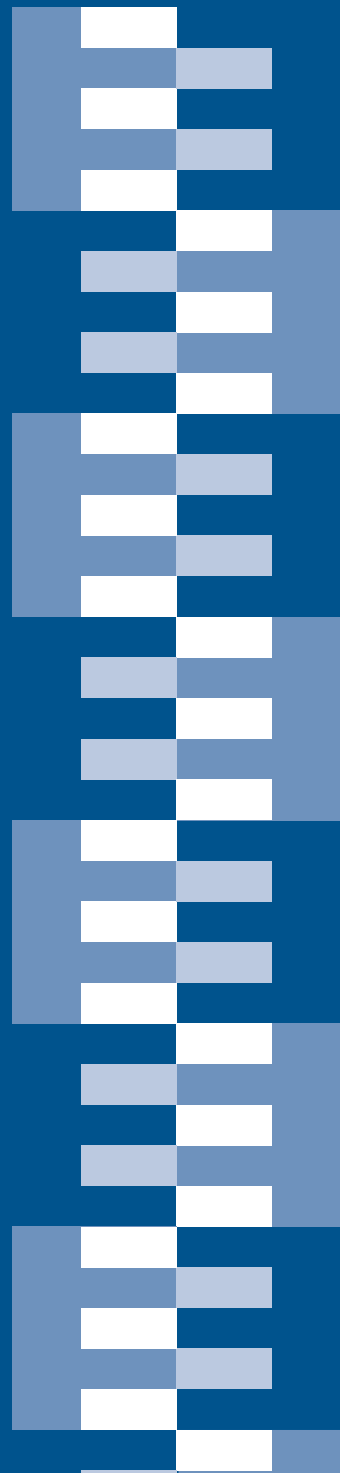


BOLLETTINO

# SIFET

SOCIETÀ ITALIANA DI FOTOGRAMMETRIA E TOPOGRAFIA



# 20

# Bollettino della Società Italiana di Fotogrammetria e Topografia

ISSN 1721-971X & ISSN 2784-9473 (online)



## Direttore responsabile | Chief Editor

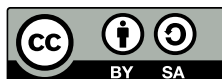
Prof. Andrea Maria Lingua

## Segreteria di Redazione | Publishing Editor

e-mail: redazione@sifet.org

I contributi di tutte le sezioni sono sottoposti a revisione tra pari, se accettati sono condivisi con licenza CC-BY-SA, il copyright rimane all'autore(i).

All the editorial sections are open submissions, peer-reviewed and indexed, copyright remains with the author(s).



## **Segmentazione semantica delle nuvole di punti utilizzando tecniche di apprendimento profondo per il patrimonio culturale**

Roberto Pierdicca, Marina Paolanti, Francesca Matrone, Massimo Martini, Christian Morbidoni, Eva Malinverni, Emanuele Frontoni, Andrea Lingua

pg.. 1-9

## **Un benchmark per la segmentazione semantica di nuvole di punti di beni culturali**

Francesca Matrone, Andrea Lingua, Roberto Pierdicca, Eva Malinverni, Marina Paolanti, Eleonora Grilli, Fabio Remondino, Arnadi Muriyoso, Tania Landes

pg.. 10-18

## **Intelligenza artificiale e realtà aumentata per la condivisione del patrimonio culturale**

Roberta Spallone, Valerio Palma

pg.. 19-26

## **Apprendimento da dati sintetici di nuvole di punti per la segmentazione semantica di edifici storici**

Christian Morbidoni, Roberto Pierdicca, Marina Paolanti, Raissa Mammoli, Ramona Quattrini

pg.. 27-37

## **Approccio gerarchico di machine learning per la segmentazione semantica di nuvole di punti 3D**

Eleonora Grilli, Simone Teruggi, Francesco Fassi, Fabio Remondino, Michele Russo

pg.. 38-46

## **Ontologie geografiche nel dominio spaziale urbano e del patrimonio costruito**

Antonia Spanò, Elisabetta Colucci

pg.. 47-56

# ORTOFOTO AD ALTA RISOLUZIONE PER INDIVIDUAZIONE DI LESIONI STRUTTURALI CON “MAGO”

## HIGH RESOLUTION ORTHOPHOTOS FOR THE RECOGNITION OF STRUCTURAL DAMAGES WITH “MAGO”

S. Gagliolo <sup>a</sup>

<sup>a</sup> Università degli Studi di Genova, DICCA – Laboratorio di Geodesia, Geomatica e GIS, Via Montallegro 1, 16145 Genoa  
sara.gagliolo @edu.unige.it

**PAROLE CHIAVE:** Ortofoto, Monitoraggio, Lesioni, Struttura, Sicurezza

**KEY WORDS:** Orthophoto, Monitoring, Damages, Structure, Security

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

### RIASSUNTO

Il presente lavoro ha l'obiettivo di illustrare un metodo innovativo per la generazione di ortofoto, identificato dall'acronimo MAGO (Mesh Adattiva per la Generazione di Ortofoto) ed implementato in ambiente C++.

Gli elementi di novità apportati da questo nuovo approccio consentono di superare alcune problematiche tipicamente riscontrabili nell'applicazione di tecniche esistenti. In particolare, i vantaggi di MAGO riguardano: (1) la possibilità di ottenere un'ortofoto alla risoluzione del GSD (Ground Sample Distance), corrispondente alla dimensione del pixel immagine originale; (2) la creazione di una mesh adattiva triangolare, generata a partire dalla nuvola di punti, in grado di rappresentare nel dettaglio la superficie dell'oggetto di interesse, da proiettare sul piano dell'ortofoto.

La procedura è stata applicata al caso studio di un ponte stradale in muratura situato in provincia di Genova, per verificare le sue potenzialità nello studio dello stato fessurativo dell'arcata di valle. Una porzione della lesione maggiormente estesa è stata infatti restituita mediante due ortofoto ad alta risoluzione, generate a partire da nuvole di punti derivanti da rilievo fotogrammetrico e laser scanner. Infine, gli output sono stati analizzati e confrontati sulla base della qualità metrica finale e della fruibilità.

### ABSTRACT

The present work aims to show an innovative method for the orthophoto generation, identified with the acronym MAGO (Adaptive Mesh for Orthophoto Generation) and implemented in C++ environment. The innovations produced by this new approach allow to overcome some problems typically observable in the application of the existing techniques. In particular, MAGO strength points are: (1) the possibility to obtain an orthophoto with the GSD (Ground Sample Distance) resolution, corresponding to the dimension of the pixel of the original image; (2) the creation of a triangular adaptive mesh, generated from the point cloud, able to represent the object of interest surface, which is projected on the orthophoto plan, in a detailed way.

The procedure has been applied to the case study of a masonry road bridge in the district of Genoa, to verify its potential in the study concerning the cracking state of the downstream arch. A portion of the wider crack is reproduced by means of high resolution orthophotos, generated starting from photogrammetric and laser scanner cloud points. Finally, the outputs are analyzed and compared, on the basis of the final metrical quality and the user-friendliness.

### 1. INTRODUZIONE

La disponibilità sempre maggiore di prodotti 3D, facilmente interrogabili e fruibili, non ha sostituito, bensì integrato l'utilizzo di prodotti bidimensionali per la rappresentazione e la descrizione della realtà che ci circonda.

L'obiettivo comune di tali modalità di restituzione è la generazione di prodotti metricamente affidabili. Tra quelli tridimensionali si annoverano: nuvole e *mesh*, rispettivamente costituite da punti e poligoni; DSM (Digital Surface Models), che discretizzano le nuvole di punti stesse attraverso celle tipicamente quadrate. Tra le rappresentazioni 2D utilizzate in ambito architettonico e strutturale, le principali sono piante, prospetti, sezioni e ortofoto.

In particolare le ortofoto, sulle quali è focalizzato il presente lavoro, sono largamente utilizzate per eseguire misure di elevata

precisione e per meglio interpretare la descrizione tematica dell'area rilevata. Esse sono realizzate a partire da immagini dell'oggetto, la cui morfologia può essere descritta attraverso un DSM o da un modello poligonale (*mesh*). Il raggiungimento della più elevata risoluzione possibile è dato dalla dimensione del cosiddetto GSD (Ground Sample Distance), corrispondente alla dimensione del pixel immagine proiettato sulla superficie dell'oggetto, e dal grado di approssimazione nella descrizione geometrica della superficie dell'oggetto mediante la *mesh*, come specificato nel paragrafo seguente.

Le ortofoto sono uno strumento essenziale per lo studio del territorio e del costruito. Le possibilità di effettuare misure precise e, parallelamente, di classificare il contenuto dell'immagine sono fondamentali nel supporto decisionale circa aspetti di produttività (Erlandsson et al., 2019), nel monitoraggio di fenomeni naturali e antropici (Niedzielski et

al., 2018; Nahhas et al., 2019), nell'archiviazione e conservazione del patrimonio culturale (Sammartano e Spanò, 2017; Perfetti et al., 2019), per la sicurezza infrastrutturale (Pérez et al., 2019).

Il presente lavoro descrive un approccio innovativo rispetto alle tecniche attualmente disponibili. Tale metodo, identificato dall'acronimo MAGO (*Mesh* Adattiva per la Generazione di Ortofoto), è stato implementato in ambiente C++, con il supporto delle librerie esistenti Cimg Library<sup>1</sup>, Image Magick<sup>2</sup>, e matrix.h<sup>3</sup> (Gagliolo et al., 2019).

La procedura consente di generare ortofoto a partire dalla nuvola di punti dell'oggetto (proveniente indifferentemente da rilievo laser scanner o fotogrammetrico) e da un'immagine, di cui siano noti l'orientamento interno ed esterno. La produzione dell'ortofoto è basata su un processo iterativo che dà luogo alla creazione di una *mesh* adattiva, generata considerando una specifica area piana triangolare come intersezione tra il raggio di collinearità e la nuvola di punti. In questo modo, è possibile limitare l'approssimazione e la perdita di definizione tipicamente dovute alla ricostruzione di una *mesh* "globale" dell'oggetto, che risulta più grossolana.

La procedura è applicata ad un caso studio, un ponte stradale in muratura sito in provincia di Genova. L'infrastruttura è di primaria importanza per gli abitanti della zona, in quanto rappresenta la via di comunicazione principale.

Lo stato fessurativo del ponte ha reso necessario eseguire un monitoraggio periodico della struttura per controllare gli eventuali avanzamenti e/o sviluppi delle lesioni. L'analisi condotta in questa sede è focalizzata alla porzione sommitale della principale lesione sull'arcata di valle.

Considerando la molteplicità di dati disponibili derivanti dal rilievo integrato, sono stati effettuati alcuni test della procedura sviluppata in differenti configurazioni al fine di comprendere quale sia il prodotto più idoneo all'analisi della lesione.

La trattazione è organizzata come segue: la procedura ed il *workflow* di MAGO sono descritti nel paragrafo 2, mentre nel paragrafo 3 viene presentato il caso studio. I risultati dei confronti sono argomentati nel paragrafo 4, che è seguito, infine, dalle conclusioni.

## 2. APPROCCIO E WORKFLOW DI MAGO

Come menzionato nell'introduzione, l'approccio di MAGO risulta innovativo rispetto ai metodi attualmente in uso. Allo stato dell'arte, infatti, sono due le principali problematiche da considerare nella generazione di ortofoto: da un lato, la possibilità di raggiungere la massima risoluzione consentita dall'immagine originale, dell'ordine del suo GSD; dall'altro, il grado di approssimazione con cui è rappresentata la superficie dell'oggetto di interesse, da proiettare sul piano dell'ortofoto.

L'approccio di MAGO è volto a risolvere tali questioni mediante la generazione di una *mesh* adattiva, costituita da specifiche aree piane triangolari i cui vertici sono ottenuti dall'intersezione tra il raggio di collinearità e la nuvola di punti. Su ciascuna di tali superfici viene quindi proiettata la porzione di immagine di competenza alla sua risoluzione originale.

Il *workflow* è articolato nelle seguenti fasi:

1. definizione del piano ortofoto;

2. acquisizione delle immagini e dei parametri di orientamento interno ed esterno provenienti da fonti esterne;
3. definizione delle dimensioni e della risoluzione dell'ortofoto;
4. processo iterativo per la determinazione dei tre punti più idonei alla definizione del piano intersezione tra la nuvola di punti e la retta di collinearità, ottenuta a partire dal pixel immagine considerato;
5. proiezione del colore dei pixel immagine sul piano ortofoto attraverso l'intersezione della retta di collinearità con il piano ottenuto al passo precedente.

### 2.1 Dati di input e impostazioni

I dati di *input* necessari per il funzionamento di MAGO sono la nuvola di punti descrivente la morfologia della superficie considerata e un'immagine della stessa, con i suoi parametri di orientamento interno ed esterno.

