POSE ESTIMATION OF A MOVING CAMERA WITH LOW-COST, MULTI-GNSS DEVICES

Manolis Lourakis[1,*], Maria Pateraki[1], Ion-Anastasios Karolos[2], Christos Pikridas[2], Petros Patias[2]

[1] Institute of Computer Science, Foundation for Research and Technology – Hellas, Heraklion, Greece
(lourakis, pateraki)@ics.forth.gr
[2] School of Rural & Surveying Engineering, Aristotle University of Thessaloniki, Thessaloniki, Greece
(ikarolos, cpik, patias)@topo.auth.gr

Commission II, WG 1

KEY WORDS: Pose, georeferencing, exterior orientation, absolute orientation, multi-GNSS, RTK

ABSTRACT:

Without additional prior information, the pose of a camera estimated with computer vision techniques is expressed in a local coordinate frame attached to the camera's initial location. Albeit sufficient in many cases, such an arbitrary representation is not convenient for employment in certain applications and has to be transformed to a coordinate system external to the camera before further use. Assuming a camera that is firmly mounted on a moving platform, this paper describes a method for continuously tracking the pose of that camera in a projected coordinate system. By combining exterior orientation from a known target with incremental pose changes inferred from accurate multi-GNSS positioning, the full 6 DoF pose of the camera is updated with low processing overhead and without requiring the continuous visual tracking of ground control points. Experimental results of applying the proposed method to a moving vehicle and a mobile port crane are reported, demonstrating its efficacy and potential.

1. INTRODUCTION

Geometric computer vision can provide a wealth of measurements about an imaged scene (Hartley, Zisserman, 2004). Without any additional prior information, these measurements are nevertheless expressed in an arbitrary local coordinate system related to the employed camera, e.g. (Snavely et al., 2008). Furthermore, monocular 3D reconstruction is possible only up to an isotropic scaling, i.e. the 3D structure and the translational component of camera motion are defined up to an unknown scale factor (Lourakis, Zabulis, 2013a). However, such a scaled, camera-centered representation is not always suitable, especially when camera measurements need to be combined with map data. To deal with this issue, the camera should be georeferenced, i.e. its local coordinate system should be aligned with a ground coordinate system (Hackeloeer et al., 2014).

The task of estimating the 6D camera position and orientation with respect to an external coordinate system is commonly referred to as exterior orientation. The computation of the exterior orientation parameters usually relies on the measurement of ground control points, which in the case of non-stationary cameras with limited control over their motion and the particularities of the imaged environment, is a practical limitation. Especially when operating outdoors, automatic feature extraction and matching is challenged by highly homogeneous areas, repetitive textures, large variations in illumination and occlusions, and its failure can severely impact the accuracy and reliability of the estimated measurements. Such obstacles can be overcome by solutions based on non-visual sensors.

Global navigation satellite systems (GNSS) devices can provide the three-dimensional geodetic coordinates of measured environment points in real-time (Fotiou, Pikridas, 2012). When incorporating differential techniques such as real-time kinematic (RTK) (Rietdorf et al., 2006) that combine measurements of the phase of the radio signal's carrier wave with real-time corrections from an accurately known reference station, GNSS devices can attain centimetre positional accuracy. Such receivers have been the standard practice in airborne mapping applications but their utilization in other application domains in need of high accuracy has been constrained by their large size and high cost. Promising solutions have emerged from the recent advent of compact, low-cost multi-GNSS devices which offer further accuracy and increased coverage benefits by intelligently using the largest number of visible satellites from different GNSS satellite constellations (U-blox, 2020).

This paper describes and evaluates an approach for georeferencing a pinhole camera rigidly mounted on a mobile platform. It employs a set of visually distinct ground control points to georeference the camera at a reference location, combined with a stream of position information acquired from a triplet of compact, low-cost GNSS receivers. The ground control points have been surveyed with an RTK GNSS receiver and their surveyed coordinates were converted to a Transverse Mercator projection coordinate system (Veis, Paradissis, 1990). Camera exterior orientation using the ground control points and their image projections is then performed. This computation is carried out in a robust regression framework to mitigate the effect of mislocalized or mismatched image points. A triplet of compact, low-cost multi-band GNSS receivers that are firmly attached to different locations on the platform provides the reference coordinates of three platform points. As the platform moves to new locations, the coordinates of these three platform points are continuously measured. Solving for the absolute orientation between the reference and the most recently measured locations of the platform points estimates the motion of the platform relative to the reference points. Finally, the camera pose at a new location is obtained by combining the camera reference pose with the platform motion.

[*] Corresponding author.

https://doi.org/10.5194/isprs-archives-XLIII-B2-2020-55-2020

The proposed method has low computational overhead, thus it can provide pose estimates at high frequency. No special constraints or adaptations of the camera installation are required. Furthermore, the ground control points need to be surveyed only once and used off-line for estimating the camera exterior orientation at the reference location. In contrast to most existing approaches for camera georeferencing, the surveyed ground points need not always be visible in images and the camera is allowed to move freely in the environment. It is only required that the ground control points are visible from the reference location. The remainder of the paper is organized as follows. Relevant previous work is briefly reviewed in Section 2. Section 3 provides an overview of some background information. The proposed approach is detailed in Section 4. Experimental results are presented in Section 5 and the paper is concluded in Section 6.

2. PREVIOUS WORK

Camera georeferencing, i.e. determining the location where an image was acquired, has been studied, often independently, by both the photogrammetry and the computer vision communities. Despite recent progress, automatic georeferencing of images remains a challenging task (Zamir et al., 2016).

