



Università degli Studi di Ferrara

DOTTORATO DI RICERCA IN
"SCIENZE DELL'INGEGNERIA"

CICLO XXVIII

COORDINATORE Prof. Stefano Trillo

LA FOTOGRAMMETRIA DIGITALE E LA COMPUTER VISION PER
L'INGEGNERIA CIVILE: TEORIA E CASI STUDIO

Settore Scientifico Disciplinare ICAR/06

Dottorando

Dott. Bolognesi Marcello

Tutore

Prof. Russo Paolo

CoTutore

Prof. Pellegrinelli Alberto

Anni 2013/2015

INDICE

PREMESSA.....	6
1 PRINCIPI di FORMAZIONE dell'IMMAGINE.....	9
1.1 LA PROIEZIONE PROSPETTICA.....	9
1.2 LA PROIEZIONE ORTOGRAFICA.....	10
1.3 LE IMMAGINI DIGITALI.....	10
2 DEFINIZIONE della GEOMETRIA di PRESA.....	12
2.1 MODELLO STENOPEICO DELLA FOTOCAMERA.....	12
2.1.1 Modello semplificato.....	13
2.1.2 Modello generale.....	14
2.1.3 Riepilogo.....	18
2.2 PROPRIETÀ DELLA MATRICE DI PROIEZIONE PROSPETTICA.....	18
2.2.1 Forma cartesiana.....	18
2.2.2 Centro di presa.....	19
2.2.3 Raggio ottico.....	20
2.3 CALIBRAZIONE DELLA FOTOCAMERA.....	20
2.3.1 Distorsione radiale.....	20
2.3.2 Self-Calibration.....	23
2.3.3 Direct Linear Transform (DLT).....	23
3 STRUCTURE FROM MOTION (S.F.M.).....	26
4 L'ESTRAZIONE DELLE CARATTERISTICHE.....	28
4.1 IL CONCETTO DI CARATTERISTICA.....	28
4.2 FEATURE DETECTION AND DESCRIPTION.....	32
4.2.1 Harris corner detector (1988).....	32
4.2.2 Shi e Tomasi (1994).....	43
4.2.3 Scale invariant feature transform: S.I.F.T. (2004).....	44
5 RICOSTRUZIONE METRICA.....	52

5.1	GEOMETRIA EPIPOLARE	52
5.2	RICOSTRUZIONE CALIBRATA	54
5.2.1	Orientamento relativo.....	54
5.2.2	Orientamento Esterno.....	57
5.3	RICOSTRUZIONE NON CALIBRATA.....	58
5.3.1	La matrice Fondamentale.....	58
5.3.2	Stima della matrice Fondamentale (F)	59
5.4	AUTOCALIBRAZIONE.....	61
5.4.1	Ricostruzione proiettiva	61
5.4.2	Promozione euclidea.....	62
5.5	TRIANGOLAZIONE (SPARSA).....	63
5.5.1	Caso semplificato: fotocamere parallele e allineate	64
5.5.2	Caso generalizzato.....	65
5.6	DENSE MATCHING.....	67
5.6.1	Metodi locali.....	68
5.6.2	Metodi globali	73
6	RICOSTRUZIONE in PRATICA.....	75
6.1	APPROCCIO SEQUENZIALE	75
6.1.1	Filtraggio outliers (RANSAC)	75
6.1.2	Ricostruzione incrementale	79
6.1.3	Bundle adjustment.....	80
6.1.4	Riepilogo.....	81
7	CASI STUDIO.....	84
7.1	INTRODUZIONE	84
7.2	PREDISPOSIZIONE DEL TEST-SITE	85
7.2.1	Il modello di riferimento	86
8	LA PRECISIONE NEL RILIEVO INTEGRATO DEI BENI CULTURALI.....	92
8.1	STRUMENTAZIONE UTILIZZATA.....	92
8.2	PIANIFICAZIONE DEL RILIEVO ED ACQUISIZIONE DEL DATO	96

8.3	ELABORAZIONE DEL DATO	98
8.3.1	Il modello Integrato.....	103
8.4	VALUTAZIONE DELLA PRECISIONE.....	105
8.4.1	Confronto automatico tra nube densa di punti fotogrammetrica e scansione laser.	106
8.4.2	Confronto automatico tra modelli tridimensionali e scansione laser.....	111
8.4.3	Confronto automatico tra modelli tridimensionali e punti fotografici di appoggio.	117
8.4.4	Confronto manuale tra nube di punti fotogrammetrica e punti di controllo.	122
8.5	CONSIDERAZIONI FINALI	129
9	IL POTENZIALE DI TECNOLOGIE LOW-COST APPLICATE AL RILIEVO DEI BENI CULTURALI	135
9.1	INTRODUZIONE	135
9.2	STRUMENTAZIONE UTILIZZATA.....	135
9.3	PIANIFICAZIONE DEL RILIEVO ed ACQUISIZIONE DEL DATO	136
9.4	ELABORAZIONE DEL DATO (IMMAGINI FOTOGRAFICHE)	138
9.4.1	Modelli realizzati con self-calibration (Photomodeler).....	139
9.4.2	Modelli realizzati con immagini ricampionate senza distorsione	141
9.4.3	Modelli realizzati con la procedura di autocalibrazione “fisheye” di Photoscan.....	144
9.5	VALUTAZIONE DELLA PRECISIONE (IMMAGINI FOTOGRAFICHE).....	145
9.6	ELABORAZIONE DEL DATO (VIDEO).....	153
9.7	VALUTAZIONE DELLA PRECISIONE (VIDEO)	156
9.7.1	Modelli realizzati con autocalibrazione “fisheye” di Photoscan.....	156
9.7.2	Modelli realizzati a partire da frames corretti con GoPro Studio	161
9.8	CONSIDERAZIONI FINALI	166
10	CONCLUSIONI FINALI	168
10.1	PIANIFICAZIONE DEL RILIEVO	168
10.1.1	Strumentazione	168
10.1.2	Pianificazione della strategia di presa	169
10.1.3	Calibrazione della camera	170
10.1.4	Rilievo dei Punti Fotografici di Appoggio	171
10.2	AQUISIZIONE DEL DATO	171

10.3	ELABORAZIONE DEL DATO	171
11	FOTOGRAMMETRIA APPLICATA ALLO STUDIO DI MECCANISMI DI COLLASSO DI TRAVI IN C.A.	177
11.1	INTRODUZIONE	177
11.2	ALLESTIMENTO DELLA PROVA.....	177
11.3	IL RILIEVO FOTOGRAMMETRICO	178
11.4	PRIMI CONFRONTI.....	181
11.5	STIMA DELLE DEFORMAZIONI	182
11.6	CONSIDERAZIONI FINALI	186
	BIBLIOGRAFIA.....	187
	Siti Consultati.....	188

PREMESSA

Vorrei iniziare la mia tesi di dottorato con una semplice definizione attorno alla quale ha ruotato tutto il percorso di studio e ricerca che ho svolto negli ultimi tre anni. Mi riferisco al termine di ***structure from motion***. Si definisce *structure from motion* una serie di processi ed algoritmi finalizzati alla ricostruzione dell'ambiente che ci circonda partendo da una sequenza di immagini. Nel dettaglio con il termine motion, si intendono i parametri di orientamento esterno della fotocamera, di fatto la loro posizione al momento dell'acquisizione delle foto, ovvero il moto della camera durante il rilievo. Si tratta quindi di una ricostruzione di tipo metrico, che passa attraverso il calcolo delle coordinate tridimensionali di punti in un sistema di riferimento noto (structure), partendo dalla posizione delle camere al momento dell'acquisizione delle immagini (motion). La stima di tale posizione è calcolata esclusivamente utilizzando le immagini acquisite, di cui non si conosce nulla a parte l'intensità luminosa dei pixel che le compongono.

In questa tesi cercherò di ricostruire passo passo l'intero processo di structure from motion cercando di focalizzare l'attenzione sulle innovazioni che tale metodologia di ricostruzione ha apportato agli schemi classici di restituzione fotogrammetrica. La prima parte di questa tesi sarà dunque una parte puramente teorica. La fotogrammetria si occupa da sempre di ricostruzione metrica, molti dei concetti che verranno trattati saranno infatti concetti già noti sui quali non verrà apportato alcun approfondimento. La vera novità di questo insieme di processi non risiede infatti nei problemi geometrici di triangolazione, calibrazione ed orientamento (anche se gli algoritmi sempre più performanti contribuiscono continuamente ad una loro evoluzione), ma nell'approccio che ci porta alla loro risoluzione.

Grazie alle tecniche sviluppate dalla ***computer vision***, il processo di ricostruzione è quasi completamente affidato ai calcolatori, l'unico dato necessario per la risoluzione del problema geometrico è di fatto l'immagine stessa. Sarà questo il vero fulcro attorno al quale ruoterà la prima parte della tesi, con l'obiettivo di fare chiarezza sull'intero processo di ricostruzione partendo semplicemente dall'acquisizione del dato. In figura 1 è possibile osservare un tipico risultato di questi processi di ricostruzione.

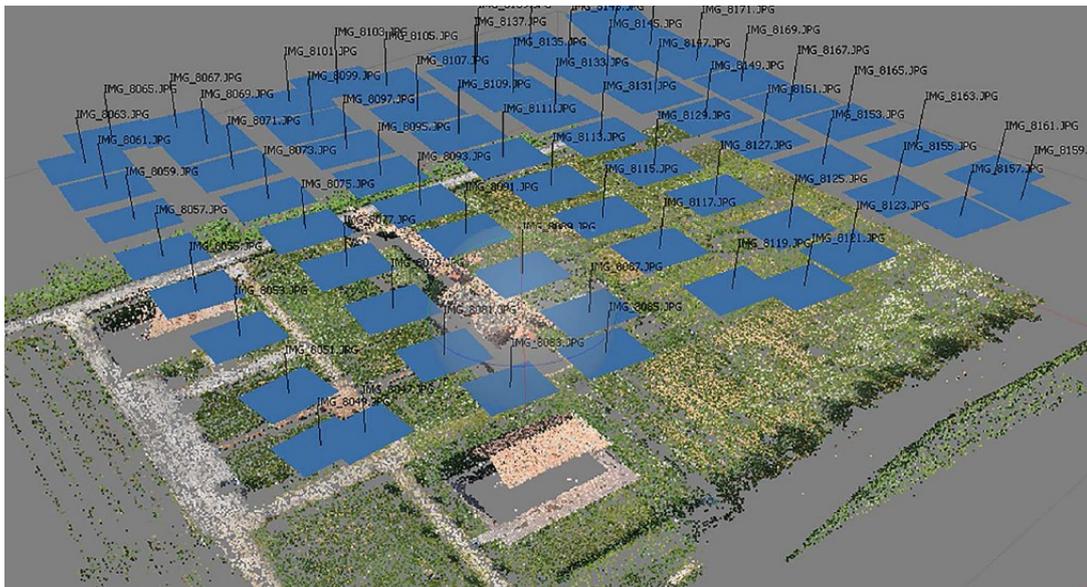


Figura 1: vista della ricostruzione del Castello del Verginese (Ferrara). In particolare si osservano le posizioni delle fotocamere al momento dell'acquisizione e la nube di punti generata dal processo di Structure from Motion. Per la sua realizzazione è stato utilizzato il software Photoscan di Agisoft.

La seconda parte avrà invece una connotazione decisamente pratica, in quanto verranno presentati ed analizzati alcuni significativi casi di studio che utilizzano proprio i processi sopraelencati, con lo scopo di delineare alcune procedure operative propedeutiche all'ottenimento di determinati standard di precisione. Saranno infine presentate alcune applicazioni di carattere innovativo che sfruttano la fotogrammetria digitale senza necessariamente utilizzare i processi di structure from motion.

Ovviamente con la presente tesi non ho la presunzione di fornire la panoramica completa dei metodi e degli algoritmi esistenti, che entrano all'interno del processo di ricostruzione. La loro varietà e la continua evoluzione che essi subiscono rende difficile una loro completa catalogazione. Da un punto di vista puramente fotogrammetrico a volte verrà presentata soltanto una descrizione di carattere generale dell'algoritmo o del procedimento trattato mentre laddove ritenuto importante verrà esplicitamente definito il particolare algoritmo e le sue principali funzioni.

PARTE PRIMA

-FONDAMENTI TEORICI-

1 PRINCIPI di FORMAZIONE dell'IMMAGINE

1.1 LA PROIEZIONE PROSPETTICA

Un qualsiasi apparato di acquisizione delle immagini (nel nostro caso una fotocamera digitale) funziona raccogliendo la luce riflessa dall'ambiente che la circonda ed imprimendola sul piano del sensore (o piano immagine) formando la classica immagine bidimensionale (matrice, nel caso di immagini digitali). Prima di procedere con la trattazione analitica vera e propria e l'individuazione delle caratteristiche geometriche che compongono il problema della restituzione è necessario dedicare qualche paragrafo alla formazione delle immagini in quanto tutto il processo che si andrà ad analizzare, si basa sull'inversione delle equazioni che regolano tale problema.

Il modello geometrico di formazione dell'immagine più semplice è il cosiddetto *modello stenopeico* o meglio noto come modello *pinhole*. Si basa sull'antico principio della *camera obscura*, di fatto non servono lenti ed è tutto affidato alla luce. La camera oscura non è altro che un locale buio in cui è stato praticato un piccolo pertugio, attraverso il quale l'immagine della realtà si rivela, andando a proiettarsi capovolta sulla parete opposta del locale oscurato. La camera stenopeica (o pinhole camera) non è altro che la riproduzione in scala di tale principio. Si immagini di prendere una scatola e di praticarvi un piccolo foro con uno spillo su uno dei due lati (**Errore. L'origine riferimento non è stata trovata.**).

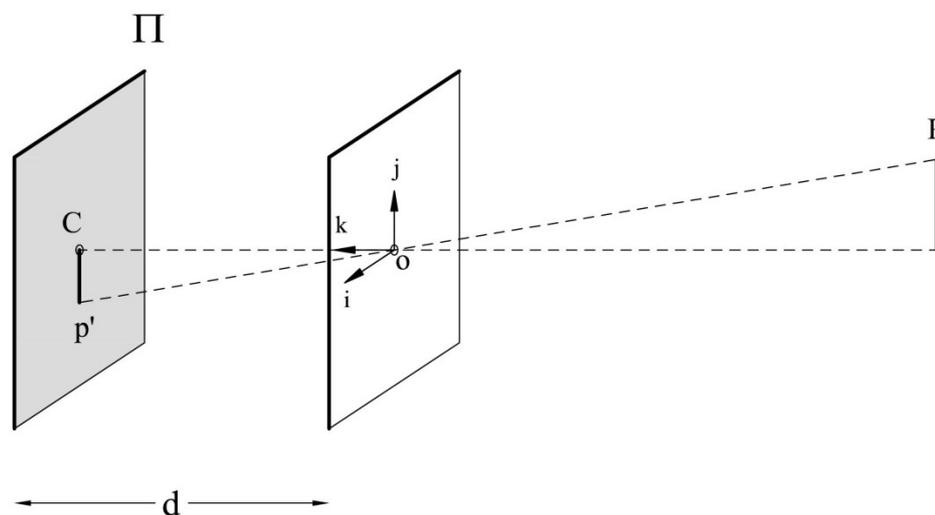


Figura 1.1: Pinhole camera, geometria di formazione dell'immagine

Consideriamo il sistema di riferimento (O, i, j, k) con centro coincidente al foro e con vettori i e j appartenenti ad un piano parallelo al piano immagine Π che si trova ad una distanza d dal foro stenopeico. La linea individuata dalla direzione k è perpendicolare al piano immagine e passante

attraverso il foro prende il nome di asse ottico. Il punto **C** intersezione dell'asse ottico con il piano immagine viene chiamato centro dell'immagine. Sia **P** un punto nello spazio dell'oggetto, di coordinate (X, Y, Z) , e sia **p'** la sua proiezione sul piano dell'immagine, di coordinate (x', y', z') . Attraverso una semplice similitudine tra triangoli possiamo ottenere:

$$\frac{x'}{d} = \frac{X}{-Z} \quad ; \quad \frac{y'}{d} = \frac{Y}{-Z} \quad (1.1)$$

da cui

$$x' = -d \frac{X}{Z}; \quad y' = -d \frac{Y}{Z}; \quad z' = d \quad (1.2)$$

si noti che l'immagine è rovesciata rispetto alla scena originale in ogni direzione. I segni meno all'interno delle equazioni indicano proprio tale peculiarità. Queste equazioni definiscono il processo di formazione dell'immagine che prende il nome di **proiezione prospettica**. Il termine Z al denominatore è responsabile dell'effetto di prospettiva, tale per cui la dimensione dell'immagine dell'oggetto varia in funzione alla sua distanza dall'osservatore.

1.2 LA PROIEZIONE ORTOGRAFICA

Se l'oggetto inquadrato ha uno sviluppo in Z trascurabile rispetto alla distanza tra l'oggetto e l'osservatore allora la *proiezione prospettica* appena definita può essere approssimata ad una **proiezione ortografica scalata**. Infatti, se la profondità dei punti oggetto variano in un intervallo $Z_0 + \Delta Z$ tale per cui $\Delta Z/Z_0 \ll Z_0$, allora il fattore di scala d/Z può essere approssimato da una costante d/Z_0 . Le equazioni di proiezione diventano:

$$x' = -d \frac{X}{Z_0}; \quad z' = -d \frac{Y}{Z_0} \quad (1.3)$$

Sono le equazioni di una proiezione ortografica con fattore di scala d/Z_0 .

1.3 LE IMMAGINI DIGITALI

In una fotocamera digitale il piano di formazione dell'immagine è costituito da un sensore CCD (Charghed-Coupled Device) o CMOS (Complementary Metal-Oxide Semiconductor). Possiamo rappresentare quest'ultimo come una matrice $n \times m$ di celle rettangolari fotosensibili, ciascuna delle quali converte l'intensità della radiazione luminosa incidente in un potenziale elettrico.

Gli elementi della matrice prendono il nome di **pixel** (picture element). Indicheremo con $I(x_i, y_j)$ il valore di luminosità associato al pixel appartenete alla riga *i-esima* ed alla colonna *j-esima* (tipicamente il sistema di riferimento x, y ha origine nell'angolo in alto a sinistra della matrice).

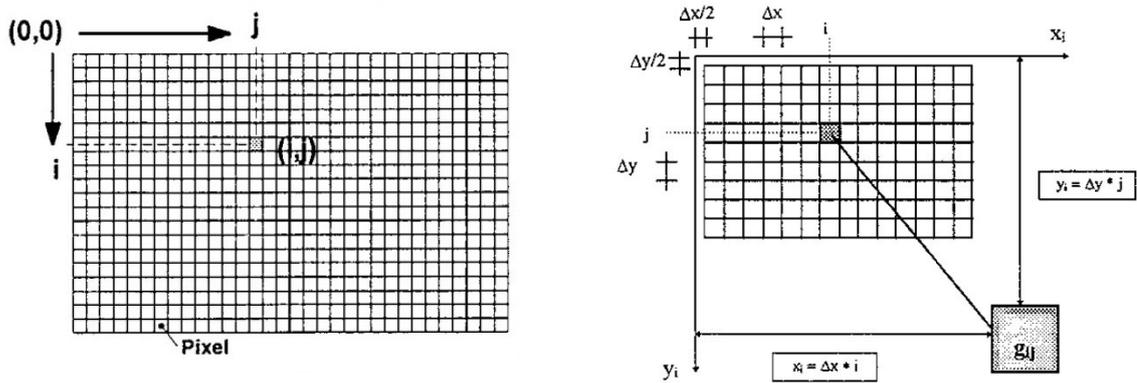


Figura 1.2: schematizzazione del sistema di riferimento di un'immagine digitale. A sinistra si osserva la griglia di elementi pixel che la compongono, con il sistema di riferimento fissato nell'angolo in alto a sinistra. Nell'immagine di destra un ingrandimento di una porzione dell'immagine.

