adjustment using image sequences from a front camera. Actually, the outputs are the results of a dead reckoning process, which are relative to starting position and attitude. However, if the outputs are combined with other sensory data provided in an absolute coordinate system, such as GPS data, they can be converted to the absolute position and attitude to be used for car navigation.

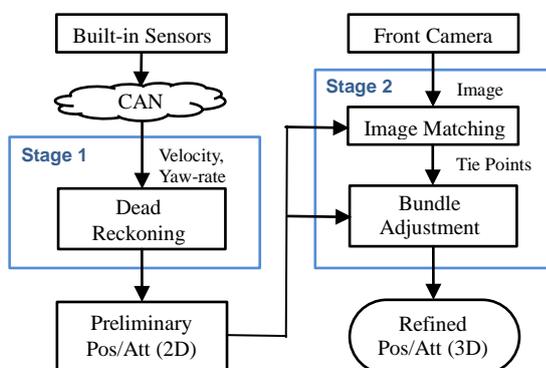


Figure 1. Overview of framework

First, the system determines relative preliminary position and attitude through dead reckoning using built-in sensors only. It receives these sensory data through a CAN (Controller Area Network) bus designed for sensor communication. Among the data, the velocity and yaw rate are used to determine two-dimensional position and attitude, based on Eq. (1)-(2). At a time epoch, if we know the current position (x_n, y_n) and attitude (θ_n) and acquire the yaw-rate $(\dot{\theta}_n)$ and velocity (v_n) , then we compute the next (t_{n+1}) position (x_{n+1}, y_{n+1}) and attitude (θ_{n+1}) .

$$\begin{bmatrix} x_{n+1} \\ y_{n+1} \end{bmatrix} = \begin{bmatrix} x_n \\ y_n \end{bmatrix} + \begin{bmatrix} \cos \theta_n & -\sin \theta_n \\ \sin \theta_n & \cos \theta_n \end{bmatrix} v_n \Delta t \quad (1)$$

$$\theta_{n+1} = \theta_n + \dot{\theta}_n \Delta t \quad (2)$$

Second, the system then refines the position and attitude using image sequences. It is assumed that the image sequences can be acquired by a front view camera, which can be an inherently built-in one or that installed by a user, such as a car black box and a smartphone. By analyzing the image sequence whenever a new image is added into the sequence, we can determine the relative positions and attitudes of the camera at the exposure time of every image in the sequence. Based on these relative orientations, we can refine the position and attitude of the car rigidly mounting the camera. This stage includes two main processes, image matching and bundle adjustment.

Image matching extracts tie points between overlapping images, which are corresponding the same object points. The tie points are then entered into the bundle adjustment process with the preliminary determined position and attitude at the first stage. This process then refines these position and attitude by adjusting the projected ray (or light bundle) to each image point so that all the rays corresponding to the same object points can intersect at a point. This method enables to acquire more precise results in comparison with the previous ones using the built-in sensors only.

As car navigation is a real-time application, all the processes in the framework should be performed in real-time. The first stage involves simple computation based on Eq. (1)-(2) and thus works in real-time. In the second stage, the bundle adjustment process has been generally performed as post-processing in a batch mode where all the images are processed at the same time. To perform it in real-time whenever a new image is acquired, we employ a new efficient sequential bundle adjustment algorithm proposed by Choi and Lee (2013). This algorithm can significantly reduce the computation time at each epoch by using the computational results from the previous epoch at most. For the bundle adjustment, the image matching process should produce a certain number of accurate tie points reliably in real-time. This process is described in detail in the next section.

3. REALTIME IMAGE MATCHING TECHNIQUE

Tie-points play a crucial part to the success of a bundle adjustment process. In computer vision, there exist many image matching techniques that can obtain tie-points automatically. To construct a car navigation system, not only does the mentioned image matching technique have to produce reliable tie-points but also have to function fast enough to achieve a real-time requirement. Kanade-Lucas-Tomasi (KLT) is a robust feature detection and tracking technique (Lucas et al., 1981; Shi and Tomasi, 1994). The technique is well-known for its

computational efficiency, and widely used for real-time applications, see (Conte and Doherty, 2008; Pollefeys et al., 2008; Clipp et al., 2009; Benfold and Reid, 2011; Tanathong and Lee, 2011).

