Accurate Positioning and Orientation Estimation in Urban Environment Based on 3D Models

Giorgio Ghinamo[1(✉)], Cecilia Corbi[1], Piero Lovisolo[1], Andrea Lingua[2], Irene Aicardi[2], and Nives Grasso[2]

[1] Telecom Italia, Torino, Italy
{giorgio.ghinamo,cecilia.corbi,piero.lovisolo}@telecomitalia.it
[2] Department of Environment, Land and Infrastructure Engineering (DIATI), Politecnico di Torino, Torino, Italy
{andrea.lingua,irene.aicardi,nives.grasso}@polito.it

Abstract. This paper describes a positioning algorithm for mobile phones based on image recognition. The use of image recognition based (IRB) positioning in mobile applications is characterized by the availability of a single camera for estimate the camera position and orientation. A prior knowledge of 3D environment is needed in the form of a database of images with associated spatial information that can be built projecting the 3D model on a set of synthetic solid images (range + RGB images). The IRB procedure proposed by the authors can be divided in two steps: the selection from the database of the most similar image to the query image used to locate the camera and the estimation of the position and orientation of the camera based on available 3D data on the reference image. The MPEG standard Compact Descriptors for Visual Search (CDVS) has been used to reduce hugely the processing time. Some practical results of the location methodology in outdoor environment have been described in terms of processing time and accuracy of position and attitude.

Keywords: Image recognition based location · Visual search · Positioning · Smartphones · Low cost

1 Introduction

As known, the positioning in indoor environments and within urban canyon is difficult and with poor accuracy through the use of common sensors GPS/GNSS. In recent years there has been evaluated the possibility of using alternative sensors that allow the positioning in these areas; among these image [1], optical [2], radio [3], magnetic [4], RFID [5] and acoustic [6] sensors were analyzed and tested.

The improvement of the sensors for image acquisitions included inside smartphones makes it possible to use these tools, more and more common, for navigation based on images; furthermore a challenge of this approach is to achieve real-time capability.

Kitanov et al. [7] compare image lines, that have been detected in images of a robot mounted camera, with a 3D vector model. Jason Zhi Liang generated a sparse

V. Murino et al. (Eds.): ICIAP 2015 Workshops, LNCS 9281, pp. 185–192, 2015.
DOI: 10.1007/978-3-319-23222-5_23

2.5D georeferenced image database using an ambulatory backpack-mounted system with two 2D laser scanners, two fish-eye cameras and one orientation sensor originally developed for 3D modeling of indoor environments.

In this paper we propose an innovative approach for positioning/navigation using a single camera of a mobile phone based on image recognition, exploiting 3D environment models in form of a set of 3D solid images (range+RGB data) and the new MPEG standard Compact Descriptors for Visual Search (CDVS). IRB positioning represents a good opportunity for Location Based Services (LBS), for example in the case of GNSS/Pseudolites denied environments as dense urban scenarios. Moreover, an advantage of IRB technology is the availability of 3D orientation of the used device (smartphone), information not available or not reliable using alternative positioning technologies.

A Terrestrial LiDAR Survey (TLS) with an associated camera can be used to acquire the 3D model of urban environment and to generate the database of reference images, a set of synthetic 3D images produced projecting the clouds of points over synthetic image plans. In this context MPEG algorithms for visual search play an important role in defining light and interoperable solution for processing and comparing the query and database images. This location procedure can be used for accurate navigation and augmented reality in urban scenarios for smart phones applications. In these use cases, to optimize battery consumption and compensating latencies of few seconds, the IRB procedure can be used jointly with inertial system, in particular with PDR (Pedestrian Dead Reckoning) technology.

The proposed procedure has already demonstrated excellent results in indoor environment [8], [9], then the purpose of this article is to validate the procedure also for the outdoor case.

2 The Positioning Procedure

The proposed procedure can be divided in two steps. As first step, a reference image is selected out of the database of images synthetically generated from the 3D environment model; this selection procedure exploits MPEG CDVS [10] visual search technology. The second step of the positioning procedure is the estimation of the camera parameters (position and orientation) based on available 3D information on the previously selected reference image according to the collinearity equations [11]. Key points and related features are extracted from query and reference images and matched for the selected set of key-points pairs.

In this section the positioning procedure is described in terms of functional steps (Image DB set up, reference image extraction, camera parameter estimation for the query image camera); the section 3 describes the test site, the image based positioning procedure and the related set up procedure; section 4 presents the performance results in terms of achieved accuracy and processing time gain.