La nuvola di punti può provenire da TLS (Terrestrial Laser Scanner), LIDAR (Light Detection and Ranging) o da fotogrammetria, mentre i parametri di orientamento dell'immagine possono essere stimati con l'uso di un software fotogrammetrico.

All'utente è richiesto di definire il piano ortofoto, attraverso le coordinate di tre punti, e la sua risoluzione, da scegliere in base al GSD dell'immagine originale.

A partire dalla nuvola di punti in *input*, MAGO calcola automaticamente le dimensioni e i limiti dell'ortofoto, come indicato nelle relazioni (1).

$$\frac{X_{MAX} - X_{MIN}}{\text{risoluzione}} = \text{larghezza ortofoto}$$
$$\frac{Y_{MAX} - Y_{MIN}}{\text{risoluzione}} = \text{altezza ortofoto}$$
(1)

Per riordinare e allocare le informazioni dei punti della nuvola e permetterne un accesso facile e veloce, viene realizzata una griglia regolare, con una risoluzione dell'ordine della spaziatura media della nuvola di punti (assegnata come *input*). Tali dati sono utilizzati per la realizzazione dell'ortofoto, mediante i passaggi presentati in modo approfondito nel paragrafo seguente; inoltre, il software è in grado di generare DSM sulla base dei dati preordinati, secondo i criteri della media, della mediana o il massimo dei valori contenuti nella cella considerata.

Infine, se non fosse necessario processare le immagini alla loro risoluzione originale per ottenere l'ortofoto desiderata, è possibile impostare un fattore di riduzione per escludere una percentuale di pixel dal processo di calcolo, accelerandolo.

Gli *input* elencati, sia quelli provenienti da fonti esterne, sia quelli stabiliti dall'utente, possono essere comunicati al software attraverso una semplice interfaccia grafica, rappresentata nella Figura 1.

<sup>1</sup> <http://cimg.eu/>

<sup>2</sup> <https://imagemagick.org/>

<sup>3</sup> <http://matrix.drque.net/>

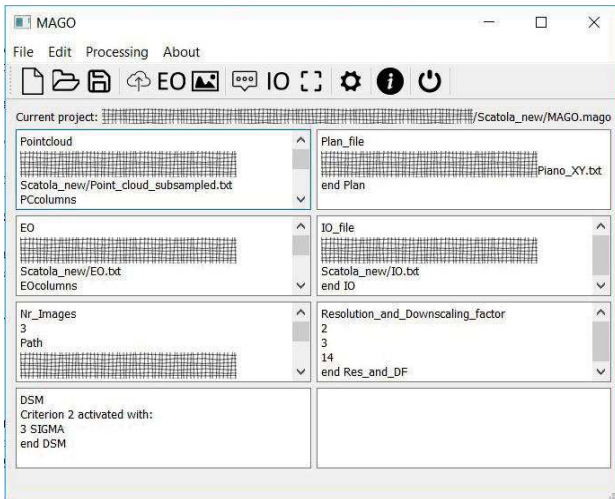


Figura 1. Interfaccia grafica di MAGO.

## 2.2 Processo iterativo per la produzione di ortofoto

Una volta caricati tutti i dati richiesti come *input*, la procedura per l'ottenimento dell'ortofoto può avere inizio.

Le equazioni di collinearità sono applicate ad ogni pixel dell'immagine considerata.

$$\begin{aligned} x &= x_0 - c \frac{r_{11}(X_p - X_0) + r_{12}(Y_p - Y_0) + r_{13}(Z_p - Z_0)}{r_{31}(X_p - X_0) + r_{32}(Y_p - Y_0) + r_{33}(Z_p - Z_0)} \\ y &= y_0 - c \frac{r_{21}(X_p - X_0) + r_{22}(Y_p - Y_0) + r_{23}(Z_p - Z_0)}{r_{31}(X_p - X_0) + r_{32}(Y_p - Y_0) + r_{33}(Z_p - Z_0)} \end{aligned} \quad (2)$$

Tra i parametri delle equazioni di collinearità riportate nell'Equazione (2) espresse in forma canonica, le coordinate immagine  $(x, y)$ , i parametri di orientamento interno (distanza focale  $c$ , e coordinate del punto principale,  $x_0, y_0$ ), e i parametri di orientamento esterno (posizione della camera  $\underline{X}_0 = (X_0, Y_0, Z_0)$  e componenti  $r_{ij}$  della matrice di rotazione di Cardano  $\mathbf{R}$ ) sono noti; invece, le coordinate del punto nel sistema oggetto  $\underline{X}_P = (X_P, Y_P, Z_P)$  sono incognite.

Pertanto, per ogni pixel il sistema di equazioni da risolvere è costituito da due equazioni con tre incognite, rappresentate dalle coordinate del punto nel sistema oggetto. Per trovare la soluzione del sistema, si dà luogo ad un processo iterativo, imponendo come  $Z_P$  di primo tentativo  $Z_{MAX}$ , ovvero la quota massima dei punti della nuvola.  $X_P$  e  $Y_P$  sono di conseguenza calcolati dall'intersezione della retta di collinearità e il piano a  $Z_P$  costante.

Le coordinate  $\underline{X}_P$  così ottenute si collocano all'interno di una cella della griglia, precedentemente costituita per organizzare e riordinare la nuvola di punti per un rapido accesso alle informazioni necessarie per il calcolo.

Se risulta che almeno un punto giace all'interno della cella indagata, si procede nel verificare la distanza in direzione  $Z$  rispetto alla  $Z_P$  fissata in precedenza. Si assume un valore di soglia  $\delta z$  definito univocamente sulla base della spaziatura della nuvola di punti  $s$  fornita in *input* e usata per formare la griglia.

Se il punto si trova nel *range* di tolleranza verticale  $\delta z$  ed ha una distanza dalla retta di collinearità inferiore al GSD, allora la corrispondenza è stata trovata ed il punto stesso viene proiettato sul piano ortofoto.

Se il *range* di tolleranza verticale  $\delta z$  è rispettato, ma il punto si trova ad una distanza maggiore del GSD, inizia la ricerca di due

punti idonei a formare la superficie triangolare adattiva di intersezione tra la nuvola di punti e il raggio di collinearità.

La ricerca dei punti avviene all'interno della cella in cui è stato trovato il primo punto e delle sue 8 celle vicine, seguendo i seguenti criteri:

1. la disomogeneità della nuvola fa sì che alcuni punti possano risultare troppo vicini o troppo lontani dal primo ritrovato; se tali punti fossero scelti come idonei, potrebbero comportare una rappresentazione non realistica della superficie. Per evitare ciò, si definisce in planimetria una corona circolare con raggio interno ( $r_i$ ) ed esterno ( $r_e$ ) rispettivamente del 40% e 180% della dimensione  $s$  e si stabilisce che i punti debbano giacere in tale fascia;
2. per evitare la definizione di piani sub-verticali, è fissato un limite di tolleranza anche in direzione verticale. Il valore di tale vincolo è proporzionale al diametro esterno dell'anello e alla tangente di  $\pm 80^\circ$ ;
3. almeno uno dei punti candidati a formare la *mesh* adattiva deve giacere dalla parte opposta della traccia della retta di collinearità rispetto al primo ritrovato;
4. i vertici devono formare un triangolo la cui proiezione planimetrica abbia un'area di almeno il 20% della dimensione  $s^2$ .

L'intersezione della superficie così definita con la retta di collinearità viene proiettata sul piano ortofoto.

I criteri delineati sono schematizzati nella Figura 2.

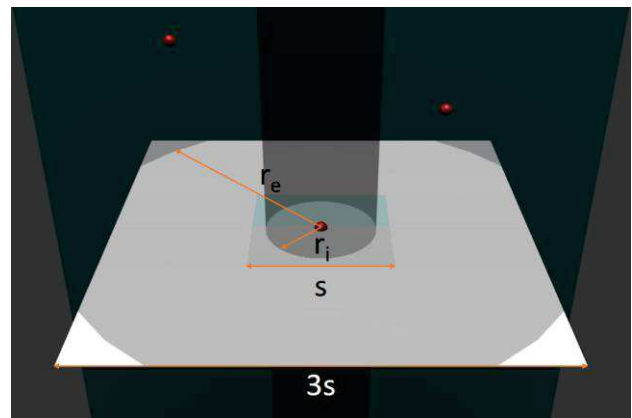


Figura 2. Ricerca dei punti per la definizione della *mesh* adattiva.

Se la cella indagata fosse vuota o vi giacessero punti che non soddisfano la tolleranza  $\delta z$ ,  $Z_P$  è decurtato automaticamente della quantità  $s$  e si ripete la procedura per il nuovo valore di  $Z_P$ .

Se nessuna corrispondenza è verificata fintanto che  $Z_P$  raggiunge il valore di  $Z_{MIN}$ , cioè il minimo valore di quota della nuvola di punti, il pixel viene scartato e la procedura prosegue con l'analisi del successivo, fino alla conclusione dell'intera immagine.

La procedura è riassunta nello schema in Figura 3.

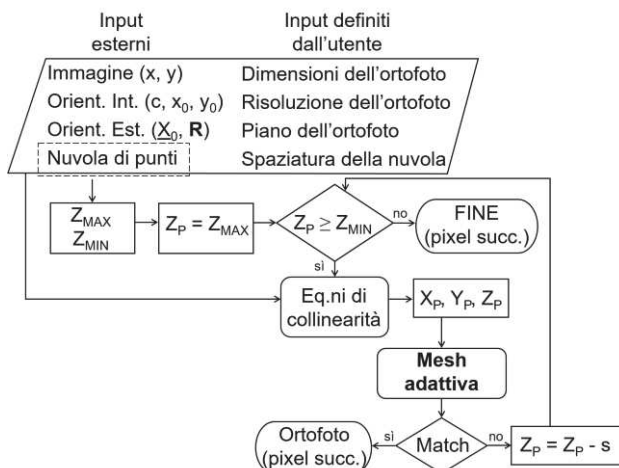


Figura 3. Schema della procedura di MAGO.

### 3. IL CASO STUDIO

Il ponte analizzato (rappresentato in Figura 4) è sito nell'entroterra di Genova. Nonostante si tratti di una piccola struttura, la sua importanza strategica è dovuta al fatto che è l'unica via di collegamento tra la zona costiera e quella a monte. Da alcuni anni lo sviluppo delle lesioni della struttura destava preoccupazione, sono stati pertanto effettuati monitoraggi periodici e, in tempi recentissimi, sono stati predisposti interventi di consolidamento.

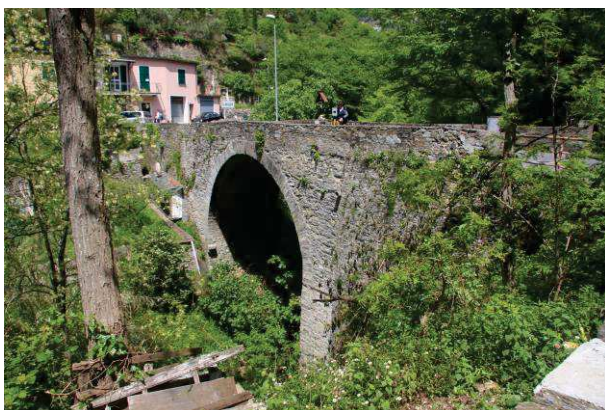


Figura 4. Vista del lato di valle del Ponte.

Le nuvole di punti considerate per l'applicazione della procedura di MAGO sono frutto di un rilievo integrato dell'intera struttura, acquisito mediante Laser Scanner, fotogrammetria e rilievo topografico dei punti di appoggio, svolto con tecnica tradizionale mediante Stazione Totale.

Gli strumenti utilizzati per la campagna di rilievo sono: la fotocamera Canon EOS 40D, il Laser Scanner Z+F Imager® 5006h e la Stazione Totale Leica TCR703.

La porzione selezionata per le analisi è la parte superiore della principale lesione, situata sul fronte di valle dell'arcata e rappresentata in Figura 5.

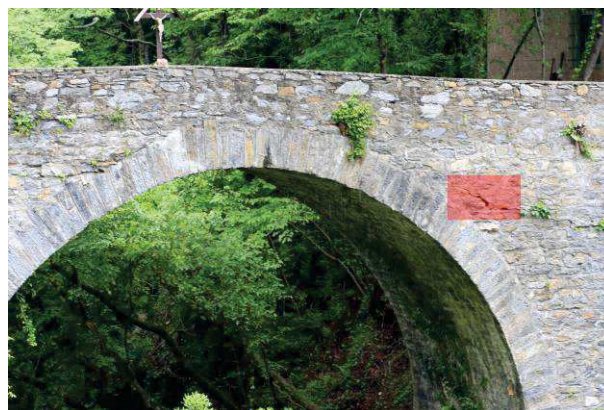


Figura 5. Immagine utilizzata per la realizzazione dell'ortofoto; in rosso, la porzione di interesse.