Although KLT is a promising approach to the real-time acquisition of tie-points, extracting tie-points from urban traffic scenes captured from a moving camera is a challenging task. To be used as a source of inputs for the bundle adjustment process, tie-points must not be acquired from moving objects but only from stationary objects. When the camera (observer) is at a fixed position, moving objects can be distinguished from stationary objects by considering the direction and magnitude of optical flow vectors. However, when the camera moves, it also induces optical flows for stationary objects. This makes it difficult to separate them from moving objects. The problem is more complicated on road scenes which involve several moving objects. At this point, the problem of image matching is not only to produce tie-points but also to discard those associated with moving objects.

This study presents an image matching system based on the KLT algorithm. To simplify the aforementioned problem, the built-in sensory data are employed. The sensors offer translation velocity and angular velocity of the camera (in fact, the vehicle that boards the camera). These data can be used to derive the position and attitude parameters of the camera, which will be referred to as preliminary exterior orientation (EO) or EO throughout this paper. The procedure of the proposed image matching for a car navigation system is illustrated as Fig. 2.

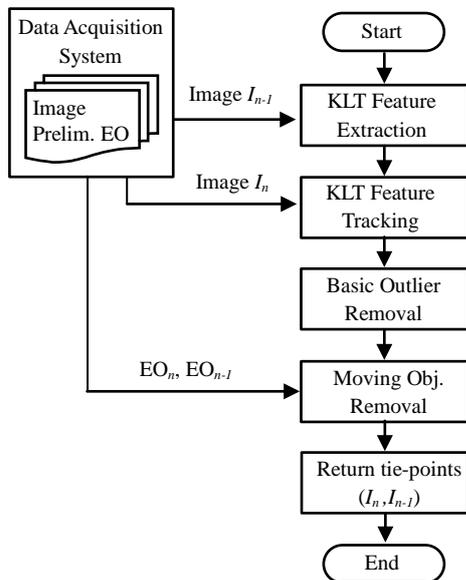


Figure 2. Image matching procedure

Although the camera can capture several images per second (let assume N fps), to avoid the computation burden of both image matching and bundle adjustment process, the image matching system is designed to deliver tie-points to the end application once a second. KLT, however, only works when the displacement between frames is small. To enable features extracted from $I_{n-1,l}$ (the first frame captured at the $n-1$ second) to be tracked on $I_{n,l}$ (the first frame of the n second) and further, we also need to perform tracking on a few intermediate images captured between them (in a pair-wise basis). Although these additional images may increase overall processing time, they

can shorten the tracking distances and improve the tracking ability of KLT. To avoid a large number of moving objects and prevent tie-points to appear along the optical axis of the camera, the middle area of the image is ignored and features are only extracted from the area on the left and right sides of the image.

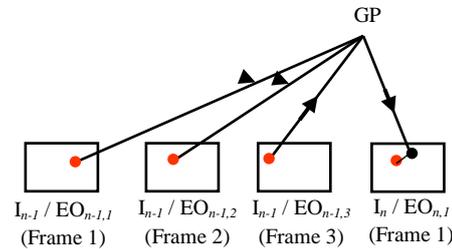


Figure 3. Moving object removal through a projection

The basic outlier removal presented in Fig. 2 to discard mismatches involves (1) detecting mismatches using normalized cross correlation coefficient, (2) performing a cross-check tracking in which tracking corresponding features in the second image should find their starting features in the first image, and (3) eliminating tie-points whose magnitude and direction of optical flow vectors do not align with the majority. To detect tie-points associated with moving object, we use the preliminary EOs derived from built-in sensory data to estimate the ground coordinate of tie-points using the collinearity equations. The underlying concept is that if the corresponding image points belong to a stationary object, they can be project to (or close to) the same ground coordinate. The tie-points corresponding with moving objects, on the other hand, will not associate with the same ground coordinate even though they belong to the same object.

