Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 15 December 2016 doi:10.20944/preprints201612.0075.v1

Peer-reviewed version available at *ISPRS Int. J. Geo-Inf.* **2017**, *6*, 56; doi:10.3390/ijgi6020056

Article

An Original Application of Image Recognition Based Location in Complex Indoor Environment

Filiberto Chiabrando [1], Youngsu Cho [2], Vincenzo Di Pietra [3,*], Andrea Lingua [3] and Juil Jeon [2]

[1] Department of Architecture and Design (DAD), Politecnico di Torino, Torino, Italy; filiberto.chiabrando@polito.it
[2] Electronic and Telecommunications Research Institute (ETRI), Daejeon, Korea; choys@etri.re.kr; seventhday07@gmail.com
[3] Department of Environment, Land and Infrastructure Engineering (DIATI), Politecnico di Torino, Torino, Italy; andrea.lingua@polito.it
* Correspondence: vincenzo.dipietra@polito.it; Tel.: +39-011-090-7657

Abstract: This paper describes the first results of an Image Recognition Based Location (IRBL) for mobile application focusing on the procedure to generate a Database of range images (RGB-D). In an indoor environment, to estimate the camera position and orientation, a prior spatial knowledge of the surrounding is needed. In order to achieve this objective a complete 3D survey of two different environment (Bangbae metro station of Seoul and E.T.R.I. building in Daejeon – Republic of Korea) was performed using LiDAR (Light Detection And Ranging) instrument and the obtained scans were processed in order to obtain a spatial model of the environments. From this, two databases of reference images were generated using a specific software realized by the Geomatics group of Politecnico di Torino (ScanToRGBDImage). This tool allow to generate synthetically different RGB-D images) centered in the each scan position in the environment. Later, the external parameters (X, Y, Z, ω, ϕ, κ) and the range information extracted from the DB images retrieved, are used as reference information for pose estimation of a set of acquired mobile pictures in the IRBL procedure. In this paper the survey operations, the approach for generating the RGB-D images and the IRB strategy are reported. Finally the analysis of the results and the validation test are described.

Keywords: image recognition bases location; indoor positioning; RGB-D images; LiDAR; DataBase; mobile computing; image retrieval

1. Introduction

In recent years, location-based services (LBS) which use data acquired from mobile devices sensors, begin to be increasingly important factor in several research conducted by the scientific community and from industry as well [1][2][3]. The growing spread and computational power of mobile phones allow to develop new application in many interesting fields with an evident increase of position accuracy and performance [4]. In this scenario, comes out our interest in indoor positioning systems that exploits as unique sensors, those found in everyday smartphones and our attention on densely lived environments that could have some critical issues.

This paper is connected to a project conducted by the Politecnico di Torino (Italy) and the Electronic and Telecommunications Research Institute (E.T.R.I. - Republic of Korea) with the aim of realizing an image recognition based location procedure useful for estimate the position and orientation of an image taken by a mobile device through the extraction of 3D information from a reference image. This approach can be a components of an hybrid navigation solution with IMU data [5][6].

The project is still in progress and at the moment have seen the validation of the first results obtained in two test site: the Bangbae metro station in Seoul and the E.T.R.I research building in Daejeon. The base of IRBL procedure is the matching between each real time acquired smartphone images and a correspondent synthetically generated 3D image extracted by a database (DB), all implemented in an automated procedure. In the next sections the entire workflow will be described: the developed algorithms will be analyzed and a complete description of the activities realized in the test sites with the validation will be reported. In Figure 1 is described the procedure in his sub sequential steps: the 3D data acquisition with a LiDAR instrument, the 3D model generation, the database of RGB-D images realization, the image retrieval with MPEG7 CDVS (Compact Descriptor for Visual Search) and the IRBL algorithms application for positioning. In the next sections the attention will be focused on data acquisition and generation of RGB-D image DB as a fundamental part for the correct application of the procedure.

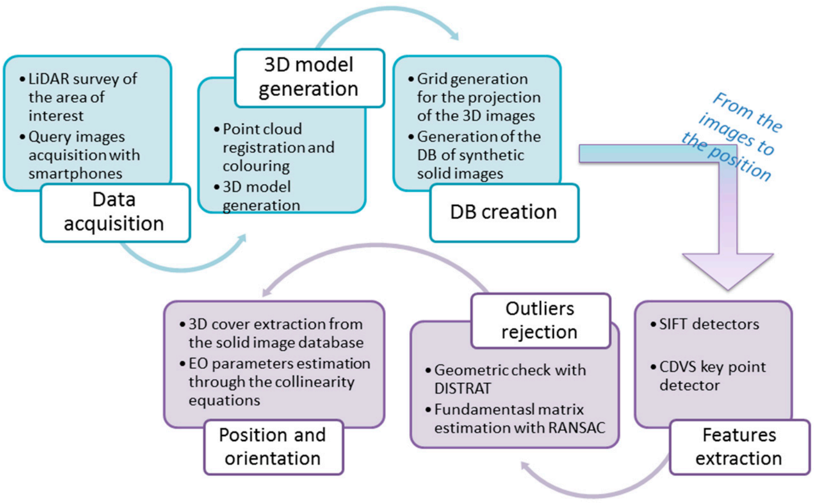


Figure 1. Workflow of the IRBL procedure

2. The proposed procedure for Image Recognition Based Location

The proposed method for IRBL is based on 3 fundamental components:

- An image DB for object area description:
  this DB uses thousands of images recorded in the particular form of RGB-D images;
- A visual search technology:
  in this study, the Compact Descriptors Visual Search (CDVS), patented by TELECOM Italia has been used to identify the reference image extracted from image DB more similar to a query image (acquired by the user with a smartphone);
- A proposed algorithm for IRBL:
  this algorithm is based on a sequence of feature matching and robust outlier rejection able to extract a set of 2D feature, homologous points between reference and query images. This 2D features can be transformed in 3D using RGB-D data for a final photogrammetric space rejection.

2.1. Generation of RGB-D Image Database

A RGB-D image is a classical RGB digital image with internal and external orientation parameters known where a distance between the center of prospective and the object acquired are recorded for each pixel. Therefore distance values are stored in an additional matrix with the same pixel size, number of columns and number of row as the RGB matrix. Additional radiometric information like NIR, MIR, TIR, multispectral or hyperspectral bands can be added in other matrix levels, defining in this way a new image that Authors can define RGB-D. Figure 2 contains a schema of RGB-D structure. With the generation of a database of RGB-D of an indoor environment it's possible to correctly represent the reality.

