ON THE USE OF INS TO IMPROVE FEATURE MATCHING

A. Masiero[a, *], A. Guarnieri[a], A. Vettore[a], F. Pirotti[a]

[a] Interdepartmental Research Center of Geomatics (CIRGEO), University of Padova,
Viale dellUniversit 16, Legnaro (PD) 35020, Italy -
masiero@dei.unipd.it
(alberto.guarnieri, antonio.vettore, francesco.pirotti)@unipd.it

KEY WORDS: Reconstruction, Calibration, Camera, Real-time, Photogrammetry

ABSTRACT:

The continuous technological improvement of mobile devices opens the frontiers of Mobile Mapping systems to very compact systems, i.e. a smartphone or a tablet. This motivates the development of efficient 3D reconstruction techniques based on the sensors typically embedded in such devices, i.e. imaging sensors, GPS and Inertial Navigation System (INS). Such methods usually exploits photogrammetry techniques (structure from motion) to provide an estimation of the geometry of the scene.
Actually, 3D reconstruction techniques (e.g. structure from motion) rely on use of features properly matched in different images to compute the 3D positions of objects by means of triangulation. Hence, correct feature matching is of fundamental importance to ensure good quality 3D reconstructions.
Matching methods are based on the appearance of features, that can change as a consequence of variations of camera position and orientation, and environment illumination. For this reason, several methods have been developed in recent years in order to provide feature descriptors robust (ideally invariant) to such variations, e.g. Scale-Invariant Feature Transform (SIFT), Affine SIFT, Hessian affine and Harris affine detectors, Maximally Stable Extremal Regions (MSER).
This work deals with the integration of information provided by the INS in the feature matching procedure: a previously developed navigation algorithm is used to constantly estimate the device position and orientation. Then, such information is exploited to estimate the transformation of feature regions between two camera views. This allows to compare regions from different images but associated to the same feature as seen by the same point of view, hence significantly easing the comparison of feature characteristics and, consequently, improving matching. SIFT-like descriptors are used in order to ensure good matching results in presence of illumination variations and to compensate the approximations related to the estimation process.

1. INTRODUCTION

Thanks to their flexibility and their ability to quickly collect a large amount of geospatial data, mobile mapping technologies are continuously increasing their importance among the currently available remote sensing systems.

Nevertheless, most of the already existing mobile mapping systems suffer of certain issues, among them: quite expensive cost, limited portability in certain environments, the quality of the acquired data can be checked only a posteriori. Despite these issues are not so relevant for a number of applications, they can limit the diffusion of mobile mapping technologies among not specialized personnel.

In order to tackle the issues mentioned above, it has been recently proposed the development of a low cost mobile mapping system, based on the sole use of a smartphone-like device (Saeedi et al., 2014, Masiero et al., 2014). The proposed systems aim at properly exploiting the sensors embedded in the device: the device is typically assumed to be provided with a standard camera and with a MEMS based Inertial Navigation System. Shots taken by the embedded camera are used to compute a 3D reconstruction of the scene by means of a photogrammetry approach.

The limited computational power and battery life of the mobile device impose challenging requirements on the algorithm used for the 3D reconstruction procedure. In order to reduce the computational load, very efficient algorithms have been recently proposed, either based on the use of Preconditioned Conjugate Gradients to speed up the bundle adjustment optimization (Agarwal et al., 2010, Byröd and Aström, 2010), or based on the Incremental Singular Value Decomposition (ISVD) to solve the Structure from Motion (SfM) problem ((Tomasi and Kanade, 1992, Brand, 2002, Kennedy et al., 2013, Masiero et al., 2014)).

Despite different approaches can be considered for the 3D reconstruction of the scene, the rationale of such procedures is that of properly matching features viewed from different cameras and exploit geometry triangulation to compute the 3D positions of the points corresponding to the matched features (Ma et al., 2003, Hartley and Zisserman, 2003, Hartley and Sturm, 1997, Triggs et al., 1999, Masiero and Cenedese, 2012) (Fig. 1).

From the above considerations, it immediately follows that the quality (and the efficiency, as well) of the reconstruction procedures strongly depends on the use of features correctly matched in different images. Due to changes of the camera position and orientation and illumination variations, the same feature ca appear significantly differently in different images. This motivates the use of proper techniques to robustly match features in different images.