The moving object removal procedure is illustrated in Fig. 3. First, image points that are corresponding to the same ground point (GP) and appear in every image frame of the $n-1$ second together with preliminary EOs are used to estimate the ground coordinate through the collinearity equations. Given the preliminary EOs of the current image frame (the first frame of the n second), the aforementioned ground point is then projected to the current image frame. The discrepancy between the image point obtained from the projection and from the image matching is then determined. If the preliminary EOs are reliable, we can expect the discrepancy to be small for the case of stationary object and may eliminate image points with large discrepancy as they are likely to associate with moving objects. If the EOs are not so reliable, the discrepancy threshold must be defined appropriately to avoid misclassification.

4. EXPERIMENTAL RESULTS

The image matching software is developed in C/C++ based on the OpenCV library (OpenCV, 2008). The testing data include an image sequence captured road scenes of a small town in Seoul, Korea, and their corresponding EO from POS/LV system. (During this experimental testing, the built-in sensory data are not available yet.) The image sequence is captured at 14.7 fps, and the image-size is 1024×768 pixels. In the experimental test, the KLT tracking window size is set to 30×30 pixels, and the number of pyramidal depth levels is assigned to 3, see (Bradski and Kaehler, 2008) for explanation. Experiments are performed on a 1.86GHz computer with 3GB RAM.

In this paper, two experiments are conducted. The first experiment is tested with a sequence of 30 images. The system is configured to process 2 frames per second (the first and the seventh frame of each second), and that results in 14 sets of tie-points. The experimental results on this dataset show that the proposed work can produce 16.4 points/image on average and their correct match ratio is measured as 99.18%. The computational cost of the system to produce tie-points is 195ms/epoch. The second experiment is adjusted to execute 3 frames per second (the first, the sixth, and the eleventh frame). For a sequence of 60 images, it produces 19 sets of tie-points. On the second dataset, the system produces 19.1points/image on average with 99.2% of correct match ratio. The system costs 220ms/epoch. Tie-points from these two experiments are not associated with moving objects. All the experimental results are illustrated in graphs as Fig. 4, and example of tie-points are presented in Fig. 5.

The experimental results suggest that, when using more number of intermediate images, as does on the second experiment, the system can attain a larger number of tie-points. However, it results in more number of mismatches because of the progressively increment of the dissimilarity of tie-points over time. Therefore, this trade-off must be carefully considered. In summary, the proposed image matching can produce very accurate tie-points, greater than 99% of correct match ratio, while the computation cost is less than 0.25sec/epoch on average.

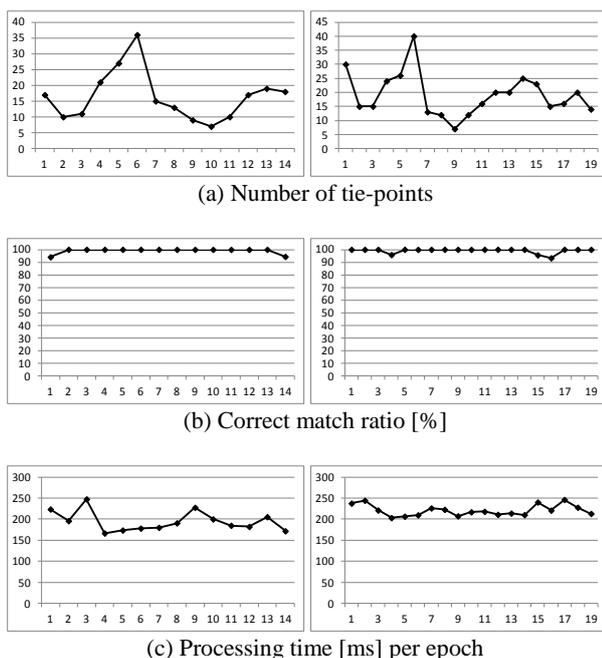


Figure 4. The results of the first experiment are presented on the left and the second experiment on the right.

5. CONCLUSIONS

In this paper, we propose an image based car navigation framework using an image sequence acquired by a front camera and built-in sensory data such as velocity and yaw rate. It consists of two main stages, first determining the preliminary position and attitude using built-in sensory data and refining them using the image sequence.

In the second stage, we particularly focus on the image matching technique that should produce a certain number of tie points in a fast and reliable manner for the subsequent bundle adjustment process. The proposed technique is basically based on the traditional KLT method. But we improve it to benefit from the preliminary EOs provided by the first stage. In addition to conventional mismatch detection procedures, we also employ a procedure to discard tie points from moving objects.