Si può determinare così una corrispondenza biunivoca tra la posizione del pixel (i,j) ed il sistema attraverso le relazioni:

$$x_i = \Delta x \cdot i \tag{1.4}$$

$$y_i = \Delta y \cdot j$$

In fotogrammetria digitale la misura tradizionale delle coordinate immagine è sostituita dall'identificazione del pixel all'interno della matrice (pixelizzazione), operazione che per quanto possibile può essere eseguita automaticamente. Più sono piccole le dimensioni dell'elemento più esso è identificabile con il vecchio punto immagine. Ad ogni pixel viene quindi associata una coppia di coordinate ed un altro valore numerico, che ne rappresenta la radiometria. Nelle immagini a colori, viene memorizzato solitamente il livello di intensità dei colori fondamentali (nel modello di colore RGB, uno dei più usati, sono tre: rosso, verde e blu, mentre nelle immagini monocromatiche in scala di grigio (dette impropriamente bianco e nero) il valore indica l'intensità del grigio, che varia dal nero al bianco. Il numero di colori o di livelli di grigio possibili dipende dal massimo numero di combinazioni permesse dalla quantità di bit utilizzata per ognuno di questi dati: un'immagine con 1 bit per pixel avrà al massimo due combinazioni possibili (0 e 1) e quindi potrà rappresentare solo due colori o solo bianco e nero; nelle immagini a 4 bit per pixel, si possono rappresentare al massimo 16 colori o 16 livelli di grigio; un'immagine a 8 bit per pixel, 256 e così via.

2 DEFINIZIONE della GEOMETRIA di PRESA

Come già accennato le immagini digitali appartengono ad uno spazio discreto in quanto sono suddivise in entità tipicamente rettangolari chiamate *pixel*. Le equazioni che abbiamo visto in precedenza sono valide se le distanze sono misurate rispetto al sistema di riferimento della camera e quando questo è centrato sull'asse principale di collimazione secondo due assi di simmetria. Nella pratica invece il sistema di riferimento della camera si discosta dal modello proposto ed è correlato al sistema di riferimento assoluto attraverso una serie di parametri chiamati parametri di orientamento. In particolare si distinguono i parametri di orientamento interno, sono quelli strettamente legati al sistema di riferimento della camera, ed i parametri di orientamento esterno che invece relazionano la posizione e l'orientamento della fotocamera rispetto un sistema di riferimento assoluto.

In questo capitolo sarà introdotto un modello geometrico per la fotocamera ed il sensore e verrà affrontato il problema della determinazione dei parametri di tale modello, attraverso il processo noto con il nome di calibrazione.

2.1 MODELLO STENOPEICO DELLA FOTOCAMERA

Sulla base dei concetti visti nel capitolo precedente, il modello geometrico di fotocamera più comune è il cosiddetto **modello stenopeico** il quale consiste di un *piano immagine* (P.I.) e di un punto *C* (*centro di presa*) distante dal piano immagine *f* (*distanza focale*). La retta passante per *C* ortogonale al piano immagine è l'*asse ottico*. La sua intersezione con il piano immagine prende il nome di *punto principale* (P.P.). Il piano passante per *C* prende invece il nome di *piano focale* (P.F.).

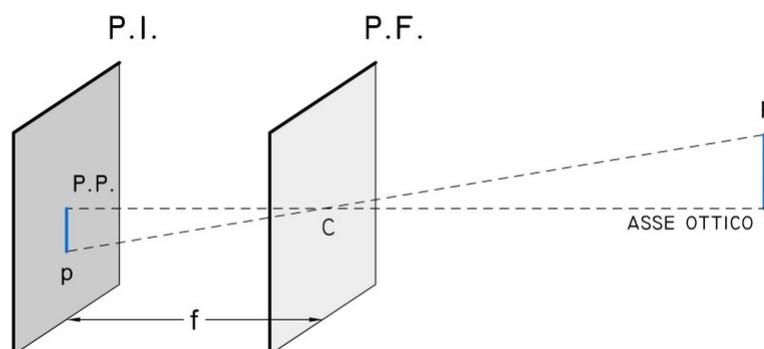


Figura 2.1. Modello stenopeico della fotocamera.

Per descrivere analiticamente questo modello geometrico, è opportuno introdurre adeguati sistemi di riferimento con cui esprimere le coordinate dello spazio oggetto e le coordinate dei punti proiettati sul piano immagine. Inizialmente si considera un modello semplificato, in cui i sistemi di

riferimento scelti sono tali da semplificare le equazioni che descrivono il problema. Solo successivamente si osserverà il caso più generale.

2.1.1 Modello semplificato

Introduciamo un sistema di riferimento (X, Y, Z) (in verde) per lo spazio oggetto centrato in C e con l'asse Z coincidente con l'asse ottico. Stiamo dunque fissando il sistema di riferimento assoluto coincidente con il sistema di riferimento della fotocamera. Introduciamo ora un altro sistema di riferimento (u, v) (in rosso) per il piano immagine centrato nel punto principale con gli assi u e v paralleli ad X, Y .

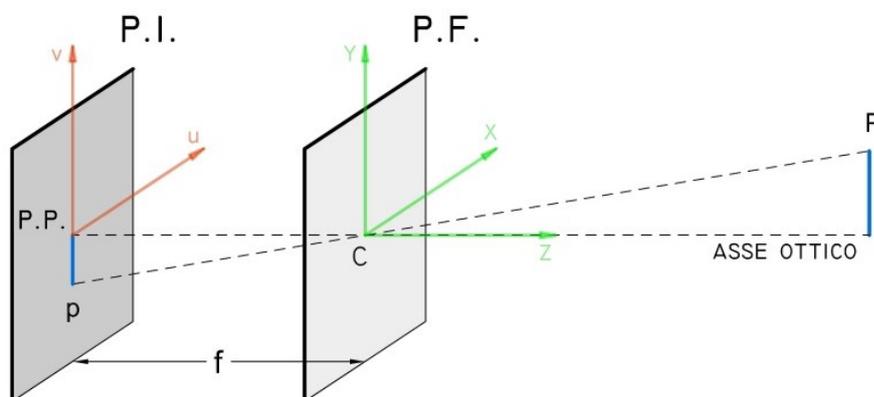


Figura 2.2. Modello stenopeico della fotocamera. Caso semplificato: il sistema di riferimento assoluto è centrato nel centro di presa della fotocamera stessa.

Consideriamo ora il punto P di coordinate (X, Y, Z) e la sua proiezione sul piano immagine p di coordinate (U, V) . Attraverso una semplice similitudine tra triangoli valgono le seguenti relazioni:

$$\frac{-U}{f} = \frac{X}{Z} \quad ; \quad \frac{-V}{f} = \frac{Y}{Z} \quad (2.1)$$

Da cui

$$U = \frac{-fX}{Z} \quad ; \quad V = \frac{-fY}{Z} \quad (2.2)$$

Sono le **equazioni di una proiezione prospettica**. La trasformazione da coordinate 3D a coordinate 2D non è lineare a causa della divisione per Z . Per eliminare questa non linearità e semplificare le trasformazioni utilizzeremo le **coordinate omogenee** (Appendice A). Come si può osservare in appendice le coordinate omogenee sono particolarmente utili soprattutto nella rappresentazione di alcune trasformazioni geometriche nello spazio per la possibilità di concatenare una sequenza di trasformazioni ed esprimerle in forma matriciale con un'unica matrice.

Vediamo ora cosa accade alle equazioni della proiezione prospettica se utilizziamo le coordinate omogenee al posto delle cartesiane. Siano dunque P di coordinate omogenee $(x, y, z, 1)$ e p di coordinate omogenee $(u, v, 1)$. Ponendo la terza coordinata uguale a 1 abbiamo escluso i punti all'infinito. L'equazione di proiezione prospettica si può riscrivere in notazione matriciale come:

$$p = \frac{1}{z} MP \quad (2.3)$$

Dove:

$$p = \begin{bmatrix} u \\ v \\ 1 \end{bmatrix}; \quad P = \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix}$$

$$\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} -fx/z \\ -fy/z \\ 1 \end{bmatrix} = \frac{1}{z} \begin{bmatrix} -fx \\ -fy \\ z \end{bmatrix} = \frac{1}{z} \begin{bmatrix} -f & 0 & 0 & 0 \\ 0 & -f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix}$$

Il parametro $1/z$ rappresenta di fatto un fattore di scala. La matrice M rappresenta un modello geometrico della fotocamera e viene chiamata **matrice di proiezione prospettica** (MPP). L'equazione può essere anche scritta come:

$$p \cong MP \quad (2.4)$$

Il simbolo (\cong) significa a meno di un fattore di scala.

2.1.2 Modello generale

Nella caso più generale il sistema di riferimento assoluto non è mai coincidente al sistema di riferimento della fotocamera. Anche il sistema di riferimento immagine non è baricentrico, spesso lo troviamo in un angolo del sensore, tipicamente quello in alto a sinistra. Oltre a queste differenze dobbiamo anche prendere in esame la pixelizzazione del sensore.

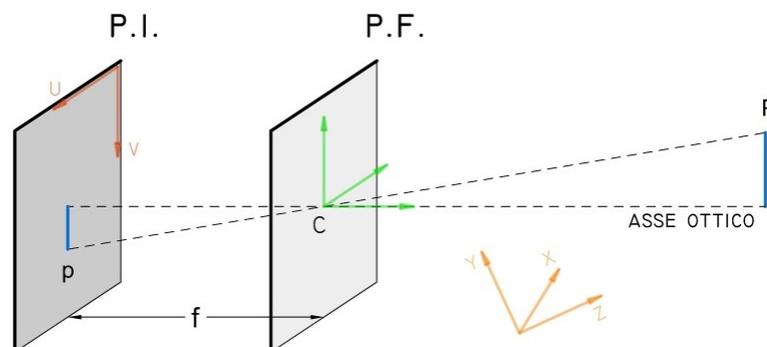


Figura 2.3. Modello stenopeico della fotocamera. Caso generalizzato: il sistema di riferimento assoluto è diverso da quello della fotocamera

Oltre che alla trasformazione prospettica descritta nel paragrafo precedente bisogna quindi tener conto di una trasformazione rigida tra il riferimento della fotocamera (in verde) e il sistema di riferimento assoluto (in giallo). Inoltre serve un cambio di coordinate da metri a pixel immagine (pixelizzazione).

2.1.2.1 Parametri di orientamento interno

È possibile associare ad una camera un piano immagine normalizzato (*P.I.N.*) parallelo al suo piano immagine fisico (*P.I.*) spostato di una unità rispetto al centro di proiezione *C*. Sul piano è materializzato un sistema di riferimento (*U', V'*) avente origine in *C'*.

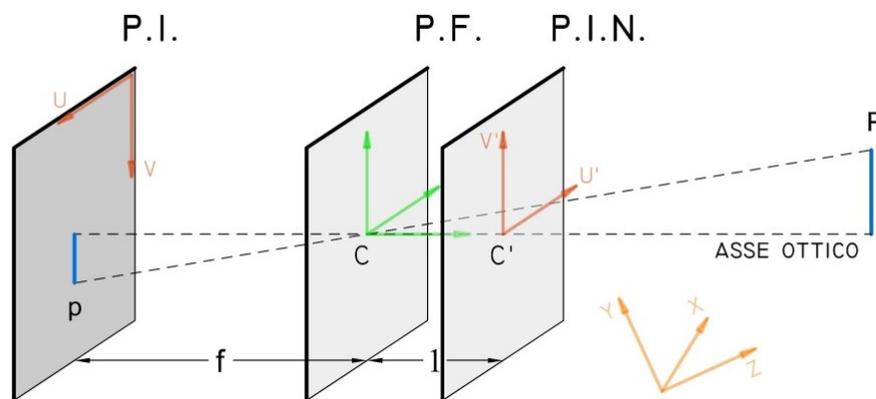


Figura 2.4. Modello stenopeico della fotocamera. In particolare si osserva il piano immagine normalizzato parallelo al suo piano immagine fisico (*P.I.*) spostato di una unità rispetto al centro di proiezione *C*. Sul piano è materializzato un sistema di riferimento (*U', V'*) avente origine in *C'*.

Se si sostituisce il piano immagine con quello normalizzato è possibile riscrivere le equazioni della prospettiva centrale nella forma:

$$U' = 1 \frac{X}{Z}; \quad V' = 1 \frac{Y}{Z}; \quad (2.5)$$

Ed in coordinate omogenee:

$$\begin{bmatrix} u' \\ v' \\ 1 \end{bmatrix} = \begin{bmatrix} x/z \\ y/z \\ 1 \end{bmatrix} = \frac{1}{z} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \frac{1}{z} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix}$$

$$p' = \frac{1}{Z} [I|0]P = \frac{1}{Z} MP, \quad \text{dove } M = [I|0] \quad (2.6)$$

Tipicamente il piano immagine si trova ad una distanza diversa dal centro di presa che viene chiamata distanza focale $f \neq 1$, e le coordinate (u, v) sono espresse in pixel. Inoltre i pixel possono

essere di forma diversa da una sensore all'altro, a volte sono infatti rettangolari altre quadrati, per questo nell'equazione della prospettiva entrano solitamente altri due parametri che chiamiamo k ed l :

$$u = kf \frac{x}{z} = kfu'; \quad v = lf \frac{y}{z} = lfv';$$

f è una distanza espressa in metri, mentre k ed l esprimono le dimensioni del pixel secondo $1/k$ e $1/l$ (*pixel/m*). Questi due parametri non sono indipendenti e possono essere sostituiti con un unico fattore di scala

$$\alpha = kf \text{ e } \beta = lf \text{ (pixel)}.$$

Ancora tipicamente l'origine del sistema di riferimento si trova in un angolo del piano immagine e il centro del sensore spesso differisce dal centro di proiezione. Si aggiungono così altri due parametri u_0 e v_0 che definiscono la posizione del centro del sensore rispetto al sistema di coordinate pixel. Pertanto :

$$\mathbf{u} = \alpha \mathbf{u}' + \mathbf{u}_0; \quad \mathbf{v} = \beta \mathbf{v}' + \mathbf{v}_0 \quad (2.7)$$

Infine il sistema di riferimento della camera potrebbe essere anche inclinato, ad esempio per errori di fabbricazione, e non formare un angolo di 90 gradi tra gli assi. È dunque necessari a un'ultima trasformazione da cui:

$$\begin{aligned} \mathbf{u} &= \alpha \mathbf{u}' - \alpha \cot \theta \mathbf{u}' + \mathbf{u}_0; \quad \mathbf{v} \\ &= \frac{\beta}{\sin \theta} \mathbf{v}' + \mathbf{v}_0 \end{aligned} \quad (2.8)$$

In forma matriciale:

$$\mathbf{p} = \mathbf{K} \mathbf{p}' \quad \text{dove} \quad \mathbf{K} = \begin{pmatrix} \alpha & -\alpha \cot \theta & u_0 \\ 0 & \frac{\beta}{\sin \theta} & v_0 \\ 0 & 0 & 1 \end{pmatrix}$$

\mathbf{K} è nota come matrice di calibrazione interna

$$\mathbf{p} = \frac{1}{Z} \mathbf{K} [\mathbf{I} | \mathbf{0}] \mathbf{P} = \frac{1}{Z} \mathbf{M} \mathbf{P}, \quad \text{dove} \quad \mathbf{M} = \mathbf{K} [\mathbf{I} | \mathbf{0}] \quad (2.9)$$

I parametri $\alpha, \beta, \theta, u_0, v_0$ prendono il nome di **parametri di orientamento interno**. La matrice $[\mathbf{I} | \mathbf{0}]$ codifica l'essenza della trasformazione prospettica senza alcun parametro. La si ottiene quando il sistema di riferimento assoluto coincide con il sistema di riferimento della camera ed i punti immagine sono espressi in *coordinate normalizzate*, cioè:

$$\mathbf{p}^* = \mathbf{K}^{-1}\mathbf{p} \quad (2.10)$$

2.1.2.2 Parametri di orientamento esterno

L'equazione (2.10) è riferita ad un sistema di riferimento solidale alla camera. Vediamo come si trasforma se il sistema di riferimento assoluto è diverso. Consideriamo un sistema di riferimento (W) diverso dal sistema di riferimento camera (C). Il cambio di coordinate tra due sistemi di riferimento (W) e (C) può essere descritto attraverso una matrice di rotazione (R) ed un vettore traslazione (\mathbf{t}). Questa trasformazione può essere descritta in coordinate cartesiane come :

$$\mathbf{P}_{(C)} = \mathbf{R}\mathbf{P}_{(W)} + \mathbf{t}, \quad \text{con } \mathbf{P}_{(C)} \text{ e } \mathbf{P}_{(W)} \in \mathbb{R}^3 \quad (2.11)$$

In coordinate omogenee invece la stessa relazione diventa:

$$\mathbf{P}_{(C)} = \mathbf{G}\mathbf{P}_{(W)}, \quad \text{dove } \mathbf{G} = \begin{pmatrix} \mathbf{R} & \mathbf{t} \\ \mathbf{0}^T & 1 \end{pmatrix} \text{ e con } \mathbf{P}_{(C)} \text{ e } \mathbf{P}_{(W)} \in \mathbb{R}^4 \quad (2.12)$$

Per passare da C a W servirà una rotazione e una traslazione. La relazione l'abbiamo già vista pertanto se

$$\mathbf{p} = \frac{1}{Z}\mathbf{M}\mathbf{P}_{(C)}$$

E

$$\begin{aligned} \mathbf{P}_{(C)} &= \mathbf{G}\mathbf{P}_{(W)}, \quad \text{dove } \mathbf{G} = \begin{pmatrix} \mathbf{R} & \mathbf{t} \\ \mathbf{0}^T & 1 \end{pmatrix} \\ \mathbf{p} &= \frac{1}{Z}\mathbf{M}\mathbf{P}_{(W)} \quad \text{dove } \mathbf{M} = \mathbf{K}[\mathbf{I}|\mathbf{0}]\mathbf{G} \end{aligned} \quad (2.13)$$

Questa è nota come *l'Equazione generale della proiezione prospettica*. Nota \mathbf{M} possiamo determinare la posizione della camera nel sistema di riferimento assoluto W .

