

# PHOTOGRAMMETRIC RECONSTRUCTION WITH BAYESIAN INFORMATION

A. Masiero<sup>a</sup>\*, F. Fissore<sup>a</sup>, A. Guarnieri<sup>a</sup>, F. Pirotti<sup>a</sup>, A. Vettore<sup>a</sup>

<sup>a</sup> Interdepartmental Research Center of Geomatics (CIRGEO), University of Padova,  
Viale dell'Università 16, Legnaro (PD) 35020, Italy -  
masiero@dei.unipd.it  
(francesca.fissore, alberto.guarnieri, francesco.pirotti, antonio.vettore)@unipd.it

**KEY WORDS:** 3D reconstruction, Structure from Motion, Feature matching, Inertial Sensors, UWB

## ABSTRACT:

Nowadays photogrammetry and laser scanning methods are the most wide spread surveying techniques. Laser scanning methods usually allow to obtain more accurate results with respect to photogrammetry, but their use have some issues, e.g. related to the high cost of the instrumentation and the typical need of high qualified personnel to acquire experimental data on the field. Differently, photogrammetric reconstruction can be achieved by means of low cost devices and by persons without specific training. Furthermore, the recent diffusion of smart devices (e.g. smartphones) embedded with imaging and positioning sensors (i.e. standard camera, GNSS receiver, inertial measurement unit) is opening the possibility of integrating more information in the photogrammetric reconstruction procedure, in order to increase its computational efficiency, its robustness and accuracy. In accordance with the above observations, this paper examines and validates new possibilities for the integration of information provided by the inertial measurement unit (IMU) into the photogrammetric reconstruction procedure, and, to be more specific, into the procedure for solving the feature matching and the bundle adjustment problems.

## 1. INTRODUCTION

Most of the Mobile Mapping Systems (MMSs) developed in the last twenty years consider the use of remote sensing/imaging devices mounted on quite expensive terrestrial or airborne vehicles.

However, thanks to the diffusion of a large quantity of devices embedded with several small sensors (typically based on MEMS), the development of small and relatively cheap mobile mapping systems have been taken in consideration in recent years. For instance, smartphones have good potential in this sense thanks to the concentration of several sensors in easily portable and widely diffused devices. However, data acquisition by means of smartphones is limited by the restricted mobility of the person carrying the device (movements are typically approximately constrained to a 2D plane/surface). Differently, drones/unmanned aerial vehicles have the advantage of allowing to acquire measurements in a much larger area in the same time lapse, while moving on (almost) all the free space 3D positions, if needed.

The availability of different types of surveying devices allow the development of a wider range of mobile mapping solutions (Chiang et al., 2003, Toth and Grejner-Brzezinska, 1997, Toth, 2001, Pirotti et al., 2014, Kraus and Pfeifer, 1998, Remondino et al., 2011), and, consequently, the availability of more spatial data that can enable the development of even more applications, e.g. Geographic information systems (GIS), location based services, object recognition (El-Sheimy and Schwarz, 1998, Guarnieri et al., 2015, Piragnolo et al., 2015, Pfeifer et al., 2004, Masiero et al., 2015b, Facco et al., 2009, Facco et al., 2010). Thanks to the small size, the cheap cost and the easy usability of cameras, independently of the specific considered device in most of the cases a standard camera is used as imaging sensor, and photogrammetry is used in order to obtain reconstructions of the observed scenes (Prodocimi et al., 2015).

Photogrammetry is one of the most used techniques for 3D reconstruction, monitoring and surveying. It is widely used in several

applications and in different working conditions. The accuracy of photogrammetry reconstruction methods changes depending on the working conditions (e.g. the number of acquired images, lighting conditions, baselines between images), and it is strictly related to the success of the solution of the Structure from Motion (SfM) problem (Hartley and Zisserman, 2003, Ma et al., 2003).

Despite its diffused use and the ever growing improvements to the reconstruction technique, photogrammetry still does not reach the same level of reliability of laser scanning surveying techniques (which can be considered the current state of the art for 3D reconstructions, (Remondino et al., 2005)) and the same level of applications: more specifically, significant issues may occur in photogrammetric reconstructions when in presence of lighting problems or when the object of interest is not sufficiently textured. However, photogrammetry has several advantages with respect to laser scanning techniques: in particular, it relies on the use of much cheaper devices, surveying is usually faster and it can be performed by less trained personnel with respect to terrestrial laser scanning.

The recent development of positioning systems based on the integration of information provided by different sensors allows to reduce the issues related to the use of GNSS as positioning approach (Lukianto and Sternberg, 2011, Masiero et al., 2014a, Saeedi et al., 2014, Toth et al., 2015): thanks to the use of inertial navigation systems, radio signals (WiFi and Bluetooth communications, UWB sensors) and (if available) a priori geometric constraints on the environment, it is possible to obtain reliable and quite accurate estimates of the position and attitude of the mobile mapping device, where the level of estimation accuracy depends on the specific considered system and the working conditions (e.g. the environment characteristics).

