



POSIZIONAMENTO INDOOR BASATO SU RICONOSCIMENTO DI IMMAGINI. APPLICAZIONE AL CASTELLO DEL VALENTINO (TORINO)

IMAGE RECOGNITION BASED LOCATION FOR INDOOR POSITIONING. APPLICATION AT THE VALENTINO CASTLE (TORINO)

Stefano Angeli

Dipartimento di Ingegneria dell'Ambiente, del Territorio e delle Infrastrutture, Politecnico di Torino, Corso Duca degli Abruzzi 24, Torino, 10129, Italia – stefano.angeli@polito.it

PAROLE CHIAVE: Posizionamento Indoor, LiDAR, immagini RGB-D, Image Recognition Based Location, database.

KEY WORDS: Indoor Positioning, LiDAR, RGB-D images, Image Recognition Based Location, database.

Questo articolo è stato presentato al Premio Giovani Autori SIFET 2019.

RIASSUNTO

L'articolo descrive il risultato dell'applicazione di un metodo di posizionamento basato sul riconoscimento di immagini (Image Recognition Based Location - IRBL), incentrato sulla creazione di un database di immagini solide (RGB-D) per un'applicazione mobile. L'attività descritta è stata realizzata per testare il metodo di posizionamento sviluppato nell'ambito di un progetto di ricerca condotto dal Politecnico di Torino (diretto dal prof. Andrea Lingua) in collaborazione con l'istituto di ricerca sud-coreano ETRI (Electronics and Telecommunications Research Institute) con l'obiettivo di realizzare una procedura di posizionamento indoor basata sul riconoscimento di immagini, per fini legati alla sicurezza in luoghi pubblici.

In una fase preliminare è stato realizzato un rilievo LiDAR (*Light Detection And Ranging*) dell'ambiente nel quale sperimentare il metodo, ottenendo un modello tridimensionale completo necessario per stimare la posizione della camera e i parametri di orientamento. Da questo modello è stato generato un database di immagini sintetiche di riferimento, creando un set di immagini in diverse direzioni e con diverse angolazioni per ognuno dei centri di scansione. Successivamente alla fase di acquisizione dei dati, le informazioni relative alle distanze e i parametri di orientamento esterno (X, Y, Z, ω, φ e κ) estratti dalle immagini presenti nel database, sono state utilizzate nella procedura di posizionamento, come riferimento per la stima della posizione di presa di immagini acquisite con un dispositivo mobile oggetto del test. Di seguito verranno descritte le operazioni di rilievo, il metodo utilizzato per la generazione del database di immagini solide con la successiva procedura di posizionamento e, infine, i risultati ottenuti in seguito alla validazione del test.

ABSTRACT

This paper describes the result of the application of an Image Recognition Based Location (IRBL) method, focused on the creation of a database of solid images (RGB-D) for a mobile application. The activity described was carried out to test the positioning method developed within a research project conducted by the Politecnico di Torino (directed by Prof. Andrea Lingua) in collaboration with the South Korean research institute ETRI (Electronics and Telecommunications Research Institute), with the aim of creating an indoor positioning procedure based on image recognition, for purposes related to security in public places.

In a preliminary stage, a LiDAR (Light Detection And Ranging) survey of the environment was carried out, obtaining a complete three-dimensional model needed to estimate the position of the camera and the orientation parameters. From this model, a database of synthetic reference images was generated, creating a set of images in different directions and with different angles for each scan centers. After the data acquisition phase, the information relating to the distances and the external orientation parameters (X, Y, Z, ω, φ and κ) extracted from the images in the database, were used in the positioning procedure, as a reference for the estimation of the camera position of the images acquired with the mobile device. The following paper will describe the survey operations, the method used to generate the database of solid images with the subsequent positioning procedure and, finally, the results obtained following the validation of the test.

1. INTRODUZIONE

I sistemi di posizionamento sono sempre stati tema di grande interesse nel corso della storia, fin dai tempi nei quali erano in uso le antiche tecniche di navigazione astronomica. Oggi sapere dove ci troviamo, non solo all'esterno ma anche all'interno di ambienti chiusi, è diventata una questione molto importante ed in questo senso, il GNSS non permette di tracciare, con una accuratezza centimetrica, la posizione in ambienti interni, dove il segnale trasmesso tra dispositivo mobile e satellite è disturbato o schermato da svariati ostacoli (alberi, edifici, barriere, etc.). Per quanto riguarda il posizionamento *indoor* il GNSS garantisce una precisione

nell'ordine di alcuni metri, ma questa tipologia di ambiente necessita, a causa delle dimensioni degli spazi, di una maggiore accuratezza (Figura 1).

I limiti di questo sistema e la diffusione di una grande quantità di LBS (*Location Based Services*), hanno portato, negli ultimi anni, allo sviluppo di diversi nuovi metodi di posizionamento e navigazione *indoor*. Ciò è stato possibile grazie all'utilizzo di dispositivi mobili come gli *smartphone*, che sono sempre più performanti in quanto dotati di vari sensori come: fotocamere digitali, ricevitori GNSS, piattaforme inerziali basate su giroscopi, accelerometri, magnetometri e sistemi per l'identificazione di radiofrequenze (RFID) (Piras et al., 2014).