Sono state prodotte due ortofoto, a partire dalla nuvola di punti del Laser Scanner e da quella fotogrammetrica. I due prodotti sono stati confrontati (paragrafo 4) per studiarne le differenze.

Il GSD dell'immagine considerata è di circa 2,5 mm (distanza focale 85 mm, distanza dall'oggetto circa 35 m), mentre la spaziatura media delle nuvole di punti è rispettivamente di 4 mm per quella da Laser Scanner e 6 mm per quella fotogrammetrica. La risoluzione dell'ortofoto è stata fissata in entrambi i casi a 4 mm, confrontabile con la risoluzione della nuvola laser e sufficiente ad apprezzare la lesione oggetto di studio. L'immagine originale è stata trattata nella sua definizione originale, senza l'applicazione di un fattore di riduzione della scala.

Il parametro  $s$ , ovvero la spaziatura della griglia per il trattamento della nuvola di punti, è stato fissato a 16 mm per la nuvola fotogrammetrica e 12 mm per quella proveniente dal Laser Scanner. Tale sovradimensionamento è cautelativo perché consente di rilevare un'effettiva mancanza di dato laddove si riscontra assenza di punti nella discretizzazione a maglia regolare  $s$ .

Le necessarie elaborazioni fotogrammetriche sono state eseguite con il software Agisoft Photoscan© (ora Metashape©, 2019), mentre la gestione delle nuvole è stata condotta grazie a CloudCompare (2019).

Il modello è stato opportunamente ruotato in modo tale che la superficie del fronte di valle risultasse parallela al piano ortofoto.

Sono in corso ulteriori sviluppi del software per consentire di scegliere liberamente il piano dell'ortofoto.

### 4. RISULTATI E DISCUSSIONE

Come menzionato in precedenza, le ortofoto consentono parallelamente all'indagine metrica dell'oggetto anche la sua analisi tematica. Per le nuvole di punti, non sempre è possibile ottenere in modo soddisfacente entrambe le caratteristiche.

In particolare, la nuvola del Laser Scanner (Fig. 6) garantisce una risoluzione geometrica migliore rispetto a quella fotogrammetrica, ma la scala in termini di valori di grigi con cui è rappresentata non permette di riconoscere facilmente gli oggetti; d'altro canto, la nuvola fotogrammetrica (Fig. 7) permette di riconoscere gli elementi grazie al colore, ma la risoluzione ottenibile è tipicamente più grossolana.

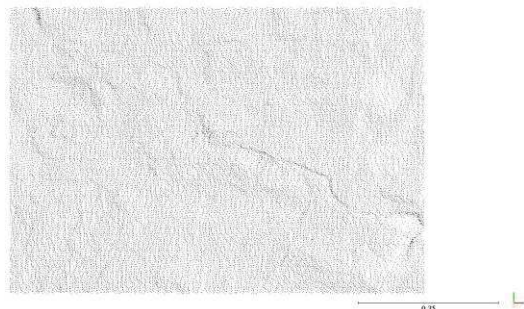


Figura 6. Nuvola Laser Scanner.  
Spaziatura 4 mm.

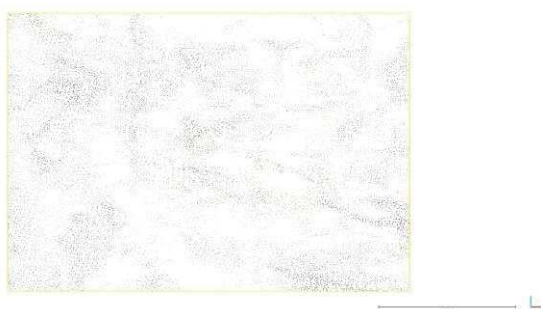


Figura 7. Nuvola fotogrammetrica.  
Spaziatura 6 mm.

L'ortofoto ricostruita con MAGO consente di ottenere da entrambe le nuvole in *input* un prodotto ad alta risoluzione. Le ortofoto ottenute rispettivamente dalla nuvola Laser Scanner e da quella fotogrammetrica sono riportate nella Figura 8 e nella Figura 9. Le porzioni di colore nero intenso, presenti nelle Figure 8 e 9, sono dovute ad assenza di dato (spaziatura della nuvola di punti superiore alla dimensione del pixel ortofoto) o a mancata corrispondenza diretta per la proiezione. Attualmente si sta lavorando ad un algoritmo di interpolazione del colore in grado di limitare tali effetti.



Figura 8. Ortofoto ottenuta dalla nuvola Laser Scanner.



Figura 9. Ortofoto ottenuta dalla nuvola fotogrammetrica.

Un dato inatteso è la presenza di un numero visibilmente più elevato di assenze di dato nell'ortofoto proveniente dalla nuvola Laser Scanner, nonostante la sua maggiore densità media rispetto alla nuvola fotogrammetrica.

Per giustificare tale comportamento, si è proceduto all'analisi dei due DSM, riportati Figura 10 per il Laser Scanner e nella Figura 11 per la fotogrammetria.

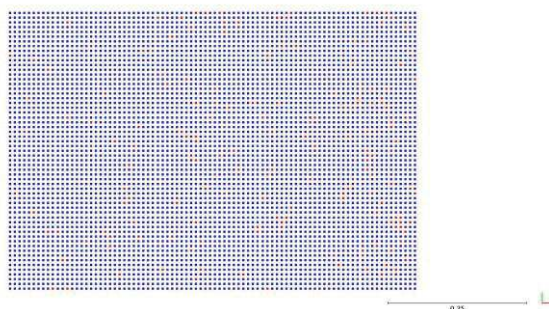


Figura 10. DSM della nuvola Laser Scanner; cella 12 mm.  
Le celle in rosso rappresentano buchi isolati.

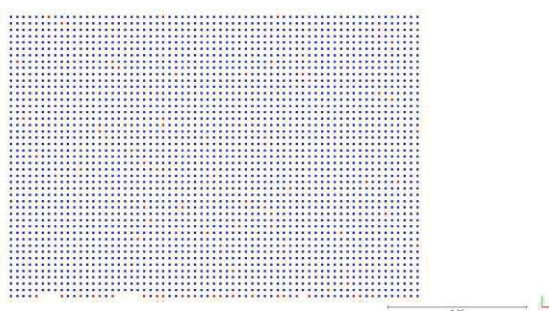


Figura 11. DSM della nuvola fotogrammetrica; cella 16 mm.  
Le celle in rosso rappresentano buchi isolati.

Nonostante la scelta della dimensione della griglia  $s$  sia stata operata cautelativamente per entrambi i casi, il DSM della nuvola ottenuta da Laser Scanner evidenzia un maggior numero di buchi isolati. Ciò riflette la disomogeneità di distribuzione dei punti, che non è stata sufficientemente compensata dalla creazione di una griglia a maglie molto più larghe della spaziatura media dei punti.

Nella Figura 12, descritta dalla Tabella 1, si riportano dati statistici relativi alle porzioni con assenza di dato delle due ortofoto. Attraverso l'uso del software GIS Open Source

GRASS<sup>4</sup>, sono state calcolate le dimensioni delle aree non ricostruite e la sovrapposizione tra tali aree nelle due ortofoto.

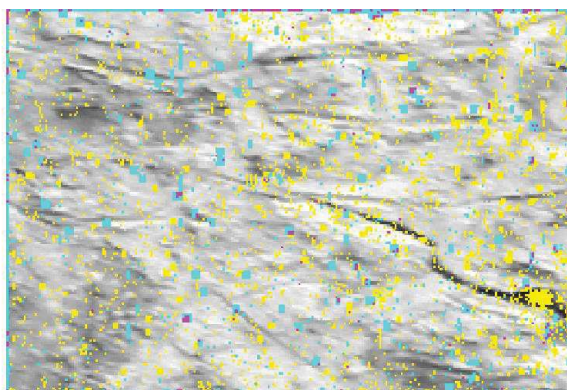


Figura 12. Individuazione delle aree non ricostruite; in giallo le aree non ricostruite a partire dalla nuvola Laser Scanner, in ciano da quella fotogrammetria, in magenta la sovrapposizione delle mancanze di dato per le due tecniche.

| Aree non ricostruite        | Num. celle |     | %   |      |
|-----------------------------|------------|-----|-----|------|
|                             | tot        | les | tot | les  |
| solo Laser Scanner (giallo) | 3431       | 173 | 7.5 | 31.6 |
| solo fotogramm. (ciano)     | 1606       | 6   | 3.5 | 1.1  |
| entrambi (magenta)          | 340        | 3   | 0.7 | 0.5  |

Tabella 1. Percentuale di ricoprimiento delle aree non ricostruite e numero di celle corrispondente per l'intera ortofoto (tot) e per la sola lesione (les).

Le statistiche riportate confermano quanto era già stato possibile valutare visivamente. Infatti, le aree con assenza di dato nell'ortofoto proveniente dal Laser Scanner sono circa il doppio di quelli evidenziati nell'ortofoto da fotogrammetria. Le aree comuni sono in percentuale esigua e localizzate maggiormente in prossimità dei bordi.

Nonostante i dati di *input* siano stati forniti per ottenere condizioni confrontabili, le significative differenze portano a considerare la diversa natura delle nuvole trattate, piuttosto che la procedura applicata.

Concentrando l'attenzione sulla lesione oggetto di studio, ad esempio, è evidente come questa sia delineata in maniera più completa in Figura 9, grazie all'utilizzo del dato fotogrammetrico.

## 5. CONCLUSIONI E PROSPETTIVE FUTURE

Il presente lavoro intende presentare un nuovo approccio per la ricostruzione di ortofoto, attraverso il software MAGO (*Mesh* Adattiva per la Generazione di Ortofoto). Tale metodo consente di ricostruire ortofoto ad alta risoluzione a partire da nuvole di punti, ottenute con tecniche di rilievo differenti, e da un'immagine orientata. L'utilizzo di una *mesh* adattiva, al posto di quella comunemente ricostruita dai software fotogrammetrici, consente di ottenere un prodotto finale che si ritiene di maggior dettaglio. Inoltre, la proiezione dell'immagine alla sua risoluzione originale consente di raggiungere il massimo grado

di dettaglio possibile, diversamente dall'ortofoto tradizionale, usualmente vincolata alla risoluzione del DSM.

La procedura è stata applicata ad un caso studio, per valutare la qualità e la fruibilità del prodotto finale, in relazione alle proprietà di partenza delle nuvole di punti, caratterizzate da una spaziatura media rispettivamente di 4 mm per quella da Laser Scanner e 6 mm per quella fotogrammetrica. La risoluzione delle ortofoto è stata fissata a 4 mm.

Il caso studio analizzato è un ponte stradale in muratura sito in provincia di Genova, del quale si è studiata una porzione della lesione localizzata sull'arcata di valle. L'analisi dei diversi prodotti 2D e 3D ha consentito di formulare le seguenti considerazioni. La nuvola Laser Scanner ottenuta dal rilievo integrato svolto ha una spaziatura media più dettagliata, che consente di apprezzare meglio la geometria dell'oggetto; tuttavia, la rappresentazione in tonalità di grigi non facilita il riconoscimento degli oggetti da esaminare. La nuvola fotogrammetrica è un prodotto meno denso, infatti presenta un numero inferiore di punti e, di conseguenza, consente un grado di risoluzione metrica inferiore; nonostante ciò, la presenza del colore facilita la localizzazione dell'oggetto di interesse.

Le ortofoto ottenute da MAGO, a partire dalle nuvole di punti del Laser Scanner e fotogrammetrica, rappresentano due prodotti con risoluzione metrica confrontabile con la spaziatura media della nuvola Laser Scanner e con il GSD, senza però trascurare l'aspetto tematico.

Le ortofoto ottenute dalle due nuvole sono state confrontate, con risultati inaspettatamente migliori per il dato fotogrammetrico.

Sono state svolte valutazioni statistiche per quantificare l'assenza di dato all'interno delle ortofoto, alla luce del fatto che attualmente il software non effettua interpolazioni per il riempimento degli stessi.

Per perfezionare la procedura, occorrerà implementare metodi di interpolazione che consentano il riempimento delle assenze di dato isolate attualmente presenti sull'ortofoto; ciò è ammissibile in aree con andamento della superficie sufficientemente liscio. Un altro miglioramento in tal senso è la possibilità di generare un ortomosaico, utile ad eliminare le eventuali lacune di dato qualora siano localizzate sul singolo fotogramma. Inoltre sarà necessario focalizzare l'attenzione sull'introduzione di metodi statistici per il filtraggio della nuvola di punti in *input*, così come nella realizzazione di un modulo per scegliere liberamente il piano di proiezione dell'ortofoto.