For more than 20 years, direct georeferencing, i.e. camera device exterior orientation via the integration of GNSS and inertial (IMU) measurements from on-board sensors, has been a standard photogrammetric procedure for airborne and satellite images (Cramer et al., 2001). This is especially the case for airborne linear imaging sensors (Pateraki, 2005), for which orientation parameters have to be estimated for each set of scan lines due to the high motion dynamics. Beyond the aerial domain, mobile mapping systems driven by high accuracy requirements have emerged in terrestrial and marine environments, employing bulky and expensive equipment (Cavegn et al., 2018). To overcome issues with GNNS signal degradation experienced in ground-based systems, (Nebiker et al., 2012, Jende et al., 2017) proposed the fusion of ground-based imagery from mobile mapping systems with aerial imagery. Despite their limited payload and battery autonomy, current unmanned aerial vehicles (UAVs) can be equipped with miniaturized high resolution cameras, accurate GNSS receivers and reliable IMUs, thereby facilitating affordable direct georeferencing (Gabrlik, 2015).

The earliest works concerned with georeferencing in the computer vision literature also focused on airborne or satellite images. For example, (Wildes et al., 2001) register a video stream augmented with telemetry to geodetically calibrated reference imagery in the form of a digital orthoimage and elevation map. (Shan et al., 2014) perform ground to aerial image matching in order to georegister ground-based multi-view stereo models. (Hourdakis, Lourakis, 2015) match a sparse, geometric representation between ground and orbital images and use it to refine the pose of ground images initially obtained via visual odometry.

The advent of consumer digital photography two decades ago combined with the ease of storing and sharing such data, has led to an explosion in consumer digital image production. As a result, attention was shifted to images obtained from ground-level viewpoints and with less controlled acquisition procedures. The prevailing paradigm has been to adopt an image retrieval strategy driven by large datasets of geo-tagged images, e.g. (Hays, Efros, 2008, Zhang, Kosecka, 2006). Often, these approaches apply structure from motion (SfM) techniques (Hartley, Zisserman, 2004, Schönberger, Frahm, 2016) to images annotated with geographic information in order to recover georeferenced scene structure and camera locations. For example, (Li et al., 2012) perform landmark recognition on large, georegistered 3D point clouds and estimate pose from the matches established among image features and 3D points. This is the most relevant work to our proposed approach, with three major differences being that our approach i) can operate in environments whose appearance can change drastically over time, ii) does not need to continuously track visual features and iii) is significantly cheaper in terms of computational cost. Georegistration of point clouds obtained by SfM is achieved by ground surveying a small set of points which are either clearly visible in a scene and its corresponding point cloud or correspond to high-contrast artificial targets introduced to a scene before image acquisition (Westoby et al., 2012).

Despite the promising results obtained with the use of geotagged objects, this approach is limited by the need to cope with voluminous amounts of data and the fact that the available geo-tagged images are clustered in urban areas with the vast majority of the Earth's surface having no such coverage. For this reason, more recent research explores cross-view image matching (Regmi, Shah, 2019, Lin et al., 2013), and attempts to directly match a ground image with aerial images. This matching has to overcome the dramatic variation in image appearance that is caused by the large disparity in viewpoints.

3. BACKGROUND

The following subsections provide brief overviews of key concepts that are of central importance for the development of the proposed method in Section 4.

3.1 Dynamic time warping

Dynamic time warping (DTW) (Sakoe, Chiba, 1978) is a technique for finding an optimal alignment between two time dependent sequences under certain constraints. An alignment is a warping that maps one time series onto another in order to facilitate their similarity comparison. DTW can be efficiently computed by using the dynamic programming paradigm, which is a general method for reducing the running time of algorithms exhibiting the properties of overlapping subproblems and optimal substructure. In our case, the GNSS position sequences are timestamped and DTW is used to synchronize them using the absolute difference of timestamps as the local cost measure. The timestamps are generated by Raspberry Pi computers attached to the employed GNSS receivers, whose clocks are assumed to be approximately synchronized with an external time server using the Network Time Protocol (NTP) (Mills, 1991).

3.2 Camera Pose Estimation

Camera pose estimation concerns the determination of a camera's position and orientation with respect to its environment given the camera intrinsic parameters and a set of correspondences between 3D features and their image projections (Hartley, Zisserman, 2004). In the photogrammetry literature, the problem is also known as exterior orientation (Grussenmeyer, Al Khalil, 2002). When the corresponding features are n pairs of points, the problem is often referred to as the Perspective-n-Point (PnP) problem. PnP is typically solved using non-iterative approaches that involve small, fixed-size sets of correspondences. For example, the basic case for triplets ($n = 3$, hence known as the P3P problem), has been studied in the nineteenth century (Grunert, 1841). P3P is known to admit up to four different solutions, whereas in practice it usually has just two. Many other solutions to PnP have been proposed over the years (Gao et al., 2003) and the problem continues to attract interest to the present day, e.g. (Hesch, Roumeliotis, 2011,Zheng et al., 2013,Nakano, 2015,Lourakis, Terzakis, 2020).

3.3 Absolute Orientation

Absolute orientation is the problem of determining the rigid transformation (i.e., rotation followed by translation) aligning two sets of corresponding 3D points. This problem manifests itself when transforming points between coordinate systems and arises often in computer vision, graphics, photogrammetry and robotics. For three points in general position, absolute orientation has a unique solution. For more than three points, a least squares problem minimizing the mean squared residual error is formed. By eliminating the impact of translation, Horn showed in (Horn, 1987) that the rotation minimizing that error corresponds to a unit quaternion which is the eigenvector associated with the largest eigenvalue of a symmetric 4×4 matrix. Improved handling of marginal cases was latter provided by Umeyama (Umeyama, 1991). Several efficient algorithms for dealing with absolute orientation are proposed and compared for their computational cost in (Lourakis, Terzakis, 2018).

3.4 Multi-band GNSS and RTK

This work employed several low-cost GNSS receiver modules, specifically the ZED-F9P from u-blox (U-blox, 2020). The ZED-F9P is a compact, high precision, high update rate positioning receiver that provides multi-band GNSS to high-volume geomatics applications. It integrates multi-band RTK technology for centimetre-accurate 3D positioning. The receiver chipset features a 184-channels engine and is capable for tracking concurrently all available GNSS constellations such as GPS, GLONASS, Galileo and BeiDou. The ability to receive multiple frequencies from multiple constellations results in improved error resolution and eventually more accurate positioning. Each of the ZED-F9P receivers was connected to a Raspberry Pi single board computer (Upton, Halfacree, 2016) in order to facilitate its configuration according to the application demands.