This paper considers the use of position and orientation information (provided by the above mentioned navigation systems) in order to improve the computational efficiency and the robustness of the photogrammetric reconstruction algorithm. Three changes

\*Corresponding author.

with respect to the standard photogrammetric reconstruction algorithm are considered:

- First, exploiting the device position and attitude information, an approximate essential matrix is computed. As shown in section 3., this allows to reduce the number of incorrect feature matches by imposing (approximate) epipolar constraints.
- Then, more accurate epipolar constraints are obtained by using the RANdom SAmple Consensus (RANSAC (Fischler and Bolles, 1981)) algorithm: the efficiency and robustness of the latter are improved thanks to the use of a two-step learning procedure, where a reduced complexity RANSAC is executed first, while (similarly to the locally optimized RANSAC (Chum et al., 2003)) the full RANSAC runs only on the inliers of the estimated reduced complexity model (4.). Differently from (Chum et al., 2003) the reduced complexity step is enabled by the use of information provided by the navigation system.
- Finally, in section 5. a modified bundle adjustment optimization problem is considered in order to improve the reconstruction results in the panoramic-like image acquisition procedure recently considered in (Piermattei et al., 2015).

It is worth to notice that in all the paper the imaging camera is assumed to be calibrated (i.e. the values of the interior calibration parameters are assumed to be known) (Heikkila and Silven, 1997, Habib and Morgan, 2003, Remondino and Fraser, 2006, Karel and Pfeifer, 2009, Balletti et al., 2014).

## 2. REVIEW OF THE 3D RECONSTRUCTION PROCEDURE

Photogrammetry aims at reconstructing the 3D structure of a scene by means of a certain number of photos (i.e. used as remote sensing measurements) taken by different points of view. Each of such photos can be approximately considered a prospective projection (on the corresponding image plane) of what visible from that point of view. Then, the rationale of photogrammetric reconstruction is as follows: given the positions of the projections (on at least two image planes) of a 3D point  $q$ , then an estimate  $\hat{q}$  of the 3D coordinates of  $q$  can be obtained by means of triangulation (Masiero and Cenedese, 2012). When the camera positions during image acquisitions are not known, then both camera positions and 3D points of the scene have to be reconstructed from the acquired images.

A more detailed description of the reconstruction procedure can be as follows:

- Feature extraction. A set of feature points is extracted in each image and their corresponding descriptors are computed (Lowe, 1999, Bay et al., 2008, Leutenegger et al., 2011, Alahi et al., 2012). Such points can be considered as the most characteristic of the images.
- Feature matching based on the appearance. Features are initially matched taking into account of their appearance similarity, based on the use of the previously computed descriptor.  $k$ - $d$  tree search algorithms are typically used in order to make this matching step fast and quite reliable (e.g. the best bin first method (Beis and Lowe, 1997)).

- Feature matching based on the estimated geometry. Depending on the specific considered case, a certain number of the matches computed at the previous step can be wrong (due to several causes, e.g. the influence of sensor noise, noise due to the atmospheric turbulence (Beghi et al., 2008), the poor texture in the images (Piermattei et al., 2015, Masiero and Chiuso, 2006), repetitive patterns (Morel and Yu, 2009, Morel and Yu, 2011)). In this step wrongly matched features are detected by taking into account of the geometry of the scene (Hartley and Zisserman, 2003). Indeed, the projections  $m_{j_1}$  and  $m_{j_2}$  (expressed in homogeneous coordinates) of a 3D point into two different image planes ( $j_1$  and  $j_2$ ) have to obey to the *epipolar constraint*, that is usually expressed by means of the fundamental matrix  $F_{j_1 j_2}$ :

$$m_{j_1}^T F_{j_1 j_2} m_{j_2} = 0 \quad (1)$$

Since also the fundamental matrix is usually estimated from the feature matches (by means of the RANSAC (Fischler and Bolles, 1981)) and of the eight-point algorithms (Hartley, 1997, Longuet-Higgins, 1981)), then the epipolar constraint is usually just an approximation of the real relation between the two projections.

- Structure from Motion. Once a set of feature matches have been computed, such matching points are used in order to estimate the 3D positions of the points (and the positions and orientations of the cameras, if necessary). This is the SfM problem, and a solution is usually computed by solving the bundle adjustment (Triggs et al., 1999) by means of an iterative optimization algorithm (Agarwal et al., 2010, Byröd and Aström, 2010). Alternatively, a solution can be obtained by means of the Tomasi and Kanades factorization (Tomasi and Kanade, 1992, Brand, 2002, Masiero et al., 2014b).
- Dense matching. Finally, starting from the set of 3D points computed at the previous step, a dense point cloud is obtained, typically by iterative (local) pixel matching taking into account of the surface regularity (Furukawa and Ponce, 2010, Furukawa et al., 2010).