## 6. BIBLIOGRAFIA

Agisoft Metashape©, 2019, <http://www.agisoft.com/> (5<sup>th</sup> March 2019)

CloudCompare, 2019, <http://www.cloudcompare.org/> (5<sup>th</sup> March 2019)

Erlandsson, R., Stoessel, M., Skånes, H., Wennbom, M., Angerbjörn, A., 2019. An innovative use of orthophotos – possibilities to assess plant productivity from colour infrared aerial orthophotos. *Remote Sensing in Ecology and Conservation*, <https://doi.org/10.1002/rse2.108>

Gagliolo, S., Federici, B., Ferrando, I., Sguerso, D., 2019. MAGO: a new approach for orthophotos production based on adaptive *mesh* reconstruction. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information*

<sup>4</sup> <https://grass.osgeo.org/>



Sciences, XLII-2/W11, 533-538, <https://doi.org/10.5194/isprs-archives-XLII-2-W11-533-2019>

GRASS GIS, 2019, <https://grass.osgeo.org/> (5<sup>th</sup> March 2019)

Nahhas, F.H., Shafri, H.Z.M., Sameen, M.I., Pradhan, B., 2018. Deep learning approach for building detection using LiDAR-Orthophoto fusion. *Journal of Sensors* 2018, <https://doi.org/10.1155/2018/7212307>

Niedzielski, T., Spallek, W., Witek-Kasprzak, M., 2018. Automated Snow Extent Mapping Based on Orthophoto Images from Unmanned Aerial Vehicles. *Pure and Applied Geophysics*, Volume 175, pp. 3285–3302. <https://doi.org/10.1007/s00024-018-1843-8>

Pérez, J.A., Gonçalves, G.R., Rangel, J.M.G., Ortega, P.F., 2019. Accuracy and effectiveness of orthophotos obtained from low cost UASs video imagery for traffic accident scenes documentation. *Advances in Engineering Software*, Volume

132, 2019, pp. 47-54, ISSN 0965-9978, <https://doi.org/10.1016/j.advengsoft.2019.03.010>

Perfetti, L., Fassi, F., Gulsan, H., 2019. Generation of Gigapixel orthophoto for the maintenance of complex buildings. Challenges and lesson learnt. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLII-2/W9, 605-614, <https://doi.org/10.5194/isprs-archives-XLII-2-W9-605-2019>

Sammartano, G., Spanò, A., 2017. High scale 3D modelling and orthophoto of curved masonries for a multipurpose representation, analysis and assessment. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLII-5/W1, 245-252, <https://doi.org/10.5194/isprs-archives-XLII-5-W1-245-2017>

Z+F LaserControl®, 2019, <https://www.zf-laser.com/> (5<sup>th</sup> March 2019)

# METODI DI CLASSIFICAZIONE AUTOMATICA TEXTURE- E GEOMETRY-BASED APPLICATI A BENI CULTURALI

## TEXTURE- AND GEOMETRY-BASED APPROACHES FOR THE CLASSIFICATION OF 3D HERITAGE

E. Grilli <sup>a,b</sup>

<sup>a</sup> 3D Optical Metrology (3DOM) unit, Bruno Kessler Foundation (FBK), Trento, Italy

<sup>b</sup> Department of Architecture, University of Bologna, Italy

grilli@fbk.eu

**PAROLE CHIAVE:** Classificazione, Nuvola di punti, Mesh, Machine Learning.

**KEY WORDS:** Classification, Point Cloud, Mesh, Machine Learning

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

### RIASSUNTO

L'applicazione delle nuove metodologie di rilievo e documentazione tridimensionale, attraverso l'utilizzo di sensori attivi e passivi, è ormai una pratica comune nel settore dei beni culturali. Per sfruttare le reali potenzialità di questa significativa quantità di dati e poter estrarre informazioni semantiche da nuvole di punti o modelli poligonali, emerge oggi la necessità di sviluppare metodi affidabili di classificazione. Questo articolo esplora l'uso di metodi di Machine e Deep Learning a supporto della classificazione 3D ai fini di studio, monitoraggio e restauro. In particolare, vengono presentati e confrontati tre diversi approcci di classificazione basati rispettivamente sull'utilizzo di features radiometriche, geometriche o una combinazione di entrambe.

### ABSTRACT

The continuous evolution in the last years of remote sensing technologies and methodologies for Cultural Heritage 3D documentation allowed to multiply photogrammetric and laser scanning acquisitions. At the same time, to exploit the real potential of this significant amount of data, the need for reliable and efficient methods to classify (i.e. semantically segment) point clouds or meshes has become a priority. This article explores the use of Machine and Deep Learning methods as support for studies, monitoring, and restoration purposes. More specifically, three different approaches based on texture, geometry, and texture plus geometry features are presented and compared.

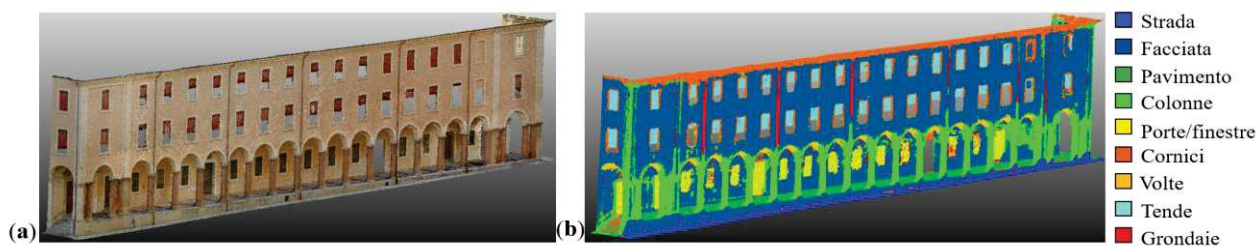


Figura 1. Nuvola di punti fotogrammetrica di un edificio rinascimentale a Bologna (85x18 m, 1.2 milioni di punti) (a); corrispondente nuvola di punti classificata automaticamente (b).

## 1. INTRODUZIONE

Il settore dei beni culturali è stato, negli ultimi decenni, uno dei campi di sperimentazione più interessanti delle nuove tecnologie e metodologie di rilievo digitale tridimensionale. L'evoluzione digitale ha profondamente trasformato le modalità di acquisizione, elaborazione e gestione dei dati in questo settore. Tuttavia, l'associazione di diverse informazioni semantiche ai prodotti del rilievo tridimensionale (nuvole di punti o modelli poligonali), attraverso metodi automatici di segmentazione e classificazione (Grilli et al., 2017), diventa fondamentale per l'interpretazione degli oggetti rilevati.

In generale, con il termine classificazione o segmentazione semantica si fa riferimento al processo di raggruppamento di dati

simili in sottoinsiemi (detti segmenti). Questi ultimi hanno in comune una o più caratteristiche (geometriche, radiometriche, ecc.), attraverso cui è possibile distinguere ed identificare le diverse parti che compongono un'immagine, una nuvola di punti o un modello poligonale. Tra le possibili applicazioni delle metodologie di classificazione 3D applicata al settore dei beni culturali troviamo:

- l'annotazione semantica dei dati 3D per scopi didattici;
- la suddivisione di grandi modelli in parti più piccole che possono essere gestite più facilmente;
- l'identificazione degli elementi architettonici, come fase preliminare per le applicazioni di Building Information Modeling (BIM) (Figura 1);

- l'identificazione e la quantificazione dei materiali da costruzione o stati di conservazione, utili per scopi di monitoraggio o restauro;
- la distinzione tra elementi architettonici strutturali e decorativi.

Tuttavia, lo sviluppo di una procedura automatica per la classificazione 3D è ancora un compito impegnativo, tenendo conto della complessità ed eterogeneità dei dati nel settore dei beni culturali. Nello specifico:

- l'identificazione delle classi non è mai univoca: a seconda dello scopo, più classi potrebbero descrivere lo stesso oggetto;
- gli elementi architettonici sono spesso scomponibili in sottoclassi (ad esempio, per le colonne possono essere individuate basi, fusti e capitelli; per le finestre, le parti decorative e strutturali, ecc.);
- è difficile classificare trasversalmente manufatti architettonici attraverso tecniche di Machine Learning, considerando le diverse peculiarità e caratteristiche dei beni di riconosciuto interesse storico e artistico.

In questo contesto, il lavoro presentato mira ad individuare soluzioni utili, affidabili, replicabili e pratiche per la classificazione sia di nuvole di punti che modelli poligonali di beni architettonici e archeologici, basandosi sull'applicazione di metodi di classificazione Machine Learning supervisionati.

L'articolo discute in primo luogo lo stato dell'arte in materia di classificazione 3D. Vengono, di seguito, valutate tre diverse metodologie di classificazione supervisionata applicate a tre diversi casi di studio: il Sarcofago degli Sposi, la tomba Bartoccini ed il Tempio di Nettuno. Infine, vengono riassunte e discusse le conoscenze acquisite.

## 2. STATO DELL'ARTE

La comunità geomatica ha mostrato, negli ultimi anni, un crescente interesse nel settore della segmentazione e classificazione dei dati tridimensionali ottenuti da rilievi reality-based. Notevoli progressi, nella ricerca di procedure automatiche di classificazione per l'analisi e la comprensione dei modelli 3D, sono attribuibili all'avvento del Machine Learning (Hackel et al., 2016; Weinmann et al., 2017; Wang et al., 2018). Machine e Deep Learning (ML / DL) sono campi di applicazione dell'Intelligenza Artificiale (AI), basati sullo sviluppo di algoritmi che permettono ai calcolatori/computer di prendere decisioni basate su dati iniziali di training. Deep Learning nello specifico viene considerato una branca del Machine Learning che si basa sull'utilizzo delle reti neurali artificiali (neural network) con due o più strati ( hidden layers ) per elaborare le informazioni in modo non lineare.

Diversi benchmark, con dati terrestri o aerei annotati, sono stati resi disponibili per testare e convalidare algoritmi. La maggior parte di questi datasets, fanno riferimento a territori rurali o scene urbane, come Semantic3d (Hackel et al., 2017) o The Cityscapes Dataset (Cordts et al., 2016). Mentre in questi scenari le classi di oggetti e le loro etichette sono ormai definite (terreno, strade, alberi, edifici, ecc.), l'identificazione di categorie precise nel settore dei beni culturali è molto più complessa. Ad oggi, le uniche banche dati del patrimonio architettonico annotato sono in 2D, come per le facciate dell'Ecole Centrale Paris (ECP) Facades dataset (Teboul et al., 2010), del eTraining for Interpreting Images of Man-Made Scenes (eTRIMS)(Korc and Forstner, 2009) o del Center for Machine Perception (CMP) Facade Database (Tyleček and Šára, 2013).

Diverse tecniche di Machine Learning sono state, tuttavia, applicate al contesto architettonico. Oses et al. (2014) applicano modelli di Machine Learning per la segmentazione di pareti in

muratura lavorando su immagini. Amato et al. (2015) affrontano il problema del riconoscimento di monumenti nelle immagini utilizzando tecniche di classificazione k-nearest neighbour (kNN). Le Convolutional Neural Networks (CNN) vengono applicate per la prima volta a scenari del patrimonio culturale in Llamas et al. (2016).

Per le applicazioni in ambito 3D, generalmente la classificazione viene eseguita manualmente e le annotazioni trasferite dal 2D al 3D (Campanaro et al., 2016; Grilli et al., 2018). Apollonio et al. (2018), nell'ambito del progetto di restauro della Fontana del Nettuno a Bologna, propongono l'annotazione diretta delle tipologie di degrado sui modelli 3D. Un esempio analogo, è dato dalla piattaforma web Aioli. L'utente, in questo caso, annotando informazioni su immagini utilizzate per la ricostruzione del modello fotogrammetrico, è in grado di trasferire in tempo reale dati semantici sul modello tridimensionale (Roussel et al., 2019). Futuri sviluppi di questa piattaforma prevedono l'utilizzo di metodi semiautomatici per i processi di annotazione.

Diversi metodi, in definitiva, sono stati sviluppati per la classificazione di immagini di edifici o monumenti di riconosciuto interesse architettonico, ma pochi sono i riferimenti a strategie di classificazione automatica di modelli tridimensionali (Poux et al., 2017; Grilli et al., 2019; Malinverni et al., 2019).