The proposed image matching technique has been applied to two sets of test data. The experimental results show that it can produce tie points with more than 99% of correct match ratio and 0.2s/epoch in average. This technique can be effectively used as the pre-processing process for the realtime sequential bundle adjustment for the image based car navigation.

6. REFERENCE

- Andrey, S., Donald, V. 2010. Integration of GPS and vision measurements for navigation in GPS challenged environments. IEEE Symposium on Position Location and Navigation, Indian Wells, pp. 826-833
- Barazzetti, L., Remondino, F., Scaioni, M., 2010. Fully automated UAV image-based sensor orientation. International Archives of IAPRS, Calgary.
- Bayoud, F.A.. 2005. Vision-aided inertial navigation using a geomatics approach. Proc. ION GNSS-2005, Long Beach, USA, pp. 2485-2493
- Benfold, B., Reid, I. 2011. Stable multi-target tracking in real-time surveillance video. Proceedings of Computer Vision and Pattern Recognition, Colorado Springs, USA, pp. 3457-3464.
- Bradski, G., Kaehler, A., 2008. Learning OpenCV: Computer Vision with the OpenCV Library. O'Reilly, California, USA., pp. 316-367.
- Chang, H., Georgy, J., El-Sheimy, N. 2013. Monitoring Cycling Performance Using a Low Cost Multi-Sensors Navigation Solution. The 8th International Symposium on Mobile Mapping Technology, Tainan, Taiwan
- Choi, K. and Lee, I., 2013. A Sequential Aerial Triangulation Algorithm for Real-time Georeferencing of Image Sequences Acquired by an Airborne Multi-Sensor System, Remote Sensing, 5(1), pp. 57-82.
- Clipp, B., Zach, C., Lim, J., Frahm, J.M., Pollefeys, M., 2009. Adaptive, real-time visual simultaneous location and mapping. Proceedings of IEEE Workshop on Applications of Computer Vision.
- Conte, G., Doherty, P., 2008. An integrated UAV navigation system based on aerial image matching. Proceedings of the IEEE Aerospace Conference, Montana, USA., pp. 1-10.
- Dissanayake, G., Sukkarieh, S., Nebot, E. 2001. The aiding of a low-cost strapdown inertial measurement unit using vehicle model constraints for land vehicle applications. IEEE Trans. Robot., 17(5), pp. 731-747
- Ebine, T., Hamada, N., 2002. Detection of moving objects using observer motion-based optical flow estimation. Systems and Computers in Japan, 33(6), pp. 83-92.