A seconda della convenienza si può considerare anche la seguente fattorizzazione:

$$\mathbf{M} = \mathbf{K}[\mathbf{R}|\mathbf{t}] \quad (2.14)$$

Una qualunque matrice di rotazione è definita attraverso 3 parametri (i tre angoli di rotazione), a cui si aggiungono i tre parametri del vettore di traslazione \mathbf{t} , per un totale di 6 parametri che prendono il nome di *parametri di orientamento esterno*. Infatti

$$\mathbf{M} = \mathbf{K}[\mathbf{R}|\mathbf{t}] = \begin{pmatrix} \alpha & -\alpha \cot \theta & x_0 \\ 0 & \frac{\beta}{\sin \theta} & y_0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} r_1^T & t_1 \\ r_2^T & t_2 \\ r_3^T & t_3 \end{pmatrix} = \begin{pmatrix} \alpha r_1^T - \alpha \cot \theta r_2^T + x_0 r_3^T & \alpha t_1 - \alpha \cot \theta t_2 + x_0 t_3 \\ \frac{\beta}{\sin \theta} r_2^T + y_0 r_3^T & \frac{\beta}{\sin \theta} t_2 + y_0 t_3 \\ r_3^T & t_3 \end{pmatrix}$$

$$\mathbf{M} = \mathbf{K}[\mathbf{R}|\mathbf{t}] = \begin{pmatrix} \alpha & -\alpha \cot \theta & x_0 \\ 0 & \frac{\beta}{\sin \theta} & y_0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} r_1^T & t_1 \\ r_2^T & t_2 \\ r_3^T & t_3 \end{pmatrix} = \begin{pmatrix} \alpha r_1^T + x_0 r_3^T & \alpha t_1 + x_0 t_3 \\ \beta r_2^T + y_0 r_3^T & \beta t_2 + y_0 t_3 \\ r_3^T & t_3 \end{pmatrix}$$

Una versione semplificata di tale matrice la si può ottenere fissando $\theta = \pi/2$.

$$\mathbf{M} = \begin{pmatrix} \alpha r_1^T + x_0 r_3^T & \alpha t_1 + x_0 t_3 \\ \beta r_2^T + y_0 r_3^T & \beta t_2 + y_0 t_3 \\ r_3^T & t_3 \end{pmatrix} \quad (2.15)$$

2.1.3 Riepilogo

Abbiamo dunque definito una relazioni importante della geometria proiettiva e cioè che dato un qualsiasi punto P appartenente alla superficie fisica di un oggetto, si definisce p la sua proiezione sul piano immagine attraverso C (centro di proiezione). Il punto p proiettato è calcolabile in via analitica moltiplicando le coordinate di P per la matrice di proiezione prospettica. La matrice \mathbf{M} applica di fatto la trasformazione prospettica alle coordinate di P . Rappresenta il legame che esiste tra i due punti secondo le regole della prospettiva. In generale, qualsiasi sia il sistema di riferimento assunto vale:

$$p \cong \mathbf{M}P$$

Dove:

$$\mathbf{M} = \mathbf{K}[\mathbf{R}|\mathbf{t}]$$

Se il sistema di riferimento assoluto coincide con il sistema di riferimento della fotocamera allora la matrice di proiezione prospettica assume la forma semplificata:

$$\mathbf{M} = \mathbf{K}[\mathbf{I}|\mathbf{0}]$$

2.2 PROPRIETÀ DELLA MATRICE DI PROIEZIONE PROSPETTICA

2.2.1 Forma cartesiana

È possibile scrivere la MPP secondo le sue righe nella forma cartesiana come:

$$\mathbf{M} = \begin{bmatrix} m_1^T \\ m_2^T \\ m_3^T \end{bmatrix} \quad (2.16)$$

Per cui la $p \cong MP$ diventa:

$$p \cong \begin{bmatrix} m_1^T \\ m_2^T \\ m_3^T \end{bmatrix} P \cong \begin{bmatrix} m_1^T P \\ m_2^T P \\ m_3^T P \end{bmatrix} \quad (2.17)$$

Quindi in coordinate cartesiane l'equazione della proiezione prospettica diventa:

$$\begin{cases} u = \frac{m_1^T P}{m_3^T P} \\ v = \frac{m_2^T P}{m_3^T P} \end{cases} \quad (2.18)$$

In fotogrammetria analitica queste due equazioni sono note con il nome di **equazioni di collinearità**.

2.2.2 Centro di presa

Per definizione il centro di presa C è l'intersezione dei piani $u = 0$; $v = 0$ e il piano focale che rappresenta il luogo dei punti che si proiettano all'infinito. Viene dunque definito partendo dalla loro intersezione:

$$0 \cong \begin{bmatrix} m_1^T \\ m_2^T \\ m_3^T \end{bmatrix} C \cong \begin{bmatrix} m_1^T C \\ m_2^T C \\ m_3^T C \end{bmatrix} \quad (2.19)$$

Ovvero $MC = 0$. Questa relazione esprime il fatto che il centro di presa è l'unico punto che ha proiezione nulla, o meglio proiezione non definita. Nei paragrafi precedenti abbiamo osservato come l'applicazione della MPP contenga tutti i parametri di tali trasformazioni e coerentemente con ciò, la sua applicazione al centro di presa restituisce il valore 0.

Potendo esprimere $M = K[R|t]$ come:

$$M = K[R|t] = [KR|Kt] = [Q|q] \quad (2.20)$$

Dove Q rappresenta la sottomatrice 3x3 e q la 3x1.

In forma cartesiana sappiamo che la relazione $MC = 0$ può essere espressa come:

$$Q\check{C} + q = 0$$

da cui

$\check{C} = -Q^{-1}q$ dove \check{C} è la rappresentazione cartesiana di

$$C = \begin{bmatrix} \tilde{C} \\ 1 \end{bmatrix} = \begin{bmatrix} -Q^{-1}q \\ 1 \end{bmatrix} \quad (2.21)$$

2.2.3 Raggio ottico

Si definisce raggio ottico del punto p , la retta passante per C ed il punto p stesso. Sul raggio ottico di p giacciono pertanto tutti i punti di cui p è proiezione. A questa retta appartiene sicuramente il punto C e sicuramente il punto ideale p_{id} , cioè quello proiettato all'infinito. Per definizione tale punto è rappresentabile da:

$$p_{id} = \begin{bmatrix} Q^{-1}p \\ 0 \end{bmatrix} \quad (2.22)$$

Come si può dimostrare dalla seguente relazione:

$$p = M p_{id} = M \begin{bmatrix} Q^{-1}p \\ 0 \end{bmatrix} = [Q|q] \begin{bmatrix} Q^{-1}p \\ 0 \end{bmatrix} = QQ^{-1}p = p \quad (2.23)$$

L'equazione della retta passante per C e per p_{id} sarà:

$$P = C + \lambda \begin{bmatrix} Q^{-1}p \\ 0 \end{bmatrix}, \lambda \in \mathbb{R} \cup \{\infty\} \quad (2.24)$$

2.3 CALIBRAZIONE DELLA FOTOCAMERA

Per risolvere le equazioni appena viste è necessario conoscere sia i parametri di orientamento interno che quelli di orientamento esterno. La calibrazione della fotocamera è un procedimento finalizzato alla misura di questi parametri. L'idea che sta alla base del procedimento è quella di invertire le equazioni della proiezione prospettica partendo da una serie di punti di coordinate note, detti **punti di calibrazione**.

2.3.1 Distorsione radiale

Un modello più accurato deve tener conto anche della distorsione ottica provocata dalle lenti. Gli effetti di tali aberrazioni sono solitamente quelli osservabili in Figura 2.7:



Figura 2.5. Tipici esempi di distorsione

Possiamo dividere la distorsione in funzione delle sue componenti principali e studiarne i contributi separatamente.

Si consideri dunque il punto immagine p_d di coordinate (u_d, v_d) affetto da distorsione radiale ed il punto p_u invece di coordinate (u_u, v_u) la posizione corretta del punto sul piano immagine. Pertanto un qualsiasi punto sul piano immagine può essere descritto dalla relazione:

$$\begin{aligned} u_d &= u_u + \Delta u \\ v_d &= v_u + \Delta v \end{aligned} \quad (2.25)$$

Come si può notare vale anche la relazione:

$$\frac{\Delta r}{r_u} = \frac{\Delta u}{u_u} = \frac{\Delta v}{v_u} \quad (2.26)$$

Sostituendo:

$$\begin{aligned} u_d &= u_u - \Delta u = u_u \left(1 - \frac{\Delta u}{u_u}\right) = u_d \left(1 - \frac{\Delta r}{r_u}\right) = u_u \delta(r_u) \\ v_d &= v_u - \Delta v = v_u \left(1 - \frac{\Delta v}{v_u}\right) = v_u \left(1 - \frac{\Delta r}{r_u}\right) = v_u \delta(r_u) \end{aligned} \quad (2.27)$$

Sostanzialmente l'errore introdotto dalla distorsione radiale è quantificabile attraverso una funzione $\delta(r_u)$ che chiameremo funzione di distorsione radiale. Si ricorda che $r_u = \sqrt{u_u^2 + v_u^2}$ rappresenta la distanza tra l'origine degli assi e il punto non affetto da distorsione. Allo stesso modo possiamo definire $r_d = \sqrt{u_d^2 + v_d^2}$ e scrivere la formulazione appena vista in forma inversa. Tradizionalmente si preferisce questa scrittura:

$$(u_u; v_u) = (u_d; v_d) \delta^{-1}(r_d) \quad (2.28)$$

Rispetto alle coordinate del punto principale le equazioni appena viste diventano:

$$(u_u - u_0) = (u_d - u_0) \delta^{-1}(r_d) \quad (2.29)$$

$$(v_u - v_0) = (v_d - v_0) \delta^{-1}(r_d)$$

La funzione di distorsione inversa più utilizzata è di tipo polinomiale ed assume la forma tipo:

$$\delta^{-1}(r_d) = 1 + k_1 r_d^2 + k_2 r_d^4 + k_3 r_d^6 + \dots \quad (2.30)$$

Dove k_i indica il coefficiente di distorsione i-esimo. In particolare se si utilizzano lenti affidabili aventi una distorsione limitata è possibile troncare il polinomio dopo il primo coefficiente:

$$\delta^{-1}(r_d) = 1 + k_1 r_d^2 \quad (2.31)$$

Il modello che ne rappresenta il comportamento è una trasformazione dalle coordinate ideali (non distorte) (u, v) a quelle osservabili nell'immagine affette da distorsione (\tilde{u}, \tilde{v}) :

$$\begin{cases} \tilde{u} = (u - u_0)(1 + k_1 r_d^2) + u_0 \\ \tilde{v} = (v - v_0)(1 + k_1 r_d^2) + v_0 \end{cases} \quad (2.32)$$

Dove

$$r_d^2 = \left(\frac{u - u_0}{\alpha_u}\right)^2 + \left(\frac{v - v_0}{\alpha_v}\right)^2 \quad (2.33)$$

E u_0, v_0 sono le coordinate del punto centrico.

Come si valuta e si calibra dunque la distorsione radiale. Supponiamo di avere \tilde{p} le coordinate del pixel affette da distorsione ed P quelle del punto di calibrazione noto. Esisterà sicuramente una matrice MPP per cui vale:

$$\mathbf{p} \cong \mathbf{M}\mathbf{P}$$

Ora supponiamo che la matrice sia nota, almeno per la trasformazione ideale, cioè per quella priva di distorsione. Per ogni punto di calibrazione possiamo dunque scrivere due equazioni:

$$\begin{cases} \tilde{u} - u = (u - u_0) \left(\left(\frac{u - u_0}{\alpha_u}\right)^2 + \left(\frac{v - v_0}{\alpha_v}\right)^2 \right) k_1 \\ \tilde{v} - v = (v - v_0) \left(\left(\frac{u - u_0}{\alpha_u}\right)^2 + \left(\frac{v - v_0}{\alpha_v}\right)^2 \right) k_1 \end{cases} \quad (2.34)$$

Il sistema è risolvibile in k_1 noti i parametri interni $u_0, v_0, \alpha_u, \alpha_v$ e noti M e P da cui è possibile ricavare p . Ancora una volta sarebbe analiticamente sufficiente un solo punto per stimare k_1 ma avendone a disposizione molti otteniamo un sistema sovradeterminato. In questo caso la soluzione si ottiene attraverso la:

$$k_1 = \frac{\sum a_i b_i}{\sum a_i} \quad (2.35)$$

Dove a_i è il coefficiente di k_1 nella i -esima equazione e b_i è il termine noto della i -esima equazione. Una volta noto k_1 possiamo compensare la distorsione radiale invertendo la trasformazione.

2.3.2 Self-Calibration

La procedura di self calibration è comunemente affidata a software che utilizzano un numero elevato di punti di coordinate note per ricostruire la geometria dei fotogrammi utilizzando particolari algoritmi basati sulle equazioni di collinearità (2.17) viste nel paragrafo precedente. Basta acquisire le immagini del foglio prese da diverse angolazioni che in automatico verrà avviata la procedura di calibrazione.

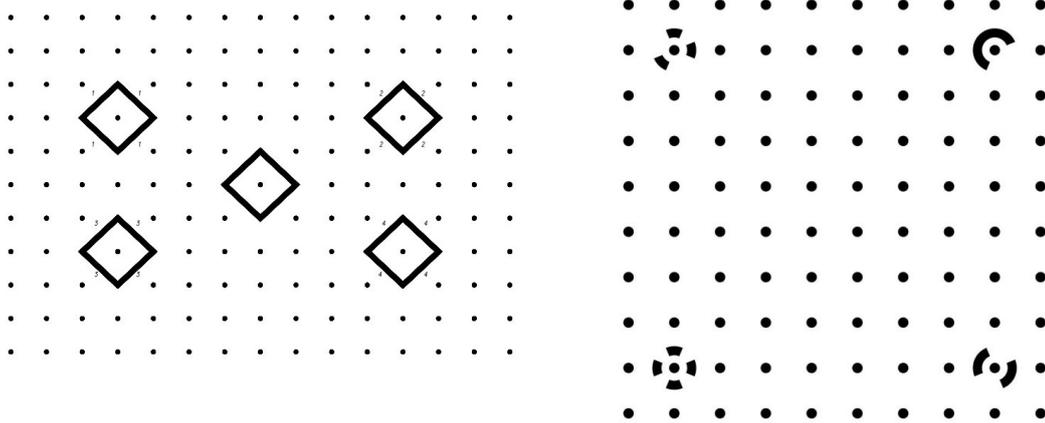


Figura 2.6. fogli di calibrazione per i software Image Master (a sinistra) e Photo Modeler (a destra)

A seconda del programma utilizzato i fogli possono avere geometrie e disegni differenti, come possono essere differenti gli algoritmi di calcolo e di compensazione ma a grandi linee il procedimento resta invariato. A calibrazione avvenuta viene mostrato un report in cui sono visualizzati tutti i risultati della procedura.

2.3.3 Direct Linear Transform (DLT)

Di seguito verrà preso in esame un processo simile anche se il suo è un approccio leggermente diverso. Il metodo è quello proposto da *Faugeras (1993)* che risolve il problema stimando la MPP contenente i parametri incogniti di orientamento. Abbiamo visto come L'equazione di proiezione prospettica sia esprimibile nella forma matriciale:

$$p = \frac{1}{Z}MP$$

Oppure a meno di un fattore di scala:

$$p \cong MP$$

Considerando il seguente modello, assumiamo come punti di calibrazione gli spigoli dei quadrati presenti sui lati della piramide. Il sistema di riferimento è fissato solidale con l'oggetto. Le coordinate dei punti di calibrazione sono note.

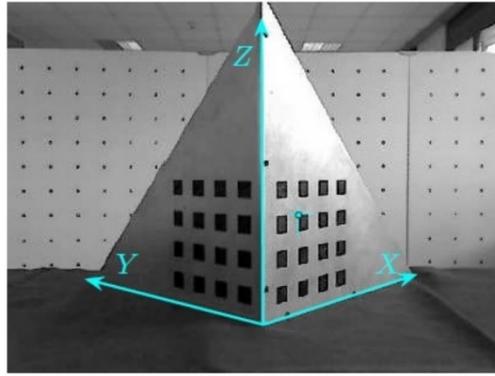


Figura 2.7. Oggetto di calibrazione

Dati n punti di calibrazione non complanari, ciascuna corrispondenza tra un punto dell'immagine $p_i = (u_i, v_i, 1)^T$ ed il punto della scena P_i è esprimibile attraverso la coppia di equazioni note con il nome di equazioni di collinearità:

$$\begin{cases} u_i = \frac{m_1^T P_i}{m_3^T P_i} \\ v_i = \frac{m_2^T P_i}{m_3^T P_i} \end{cases}$$

$$\begin{cases} m_1^T P_i - u_i m_3^T P_i = 0 \\ m_2^T P_i - v_i m_3^T P_i = 0 \end{cases} \quad (2.36)$$

In forma matriciale questa relazione è esprimibile attraverso:

$$\begin{bmatrix} P_i^T & \mathbf{0} & -u_i P_i^T \\ \mathbf{0} & -P_i^T & v_i P_i^T \end{bmatrix} \begin{bmatrix} m_1 \\ m_2 \\ m_3 \end{bmatrix} = \mathbf{0}_{[2 \times 1]} \quad (2.37)$$

Per n punti otteniamo un sistema di $2n$ equazioni lineari omogenee che possiamo scrivere come:

$$Avec(M^T) = \mathbf{0} \quad (2.38)$$

Dove A è la matrice $[2n \times 12]$ dei coefficienti e dipende dalle coordinate dei punti di calibrazione (parametri noti) mentre il vettore delle incognite $vec(M^T)$ contiene i 12 elementi della matrice di proiezione prospettica M letti per righe. In teoria quindi 6 punti non complanari sono sufficienti per risolvere M , anche se in pratica è preferibile utilizzarne molti di più in maniera da compensare ai minimi quadrati gli inevitabili errori di misura.

2.3.3.1 Raffittimento non lineare ai minimi quadrati

Il metodo lineare appena descritto è semplice e veloce ma minimizza uno scarto algebrico che non ha nessun significato geometrico. Per migliorare accuratezza (errore sistematico) e precisione (varianza della stima) è necessario minimizzare un residuo geometrico. La funzione ottenibile è però di tipo non lineare:

$$\varepsilon(P) = \sum_{i=1}^n \left[\left(\frac{m_1^T P_i}{m_3^T P_i} - u_i \right)^2 + \left(\frac{m_2^T P_i}{m_3^T P_i} - v_i \right)^2 \right] \quad (2.39)$$

Si tratta della somma dei quadrati delle distanze tra i punti P_i e la loro proiezione p_i nell'immagine. Essendo l'equazione di tipo non lineare la minimizzazione di $\varepsilon(P)$ avviene attraverso tecniche di risoluzione ai minimi quadrati non lineari le quali convergono solo localmente e pertanto necessitano di una stima iniziale della soluzione, che può essere ottenuta con il metodo lineare.

2.3.3.2 Estrazione dei parametri

Ottenuta dunque la nostra matrice M sappiamo che per estrarre i parametri di orientamento dobbiamo decomporla nella forma:

$$M = K[R|t] = [KR|Kt] = [Q|q] \quad (2.40)$$

Per confronto si ottiene $Q = KR$, con K triangolare superiore ed R ortogonale. Applicando la fattorizzazione QR alla matrice Q possiamo dunque scrivere:

$$Q^{-1} = US = K^{-1}R^{-1} \text{ con } R = U^{-1} \text{ e } K = S^{-1} \quad (2.41)$$

in quanto U ortogonale ed S triangolare superiore.

Infine per ricavare t basta fare:

$$t = K^{-1}q = Sq \quad (2.42)$$

3 STRUCTURE FROM MOTION (S.F.M.)

All'inizio degli anni '90, con il diffondersi delle prime telecamere digitali, nel settore della computer vision l'attenzione si è concentrata sull'analisi di sequenze di immagini, affrontando la ricostruzione della Structure from Motion (*Dhond et al.*, 1989), (*Faugeras*, 1993). Il problema principale era quello di stimare la struttura o la scena inquadrata dalle fotocamere avendo a disposizione un numero molto limitato di informazioni. Nei capitoli precedenti è stato affermato che per poter ricostruire la geometria del problema sono necessari almeno i parametri di orientamento. Questa affermazione non è completamente vera, in quanto come osserveremo nel proseguo della trattazione, anche senza disporre di informazioni precise riguardo l'orientamento delle camere è possibile estrarre direttamente dalle immagini acquisite alcuni importanti aspetti riguardo la geometria di presa. Attraverso queste considerazioni sarà poi possibile ricostruire l'intera struttura della scena che stiamo fotografando o riprendendo.