Several approaches have been proposed in the literature to extract and properly match features. Among them, Scale-Invariant Feature Transform (SIFT) is widely known for its ability in computing features invariant to certain transformations (e.g. scale, rotation and illumination changes). In order to take into account of object deformations due to perspective changes as well (i.e. deformations related to tilts, shifts and camera projection), recently methods that approximate such deformations with affine transformations have been considered (Affine SIFT (Morel and Yu, 2009, Morel and Yu, 2011)), Hessian affine and Harris affine de-

[*] Corresponding author.

This contribution has been peer-reviewed.
doi:10.5194/isprsarchives-XL-1-227-2014

tectors (Mikolajczyk and Schmid, 2002), Maximally Stable Extremal Regions (MSER) (Matas et al., 2004)).

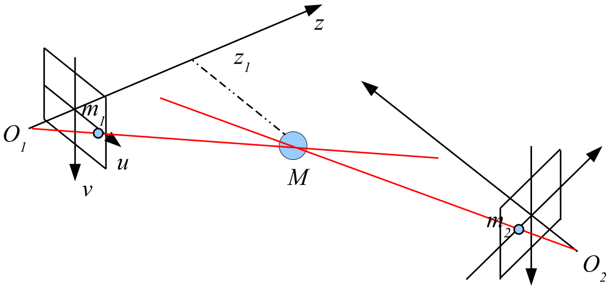

Figure 1: 3D reconstruction based on triangulation.

Such methods for computing invariant features are based on the information provided by the acquired images. However, in our case of interest information provided by the INS is available as well: the goal of this work is that of integrating the information about device position and orientation in the 3D reconstruction procedure, in order to improve the obtained results. To be more specific, here we focus on the feature matching stage.

In this paper we take advantage of a previously developed navigation technique (Masiero et al., 2013): measurements provided by the Inertial Navigation System are used to estimate the device position and orientation at the shooting instants. This allows us to compute the approximate coordinate system transformation between different image views. Then, feature regions can be mapped into the coordinate system of the other camera views: hence feature characteristics are compared as seen by the same point of view.

Actually, a calibrated should be required in order to obtain a very accurate feature region transformation in other camera views, however, as shown in the simulation section, quite good results can be obtained by using uncalibrated systems (by means of a very rough approximation of the camera interior parameters).

The paper is organized as follows: a brief description of the considered system and the 3D reconstruction procedure are given in Section 2. The proposed feature matching procedure is presented in Section 3. Finally, some simulation results are presented and discussed in Section 4.

2. SYSTEM DESCRIPTION

This work assumes the use of a (typically low cost) mobile device (e.g. a smartphone). Such device has to be provided of an imaging sensor (i.e. a camera), and of a navigation system. Specifically, our tests of the simulation section have been performed by using a low-cost smartphone, Huawei Sonic U8650. The considered smartphone is provided with a 3-axis accelerometer (Bosh, BMA150, resolution 0.15m/s^2, maximum range ±39.24m/s^2), a 3-axis magnetic field sensor (Asahi Kasei, AK8973, resolution 0.0625μT, maximum range ±2000μT), and a 3 Megapixel camera.

During the data acquisition process, the device is moved on several locations, where the user takes shots of the scene by means of the camera embedded in the device. Then, images acquired by the embedded camera are used to estimate a 3D reconstruction of the scene: 3D reconstruction is accessed by relating with each others features in shots taken from different point of views (Structure from Motion (SfM) approach (Hartley and Zisserman, 2003, Ma et al., 2003)).

Figure 2: Smartphone used to test the reconstruction system: Huawei U8650 Sonic.

Fig. 3 shows the scheme of the 3D reconstruction procedure: first, features are extracted from the acquired images; then features from different images are compared and properly matched; finally, matched features are used to compute the 3D reconstruction by solving the SfM problem. An efficient solution of SfM problem can be obtained either using optimized techniques of bundle adjustment (Agarwal et al., 2010, Byröd and Aström, 2010), or by incremental factorization methods (Brand, 2002, Kennedy et al., 2013).

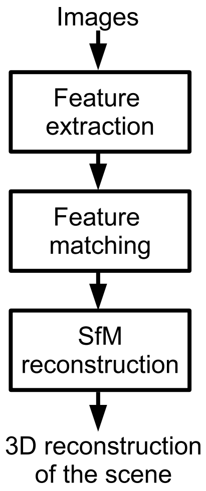

Figure 3: Reconstruction procedure scheme.