2.1 High Level Functional Steps

The proposed location methodology consists of the following parts (Fig. 1):

- acquisition of a 3D model of the area where the positioning service is offered: the model is used to generate a synthetic images database with related 3D information. Due to the properties of image recognition algorithms based on local descriptors, whose performances rely on the similar perspective of details (i.e. key points), the database should provide an exhaustive coverage of the area where the service is offered, in terms of a grid of camera positions, orientations and focal length;

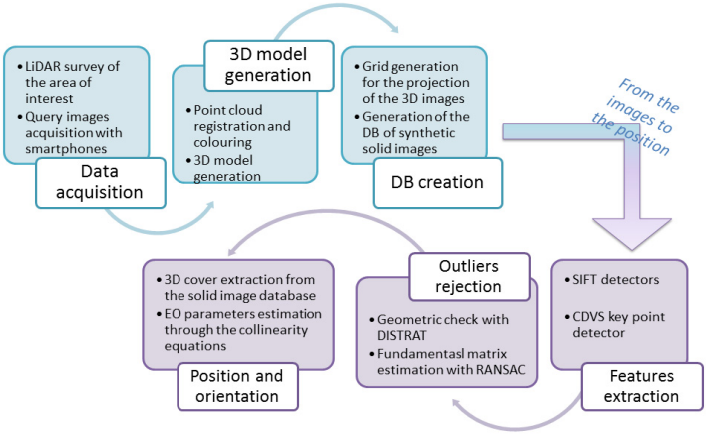

Fig. 1. Workflow of the Image Recognition Based procedure

- mobile phone takes a query picture used for locating the camera: a reference image, that is the most similar to the query one, is extracted from the database. For this task, the MPEG CSVS technology is used, with a minor changes regarding the distribution of selected most significant key points;
- using 3D information available for the selected reference image, external orientation parameters of camera (3D position and attitude angles) are estimated.

2.2 The 3D Model and the Synthetic Images Database

All the images of the database with related 3D information are created processing a colored 3D model of the environment. The 3D model can be generated with a TLS system, that also allows the image acquisitions with an integrated camera. The acquired point cloud are colored using the camera associated to the LiDAR instrument.

Numerous different scans are acquired and all of them are mounted in a single model and reported in used coordinates system (Fig. 2 left). As a result of the process, a geo-referenced RGB point cloud of the environment is made, on which you can directly read 3D coordinates/color of object points.

From the 3D model of the environment, a database of Solid Images (SIs) is created. SIs are synthetic RGB color images with the additional information about the distance from the camera center of the spatial point represented in each pixel [12]. Combining the camera parameters information together with the distance of object represented in the pixel from the camera, the 3D position of points is estimated, in terms of 3D coordinates of key points in the model reference system.

2.3 The Retrieval of Reference Image out of the Reference Database

The goal of the retrieval procedure is to select a reference image out of the images database with the highest level of similarity with the query image acquired by the terminal camera. To quickly select out of a database the most similar image, the following operations have been defined by MPEG CDVS:

- the images of the database are preliminary ranked based on a global descriptors similarity score when compared with the query image. Global descriptors provide a statistical representation of a set of most significant local descriptors extracted from the two images;
- for the images selected in previous step, the pairwise matching procedure with query image is executed between a limited number of most significant extracted key points. Trying to couple similar key points present in both images, the matched key points are validated by a geometric check [13] based on the concept that the statistical properties of the log distance ratio for pairs of incorrect matches are distinctly different from the properties of that for correct matches.

2.4 The Estimation Procedure of Camera Parameters

The second step of the location procedure, EO (External Orientation) parameters estimation (position and orientation), is based on the resolution of collinearity equations where key points of the query image are associated with 3D position information available in the reference image extracted from the database, with related spatial information (see 2.2). The 3D information is stored in construction of the solid images where for each pixel the distance (range) of the obstacle depicted in the image is reported, together with internal/external orientation parameters of SI in terms of orientation, focal length and sensor position.

The information that should be estimated in the location procedure are the EO parameters of the query image camera. The procedure to estimate these parameters from a solid image consists of the following steps:

1. features extraction from query and reference images using SIFT detector [14] or CDVS key point detector [10];
2. CDVS geometric check (DISTRAT) [13] is used for a coarse preliminary rejection of matched outliers; the use of DISTRAT is required to speed up outliers rejection procedure. However, the DISTRAT output still contains few percentiles of outliers in the selected set of paired features;

3. given the set of common features selected out of DISTRAT geometric check, the fundamental matrix is estimated with a RANSAC procedure, where the fundamental matrix is a representation of the rototraslation of the camera between query image and reference image [15]. This step allows to exclude remaining outliers out of DISTRAT check. In fact, the preliminary use of DISTRAT reduces the percentage of outliers from 30-70% order to few percentiles, this allows to strongly reduce the RANSAC execution time, approximately of 100 times (at this stage focal length is assumed to be similar in both images out of retrieval step and the camera distortion model are not taking into account);
4. the spatial information (3D coordinates) of the common features between query and reference image is retrieved using the solid image information available for the reference images of the DB, derived from the 3D model of the scene;
5. using the collinearity equations the EO parameters are estimated ([15], [16]), as this step is implemented through a non-linear Least Square estimation, as starting solution the EO parameters of the selected SI out of step 2.3 are considered.