From the above summary, it is clear that the core of the reconstruction procedure is given by the correct matching of corresponding points in different images. As depicted above, this is usually done by using only the information provided by the camera. However, the diffusion of smart devices (embedded with several sensors) makes realistic of use the information provided by other sensors (e.g. the inertial navigation system (Masiero et al., 2014c, Masiero et al., 2015a)) in order to improve the overall procedure, in terms of both speed and reliability, if possible. Accordingly with this observation, this paper investigates in the next sections the integration of the information provided by the navigation system in certain of the steps of the feature matching procedure and in the bundle adjustment problem.

Nowadays, several feature extractors and descriptors have been considered in the literature. Among this large variety of choices, in this paper the scale-invariant feature transform (SIFT, (Lowe, 1999)), which can be considered as the state of the art in the last decade, has been used. Nevertheless, the proposed approach can be easily adapted in order to work with all the other possible alternatives.

## 3. APPROXIMATE EPIPOLAR CONSTRAINT

Despite the epipolar constraint (1) is expressed by using the fundamental matrix  $F_{j_1 j_2}$ , if the values of the camera interior parameters are known, then a similar constraint can be expressed

by means of the essential matrix:

$$m'_{j_1}{}^T E_{j_1 j_2} m'_{j_2} = 0 \quad (2)$$

where  $E_{j_1 j_2}$  stands for the essential matrix corresponding to the  $j_1$ -th and  $j_2$ -th views, and  $m'_{j_1}$ ,  $m'_{j_2}$  are the normalized homogeneous coordinates of the considered feature point in the two image planes. Even if the camera has not been calibrated a rough estimate of the interior camera parameters can typically be obtained from the operative system of the device (Fusiello and Irsara, 2011). When lens distortion is not negligible, then the normalized coordinates are assumed to have been computed in order to compensate distortion as well (Heikkila and Silven, 1997, Claus and Fitzgibbon, 2005).

For simplicity of exposition, let the coordinate system be positioned in the optical center of the camera during the acquisition from the  $j_2$ -th point of view, and let it have the same orientation of the camera during such acquisition. Furthermore, let  $t_{j_1 j_2}$  and  $R_{j_1 j_2}$  be translation and rotation from the coordinate system of  $j_2$  to that of  $j_1$ , respectively. Then, the essential matrix  $E_{j_1 j_2}$  can be expressed as follows:

$$E_{j_1 j_2} = [t_{j_1 j_2}]_{\times} (-R_{j_1 j_2} t_{j_1 j_2}) \quad (3)$$

where  $[t_{j_1 j_2}]_{\times}$  is the skew-symmetric matrix corresponding to the cross-product operator of  $t_{j_1 j_2}$  i.e.  $[t_{j_1 j_2}]_{\times} x = t_{j_1 j_2} \times x$ , for all the possible values of the vector  $x$ .

In this paper the device considered for image acquisition is assumed to be provided of a navigation system running during the survey. In particular, the device is assumed to be provided of embedded sensors in order to be able to estimate both the position and orientation changes (nowadays smartphones are typically provided of GNSS receiver, and an inertial measurement unit (IMU), e.g. 3-axis accelerometer, gyroscope and magnetometer). Hence, in this paper estimates ( $\hat{t}_{j_1 j_2}$ ,  $\hat{R}_{j_1 j_2}$ ) of the translation vector and of the rotation matrix between two camera views ( $j_1$ ,  $j_2$ ) are assumed to be provided by the navigation system. It is worth to notice that, being the IMU measurements independent of any other external device the estimates  $\hat{t}_{j_1 j_2}$  and  $\hat{R}_{j_1 j_2}$  are available also in challenging environments for the GNSS positioning method.

Exploiting the information provided by the navigation system an estimate  $\hat{E}_{j_1 j_2}$  of the essential matrix can be obtained by simply substituting the values of the estimates in (3). Then, (2) (with the estimated value of the essential matrix  $\hat{E}_{j_1 j_2}$ ) can be used in order to discard wrong matches. Since  $\hat{E}_{j_1 j_2}$  has been obtained by noisy measurements of the translation vector and of the rotation matrix, a not so stringent threshold has to be used in order to discriminate wrong matches, i.e. only a subset of the wrong matches can be detected in this way.

#### 4. ESTIMATION OF THE ESSENTIAL MATRIX

Equation (2), with the essential matrix estimated from the matched features, can be used in order to obtain a more robust wrong feature matching detector. In this case the estimation of the essential matrix can be done by using the RANSAC and (for instance) the eight point algorithm (Hartley, 1997).

More specifically, such estimation procedure can be summarized as follows: let  $n$  be the number of couples of feature matching candidates. Then the algorithm repeats for  $N$  iterations the following procedure:

- randomly sample 8 couples of feature matching candidates among the  $n$  possible ones.
- compute an essential matrix estimate from the 8 sampled couples by using the eight point algorithm.
- compute the fitting error (usually expressed taken into account of the inliers of the current model) of the estimated essential matrix on the  $n$  possible couples.
- if the fitting error of the current estimate of the essential matrix is better than the previous best fitting error, than the current estimate is set as the best estimate of the essential matrix and the current set of inliers is stored.