For high precision position estimation, the RTK technique was employed. Using a fixed base station and a mobile rover, RTK reduces the rover's position error by transmitting real-time corrections from the base to the rover. A data format designed to support RTK operation is RTCM (O'Keefe, Lachapelle, 2007). Format versions 3.0 or later have significantly reduced network bandwidth demands, a feature that is particularly attractive both in terms of preserving bandwidth and reducing costs when operating over mobile IP networks like GSM, GPRS or UMTS. NTRIP (Networked Transport of RTCM via Internet Protocol) is a HTTP-based, application level protocol for streaming RTCM data (Weber et al., 2005). The present study used a NtripCaster server, set up and operated by the Aristotle University of Thessaloniki (Fotiou et al., 2009). The server implements NTRIP and streams RTCM data in various versions over the Internet, using a network of permanent GNSS stations covering a large part of Greece that includes the areas of interest to this study. The ZED-F9P has built-in support for standard RTCM corrections, from either a local base station or virtual reference stations (VRS) in a network setup. In our case, RTCM data were obtained via NTRIP from the nearest GNSS base station which is equipped with a high quality geodetic receiver and antenna. Thus, each position measurement derived by the ZED-F9P receiver, delivers centimetre-level accuracy by combining GNSS signals from multiple frequency bands (i.e., L1/L2/L5).

3.5 Coordinate System

Our georeferencing employs a projected coordinate system, defined by the Hellenic Geodetic Reference System 1987 (HGRS87 for short, or ΕΓΣΑ87 in Greek). HGRS87 specifies a local geodetic datum and a projection (Veis, Paradissis, 1990). The HGRS87 datum is implemented by a first order geodetic network, consisting of several tens of triangulation stations throughout Greece. HGRS87 uses the GRS80 ellipsoid (National Imagery and Mapping Agency, 2000) with the axes origin shifted relative to the GRS80 geocenter, so that the ellipsoidal surface is best for Greece. The HGRS87 constitutes the official datum of the Hellenic Cadastre and is widely used for most civilian applications. HGRS87 also specifies the TM87 projection system, a transverse Mercator cartographic projection covering six degrees of longitude on either side of the central meridian at 24 degrees east (18-30 degrees east). In this manner, the entire Greek territory (stretching to approximately 9° of longitude) is projected in one zone. The corresponding coordinates (E, N) are in meters and rely on TM87. Northings are measured from the equator. A false easting of 500000 m is assigned to the central meridian (24° east), so that eastings (E) are always positive.

4. THE PROPOSED METHOD

We begin by developing the formulas that define how the pose of a camera at a certain location transforms when the platform on which it is mounted moves. Let $\mathbf{R}_c$, $\mathbf{t}_c$ be a 3×3 rotation matrix and a 3×1 translation vector defining the camera pose at some reference location. This means that a point $\mathbf{M} \in \mathbb{R}^3$ is transformed to the camera coordinate frame by

$$\mathbf{M}_c = \mathbf{R}_c \, \mathbf{M} + \mathbf{t}_c \tag{1}$$

Assume next that the camera moves to a new location with rotation $\mathbf{R}$ and translation $\mathbf{t}$. To calculate the camera pose at this new location, we can equivalently assume that the camera is stationary and its surroundings move with the reverse motion, i.e. $\mathbf{M}$ moves to $\mathbf{R}^{\mathsf{T}}(\mathbf{M}-\mathbf{t})$. Thus, substitution in eq. (1) yields

$$\mathbf{M}_c = \mathbf{R}_c \, \mathbf{R}^{\mathsf{T}}(\mathbf{M} - \mathbf{t}) + \mathbf{t}_c, \tag{2}$$

from which the camera pose in the new location becomes

$$\mathbf{R}_c\mathbf{R}^{\mathsf{T}} \;\text{ and }\; \mathbf{t}_c - \mathbf{R}_c\mathbf{R}^{\mathsf{T}}\mathbf{t} \tag{3}$$

For future use, the optical center (i.e., center of projection) $\mathbf{C}$ of a pinhole camera can be computed by setting the left side of eq. (1) to zero and solving for $\mathbf{M}$, i.e.

$$\mathbf{0} = \mathbf{R}\,\mathbf{C} + \mathbf{t} \;\Leftrightarrow\; \mathbf{C} = -\mathbf{R}^{\mathsf{T}}\mathbf{t} \tag{4}$$

The rest of the section describes in more detail how a reference camera pose is estimated from visual ques and then updated by integrating GNSS location measurements as the camera platform moves. These two steps respectively provide the motions $\mathbf{R}_c$, $\mathbf{t}_c$ and $\mathbf{R}$, $\mathbf{t}$ appearing in eqs. (1) and (2). The camera pose at a certain location (cf. eq. (1)) is obtained via pose estimation from a number of ground control points (GCPs). Specifically, our approach to pose estimation from a single image uses a set of 3D-2D point correspondences to compute a preliminary pose estimate and then refine it iteratively. This is achieved by embedding the PnP solver of (Lourakis, Terzakis, 2020) into a RANSAC stochastic sampling framework (Fischler, Bolles, 1981) and using random triplets to compute an initial pose estimate along with a classification of correspondences into inliers and outliers. The pose corresponding to the maximal consensus set computed by RANSAC is next refined to take into account all inlying correspondences by using the Levenberg-Marquardt algorithm to minimize a non-linear cost function corresponding to the cumulative reprojection error. This minimization is made more immune to noise caused by mislocalized image points by substituting the squared distance (L_2 norm) of the reprojection errors with a robust cost function (i.e., M-estimator). Our pose estimation approach is presented in more detail in (Lourakis, Zabulis, 2013b) and is implemented by (Lourakis, 2014).

The motion of the camera to a new location equals the relative motion between the corresponding positions of the u-blox receivers, which is estimated by solving the absolute orientation problem (cf. Sec. 3.3). Using algorithms such as those in (Lourakis, Terzakis, 2018), this problem can be solved with very small computational overhead, even on modest hardware.