In this work we consider the use of INS in addition to the camera images: the considered positioning system is as in (Masiero et al., 2013), where simultaneous measurements by the embedded sensors (e.g. accelerometer and magnetometer) allow the estimation of both the movements of the mobile device, and the attitude of the device during the camera shot. However, notice that, depending on the working conditions of interest, different choices can be considered without affecting the effectiveness of the procedure (Azizyan et al., 2009, Bahl and Padmanabhan, 2000, Cenedese et al., 2010, Foxlin, 2005, El-Sheimy et al., 2006, Guarnieri et al., 2013, Lukianto and Sternberg, 2011, Masiero et al., 2013, Ruiz et al., 2012, Youssef and Agrawala, 2005, Wang et al., 2012, Widyawan et al., 2012, Piras et al., 2010, Saeedi et al., 2014).

In order to simplify the use of the system to the user, both imaging and INS sensors are uncalibrated (no specific calibration procedure is required). However, during data acquisition a rough calibration of the magnetometer can be considered in order to reduce the error on the estimated device orientation (Masiero et al., 2013).

In addition, a rough estimate of the interior camera parameters is usually available. Assume that the camera is well approximated by a pinhole (projective) camera (no radial distortion is considered), then its measurements can be expressed as follows (Ma et al., 2003):

$$m_i \simeq P_i M \tag{1}$$

where the measurement m_i of a feature on the image plane of camera view i is related with its corresponding 3D point M. In the above equation, P_i is the projective matrix of camera view i, m_i and M are written by using homogeneous coordinate notation (Ma et al., 2003). As usually done when dealing with homogeneous coordinates, $\simeq$ stands for the equality up to a scale factor. The effect of the projection is graphically depicted in Fig. 1.

The projection matrix P_i can be expressed as follows:

$$P_i = K_i \left[\begin{array}{c|c} R_i & -R_i t_i \end{array} \right] \tag{2}$$

where K_i is the matrix containing the interior camera parameters, t_i and R_i correspond to the position and rotation matrix (related to the camera orientation) of the device, respectively. Hereafter R_i and t_i are assumed to be (approximately) known, thanks to the estimates provided by the navigation system. Assuming to use the same device (with fixed focal length lens) during all image acquisitions, then $K_i = K, \forall i$.

Despite the real value of the interior parameter matrix K_i is unknown, it can be roughly approximated as follows:

$$K \approx \begin{bmatrix} a & 0 & u_0 \\ 0 & a & v_0 \\ 0 & 0 & 1 \end{bmatrix} \tag{3}$$

where pixels are assumed to be approximately squares, sensor axes are assumed to be orthogonal, and the displacement of the sensor center with respect to the optical axis is approximated with $(u_0, v_0) \approx (\frac{c}{2}, \frac{r}{2})$, where r and c are the number of image rows and of columns, respectively (camera coordinate system can be seen in Fig. 1). The parameter a is related to the focal length and to the pixel size. When the characteristics of the device are available the value of a can be quite easily approximated. When no information on such characteristics is available, the procedure described in the following section shall be repeated[1] for different values of a ranging in the following interval $(\frac{1}{3}(r+c), 3(r+c))$, where the real value of a is supposed to be (Heyden and Pollefeys, 2005, Fusiello and Irsara, 2011).

In the following we will focus on the integration of the device position and orientation information in the feature matching step of the procedure shown in Fig. 3.

3. FEATURE MATCHING

Reconstruction procedures based on a feature matching approach estimate the geometry of the scene by analyzing the position of features from different point of views and triangulating the 3D feature position. Hence, in order to properly estimate the geometry of the scene the same spatial point has to be recognized and matched in the images where it is visible: the use of a proper feature matching technique is of fundamental importance in order to ensure the effectiveness of the reconstruction procedure.

3.1 Approximate epipolar constraints

Let m_1 and m_2 be the homogeneous coordinates of the features to be compared (from camera views 1 and 2, respectively). In addition, let P_1, P_2, R_1, R_2, t_1, t_2 be the corresponding projection matrices, camera rotation matrices and positions. Notice that, according with the assumptions presented in the previous section, here such variables are approximately known.

Given the above information it is possible to compute an approximate fundamental matrix F (Hartley and Zisserman, 2003): hence, as usual, according with the (approximate) geometry of the system m_1 and m_2 can be considered to be matched only if $m_1^\top F m_2 \approx 0$. Since F is only an approximation of the real fundamental matrix, then such condition shall be significantly weakened, i.e. $|m_1^\top F m_2| < \sigma_F$, where σ_F is a proper threshold that should be set to a value (significantly) larger than 0 (such value depends on the quality of the approximations done to compute F and on the value of measurement noise). Hence the search for features matching m_1 cannot be reduced to a 1D search as in the case of availability of the perfect F. Nevertheless, such (approximate) epipolar constraint can be used to substantially reduce the number of comparisons to be done.