After  $N$  iterations a more robust estimate of the essential matrix is usually obtained by recomputing it on all the set of inliers of the best obtained estimate.

The number of iterations  $N$  is usually determined taking into account of the probability  $\beta$  of a couple of matching features to be wrong (Chum et al., 2003). To be more specific,  $N$  is computed by assuming that whenever a set of 8 correct couples is drawn, then a good estimate of the essential matrix is obtained (e.g. the inliers of such estimate are (most of all) the correct couples of matching features). Unfortunately, this is typically an optimistic assumption, and hence the RANSAC algorithm has to be iterated for a larger number of iterations with respect to the expected one in order to obtain a reliable estimation.

The rationale of this section is that of exploiting the estimate of the change of orientation provided by the navigation system in order to make the estimation of the essential matrix faster and possibly more robust. The RANSAC-like procedure considered here is a two-step algorithm.

- First, the orientation information provided by the navigation system is used in order to compute  $\hat{R}_{j_1 j_2}$ , and the RANSAC algorithm is used to estimate only  $\hat{t}_{j_1 j_2}$ . Since  $\hat{t}_{j_1 j_2}$  can be linearly estimated by using only 3 couples of candidate matching features, this part of the procedure is much faster with respect to that previously considering the eight point algorithm.
- Similarly to the locally optimized RANSAC (Chum et al., 2003), only when a new best model has been obtained, then a RANSAC with the eight point algorithm is used to estimate the complete essential matrix by using only the inliers of the estimate of the previous step. Since this second step is performed only a small number of times, and the cardinality of the inliers used in this step is smaller with respect to  $n$ , then the overall computational complexity of the algorithm is usually much smaller than that of the original RANSAC.

#### 5. BUNDLE ADJUSTMENT REVISED

A panoramic-like image acquisition procedure has been recently considered in (Piermattei et al., 2015): in such procedure several images are acquired in each of the considered acquisition positions by simply changing the orientation of the camera. To be more specific, in each considered position a sequence of images is typically acquired in to have a panoramic-like view of the scene, as shown for instance in Fig. 1.

On the one hand, as shown in (Piermattei et al., 2015) the panoramic-like image acquisition has the advantage of ensuring a better coverage of the scene. However, on the other hand, since several



Figure 1: Panoramic-like image acquisition of the façade of Villa XXV Aprile, Mirano, Italy.

images acquired from approximately the same position, this kind of acquisition procedure provide some information that can be exploited in the solution of the SfM problem.

The standard bundle adjustment optimization problem used to estimate the parameters of interest is:

$$\{(\hat{t}_j, \hat{R}_j), \forall j, \hat{q}_i, i = 1, \dots, n\} = \arg \min \sum_{ij} \|m_{ij} - \hat{m}_{ij}\|^2 \quad (4)$$

where  $\hat{q}_i$  is the estimated 3D position of the  $i$ -th feature point,  $(\hat{t}_j, \hat{R}_j)$  are the estimated position and orientation matrix of the camera during the  $j$ -th image acquisition,  $m_{ij}$  are the coordinates of the  $i$ -th feature on the image plane during the  $j$ -th image acquisition, and  $\hat{m}_{ij}$  are the corresponding reprojected coordinates accordingly with the current values of the estimation parameters  $\{(\hat{t}_j, \hat{R}_j), \forall j, \hat{q}_i, i = 1, \dots, n\}$ . In order to simplify the notation, conventionally the sum in (4) is assumed to be limited to all (and only) the values of  $i$  and  $j$  for which a corresponding measurement  $m_{ij}$  is available.

In the panoramic-like image acquisition it is possible to group the image indexes accordingly to their acquisition positions: if the  $j$ -th and  $j'$ -th images have been acquired in (approximately) the same position then  $t_j \approx t_{j'}$ , hence  $t_j = t_{j'} + \Delta t_j$ , where  $\Delta t_j$  is defined as the difference between  $t_j$  and  $t_{j'}$ . Let  $\sigma_{\Delta t}$  be the uncertainty on the value of  $\Delta t_j$  (i.e. on the (small) translation between  $t_j$  and  $t_{j'}$ ), and let  $\sigma_{\text{pix}}$  be the feature measurement error, then, accordingly to the prior provided by the navigation system, (4) can be reformulated as follows:

$$\{(\hat{t}_j, \hat{R}_j), \forall j, \hat{q}_i, i = 1, \dots, n\} = \arg \min \sum_{ij} \frac{\|m_{ij} - \hat{m}_{ij}\|^2}{\sigma_{\text{pix}}^2} + \sum \frac{\|\hat{t}_j - \hat{t}_{g(j)}\|}{\sigma_{\Delta t}} \quad (5)$$

where  $g(j)$  is the index of the first image acquired from (approximately) the position  $t_j$ , e.g.  $g(j) = j$  if the  $j$ -th image is the first (or the only) image acquired from  $t_j$ .