## 3. METODI DI CLASSIFICAZIONE 3D

In questo lavoro, vengono presentate diverse strategie di classificazione applicabili a nuvole di punti o mesh, basate sull'utilizzo di algoritmi di Machine e Deep Learning con apprendimento supervisionato. Con il termine "Machine Learning" ("apprendimento automatico"), si fa riferimento all'abilità delle macchine (computer) di apprendere senza essere state esplicitamente e preventivamente programmate. L'apprendimento supervisionato è una categoria del Machine Learning che prevede l'utilizzo di un set di dati in input e di informazioni relative ai risultati attesi, per l'identificazione di una regola generale che colleghi i dati in ingresso con quelli in uscita.

Per la classificazione di nuvole di punti o modelli poligonali di edifici o elementi architettonici, le categorie semantiche vengono apprese dai dati di input forniti manualmente sotto forma di patch (nel caso vengano annotate le texture) o porzioni annotate di nuvola (nel caso si lavori sul 3D). Oltre alle classi semantiche, ai dati di training sono associate le cosiddette features, particolari caratteristiche radiometriche o geometriche che permettono di distinguere una classe dall'altra. Queste informazioni vengono utilizzate per addestrare un modello predittivo specifico (ad esempio Random Forest) o una rete neurale, in grado di predire con un certo livello di accuratezza la segmentazione semantica dell'intero dataset.

Di seguito vengono presentati tre diversi approcci implementati per facilitare le esigenze di classificazione di scenari eterogenei:

- approccio texture-based, basato sull'utilizzo dei dati radiometrici delle texture dei modelli (sezione 3.1);
- approccio geometry-based, che sfrutta le proprietà geometriche delle nuvole di punti (sezione 3.2);
- approccio combinato texture- e geometry-based (sezione 3.3).

### 3.1 Classificazione texture-based

Considerando la vasta disponibilità in letteratura di sistemi Machine Learning applicati alla classificazione di immagini, il primo metodo proposto per la classificazione di modelli 3D si basa sull'individuazione delle classi a partire da dati 2D.

In base alla complessità dell'oggetto trattato, questa metodologia è applicabile ad ortofoto (nel caso di oggetti planari quali murature) o mappe UV, ovvero proiezioni su 2D della superficie dei modelli 3D.

Partendo da modelli 3D texturizzati, la pipeline sviluppata si basa su 3 step:

- generazione di ortofoto o mappe UV;
- classificazione supervisionata delle ortofoto o mappe UV;
- ri-proiezione dei risultati della classificazione 2D sui modelli 3D.

Per la classificazione delle texture, è stato utilizzato il plugin Trainable WeKa Segmentation (Witten et al., 2016), disponibile all'interno della distribuzione Fiji di ImageJ (Schindelin et al., 2012). Weka offre la possibilità di annotare manualmente piccole porzioni dell'immagine, estrarre features da informazioni radiometriche ed estendere la classificazione istruendo uno dei diversi modelli predittivi disponibili.

In questo lavoro, viene utilizzato un classificatore Random Forest (Breiman, 2001). Un esempio applicativo è fornito alla Sezione 4.1.

### 3.2 Classificazione geometry-based

In assenza di particolari caratteristiche colorimetriche e materiche che consentano di evidenziare le classi/componenti sulle texture, la classificazione deve tener conto delle proprietà geometriche del dato tridimensionale.

L'approccio geometry-based viene riassunto alla Figura 2.

In questo caso, le tecniche supervisionate di machine/deep learning vengono applicate alle nuvole di punti. Per la fase di training, questo approccio prevede l'identificazione e la segmentazione manuale delle classi.

L'estrazione delle covarianza features consente l'individuazione delle caratteristiche geometriche locali del dato 3D. Queste features, derivate dalla matrice di covarianza (Cheata et al., 2009), sono in grado di descrivere la distribuzione dei punti della nuvola in un determinato intorno. Vengono, in questo modo, evidenziati i comportamenti locali della nuvola (la linearità, planarità e sfericità o ancora misure quali l'onnivarianza e l'anisotropia (Blomey et al., 2014).

La scelta delle features geometriche ed il raggio di intorno del punto è argomento trattato dall'autore in un precedente articolo

(Grilli et al., 2019). A partire dal dato di training, possono quindi essere istruiti diversi modelli predittivi (machine o deep learning) che hanno il compito di estendere la classificazione al resto del dataset.

In questo lavoro, viene utilizzata un'implementazione Machine Learning Random Forest disponibile nella libreria Python di Scikit-learn (versione 0.21.1) (Pedregosa et al., 2011) ed una rete neurale 2D Convolutional Neural Network (CNN) (Ozdemir and Remondino, 2019a; Ozdemir and Remondino, 2019b).

Come strumento di valutazione dei risultati della classificazione, è valutata la matrice di confusione. Nell'ambito dell'Intelligenza Artificiale, la matrice di confusione restituisce una rappresentazione dell'accuratezza della classificazione statistica. Presa in considerazione una porzione del dataset, per ogni classe segmentata, la colonna della matrice rappresenta i valori predetti ed ogni riga rappresenta i valori reali annotati manualmente. Questa tabella evidenzia le performance del modello predittivo di classificazione, in modo da determinare quanto questo modello sia accurato ed efficace, considerando le seguenti metriche:

$$Accuracy = \frac{Tp + Tn}{Tp + Fp + Tn + Fn} \quad (1)$$

$$Precision = \frac{Tp}{Tp + Fp} \quad (2)$$

$$Recall = \frac{Tp}{Tp + Fn} \quad (3)$$

$$F1\ score = 2 * \frac{Recall * Precision}{Recall + Precision} \quad (4)$$

Tali metriche fanno riferimento non solo ai valori correttamente classificati (TP = true positive e TN = true negative) ma anche a quelli erronei (FP = false positive e FN = false negative).

Alla Sezione 4.2 è possibile trovare un esempio pratico di classificazione Machine e Deep Learning basato sull'uso di sole features geometriche.

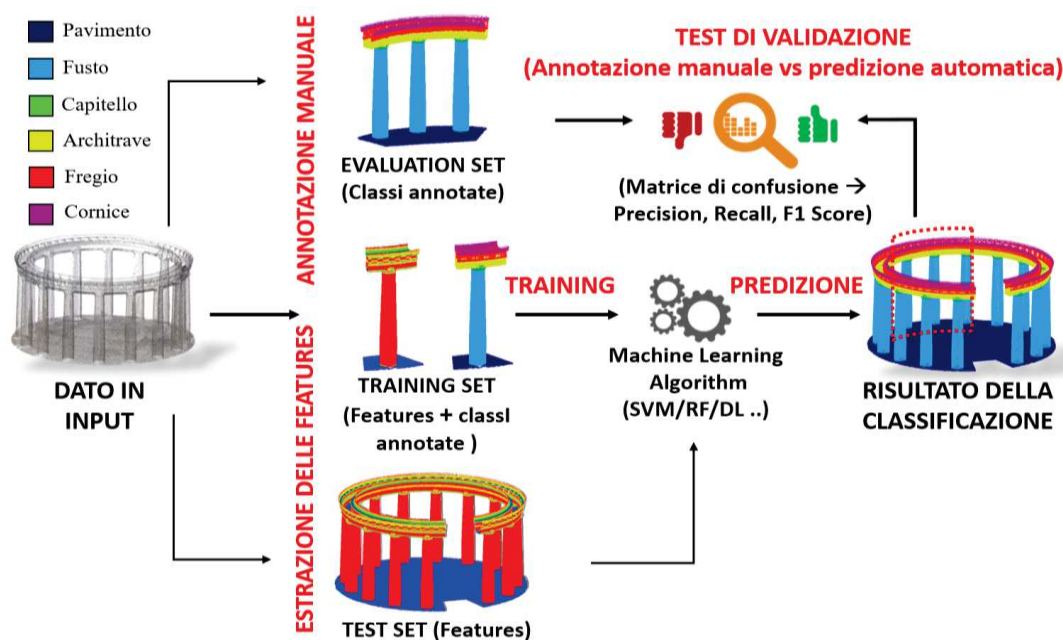


Figura 2. Workflow esemplificativo dell'approccio di classificazione supervisionata geometry-based.

### 3.3 Classificazione texture- & geometry-based

In alcuni casi, l'utilizzo in input delle sole informazioni radiometriche o geometriche può risultare insufficiente per classificare in maniera soddisfacente il dato 3D. Il tool di classificazione disponibile all'interno della libreria Computational Geometry Algorithms Library (CGAL)(Giraudot and Lafrange, 2019) permette la combinazione di entrambe. Come nell'approccio precedente, le classi di interesse vengono manualmente annotate sulla nuvola di punti. Le features geometriche (distanza dal piano, elevazione, dispersione verticale, verticalità, ecc.) vengono calcolate sulla nuvola a diverse scale di dettaglio, in modo da poter estrapolare ad esempio cambiamenti di direzione dei punti della nuvola sia macro che micro (Brodu and Lague, 2012). Insieme alle features geometriche vengono inoltre considerate quelle colorimetriche, convertendo i valori RGB della nuvola di punti allo spazio colore HSV, meno sensibile ai cambi di illuminazione (Sural et al., 2002). Preparati i dati di training, viene istruito un algoritmo Random Forest (ETH Zurich Random Forest Template, 2015), che estende la classificazione all'intero dataset. Un esempio applicativo di questo tipo di classificazione è riportato nella Sezione 4.3.

## 4. CASI STUDIO E RISULTATI

La scelta della migliore metodologia di classificazione di modelli archeologici/architettonici è funzione della complessità dell'oggetto e dello scopo della segmentazione. I metodi di classificazione descritti nella Sezione 3 sono stati applicati a tre diversi scenari archeologici:

- Sarcofago degli Sposi (Figura 3), situato nel museo di Villa Giulia a Roma (Italia): si tratta di un sarcofago etrusco di fine VI secolo a.C., alto 1,14 m per 1,9 m di larghezza e 0,8 m di profondità, in terracotta un tempo dipinta con colori vivaci e brillanti. L'approccio texture-based (Sezione 3.1) è stato applicato per identificare le anomalie superficiali (cioè fratture e decadimenti) e quantificare la quantità di cemento mimetico utilizzato per ri-assemblare il manufatto.



Figura 3. Sarcophago degli Sposi, Roma, Italy.

- Tempio di Nettuno (Figura 4) a Paestum (Italia): misura circa 24,5 m x 60 m x 17 m di altezza nel suo punto più alto. Si compone di 6 colonne frontali e 14 colonne laterali, mentre nella zona interna presenta due file di colonne a doppio ordine che dividono il naos in tre parti. L'approccio geometry-based mira ad identificare i principali elementi architettonici che compongono il tempio greco.



Figura 4. Tempio di Nettuno a Paestum, Italy.

- Tomba Bartoccini (Figura 5) a Tarquinia (Italia): la tomba etrusca, risalente al IV secolo a.C., si compone di quattro stanze ancora interamente affrescate - una centrale (circa 5 m per 4 m) e tre ai lati (circa 3 m per 3 m) di altezza non superiore a 3 m. Combinando l'uso di features geometriche e colorimetriche, la classificazione mira all'individuazione e alla quantificazione delle aree che presentano degradi dell'intonaco o hanno subito infiltrazioni.