5. EXPERIMENTS

This section reports experimental results from the deployment of the proposed method to two different moving platforms. Before going into details, we briefly explain the difference between fixed and float RTK solutions (Jensen, Cannon, 2000). RTK GNSS typically returns results of three different types, which in order of increasing accuracy are i) autonomous, ii) RTK float and iii) RTK fixed. Autonomous means that the mobile rover is not receiving corrections from the base station, due to problems related to the base station, communications link, or distance. RTK float indicates that while the rover is receiving corrections from the base station, these are not sufficient for accurate carrier phase ambiguity resolution due to a low number of visible satellites, poor satellite constellation geometry or large distance to base station. RTK fixed means that the mobile rover is receiving corrections from the base station which are based on sufficient satellites received in common. The accuracy of float solutions is in the range of several decimetres and of fixed solutions a few centimetres.

The first experiment aims at verifying the accuracy of the u-blox receivers and their sufficiency for estimating a moving platform's pose. Towards this end, the antennas of three u-blox ZED-F9P receivers were magnetically mounted on a car. The car was then driven in the vicinity of FORTH's premises with a speed of up to 60km/h along a closed, 5.5 kilometers long route. A sequence of position measurements was logged from each of the u-blox receivers with a frequency of 2Hz. Each measurement was timestamped with the number of milliseconds elapsed since the Unix epoch. Less than 1.4% of the total measurements were obtained in the RTK float mode (primarily due to tree foliage) and the remainder were in RTK fixed mode.

Figure 1 illustrates the positions calculated by one of the u-blox superimposed on a street map. As can be seen, the estimated path is qualitatively correct, as it aligns well with the road network. To examine further the accuracy of position measurements, we synchronized the position sequences as explained in Sec. 3.1 and used corresponding triplets from all u-blox receivers to estimate the poses of the car with respect to the route starting point. The estimated poses are illustrated in Figure 2 using right-handed triads of mutually orthogonal vectors; red vectors pointing sideways correspond to the x-axis, green vectors pointing forward (and to the left of the x-axis) correspond to the y-axis and blue vectors pointing upwards correspond to the z-axis. Clearly, the orientation of the vectors along the route are consistent with that of the car.

In a second experiment, the three u-blox receivers were installed around the operator's cabin of a wheeled container quay crane at the commercial port of Heraklion (see Fig. 3 (a)). The u-blox receivers were at an approximate height of 19m from which they had open views of the sky apart from areas obstructed by the crane's mast and boom. Five full revolutions of the crane were performed during which readings from each u-blox were timestamped and logged every 0.5s. Post processing and plotting of the recorded positions revealed that RTK float measurements amounted to 36% of the total. Due to the COVID-19 lock-down and the resulting inaccessibility to the site, we have been unable to perform further tests in order to determine whether this increased occurrence of float measurements was due to transient causes (e.g., radio interference, visible satellites alignment, atmospheric conditions) or due to interference by the crane's metallic structure (e.g. multipath propagation). Nevertheless, we have been able to rectify the float measurements as follows. We observed that at least one of the measurements of each triplet was in RTK fixed mode, meaning that it is reliable. Since the crane performs a planar rotation, the u-blox trajectories form concentric circles which can be estimated analytically by fitting circles to the RTK fixed measurements of each u-blox. Then, by noticing that the relative angles formed by the common circles center and each u-blox are constant, we estimated them using a median filter. Finally, given one fixed u-blox location, any RTK float locations in the remaining two of a triplet are calculated with simple trigonometry from the estimated relative angles.

In addition to the u-blox receivers, a FLIR Blackfly GigE camera was also installed to the crane's cabin, pointing forward and being tilted with respect to the horizon. It is noted that a large part of the crane's surroundings consists of sea water or a vessel being served whose appearance cannot be anticipated in advance. Hence, geo-tagged approaches such as (Li et al., 2012) are not applicable in this setting. The employed camera had a resolution of 2048×1536 pixels and was equipped with a 6mm lens. The camera was intrinsically calibrated using Zhang's planar checkerboard method (Zhang, 2000). The quay has grids painted on its surface to assist the crane operator in aligning the containers in rows. The junctions of these grids are clearly discernible in images and hence suitable to serve as GCPs (see Fig. 3 (c)). We surveyed 30 of these junctions in the HGRS87 projected coordinate system using a pole-mounted Leica Viva GS08plus RTK GNSS receiver. The image projections of the GCPs are obtained manually via mouse clicking. This process is accelerated by first identifying the four extremal grid junctions in an image, estimating the world to image plane homography (Liebowitz, Zisserman, 1998) with them, and eventually using the estimated homography to project all grid points to the image. In a last step, the projected grid points were manually adjusted to their correct image locations.