As shown in section 6., the prior information about camera positions introduced in the revised bundle adjustment optimization problem (5) allows to reduce the estimation errors on the reconstruction parameters.

## 6. RESULTS AND DISCUSSION

The results shown in this section have been obtained by acquiring measurements (images and data provided by the IMU) with an LG Google Nexus 5 (Fig. 2), which is provided with 3-axis accelerometer, 3-axis gyroscope, 3-axis magnetometer, barometer, GNSS and WiFi receiver, and 8 Mpixel camera. Hence,

the navigation system, similar to that in (Masiero et al., 2014a), has been implemented in the Android environment (Android 4.4 KitKat). Nevertheless, similar results can be obtained by considering different devices and different operative systems. Furthermore, similar results are also expected if using different navigation systems (e.g. other pedestrian navigation systems (Ruiz et al., 2012, Widyawan et al., 2012, Foxlin, 2005)).



Figure 2: LG Google Nexus 5.

As previously assumed the camera embedded in the device has been calibrated by using the OpenCV camera calibration toolbox (Intel Research, 2000), and hence feature point coordinates have been expressed as normalized homogeneous coordinates.

First, the approximate epipolar constraint approach, presented in section 3., for detecting wrong feature matches is investigate in a case study (Villa XXV Aprile, Mirano, Italy). As shown in Fig. 3(a) and (b), 5 wrong matches have been correctly detected. It is worth to notice that this method can hardly detect all the wrong feature matches: this is due to the use of a rough approximate of the essential matrix  $\hat{E}_{j_1 j_2}$ , that in this approach has been directly computed by noisy measurements of the translation vector and of the rotation matrix.

Since the approximate essential matrix  $\hat{E}_{j_1 j_2}$  computed as described in section 3. is usually a quite rough approximation of the real one, then a more accurate estimation should be used for obtaining a more reliable description of the epipolar geometry of the scene, and, consequently, a more trustworthy wrong feature matching detection. To this purpose, the procedure described in section 4. has been applied to the same case study previously considered.

Fig. 4 compares the computational complexity of the standard



(a)



(b)

Figure 3: Wrong feature matches detected by using an approximate epipolar constraint.

RANSAC with the linear (eight point) algorithm for the estimation of the essential matrix with that of the approach proposed in section 4.. It is worth to notice that computational times depend on the specific implementation of the algorithms, on the considered programming language and on the machine used to execute them. The results reported in the figure have been obtained by running the algorithms in Matlab®, on an Intel® Core™ i7-4790 (3.60 GHz).

The values of  $\beta$  in Fig. 4 have been chosen in order to allow a correct estimation of the essential matrix (smaller values of  $\beta$  allows to significantly reduce the computational burden, however the value of the estimated essential matrix is frequently far from the correct one).

Finally, the bundle adjustment formulation for panoramic-like image acquisitions is tested in a synthetic example in order to validate just the bundle adjustment procedure, without having to deal with other external factors. In the simulation considered here 100 feature points randomly distributed are assumed to be visible in the observed scene. Image measurements are assumed to be taken from 3 different positions randomly drawn at a mean distance of 15 m, approximately, from the feature points. Then, to simulate the panoramic-like image acquisition method, in each position 3 images have been acquired changing the camera orientation. In order to make the simulation more statistically robust, 100 Monte Carlo simulations have been considered. Fig. 5 shows the comparison of the root mean square error (RMSE) of the standard bundle adjustment (blue dashed line) with that proposed in section 5. (red solid line) in the 100 independent simulations. The

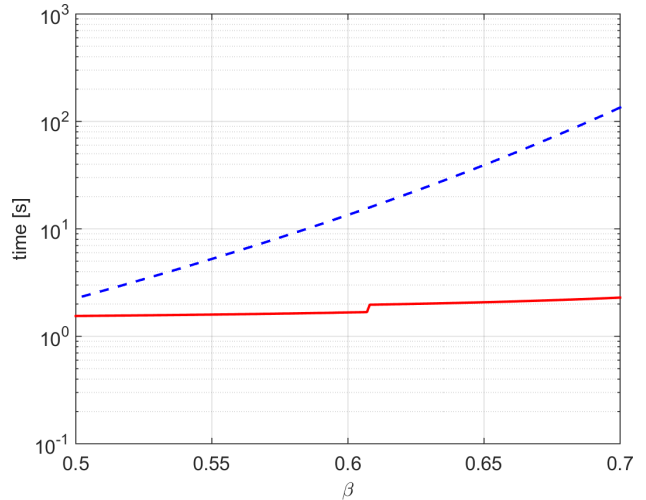


Figure 4: Comparison of the computational burden required by the standard RANSAC for the linear algorithm for the estimation of the essential matrix (blue dashed line) with that of the method proposed in section 4. (red solid line).

proposed method allow to ensure an average 38% reduction of the RMSE in the 100 simulations.