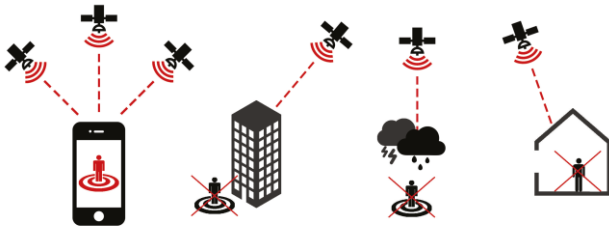


Figura 1. Limiti nel posizionamento con l'utilizzo del GNSS.

Esistono diverse tipologie di LBS (Figura 2) che utilizzano differenti tecnologie e strategie: *WiFi* (Bumgon et al., 2011), *Bluetooth*, *VLC (Visible Light Communication)* (Gosze, 2015; Hyassan et al., 2015), immagini digitali (Liu et al., 2010; Anwar et al., 2013), modelli tridimensionali e BIM (Building Information Modeling) (Faragher et al., 2015; Gupta et al., 2016), ognuno con un proprio campo di applicazione ottimale in base alle necessità dell'utente e alle possibilità che l'ambiente fornisce.

Le tecnologie WiFi e Bluetooth utilizzano entrambe il metodo del *Fingerprinting* secondo il quale il dispositivo mobile invia in rete una stringa alfanumerica identificativa e la posizione della fonte del segnale viene calcolata con algoritmi di trilaterazione, tracciando in ogni istante la distanza tra un dispositivo mobile e diversi punti di accesso WiFi installati nell'ambiente.

La tecnologia VLC prevede l'analisi di dati ricevuti da fonti luminose sfruttando lampade fluorescenti per la trasmissione di segnali a 10 kbit/s o LED per la trasmissione di segnali fino a 500Mbit/s. Ogni fonte emette un flusso luminoso caratterizzato da uno sfarfallio impercettibile che può essere rilevato dalla fotocamera di uno *smartphone* o da un sensore.

La tecnica che prevede l'utilizzo dei modelli digitali sfrutta le proprietà dei modelli BIM che vengono utilizzati per tracciare nell'ambiente una serie di percorsi navigabili. Il percorso adottato da un dispositivo mobile in ogni ambiente dipende alla posizione di tutti gli oggetti statici o in movimento presenti in esso e, di conseguenza, il percorso del sensore è limitato ad una serie di tracciati intercettati da diversi checkpoint nei punti di intersezione tra essi.

Technology	Indoor/Outdoor	Accuracy	Range	Cross-Platform	Power supply
GPS/GNSS		5 - 20m	worldwide		
WiFi		5 - 15m	< 150m		
Bluetooth		1 - 3m	< 30m		
VLC		< 50 cm	< 8m		
BIM		< 50 cm			
Image Based		20 - 50cm	worldwide		

Figura 2. Comparazione delle possibili tecnologie alla base degli LBS.

In particolare, il posizionamento *Image Based* sta diventando, negli anni, la tecnica sempre più dominante in un ampio campo di applicazioni grazie al miglioramento dei sistemi di rilevamento e all'avanzamento nella tecnologia dei sensori. Oltre a questo, si aggiunge un aumento della velocità di trasmissione dei dati e delle capacità di calcolo dei dispositivi utilizzati per l'elaborazione delle immagini. Il metodo si basa sull'individuazione di oggetti di posizione nota in immagini di riferimento e la loro corrispondenza nell'immagine digitale acquisita dall'utente con il dispositivo mobile. I riferimenti

possono provenire da immagini di modelli 3D digitali, da target codificati o proiettati e dal calcolo delle corrispondenze in una sequenza di immagini (Mautz et al., 2011).

A differenza degli altri sistemi citati in precedenza, questi metodi hanno il vantaggio di fornire, oltre alla posizione, anche l'orientamento dell'utente e non necessitano dell'installazione di dispositivi elettronici negli ambienti.

2. METODOLOGIA

Il metodo proposto per il posizionamento consente di stimare la posizione e l'orientamento di un'immagine scattata dalla camera di un dispositivo mobile attraverso l'estrazione di informazioni 3D da un'immagine di riferimento. La base del posizionamento è il *matching* tra l'immagine acquisita in tempo reale e la corrispondente immagine generata sinteticamente ed estratta da un *database*.

Il *workflow* (Figura 3) prevede una completa integrazione tra le operazioni di rilievo tradizionale (georeferenziazione dell'area), la raccolta di dati 3D (scansioni con laser scanner terrestre), l'elaborazione dei dati raccolti (generazione del database di immagini sintetiche) e applicazione di algoritmi di *computer vision* per estrarre le informazioni necessarie per il posizionamento (Chiabrando et al., 2016).

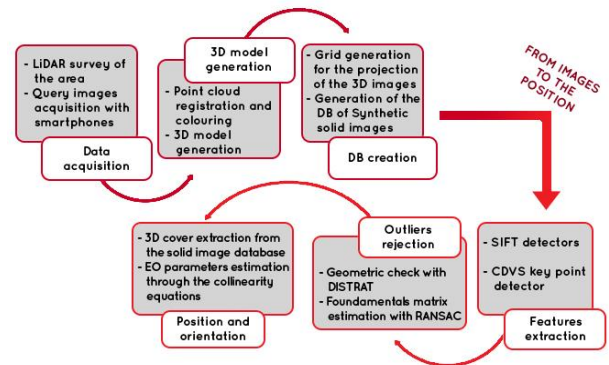


Figura 3. Workflow procedura IRBL.