I metodi di risoluzione della Structure from Motion sono numerosi, come numerosi sono gli algoritmi che concorrono alla risoluzione delle varie fasi del processo. Come accennato nell'introduzione non sarà scopo di questa tesi quello di fornire un elenco esaustivo di tutte le metodologie preposte a tale compito, ma piuttosto fornire una metodologia generale così da comprenderne meglio il funzionamento. Di fatto tutti i metodi di risoluzione si basano su metodologie simili che possono essere riassunte nel seguente elenco:

- Il primo passo generalmente consiste nell'estrazione delle caratteristiche (features) rilevabili nelle diverse immagini. Le features sono una serie di elementi (solitamente punti) che, con un certo grado di probabilità siano riscontrabili su diversi fotogrammi.
- Una volta individuato un numero sufficientemente elevato di features, per mezzo di opportuni criteri geometrici e radiometrici, si stimano delle corrispondenze che formino un primo tentativo di catalogazione di punti coniugati sulle varie immagini.
- A questo punto è possibile stimare una geometria di presa preliminare ed eliminare le corrispondenze sbagliate.
- Al termine delle stime la geometria individuata non strettamente verifica le condizioni reali del blocco fotogrammetrico, nel senso che fornisce una geometria proiettivamente equivalente a quella reale. Per ottenere una ricostruzione reale (metrica) del blocco e della scena è sufficiente la conoscenza dei parametri di orientamento interno. Tali parametri possono essere noti a priori ad esempio attraverso una procedura di self-calibration, oppure possono essere stimati a posteriori durante il processo di ricostruzione attraverso una metodologia nota con il termine di autocalibrazione.
- Infine, è possibile ottenere una stima ottimale utilizzando i risultati ottenuti come parametri approssimati di un sistema risolvibile a minimi quadrati (*bundle adjustment*) in cui possono essere inserite informazioni aggiuntive sul blocco (punti d'appoggio, interdistanze note, orientamenti etc...).

- Resta da completare il processo di ricostruzione attraverso una procedura di triangolazione pixel dopo pixel in tutte le immagini disponibili. Si parla di Dense matching. Al termine di questa fase avremo a disposizione la classica nube di punti su cui poi sarà possibile eventualmente modellare una struttura solida formata da mesh.

I passaggi principali sono riassunti nel seguente schema a blocchi:

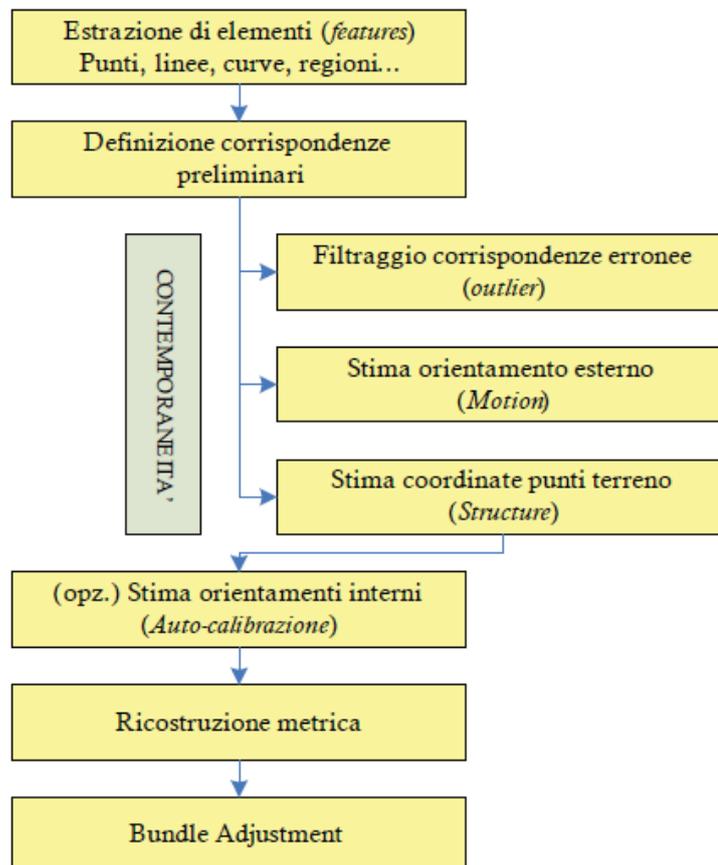


Figura 3.1. il diagramma a blocchi riportato rappresenta le fasi principali che compongono un processo di structure from motion classico. Il diagramma è stato estratto dall'articolo: "Sviluppo e applicazioni di tecniche di automazione in fotogrammetria dei vicini" di A.Roncella, Università di Parma.

Di seguito verranno analizzate in dettaglio le fasi principali del procedimento ponendo particolare attenzione alle procedure di ricostruzione multivista non calibrata, che rappresenta il caso più generale dell'applicazione. Per una miglior comprensione del processo finale verranno suddivise in capitoli le principali fasi partendo da quella forse più importante ed innovativa che è il riconoscimento automatico delle features (punti caratteristici), su cui per altro si basa tutta la ricostruzione.

4 L'ESTRAZIONE DELLE CARATTERISTICHE

4.1 IL CONCETTO DI CARATTERISTICA

Il rilevamento delle cosiddette *features* (o caratteristiche) è sicuramente una componente essenziale per numerose applicazioni della computer vision ed in particolare per tutti processi di structure from motion. É il primo passo che una macchina (o computer) deve compiere, se vuole interpretare nella maniera corretta i fotogrammi a disposizione e se vuole ricostruire nelle tre dimensioni l'intera scena ripresa. La domanda nasce quindi spontanea; quali sono gli elementi, le *features* (o caratteristiche) distintive delle immagine che abbiamo di fronte? Purtroppo una vera e propria definizione di caratteristica è difficile da dare, soprattutto se si osservano le immagini dal punto di vista dei computer (immagine=matrice). Per capire meglio questo concetto e comprenderne il significato si osservi il seguente esempio (figura 4.1.):



Figura 4.1. Esempio di features. Dall'immagine digitale sono state estratte 6 possibili features elencate nei riquadri in alto con le lettere A,B,C,D,E ed F. Sono tutte regioni dell'immagine che possono concorrere a diventare caratteristiche.

Sono state selezionate alcune regioni dell'immagine, tutte possibili candidate a diventare feature. Ora si immagini di dover riposizionare sull'immagine originale ciascuna delle regioni estratte. Sicuramente chiunque sarebbe in grado di posizionare in maniera corretta le regioni (E) e (F), mentre risulterà impossibile ricostruire la posizione della (A) e della (B). Per le restanti (C) e (D) pur sapendo di quale porzione di immagine si tratti, è difficile risalire alla loro posizione esatta. Tra le regioni

proposte la scelta delle feature ricadrebbe sicuramente sulla (E) e sulla (F). Sono le uniche facilmente ed univocamente riconoscibili poiché rappresentano gli angoli (**corners**) dell'edificio rappresentato in figura. Le regioni che presentano una texture omogenea o ricorrente, come la (A) e la (B) sono le cosiddette regioni **flat**, per le quali trovare una corrispondenza è impossibile. Nelle restanti regioni si trovano i bordi (**edges**), sappiamo dove sono ma non riusciamo a trovarne con esattezza la posizione.

Capito il concetto di feature la domanda successiva che ci si pone è: come riesce un computer a trovare i punti caratteristici all'interno di una immagine? La brillantezza di un singolo pixel risulta poco discriminativa ed in generale si ricerca qualche proprietà funzione dei livelli di grigio di una regione circostante il punto indagato. Il metodo più semplice e immediato per rilevare la presenza di punti caratteristici è quindi quello di studiare la luminosità di un insieme di pixel osservandone eventuali variazioni in un suo preciso intorno. Prima di darne una formulazione analitica cerchiamo di capire meglio il concetto attraverso un altro esempio. Si osservi la figura 4.2.:

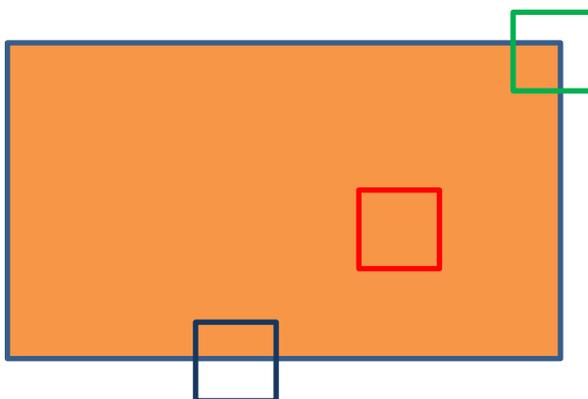


Figura 4.2. Le tre regioni quadrate che si distinguono sull'immagine rappresentano tre ipotetiche finestre all'interno delle quali può essere presente una feature. Riprendono direttamente le tre zone definite nel paragrafo precedente: regioni flat, edges e corners

Supponiamo di poter osservare l'immagine un pezzo alla volta attraverso apposite finestre di dimensioni finite. Pixel dopo pixel viene analizzata tutta l'immagine semplicemente spostando la finestra nelle due direzioni principali. Quando la finestra si troverà in corrispondenza di una regione **flat** (quadrato rosso), ogni piccolo spostamento non comporterà nessuna importante variazione in termini di intensità luminosa, pertanto la macchina in quelle zone non rileverà nessun punto di interesse. Se ci si trova in corrispondenza di un bordo (quadrato blu) spostamenti in direzione parallela al bordo non sortiranno nessun cambio di intensità, mentre per spostamenti in direzione ortogonale al bordo si noterà un brusco cambio di intensità. Tutti i pixel soggetti a tale fenomeno saranno inquadrati come **edges**. Vediamo ora il caso dell'angolo (**corner**) (verde), ogni piccola variazione in una qualunque direzione comporta sempre un significativo cambio di luminosità, è stato trovato il punto di interesse. La procedura appena descritta deve essere espressa in termini informatici attraverso particolari algoritmi che consentano ad un computer di analizzare e riconoscere le possibili feature presenti nella scena. Questo processo è denominato **feature detection**.

Quelle proposte sono solo alcune delle famiglie di caratteristiche distinguibili all'interno di un'immagine, ne esistono altre definite in maniera più astratta con il termine **blob**. Sono regioni o parti di un'immagine caratterizzate da una o più proprietà che restano costanti, differenziandole in maniera univoca dalle altre regioni presenti nell'immagine. Non hanno una forma o una dimensione standard, si distinguono proprio in funzione di una determinata proprietà. Il processo che ne determina l'individuazione è noto come **blob detection**.

Una volta trovate tutte le feature all'interno di una immagine, si passa all'immagine successiva e così via per tutte le immagini a disposizione. Per poter poi trovare le giuste corrispondenze tra le varie feature è necessario associare ad esse una breve "descrizione" ad esempio della regione che le circonda. A parole risulterebbe piuttosto semplice. In riferimento alla figura (4.1.), se la feature che desideriamo descrivere è la (F) basterebbe dire che si tratta di un angolo della copertura dell'edificio, che la parte superiore della regione è di colore azzurro, che si trova nella parte superiore dell'immagine. Tutto ciò può essere tradotto in un linguaggio informatico attraverso un processo definito **feature description**, gli algoritmi preposti a tale compito vengono definiti *descrittori*.

Solo dopo aver trovato le caratteristiche e dopo averle accuratamente descritte è possibile confrontarle tra loro e stabilirne le eventuali corrispondenze (**feature matching**). Esistono due approcci principali per la ricerca dei punti di interesse e la stima delle loro corrispondenze. Il primo consiste nel rilevare una caratteristica e tracciarne il movimento nelle immagini successive (**tracking**) utilizzando ad esempio tecniche di ricerca localizzate, la seconda invece consiste nel rilevare prima le caratteristiche in tutte le immagini e successivamente cercare le corrispondenze sulla base della loro descrizione (**matching**). La prima metodologia è particolarmente utile se si dispone di un dataset di immagini acquisito in successione (ad esempio le classiche strisciate aerofotogrammetriche), mentre la seconda è probabilmente più adatta qualora si disponga di fotogrammi acquisiti con importanti variazioni di scala e di orientamento.

Il processo di estrazione delle feature si articola dunque in tre momenti principali:

- **Feature detection:** è il rilevamento delle caratteristiche e racchiude una serie di metodi per l'estrapolazione di informazioni da una immagine e per prendere decisioni locali sull'esistenza o meno di una caratteristica in quel determinato punto. Le caratteristiche risultanti saranno un sottoinsieme del dominio dell'immagine;
- **Feature description:** letteralmente una descrizione delle caratteristiche. Tipicamente viene assolta con l'indicazione della loro posizione, di una qualche funzione dell'intensità luminosa e di un angolo, che ne computa la direzione. Queste informazioni sono racchiuse all'interno di un vettore che prende il nome di descrittore.
- **Feature tracking or feature matching:** è di fatto l'accoppiamento delle singole caratteristiche nelle diverse immagini.

La vera innovazione del processo di restituzione sta proprio in questa prima fase del lavoro, quelli che un tempo erano punti omologhi riconosciuti a mano da un operatore, grazie alla computer vision, possono essere oggi computati in maniera automatica. L'esempio sicuramente migliore di feature detection, comprensivo di tutte le fasi sopracitate, lo si può trovare nell'articolo di David Lowe (2004) in cui presenta una descrizione dettagliata del suo Scale Invariant Feature Transform (SIFT). Nei paragrafi successivi osserveremo in dettaglio le diverse fasi del lavoro ed in particolare alcuni algoritmi per l'estrazione delle caratteristiche.

4.2 FEATURE DETECTION AND DESCRIPTION

Tra le famiglie di feature che si possono distinguere all'interno di una immagine digitale, rivestono sicuramente un ruolo importante i cosiddetti punti di interesse o *keypoint*. Si tratta formalmente di punti, definiti attraverso particolari proprietà, come ad esempio l'intensità luminosa. Sono particolarmente utili perché consentono una stima preliminare delle corrispondenze anche in presenza di occlusioni o di immagini acquisite con scale differenti e variazioni importanti di orientamento. Inoltre sono complementari a tutte le altre famiglie di caratteristiche (edges) e possono tranquillamente essere utilizzate insieme.

Gli angoli fanno sicuramente parte di questa famiglia. Pertanto, di seguito si osserva probabilmente l'algoritmo più famoso per l'individuazione di tali caratteristiche, noto con il nome di **Harris Corner Detector** (Harris e Stephens: A combined Corner and Edge detector-1988).

4.2.1 Harris corner detector (1988)

Per comprendere meglio il funzionamento dell'algoritmo si osservi il ragionamento che vi è alla base. Moravec nel 1977 aveva formulato durante i suoi studi un semplice metodo per il riconoscimento di entità particolari (ostacoli) presenti sulla scena di una immagine. Diventò probabilmente il primo algoritmo di corner detection, da cui proprio Harris nel 1988 trasse ispirazione per il le sue teorie. Di fatto l'Harris corner detector può anche essere visto come una rivisitazione delle teorie di Moravec, mirata alla correzione di alcune problematiche rilevate. Partiamo dunque dalle considerazioni di Moravec.

Si consideri un'immagine in scala di grigi. Sia $I_{u,v}$ l'intensità luminosa di un qualsiasi punto di coordinate (u, v) all'interno dell'immagine. Si definisce la variazione di luminosità $E_{x,y}$ prodotta da un piccolo vettore spostamento (x, y) noto, attraverso la relazione:

$$E_{x,y} = I_{(x+u,y+v)} - I_{u,v} \quad (4.1)$$

Tipicamente gli spostamenti considerati sono unitari e lungo le 8 direzioni principali $(1,0)(1,1)(0,1)(-1,0)(-1,-1)(0,-1)(-1,1)(1,-1)$. La formulazione così presentata è piuttosto debole. Le immagini sono affette da rumore, e spesso le differenze in termini di luminosità non sono così brusche da consentirne una stima accurata. Per rafforzare leggermente questa relazione, dal momento che non ci interessa la direzione dell'intensità ma soltanto il suo modulo possiamo elevare al quadrato la relazione appena vista.

$$E_{x,y} = (I_{(x+u,y+v)} - I_{u,v})^2 \quad (4.2)$$

Il calcolo così composto deve essere eseguito per ogni pixel (u, v) dell'immagine.

$$E_{x,y} = \sum_{u,v} (I_{(x+u,y+v)} - I_{u,v})^2 \quad (4.3)$$

Formalmente calcolare tale somma per tutti i pixel dell'immagine non è molto utile, basta valutare quello che accade in un preciso intorno del pixel. Per questo si utilizzano delle finestre centrate nel punto di interesse $w_{u,v}$ di dimensioni note nelle quali calcolare la variazione di intensità. La finestra in questo caso è di dimensioni quadrate, e assume valore unitario all'interno della regione fissata e zero in tutte le altre parti. Premoltiplicando la formula di cui sopra per la finestra così composta si ottiene:

$$E_{x,y} = \sum_{u,v} w_{u,v} (I_{(x+u,y+v)} - I_{u,v})^2 \quad (4.4)$$

L'algoritmo proposto da Moravec si compone dunque dei seguenti passaggi:

- Per ogni pixel (u, v) dell'immagine viene calcolato attraverso la 4.2 la variazione di intensità $E_{x,y}$ in ogni direzione. Si avranno pertanto per ciascun pixel 8 valori di $E_{x,y}$
- Si costruisce la cosiddetta "cornerness map" estraendo per ciascun pixel il valore minimo di intensità trovato nelle varie direzioni:

$$C(u, v) = \min(E_{x,y}(u, v)) \quad (4.5)$$

- Si fissa una determinata soglia T sotto la quale tutti i valori di C compresi tra tale soglia e zero vengano posti uguali a zero. Tipicamente la soglia è una frazione del valore massimo (preferibile);
- Si effettua la soppressione dei non-massimi restanti. Si fa scorrere una finestra di dimensioni prefissate su tutta l'immagine e si forza il pixel centrale a zero se non è il massimo del suo intorno, altrimenti si lascia inalterato. Così facendo si mettono in evidenza effettivamente solo i massimi;
- I punti restanti sulla mappa "non nulli" sono gli angoli ricercati.