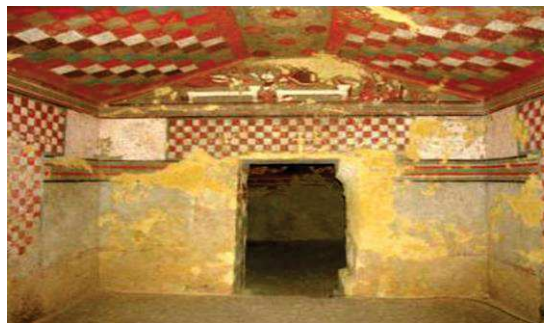


Figura 5. Tomba Bartoccini a Tarquinia, Italy

### 4.1 Sarcophago degli Sposi, museo di Villa Giulia, Roma

Il "Sarcophago degli Sposi", di epoca etrusca, è stato ritrovato nel 1881 a Tarquinia (Italia), frammentato in più di 400 pezzi. Gli interventi successivi di restauro, con l'utilizzo di malta cementizia mimetica, hanno consentito la ricomposizione del sarcofago. Diverse metodologie e strumentazioni di rilievo reality-based (fotogrammetria, scansione laser a tempo di volo e a triangolazione), nel 2013, sono state impiegate per la realizzazione del modello tridimensionale (Menna et al., 2013). La mappa UV ad alta risoluzione (4096x4096 pixel), ottenuta dal complesso modello 3D (Figura 6a) (cinque milioni di triangoli), è stata manualmente annotata, identificando tre categorie principali ed una accessoria per la classificazione (Figura 6b). L'annotazione manuale delle patch necessarie al training ha richiesto circa 15 minuti ed è stata assistita da esperti di restauro. Una volta estratte le features radiometriche, il classificatore Random Forest ha esteso la classificazione all'intera mappa UV (Figura 6c). Una volta classificata, la texture è stata ri-proiettata sulla geometria 3D, permettendo di evidenziare i diversi dettagli del manufatto etrusco ricomposto (Figura 7). I piedi di sostegno del sarcofago (le uniche parti del sarcofago con visibili decorazioni a pigmento) sono stati esclusi dalla classificazione. Il modello classificato è stato, infine, analizzato per individuare la distribuzione delle diverse classi individuate. Il dato più interessante emerso è che il 12% della superficie del sarcofago sarebbe composta da malta mimetica. Conoscendo inoltre la superficie complessiva del sarcofago (6.8 m<sup>2</sup>), misurata a partire dal modello poligonale, si deriva che circa 0.8m<sup>2</sup> sono occupati da parti ricostruite. Un video rappresentativo della classificazione del Sarcophago è disponibile al seguente link: <https://www.youtube.com/watch?v=ZmjUh3xn6eA>

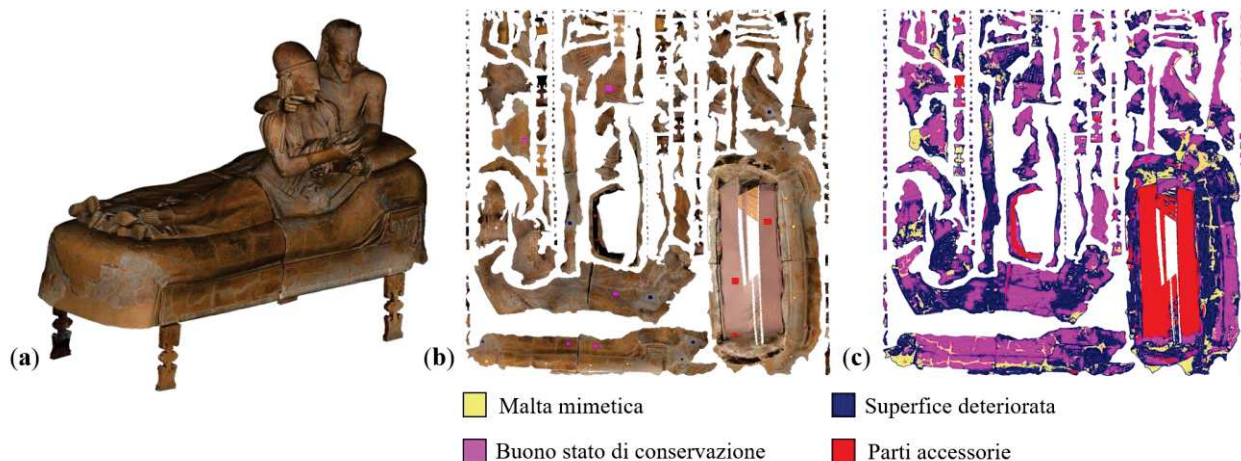


Figura 6. Modello fotogrammetrico ad alta risoluzione del Sarcofago degli sposi (a); classi manualmente annotate sulla mappa UV (b); risultati della classificazione sulla mappa UV (c).



Figura 7. Modello 3D classificato del Sarcofago degli Sposi: evidenziate in giallo le aree occupate da malta mimetica.

#### 4.2 Tempio di Nettuno, parco archeologico di Paestum

Il tempio di Nettuno, costruito nell'antica polis di Paestum intorno alla metà del V secolo a.C., viene considerato tra gli esempi perfetti dell'architettura dorica templare in Italia e in Grecia. La nuvola di punti è il risultato di un rilievo fotogrammetrico con immagini acquisite da piattaforme UAV e terrestri (Fiorillo et al., 2013). La nuvola sotto-campionata, per migliorare le prestazioni computazionali, è costituita da circa 2.2 milioni di punti. Le classi semantiche del tempio sono state individuate e manualmente annotate su piccole porzioni della nuvola di punti (Figura 8). L'operazione di segmentazione manuale ai fini di annotazione ha richiesto all'operatore circa mezz'ora.

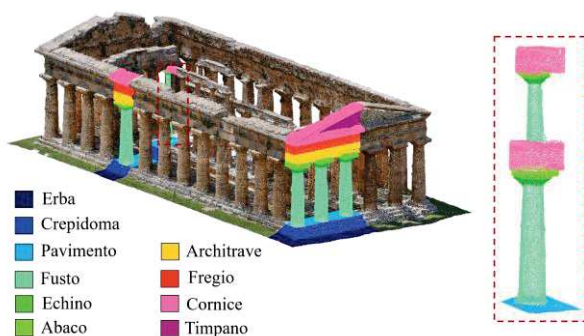


Figura 8. Nuvola di punti con classi annotate manualmente.

Le features geometriche significative per la distinzione degli elementi architettonici sono state estratte utilizzando CloudCompare (Hackel et al., 2016) (Figura 9).

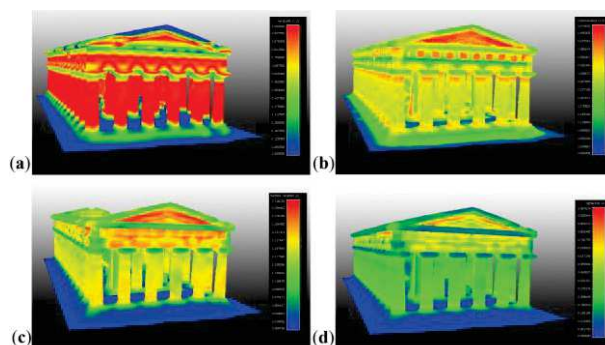


Figura 9. Alcune features geometriche estratte sul dataset del Tempio: verticalità (a), onnivarianza (b), sfericità (c), variazione di superficie (d).

Preparato (in un'ora circa) il cosiddetto training set con classi annotate e features estratte, la classificazione è stata estesa all'intero dataset mediante i modelli predittivi Random Forest e CNN 2D.

Classificato il dato 3D, è possibile visualizzare le diverse classi semantiche separatamente. In previsione di un riutilizzo del dato segmentato in ambiente BIM, questa possibilità è particolarmente rilevante (Figura 10). Le tabelle 2 e 3 (situate al

termine dell'articolo) riportano le matrici di confusione e le metriche di precisione estratte rispettivamente con RF e CNN. In termini di precisione, i risultati ottenuti sono comparabili. In riferimento ai tempi di elaborazione, l'approccio Random Forest ha impiegato pochi minuti per classificare l'intero dataset, mentre la CNN ha richiesto circa 1 ora. Per vedere nel dettaglio la nuvola di punti, le features estratte e i risultati della classificazione è disponibile un video al seguente link:

<https://www.youtube.com/watch?v=8-muH633ud8&t=3s>

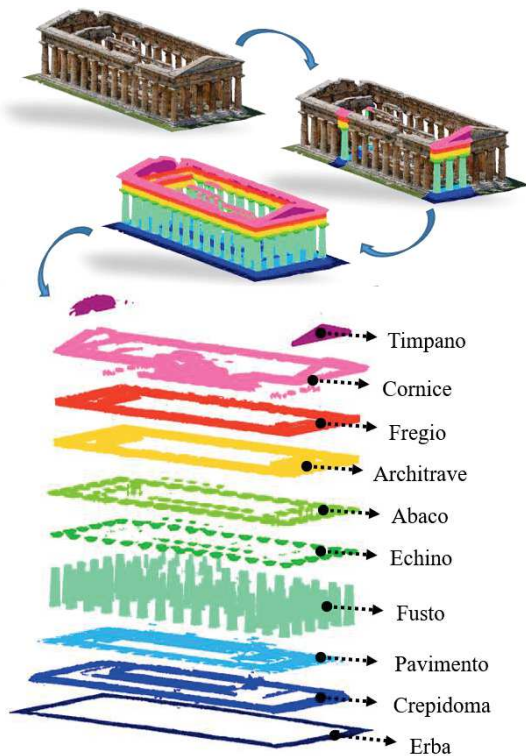


Figura 10. Risultato della classificazione geometry-based applicata alla nuvola del Tempio di Nettuno a Paestum.

### 4.3 Tomba Bartoccini

La tomba Bartoccini, risalente al IV secolo a.C., fu scoperta nel 1959 a Tarquinia (Italia). Nel 2013, un modello 3D completo e

texturizzato della tomba (Fernandez Jimenez et al., 2013) è stato ottenuto con tecniche integrate di rilievo reality-based. L'esposizione a diversi agenti atmosferici e fenomeni di umidità di risalita hanno compromesso nei secoli lo stato di conservazione della tomba, che presenta fenomeni di degrado superficiali e profondi sulle murature. La classificazione, in questo caso, ha avuto come obiettivo l'identificazione e la quantificazione delle superfici deteriorate presenti sulle pareti affrescate.

Applicando il metodo descritto alla Sezione 3.2 ed utilizzando features geometriche e radiometriche, sono state annotate in tempi ridotti (circa 10 minuti) piccole porzioni ben distribuite della nuvola di punti (Figura 12a). Il classificatore Random Forest ha, quindi, predetto la classificazione sull'intero dataset (Figura 12b). L'intera elaborazione ha richiesto circa un'ora di tempo. La decisione di combinare features geometriche e radiometriche ha due motivazioni:

- considerando la sola texture, sono emersi alcuni errori di classificazione dovuti alla somiglianza cromatica tra le superfici deteriorate e alcuni dipinti murali;
- utilizzando solo la geometria, è difficile identificare i degradi più superficiali.

Per valutare matematicamente le prestazioni della classificazione, è stata considerata una singola parete della tomba (Figura 11a), annotata manualmente (Figura 11b) e confrontata con i risultati della classificazione automatica (Figura 11c). È stato, dunque, calcolato il rapporto tra il numero di punti correttamente classificati ed il numero totale di punti:

$$Accuracy = \frac{N \text{ of point correctly classified}}{\text{total number of point}} \quad (5)$$

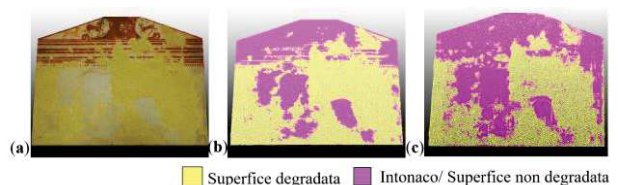


Figura 11. Test di validazione della tomba Bartoccini (a), annotata manualmente (b), classificata automaticamente (c).

Come risultato, è emerso che il 90% della previsione era corretto.

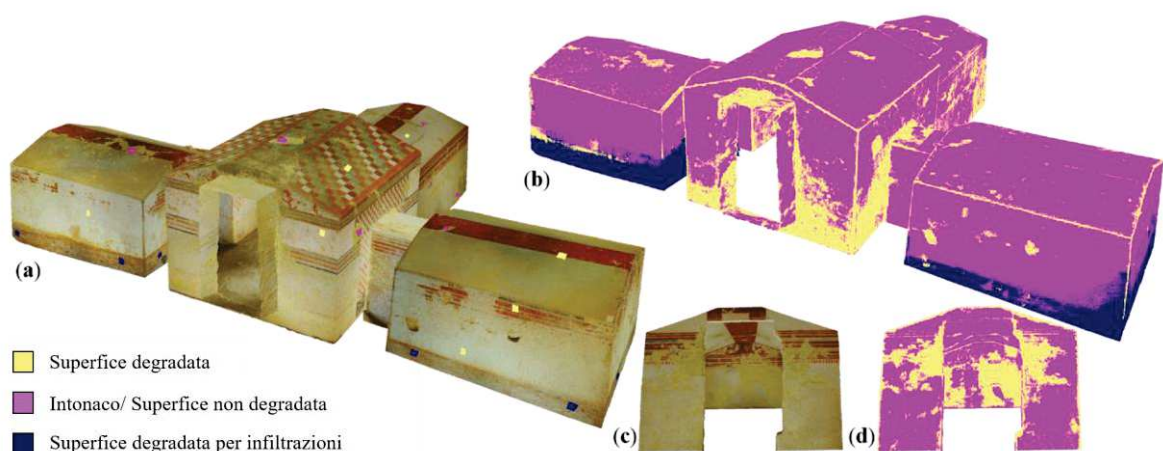


Figura 12. Nuvola di punti della tomba Bartoccini (circa 5 milioni di punti) con evidenziate le parti annotate manualmente (a); risultato della classificazione ottenuto utilizzando features geometriche e radiometriche (b). Dettaglio di un ambiente (c)(d).