2.1 Generazione del database di immagini RGB-D

Un'alternativa al metodo classico di restituzione fotogrammetrica è rappresentata dalla conoscenza, per ogni pixel dell'immagine, della distanza tra il centro di presa della camera ed il singolo pixel in ciascuna delle direzioni nello spazio. Noti i due angoli che definiscono una direzione nello spazio e una distanza, è possibile ricostruire la posizione di ciascuno di questi punti rappresentati nell'immagine. La valutazione della distanza tra il centro di presa e ogni punto dell'oggetto rappresentato è possibile ottenerla utilizzando un modello digitale di tipo denso dello stesso (Dense Digital Elevation Model - DDEM). Tali modelli digitali sono facilmente ottenibili con l'utilizzo della tecnologia Lidar. Il concetto di immagine RGB-D (Bornaz et al. 2003) prevede quindi l'integrazione del modello digitale dell'oggetto e l'informazione fotografica corrispondente (Figura 4). Una classica immagine digitale è generalmente formata dalla sovrapposizione di tre matrici contenenti i valori radiometrici RGB (Red, Green e Blue), mentre all'immagine RGB-D viene attribuita in aggiunta una matrice, delle stesse dimensioni in termini di numero di righe e colonne, contenente l'informazione ottenuta dal DDEM. In un'immagine digitale, conoscendo i parametri di orientamento interno ed esterno, è possibile stabilire una direzione nello spazio; quindi, se per ogni pixel dell'immagine, si conosce anche la distanza tra il centro della prospettiva e l'oggetto, è possibile calcolare, per

ognuno di questi, la posizione nello spazio all'interno di un sistema di riferimento noto.

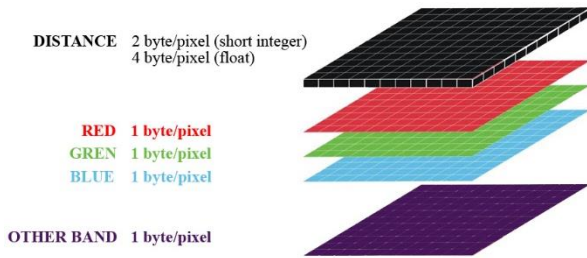


Figura 4. Struttura dell'immagine RGB-D.

Per riempire la matrice "distances" è necessario calibrare l'immagine. Il processo di calibrazione consiste nella stima dei parametri di distorsione della lente e di orientamento interno ed esterno della camera. Questa operazione avviene misurando le coordinate immagine di diversi punti di controllo dei quali si conosce la posizione spaziale nel sistema di riferimento. Una volta noti i parametri di orientamento interno ed esterno dell'immagine, è possibile proiettare il DDEM (derivato dalla nuvola di punti LiDAR) sull'immagine digitale.

Ottenuto l'intero modello dell'ambiente si procede alla generazione delle immagini RGB-D orientate che andranno a definire il database. La completezza della rappresentazione dipende dal numero di immagini sintetiche che vengono generate e che sono in grado di descrivere e rendere visibili tutti i dettagli più significativi presenti nell'ambiente. Più è complesso l'ambiente, maggiore sarà il numero di immagini richiesto per rappresentarlo in ogni suo dettaglio. Il modello 3D dal quale viene generato il database, quindi, deve avere sia informazioni geometriche che radiometriche.

Le immagini RGB-D vengono generate in base ai parametri della camera che vengono definiti inizialmente:

- I parametri di orientamento esterno corrispondenti alla posizione e di orientamento della camera ($X_0, Y_0, Z_0, \omega, \phi, \kappa$);
- I parametri di orientamento interno corrispondenti alla focale e alla posizione del punto principale della camera rispetto all'immagine sintetica (c, ζ_0, η_0);
- Il numero di pixel per ogni riga e colonna dell'immagine RGB-D (n_{col}, n_{row}), e la dimensione del pixel nell'immagine (d_{pix}).

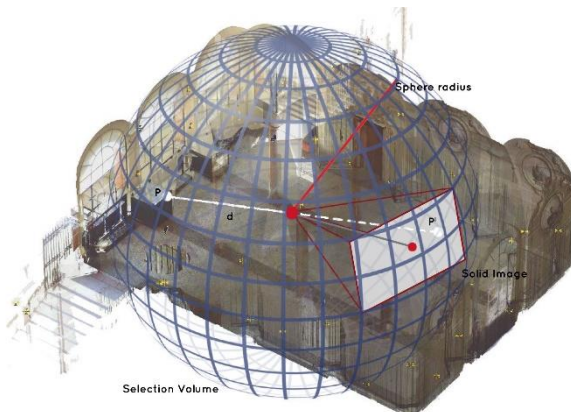


Figura 5. Sfera di selezione per la creazione delle immagini sintetiche.