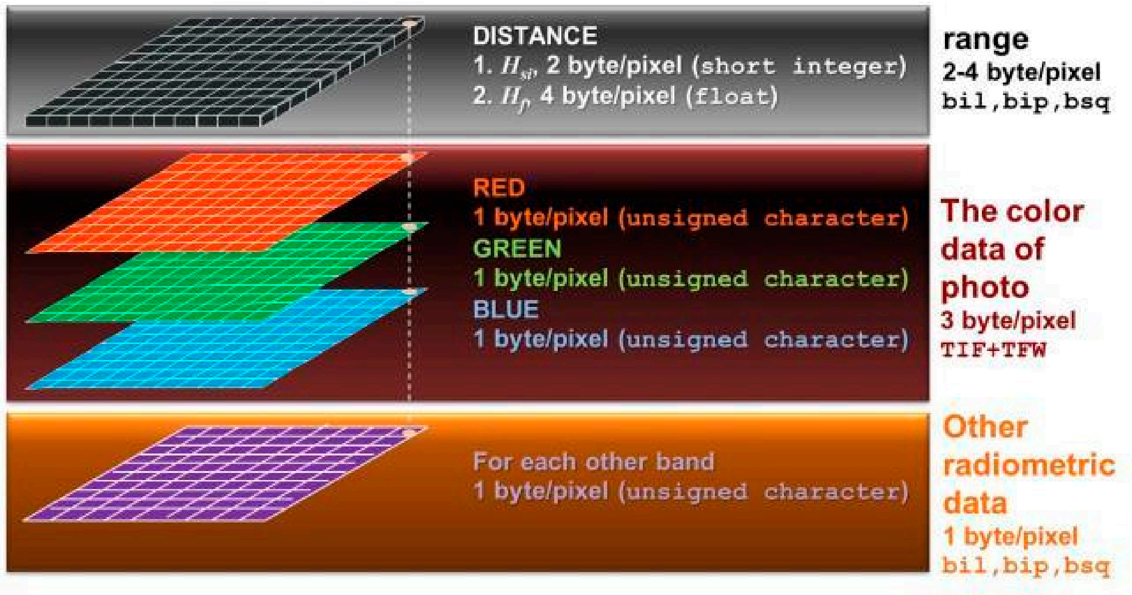


Figure 2. RGB-D structure.

To generate a DB of RGB-D images automatically, a realistic 3D model of the area of interest, with both geometric and color information, is required as input data. This model could be extracted from an existing 3D model, generated by a terrestrial or aerial survey or obtained through a Mobile Mapping System. Once the 3D model is generated, the RGB-D images can be automatically realized by means of the developed software, ScanToRGBDImage, that execute the following steps [7]:

1. an empty image (RGB and range matrix levels) is generated using (n_{col}, n_{row});
2. a subset of colored points (X_i, Y_i, Z_i) with I = 1:n, (n = number of selected points) can be extracted from the original RGB point cloud according to a selection volume that can be defined by a sector of a sphere with:
    a. center in the location of generated RGB-D image;
    b. axis direction coincident with the optical axis of synthetic image;
    c. radius R;
    d. amplitude defined by an angle (≤ 90°) that is half the cone angle measured from direction axis;
3. for each selected colored point, a distance d_i respect the location of the generated image is calculated:

$$d_i = \sqrt{(X_i - X_0)^2 + (Y_i - Y_0)^2 + (Z_i - Z_0)^2} \quad (1)$$

4. each selected RGB point is projected on the synthetic image defining its image coordinates (ξ_i, η_i) by means of the internal and external orientation parameters inside the collinearity equations:

$$\xi = \xi_0 - c\frac{r_{11}(X - X_0) + r_{21}(Y - Y_0) + r_{31}(Z - Z_0)}{r_{13}(X - X_0) + r_{23}(Y - Y_0) + r_{33}(Z - Z_0)} \quad (2)$$

$$\eta = \eta_0 - c\frac{r_{12}(X - X_0) + r_{22}(Y - Y_0) + r_{32}(Z - Z_0)}{r_{13}(X - X_0) + r_{23}(Y - Y_0) + r_{33}(Z - Z_0)} \quad (3)$$

where $(r_{11}, r_{12}, r_{13}, r_{21}, r_{22}, r_{2,3}, r_{31}, r_{32}, r_{33})$ are the coefficients of a 3x3 spatial rotation matrix depending from the camera attitude (ω, φ, K)

Peer-reviewed version available at *ISPRS Int. J. Geo-Inf.* **2017**, *6*, 56; doi:10.3390/ijgi6020056

$$\mathbf{R}_{\omega\varphi\kappa} = \begin{pmatrix} \cos\phi\cos\kappa & -\cos\phi sen\kappa & sen\phi \\ \cos\omega sen\kappa + sen\omega sen\phi\cos\kappa & \cos\omega\cos\kappa - sen\omega sen\phi sen\kappa & -sen\omega\cos\phi \\ sen\omega sen\kappa - \cos\omega sen\phi\cos\kappa & sen\omega\cos\kappa + \cos\omega sen\phi sen\kappa & \cos\omega\cos\phi \end{pmatrix} \tag{4}$$

5. the image coordinates (ξ_i, η_i) are converted in pixel coordinates (c_i, r_i) using:

$$c_i = \frac{\xi_i}{d_{pix}} + \frac{n_{col}}{2} \tag{5}$$

$$r_i = -\frac{\eta_i}{d_{pox}} + \frac{n_{row}}{2} \tag{6}$$

6. the RGB values of each point are wrote inside the cells of the image matrix in the position (c_i, r_i);
7. the distance value d_i is wrote inside the cell of range image matrix in the position (c_i, r_i);
8. at the end of the procedure, pixels still void are filled by means of an interpolation algorithm based on nearest filled pixels.

After the process, ScanToRGBDImage generate a set of images with the information about position and attitude, i.e. the RGB-D images database.

The Figure 3 shows an example of a set of RGB-D images connected to a scan position in Bangbae metro station (X= 322920.858, Y = 4150175.414, Z = 45.967 in meters – UTM-WGS84, 52S)

(a)

(b)

Figure 3. (a) Example of six RGB-D images generated with the software ScanToRGBDImage in RGB visualization. (b) Example of six RGB-D images in depth map visualization.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 15 December 2016 doi:10.20944/preprints201612.0075.v1

Peer-reviewed version available at ISPRS Int. J. Geo-Inf. 2017, 6, 56; doi:10.3390/ijgi6020056

2.2. The Compact Descriptor Visual Search

The goal of the retrieval procedure is to select a reference image out of the images DB (Figure 4) with the highest level of similarity with the image acquired by the smartphone camera, target of positioning procedure. For the retrieval procedure the adopted solution is defined by MPEG7 CDVS (Compact Descriptors for Visual Search) [8] with minor optimization. To select out of a DB the most similar image the following operations have been defined by MPEG7 -CDVS:

1. local descriptors in query and database images are extracted and compressed
2. the images are preliminary ranked based on global descriptors [9] similarity score between the query image. Global descriptors provide a statistical representation of a set of most significant local descriptors extracted from the two images. As a result of global descriptor preliminary screening, a number of potentially similar images are than selected out of the DB.
3. for the selected best ranked images by the global descriptor similarity test, the pairwise matching procedure between the extracted key points in couple of images is executed, trying to couple similar key points present in both images. For each feature descriptor of the query image, one and only one similar feature descriptor is searched in each single images part of the DB (Figure 4).
4. the matched key points are validated by a geometry check 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.

Based on a statistical model, a set of good matches can be ranked by a similarity score given by:

5. the number of correct pairwise key points out of DISTRAT check;
6. the reliability of each selected match, given by the distance ratio between the first and the second closest descriptors detected in the reference image.

Due to the potential large number of images in the DB, to speed up the retrieval process CDVS uses compressed descriptors [10] to speed up the entire process. For this reason only a limited number of key points are used in the image search procedure and moreover the CDVS gives more priority to the points located in the center of the image. It's evident that in some common view, the center of the picture represents the infinite point of the prospective view so the selected points could be far away from the camera, causing a loss of accuracy in the next step of location. In order to enhance the accuracy level of location procedure the criteria for ranking and selecting the key points should be modified: in particular there is a need for homogeneous distribution of key points in the overall picture, not giving priority to those concentrated in the image center.