Infine, sono state quantificate le superfici erose dell'intera tomba, calcolando la percentuale che ogni classe segmentata occupava rispetto alla superficie totale. La Tabella 1 riporta per ogni classe l'area occupata in termini di numero di punti, percentuale occupata e area in m<sup>2</sup>. Questo tipo di risultato potrebbe significativamente migliorare la pianificazione di interventi di restauro o monitoraggio dello stato di conservazione della tomba. Un video sui risultati della classificazione della tomba di Batoccini è disponibile al seguente link: <https://www.youtube.com/watch?v=omI9xLroSDk&t=4s>

| CLASSE                      | N° punti | % occupata | Area (m <sup>2</sup> ) |
|-----------------------------|----------|------------|------------------------|
| <b>Superficie degradata</b> | 3539781  | 22.62%     | 42                     |
| <b>Intonaco</b>             | 11014402 | 70.38%     | 131                    |
| <b>Infiltrazioni</b>        | 1095229  | 7.00%      | 13                     |
| <b>TOT</b>                  | 15649412 | 100.00%    | 186                    |

Tabella 1. Misura in percentuale e m<sup>2</sup> delle aree occupate da ciascuna classe.

## 5. CONCLUSIONI

In questa ricerca sono stati sviluppati e testati diversi metodi di classificazione che permettono di arricchire semanticamente dati 3D appartenenti alla sfera dei beni culturali, lavorando sulla texture, sulla geometria 3D o combinando entrambe le informazioni. Risultati interessanti sono stati raggiunti in tempi di elaborazione ragionevoli rispetto ad un approccio manuale:

- Sarcofago degli sposi: partendo da un modello complesso, utilizzando la texture ad alta risoluzione, è stato possibile distinguere diversi stati di conservazione e stimare la quantità di malta cementizia mimetica utilizzata durante la ricomposizione del sarcofago etrusco. Questi risultati potrebbero aiutare studi sul processo di ricomposizione del manufatto o per monitorarne lo stato di conservazione. L'intero processo di classificazione ha richiesto circa 1 ora, con un input manuale di 10 minuti.
- Tempio di Nettuno: attraverso un'analisi dettagliata delle features geometriche (Grilli et al., 2019) ed un'accurata annotazione manuale di alcune porzioni della nuvola (20 minuti richiesti), è stato possibile ottenere in tempi brevi (RF 12 minuti, CNN circa un'ora) ottimi risultati di segmentazione semantica, sia utilizzando algoritmi machine che deep learning.
- Tomba Bartoccini: dopo aver individuato la migliore strategia per classificare il caso studio, la nuvola di 5 milioni di punti è stata interamente classificata in meno di un'ora, di cui 10 minuti sono stati dedicati all'annotazione manuale. L'accuratezza dei risultati calcolati su un set di prova ha dato come risultato una precisione della previsione del 90%. I risultati della classificazione permettono di calcolare la quantità di superfici deteriorate, fornendo informazioni utili in caso di futuri lavori di restauro.

Considerando i risultati presentati e le precedenti esperienze (Grilli and Remondino, 2019; Grilli et al., 2019), l'autore ritiene che, nel settore dei beni culturali, ogni caso di studio debba essere considerato singolarmente. Per ogni oggetto trattato è essenziale capire quali siano gli scopi della classificazione, le classi ricercate e il livello di dettaglio da raggiungere. La complessità dell'oggetto, dimensioni e caratteristiche materiche e colorimetriche, contribuiscono alla scelta della migliore metodologia da applicare (texture- o geometry-based).

## BIBLIOGRAFIA

- Amato, G., Falchi, F. and Gennaro, C., 2015. Fast image classification for monument recognition. *Journal on Computing and Cultural Heritage (JOCCH)*, 8(4), p.18.
- Apollonio, F.I., Basilissi, V., Callieri, M., Dellepiane, M., Gaiani, M., Ponchio, F., Rizzo, F., Rubino, A.R. and Scopigno, R., 2018. A 3D-centered information system for the documentation of a complex restoration intervention. *Journal of Cultural Heritage*, 29, pp.89-99.
- Blomley, R., Weinmann, M., Leitloff, J., and Jutzi, B., 2014. Shape distribution features for point cloud analysis-a geometric histogram approach on multiple scales. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 2(3), p. 9.
- Breiman, L., 2001. Random forests. *Machine learning*, 45(1), pp. 5-32.
- Brodu, N. and Lague, D., 2012. 3D terrestrial lidar data classification of complex natural scenes using a multi-scale dimensionality criterion: Applications in geomorphology. *ISPRS Journal of Photogrammetry and Remote Sensing*, 68, pp.121-134.
- Campanaro, D. M., Landeschi, G., Dell'Unto, N. and Touati, A.-M. L., 2016. 3D GIS for cultural heritage restoration: A 'white box' workflow. *Journal of Cultural Heritage*, 3, Volume 18, pp. 321-332.
- Chehata, N., Guo, L. and Mallet, C., 2009. Airborne lidar feature selection for urban classification using random forests. *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, 38(3/W8).
- Cordts, M., Omran, M., Ramos, S., Rehfeld, T., Enzweiler, M., Benenson, R., Franke, U., Roth, S. and Schiele, B., 2016. The cityscapes dataset for semantic urban scene understanding. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 3213-3223.
- ETH Zurich Random Forest Template, 2015. Available online: [www.prs.igp.ethz.ch/research/Source\\_code\\_and\\_datasets.html](http://www.prs.igp.ethz.ch/research/Source_code_and_datasets.html) (accessed on 15 May 2019).
- Fernández-Palacios, B.J., Rizzi, A. and Remondino, F., 2013. Etruscans in 3D-Surveying and 3D modeling for a better access and understanding of heritage. *Virtual Archaeology Review*, 4(8), pp.85-89.
- Fiorillo, F., Fernández-Palacios, B. J., Remondino, F., and Barba, S., 2013. 3D Surveying and modelling of the Archaeological Area of Paestum, Italy. *Virtual Archaeology Review*, Vol. 4(8), pp. 55-60.
- Giraudot, S. and Lafarge, F., 2019. Classification. In *CGAL User and Reference Manual*. CGAL Editorial Board, 4.14 edition.
- Grilli, E., Menna, F. and Remondino, F., 2017. A review of point clouds segmentation and classification algorithms. *The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, 42, p.339.
- Grilli, E., Dinunno, D., Petrucci, G. and Remondino, F., 2018. From 2D to 3D supervised segmentation and classification for cultural heritage applications. *International Archives of the Photogrammetry, Remote Sensing & Spatial Information Sciences*, 42(2), pp. 399-406.
- Grilli, E. and Remondino, F., 2019. Classification of 3D Digital Heritage. *Remote Sensing*, 11(7), p.847.



- Grilli, E., M. Farella, E., Torresani, A. and Remondino, F., 2019. Geometric features analysis for the classification of cultural heritage point clouds. *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, Vol. XLII-2/W15, pp 541-548.
- Hackel, T., Wegner, J.D., Schindler, K., 2016. Fast semantic segmentation of 3D point clouds with strongly varying density. *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences*, Vol. III(3), pp. 177-184.
- Hackel, T., Savinov, N., Ladicky, L., Wegner, J.D., Schindler, K. and Pollefeys, M., 2017. Semantic3d. net: A new large-scale point cloud classification benchmark. *arXiv preprint arXiv:1704.03847*.
- Korc, F. and Förstner, W., 2009. eTRIMS Image Database for interpreting images of man-made scenes. *Dept. of Photogrammetry, University of Bonn, Tech. Rep. TR-IGG-P-2009-01*.
- Llamas, J., Leronés, P. M., Zalama, E., and Gómez-García-Bermejo, J., 2016, October: Applying deep learning techniques to cultural heritage images within the INCEPTION project. In *Euro-Mediterranean Conference* (pp. 25-32). Springer, Cham.
- Malinverni, E. S., Pierdicca, R., Paolanti, M., Martini, M., Morbidoni, C., Matrone, F., and Lingua, A., 2019. Deep learning for semantic segmentation of 3D point cloud. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, XLII-2/W15, pp. 735–742.
- Menna, F., Nocerino, E., Remondino, F., Dellepiane, M., Callieri, M. and Scopigno, R., 2016. 3D digitization of a heritage masterpiece- A critical analysis on quality assessment. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 41(B5).
- Oses, N., Dornaika, F., and Moujahid, A., 2014. Image-based delineation and classification of built heritage masonry. *Remote Sensing*, 6(3), pp 1863-1889.
- Ozdemir, E., Remondino, F., 2019a. Classification of aerial point clouds with deep learning. *Int. Archives of the Photogrammetry, Remote Sensing and Spatial Information Science*. Proceedings Geospatial Week 2019, in press.
- Ozdemir, E., Remondino, F., 2019b. Aerial point cloud classification with deep learning and machine learning algorithms. *Int. Archives of the Photogrammetry, Remote Sensing and Spatial Information Science*. Proceedings Geospatial Conference 2019, in press.
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V. and Vanderplas, J., 2011. Scikit-learn: Machine learning in Python. *Journal of machine learning research*, 12(Oct), pp.2825-2830.
- Poux, F., Hallot, P., Neuville, R. and Billen, R., 2016. Smart Point Cloud: Definition and Remaining Challenges. *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences*, 10, Volume IV-2/W1, pp. 119-127.
- Roussel R., Bagnéris M., De Luca L. and Bomblet P., 2019. A digital diagnosis for the <<autumn>> statue (Marseille, France): photogrammetry, digital cartography and construction of a thesaurus. *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, Vol. XLII-2/W15, pp. 1039–1046.
- Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., Preibisch, S., Rueden, C., Saalfeld, S., Schmid, B. and Tinevez, J.Y., 2012. Fiji: an open-source platform for biological-image analysis. *Nature methods*, 9(7), p.676.
- Sural, S., Qian, G. and Pramanik, S., 2002. Segmentation and histogram generation using the HSV color space for image retrieval. In *Proceedings. International Conference on Image Processing*, Vol. 2, pp. 589-592. IEEE.
- Teboul, O., Simon, L., Koutsourakis, P. and Paragios, N., 2010, June. Segmentation of building facades using procedural shape priors. In *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 3105-3112.
- Teboul, O., Kokkinos, I., Simon, L., Koutsourakis, P. and Paragios, N., 2013. Parsing facades with shape grammars and reinforcement learning. *IEEE transactions on pattern analysis and machine intelligence*, 35(7), pp.1744-1756.
- Tyleček, R. and Šára, R., 2013, September. Spatial pattern templates for recognition of objects with regular structure. In *German Conference on Pattern Recognition* (pp. 364-374). Springer, Berlin, Heidelberg.
- Wang, Y., Sun, Y., Liu, Z., Sarma, S.E., Bronstein, M.M. and Solomon, J.M., 2018. Dynamic graph CNN for learning on point clouds. *arXiv preprint arXiv:1801.07829*.
- Weinmann, M. and Weinmann, M., 2017. Geospatial Computer Vision Based on Multi-Modal Data - How Valuable Is Shape Information for the Extraction of Semantic Information? *Remote Sensing*, 10(1), p.2.
- Witten, I.H., Frank, E., Hall, M.A. and Pal, C.J., 2016. *Data Mining: Practical machine learning tools and techniques*. Morgan Kaufmann.

| CLASS NAME     | Erba  | Crep. | Pavim. | Fusto  | Echino | Abaco | Architr | Fregio | Cornice | Timp. | Prec.         | Recall        | F1            |
|----------------|-------|-------|--------|--------|--------|-------|---------|--------|---------|-------|---------------|---------------|---------------|
| Erba           | 56998 | 1647  | 0      | 0      | 0      | 0     | 0       | 0      | 0       | 0     | 97.19%        | 99.88%        | 98.52%        |
| Crepidoma      | 67    | 38389 | 484    | 497    | 0      | 0     | 0       | 0      | 0       | 0     | 97.34%        | 93.43%        | 95.34%        |
| Pavimento      | 0     | 958   | 62211  | 993    | 0      | 0     | 0       | 0      | 0       | 0     | 96.96%        | 99.03%        | 97.98%        |
| Fusto          | 0     | 54    | 125    | 169950 | 172    | 56    | 4       | 0      | 268     | 0     | 99.60%        | 98.49%        | 99.04%        |
| Echino         | 0     | 0     | 0      | 966    | 18717  | 988   | 4       | 0      | 40      | 0     | 90.35%        | 86.62%        | 88.45%        |
| Abaco          | 0     | 0     | 0      | 63     | 2651   | 23941 | 3998    | 0      | 636     | 0     | 76.52%        | 87.54%        | 81.66%        |
| Architrave     | 0     | 0     | 0      | 0      | 0      | 2140  | 50903   | 2600   | 1       | 0     | 91.48%        | 90.84%        | 91.16%        |
| Fregio         | 0     | 0     | 0      | 0      | 0      | 0     | 1036    | 34609  | 334     | 33    | 96.10%        | 79.98%        | 87.31%        |
| Cornice        | 0     | 42    | 0      | 90     | 67     | 223   | 90      | 4569   | 106415  | 2071  | 93.70%        | 95.41%        | 94.55%        |
| Timpano        | 0     | 0     | 0      | 0      | 0      | 0     | 4       | 1492   | 3841    | 21393 | 80.03%        | 91.05%        | 85.19%        |
| <b>AVERAGE</b> |       |       |        |        |        |       |         |        |         |       | <b>91.93%</b> | <b>92.23%</b> | <b>91.92%</b> |