Le immagini RGB-D vengono generate secondo il seguente procedimento (Bornaz et al. 2003):

- Viene creata un'immagine vuota (RGB e distance) usando (n_{col}, n_{row});

- Viene selezionato un sottoinsieme di nuvole di punti RGB definendo il raggio massimo di ricerca per creare una sfera di selezione centrata rispetto alla posizione in cui viene generata l'immagine (Figura 5);

- Per ogni punto selezionato è calcolata una distanza rispetto alla posizione dell'immagine RGB-D;

- Ogni punto RGB selezionato viene proiettato sull'immagine solida (Figura 6), definendo le proprie coordinate immagine (ζ_i, η_i) attraverso i parametri di orientamento interno ed esterno nelle equazioni:

$$\xi = \xi_0 - c \frac{r_{11}(X - X_0) + r_{21}(Y - Y_0) + r_{31}(Z - Z_0)}{r_{13}(X - X_0) + r_{23}(Y - Y_0) + r_{33}(Z - Z_0)} \quad (1)$$

$$\eta = \eta_0 - c \frac{r_{11}(X - X_0) + r_{21}(Y - Y_0) + r_{31}(Z - Z_0)}{r_{13}(X - X_0) + r_{23}(Y - Y_0) + r_{33}(Z - Z_0)}$$

- Le coordinate immagine (ζ_i, η_i) vengono convertite in coordinate pixel (c_i, r_i);

$$c_i = \frac{\xi_i}{d_{pix}} + \frac{n_{col}}{2} \quad r_i = \frac{\eta_i}{d_{pix}} + \frac{n_{row}}{2} \quad (2)$$

- Il valore RGB del punto viene scritto all'interno della matrice immagine RGB, nelle posizioni (c_i, r_i);

- Il valore di distanza di viene scritto all'interno della cella nella matrice delle distanze, nella posizione (c_i, r_i);

- Al termine del procedimento, i pixel ancora vuoti vengono riempiti grazie ad un algoritmo di interpolazione sulla base dei pixel pieni più vicini.

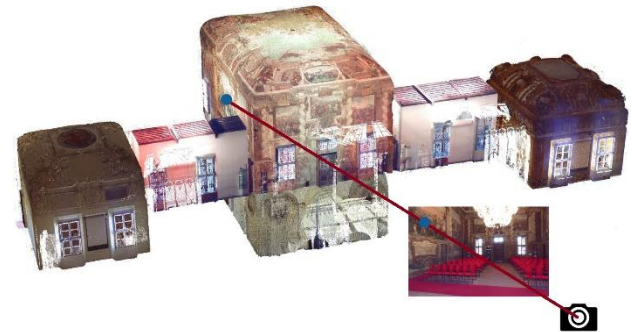


Figura 6. Procedura di generazione delle immagini RGB-D.

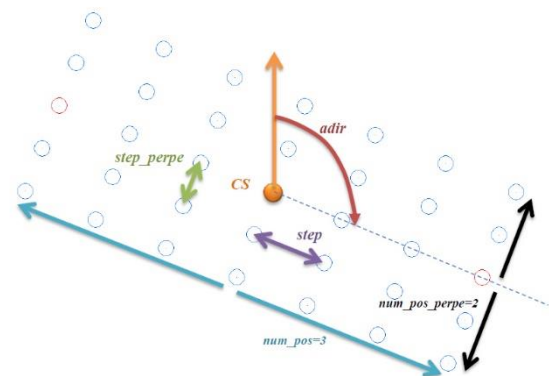


Figura 7. Definizione delle posizioni.

Le immagini RGB-D sono state generate automaticamente utilizzando un software sviluppato dal gruppo di ricerca di

Geomatica del Politecnico di Torino (ScanToRGBDImage) con il seguente schema (Figura 7). La definizione della direzione degli assi per ogni posizione segue lo schema in Figura 8.

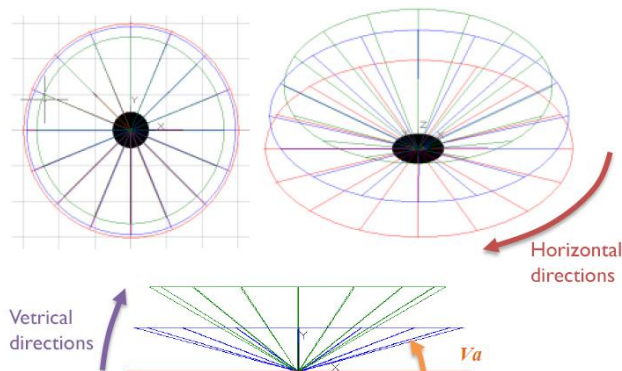


Figura 8. Definizione della direzione degli assi per ogni posizione.

Al termine del processo il software genera un set di immagini sintetiche con le informazioni riguardo la posizione e l'orientamento delle immagini (Figura 9). Per ognuna delle immagini prodotte viene generato un file .ori contenente i parametri di orientamento interno ed esterno della camera, un'immagine raster in formato .jpg in cui è registrata la matrice RGB ed un file .hdr in cui è registrata la matrice delle distanze.



Figura 9. Esempio di immagine generata dal software ScanToSolidRGBDImage nella visualizzazione RGB e indicazione delle profondità.

2.2 CDVS (Compact Descriptor Visual Search)

Una volta realizzato il database di immagini sintetiche, l'obiettivo è stato quello di selezionare un'immagine di riferimento (*reference*), all'interno del database, con il più alto livello di somiglianza ad ognuna delle immagini acquisite con la camera dello smartphone (*query image*). Per questa operazione è stata utilizzata la soluzione fornita dal MPEG7 CDVS (Compact Descriptor for Visual Search) (Sikora, 2001) con il quale viene svolta una prima classificazione delle immagini attraverso i descrittori globali, consentendo di estrarre dal database un certo numero di immagini definite potenzialmente simili rispetto a quella scattata dalla camera. La procedura si sviluppa secondo le seguenti fasi:

- i descrittori locali, nelle immagini *query* e in quelle presenti nel database (*reference*), vengono estratti e compressi;
- le immagini del database vengono classificate preliminarmente in base ai descrittori globali (Sikora, 2001) di punteggio di somiglianza con le immagini *query*. I descrittori globali forniscono una rappresentazione statistica di un insieme di descrittori locali estratti dalle due immagini. Come risultato di questa operazione vengono selezionate nel database tre immagini potenzialmente simili all'immagine *query*;