Esempio di calcolo (realizzato con un foglio excel): si consideri la seguente immagine:



255	255	255	255	255	255
255	255	255	255	255	255
255	255	255	255	255	255
0	0	0	255	255	255
0	0	0	255	255	255
0	0	0	255	255	255

Sia data come finestra dell'immagine la funzione peso w

0	0	0	0	0
0	1	1	1	0
0	1	1	1	0
0	1	1	1	0
0	0	0	0	0

Eseguiamo quindi il calcolo dell'intensità luminosa nei pixel dell'immagine. Partiamo dal pixel che si trova in posizione (3,3):

255	255	255	255	255	255
255	255	255	255	255	255
255	255	255	255	255	255
0	0	0	255	255	255
0	0	0	255	255	255
0	0	0	255	255	255

A1	A2	A3
A4	A5	A6
A7	A8	A9

B1	B2	B3
B4	B5	B6
B7	B8	B9

La finestra in rosso indica la posizione originale della finestra di ricerca. Quella in verde rappresenta la posizione della finestra dopo un spostamento di (1,1). Il calcolo dell'intensità avviene analiticamente secondo la formula:

$$E_{x,y} = \sum_{u,v} w_{u,v} (I_{(x+u,y+v)} - I_{u,v})^2$$

Da cui:

$$E_{x,y} = \sum_{i=1}^9 w_i (A_i - B_i)^2$$

Di seguito vediamo i conti eseguiti per il pixel in posizione (3,3) e le 8 direzioni di spostamento:

Direzione (1,1)

255	255	255
0	0	255
0	0	255

Direzione (1,0)

255	255	255
0	0	255
0	0	255

Direzione (1,-1)

255	255	255
0	0	255
0	0	255

Direzione (0,-1)

255	255	255
0	0	255
0	0	255

255	255	255
255	255	255
0	255	255

$$E = 3 \cdot 255^2$$

$$= 195075$$

255	255	255
0	255	255
0	255	255

$$E = 2 \cdot 255^2$$

$$= 130050$$

0	255	255
0	255	255
0	255	255

$$E = 3 \cdot 255^2$$

$$= 195075$$

0	0	255
0	0	255
0	0	255

$$E = 2 \cdot 255^2$$

$$= 130050$$

Direzione (-1,-1)

Direzione (-1,0)

Direzione (-1,1)

Direzione (0,1)

255	255	255
0	0	255
0	0	255

255	255	255
0	0	255
0	0	255

255	255	255
0	0	255
0	0	255

255	255	255
0	0	255
0	0	255

0	0	0
0	0	0
0	0	0

$$E = 5 \cdot 255^2$$

$$= 325125$$

255	255	255
0	0	0
0	0	0

$$E = 2 \cdot 255^2$$

$$= 130050$$

255	255	255
255	255	255
0	0	0

$$E = 3 \cdot 255^2$$

$$= 195075$$

255	255	255
255	255	255
0	0	255

$$E = 2 \cdot 255^2$$

$$= 130050$$

Pertanto

$$C(u, v) = \min(E_{x,y}(u, v)) = \min(E_{x,y}(3,3)) = 130050$$

Eseguendo gli stessi calcoli per ciascun pixel dell'immagine otteniamo la cornerness map. La riportiamo di seguito sogliata con un valore $T = 255^2$ (cioè dividiamo i valori ottenuti per $1/255^2$). Il risultato sarà il seguente:

X	X	X	X	X	X
X	X	X	X	X	X
X	X	1	1	X	X
X	X	2	1	X	X
X	X	X	X	X	X
X	X	X	X	X	X

Le X rappresentano quei punti dell'immagine in cui non è stato possibile calcolare la variazione in quanto lo spostamento della finestra in talune direzioni non era consentito (una parte della finestra

di ricerca usciva dall'immagine). Nella fase successiva dove andremo a sopprimere tutti i valori non massimi della mappa al posto delle X verranno inseriti degli zeri. Il risultato finale sarà dunque una mappa del tipo:

0	0	0	0	0	0
0	0	0	0	0	0
0	0	0	0	0	0
0	0	2	0	0	0
0	0	0	0	0	0
0	0	0	0	0	0

Tutti i punti diversi da zero rimasti all'interno della mappa sono angoli.

Punti deboli dell'algoritmo:

- *Risposta anisotropa:* è forse uno dei difetti più evidenti dell'algoritmo proposto, poiché la finestra si muove soltanto lungo le direzioni principali. Pertanto tutti gli edge che hanno una direzione differente da quelle predefinite presentano valori di C molto elevati pur non essendo di fatto corner.

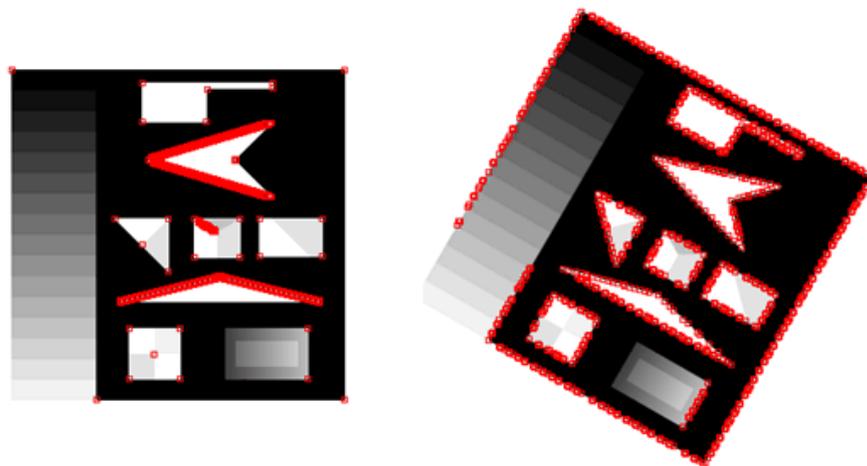


Figura 4.3. esempio di features estratte da immagini identiche semplicemente ruotate. Nella seconda la stima dei corner è molto diversa dalla prima. Come si osserva tutti i pixel identificativi di edges sono stati erroneamente scambiati per corner. Allo stesso modo nella prima immagine, tutti gli oggetti con edges ruotati rispetto le direzioni di ricerca sono stati erroneamente computati come corner.

L'operatore avendo un insieme finito di spostamenti presenta un basso tasso di ripetibilità in quanto non è invariante rispetto alla rotazione. Immagini acquisite con angoli di rotazione diversi darebbero origine a cornerness map molto diverse rendendo difficile la corretta stima delle corrispondenze.

- *Risposta rumorosa:* La finestra w usata come peso è quadrata e binaria. La distanza dei pixel lungo i bordi non è costante, ma aumenta con l'avvicinarsi agli angoli della finestra. Per ottenere una stima più accurata della variazione di intensità sarebbe più adatto utilizzare una

finestra circolare, in maniera che la distanza dal pixel centrale sia la stessa in tutti i punti. Se poi al posto di una griglia discreta avessi una funzione peso diversa la stima sarebbe ancora migliore.

- Elevata risposta in presenza di bordi: Per definizione le zone di bordo (edges) sono zone caratterizzate da un'elevata variazione di intensità, soprattutto nelle immagini in scala di grigi. Questo fa sì che ogni piccola imperfezione dei bordi causata ad esempio da rumore, pixelizzazione errata ed imperfezioni, provochi una variazione di intensità in tutte le direzioni, proprio come nel caso della rotazione.

Fu proprio Harris a proporre nel 1988 un operatore leggermente diverso in grado di correggere queste problematiche. Ripartiamo dalla scrittura già vista:

$$E_{x,y} = \sum_{u,v} w_{u,v} (I_{(x+u,y+v)} - I_{u,v})^2 \quad (4.6)$$

La correzione al primo problema avviene sostituendo la forma $I_{(x+u,y+v)}$ con una funzione del gradiente del primo ordine dell'immagine. Analiticamente tale sostituzione è lecita in quanto sviluppando in serie di Taylor la funzione somma si ottengono proprio le due componenti del gradiente:

$$I_{(x+u,y+v)} \approx I_{u,v} + \frac{\partial I_u}{\partial u} x + \frac{\partial I_v}{\partial v} y \quad (4.7)$$

In maniera più intuitiva si può notare come le due scritture siano effettivamente una l'approssimazione dell'altra in quanto il gradiente può essere approssimato con:

<p>Mask Labeling</p> <table border="1" style="border-collapse: collapse; text-align: center;"> <tr><td>A1</td><td>A2</td><td>A3</td></tr> <tr><td>A4</td><td>A5</td><td>A6</td></tr> <tr><td>A7</td><td>A8</td><td>A9</td></tr> </table>	A1	A2	A3	A4	A5	A6	A7	A8	A9	<p>X-direction Mask</p> <table border="1" style="border-collapse: collapse; text-align: center;"> <tr><td>-1</td><td>0</td><td>1</td></tr> </table> <p>Horizontal gradient at A_5 :</p> $\frac{\partial I_{A5}}{\partial x} \approx (I_{A6} - I_{A4}) = I_{A5} \otimes (-1, 0, 1)$	-1	0	1
A1	A2	A3											
A4	A5	A6											
A7	A8	A9											
-1	0	1											
	<p>Y-direction Mask</p> <table border="1" style="border-collapse: collapse; text-align: center;"> <tr><td>1</td></tr> <tr><td>0</td></tr> <tr><td>-1</td></tr> </table> <p>Vertical gradient at A_5 :</p> $\frac{\partial I_{A5}}{\partial y} \approx (I_{A2} - I_{A8}) = I_{A5} \otimes (-1, 0, 1)^T$	1	0	-1									
1													
0													
-1													
	<p>Upper-right Horizontal Mask</p> <table border="1" style="border-collapse: collapse; text-align: center;"> <tr><td></td><td></td><td>1</td></tr> <tr><td></td><td>0</td><td></td></tr> <tr><td>-1</td><td></td><td></td></tr> </table> <p>Upper-right diagonal gradient at A_5 :</p> $\frac{\partial I_{A5}}{\partial h} \approx (I_{A3} - I_{A7}) = I_{A5} \otimes \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ -1 & 0 & 0 \end{bmatrix}$			1		0		-1					
		1											
	0												
-1													

Pertanto sostituendo tale formulazione otteniamo:

$$E_{x,y} = \sum_{u,v} w_{u,v} (I_{(x+u,y+v)} - I_{u,v})^2 \approx \sum_{u,v} w_{u,v} \left(I_{u,v} + \frac{\partial I_u}{\partial u} x + \frac{\partial I_v}{\partial v} y - I_{u,v} \right)^2$$

$$E_{x,y} = \sum_{u,v} w_{u,v} \left(\frac{\partial I_u}{\partial u} x + \frac{\partial I_v}{\partial v} y \right)^2 \quad (4.8)$$

È bene osservare che le scritte non sono del tutto uguali ma bensì simili. Il vantaggio di utilizzare il gradiente in vece della somma è che è possibile misurare il gradiente in qualsiasi direzione variando opportunamente le direzioni di calcolo.

A questo punto Harris si pone il problema della funzione peso w . Come accennato in precedenza sarebbe più adeguata una funzione simmetrica che magari aumentasse l'importanza delle zone vicine al centro piuttosto che in quelle di bordo in modo da rendere più robusta la stima del punto osservato. Al posto della finestra rettangolare discreta si passa dunque ad una finestra pesata con una funzione di tipo Gaussiano.

$$w_{u,v} = e^{-(u^2+v^2)/2\sigma^2} \quad (4.9)$$

Gaussian Window

w1 .04	w2 .12	w3 .04
w4 .12	w5 .36	w6 .12
w7 .04	w8 .12	w9 .04

Weighted Horizontal Intensity Variation:

$$V_x = \sum_{i=1}^9 w_i (A_i - B_i)^2 = \sum_{i=1}^9 w_i (B_i - A_i)^2 \approx \sum_{i=1}^9 w_i \left(\frac{\partial I_i}{\partial x} \right)^2$$

where: $\frac{\partial I_i}{\partial x} \equiv I_i \otimes (-1, 0, 1) \approx B_i - A_i$

X-direction Shift

A1	A2 B1	A3 B2	B3
A4	A5 B4	A6 B5	B6
A7	A8 B7	A9 B8	B9

Per porre rimedio infine all'ultima problematica riscontrata, è cioè l'elevata risposta in presenza di bordi l'operatore proposto da Harris riformula la stima del valore di cornerness considerando non più il minimo valore a disposizione ma bensì considerando la differenza misurata tra variazioni di intensità calcolate secondo diverse direzioni.

Possiamo dunque riscrivere l'equazione:

$$\begin{aligned}
E_{x,y} &= \sum_{u,v} w_{u,v} \left(\frac{\partial I_u}{\partial u} x + \frac{\partial I_v}{\partial v} y \right)^2 \\
&= \sum_{u,v} w_{u,v} \left(\frac{\partial I_u}{\partial u}^2 x^2 + 2 \frac{\partial I_u}{\partial u} \frac{\partial I_v}{\partial v} xy + \frac{\partial I_v}{\partial v}^2 y^2 \right) = Ax^2 + 2Cxy + By^2
\end{aligned} \tag{4.10}$$

Dove

$$A = \left(\frac{\partial I_u}{\partial u}^2 x^2 \right) \otimes w_{u,v}; \quad B = \left(\frac{\partial I_v}{\partial v}^2 y^2 \right) \otimes w_{u,v}; \quad C = \left(\frac{\partial I_u}{\partial u} \frac{\partial I_v}{\partial v} \right) \otimes w_{u,v} \tag{4.11}$$

Tale formulazione può essere riscritta in forma matriciale:

$$E_{x,y} = [x \quad y] \begin{bmatrix} A & C \\ C & B \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = [x \quad y] \mathbf{M} \begin{bmatrix} x \\ y \end{bmatrix} \tag{4.12}$$

La matrice M è semi-definita positiva e simmetrica, perciò i suoi autovalori (λ_1, λ_2) saranno non negativi. Gli autovalori sono un indice della variazione di intensità dell'immagine essendo proporzionali alle curvature principali dell'autocorrelazione locale e formano un descrittore invariante per rotazione della matrice M , infatti:

- Se gli autovalori sono entrambi nulli o molto vicini a zero nel punto considerato, l'immagine sarà uniforme e quindi è presente un blob.
- Se soltanto uno è prossimo a zero, mentre l'altro è molto più grande si ha un edge. L'autovettore associato all'autovalore più grande è perpendicolare all'edge
- Se entrambi gli autovalori sono grandi si ha un corner. Essendo una matrice 2×2 , gli autovalori sono facilmente calcolabili in forma chiusa dal calcolo del polinomio caratteristico:

$$\lambda_{1,2} = \frac{1}{2} \left(A + C \pm \sqrt{(A - C)^2 + 4B^2} \right) \tag{4.13}$$

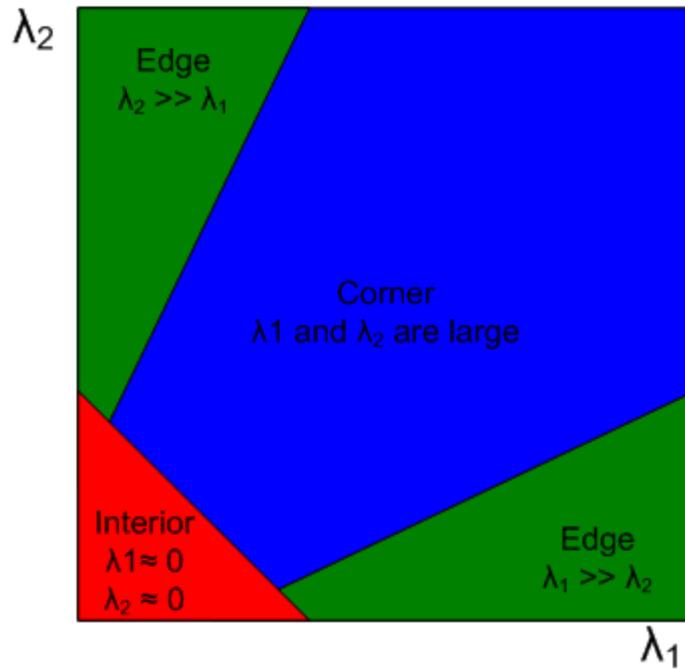


Figura 4.4. Grafico che rappresenta il modulo dei due autovalori della matrice M . In ascissa il primo autovalore, in ordinata il secondo. Si osserva come in funzione della loro combinazione si possano ottenere caratteristiche differenti.

Una valutazione del valore di cornerness può essere dunque fatta partendo da:

$$C(u, v) = \det(M) - k(\text{Tr}(M))^2 \quad (4.14)$$

$$\det(M) = \lambda_1 \lambda_2 = AB - C^2$$

$$\text{Tr}(M) = \lambda_1 + \lambda_2 = A + B$$

Da cui

$$C(u, v) = \lambda_1 \lambda_2 - k(\lambda_1 + \lambda_2)^2 \quad (4.15)$$

K è una costante che attenua o accentua la curvatura della funzione di autocorrelazione degli autovalori.

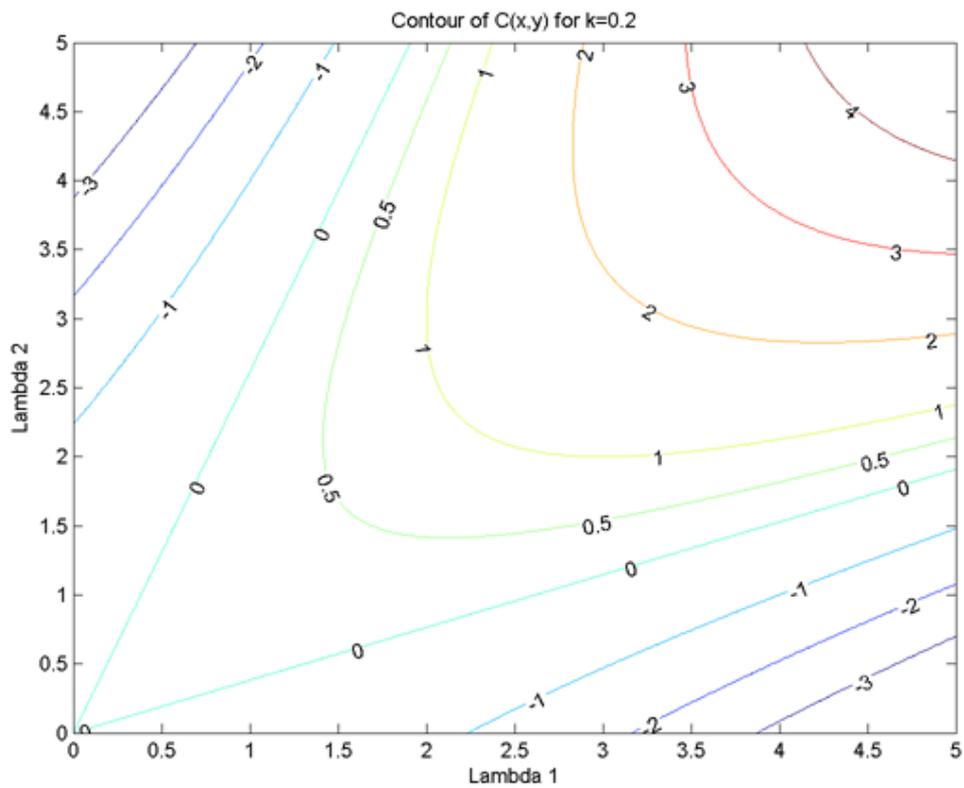


Figura 4.5. Nell'immagine è rappresentata la funzione di Cornerness, funzione della correlazione tra i due autovalori di M . In particolare si osservano le curvature per una costante K pari a 0,2

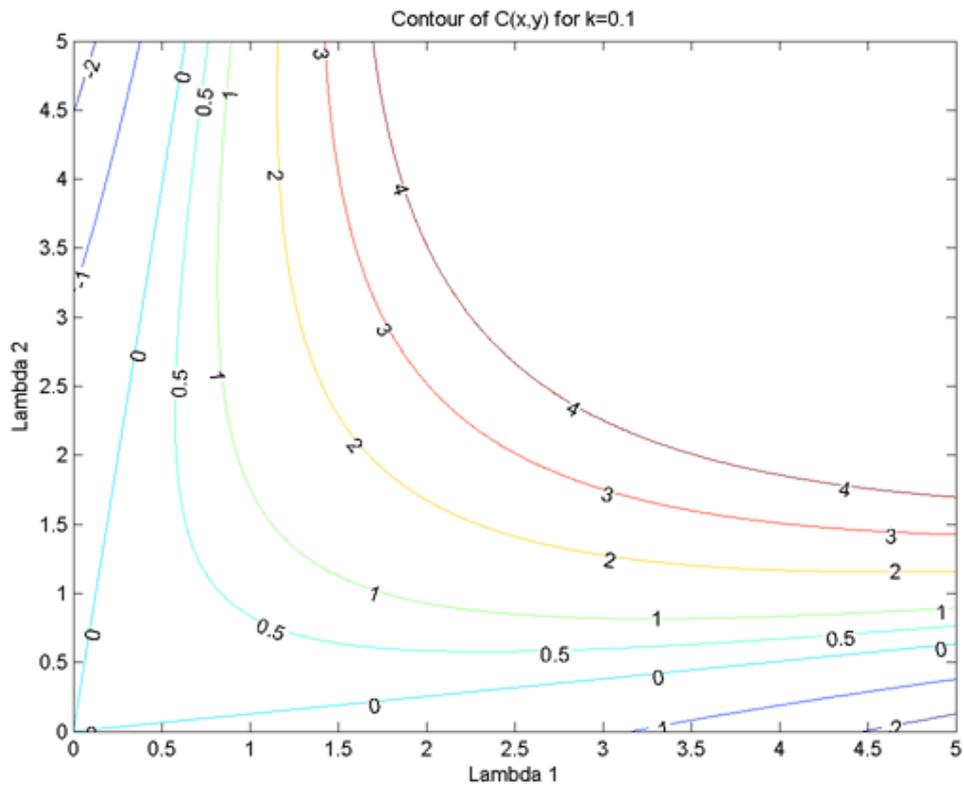


Figura 4.6. Nell'immagine è rappresentata la funzione di Cornerness, funzione della correlazione tra i due autovalori di M . In particolare si osservano le curvature per una costante K pari a 0,1