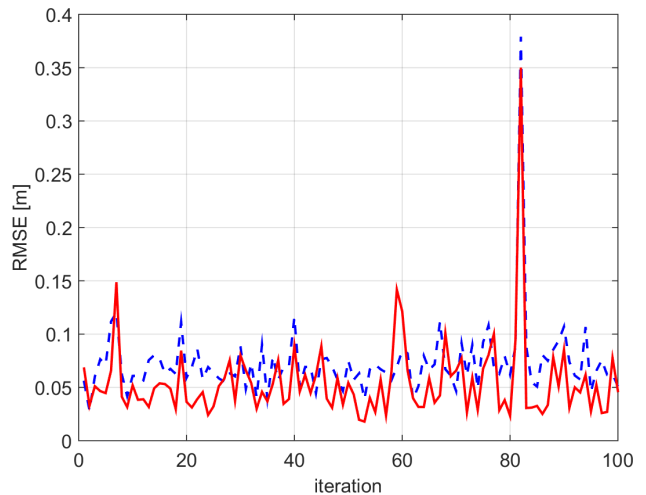


Figure 5: Comparison of the RMSE of the standard bundle adjustment (blue dashed line) with that proposed in section 5. (red solid line). The reported RMS errors have been computed in 100 independent simulations.

In our future works the modified bundle adjustment method proposed in section 5. will be stressfully validated on a large variety of cases, including experimental data of natural environments and human buildings/infrastructures.

Furthermore, the proposed two-step RANSAC procedure will be compared with other recent Bayesian developments of the RANSAC algorithm (Kang, 2015).

## 7. CONCLUSIONS

This paper proposed certain changes with respect to the standard photogrammetric reconstruction procedure in order to take properly advantage of the information provided by the navigation system, that can be included in most of the currently used smart mobile devices.

Three changes have been considered: the use of an approximate epipolar constraint (completely derived by means of the navigation system information), a computationally efficient method to estimate the essential matrix and a slight change of the bundle adjustment optimization function in order to exploit a priori information on camera positions when dealing with panoramic-like image acquisitions.

The proposed methods have been validated on experimental data and on a synthetic case (in the modified bundle adjustment case) showing a significant performance improvement with respect to the classical methods in terms of either error reduction or computational time.

Finally, it is worth to notice that the use of the information provided by the navigation system as a prior for the estimation procedure forces the solution to be relatively close to the prior, ensuring an increase of robustness in the estimation.