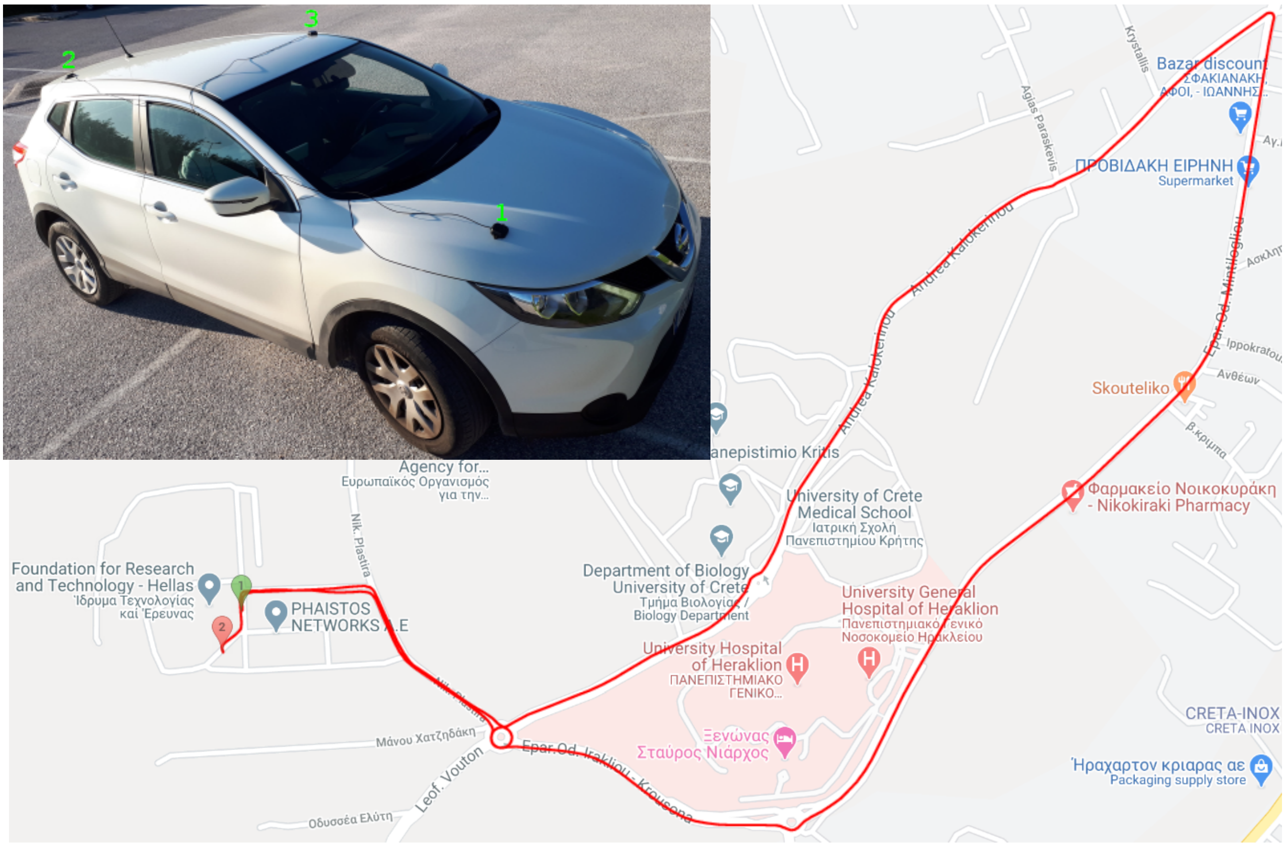


Figure 1. Closed route traversed in a counterclockwise direction overlaid on a Google street map. Start and end points are indicated with green and red pins, respectively. The inset image shows the antennas of the three u-blox receivers attached on the car used.

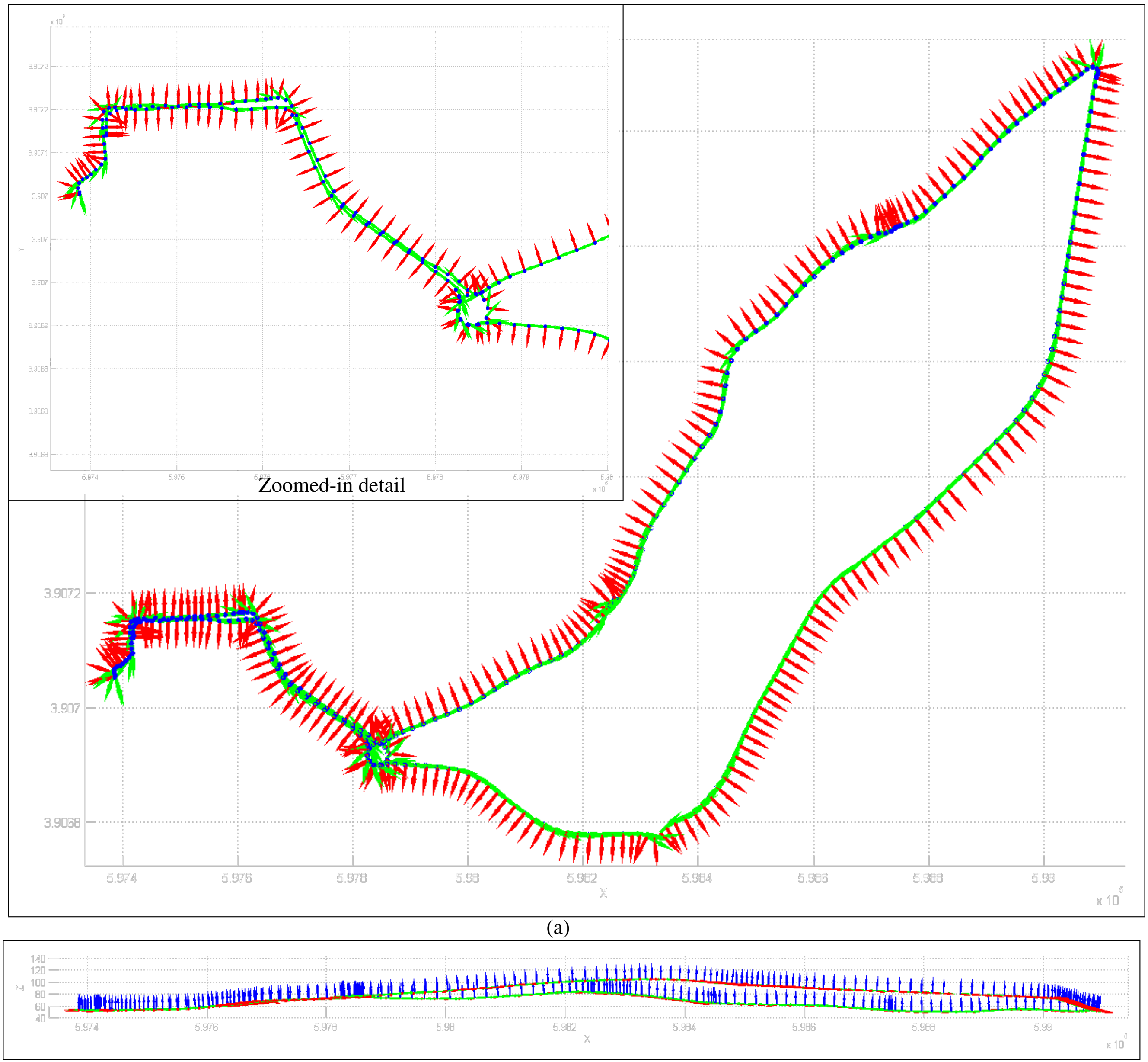
The surveyed HGRS87 coordinates of the junction GCPs are expressed in meters and in our case were of the form (6042XX, 39115XX, 0). It is well-known that when performing geometric computations that involve coordinates whose absolute values are far from 1, it is recommended to improve conditioning via suitable transformations, e.g. see (Förstner, Wrobel, 2016). In our case, we subtracted (60000, 3900000, 0) from all GCP coordinates, effectively transforming them to have the former as their origin; u-blox positions were also translated accordingly.