- per ogni immagine selezionata, la procedura di *matching* a coppie con l'immagine *query*, viene eseguita tra i *key points* estratti, cercando di accoppiare le coppie di *key points* individuati nelle due immagini. Dopodiché, per ogni *feature* dell'immagine *query*, viene ricercata una sola *feature* simile nell'immagine sintetica di riferimento;

- i *key points* riconosciuti vengono validati da un controllo geometrico (DISTRAT - DISTance RATio) (Lepsoy et al., 2011; Francini et al., 2016) basato sul concetto che le proprietà statistiche del rapporto di distanza per un *matching* errato sono nettamente differenti dalle proprietà relative ad un *matching* corretto.

Dato l'eventuale elevato numero di immagini all'interno del database, il CDVS per accelerare il processo di recupero utilizza dei descrittori compressi, con priorità ai punti situati al centro dell'immagine.

2.3 L'Algoritmo di IRBL

Dopo aver trovato l'immagine di riferimento, è possibile estrarre da questa le informazioni tridimensionali delle *features* selezionate per stimare i parametri di orientamento dell'immagine *query*. In seguito all'estrazione delle immagini di riferimento, i *keypoints* e le relative *features* vengono estratte dalla *query* e dalla immagine *reference* utilizzando l'algoritmo SIFT (Scale-Invariant Feature Transform) (Lowe D. G., 2004) il quale consente l'associazione preliminare tra i *keypoints* delle due immagini. Successivamente viene eseguito lo scarto di un'alta percentuale di *outliers* secondo un approccio in due fasi. In un primo momento le corrispondenze corrette vengono selezionate con l'algoritmo DISTRAT (Lepsoy et al. 2011) (Francini et al. 2016) con un controllo geometrico basato sui rapporti di distanza tra le coppie di punti nelle immagini analizzate. Dopodiché viene effettuato un controllo RANSAC (RANdom Sample Consensus) (Hartley et al., 2004) su un set più ristretto di corrispondenze, con la stima della matrice fondamentale tra l'immagine *query* e l'immagine di riferimento, consentendo l'esclusione di restanti *outliers*. In questo modo si riduce di circa dieci volte il tempo di elaborazione rispetto al solo utilizzo dell'approccio RANSAC. Questa operazione avviene in tempi molto ristretti, infatti per ogni ambiente nel quale è stato testato il metodo, il tempo di estrazione delle immagini *reference* dal database di immagini sintetiche è stato di poche decine di secondi. Le *features* dell'immagine *query* e dell'immagine *reference* vengono trasformate in informazioni dimensionali utilizzando l'immagine RGB-D derivata dal modello digitale dell'ambiente.

3. ACQUISIZIONE DATI E PROCESSING

Per l'applicazione del metodo di posizionamento indoor, tema della ricerca presentata, è stato scelto come caso studio il Castello del Valentino (Figura 10), una delle Residenze Sabaude incluse nella "Corona delle Delitiae" dichiarate Patrimonio dell'Umanità dall'UNESCO nel 1977.

Il Castello, oggi sede della facoltà di Architettura del Politecnico di Torino, è sempre più frequentemente sede di importanti eventi.



Figura 10. Caso studio: il Castello del Valentino, Torino.

Come prima operazione è stata realizzata una rete topografica dell'area e dell'ambiente circostante per definire un sistema di coordinate di riferimento con una tecnica mista GNSS e Stazione Totale. La rete è stata realizzata nel piano terreno e nel piano nobile del Castello come mostrato in Figura 11.

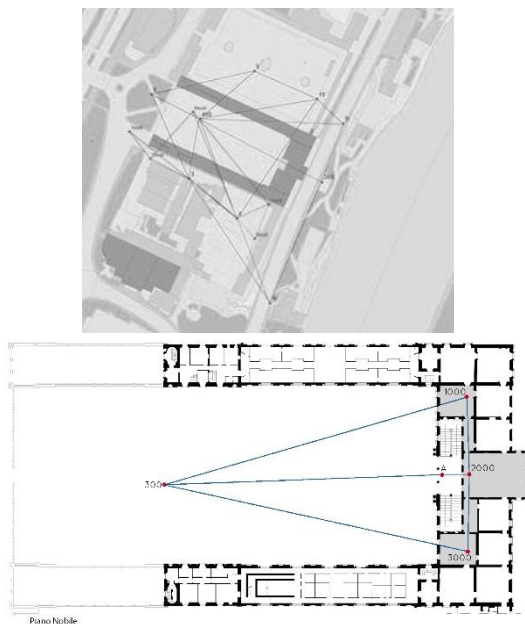


Figura 11. Indicazione della rete topografica.

Successivamente sono stati posizionati 66 marker a scacchiera nelle quattro sale del Castello oggetto di studio (Sala delle Colonne, Salone d'Onore, Stanza delle Feste e dei Fasti e Stanza dei Gigli) misurati con stazione totale e utilizzati in fase di registrazione delle scansioni laser. In ultimo è stato effettuato un rilievo LiDAR completo delle quattro sale e dei relativi spazi di collegamento (Figura 12), utilizzando uno scanner Faro Focus 3D X130 e acquisendo anche le immagini digitali in modo da ottenere una nuvola di punti colorata e il più possibile simile all'ambiente reale. Con la tradizionale procedura (Balletti et al. 2017) sono stati processati i dati acquisiti e realizzato un modello tridimensionale completo, formato da 27 scansioni (ogni scansione formata, in media, da circa 30 milioni di punti). Utilizzando i marker in fase di registrazione è stato possibile calcolare inoltre la tensione sui punti misurati, ottenendo una media di 6.5 mm sul totale dei 66 punti acquisiti.