3.2 Feature matching by region transform

Feature matching is based on the appearance of the 2D image regions in the neighborhood of the considered features: since images are taken from different point of views the same feature can undergo certain appearance changes, then the goal of several matching techniques recently proposed in the literature is that of extracting features invariant to such deformations (SIFT ensure invariance for scale and illumination variations and rotations on the image plane, while in ASIFT, Harris and Hessian affine detectors tilts and projection deformations are locally modeled as affine transforms).

Such effort on the formulation of invariant features is motivated by the usual impossibility of comparing features in similar conditions, while, ideally speaking, the optimal condition for comparing two features is that of looking at them from the same point of view: when such condition holds, then invariance to several transforms (e.g. rotations, translations) becomes a not desired condition.

In the working conditions considered here both K and the change of coordinates between the camera views are approximately known, hence they can be used to look at the features from approximately the same point of view as described in the following:

- The 3D point M corresponding to the features m_1, m_2 can be computed by triangulation (Hartley and Sturm, 1997, Hartley and Zisserman, 2003, Masiero and Cenedese, 2012) (Fig. 1). For such point (1) can be rewritten as follows

$$m_i z_i = P_i M \quad , \quad \text{for} \quad i = \{1,\, 2\} \tag{4}$$

[1] The value of a that allows to obtain the largest number of matching features is expected to be the most probable. Since the feature matching procedure has typically to be repeated for different couples of images, then the previously estimated value of a can be exploited in order to reduce the overall computational load.

This contribution has been peer-reviewed.
doi:10.5194/isprsarchives-XL-1-227-2014

where z_i is the coordinate of M along the optical axis direction of the ith camera view.

- The projection procedure can be (partially) inverted for camera view 1:

$$\bar{M_1} = z_1 R_1^{-1} K^{-1} m_1 \tag{5}$$

Then M_1 is obtained by translating $\bar{M_1}$ of $\Delta = (M - \bar{M_1})$, in order to obtain $M_1 = M$.

- M_1 is projected on the image plane of the second camera view:

$$m_{12} \simeq P_2 M_1 \tag{6}$$

where m_{12} are the homogeneous coordinates of the first feature projected on the image plane of the second camera view.

Since the feature comparison is done taking into account of the feature local appearance, then, once Δ has been computed, the last two steps of the above procedure should be repeated for the points in the neighborhood of m_1. Let m_1' be a point in the neighborhood of m_1, then its projection on the image plane of the second camera view can be obtained as follows:

- The projection procedure can be (partially) inverted for camera view 1:

$$\bar{M_1'} = z_1 R_1^{-1} K^{-1} m_1' \tag{7}$$

Then, $M_1' = \bar{M_1'} + \Delta$.

- M_1' is projected on the image plane of the second camera view:

$$m_{12}' \simeq P_2 M_1' \tag{8}$$

where m_{12} are the homogeneous coordinates of the first feature projected on the image plane of the second camera view.

Then the SIFT descriptor of the first feature is computed on its version projected on the image plane of the second camera view. Since the points mapped in this way on the image plane of the second camera view are typically not disposed on a grid, then, for convenience of computation of the SIFT descriptor, an approximation of the image values corresponding to a grid disposition can be easily obtained by interpolation. The resulting descriptor can be compared as usually with the SIFT descriptor of the second feature.

4. RESULTS AND CONCLUSIONS

In this section SIFT and the approach presented here are compared on the task of matching features computed on images of the facade of a building of the University of Padova. Images are taken from 3 different positions and orientations (with mean distance of approximately 12 m). The mean distance from the facade is of approximately 20 m.

The sensors of the considered device have not been calibrated and the parameter a have been varied among 10 possible values in the interval of values most commonly used.

Feature points to be matched are extracted as in the SIFT algorithm: hence the same feature points to be compared have been considered by both the approaches. Features have been matched to their most similar feature on the other image, according with the distance between the descriptors.

The proposed approach has correctly matched approximately 30% more couples of features. Fig. 4 shows the features properly matched by the proposed algorithm in two images of the building.