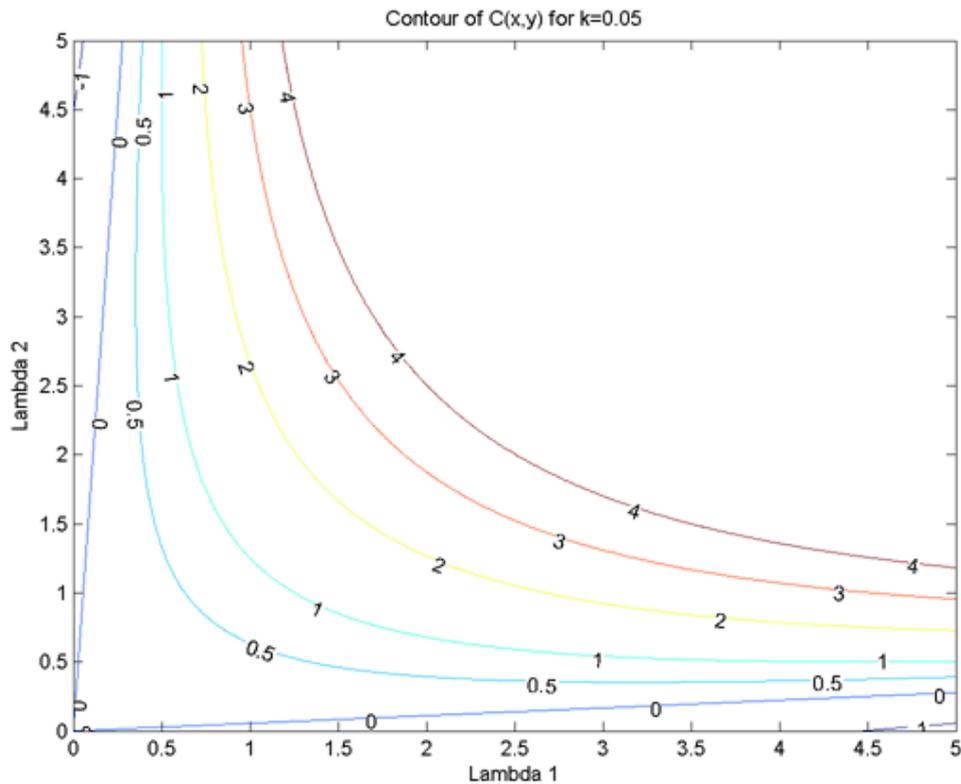


Figura 4.7. Nell'immagine è rappresentata la funzione di Cornerness, funzione della correlazione tra i due autovalori di M . In particolare si osservano le curvature per una costante K pari a 0,05

L'algoritmo proposto da Harris presenta gli stessi passaggi di quello proposto da Moravec, cambia nella forma:

- Per ogni pixel (u, v) dell'immagine viene calcolata la matrice di autocorrelazione M attraverso la stima del gradiente di intensità dell'immagine.

$$E_{x,y} = [x \quad y] \begin{bmatrix} A & C \\ C & B \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = [x \quad y] M \begin{bmatrix} x \\ y \end{bmatrix}$$

- Si costruisce la cosiddetta "cornerness map" estraendo per ciascun pixel :

$$C(u, v) = \det(M) - k(\text{Tr}(M))^2$$

$$C(u, v) = \lambda_1 \lambda_2 - k(\lambda_1 + \lambda_2)^2$$

- Si fissa una determinata soglia T sotto la quale tutti i valori di C compresi tra tale soglia e zero vengano posti uguali a zero. Tipicamente la soglia è una frazione del valore massimo (preferibile);
- Si effettua la soppressione dei non-massimi restanti. Si fa scorrere una finestra di dimensioni prefissate su tutta l'immagine e si forza il pixel centrale a zero se non è il massimo del suo intorno, altrimenti si lascia inalterato. Così facendo si mettono in evidenza effettivamente solo i massimi;

I punti restanti sulla mappa "non nulli" sono gli angoli ricercati.

4.2.2 Shi e Tomasi (1994)

Nel corso degli anni a questo algoritmo sono state apportate alcune modifiche in maniera da renderlo più robusto. Una di queste fu la correzione apportata da J.Shi e C.Tomasi nel 1994. Invece di utilizzare la formulazione (4.15) proposero la seguente:

$$C(u, v) = \min(\lambda_1 \lambda_2) \quad (4.16)$$

Quando tale valore è maggiore di una determinata soglia allora siamo in presenza di un angolo. Plottando tale relazione in uno spazio (λ_1, λ_2) si ottiene:

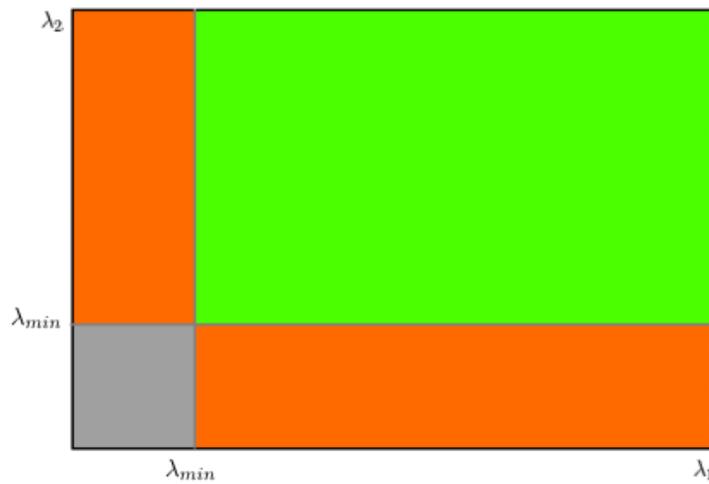


Figura 4.8. Plot della funzione di cornerness. Se gli autovalori presentano un valore maggiore di un prefissato valore minimo allora siamo in presenza di corner (zona verde)

Come si può facilmente notare quando gli autovalori sono maggiori dei loro valori minimi siamo in presenza di un angolo (zona verde). Questa funzione risulta più appropriata in caso di tracking delle caratteristiche.

4.2.3 Scale invariant feature transform: S.I.F.T. (2004)

In questo paragrafo cercheremo di capire come funziona l'algoritmo proposto da Lowe e come descrive i punti caratteristici trovati. Nei paragrafi precedenti abbiamo visto alcuni algoritmi di corner detection invarianti rispetto alle rotazioni. Anche ruotando l'immagine infatti l'angolo rimane sempre un angolo e l'algoritmo allo stesso modo è in grado di riconoscerlo all'interno del fotogramma. Il problema si pone se tra due immagini dello stesso oggetto vi è invece una forte variazione di scala. Osservando l'immagine seguente ci si accorge di come gli algoritmi proposti fino adesso non siano invarianti rispetto alla scala.

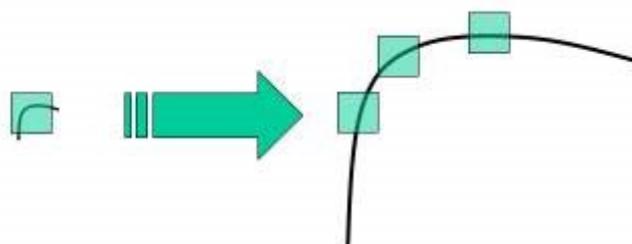


Figura 4.9. Nel riquadro di sinistra si osserva una regione dell'immagine entro la quale è contenuto un angolo. Aumentando la scala dell'immagine e mantenendo costante la dimensione della finestra (della funzione peso) ciò che prima era inequivocabilmente un corner ora è computato come un edge.

Ciò che era riconosciuto come un angolo nel dettaglio di sinistra, non lo è più se la sua scala aumenta come nell'immagine proposta a destra. Per risolvere questo problema di indeterminazione D.Lowe nel 2004 ha proposto un nuovo algoritmo definito Scale-Invariant Feature Transform, ampiamente descritto all'interno del suo articolo "Distinctive Image Features from Scale-Invariant Keypoints". Possiamo suddividere l'algoritmo in quattro momenti fondamentali:

- Definizione dello spazio-scala (scale-space extrema detection);
- Localizzazione dei punti di interesse;
- Orientamento dei punti di interesse;
- Descrizione dei punti di interesse.

4.2.3.1 Definizione dello Spazio-Scala

Si osservi di nuovo l'immagine di cui sopra. È ovvio che non si può utilizzare la stessa finestra di ricerca su immagini a scale diverse. Anche questa deve essere adattata alla scala. Per angoli di piccole dimensioni si può utilizzare una finestra piccola, mentre angoli di grandi dimensioni necessitano per forza di finestre più ampie. Tipicamente per risolvere questo problema si utilizzano dei filtri in maniera tale da creare per ciascuna immagine una serie di versioni smoothed secondo la dimensione del nucleo di filtraggio utilizzato. Tale nucleo funziona dunque come un fattore di scala. Si opera in questo senso per cercare di individuare anche nelle versioni dell'immagine smoothed il maggior numero di punti di interesse così da poterli eventualmente confrontare anche con immagini che

presentano una scala differente. Il filtro più utilizzato è solitamente quello rappresentato da un nucleo Gaussiano.

È dunque possibile definire un punto di interesse attraverso tre parametri (x,y,σ) , dove due ne definiscono la posizione nello spazio (x,y) ed il terzo σ la scala. Si definisce quindi con il termine **spazio-scala** la funzione:

$$L(x, y, \sigma) = G(x, y, \sigma) * I(x, y) \quad (4.17)$$

Calcolata dal prodotto di convoluzione di una gaussiana $G(x, y, \sigma)$ con una immagine di partenza data $I(x, y)$.



Figura 4.10. Spazio scala generato da nuclei di Gaussiana con σ sempre maggiori. Più il valore della deviazione standard è grande più l'immagine iniziale risulta sfuocata. Immagine tratta da: <http://aishack.in/tutorials/sift-scale-invariant-feature-transform-scale-space/>

Uno dei rivelatori più comuni di punti caratteristici si basa sull'operatore Laplaciano della Gaussiana:

$$\nabla^2 L = \left(\frac{\partial^2 L}{\partial x^2} + \frac{\partial^2 L}{\partial y^2} \right) \quad (4.18)$$

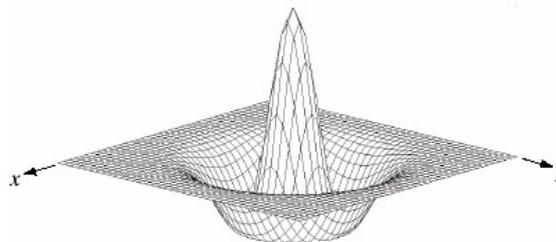


Figura 4.11. Funzione Laplaciano di Gaussiana

I punti caratteristici sono I massimi o I minimi locali della funzione. Da un punto di vista frequenziale il LoG è un filtro passa-banda. In altre parole rivela macchie (blob) ad una determinata scala (fig. 4.12).

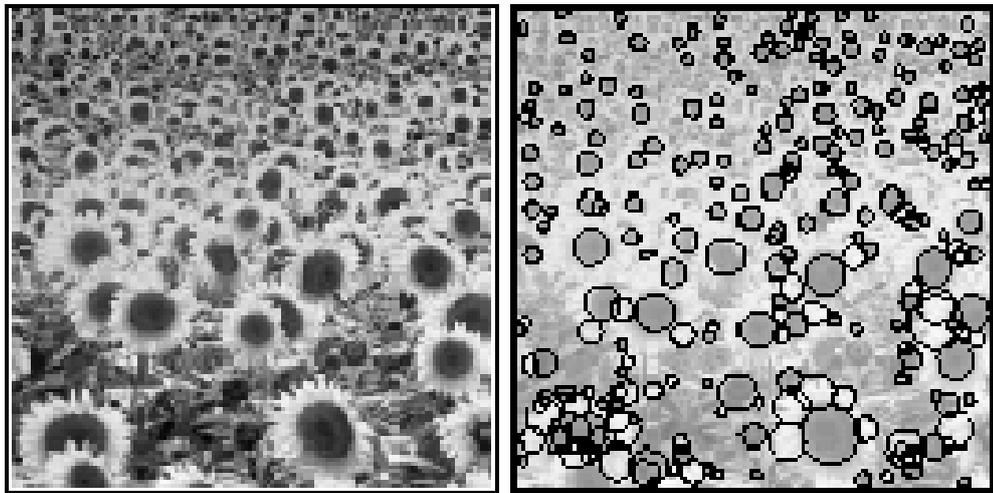


Figura 4.12. Blob rivelati dal Laplaciano di una immagine filtrata ad una determinata scala. (www.nada.kth.se)

L'applicazione di questo tipo di filtro ha però un costo computazionale abbastanza elevato, per questo il SIFT ragiona in maniera leggermente diversa.

Dopo aver creato una sequenza di immagini smoothed di una immagine riscalda l'immagine con un fattore di scala ad esempio pari a 2σ , creando quella che nel gergo si definisce un'ottava superiore. Come per la precedente anche per l'ottava superiore crea una sequenza di immagini smoothed. Viene così costruita una piramide la cui dimensione cambia ogni 4 livelli (Fig. 4.13):

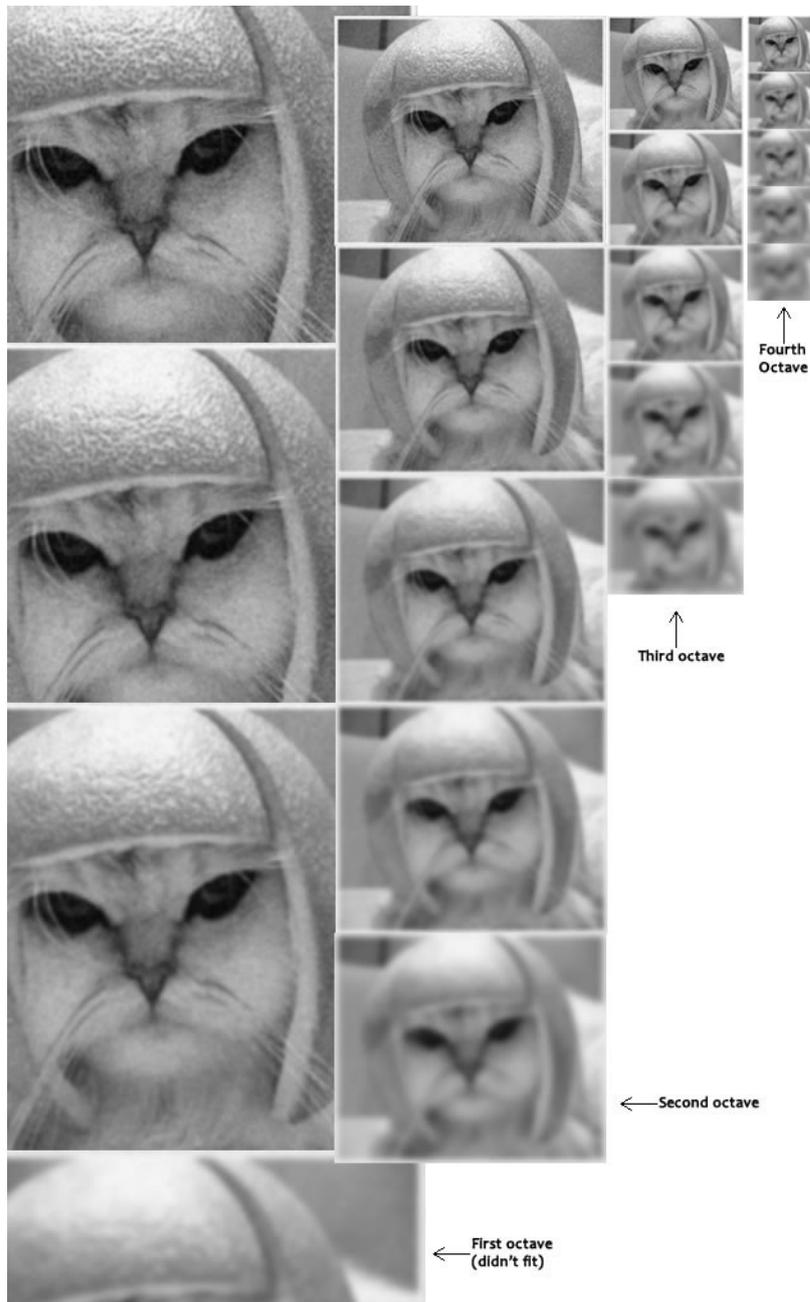


Figura 4.13. Piramide Gaussiana creata dal SIFT. Ad ogni ottava corrisponde una sequenza di immagini smoothed. (<http://aishack.in/tutorials/sift-scale-invariant-feature-transform-scale-space/>)

A questo punto interviene il rivelatore. Al posto del LoG il SIFT utilizza una sua approssimazione definita Differenza di Gaussiane (DoG). Questa differenza si ottiene sottraendo due Gaussiane con diversi σ , in particolare uno σ e l'altro $k\sigma$.

$$DoG(x, y, \sigma) = (G(x, y, k\sigma) - G(x, y, \sigma)) * I(x, y) = L(x, y, k\sigma) - L(x, y, \sigma) \quad (4.19)$$

Il processo è ripetuto per ogni ottava della piramide gaussiana dell'Immagine.

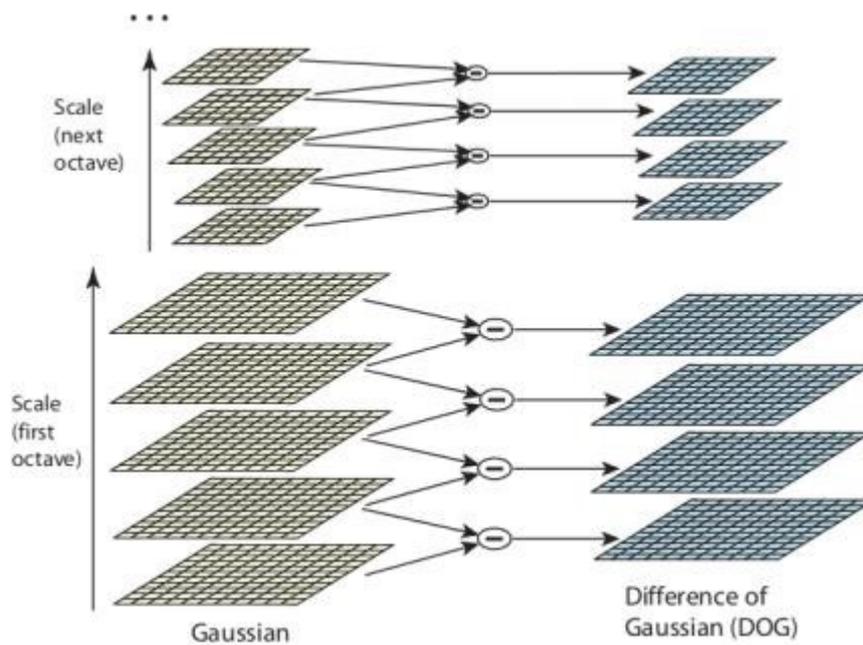


Figura 4.14. Per ogni ottava dello spazio scala, viene eseguita la convoluzione dell'immagine iniziale attraverso un filtro di tipo Gaussiano producendo il set di immagini sulla sinistra. Successivamente per differenza si crea la DoG. Dopo ogni ottava l'immagine viene ridotta della metà e il processo si ripete.