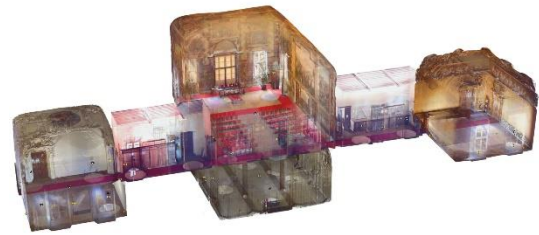


Figura 12. Nuvola di punti completa.

L'ultima operazione, per quanto riguarda la fase di processing dei dati acquisiti, consiste nella generazione del database di immagini sintetiche. Con i file .xyz contenenti le coordinate di ciascun punto delle scansioni effettuate ed i rispettivi valori RGB, utilizzando il software ScanToRGBDImages, secondo il procedimento descritto precedentemente, è stato generato automaticamente il database di immagini RGB-D (esempio in Figura 9) partendo dalla nuvola di punti LiDAR. In totale sono state generate 5184 immagini sintetiche, 192 per ciascuna posizione di scansione (6 direzioni sul piano verticale verticale per 32 direzioni sul piano orizzontale) con dimensioni 2500x1600 pixel, dimensione del pixel su sensore di 3µm e distanza focale di 4.667 mm, in modo da avere una buona copertura in un ambiente di dimensioni ridotte.

4. ACQUISIZIONE IMMAGINI CON SMARTPHONE

Nelle sale oggetto di studio sono state acquisite 90 immagini digitali con uno smartphone Apple iPhone 5s utilizzate come immagini *query*. Le immagini sono state volontariamente scattate sia con orientamento verticale che orizzontale, in modo da testare il corretto funzionamento della procedura di posizionamento in entrambi i casi. La camera del dispositivo utilizzato richiede una calibrazione attraverso procedure analitiche per definire le caratteristiche dell'ottica e per valutarne i parametri di distorsione coinvolti nella definizione dell'orientamento interno. Come approssimazione è possibile considerare solo gli effetti della distorsione radiale espressa da k_1 e k_2 (Chiabrando et al. 2017). La camera dello smartphone è stata calibrata con l'utilizzo del *tool* Camera Calibrator del software MatLAB per ricavare distanza focale f_x ed f_y , punto principale ζ_0 e η_0 , distorsioni radiali k_1 e k_2 (Figura 13). Per effettuare questa operazione sono state utilizzate 20 immagini di una scacchiera acquisite a distanze diverse e con angolazioni diverse.

```

cameraParams =
  cameraParameters with properties:
    Camera Intrinsic
      IntrinsicMatrix: [3x3 double]
      FocalLength: [2.7802e+03 2.7877e+03]
      PrincipalPoint: [1.5915e+03 1.2413e+03]
      Skew: 0
    Lens Distortion
      RadialDistortion: [0.0701 0.2122]
      TangentialDistortion: [0 0]
    Camera Extrinsic
      RotationMatrices: [3x3x7 double]
      TranslationVectors: [7x3 double]
    Accuracy of Estimation
      MeanReprojectionError: 1.0085
      ReprojectionErrors: [40x2x7 double]
      ReprojectedPoints: [40x2x7 double]
    Calibration Settings
      NumPatterns: 7
      WorldPoints: [40x2 double]
      WorldUnits: 'mm'
      EstimateSkew: 0
      NumRadialDistortionCoefficients: 2
      EstimateTangentialDistortion: 0
  
```

Figura 13. Calibrazione della camera.

4.1 Matching tra immagine digitale ed immagine RGB-D di riferimento

Come descritto nella sezione 2.2, la tecnologia del riconoscimento visivo consente di individuare la migliore corrispondenza tra le immagini presenti nel database, classificandole in base ad un punteggio. Un esempio dei risultati della procedura applicata in questo caso sono riportati in Figura 14 e Tabella 1.



Figura 14. Immagine *query* (sx), immagini *reference* corrispondenti (dx) e indicazione dei punteggi per l'individuazione della corrispondenza nelle immagini del database (basso).

Il tempo di calcolo per il completamento della procedura si aggira intorno a una decina di secondi ed i punteggi maggiori di 5 indicano una buona corrispondenza.