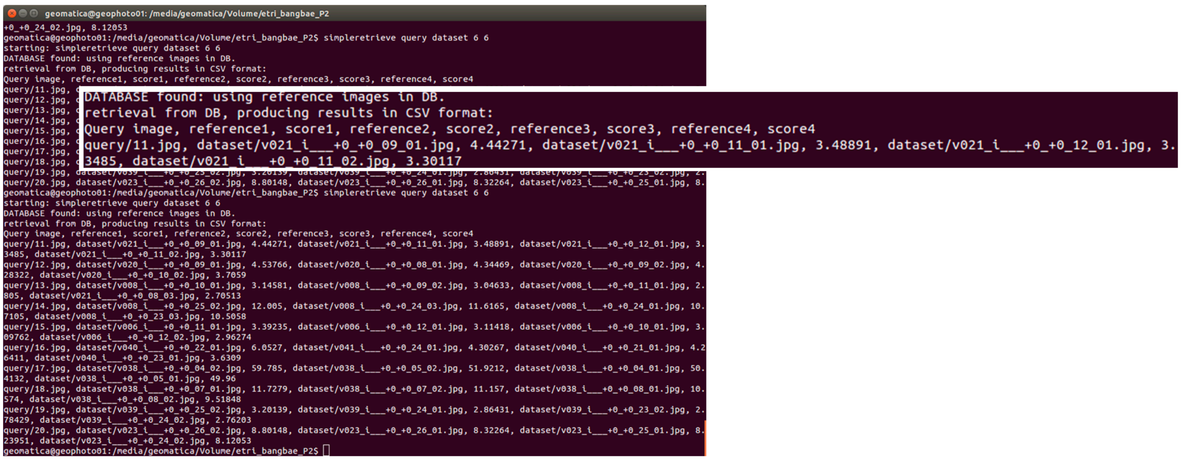


Figure 4. Example of image retrieval with ranking between query and reference images.

2.3. The Image Recognition Based Loction algorithm

Once the retrieval of the reference image is completed, from the image is possible to extract the 3D information of the selected features in order to estimate the camera parameters (position and

Peer-reviewed version available at ISPRS Int. J. Geo-Inf. 2017, 6, 56; doi:10.3390/ijgi6020056

attitude) of the query image. In details, those 3D information are stored inside the DB of RGB-D images where for each pixel the distance (range) of the obstacle depicted in the image is reported, together with internal and external orientation parameters (IO/EO). The key points and related features are extracted from query and reference images using a state of the art solution [11] that allow a preliminary association between key-points of the two images. After that an high percentage outliers rejection is executed according to a new proposed two step approach: at first good matches are selected using DISTRAT algorithm [12][13][14], using a geometric check based on the distances ratio between pairs of points in the two analyzed images, then a RANSAC check is executed over a quality improved set of matches. The proposed outlier rejection approach, when applied to real working conditions, reduces the processing time by a factor of 10, with respect to the use of a standard RANSAC approach [15]. Finally, camera parameters are estimated, based on 3D information available on the reference image for the selected set of key-points pairs according to the collinearity equations [16].

In order to analyze the detail of the processing the next list specify each step of the IRBL algorithm :

1. extraction of features from query and reference images using SIFT detector [17] or CDVS key point detector;
2. key points matching procedure where only query image key points that have one and only one similar descriptors among key points in reference image are selected, according to an approach slightly modified with respect to the one proposed in [11];
3. a geometric check (DISTRAT) is used for a coarse preliminary rejection of matched outliers, the use of DISTRAT is required to speed up outliers rejection procedure;
4. given the set of common features selected out of DISTRAT geometric check, the fundamental matrix between query image and reference image is estimated with a RANSAC procedure, allowing to exclude remaining outliers out of DISTRAT check. RANSAC is a robust iterative method to estimate parameters of a mathematical model from a set of observed data which contains outliers, in DISTRAT output few percentiles of outliers are present in the selected set of common features. RANSAC is a non-deterministic algorithm in the sense that it produces a reasonable result only with a certain probability, with this probability increasing as more iterations are permitted. The preliminary use of DISTRAT reduces the percentage of outliers from 70% order to few percentiles, this allows to dramatically 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);
5. the common features are transformed into 3D information by the RGB-D image associated with the image derived from the three-dimensional model of the scene;
6. it's estimated the DLT (Direct Linear Transformation) associated with the reference image based on the support points obtained through the 3D features extracted in the previous step [18], with this procedure camera parameters position orientation, focal length and sensor position is roughly estimated;
7. the 11 DLT parameters, that are a mathematical combination of physical parameters, combined to linearize the equations system are decoded to obtain the EO parameters in the first approximation;
8. rejection of outliers not detected by step 3 and 4 are processed by a data snooping process [19]: for the given 11 DLT estimations, the post fit residuals are calculated in terms of the distance between the projection of the solid point on the query image pair and matched key point coordinates, if the largest residual exceeds a threshold the worst point is discarded and the DLT parameters are estimated again;
9. using the collinearity equations the EO parameters are refined [18][16], this step requires to know the focal length out of a calibration process;

The reliability of the final estimated location can be validated using the variance covariance matrix of the collinearity equation [19] and checking again the post fit residuals;

This algorithms has been actually implemented in MATLAB environment.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 15 December 2016 doi:10.20944/preprints201612.0075.v1

Peer-reviewed version available at ISPRS Int. J. Geo-Inf. 2017, 6, 56; doi:10.3390/ijgi6020056

3. Data acquisition and processing for image database construction

The research project between Politecnico di Torino and Electronic and Telecommunication Research Institute is based on the validation of the proposed procedure on two different test sites that has been chosen in order to have two different indoor scenario with some specific issues. The first environment, the Bangbae metro station of Seoul (Republic of Korea), it's an important public infrastructure of interest, where a LBS can better express its usefulness. It presents various indoor spaces with different furniture, but also a railway floor very repetitive. It is also very populated, an important issues in a IRBL system. The second test site is the research department of E.T.R.I. building in Daejeon (Republic of Korea) where, according to the function (research office), the internal areas are repetitive: each floor has the same aisle with the same color and the same furniture. The reason of the different scenario is based on the evaluation of the procedure of indoor localization in noisy areas (very popular with a lot of people), and in similar areas where from a first view is difficult to find differences between different floor (Figure 5).

From the operative point of view, the first step of work was the realization of a complete survey of the two test area through the use of traditional LiDAR instrument and procedure [21]. In order to guarantee a continuity of the data in all the environment, also several images for a typical photogrammetric approach based on Structure from Motion (SfM) algorithms were acquired with the idea of combining the data in case of loss of information [22]. As the LiDAR acquisition was suitable for the entire representation of the two environment, the photogrammetric elaboration was actually not used for the generation of the RGB-D database.

Another aspect that need to be underline is is that actually the survey at E.T.R.I. building was not geo-referenced using a topographic network. This lack does not degrade the indoor positioning procedure that will present in this case a relative reference of the camera towards the surrounding environment.

(a) (b)

Figure 5. (**a**) An indoor Bangbae station view; (**b**) and a typical aisle in ETRI building.

For the metric survey of Bangbae metro station, first of all a general topographic network of the area and the surrounding was realized in order to define a common reference coordinates system. In this case a mixed GNSS and Total Station (TS) survey strategy were employed. The network was realized on three main level of the subway Station. The GNSS measurement naturally were acquired in outdoor condition, furthermore the 2 vertexes were connected to the level -1 and -2 with traditional Total station (TS) measurements, as shown in Figure 6 (a and b). For the GNSS survey a Geomax Zenith 35 receiver was employed, for the TS network a Leica TS06 was used.