Una volta trovato il DoG si va alla ricerca per ciascun spazio scala dei massimi e minimi locali. Ad esempio, un pixel in una immagine e i suoi 8 pixel vicini vengono comparati con i 9 pixel della scala superiore ed inferiore. Se risulta un massimo potrebbe essere un potenziale punto di interesse.

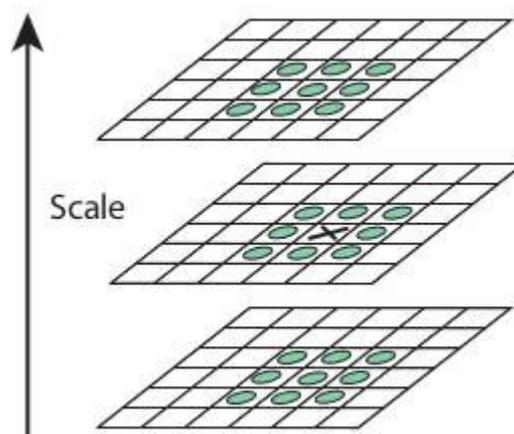
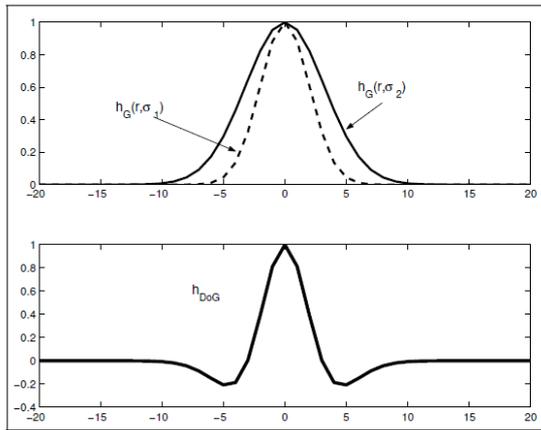


Figura 4.15. I massimi ed i minimi della DoG sono valutati confrontando il pixel X con i suoi 8 pixel vicini nella scala corrente e per le scale precedente e successiva.

Il DoG è utilizzato dunque al posto del Laplaciano in quanto ritenuto una sua approssimazione ottimale. Lo si può osservare nella figura 4.16.

DIFFERENZA DI GAUSSIANE



LAPLACIANO DI GAUSSIANA

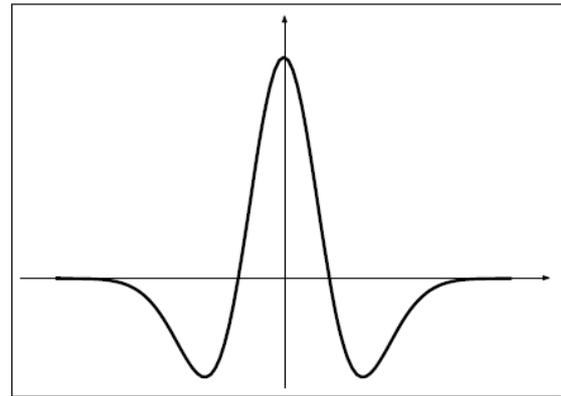


Figura 4.16. In figura si osserva come la differenza tra due Gaussiane (immagine a sinistra) sia molto simile alla funzione di Laplaciano di una Gaussiana (immagine a destra).

Tale approssimazione è dimostrabile anche per via analitica. Si dimostra che:

$$DoG(x, y, \sigma) \approx \frac{(k^2 - 1)}{2} \nabla^2_{norm} L(x, y, \sigma) \quad (4.20)$$

4.2.3.2 Localizzazione dei punti di interesse

Una volta riconosciuti i potenziali punti di interesse devono essere “raffinati” in modo da ottenere risultati più accurati. Serve eliminare i punti di interesse caratterizzati da un basso contrasto o che si trovino in prossimità di edges. Per eliminare i blob a basso contrasto, si rigettano tutti i punti estremanti che hanno una risposta sotto una determinata soglia (in valore assoluto).

Per eliminare invece i punti in corrispondenza di edges (poco significativi per la stima delle features) si utilizza un procedimento simile a quello visto nell’algoritmo di Harris. Si utilizza cioè una matrice 2x2 per computare la curvatura principale. Abbiamo visto come per Harris gli edges erano caratterizzati da un autovalore più grande rispetto all’altro. Ancora se tale rapporto è maggiore di una determinata soglia allora il punto viene scartato.

In questo modo vengono eliminati sia i punti appartenenti agli edges che i punti caratterizzati da un basso contrasto, tutto ciò che resta sono i punti di interesse più rappresentativi.

4.2.3.3 Assegnazione dell’orientamento

Ad ogni punto di interesse viene assegnato un orientamento così da raggiungere l’invarianza nei confronti della rotazione dell’immagine. Per ottenere invarianza rispetto la scala, viene normalizzata la dimensione delle immagini secondo il valore di scala a cui è stato rivelato il punto di interesse. Viene quindi considerata una finestra di 16x16 attorno al punto rivelato. In tale finestra viene

computata l'intensità del gradiente $m(x, y)$ e il suo orientamento $\theta(x, y)$ per differenza di pixel secondo le seguenti formule:

$$m(x, y) = \sqrt{(L(x + 1, y) - L(x - 1, y))^2 + (L(x, y + 1) + L(x, y - 1))^2} \quad (4.21)$$

$$\theta(x, y) = \tan^{-1}((L(x, y + 1) + L(x, y - 1))/(L(x + 1, y) - L(x - 1, y))) \quad (4.22)$$

Si crea così un istogramma in cui il dominio angolare è discretizzato secondo 36 intervalli (circa uno ogni 10 gradi) ed i voti sono dati dall'intensità del gradiente nella specifica direzione. Il massimo dell'istogramma identifica l'orientazione dominante. Il valore viene poi raffinato interpolando tale valore con i valori degli intervalli vicini. Vengono considerate orientazioni multiple se vi sono altri picchi maggiori dell'80% del picco massimo.

4.2.3.4 Descrizione dei punti di interesse

Si possiedono a questo punto tutte le informazioni necessarie a descrivere il nostro punto di interesse. Come viene creato il descrittore (vettore)? Si considera sempre una matrice 16x16 nell'intorno del punto di interesse e viene suddivisa in 16 blocchi di 4x4.

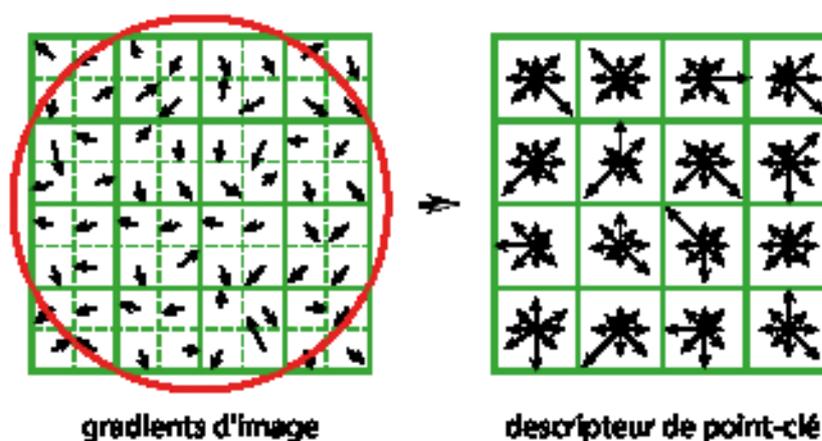


Figura 4.17. in ogni quadrante della finestra 16x16 viene calcolato un istogramma delle orientazioni a 8 bin

Per ciascun blocco viene creato un istogramma di orientamento a 8 bin, per un totale di 128 bin. Il tutto è rappresentato attraverso un vettore che formerà proprio la descrizione del punto di interesse. Oltre a questo verranno eseguite altre misure in maniera da rendere più robusta tale descrizione nei confronti dei cambi di illuminazione, rotazione, ecc. ecc. ...

Il risultato di questo insieme di procedure è riportato nelle immagini di Figura 4.18.

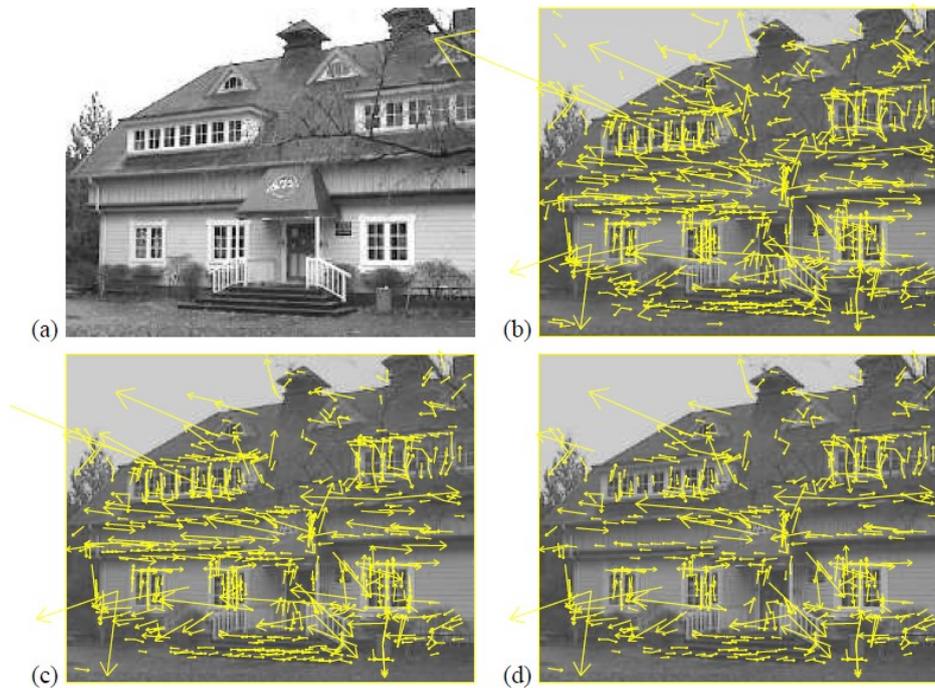


Figura 4.18. In figura sono rappresentate le fasi di selezione delle features. (a) è l'immagine originale, (b) rappresenta invece 832 punti localizzati attraverso il DoG. Sono rappresentati attraverso dei vettori che indicano la scala, l'orientamento e la posizione. (c) è il risultato dopo l'applicazione del raffinamento a contrasto. Restano solo 729 punti. (d) è invece la configurazione finale dopo l'eliminazione dei punti corrispondenti agli edges. Sono rimasti soltanto 536 punti.

4.2.3.5 Matching dei punti di interesse

I punti di interesse appartenenti ad immagini diverse sono correlati attraverso una procedura di nearest neighbours. Per ciascun descrittore si cerca il descrittore che minimizza la distanza euclidea tra i descrittori (vettori 128 bin). In alcuni casi il secondo candidato più vicino può essere molto vicino al primo, (a causa del rumore) per questo viene considerato il rapporto tra la distanza più vicina e la seconda distanza più vicina. Se tale valore è maggiore di 0,8 questi vengono rigettati. Così vengono eliminati circa il 90% di false corrispondenze e vengono scartate soltanto il 5% tra quelle corrette.

5 RICOSTRUZIONE METRICA

Nel capitolo 2 è stata presentata la geometria del problema. Per poter risolvere le equazioni di collinearità e quindi calcolare le coordinate tridimensionali dei punti oggetto partendo dai pixel immagine, è necessario conoscere i parametri interni ed esterni della nostra fotocamera.

Si consideri un insieme di punti tridimensionali P^j visti da n fotocamere, ciascuna con matrice di proiezione prospettica $\{M_i\}_{i=1\dots n}$. Siano p_i^j le coordinate pixel (omogenee) della proiezione del j -esimo punto nella i -esima fotocamera. È stato dimostrato come il problema della ricostruzione può essere rappresentato dalla relazione:

$$p_i^j \cong M_i P^j \quad (5.1)$$

Conclusa la ricerca automatica dei punti caratteristici (feature detection) numerose p_i^j sono note nelle diverse immagini, mentre restano incognite le matrici di proiezione. Si ricorda inoltre che:

$$M_i = K_i [R_i | t_i] = [K_i R_i | K_i t_i] = [Q_i | q_i] \quad (5.2)$$

Quando i parametri di orientamento interno (K_i) ed esterno ($R_i | t_i$) sono completamente incogniti, le M_i devono essere stimate. È possibile in ogni caso ottenere una ricostruzione definita a meno di una proiezione arbitraria. Si parlerà di **ricostruzione proiettiva**. Se vale la (5.1) deve essere sempre valida anche:

$$p_i^j = M_i T T^{-1} P^j \quad (5.3)$$

Essendo T una matrice (4x4) non singolare rappresentativa di una trasformazione lineare tridimensionale.

Se il problema è condizionato da alcune ipotesi (come ad esempio se almeno 3 fotocamere presentano parametri di orientamento interno costanti) sarà poi sempre possibile passare dalla ricostruzione proiettiva alla **ricostruzione euclidea** (ricostruzione metrica). Il problema non si pone quando sono noti invece i parametri di orientamento interno della camera, ad esempio a seguito di una procedura di calibrazione della stessa. Di seguito definiremo le relazioni che legano tra loro i punti caratteristici ed osserveremo i metodi per il calcolo delle matrici di proiezione prospettica in entrambi i casi, calibrato e non.

5.1 GEOMETRIA EPIPOLARE

Date due immagini che riprendono la stessa scena, le sole informazioni che si possono estrarre sono quelle relative al riconoscimento ed all'accoppiamento delle features. Esistono però alcuni vincoli tra

i punti caratteristici individuati, che ne fissano la geometria. Uno di questi è rappresentato ad esempio dalla geometria epipolare, che da sempre descrive le relazioni che intercorrono tra due immagini di una stessa scena. In particolare è noto che p' il punto coniugato di p deve giacere su di una linea retta definita **retta epipolare** di p .

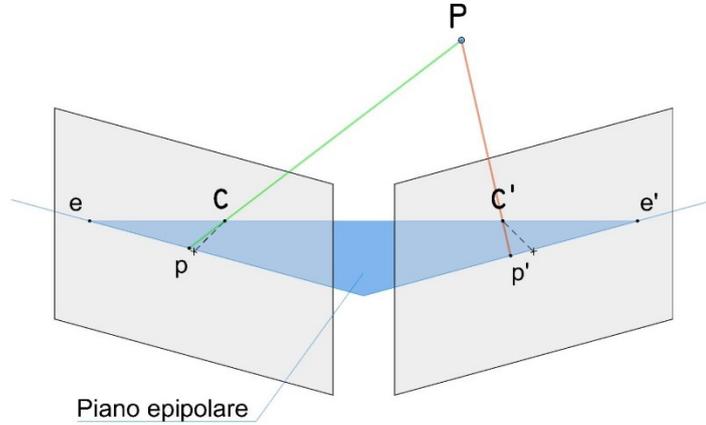


Figura 5.1. Geometria epipolare

Fissato un qualsiasi sistema di riferimento, e date le due matrici di proiezione prospettica tali che:

$$\begin{aligned} p &\cong MP \\ p' &\cong M'P \end{aligned} \quad (5.4)$$

La retta epipolare di p è di fatto la proiezione secondo M' del raggio ottico di p . È pertanto possibile esprimere la relazione tra i due punti p ed p' proprio grazie a questa definizione. Il raggio ottico di p ha equazione:

$$P = C + \lambda \begin{bmatrix} Q^{-1}p \\ 0 \end{bmatrix} \quad (5.5)$$

Proiettandolo secondo M' (moltiplicandolo per M') si ottiene:

$$M'P = M'C + M'\lambda \begin{bmatrix} Q^{-1}p \\ 0 \end{bmatrix} \quad (5.6)$$

Sostituiamo ricordando che: $C = \begin{bmatrix} \check{C} \\ 1 \end{bmatrix} = \begin{bmatrix} -Q^{-1}q \\ 1 \end{bmatrix}$ dove \check{C} rappresenta la parte cartesiana di C e che $M' = [Q'|q']$ ed osserviamo ciascun termine separatamente:

$$\underbrace{M'P}_{(i)} = \underbrace{M'C}_{(ii)} + \underbrace{M'\lambda \begin{bmatrix} Q^{-1}p \\ 0 \end{bmatrix}}_{(iii)} \quad (5.7)$$

(i) $M'P \cong p'$

$$(ii) \mathbf{M}'\mathbf{C} = \mathbf{M}' \begin{bmatrix} -\mathbf{Q}^{-1}\mathbf{q} \\ 1 \end{bmatrix} = \mathbf{q}' - \mathbf{Q}'\mathbf{Q}^{-1}\mathbf{q} = \mathbf{e}' \text{ (epipolo)}$$

$$(iii) \mathbf{M}'\lambda \begin{bmatrix} \mathbf{Q}^{-1}\mathbf{p} \\ 0 \end{bmatrix} = \lambda\mathbf{Q}'\mathbf{Q}^{-1}\mathbf{p}$$

Pertanto sostituendo:

$$\mathbf{M}'\mathbf{P} = \mathbf{M}'\mathbf{C} + \mathbf{M}'\lambda \begin{bmatrix} \mathbf{Q}^{-1}\mathbf{p} \\ 0 \end{bmatrix} \longrightarrow \mathbf{p}' \cong \lambda\mathbf{Q}'\mathbf{Q}^{-1}\mathbf{p} + \mathbf{e}' \quad (5.8)$$

Questa è l'equazione della retta epipolare di \mathbf{p} , nota anche come equazione di Longuet-Higgins e può essere scritta anche nella forma:

$$\mathbf{0} = \mathbf{p}'^T [\mathbf{e}']_x \mathbf{Q}'\mathbf{Q}^{-1}\mathbf{p} \quad (5.9)$$

Questa è una equazione molto importante perché fissa la relazione che esiste tra i punti immagine coniugati, trovati tramite feature detection. Questa relazione sarà la base per la nostra ricostruzione metrica in quanto sarà proprio grazie a questa che riusciremo a stimare le matrici di proiezione prospettica delle fotocamere.