Fig. 5 compares the image region in the neighborhood of a feature in different cases: the regions on the images taken by camera 1 (left) and 2 (right), while the image in the middle corresponds to the region taken by camera 1 projected on the image plane of camera 2 as described in the previous section. In practice the image in the middle can be considered as an estimation of the image region in the right obtained by using only the information of the left image. It is clear that the use of uncalibrated sensors ensure lower quality results with respect to those expected in a calibrated case. Nevertheless, the synthetic projection described in the previous section has partially succeed in producing an image more similar to the right one with respect to the left one, in particular close to the feature position.

Some observations are now in order: first, the procedure described in the previous section is based on a local description of the image region in the feature neighborhood as a planar surface parallel to the image plane. Clearly this is an approximation that can be more or less realistic depending on the considered case. Furthermore, in accordance with the above consideration, such approximation is expected to become less reliable as the distance from the feature increases.

Interestingly, by comparing the left border of the synthetic window (middle image in Fig. 5) with that in the left and right windows, it is possible to notice that obviously the system cannot properly estimate image parts that are not visible in the original image (e.g. the internal left border of the window is not visible in the left image, and, consequently, it cannot be estimated in the middle one as well).

Since by means of the procedure previously presented the features are compared approximately from the same point of view, then, as previously claimed, in these working conditions the feature descriptor invariance (for instance) to camera rotations and translations is not a condicio sine qua non. However, in practice the use of a SIFT descriptor is still useful to compensate the effects of the considered approximations (e.g. the presented procedure is applied by using approximate values of the real parameters).

According to our simulations the proposed algorithm allows to improve the matching results of the SIFT technique in presence of camera tilts. However, this is obtained at the cost of an increase of the computational complexity: our future work will focus on the optimization of the technique in order to make it more computationally efficient.

Finally, it is worth to notice that, despite the considered approach allows to improve the SIFT results on compensating changes in the feature appearance (due to perspective change), unfortunately mismatches are unavoidable, in particular when dealing with repetitive structures, e.g. in human buildings. In order to make feature matching more reliable in such critical conditions, another matching step based on the reconstructed system geometry can be useful: after matching points between two images, a RANSAC approach (Fischler and Bolles, 1981) shall be used in order to obtain a more reliable estimate of the fundamental matrix (i.e. by using

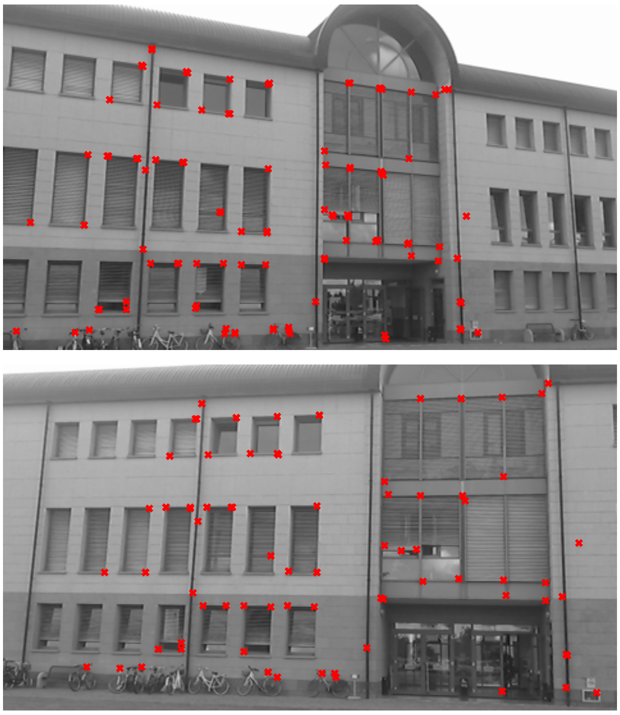

Figure 4: Example of different camera views: correctly matched features by the proposed algorithm are shown (red crosses).

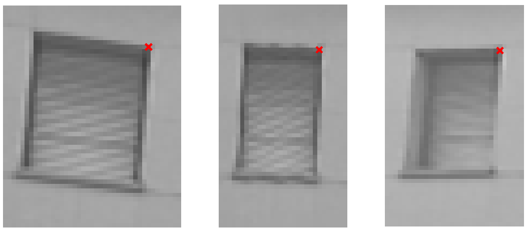

Figure 5: Feature matching: feature region in the first camera view (left), feature region mapped from the first to the second camera view (middle), feature region in the second camera view (right). The considered feature reported (red crosses) in all the images.

the eight-point algorithm as in (Longuet-Higgins, 1981, Hartley, 1997, Hartley and Zisserman, 2003)), then such matrix can be used to make a more robust feature selection based on a better knowledge of the system geometry.