In post-processing the network has been adjusted with Leica Geo-office and Microsurvey Starnet software using as reference point, the GNSS permanent station of Suwon (a station of the International GNSS Service network). According to the achieved accuracy on each vertex (less than 1 cm), the next step was the survey of the markers positioned on the station area this operation was performed with the TS using traditional side shot measurements. The markers, in this case black and

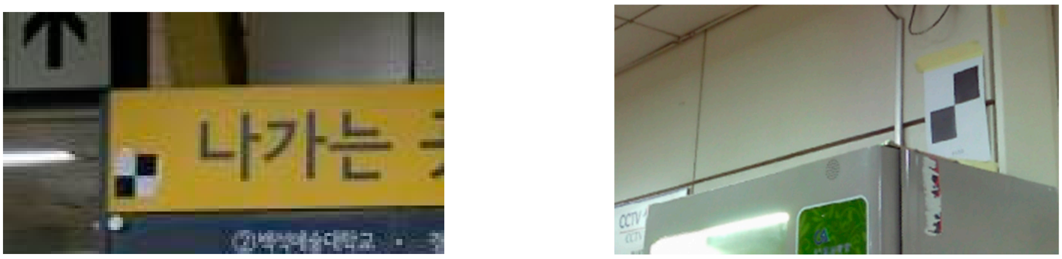
Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 15 December 2016 doi:10.20944/preprints201612.0075.v1

Peer-reviewed version available at *ISPRS Int. J. Geo-Inf.* **2017**, *6*, 56; doi:10.3390/ijgi6020056

white checkerboards, are commonly used for the registration of the scans and for geo-referencing the final model (Figure 6, c).

(a)

(b)



(c)

Figure 6. (**a**) GNSS acquisition (**b**) and Total Station Measurements.
(**c**) An example of two markers positioned in the surveyed area.

Finally for the LiDAR acquisitions two Faro Focus3d X130 were employed. The instrument is a phase shift laser that allow to acquire 3D point clouds with an accuracy of ± 2 mm in the following range: 0.30 – 130 m. During the point cloud acquisition, thanks to the included digital camera, is possible to acquire the images of the scanned area as well. In the test field the acquisition were performed with a resolution of 1/5 (a point each 9 mm at 10 m) and quality of 4x (points measured four times). For the complete LiDAR survey of the Bangbae subway station 114 scans were acquired (55 at Level -1 and 59 at level -2). According to the afore mentioned setting of the scanner, each scan contains 26 million of points approximately, in total about 3 billion of points were measured. The LiDAR data were processed according to the traditional approach [23] using Scene software by Faro, which includes the following main step: point cloud colouring, scan registration and scan georeferencing. Naturally using the markers is possible to evaluate the accuracy of the georeferencing according to the residual on the measured point. The mean RMS on the measured markers (85 were employed) was 1.56 cm. In Figure 7, three views of the complete point cloud (114 merged point clouds).

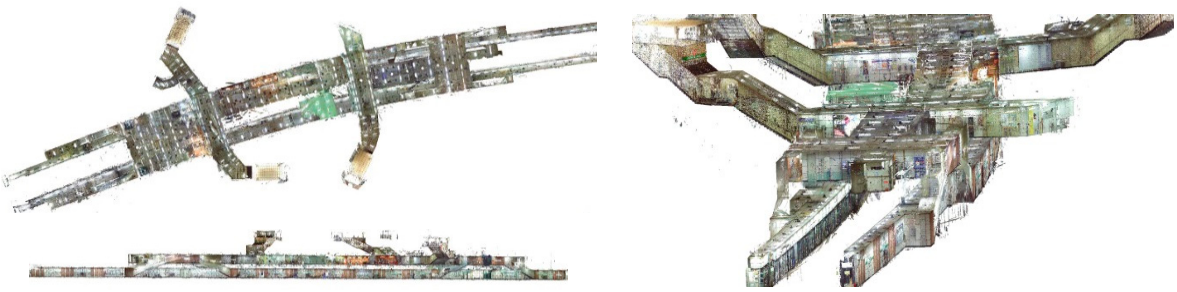


Figure 7. Some views of complete point cloud of Bangbae Station.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 15 December 2016 doi:10.20944/preprints201612.0075.v1

Peer-reviewed version available at ISPRS Int. J. Geo-Inf. 2017, 6, 56; doi:10.3390/ijgi6020056

The ETRI building was only surveyed by the LiDAR in a local reference system. All the acquisitions were realized without the usually required topographic network and without the markers for the registration of the clouds. As a consequence the final point cloud is not located in a known cartographic reference system.

As for the Bangbae Station the LiDAR acquisitions have been performed using the afore mentioned Faro Focus 3d X130 that was used at a quite higher resolution: 1/4 (a point each 5 mm at 10 m) with the same quality (4x) of the Bangbae settings. The complete building (seven floors) was completely scanned with 111 scans, that according to the setting of the scanner delivered each scan with 40 million of points approximately, in total about 4.5 billion of points were measured.

Also in the case of ETRI building the data were processed using Scene software by Faro but the scan registration was realized using the cloud to cloud approach [24]. This approach based on the Iterative Closest Point (ICP) well known algorithms [25][26][27] has been implemented starting from the version 5.5 of the Scene Software and nowadays working very well in the pipeline of Scene LiDAR data processing. Using this approach firstly is important to define an initial setting of the several scan position and after the initial position the algorithm allow to improve the position of the adjacent scans using the shape of the different clouds. In terms of accuracy in this case is possible to understand only the discrepancy between the adjacent clouds that in the case of the ETRI bulding was for all the registered scans under 1 cm. Naturally, as is reported above with the cloud to cloud approach the geo-referencing were not allowed since no GCPs were measured on the area. All the point clouds were referenced to a local system that start from an arbitrary position of the first achieved scan in the building. In Figure 8 two views of the complete point cloud are showed.

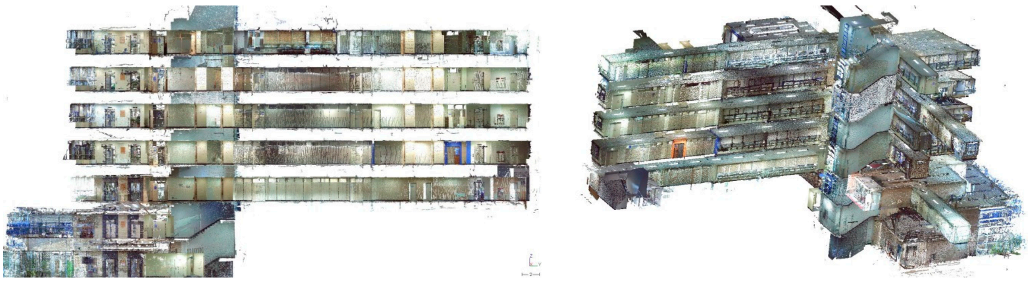
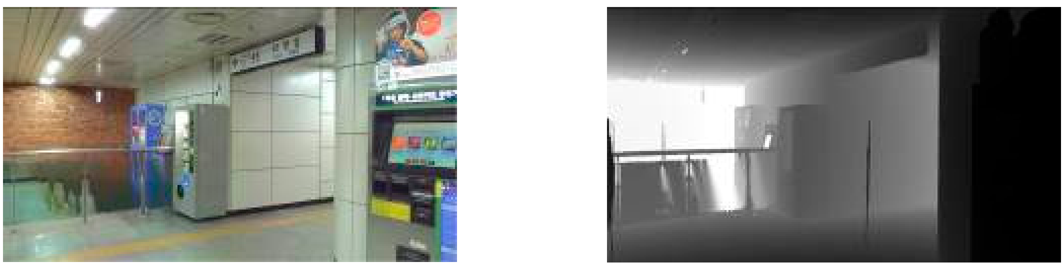


Figure 8. Lateral view and 3D view of the ETRI building point cloud.

The final step for both the building was the generation of the *.xyz file, this ASCII file contain the X, Y, Z coordinate of each point and the R, G, B value extracted from the LiDAR internal camera. This file were used for the generation of the RGB-D images .