3 The Test Site

An outdoor test has been done to validate the procedure analyzing the result accuracy. The defined area is along three blocks of via Garibaldi, an historical central pedestrian road, in Torino (Piedmont, Italy). A Faro laser scanner (series Cam2 Focus 3D) has been used for a TLS survey composed by six scans for a length of about 150 m. Examples of used scans are presented, as spherical images, in fig. 2 (right). However, if the interest area is larger, it is possible to acquire the colored point cloud and to generate the 3D model of the environment using Mobile Mapping Systems techniques (MMS), that allows to get information over a large area in a short time [17].

Fig. 2. Example of spherical image in pedestrian downtown road

In order to generate a reference images database characterized by an exhaustive cove-rage of all the possible perspective of the environment, a grid of points have been taken into account. Points are spaced 2 meters on the direction orthogonal to buildings front and 3 meter on the direction parallel to building front. For each points 16 different headings on the horizontal plane and 4 different inclination of the vertical axe (0, 5, 10 and 15 degrees) have been considered for a total of 64 images (Fig. 3, left), 1826 images for a single laser scanner position [18]. Fig. 3 (right) shows a snapshot of some synthetic images of the DB.

4 Trial Results

The proposed location procedure has been tested in terms of processing load benefit and the accuracy of the approach. For accuracy estimation we have considered a set of 20 smartphone images geo-located with photogrammetric techniques, in terms of position and attitude, as described in Section 2.

Fig. 3. The schema for the synthetic images generation (left) and examples of a part of the database of synthetic images (right)

4.1 Accuracy

Twenty images captured with a Samsung S4 smartphone have been considered in the test. The true positions and EOs of S4 have been acquired using an "ad hoc" calibrated system (the butterfly, Fig. 4) that consists on a car support for mobile phone mounted on a plate with 4 colored spheres (red, blu, yellow and grey). The system can be placed on a tripod that allows rotations for vertical and horizontal camera images. Thus, the spheres have been measured in a carthographic reference frame with high accuracy (mm) using surveying techniques.

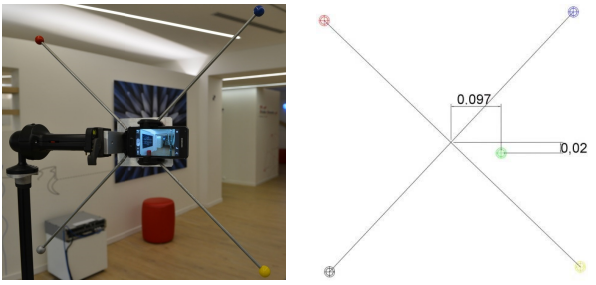

Fig. 4. The butterfly system, in green the camera projection center

Table 1 summarizes the accuracy results in terms of discrepancies from ground truth and estimated values in case of good level of similarity between the query image and the reference one extracted out of the database. The standard deviations of discrepancies are about 30 cm in position and about 0.15 radians in attitude. No systematic errors are present.

Fig. 5 describes an examples of results of key points detection, matching and outliers rejection in a couple of query and reference images. Query images and TLS are acquired in not ideal condition including people and cars randomly present in the scene, good level of similarity is detected between the two images. A maximum number of 2000 key points have been selected, ranked by the absolute value in descending order of the response of the keypoint to the Laplacian of Guassian filtering (peak), bring benefit in terms of reduction of number of key points.

Table 1. Accuracy results in outdoor trial for position (ΔX ΔY ΔZ) and attitude (Δω Δφ Δk)

Param	ΔX [m]	ΔY [m]	ΔZ [m]	Δω [rad]	Δφ [rad]	Δk [rad]
|Max|	0.420	0.500	0.320	0.139	0.037	0.118
Mean	0.059	-0.023	0.045	0.035	-0.012	-0.081
Dev.St.	0.249	0.383	0.191	0.098	0.047	0.093

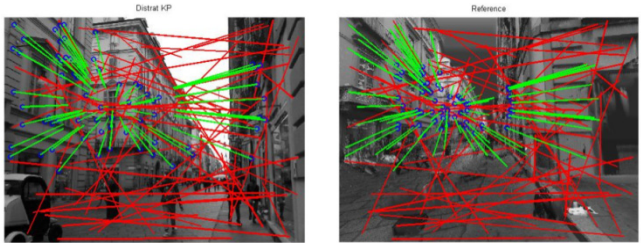

Fig. 5. Example of query and reference images key points matching results.