Tabella 2. Matrice di confusione e metriche di precisione ottenute dalla classificazione con Random Forest.

| CLASS NAME     | Erba  | Crep. | Pavim. | Fusto  | Echino | Abaco | Architr | Fregio | Cornice | Timp. | Prec.         | Recall        | F1            |
|----------------|-------|-------|--------|--------|--------|-------|---------|--------|---------|-------|---------------|---------------|---------------|
| Erba           | 57429 | 1216  | 0      | 0      | 0      | 0     | 0       | 0      | 0       | 0     | 97.93%        | 99.77%        | 98.84%        |
| Crepidoma      | 133   | 38380 | 283    | 641    | 0      | 0     | 0       | 0      | 0       | 0     | 97.32%        | 93.77%        | 95.51%        |
| Pavimento      | 0     | 1256  | 61923  | 983    | 0      | 0     | 0       | 0      | 0       | 0     | 96.51%        | 99.05%        | 97.77%        |
| Fusto          | 0     | 77    | 110    | 169830 | 233    | 102   | 0       | 0      | 277     | 0     | 99.53%        | 97.46%        | 98.48%        |
| Echino         | 0     | 0     | 0      | 1326   | 17857  | 894   | 0       | 0      | 638     | 0     | 86.20%        | 78.97%        | 82.43%        |
| Abaco          | 0     | 0     | 0      | 92     | 3899   | 24390 | 2189    | 0      | 719     | 0     | 77.95%        | 78.26%        | 78.11%        |
| Architrave     | 0     | 0     | 0      | 9      | 200    | 5402  | 49168   | 752    | 113     | 0     | 88.36%        | 87.52%        | 87.94%        |
| Fregio         | 0     | 0     | 0      | 0      | 0      | 14    | 4724    | 31047  | 196     | 31    | 86.21%        | 73.33%        | 79.25%        |
| Cornice        | 0     | 0     | 197    | 1378   | 423    | 363   | 98      | 7222   | 102117  | 1769  | 89.92%        | 94.85%        | 92.32%        |
| Timpano        | 0     | 0     | 1      | 0      | 0      | 0     | 0       | 3316   | 3597    | 19816 | 74.13%        | 91.67%        | 81.98%        |
| <b>AVERAGE</b> |       |       |        |        |        |       |         |        |         |       | <b>89.41%</b> | <b>89.47%</b> | <b>89.26%</b> |

Tabella 3. Matrice di confusione e metriche di precisione ottenute dalla classificazione con CNN 2D.

# 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 ( $X, Y, Z, \omega, \varphi$  e  $\kappa$ ) 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 ( $X, Y, Z, \omega, \varphi$  and  $\kappa$ ) 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).

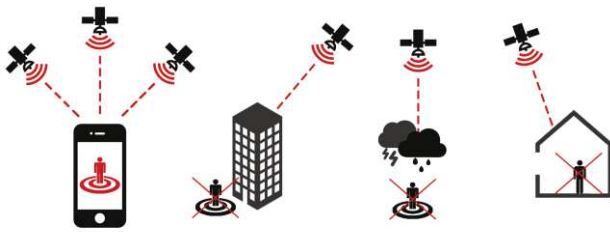


Figura 1. Limiti nel posizionamento con l'utilizzo del GNSS.

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    |                | <br>5 - 20m    | <br>worldwide |                |              |
| WiFi        |                | <br>5 - 15 m   | <br>< 150 m   |                |              |
| Bluetooth   |                | <br>1 - 3m     | <br>< 30 m    |                |              |
| VLC         |                | <br>< 50 cm    | <br>< 8 m     |                |              |
| BIM         |                | <br>< 50 cm    |               |                |              |
| Image Based |                | <br>20 - 50 cm | <br>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).

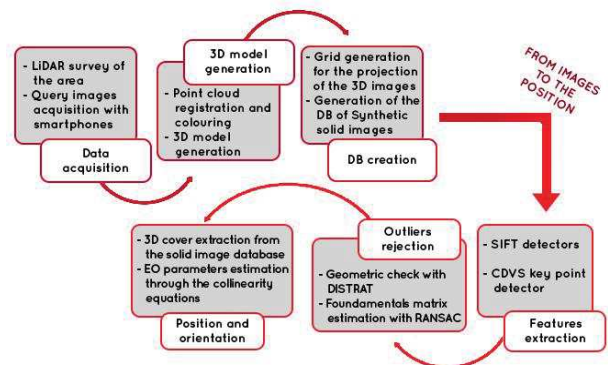


Figura 3. *Workflow* procedura IRBL.

### 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.

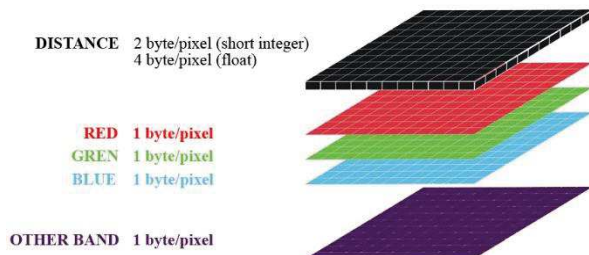


Figura 4. Struttura dell'immagine RGB-D.

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 ( $X_0, Y_0, Z_0, \omega, \varphi, \kappa$ );
- I parametri di orientamento interno corrispondenti alla focale e alla posizione del punto principale della camera rispetto all'immagine sintetica ( $c, \zeta_0, \eta_0$ );
- Il numero di pixel per ogni riga e colonna dell'immagine RGB-D ( $n_{col}, n_{row}$ ), e la dimensione del pixel nell'immagine ( $d_{pix}$ ).

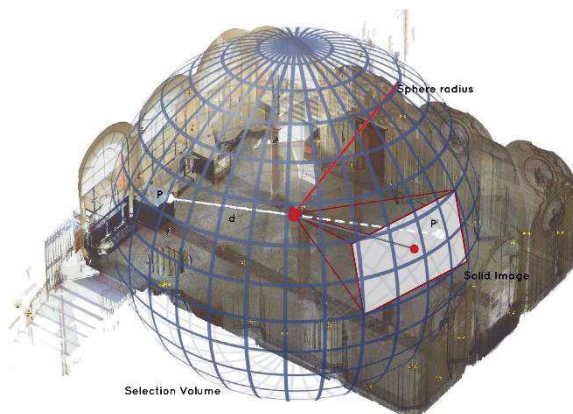


Figura 5. Sfera di selezione per la creazione delle immagini sintetiche.

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 ( $\zeta_i, \eta_i$ ) 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 ( $\zeta_i, \eta_i$ ) 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.

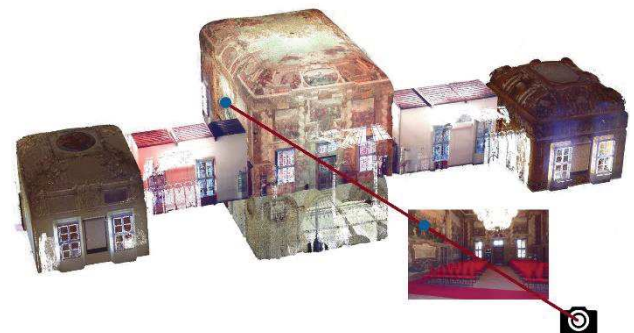


Figura 6. Procedura di generazione delle immagini RGB-D.

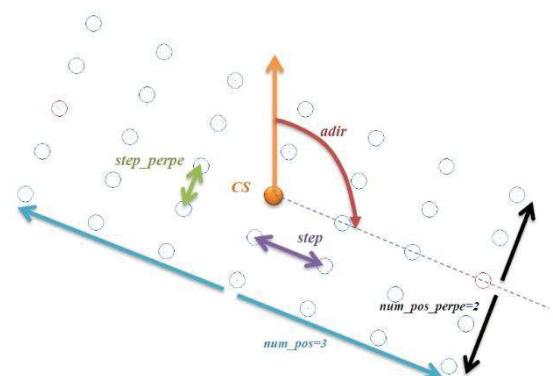


Figura 7. Definizione delle posizioni.

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.

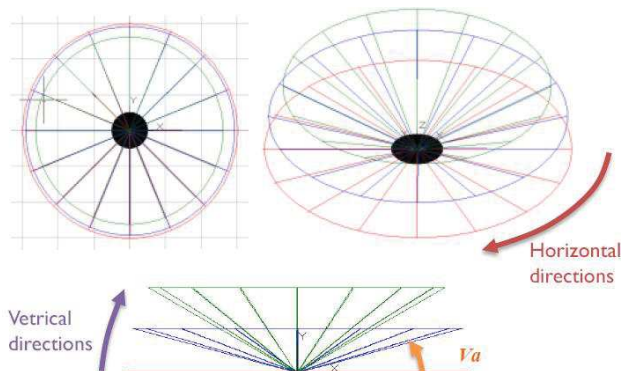


Figura 8. Definizione della direzione degli assi per ogni posizione.

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.

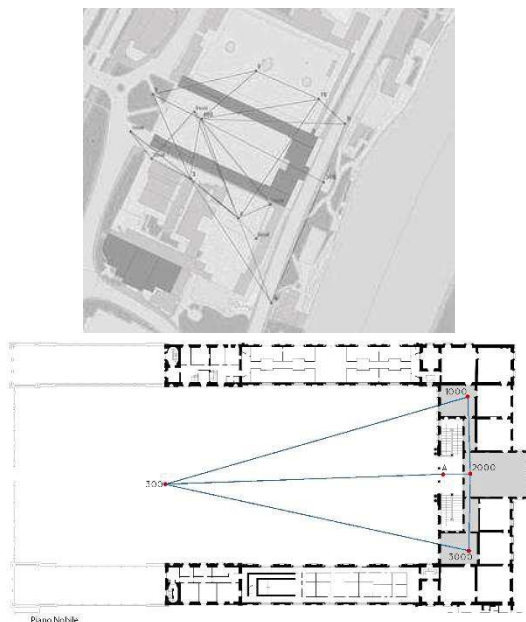


Figura 11. Indicazione della rete topografica.

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.

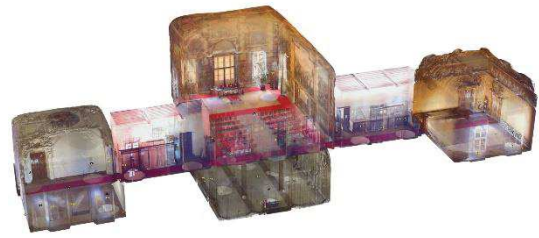


Figura 12. Nuvola di punti completa.

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  $\zeta_0$  e  $\eta_0$ , 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: [3x3 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.

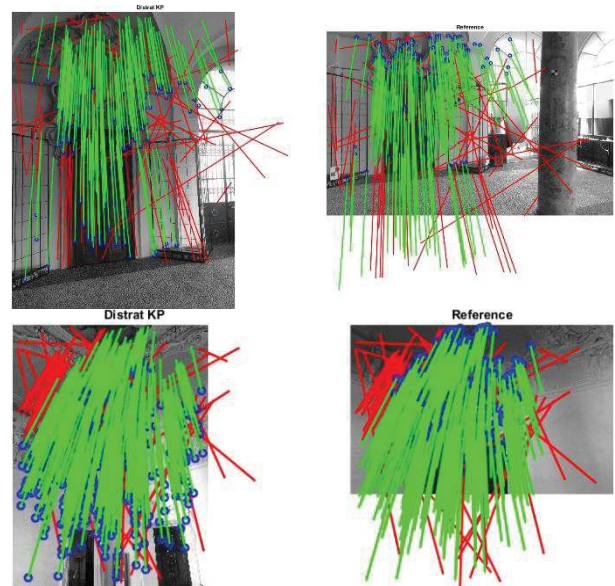


Figura 15. Indicazione delle *features* e degli *outliers* estratti.

#### 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 Coputer 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.