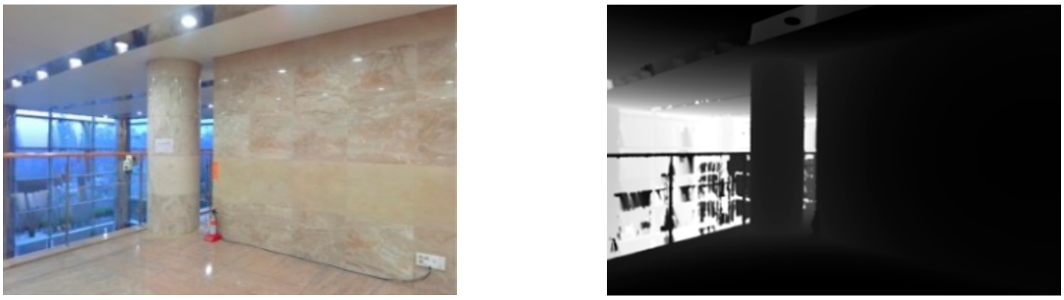
The synthetic RGB-D image can be automatically generated by means of ScanToRGBDImage software tools (developed by the Geomatics research group of the Politecnico di Torino in Intel Visual Fortran) starting from the LiDAR point cloud. ScanToRGBDImage software generate a set of "synthetic" *.JPG images with correspondent range image (Figure 9). For each scan position, 96 images have been generated: 32 horizontal direction for 3 different inclination of 0, 10°, 20° respect the horizontal plane with 2500x1600 pixels, 3 μm of pixel size, focal length of 4.667 mm. For the Bangbae DB almost 9700 RGB-D images have been produced in about 36 hours of batch processing time with desktop computer (i7 5600 U 2.66 GHz 32 Gb RAM) while for the E.T.R.I. building 10700 images have been produced in about 40 hours with a PC with the same characteristic.



(**a**) (**b**)

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 15 December 2016 doi:10.20944/preprints201612.0075.v1

Peer-reviewed version available at *ISPRS Int. J. Geo-Inf.* **2017**, *6*, 56; doi:10.3390/ijgi6020056



(**c**) (**d**)

Figure 9. (**a**) An example of RGB image in Bangbae (**b**) and the correspondent range image. (**c**) An example in E.T.R.I building (**d**) and the correspondent range image

4. Smartphone image acquisition for retrieval procedure and definition of ground truth

On site, with aim to the purpose to evaluate the retrieval procedure, several pictures of the test areas has been taken with commercial mobile devices, in particular the Samsung Galaxy A5, Galaxy S5 and Galaxy S7 Edge were used in order to compare different sensors.

The device used for the acquisitions are smartphones with an integrated non-metric camera that requires a calibration through analytical procedures to define the characteristics of the optical-digital system, to evaluate the distortion parameters and other errors. In particular, the calibration allow to evaluate the effects of the radial and tangential distortion of the sensor, that are involved in the definition of the camera internal orientation using the collinearity equations. However, as approximation, it's possible to consider only the effects of the radial distortion, expressed in this case by two parameters K_1, K_2.

Knowing the object coordinates of some points acquired by the camera it is possible to obtain the unknown parameters solving the bundle-adjustment calculation. The unknown are the six external orientation parameters of the images and the five parameters of the camera (ξ_0, η_0, c, K_1, K_2). The object on which the calibration usually is made is a calibration grid, specifically made, in which the coordinates of the grid points are known with high precision. This procedure is known as Self-Calibration of the camera sensor.

In order to include the calibration process in the IRBP procedure, the 'CameraCalibrator' tool of MATLAB was tested. This tool is able to estimate intrinsic, extrinsic and lens distortion parameters in order to remove the distorsion effects and to reconstruct the 3D scene. The application requires the use of a specific checkerboard pattern that must not be square (Figure 10). The images of the pattern must be acquired with a fix zoom and focus. The calibration requires at least three images, but it is suggested the use of 10-20 images from different distances and orientations in order to obtain best results. The tool's Data Browser displays the images with the detected points, thanks to the not square checkerboard pattern, a reference system is also defined using the different number of square in the two directions. The calibration algorithm assumes a pinhole camera model and after the processing the applications displays the results and the accuracies of the process.

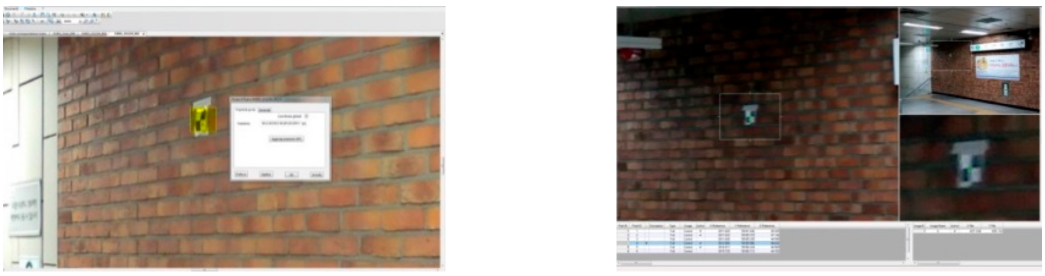
Figure 10 Some images of the checkerboard acquired by a smartphone for camera calibration

In this work the self-calibration was made on the three different smartphone used in the procedure of IBRP and the results are shown above (Table 1).

Table 1. Internal calibration parameters

	Samsung Galaxy A5	Samsung Galaxy S5	Samsung Gal. S7 Edge
Pixel size	1 μm	1,14 μm	1,4 μm
Focal lengths fx [pixels]	3706.0	4290.8	3168.7
Focal lengths fy [pixels]	3722.6	4282.8	3178.9
Princ. Point ξ0 [pixels]	2070.1	2667.8	1995.3
Princ. Point η0 [pixels]	1135.4	1477.8	1204.4
Radial distortion K1	0.1386	0.1148	0.3444
Radial distortion K2	-0.2587	0.0100	-0.6117
Focal length [mm]	3.714	4.801	4.446
Princ. Point ξ0 [mm]	0.006	0.014	-0.087
Princ. Point η0 [mm]	0.026	0.018	-0.179

After the internal calibration, in order to define the position and attitude of the acquired smartphone images and then use it as "groundtruth", a photogrammetric process was employed. In the case of single shot acquisition is possible to perform the single image adjustment (or pyramid vertex) that allow to evaluate the coordinate of the acquisition point X_0, Y_0, Z_0 and the asset as well (ω, φ, k). For this task, at least six collinearity equations must be written which mean that for performing this process three plano-altimetric GCPs are required. The coordinates of the GCPs were extracted directly from the previous LiDAR point clouds using Scene. Firstly a well visible point was selected on the smartphone image, afterwards the same point was measured on the point cloud and the coordinate were extracted. These values (coordinates) were used as GCPs in the employed photogrammetric software (Figure 11). In the present research Erdas Imagine by Hexagon Geospatial was employed for the process. In order to have an accurate control of the results at least six points were used as GCPs. The final precision for all the analyzed images was around 5 cm for the position and around 10 mgon for the angular values. The total number of query images used for the check were: 20 images for Bangbae station (10 images for each floor) and 10 images for E.T.R.I. building.



(**a**) (**b**)

Figure 11. (**a**) GCP coordinate extraction from LiDAR data, (**b**) GCP measurement in Erdas.