## REFERENCES

- Agarwal, S., Snavely, N., Seitz, S. and Szeliski, R., 2010. Bundle adjustment in the large. In: *European Conference on Computer Vision, Lecture Notes in Computer Science*, Vol. 6312, pp. 29–42.
- Alahi, A., Ortiz, R. and Vandergheynst, P., 2012. Freak: Fast retina keypoint. In: *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*, Ieee, pp. 510–517.
- Balletti, C., Guerra, F., Tsioukas, V. and Vernier, P., 2014. Calibration of action cameras for photogrammetric purposes. *Sensors* 14(9), pp. 17471–17490.
- Bay, H., Ess, A., Tuytelaars, T. and Van Gool, L., 2008. Speeded-up robust features (surf). *Computer vision and image understanding* 110(3), pp. 346–359.
- Beghi, A., Cenedese, A. and Masiero, A., 2008. A Markov-random-field-based approach to modeling and prediction of atmospheric turbulence. In: *Control and Automation, 2008 16th Mediterranean Conference on*, IEEE, pp. 1735–1740.
- Beis, J. S. and Lowe, D. G., 1997. Shape indexing using approximate nearest-neighbour search in high-dimensional spaces. In: *Computer Vision and Pattern Recognition, 1997. Proceedings., 1997 IEEE Computer Society Conference on*, IEEE, pp. 1000–1006.
- Brand, M., 2002. Incremental singular value decomposition of uncertain data with missing values. In: *European Conference on Computer Vision, Lecture Notes in Computer Science*, Vol. 2350, pp. 707–720.
- Byröd, M. and Aström, K., 2010. Conjugate gradient bundle adjustment. In: *European Conference on Computer Vision, Lecture Notes in Computer Science*, Vol. 6312, pp. 114–127.
- Chiang, K., Noureldin, A. and El-Sheimy, N., 2003. Multisensor integration using neuron computing for land-vehicle navigation. *GPS solutions* 6(4), pp. 209–218.
- Chum, O., Matas, J. and Kittler, J., 2003. Locally optimized RANSAC. In: *Pattern recognition*, Springer, pp. 236–243.
- Claus, D. and Fitzgibbon, A., 2005. A rational function lens distortion model for general cameras. In: *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, Vol. 1, pp. 213–219 vol. 1.
- El-Sheimy, N. and Schwarz, K., 1998. Navigating urban areas by VISAT—A mobile mapping system integrating GPS/INS/digital cameras for GIS applications. *Navigation* 45(4), pp. 275–285.
- Facco, P., Bezzo, F., Barolo, M., Mukherjee, R. and Romagnoli, J., 2009. Monitoring roughness and edge shape on semiconductors through multiresolution and multivariate image analysis. *AICHe Journal* 55, pp. 1147–1160.
- Facco, P., Tomba, E., Roso, M., Modesti, M., Bezzo, F. and Barolo, M., 2010. Automatic characterization of nanofiber assemblies by image texture analysis. *Chemometrics and Intelligent Laboratory Systems* 103, pp. 66–75.
- Fischler, M. and Bolles, R., 1981. Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM* 24(6), pp. 381–395.
- Foxlin, E., 2005. Pedestrian tracking with shoe-mounted inertial sensors. *Computer Graphics and Applications, IEEE* 25(6), pp. 38–46.
- Furukawa, Y. and Ponce, J., 2010. Accurate, dense, and robust multiview stereopsis. *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 32(8), pp. 1362–1376.
- Furukawa, Y., Curless, B., Seitz, S. and Szeliski, R., 2010. Towards internet-scale multi-view stereo. In: *Proceedings of the 23rd IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1434–1441.
- Fusiello, A. and Irsara, L., 2011. Quasi-euclidean epipolar rectification of uncalibrated images. *Machine Vision and Applications* 22(4), pp. 663–670.
- Guarnieri, A., Masiero, A., Vettore, A. and Pirotti, F., 2015. Evaluation of the dynamic processes of a landslide with laser scanners and bayesian methods. *Geomatics, Natural Hazards and Risk* 6(5–7), pp. 614–634.
- Habib, A. and Morgan, M., 2003. Automatic calibration of low-cost digital cameras. *Optical Engineering* 42(4), pp. 948–955.
- Hartley, R., 1997. In defense of the eight-point algorithm. *IEEE Transaction on Pattern Recognition and Machine Intelligence* 19(6), pp. 580–593.
- Hartley, R. and Zisserman, A., 2003. *Multiple View Geometry in Computer Vision*. Cambridge University Press.
- Heikkila, J. and Silven, O., 1997. A four-step camera calibration procedure with implicit image correction. In: *Computer Vision and Pattern Recognition, 1997. Proceedings., 1997 IEEE Computer Society Conference on*, pp. 1106–1112.
- Intel Research, 2000. OpenCV. In: <http://opencv.org/>.
- Kang, Z., 2015. The applications of robust method baysac in 3d indoor scene reconstruction fusing kinect point clouds and videocam images. In: *The 9th International Symposium on Mobile Mapping Technology, MMT 2015*.
- Karel, W. and Pfeifer, N., 2009. Range camera calibration based on image sequences and dense comprehensive error statistics. In: *IS&T/SPIE Electronic Imaging, International Society for Optics and Photonics*, pp. 72390D–72390D.
- Kraus, K. and Pfeifer, N., 1998. Determination of terrain models in wooded areas with airborne laser scanner data. *ISPRS Journal of Photogrammetry and remote Sensing* 53(4), pp. 193–203.
- Leutenegger, S., Chli, M. and Siegwart, R. Y., 2011. Brisk: Binary robust invariant scalable keypoints. In: *Computer Vision (ICCV), 2011 IEEE International Conference on*, IEEE, pp. 2548–2555.