- El-Sheimy, N., Nassar, S., Shin, E.-H., Niu, X. 2006. Analysis of various kalman filter algorithms using different inertial navigation systems integrated with the global positioning system. *Can. Aeronaut. Space J.*, 52(2), pp. 59–67
- Goldshtein, M., Oshmant, Y., Efrati, T. 2007. Seeker gyro calibration via model-based fusion of visual and inertial data. *Information Fusion, 2007 10th International Conference*, Quebec, Canada, pp. 1-8.
- Hasan, A.M., Samsudin, K., Ramli, A.R., Azmir, R.S. 2010. Comparative study on wavelet filter and thresholding selection for GPS/INS data fusion. *Int. J. Wavelets Multiresolution Inf. Process.*, 8(3), pp. 457–473.
- Honkavaara, E., Hogholen, A., 1996. Automatic tie point extraction in aerial triangulation. *International Archives of Photogrammetry and Remote Sensing*, Vienna, pp. 337-342.
- Hwangbo, M., Kim, J., Kanade, T., 2011. Gyro-aided feature tracking for a moving camera: fusion, auto-calibration and GPU implementation. *The International of Robotics Research*, 30(14), pp. 1755-1774.
- Jin, H., Favaro, P., Soatto, S., 2001. Real-time feature tracking and outlier rejection with changes in illumination. *Proceedings of International Conference on Computer Vision*, pp. 684-689.
- Jo, K., Chu, K., Kim, J. and Sunwoo, M. 2011. Distributed Vehicle State Estimation System Using Information Fusion of GPS and In-vehicle Sensors for Vehicle Localization, *IEEE Conference on Intelligent Transport Systems*, pp. 2009-2014.
- Jung, B., Sukhatme, G.S., 2004. Detecting moving objects using a single camera on a mobile robot in an outdoor environment. *Proceedings of the 8th Conference on Intelligent Autonomous Systems*, Amsterdam, pp. 980-987.
- Jung, H., Lee, D., Kim, S., Kim, C., Huh, K., 2012. Development of Vehicle Position Estimator using Dead Reckoning Algorithm. *KSAE 2012 Annual Conference and Exhibition*, pp. 1255-1257.
- Kim, M., Lim, J., Park, J., Kim, H., Lee, H. 2012. Vehicle Displacement Estimation by GPS and Vision Sensor. *The Journal of Korea Navigation Institute*, 16(3), pp. 417-425.
- Kim, S., Bazin, J., Lee, H., Choi, K. and Park, S. 2011. Ground vehicle navigation in harsh urban conditions by integration inertial navigation system, global positioning system, odometer and vision data. *IET Radar, Sonar, & Navigation*, 5(8), pp. 814-823.
- Langel Steven E., Samer M., Chan Fang-Cheng, Pervan Boris S. 2010. Tightly coupled GPS/INS integration for differential carrier phase navigation systems using decentralized estimation. *Proceedings of IEEE/ION PLANS 2010*, Palm Springs, CA, pp. 397-409.
- Limsoonthrakul, S., Dailey, M.N., Parnichkun, M. 2009. Intelligent vehicle localization using GPS, compass, and machine vision'. *IEEE RSJ, Missouri, USA*, pp. 3981–3986.
- Lucas, B., Kanade, T., 1981. An iterative image registration technique with an application to stereo vision. *Proceedings of the International Joint Conference on Artificial Intelligence*, pp. 674-679.
- Madison, R., Andrews, G., Debitetto, P., Rasmussen, S., Bottkol, M. 2007. Vision-aided navigation for small UAVs in GPS-challenged environments. *AIAA InfoTech at Aerospace Conf.*, pp. 2733–2745.
- Nettleton, E., Sukkarieh, S. 2004. Real-time experiment of feature tracking/mapping using a low-cost vision and GPS/INS system on an UAV platform. *J. Glob. Posit. Syst.*, 3(1), pp. 167–172.
- OpenCV, 2008. Open Source Computer Vision Library, <http://sourceforge.net/projects/opencvlibrary/files/opencv-win/1.1pre1/> (18 June 2013).
- Pollefeys, M., Nister D., Frahm J.M., Akbarzadeh, A., Mordohai, P., Clipp B., Engels, C., Gallup, D., Kim, S.J., Merrell, P., Salmi, C., Sinha, S., Talton, B., Wang, L., Yang, Q., Stewenius, H., Yang, R., Welch, G., Towles, H., 2008. Detailed real-time urban 3D reconstruction from video. *International Journal of Computer Vision*, 78(2-3), pp. 143-167.
- Retscher, G., Hecht, T., Mok, E. 2013. Location Capabilities and Performance of Smartphones for LBS Navigation Applications. *The International Symposium on Mobile Mapping Technology*, Tainan, Taiwan.
- Shi, J., Tomasi, C., 1994. Good features to track. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 593-600.
- Shibata, M., Yasuda, Y., Ito, M., 2008. Moving object detection for active camera based on optical flow distortion. *Proceedings of the 17th World Congress – The International Federation of Automatic Control*, Seoul, pp. 14720-14725.
- Tanathong, S., Lee, I., 2011. An automated real-time image georeferencing system. *Proceedings of International Conference of Image Processing, Computer Vision, and Pattern Recognition*, Las Vegas, USA, pp. 68-74.
- Yoo, J., Choi, J., Sung, K., Kim, J. 2005. Vehicular image based geographic information system for telematics environments. *Geoscience and Remote Sensing Symposium, 2005. IGARSS '05 Proceedings, 2005 IEEE International*, 2, pp. 1206-1209.



(a) Tie-points from the first experiment



(b) Tie-points from the second experiment

Figure 5. Examples of tie-points