As stated in Cap. 2.2. the visual search technology allow to retrieve the best reference images form the RGB-D images database and ranked it with a priority score. These procedure was applied on the selected query images for both the test site and the results of the extraction are shown in Table 2 for Bangbae metro station and in Table 3 for E.T.R.I. building. In these tables the obtained scores of the 1^{st} ranked image selected by the CDVS server are reported. This is the best solutions from the 3 possible candidates proposed by CDVS. As is shown in Table 2 the score are always greater than 3, indicating a quite good solutions. In most cases, score is greater than 5 indicating a good solution. The time spending for the query retrieval process is estimated in about 3 seconds. In the second test site, 10 check images have been acquired by a smartphone Samsung S7. The results of reference images extraction using CDVS are greater than 3, indicating a quite good solutions, excluding image n. 2 (score=2.54) that was ignored since the result IRBL solution was incorrect.

Table 2. Results of reference image extraction from the image DB of Bangbae Station using CDVS

Query im.	Reference image 1	Sc.
query/1.jpg	dataset/b022_i___+0_+0_24_02.jpg	8.3
query/10.jpg	dataset/b011_i___+0_+0_18_02.jpg	20.2
query/2.jpg	dataset/b012_i___+0_+0_09_01.jpg	8.9
query/3.jpg	dataset/b002_i___+0_+0_10_01.jpg	7.4
query/4.jpg	dataset/b004_i___+0_+0_10_01.jpg	17.7
query/5.jpg	dataset/b006_i___+0_+0_07_03.jpg	15.3
query/6.jpg	dataset/b013_i___+0_+0_27_01.jpg	31.9
query/7.jpg	dataset/b013_i___+0_+0_25_01.jpg	9.5
query/8.jpg	dataset/b007_i___+0_+0_16_03.jpg	4.5
query/9.jpg	dataset/b011_i___+0_+0_15_02.jpg	48.7
query/11.jpg	dataset/v004_i___+0_+0_15_02.jpg	42.2
query/12.jpg	dataset/v020_i___+0_+0_09_01.jpg	4.5
query/13.jpg	dataset/v008_i___+0_+0_10_01.jpg	3.1
query/14.jpg	dataset/v008_i___+0_+0_25_02.jpg	12.0
query/15.jpg,	dataset/v006_i___+0_+0_11_01.jpg,	3.3
query/16.jpg,	dataset/v040_i___+0_+0_22_01.jpg,	6.0
query/17.jpg,	dataset/v038_i___+0_+0_04_02.jpg,	59.7
query/18.jpg,	dataset/v038_i___+0_+0_07_01.jpg,	11.7
query/19.jpg,	dataset/v039_i___+0_+0_25_02.jpg,	3.2
query/20.jpg,	dataset/v023_i___+0_+0_26_02.jpg,	8.8

Table 3. Results of reference image extraction from the image DB of ETRI building using CDVS.

Query im.	Reference image 1	Sc.
query/1_01.jpg	dataset/s020_i___+0_+0_25_01.jpg	7.27
query/1_02.jpg	dataset/s021_i___+0_+0_12_02.jpg	2.54
query/2_03(2).jpg	dataset/s011_i___+0_+0_12_03.jpg	4.54
query/3_03(2).jpg	dataset/d011_i___+0_+0_31_01.jpg	6.02
query/3_05.jpg	dataset/d008_i___+0_+0_29_01.jpg	8.51
query/4_01.jpg	dataset/s066_i___+0_+0_18_01.jpg	6.46
query/4_03(3).jpg	dataset/d008_i___+0_+0_17_01.jpg	6.56
query/5_01(2).jpg	dataset/d012_i___+0_+0_03_01.jpg	9.6
query/5_04(2).jpg	dataset/d011_i___+0_+0_19_01.jpg	6.25
query/5_06.jpg	dataset/d012_i___+0_+0_03_01.jpg	8.10

5. Results

After the data acquisition and processing useful for the DB generation, the image retrieval and the ground truth definition, the next step was the applying the IRBL algorithmin in order to define the position and orientation of the acquired smartphone camera . This steps was applied on the 30 acquired images using as query and was complete automatically using the proposed algorithm described in Chapter 2.3.

5.1 Accuracy evaluation

The images have been located in few seconds using the RGB-D images extracted by CDVS, as reference image.

The results for the Bangbae test site have been summarized in Table 4 (main floor, A5 smartphone) and Table 5 (train floor, S5 smartphone) according to the Table is possible to notice that:

- the IRBL and ground truth results for the best solutions (image n. 4 and n. 17) and the worst solutions (images n. 8 and n. 16) for 2 analyzed floors;
- the discrepancies between IRBL solutions and ground truth for the best solutions and the worst solutions expressed by the differences from the 6 external orientation parameters of IRBL and ground truth results;
- some statistical parameter (min, max, mean, and root mean squares error=RMSE) of discrepancies.

Peer-reviewed version available at ISPRS Int. J. Geo-Inf. 2017, 6, 56; doi:10.3390/ijgi6020056

As general comments is possible to state that:

- the discrepancies in X, Y, Z are always lower than 1.5 m in absolute value, excluding the gross error of image 12 in X. According to the shape of train floor (long and narrow) some critical problem of incorrect geometry of feature points were founded;
- the standard deviations of the discrepancies in X is about 1 m and in Y is about 50 cm, this is the XY quality of DB;
- the standard deviations of the discrepancies in Z are of about 40 cm, this is the Z quality of DB;
- the angular values are estimated with a precision of about 10 gon;
- the estimated averages are not significant for all the parameter, than there are not any systematic estimation

Calculating the relative frequencies of 3D discrepancies, it is possible to define that the 25% of IRBL solutions have discrepancies less than 0.5 m, the 65% less than 1 m, the 95% less than 2 m. An example of good solution has shown in Figure 12 with a quite good solution in Figure 13.

Table 4. Accuracy in Bangbae station with Samsung A5

im	np	X [m]	Y [m]	Z [m]	ω [gon]	φ [gon]	k [gon]
				IRBL algorithm			
4	71	3007.48	50177.3	46.10	147.638	309.179	49.148
8	19	3021.65	50178.60	46.32	101.725	373.332	198.649
				Ground truth			
4		3007.43	50177.38	46.12	-59.216	-110.713	242.225
8		3020.73	50179.37	46.15	305.1	-45.201	204.819
				Discrepancies			
		ΔX [m]	**ΔY [m]**	**ΔZ [m]**	**ω [gon]**	**φ [gon]**	**k [gon]**
4		0.055	-0.019	-0.015	6.854	19.891	6.922
8		0.930	-0.764	0.171	-3.375	18.533	-6.170
min		-1,001	-0,764	-0,480	-21,016	-2,727	-6,170
max		0,930	0,548	0,171	11,442	30,885	21,047
mean		0,195	-0,005	-0,089	-1,881	10,884	4,295
RMSE		0,573	0,458	0,182	9,228	13,367	7,998

Table 5. Accuracy in Bangbae station with Samsung S5

im	np	X [m]	Y [m]	Z [m]	ω [gon]	φ [gon]	k [gon]
				IRBL algorithm			
16	19	2959.17	50171.21	42.28	226.351	84.371	177.549
17	314	2983.93	50177.03	41.33	101.543	-56.282	-0.14
				Ground truth			
16		2960.704	50171.93	41.016	244.879	74.385	196.816
17		2984.193	50177.19	41.322	101.869	-55.63	0.233
				Discrepancies			
		ΔX [m]	**ΔY [m]**	**ΔZ [m]**	**ω [gon]**	**φ [gon]**	**k [gon]**
16		-1.52	-0.71	1.272	18.528	9.986	-19.267
17		-0.26	-0.16	0.008	-0.326	-0.652	-0.373
min		-1.52	-0.71	-0.053	-18.989	-0.652	-19.267
max		1.08	1.44	1.272	18.528	9.986	13.697
mean		0.38	0.32	0.206	1.039	2.233	-1.492
RMSE		0.91	0.61	0.417	10.849	3.242	10.95

For the E.T.R.I. building in the present paper only the discrepancies are reported in Table 6. Excluding image n. 2, the discrepancies between IRBL solutions and ground truth are similar to Bangbae Station with some differences: in 2 cases the results presents outliers in Y direction (over 3 m of discrepancies) due to a low number of feature points rather close. In other 2 cases, there are discrepancies in Z very high (over 8 m) due to similarity between different floors in ETRI building cause of an wrong retrieval (Figure 14).