- Longuet-Higgins, H., 1981. A computer algorithm for reconstructing a scene from two projections. *Nature* 293(5828), pp. 133–135.
- Lowe, D., 1999. Object recognition from local scale-invariant features. In: *Computer Vision, 1999. The Proceedings of the Seventh IEEE International Conference on*, Vol. 2, pp. 1150–1157 vol.2.
- Lukianto, C. and Sternberg, H., 2011. Stepping – smartphone-based portable pedestrian indoor navigation. *Archives of photogrammetry, cartography and remote sensing* 22, pp. 311–323.
- Ma, Y., Soatto, S., Košecká, J. and Sastry, S., 2003. *An Invitation to 3D Vision*. Springer.
- Masiero, A. and Cenedese, A., 2012. On triangulation algorithms in large scale camera network systems. In: *Proceedings of the 2012 American Control Conference, ACC 2012, Montréal, Canada*, pp. 4096–4101.
- Masiero, A. and Chiuso, A., 2006. Non linear temporal textures synthesis: a Monte Carlo approach. In: *Lecture Notes in Computer Science (Computer Vision–ECCV 2006)*, Springer, pp. 283–294.
- Masiero, A., Guarnieri, A., Pirotti, F. and Vettore, A., 2014a. A particle filter for smartphone-based indoor pedestrian navigation. *Micromachines* 5(4), pp. 1012–1033.
- Masiero, A., Guarnieri, A., Pirotti, F. and Vettore, A., 2015a. Painting recognition with smartphones equipped with inertial measurement unit. In: *Proceedings of SPIE, Optical Metrology*, Vol. 9527, pp. 95270X–95270X–10.
- Masiero, A., Guarnieri, A., Pirotti, F. and Vettore, A., 2015b. Semi-automated detection of surface degradation on bridges based on a level set method. *The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences* 40(3), pp. 15–21.
- Masiero, A., Guarnieri, A., Vettore, A. and Pirotti, F., 2014b. An ISVD-based euclidian structure from motion for smartphones. *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences XL-5*, pp. 401–406.
- Masiero, A., Guarnieri, A., Vettore, A. and Pirotti, F., 2014c. On the use of INS to improve feature matching. *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences XL-1*, pp. 227–232.
- Morel, J. and Yu, G., 2009. ASIFT: A new framework for fully affine invariant image comparison. *SIAM Journal on Imaging Sciences* 2(2), pp. 438–469.
- Morel, J. and Yu, G., 2011. Is SIFT scale invariant? *Inverse Problems and Imaging* 5(1), pp. 115–136.
- Pfeifer, N., Gorte, B., Winterhalder, D. et al., 2004. Automatic reconstruction of single trees from terrestrial laser scanner data. In: *Proceedings of 20th ISPRS Congress*, pp. 114–119.
- Piermattei, L., Carturan, L. and Guarnieri, A., 2015. Use of terrestrial photogrammetry based on structure-from-motion for mass balance estimation of a small glacier in the Italian Alps. *Earth Surface Processes and Landforms* 40(13), pp. 1791–1802.
- Piragnolo, M., Masiero, A., Fissore, F. and Pirotti, F., 2015. Solar irradiance modelling with NASA WW GIS environment. *ISPRS International Journal of Geo-Information* 4(2), pp. 711–724.
- Pirotti, F., Laurin, V., Vettore, A., Masiero, A. and Valentini, R., 2014. Small footprint full-waveform metrics contribution to the prediction of biomass in tropical forests. *Remote Sensing* 6(10), pp. 9576–9599.
- Prosdocimi, M., Calligaro, S., Sofia, G., Dalla Fontana, G. and Tarolli, P., 2015. Bank erosion in agricultural drainage networks: new challenges from structure-from-motion photogrammetry for post-event analysis. *Earth Surface Processes and Landforms* 40(14), pp. 1891–1906.
- Remondino, F. and Fraser, C., 2006. Digital camera calibration methods: considerations and comparisons. *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences XXXVI-5*, pp. 266–272.
- Remondino, F., Barazzetti, L., Nex, F., Scaioni, M. and Sarazzi, D., 2011. UAV photogrammetry for mapping and 3D modeling—current status and future perspectives. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* 38(1), pp. C22.
- Remondino, F., Guarnieri, A. and Vettore, A., 2005. 3D modeling of close-range objects: photogrammetry or laser scanning? In: *Electronic Imaging 2005, International Society for Optics and Photonics*, pp. 216–225.
- Ruiz, A., Granja, F., Prieto Honorato, J. and Rosas, J., 2012. Accurate pedestrian indoor navigation by tightly coupling foot-mounted IMU and RFID measurements. *Instrumentation and Measurement, IEEE Transactions on* 61(1), pp. 178–189.
- Saeedi, S., Moussa, A. and El-Sheimy, N., 2014. Context-aware personal navigation using embedded sensor fusion in smartphones. *Sensors* 14(4), pp. 5742–5767.
- Tomasi, C. and Kanade, T., 1992. Shape and motion from image streams under orthography: a factorization method. *International Journal of Computer Vision* 9(2), pp. 137–154.
- Toth, C., 2001. Sensor integration in airborne mapping. In: *Instrumentation and Measurement Technology Conference, 2001. IMTC 2001. Proceedings of the 18th IEEE*, Vol. 3, IEEE, pp. 2000–2005.
- Toth, C. and Grejner-Brzezinska, D., 1997. Performance analysis of the airborne integrated mapping system (AIMS). *International Archives of Photogrammetry and Remote Sensing* 32, pp. 320–326.
- Toth, C., Jozkow, G., Ostrowski, S. and Grejner-Brzezinska, D., 2015. Positioning slow moving platforms by uwb technology in GPS-challenged areas. In: *The 9th International Symposium on Mobile Mapping Technology, MMT 2015*.
- Triggs, B., McLauchlan, P., Hartley, R. and Fitzgibbon, A., 1999. Bundle adjustment - a modern synthesis. In: *Proceedings of the International Workshop on Vision Algorithms: Theory and Practice*, Springer Lecture Notes on Computer Science, Springer Verlag, pp. 298–372.
- Widyawan, P., Pirkel, G., Munaretto, D., Fischer, C., An, C., Lukowicz, P., Klepal, M., Timm-Giel, A., Widmer, J., Pesch, D. and Gellersen, H., 2012. Virtual lifeline: Multimodal sensor data fusion for robust navigation in unknown environments. *Pervasive and Mobile Computing* 8(3), pp. 388–401.