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.

Azizyan, M., Constandache, I. and Choudhury, R., 2009. Surroundsense: mobile phone localization via ambience fingerprinting. In: Proceedings of the 15th annual international conference on Mobile computing and networking, MobiCom '09, pp. 261–272.

Bahl, P. and Padmanabhan, V., 2000. RADAR: An in-building RF-based user location and tracking system. In: IEEE INFOCOM, Vol. 2, pp. 775–784.

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.

Cenedese, A., Ortolan, G. and Bertinato, M., 2010. Low-density wireless sensor networks for localization and tracking in critical environments. Vehicular Technology, IEEE Transactions on 59(6), pp. 2951–2962.

El-Sheimy, N., Kai-wei, C. and Noureldin, A., 2006. The utilization of artificial neural networks for multisensor system integration in navigation and positioning instruments. Instrumentation and Measurement, IEEE Transactions on 55(5), pp. 1606–1615.

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., 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.

Goesele, M., Snavely, N., Curless, B., Hoppe, H. and Seitz, S., 2007. Multi-view stereo for community photo collections. In: Proceedings of the 11th IEEE International Conference on Computer Vision (ICCV).

Guarnieri, A., Pirotti, F. and Vettore, A., 2013. Low-cost mems sensors and vision system for motion and position estimation of a scooter. Sensors 13(2), pp. 1510–1522.

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 Sturm, P., 1997. Triangulation. Computer Vision and Image Understanding 68(2), pp. 146–157.

Hartley, R. and Zisserman, A., 2003. Multiple View Geometry in Computer Vision. Cambridge University Press.

Heyden, A. and Pollefeys, M., 2005. Multiple view geometry. In: G. Medioni and S. B. Kang, editors, Emerging Topics in Computer Vision, Prentice Hall, pp. 45–107.

Kennedy, R., Balzano, L., Wright, S. and Taylor, C., 2013. Online algorithms for factorization-based structure from motion. ArXiv p. 1309.6964.

Longuet-Higgins, H., 1981. A computer algorithm for reconstructing a scene from two projections. Nature 293(5828), pp. 133–135.

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 Cenedese, A., 2013. Affinity-based distributed algorithm for 3d reconstruction in large scale visual sensor networks. In: Proceedings of the 2014 American Control Conference, ACC 2014, Portland, USA.

Masiero, A., Guarnieri, A., Vettore, A. and Pirotti, F., 2013. An indoor navigation approach for low-cost devices. In: Indoor Positioning and Indoor Navigation, IPIN 2013, Montbeliard, France.

Masiero, A., Guarnieri, A., Vettore, A. and Pirotti, F., 2014. 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.

Matas, J., Chum, O., Urban, M. and Pajdla, T., 2004. Robust wide-baseline stereo from maximally stable extremal regions. Image and vision computing 22(10), pp. 761–767.

Mikolajczyk, K. and Schmid, C., 2002. An affine invariant interest point detector. In: Proceedings of the 8th International Conference on Computer Vision (ICCV), pp. 128–142.

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.

Piras, M., Marucco, G. and Charqane, K., 2010. Statistical analysis of different low cost GPS receivers for indoor and outdoor positioning. In: IEEE Position Location and Navigation Symposium (PLANS), MobiSys '12, pp. 838–849.

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.

Snavely, N., Seitz, S. and Szeliski, R., 2008. Modeling the world from internet photo collections. International Journal of Computer Vision 80(2), pp. 189–210.

Tomasi, C. and Kanade, T., 1992. Share and motion from image streams under orthography: a factorization method. International Journal of Computer Vision 9(2), pp. 137–154.

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.

Wang, H., Sen, S., Elgohary, A., Farid, M., Youssef, M. and Choudhury, R., 2012. No need to war-drive: unsupervised indoor localization. In: Proceedings of the 10th international conference on Mobile systems, applications, and services, MobiSys '12, pp. 197–210.

Widyawan, Pirkl, 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.

Youssef, M. and Agrawala, A., 2005. The horus WLAN location determination system. In: Proceedings of the 3rd international conference on Mobile systems, applications, and services, MobiSys '05, pp. 205–218.