Smartphone Image	Reference Image n°1	Score
query/IMG_2740.JPG	references_colonne/s004-i+0+0-27-04.jpg	6.65771
query/IMG_2744.JPG	references_colonne/s004-i+0+0-16-01.jpg	3.8451
query/IMG_2755.JPG	references_colonne/s004-i+0+0-30-03.jpg	2.73575
query/IMG_2758.JPG	references_colonne/s004-i+0+0-30-03.jpg	10.906
query/IMG_2764.JPG	references_colonne/s004-i+0+0-13-01.jpg	3.71374
query/IMG_2769.JPG	references_colonne/s004-i+0+0-29-01.jpg	3.56258
query/IMG_2776.JPG	references_colonne/s004-i+0+0-12-02.jpg	2.1087
query/IMG_2786.JPG	references_colonne/s004-i+0+0-24-02.jpg	4.87237
query/IMG_2793.JPG	references_colonne/s004-i+0+0-05-01.jpg	29.2305
query/IMG_2795.JPG	references_colonne/s004-i+0+0-29-02.jpg	13.7709
query/IMG_2800.JPG	references_colonne/s004-i+0+0-22-04.jpg	17.2807
query/IMG_2812.JPG	references_colonne/s004-i+0+0-10-01.jpg	4.79915
query/IMG_2821.JPG	references_colonne/s004-i+0+0-11-03.jpg	2.67447
query/IMG_2826.JPG	references_colonne/s004-i+0+0-30-02.jpg	2.81401
query/IMG_2827.JPG	references_colonne/s004-i+0+0-18-02.jpg	4.59416

Tabella 1. Indicazione dei punteggi per l'individuazione della corrispondenza nelle immagini del database.

4.2 Posizionamento e orientamento

Come descritto nella sezione 2.3, sono stati stimati i parametri di orientamento esterno delle immagini *query* prodotte dallo smartphone ed estratte le *features* dalle immagini *query* e *reference* e, dopo aver ottenuto le coordinate spaziali, le *features* vengono trasformate in informazioni tridimensionali grazie all'immagine RGB-D. Nelle immagini che seguono (Figura 15) sono riportati i risultati dell'individuazione dei *key points* nelle immagini scattate dallo *smartphone* e le relative immagini sintetiche associate. I vettori in colore verde indicano le rispettive posizioni dei *key points* riconosciuti in

entrambe le immagini, mentre i vettori in colore rosso indicano i punti scartati. Il procedimento è stato effettuato con l'utilizzo del software MatLAB.

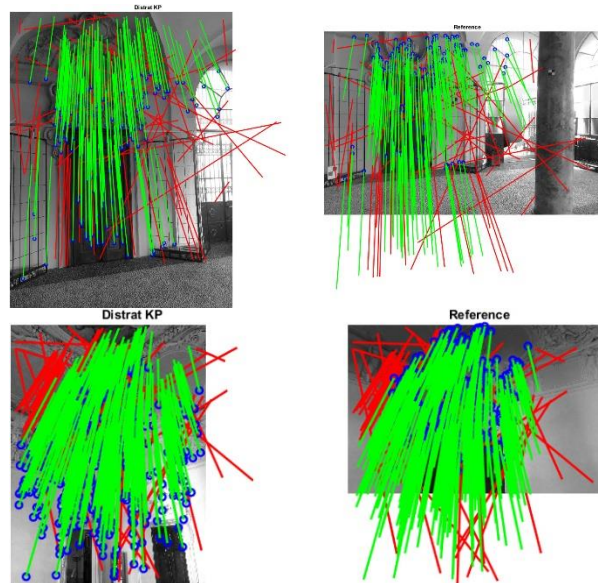


Figura 15. Indicazione delle *features* e degli *outliers* estratti.

4.3 Ground Truth

Le coordinate del centro della camera con cui sono state scattate le immagini *query* ottenute dal procedimento di IRBL completato con il software MatLAB, sono state in seguito verificate confrontandole con le coordinate calcolate attraverso la procedura fotogrammetrica classica. Con l'utilizzo del software ERDAS Imagine Photogrammetry, su un campione di diciotto immagini (quattro per ogni Sala e una per ognuno dei corridoi di collegamento tra il Salone d'Onore e le Stanze delle Feste e dei Fasti e dei Gigli) sono stati calcolati i parametri di orientamento esterno. Per ognuna delle diciotto immagini campione, sono stati individuati sei punti di coordinate note, reperite dalle nuvole di punti acquisite e georiferite, in modo da avere un riferimento spaziale nel sistema di coordinate utilizzato (Figura 16).



Figura 16. Estrazione diretta di punti di controllo per verificare la correttezza nel posizionamento.

Dopo aver individuato i punti di coordinate note, sono stati inseriti nel software i parametri di orientamento esterno ricavati dal software ScanToRGBDImage.

Avviato e concluso il calcolo, viene generato un file .xps contenente il report del processo. Questo contiene le coordinate dei punti individuati, gli errori residui, i parametri di orientamento interno impostati in precedenza e i parametri di orientamento esterno calcolati sull'immagine scattata con lo smartphone.

Di seguito sono riportati i risultati del confronto tra le due soluzioni con l'indicazione degli errori ottenuti (Tabella 2).

Viene riportato un campione caratteristico e rappresentativo dei risultati ottenuti sul totale delle diciotto verifiche effettuate. Gli scarti in x e in y, in valore assoluto, non superano mai i 70 cm, con una media di circa 40 cm e una deviazione standard di circa 15 cm. In z, invece, gli scarti non superano i 90 cm fatta eccezione per un caso relativo al Salone d'Onore, in cui si ha una differenza di 1,4 m circa tra le coordinate del posizionamento fornito dal procedimento di IRBL e il procedimento fotogrammetrico classico. La media dei valori di scarti in z è di circa 70 cm, mentre la deviazione standard di 50 cm circa.

IMMAGINE	Scarti		
	x (m)	y (m)	z (m)
Colonne_1	0.192	0.357	0.287
Colonne_2	0.223	0.201	0.133
Onore_1	0.221	0.198	0.030
Onore_2	0.669	0.556	1.396
Feste_1	0.308	0.046	0.079
Feste_2	0.251	0.020	0.365
Gigli_1	0.272	0.297	0.864
Gigli_2	0.302	0.374	0.509
min	0.192	0.020	0.030
max	0.669	0.556	1.396
med	0.431	0.288	0.713
dev. std.	0.153	0.178	0.466

Tabella 2. Verifica della precisione nel posizionamento.