Table 6. Accuracy in ETRI building with Samsung Galaxy S7 Edge.

	Discrepancies					
	ΔX[m]	**ΔY[m]**	**ΔZ[m]**	**ω[gon]**	**φ[gon]**	**κ[gon]**
min	-0,83	-3,57	-0,81	-9,414	-1,705	-4,938
max	0,34	3,73	0,56	2,666	3,986	4,446
mean	-0,26	-0,27	0,01	-1,193	1,052	0,334
RMSE	0,34	2,13	0,42	3,593	1,995	3,379

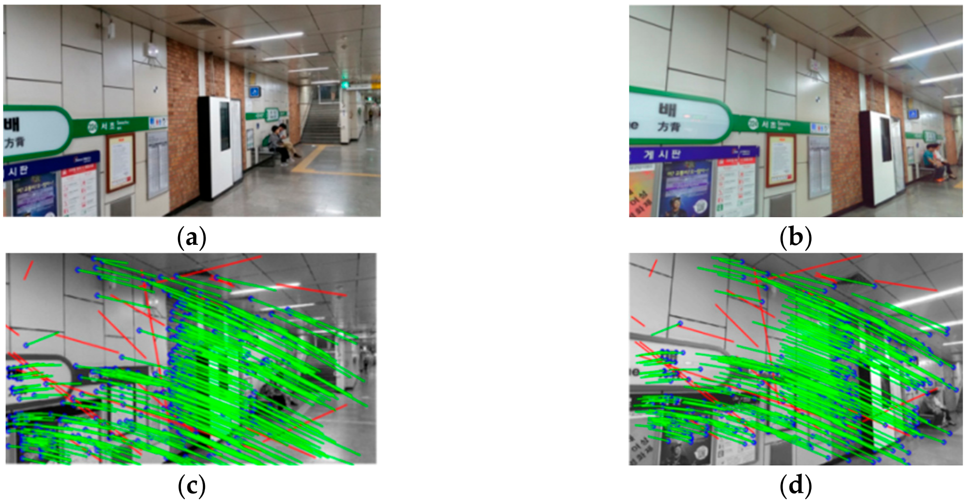


Figure 12. A good solution for Bangbae station, image n. 17: (**a**) query image, (**b**) reference image from DB, (**c**) query image and (**d**) reference image with used feature points; In red the matched points that have been rejected.

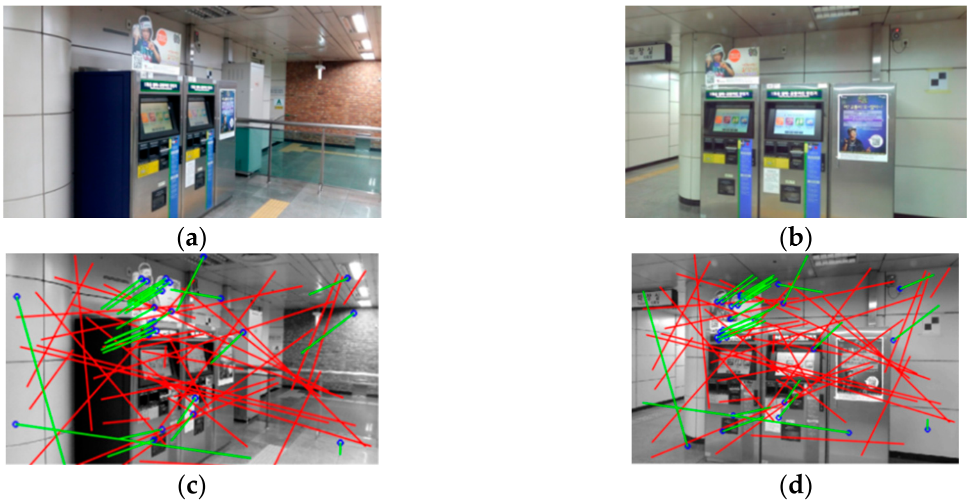
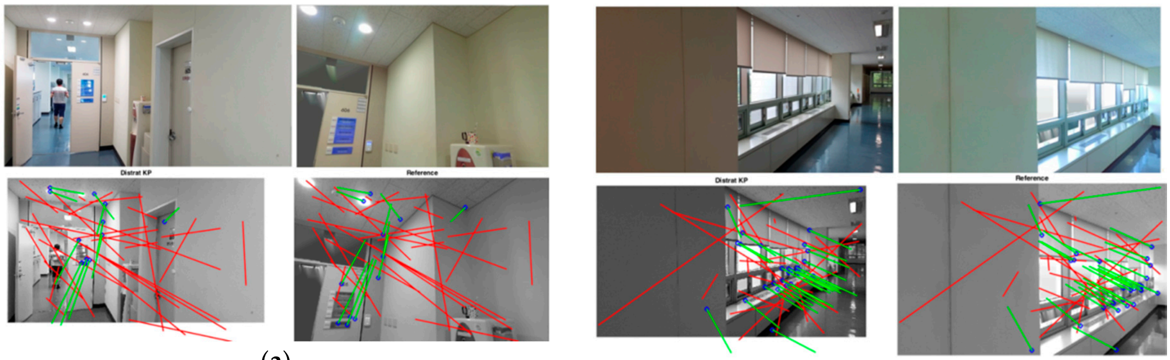


Figure 13. An example of quite good solution for Bangbae station, image n. 8: (**a**) query image, (**b**) reference image from DB, (**c**) query image and (**d**) reference image with used feature points; In red the matched points that have been rejected.

Peer-reviewed version available at ISPRS Int. J. Geo-Inf. 2017, 6, 56; doi:10.3390/ijgi6020056



(a) (b)

Figure 14. (**a**) Worst solution for ETRI Building due to a low number of feature point rather close; (**b**) Worst solution in Z for ETRI Building due to image similarity from different floors.

5.2. IRBL reliability

In the proposed procedure the estimation of Fundamental matrix between reference and query images uses a robust estimation algorithms based on RANSAC. This technique can be causes a certain variability of final solution based on the number of sample extractions used. For some images the IRBL procedure has been repeated 20 times to define the reliability of solutions (min. max. mean. RMSE) for the 6 external orientation parameters. The results are reported in Table 7 to demonstrate the substantial reliability of estimated solutions. In particular the RMSE correspondent to the nominal precision of this method are less than the estimated accuracy..

Table 7. Reliability analysis in Bangbae station data set.

	X[m]	Y[m]	Z[m]	ω[gon]	φ[gon]	k[gon]
			An example: Image n.11			
mean	3021.42	50173.61	46.21	1.385	-0.905	-0.116
min	3020.87	50173.39	46.48	1.201	-1.069	-0.215
max	3021.89	50174.07	46.69	1.648	-0.804	0.000
RMSE	0.26	0.28	0.23	0.205	0.071	0.103
			Summary of all the query image			
RMSE	0.12	0.18	0.11	0.053	0.051	0.158

Conclusions and future works

The procedure that is reported in the article is nowadays well tested and allow to state that especially the first part of the proposed workflow could be successfully performed without problems naturally according to the area that need to be surveyed. The realized software is stable and efficiently work with large dataset as well; a new version in C++ is under development in order to speed up the computational time for the generation of the RGB-D images. Actually the resolution of the generated images is connected to the LiDAR model and especially to the on board camera that is used for acquiring the RGB information after the scans.