At the beginning and end of each of the five crane revolutions, an image of the quay with the GCPs visible was acquired. Camera poses were then estimated with (Lourakis, 2014) using the translated 3D coordinates of the GCPs (all lying on the $z = 0$ plane) and their image projections, after distortion removal. The average reprojection error corresponding to the inliers for the estimated poses was in the order of 1.5–2.5 pixels. Using the image before the first revolution as reference, we combined its pose and the crane platform's displacements with eq. (3) to transform it to the poses of the remaining nine crane locations. To compare the transformed rotations with those estimated from the GCPs, we used the metric $\arccos((\mathrm{trace}(\mathbf{R}_2^T \mathbf{R}_1) - 1)/2)$ that corresponds to the angle of rotation about a unit vector that transfers $\mathbf{R}_2$ to $\mathbf{R}_1$ (Huynh, 2009). For translations, we used the Euclidean distance of camera centers computed with eq. (4). We found that the average difference in rotation angles was $0.49°$ (SD $0.28°$) and the average distance of camera centers 0.31m (SD 0.15m). Note that these differences incorporate several sources of error such as calibration, pose estimate and GNSS position inaccuracies, numerical round-off errors, etc.

Figures 3 (b) and (c) illustrate an example of using the proposed method when the GCPs are not visible in the new location. The image in Fig. 3 (c) corresponds to a reference view from which a reference pose is estimated with the projections of the shown GCPs. Figure 3 (b) was obtained after the crane performed a counterclockwise rotation of about $120°$. Using the proposed method, the camera pose for the image in Fig. 3 (b) was calculated from the movement of the u-blox receivers. Measurements from all three u-blox receivers were used, however due to the platform performing a 2D rotation only, two or even a single one would suffice. The frustums of the camera for the two images were computed from their poses and are projected on the quay's plane in Fig. 3 (a). To quantitatively assess the accuracy of the pose estimated for Fig. 3 (b), the corners of the five nearest concrete slabs on the quay's surface were surveyed and the camera pose was estimated using them as GCPs. The pose estimate obtained in this manner was then compared with that computed with the proposed method. The difference in rotations using the aforementioned metric was $1°$ whereas the distance between the camera centers was 0.35m.

6. CONCLUSIONS

This paper has presented a technique for continuously estimating the full 6D pose of a pinhole camera mounted on a mobile platform. It overcomes the need for incessant tracking of visual features by estimating the camera pose at a reference location only and then incorporating position measurements from three low-cost, multi-GNSS receivers attached on the platform to update the camera pose as the platform moves. No special fixtures or arrangements are required for the camera or the GNSS receivers. Experiments with two different setups have demonstrated the effectiveness of the proposed solution.



(a)

(b)

Figure 2. Moving trihedrals for the 3D poses estimated for the route of Figure 1. A top view of the x, y plane is in (a) and a side view towards the positive y-axis, demonstrating altitude differences, in (b). The inset image in (a) shows its lower left part in more detail.

ACKNOWLEDGEMENTS

This work has been partially supported by the European Union's Horizon 2020 Research and Innovation programme under Grant Agreement No. 826506 (EU project sustAGE, `https://www.sustage.eu`). We thank Evangelos Loutsetis for his support with data acquisition and the Heraklion Port Authority for granting us quay and crane access.

REFERENCES

Cavegn, S., Blaser, S., Nebiker, S., Haala, N., 2018. Robust and accurate image-based georeferencing exploiting relative orientation constraints. *ISPRS Ann. Photogramm. Remote Sens. Spatial Inf. Sci.*, IV-2, 57–64.

Cramer, M., Stallmann, D., Haala, N., 2001. Direct Georeferencing Using GPS/Inertial Exterior Orientations for Photogrammetric Applications. *International Archives of Photogrammetry and Remote Sensing*, 33.

Fischler, M. A., Bolles, R. C., 1981. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Commun. of the ACM*, 24(6), 381–395.

Förstner, W., Wrobel, B. P., 2016. *Photogrammetric Computer Vision: Statistics, Geometry, Orientation and Reconstruction*. Springer International Publishing, Cham.

Fotiou, A., Pikridas, C., 2012. *GPS and Geodetic Applications*. Second edn, Ziti Editions, ISBN: 978-960-456-346-3.

Fotiou, A., Pikridas, C., Rossikopoulos, D., Spatalas, S., Tsioukas, V., Katsougiannopoulos, S., 2009. The Hermes GNSS NtripCaster of AUTh. *Bulletin of Geodesy and Geophysics*.

Gabrlik, P., 2015. The Use of Direct Georeferencing in Aerial Photogrammetry with Micro UAV. *IFAC-PapersOnLine*, 48(4), 380–385. 13th IFAC and IEEE Conference on Programmable Devices and Embedded Systems.

Gao, X.-S., Hou, X.-R., Tang, J., Cheng, H.-F., 2003. Complete solution classification for the perspective-three-point problem.

IEEE Transactions on Pattern Analysis and Machine Intelligence, 25(8), 930–943.

Grunert, J., 1841. Das pothenotische Problem in erweiterter Gestalt nebst über seine Anwendungen in Geodäsie. *Grunerts Archiv für Mathematik und Physik*.

Grussenmeyer, P., Al Khalil, O., 2002. Solutions for Exterior Orientation in Photogrammetry: A Review. *The Photogrammetric Record*, 17(100), 615–634.