5. CONCLUSIONI

Il sistema testato ha prodotto risultati soddisfacenti per quanto riguarda l'accuratezza raggiunta nel posizionamento e i tempi impiegati per la fase di ricerca di immagini *reference* nel database. La fase di generazione del database è quella che comporta maggiore dispendio in termini di tempo, potrebbe essere quindi migliorata con una nuova versione del software ScanToRGBDImage in un differente linguaggio di programmazione. Inoltre sarebbe possibile lo sviluppo di un'applicazione per permettere di estrarre le informazioni relative alla posizione direttamente sul dispositivo mobile ed effettuare le operazioni di calcolo *on-line*.

6. RINGRAZIAMENTI

Il lavoro presentato è stato possibile grazie all'utilizzo di strumenti sviluppati dal gruppo di ricerca di Geomatica del Politecnico di Torino e alle ricerche sviluppate nell'ambito del progetto, condotto dal prof. A.M. Lingua, in collaborazione con l'istituto sud-coreano ETRI.

7. BIBLIOGRAFIA

Anwar Q., Malik A.W., Thornberg B., *Design of coded reference labels for indoor optical navigation using monocular camera*, in Proceedings of the International

Conference Indoor Positioning and Indoor Navigation (IPIN), Montbeliard, France, 28–31 October 2013; pp. 1–8.

Balletti C., Guerra F., Vernier P., Studnicka N., Riegl J., Orlandini S., *Practical Comparative Evaluation of an Integrated Hybrid Sensor Based on Photogrammetry and Laser Scanning for Architectural Representation*, in ISPRS proceedings, XXXV congress, comm5, January 2017.

Bornaz L., Dequal S., *A new concept: The solid image*, in Proceedings of the XIXth International Symposium, CIPA 2003: New Perspectives to Save Cultural Heritage, Antalya, Turkey, 30 September–4 October 2003.

Bumgon K., Wonsun B., Kim Y.C. *Indoor localization for Wi-Fi devices by cross-monitoring AP and weighted triangulation*, in Proceedings of the IEEE Consumer Communications and Networking Conference (CCNC), Las Vegas, NV, USA, 9–12 January 2011.

Chiabrando F., Di Pietra V., Lingua A.M., Cho Y., Jeon J., *An Original Application of Image Recognition Based Location in Complex Indoor Environments*, in International Journal of Geo-Information, 2017.

Chiabrando F., Di Pietra V., Lingua A.M., Jeon J. I., Cho Y., *Solid Image Generation for Indoor Image Based Positioning, the Bangbae Metro Station in Seoul*, in 2016 International Conference on Indoor Positioning and Indoor Navigation (IPIN), Alcalá de Henares (Spain) 4 - 7 October 2016.

Faragher R. M., Harle R. K., *Towards an Efficient, Intelligent, Opportunistic Smartphone Indoor Positioning System*, in NAVIGATION: Journal of The Institute of Navigation, vol. 62, no. 1, 2015.

Francini G., Lepsoy S., *Method and System for Comparing Images*, US 9245204 B2, 26 January 2016.

Gozse I., *Optical Indoor Positioning System Based on TFT Technology*, in MDPI Sensors 16,19, 2015.

Gupta A., Yilmaz A., *Indoor Positioning using Visual and Inertial Sensors*, in IEEE Sensors, 2016.

Hassan N.U., Naeem A., Pasha M.A., Jadoon T., *Indoor Positioning Using Visible LED Lights: A Survey*, in ACM Computing Surveys, Vol. 48, No. 2, Article 20, November 2015.

Hartley, R., Zisserman, A., *Multiple View Geometry in Computer Vision*, 2nd ed.; Cambridge University Press: Cambridge, UK, 2004.

Lepsoy S., Francini G., Cordara G., de Gusmao P.P., *Statistical modelling of outliers for fast visual search*, in Proceedings of the IEEE International Conference on Multimedia and Expo (ICME), Barcelona, Spain, 11–15 July 2011; pp. 1–6.

Liu J.J., Philips C., Daniilidis K. *Video-based localization without 3D mapping for the visually impaired*, in Proceedings of the IEEE Computer Society Conference Computer Vision and Pattern Recognition Workshops (CVPRW), San Francisco, CA, USA, 13–18 June 2010; pp. 23–30.

Lowe D. G., *Distinctive Image Features From Scale-Invariant Keypoint*, in International Journal of Computer Vision, vol. 60, 2 - 5 Gennaio 2004, pp. 91-110.

Mautz R., Tilch S., *Survey of Optical Indoor Positioning Systems*, in International Conference on Indoor Positioning and Indoor Navigation (IPIN), 21 - 23 September 2011.

Piras M., Lingua A.M., Dabove P., Aicardi I., *Indoor Navigation Using Smartphone Technology: A Future*

Challenge Or An Actual Possibility?, in 2014 IEEE/ION Position, Location and Navigation Symposium-PLANS, IEEE, 2014, pp. 1343-1352.

Sikora T., *The MPEG-7 Visual Standard for Content Description - An Overview*, in IEEE transactions on circuits and systems for video technology, vol. 11, no. 6, June 2001, pp. 696-702.