The developed CDVS procedure is efficient and actually work without any problem and deliver excellent results very quickly during the retrieval process.

The approach of the IRBL procedure is new and delivery excellent results as is possible to notice from the accuracy results reported, is possible to state that using this approach is possible to obtain a correct indoor positioning using smartphoneimages with metrical/sub-metrical accuracy for the position and a few gons for attitude. IRBL is able to obtain a correct solution also in complex condition (noise due to people, narrow corridors, artificial light and other environment problems).

Peer-reviewed version available at ISPRS Int. J. Geo-Inf. 2017, 6, 56; doi:10.3390/ijgi6020056

Actually since the IRBL procedure is under testing the available application is only available in Matlab and need to be improved using other programming languages in order to obtain a products with higher performance in terms of usability and speed.

The research is still in progress, first of all an integration of the survey operation connecting to photogrammetric data and LiDAR data is under evaluation. Moreover, according to the next steps of the common research project (Politecnico and E.T.R.I) a server will be realized with the CDVS technology in order to allow to realize the retrieval using the web.

Finally the IRBL algorithm will be further developed with the realization of some API (Application programming interface) that will allow to extract the needed information for the indoor positioning and deliver the final results directly on the used smartphone.

Acknowledgments: This work was supported by Institute for Information & communications Technology Promotion (IITP) grant funded by the Korea government (MSIP) (R101-16-0306, Development of Autonomous Location Infrastructure DB Update Technology based on User Crowd-sourcing).

Author Contributions: All the author have equally contribute to the research and to the realization of the present article.

Conflicts of Interest: The authors declare no conflict of interest

References

1. J.Z. Liang. N. Corso. E. Turner. and A. Zakhor. "Image-based positioning of mobile devices in indoor environments". in Multimodal Location Estimation of Videos and Images. Springer International Publishing. 2005. pp. 85-99.
2. J. Dong. Y. Xiao. M. Noreikis. Z. Ou and A. Ylä-Jääski. " Demo: iMoon: Using Smartphones for Image-based Indoor Navigation". in Proceedings of the 13th ACM Conference on Embedded Networked Sensor Systems. 2015. pp. 449-450
3. Y.S. Cho. M. Ji. J.Y. Kim. and J. I. Jeon. "High-scalable 3D indoor positioning algorithm using loosely-coupled Wi-Fi/sensor integration". in 2015 17th International Conference on Advanced Communication Technology (ICACT). IEEE. 2015. pp. 96-99.
4. Dabove, Paolo; Ghinamo, Giorgio; Lingua, Andrea Maria (2015) Inertial sensors for smartphones navigation. In: SPRINGERPLUS, vol. 4, pp. 1-18. - ISSN 2193-1801
5. M. Piras; P. Dabove "Comparison of two different mass-market IMU generations: bias analyses and real time applications". In: 2016 IEEE/ION Position, Location and Navigation Symposium, PLANS 2016, Savannah (USA), April 11-14, 2016. pp. 34-41
6. Dabove, Paolo; Aicardi, Irene; Grasso, Nives; Lingua, Andrea; Ghinamo, Giorgio; Corbi, Cecilia (2016) Inertial Sensors Strapdown Approach for Hybrid Cameras and MEMS Positioning. In: 2016 IEEE/ION Position, Location and Navigation Symposium, PLANS 2016, Savannah (USA), April 11-14, 2016. pp. 994-1000
7. Bornaz L., Dequal S. (2003) - A new concept: the solid image. CIPA 2003 Proceedings of XIXth International Symposium: 169-174
8. Lingua, Andrea; Aicardi, Irene; Ghinamo, Giorgio; Corbi, Cecilia; Francini, Gianluca; Lepsoy, Skjalg; Lovisolo, Piero (2014). The MPEG7 Visual Search Solution for image recognition based positioning using 3D models. In: 27th International Technical Meeting of The Satellite Division of the Institute of Navigation (ION GNSS+ 2014), Tampa, Florida, September 8 - 12, 2014. pp. 2078-2088
9. Visual Standard for Content Description—An Overview, Sikora, IEEE transactions on circuits and systems for video technology, VOL. 11, NO. 6,
10. ISO/IEC JTC1/SC29/WG11/W13564 "Test Model 6: Compact Descriptors for Visual Search", Incheon, Korea, April 2013.
11. Lowe D., 2004. Distinctive image features from scale-invariant keypoints. International, Journal of Computer Vision 60(2): 91-110
12. S. Lepsoy, G. Francini, G. Cordara, and P. P. de Gusmao, "Statistical modelling of outliers for fast visual search," in IEEE International Conference on Multimedia and Expo (ICME), pp. 1–6,IEEE, 2011
13. PCT/EP2011/050994 "Method and system for comparing images"
14. PCT/EP2012/050025 "Method and system for comparing images"
15. Multiple View Geometry in Coputer Vision Second Edition. Richard Hartley and Andrew Zisserman, Cambridge University Press, March 2004.

16. Editor in Chief C. Mc Glone Manual of Photogrammetry 5th edition ASPRS, p 280-281
17. CDVS (2014). ISO/IEC DIS 15938-13 Compact Descriptors for Visual Search
18. Editor in Chief H.M. Karara, Non Topography Photogrammetry Second Edition ASPRS pp46-481
19. W. Baarda, (1968). A testing procedure for use in geodetic networks, Publications on Geodesy 9 (Vol. 2 Nr. 5), Delft, 1968. 97 pagina's.I SBN-13: 978 90 6132 209 2. ISBN-10: 90 6132 209 X, pag. 53-55
20. K. Kraus Photogrammetry: Geometry from Images and Laser Scans, Volume 1, pag 21-29, 184-189
21. C. Balletti, F. Guerra, P. Vernier, N. Studnicka, J. Riegl, S. Orlandini "Practical comparative evaluation of an integrated hybrid sensor based on Photogrammetry and Laser Scanning for Architectural Representation, ISPRS, Int. Arch. Photogramm.Remote Sens, 2004, pp. 12-23
22. F. Noardo "Dense Image matching per il recupero di contenuto metrico da immagini di documentazione e camere non metriche". In: BOLLETTINO SIFET, vol. 1/2015. - ISSN 1721-971X
23. P. Bastonero, E. Donadio, F. Chiabrando, A. Spanò "Fusion of 3d models derived from tls and image-based techniques for CH enhanced documentation. In: ISPRS Ann. Photogramm. Remote sens. Spatial inf. Sci, riva del garda, 23 – 25 june 2014. Pp. 73-80
24. Hao Men, Biruk Gebre, Kishore Pochiraju "Color point cloud registration with 4D ICP algorithm". Conference Paper in Proceedings - IEEE International Conference on Robotics and Automation · June 2011.
25. Besl, Paul J.; N.D. McKay (1992). "A Method for Registration of 3-D Shapes". IEEE Trans. on Pattern Analysis and Machine Intelligence (Los Alamitos, CA, USA: IEEE Computer Society) 14 (2): 239–256.
26. Chen, Yang; Gerard Medioni (1991). "Object modelling by registration of multiple range images". Image Vision Comput. (Newton, MA, USA: Butterworth-Heinemann): 145–155.
27. Zhang, Zhengyou (1994). "Iterative point matching for registration of free-form curves and surfaces". International Journal of Computer Vision (Springer) 13 (12): 119–152.