Hackeloeer, A., Klasing, K., Krisp, J. M., Meng, L., 2014. Georeferencing: a review of methods and applications. *Annals of GIS*, 20(1), 61–69.

Hartley, R., Zisserman, A., 2004. *Multiple View Geometry in Computer Vision*. Second edn, Cambridge University Press, ISBN: 0521540518.

Hays, J., Efros, A. A., 2008. IM2GPS: estimating geographic information from a single image. *IEEE Conference on Computer Vision and Pattern Recognition*, 1–8.

Hesch, J. A., Roumeliotis, S. I., 2011. A direct least-squares (DLS) method for PnP. *International Conference on Computer Vision*, IEEE, 383–390.

Horn, B. K. P., 1987. Closed-form Solution of Absolute Orientation Using Unit Quaternions. *J. Optical Society of America A*, 4(4), 629–642.

Hourdakis, E., Lourakis, M., 2015. Countering drift in visual odometry for planetary rovers by registering boulders in ground and orbital images. *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 111–116.

Huynh, D. Q., 2009. Metrics for 3D Rotations: Comparison and Analysis. *J. Math. Imaging Vis.*, 35(2), 155–164.

Jende, P., Nex, F., Gerke, M., Vosselman, G., 2017. Fully automatic feature-based registration of mobile mapping and aerial nadir images for enabling the adjustment of mobile platform locations in GNSS-denied urban environments. *Proceedings of ISPRS Hannover Workshop: HRIGI 17 – CMRT 17 – ISA 17 – EuroCOW 17*, ISPRS Archives, 317–323.

Jensen, A., Cannon, M., 2000. Performance of network RTK using fixed and float ambiguities. *Proceedings of the National Technical Meeting of the Institute of Navigation*, Institute of Navigation, 797–805.

Li, Y., Snavely, N., Huttenlocher, D., Fua, P., 2012. Worldwide pose estimation using 3D point clouds. *European Conference on Computer Vision*, I, Springer-Verlag, Berlin, Heidelberg, 15–29.

Liebowitz, D., Zisserman, A., 1998. Metric rectification for perspective images of planes. *IEEE Conference on Computer Vision and Pattern Recognition*, 482–488.

Lin, T.-Y., Belongie, S., Hays, J., 2013. Cross-view image geolocalization. *IEEE Conference on Computer Vision and Pattern Recognition*, 891–898.

Lourakis, M., 2014. posest : A C/C++ library for robust 6DoF pose estimation from 3D-2D correspondences. (8 Apr. 2020).

Lourakis, M., Terzakis, G., 2018. Efficient absolute orientation revisited. *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 5813–5818.

Lourakis, M., Terzakis, G., 2020. A globally optimal method for the PnP problem with MRP rotation parameterization. Submitted to the International Conference on Pattern Recognition.

Lourakis, M., Zabulis, X., 2013a. Accurate scale factor estimation in 3D reconstruction. *Computer Analysis of Images and Patterns*, Springer Berlin Heidelberg, 498–506.

Lourakis, M., Zabulis, X., 2013b. Model-based pose estimation for rigid objects. *International Conference on Computer Vision Systems*, Springer, 83–92.

Mills, D. L., 1991. Internet time synchronization: the network time protocol. *IEEE Transactions on Communications*, 39(10), 1482-1493.

Nakano, G., 2015. Globally optimal DLS method for PnP problem with Cayley parameterization. *British Machine Vision Conference*, 78.1–78.11.

National Imagery and Mapping Agency, 2000. Department of defense world geodetic system 1984: its definition and relationships with local geodetic systems. Technical Report TR8350.2, NIMA, St. Louis, MO, USA.

Nebiker, S., Cavegn, S., Eugster, H., Laemmer, K., Markram, J., Wagner, R., 2012. Fusion of airborne and terrestrial image-based 3D modelling for road infrastructure management - vision and first experiments. *Proceedings of Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, 79–84.

O'Keefe, K., Lachapelle, G., 2007. Network Real-Time Kinematic Performance Analysis Using RTCM 3.0 and the Southern Alberta Network. *GEOMATICA*, 61(1), 29–41.

Pateraki, M., 2005. Adaptive multi-image matching algorithm for DSM generation from airborne linear array CCD data. PhD thesis, Diss. ETH Zurich, Nr. 15915, Institute of Geodesy and Photogrammetry, Mitteilung No. 86.

Regmi, K., Shah, M., 2019. Bridging the domain gap for ground-to-aerial image matching. *IEEE/CVF International Conference on Computer Vision*, 470–479.

Rietdorf, A., Daub, C., Loef, P., 2006. Precise positioning in real-time using navigation satellites and telecommunication. *Proceedings of the 3rd Workshop on Positioning, Navigation and Communication*.

Sakoe, H., Chiba, S., 1978. Dynamic programming algorithm optimization for spoken word recognition. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 26(1), 43–49.

Schönberger, J. L., Frahm, J.-M., 2016. Structure-from-motion revisited. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 4104–4113.

Shan, Q., Wu, C., Curless, B., Furukawa, Y., Hernandez, C., Seitz, S. M., 2014. Accurate geo-registration by ground-to-aerial image matching. *International Conference on 3D Vision*, 1, 525–532.

Snavely, N., Seitz, S. M., Szeliski, R., 2008. Modeling the World from Internet Photo Collections. *International Journal of Computer Vision*, 80(2), 189–210.

U-blox, 2020. ZED-F9P module. `https://www.u-blox.com/en/product/zed-f9p-module`. (8 Apr. 2020).

Umeyama, S., 1991. Least-squares Estimation of Transformation Parameters Between Two Point Patterns. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13(4), 376–380.

Upton, E., Halfacree, G., 2016. *Raspberry Pi User Guide*. 4th edn, Wiley Publishing.

Veis, G., Paradissis, D., 1990. GPS activities for geodesy and geodynamics in Greece. *2nd GPS Conference*, Ottawa.