4.2 Processing Load

The tests have shown that the use of geometric check in pairwise matches outliers rejection allows to reduce processing time of a factor of 10 times or more with respect of pure RANSAC procedure, in case of medium degree of similarity between the query and the reference image, guaranteeing at the meantime an good accuracy.

Table 2 describes the processing load results of outlier rejection procedure for pure RANSAC procedure versus the hybrid DISTRAT and RANSAC approach, for an Intel Core 2 T7500 processor. When the similarity between query and reference image is not so high (the rate of good matches is around 30%) the hybrid approach provide strong benefit. Otherwise the processing load for the 2 approaches is similar.

Table 2. Processing load gain: hybrid DISTRAT and RANSAC vs RANSAC

Rate of inliers	RANSAC only	DISTART+RANSAC
35%	10 sec	0.6 sec
70%	0.5 sec	0.6 sec

5 Conclusions

The proposed location procedure offer a good level of accuracy with a standard error of few decimeters in the described scenarios. It is a very good results if we remember

that the approach is based on a single camera available on the smartphones, whose characteristics are significantly different respect to the cameras used for photogrammetric purpose. In the collinearity equations the focal length and the principal point are assumed known (it is also possible to take the nominal focal length, but it was not used in these tests). The proposed technique can be a component of image based navigation that we are now analyzing introducing the integration of the internal sensors data acquired from the smartphones or video odometry for consumer camera.

References

1. Nishkam, R., Pravin, S., Andrew, F., Ahmed, E., Liviu, I.: Indoor localization using camera phones. In: Mobile Computing Systems and Applications (2006)
2. Mautz, R., Tilch, S.: Survey of optical indoor positioning systems. In: International Conference on Indoor Positioning and Indoor Navigation (IPIN), September 21-23, 2011
3. Biswas, J., Veloso, M.: WiFi localization and navigation for autonomous indoor mobile robots. In: International Conference on Robotics and Automation (2010)
4. Chung, L., Donahoe, M., Schmandt, C., Kim, I., Razavai, P., Wiseman, M.: Indoor location sensing using geomagnetism. In: Proceedings of the 9th International Conference on Mobile Systems, Applications, and Services, pp. 141–154 (2011)
5. Schneegans, S., Vorst, P., Zell, A.: Using RFID snapshots for mobile robot self-localization. In: European Conference on Mobile Robots (2007)
6. Hong-Shik, K., Jong-Suk, C.: Advanced indoor localization using ultrasonic sensor and digital compass. In: International Conference on Control, Automation and Systems (2008)
7. Kitanov, A., Bišević, S., Petrović, I.: Mobile robot self-localization in complex indoor environments using monocular vision and 3D model. In: IEEE/ASME International Conference on Advanced Intelligent Mechatronics, Zürich, Switzerland (2007)
8. Piras, M., Dabove, P., Lingua, A.M., Aicardi, I.: Indoor navigation using smartphone technology: a future challenge or an actual possibility? In: IEEE/ION Position, Location Proceedings of the and Navigation Symposium, May 5-8, 2014
9. Lingua, A.M., Aicardi, I., Ghinamo, G., Francini, G., Lepsoy, S.: The MPEG7 visual search solution for image recognition based positioning using 3D models. In: Proceedings of the 27th International Technical Meeting of the Satellite Division of the Institute of Navigation (ION GNSS+ 2014), September 8-12, 2014
10. CDVS. ISO/IEC DIS 15938-13 Compact Descriptors for Visual Search (2014)
11. McGlone, C. (ed.): Manual of Photogrammetry, 5th edn., pp. 280–281. ASPRS
12. Bornaz, L., Dequal, S.: A new concept: the solid image. In: CIPA 2003 Proceedings of XIXth International Symposium, pp. 169–174 (2003)
13. PCT/EP2011/050994 Method and system for comparing images
14. Lowe, D.: Distinctive image features from scale-invariant keypoints. International, Journal of Computer Vision **60**(2), 91–110 (2004)
15. Hartley, R., Zisserman, A.: Multiple View Geometry in Computer Vision, 2nd edn. Cambridge University Press, March 2004
16. Karara, H.M. (ed.): Non Topography Photogrammetry, 2nd edn., pp. 46–48. ASPRS
17. De Agostino, M., Lingua, A., Marenchino, D., Nex, F., Piras, M.: GIMPHI: a new integration approach for early impact assessment. Applied Geomatics **3**(4), 241–249. ISSN 1866-9298
18. Fusiello, A.: Visione computazionale. Tecniche di ricostruzione tridimensionale (2013)