Weber, G., Dettmering, D., Gebhard, H., 2005. Networked transport of RTCM via Internet Protocol (NTRIP). F. Sansò (ed.), *A Window on the Future of Geodesy*, Springer Berlin Heidelberg, 60–64.

Westoby, M., Brasington, J., Glasser, N., Hambrey, M., Reynolds, J., 2012. 'Structure-from-Motion' photogrammetry: A low-cost, effective tool for geoscience applications. *Geomorphology*, 179, 300–314.

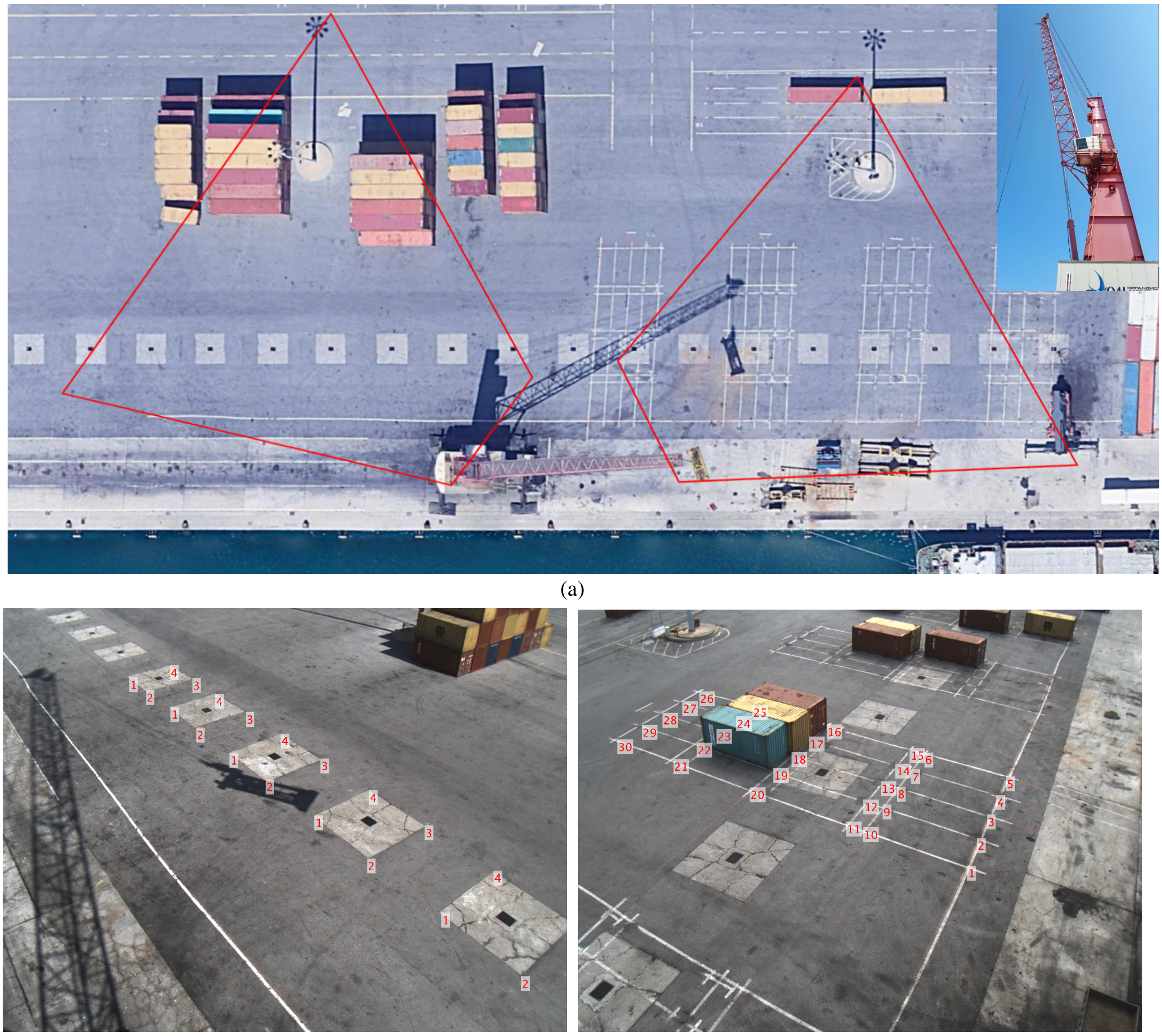
Wildes, R. P., Hirvonen, D. J., Hsu, S. C., Kumar, R., Lehman, W. B., Matei, B., Zhao, W. ., 2001. Video georegistration: algorithm and quantitative evaluation. *IEEE International Conference on Computer Vision*, 2, 343–350.

Zamir, A. R., Hakeem, A., Van Gool, L., Shah, M., Szeliski, R., 2016. *Introduction to Large-Scale Visual Geo-localization*. Springer International Publishing, Cham, 1–18.

Zhang, W., Kosecka, J., 2006. Image based localization in urban environments. *Third Int'l Symposium on 3D Data Processing, Visualization, and Transmission (3DPVT'06)*, 33–40.

Zhang, Z., 2000. A Flexible New Technique for Camera Calibration. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(11), 1330–1334.

Zheng, Y., Kuang, Y., Sugimoto, S., Åström, K., Okutomi, M., 2013. Revisiting the PnP problem: A fast, general and optimal solution. *Proceedings of the IEEE International Conference on Computer Vision*, 2344–2351.



(a)

(b)

(c)

Figure 3. (a) Non-orthorectified satellite view of the port quay with the crane in the bottom middle and the inset image. The two red quadrangles correspond to the intersections of the quay's plane with the camera frustums of the images in (b) and (c) *[the position of the crane shown in image (a) does not coincide with that from which images (b), (c) were taken]*. (b), (c) Views of the quay from the camera installed on the crane. The image in (c) was obtained from the reference location and has the employed GCPs superimposed. The four extremal junctions used for the world to image homography estimation are GCPs 1, 5, 26 and 30. Image (b) is a view from the camera after the crane has rotated about 120° in a counterclockwise direction. The GCPs shown are used for pose verification.