5.2 RICOSTRUZIONE CALIBRATA

5.2.1 Orientamento relativo

Si immagini dunque di avere una camera di cui siano noti i parametri di orientamento interno $[\mathbf{K}]$, ma non quelli di orientamento esterno, e si considerino due immagini di una stessa scena acquisite con tale fotocamera. Si assuma inoltre di disporre di numerosi punti caratteristici estratti con le procedure presentate nel capitolo 4. Se si fissa il sistema di riferimento assoluto nel centro di presa di una delle due immagini le due matrici di proiezione prospettica diventano:

$$\mathbf{M} = \mathbf{K}[\mathbf{I}|\mathbf{0}]; \quad \mathbf{M}' = \mathbf{K}[\mathbf{I}|\mathbf{0}]\mathbf{G} = \mathbf{K}[\mathbf{R}|\mathbf{t}] \quad (5.10)$$

In coordinate normalizzate ($\mathbf{m}^* = \mathbf{K}^{-1}\mathbf{m}$):

$$\mathbf{M} = [\mathbf{I}|\mathbf{0}]; \quad \mathbf{M}' = [\mathbf{I}|\mathbf{0}]\mathbf{G} = [\mathbf{R}|\mathbf{t}] \quad (5.11)$$

Se si sostituiscono queste due matrici all'interno dell'equazione di Longuet-Higgins otteniamo la relazione che lega i punti coniugati in coordinate normalizzate. Essendo:

$$\mathbf{Q}' = \mathbf{R}; \quad \mathbf{Q} = \mathbf{I}; \quad \mathbf{e}' = \mathbf{P}'\mathbf{C} = [\mathbf{R}|\mathbf{t}] \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix} = \mathbf{t} \quad (5.12)$$

Si riporta l'equazione (5.9)

$$0 = p^{iT} [e']_x Q' Q^{-1} p$$

Sostituendo diventa:

$$0 = p^{iT} [t]_x R p \quad (5.13)$$

Si definisce **matrice essenziale**:

$$E = [t]_x R \quad (5.14)$$

La matrice essenziale E contiene i 3 parametri della rotazione e i due della traslazione. Il modulo di t non conta, infatti la relazione è nota a meno di un fattore di scala. Per poterla conoscere è necessario un ulteriore parametro come ad esempio la distanza tra due punti. Pertanto una matrice essenziale ha solo 5 gradi di libertà, che tengono conto della rotazione (3 parametri) e della traslazione a meno di un fattore di scala.(2 parametri).

Se i parametri interni $[K]$ e $[K']$ delle camere sono noti, attraverso E si possono ricavare le due matrici di proiezione prospettica delle immagini $[P]$ e $[P']$. Si può dunque risolvere l'orientamento relativo tra le due immagini.

5.2.1.1 Stima della matrice E

5.2.1.1.1 Criterio lineare – l'algoritmo degli 8 punti

Sia dato un insieme di corrispondenze $(p_i; p'_i)$ sufficientemente grande, in coordinate normalizzate.

Per ciascuna corrispondenza possiamo scrivere una relazione del tipo:

$$p_i'^T E p_i = 0 \quad (5.15)$$

La matrice essenziale incognita può essere calcolata grazie al *prodotto di Kronecker* e al passaggio al *vec*:

$$p_i'^T E p_i = 0 \iff (p_i^T \otimes p_i'^T) \text{vec}(E) = 0 \quad (5.16)$$

Per ogni corrispondenza trovata possiamo scrivere un'equazione omogenea lineare nei 9 elementi incogniti di E . Da n punti corrispondenti si ottiene un sistema di n equazioni lineari.

$$\underbrace{\begin{pmatrix} p_1^T \otimes p_1'^T \\ p_2^T \otimes p_2'^T \\ \dots \\ p_n^T \otimes p_n'^T \end{pmatrix}}_{U_n} \text{vec}(E) = 0 \quad (5.17)$$

La soluzione al sistema è rappresentata dal nucleo di U_n . Con $n=8$ corrispondenze il nucleo ha dimensione 1, quindi la soluzione è determinata a meno di un fattore di scala.

Nella pratica sono disponibili molte corrispondenze, pertanto $n \gg 8$. Il problema della stima di E può essere risolto ai minimi quadrati. In realtà poiché la matrice E ha soltanto 5 gradi di libertà, è possibile calcolarla a partire da sole 6 corrispondenze. Quello appena mostrato voleva essere un metodo generale, per la soluzione di sistemi a 9 parametri incogniti, tipici di matrici 3×3 come E ed F . La presenza di vincoli ovviamente riduce i gradi di libertà di tali matrici e ne semplifica la risoluzione.

5.2.1.2 Fattorizzazione di E

La matrice E così calcolata non soddisfa in generale l'equazione (5.11), ma deve essere fattorizzata nel prodotto tra una matrice di rotazione R e una traslazione t rappresentata da una matrice antisimmetrica a meno di un fattore di scala. *Huang e Faugeras nel 1989* proposero il seguente metodo:

Teorema: Una matrice reale $E(3 \times 3)$ può essere fattorizzata come prodotto di una matrice non nulla antisimmetrica e di una matrice di rotazione se e soltanto se E ha due valori singolari non nulli ed un valore singolare uguale a zero.

Sia UDV^T la SVD (decomposizione in valori singolari) di E . Dove D è una matrice diagonale 3×3 i cui elementi prendono il nome di valori singolari, U e V sono due matrici ortogonali 3×3 . In particolare si osserva che:

$$D = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} = \underbrace{\begin{bmatrix} 0 & 1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}}_{S'} \underbrace{\begin{bmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}}_{R'} \quad (5.18)$$

S' è antisimmetrica e R' rappresenta una rotazione in torno a Z di $-\pi/2$. Pertanto

$$E = UDV^T = US'R'V^T = (US'U^T)(UR'V^T) \cong \underbrace{(US'U^T)}_S \underbrace{\det(UV^T)(UR'V^T)}_R \quad (5.19)$$

$$E = SR$$

Questa fattorizzazione non è unica, in quanto è possibile trasporre sia S' che R' ottenendo in tutto 4 combinazioni. L'interpretazione geometrica di tali soluzioni è la seguente: il cambio di segno di S inverte la fotocamera destra con la sinistra mentre la scelta su R determina la rotazione della fotocamera di 180° intorno alla linea di base. In ognuno di questi casi è possibile triangolare la posizione delle corrispondenze ma soltanto in un caso questa avviene con il punto tridimensionale giacente davanti alle due fotocamere. Questa rappresenta l'unica soluzione fisica realizzabile.

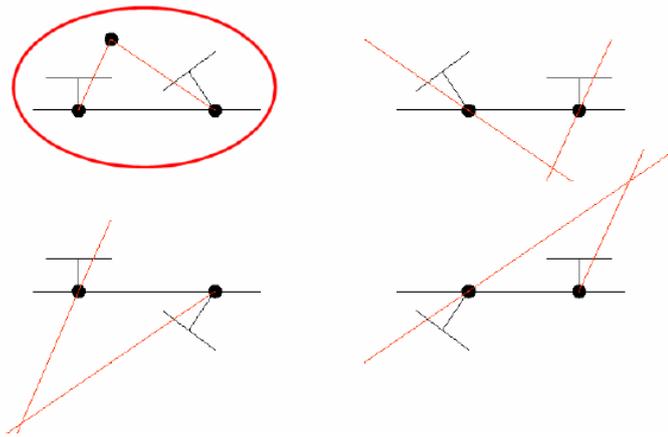


Figura 5.2. In figura sono rappresentate le 4 soluzioni della fattorizzazione di E. Solo nel caso cerchiato di rosso il punto triangolato si trova di fronte alle due fotocamere.

L'insieme delle operazioni viste fino adesso possono dunque essere semplificate dai seguenti passaggi:

- Stima della matrice essenziale E con l'algoritmo degli 8 punti
- Fattorizzazione di E in SR
- Calcolo delle due MPP: $P = [I|0]$ e $P' = [I|0]G = [R|t]$

È stato così risolto l'orientamento relativo tra i fotogrammi, avendo assunto come sistema di riferimento assoluto il centro di presa di una delle due camere.

5.2.2 Orientamento Esterno

Supponiamo di conoscere la posizione tridimensionale di alcuni punti sulla scena e la loro proiezione sul piano dell'immagine della fotocamera, di cui conosciamo i parametri di orientamento interno. Il problema dell'orientamento esterno consiste di fatto nel calcolo della matrice $[R|t]$ contenente i parametri incogniti della rototraslazione rigida tra l'origine del nostro sistema di riferimento assoluto e il centro di presa della camera. La relazione tra un punto oggetto e la sua proiezione è espressa dalla:

$$p_i \cong K[R|t]P_i \quad (5.20)$$

Essendo K nota possiamo riscriverla in coordinate normalizzate, essendo $p_i^* = K^{-1}p_i$

$$p_i^* = K^{-1}p_i \cong [R|t]P_i \quad (5.21)$$

In coordinate cartesiane si osserva come secondo la relazione appena vista ciascuna corrispondenza genera due equazioni del tipo:

$$\begin{cases} x_i = \frac{r_1^T \tilde{P}_i + t_1}{r_3^T \tilde{P}_i + t_3} \\ y_i = \frac{r_2^T \tilde{P}_i + t_2}{r_3^T \tilde{P}_i + t_3} \end{cases} \quad (5.22)$$

Dove $R = [r_1, r_2, r_3]^T$ è la matrice di rotazione e $t = [t_1, t_2, t_3]^T$ il vettore di traslazione. I parametri di R sono espressi in funzione di tre angoli di rotazione. Il sistema di equazioni che si genera è di tipo non lineare e può essere risolto soltanto se sono note le corrispondenze di almeno 3 punti (6 incognite in 6 equazioni). Nella pratica è consigliabile usare più corrispondenze e al fine di minimizzare gli errori dovuti a corrispondenze poco accurate il sistema viene risolto ai minimi quadrati. Lowe nel 1991 propose di applicare il metodo di Gauss Newton per determinare la soluzione ai minimi quadrati del sistema non lineare.

Si fissa un punto P_i e si definisce un'applicazione $f_i(a): \mathbb{R}^6 \rightarrow \mathbb{R}^2$ dallo spazio dei 6 parametri che regolano la rototraslazione ai due delle coordinate immagine. Possiamo riscrivere in forma compatta il sistema di equazioni:

$$p_i - f_i(a) = 0 \quad (5.23)$$

Dove $a = [t_x, t_y, t_z, \varphi_x, \varphi_y, \varphi_z]^T$. Si ottiene un sistema di $2n$ equazioni che si risolve ai minimi quadrati calcolando:

$$\min \left[F(a) = \frac{1}{2} \sum f_i(a)^2 \right] \quad (5.24)$$

Trovati i parametri di R e t possiamo risalire alle MPP rispetto il sistema di riferimento assoluto.

5.3 RICOSTRUZIONE NON CALIBRATA

5.3.1 La matrice Fondamentale

Fondamentalmente il caso non calibrato è simile a quello appena visto. La differenza principale sta nella stima dei parametri interni della fotocamera attraverso una procedura nota con il nome di autocalibrazione. Riportiamo per comodità la relazione già vista tra i punti corrispondenti nelle diverse immagini. L'equazione nota come equazione di Longuet-Higgins è:

$$0 = p^{iT} [e']_x Q' Q^{-1} p \quad (5.25)$$

Si definisce *matrice fondamentale*, la matrice:

$$[e']_x Q' Q^{-1} = F \quad (5.26)$$

Da cui:

$$p'^T F p = 0 \quad (5.27)$$

In F sono contenute le informazioni relative alla geometria epipolare. In principio matrici come queste (3x3) hanno 9 gradi di libertà. Abbiamo già accennato al fatto che la ricostruzione è valida anche a meno di un fattore di scala, quindi uno dei parametri della matrice può essere scelto in maniera arbitraria, inoltre si dimostra che il $\det(F) = 0$ e che il rango di F è uguale a 2. Questa relazione che può essere utilizzata come un vincolo di condizione. Ogni vincolo elimina un grado di libertà pertanto la matrice fondamentale F ha soltanto 7 gradi di libertà. Ha pertanto 7 parametri incogniti che devono essere stimati. Anche la matrice fondamentale, così come quella essenziale è una funzione delle due MPP, pertanto una volta che tale matrice è nota è sempre possibile risalire alle matrici di proiezione prospettica delle due fotocamere e da lì iniziare il processo di ricostruzione della struttura.

Esiste dunque una relazione tra le due matrici appena viste, lo si può dedurre dal fatto che entrambe contengono le informazioni riguardo al movimento rigido esistente tra le due fotocamere. La matrice essenziale E mette in relazione le coordinate normalizzate dei punti coniugati, mentre la seconda mette in relazione le coordinate pixel dei punti coniugati. Pertanto si dimostra che:

$$F = K'^{-T} E K^{-1} \quad (5.28)$$

5.3.2 Stima della matrice Fondamentale (F)

5.3.2.1 Criterio lineare – l' algoritmo degli 8 punti

Così come nel caso calibrato anche quando i parametri interni della camera sono incogniti è possibile stimare la matrice fondamentale utilizzando un approccio lineare uguale a quello proposto nel paragrafo precedente. Se F possiede 7 gradi di libertà, 7 corrispondenze sono sufficienti per la stima dei parametri incogniti. Data quindi una corrispondenza $p(x, y) \Leftrightarrow p'(x', y')$ è possibile dunque scrivere una equazione del tipo:

$$(x'x, x'y, x'x', y'x, y'y, y'y', x, y, 1) \begin{bmatrix} f_{11} \\ f_{12} \\ f_{13} \\ f_{21} \\ f_{22} \\ f_{23} \\ f_{31} \\ f_{32} \\ f_{33} \end{bmatrix} = 0 \quad (5.29)$$

Disponendo di almeno 7 punti si ottiene n sistema di equazioni del tipo $AF = 0$

$$\underbrace{\begin{pmatrix} x'_i x_i & x'_i y_i & x'_i & y'_i x_i & y'_i y_i & y'_i & x_i & y_i & 1 \\ \vdots & \vdots \\ x'_n x_n & x'_n y_n & x'_n & y'_n x_n & y'_n y_n & y'_n & x_n & y_n & 1 \end{pmatrix}}_A \begin{matrix} f_{11} \\ f_{12} \\ f_{13} \\ f_{21} \\ f_{22} \\ f_{23} \\ f_{31} \\ f_{32} \\ f_{33} \end{matrix} = 0 \quad (5.30)$$

Nella pratica si dispone di un numero di corrispondenze molto elevato $n \gg 7$, pertanto la soluzione del sistema avviene tramite una soluzione standard ai minimi quadrati. Tuttavia la F così stimata non risolve la condizione $\det(F) = 0$ pertanto è necessaria una fattorizzazione di F attraverso una decomposizione in valori singolari (SVD – Singular Value Decomposition)

Sia UDV^T la SVD di F. Dove D è una matrice diagonale 3x3 i cui elementi prendono il nome di valori singolari, U e V sono due matrici ortogonali 3x3. In particolare si osserva che:

$$F = UDV^T = U \begin{bmatrix} d_1 & 0 & 0 \\ 0 & d_2 & 0 \\ 0 & 0 & d_3 \end{bmatrix} V^T \quad (5.31)$$

Il vincolo di singolarità $\det(F) = 0$ viene imposto ponendo $d_3=0$

5.3.2.2 Raffinamento non lineare

La stima ottenuta dall'algorithmo lineare, essendo particolarmente sensibile al rumore può essere raffinata mediante la minimizzazione di un opportuno residuo geometrico, pari alla somma delle distanze tra i punti e le rette epipolari coniugate (Luong e Faugeras 1996). Si ottiene una soluzione più robusta.

La F si ottiene risolvendo il problema ai minimi quadrati non lineare:

$$\min \sum_i d(Fm_i, m'_i)^2 + d(F^T m'_i, m_i)^2 \quad (5.32)$$

Dove d è la distanza tra il punto e la retta nel piano euclideo. Ancora per garantire la singolarità è necessario parametrizzare la F combinando linearmente due righe o due colonne.

$$F = \begin{bmatrix} f_1 & f_4 & \alpha f_1 + \beta f_4 \\ f_2 & f_5 & \alpha f_2 + \beta f_5 \\ f_3 & f_6 & \alpha f_3 + \beta f_6 \end{bmatrix} \quad (5.33)$$

Sono 8 parametri e fissandone in maniera arbitraria uno per bloccare la scala, diventano 7 come i gradi di libertà. Il metodo non lineare è migliore rispetto a quello lineare degli 8 punti, tuttavia risulta computazionalmente più oneroso. Tipicamente per migliorare la precisione dei risultati ottenuti si tende a utilizzare l'approccio non lineare a valle di quello lineare.

5.4 AUTOCALIBRAZIONE

Il problema di calibrazione della camera può essere risolto con i metodi classici proposti nel capitolo 3 oppure con una procedura nota con il termine di **autocalibrazione**. Questo processo può avvenire in due modi differenti: si può partire dalla ricostruzione proiettiva e poi passare a quella euclidea oppure si possono recuperare i parametri di orientamento interni a partire dalle matrici fondamentali e ricondurre il tutto al caso calibrato.

5.4.1 Ricostruzione proiettiva

La ricostruzione che verrà affrontata nel presente paragrafo è quella relativa a m fotocamere ($M_1 \dots M_m$) che riprendono n punti dello spazio tridimensionale ($P^1 \dots P^n$). Ancora si riporta l'equazione della proiezione prospettica esplicitando questa volta il fattore di scala λ .

$$\lambda_i^j p_i^j = M_i P^j \quad (5.34)$$

In forma matriciale si traduce in:

$$\underbrace{\begin{bmatrix} \lambda_1^1 p_1^1 & \lambda_2^1 p_2^1 & \dots & \lambda_m^1 p_m^1 \\ \vdots & \vdots & \ddots & \vdots \\ \lambda_1^2 p_1^2 & \lambda_2^2 p_2^2 & \dots & \lambda_m^2 p_m^2 \\ \vdots & \vdots & \ddots & \vdots \\ \lambda_1^n p_1^n & \lambda_2^n p_2^n & \dots & \lambda_m^n p_m^n \end{bmatrix}}_W = \underbrace{\begin{bmatrix} M_1 \\ M_2 \\ \vdots \\ M_m \end{bmatrix}}_M \underbrace{\begin{bmatrix} P^1 & P^2 & \dots & P^n \end{bmatrix}}_P \quad (5.35)$$

L'unico parametro noto del sistema sono le coordinate pixel delle corrispondenze. Supponendo il fattore di scala sia noto la matrice W diviene nota ed è possibile calcolarne la SVD.

$$W = UDV^T \quad (5.36)$$

Nel caso ideale, essendo W di rango 4, la matrice diagonale contenente i valori singolari è esprimibile come $D = \text{diag}(d_1, d_2, d_3, d_4, 0, \dots, 0)$. Solo le prime 4 colonne di U e di V contribuiscono al prodotto tra matrici. Quindi:

$$W = U_{[3m \times 4]} D V_{[3n \times 4]}^T \quad (5.37)$$

Da cui:

$$M = U_{[3m \times 4]} \text{diag}(d_1, d_2, d_3, d_4)$$

$$P = V_{[3n \times 4]}^T \quad (5.38)$$

Nel caso reale il rango di W non è 4, forzando comunque $D = \text{diag}(d_1, d_2, d_3, d_4, 0, \dots, 0)$ si ottiene una soluzione minimizzando l'errore in norma:

$$\|W - MP\|_F^2 = \sum_{i,j} \|(\lambda_i^j p_i^j - \mathbf{M}_i P^j)\|^2 \quad (5.39)$$

Essendo il fattore di scala incognito la soluzione si ottiene per via iterativa, stimando dapprima la scala dati M e P (dalla fattorizzazione di W nota) e poi stimando M e P partendo dal valore della scala calcolato al passo 1.

5.4.2 Promozione euclidea

Abbiamo visto come il problema sia descritto dalla relazione:

$$p_i^j = \mathbf{M}_i P^j$$

E che in generale, con i parametri di orientamento interno (K_i) ed esterno ($R_i|t_i$) incogniti, è possibile ottenere una ricostruzione definita a meno di una proiettività arbitraria. Prende il nome di **ricostruzione proiettiva**. In generale è infatti sempre valida anche:

$$p_i^j = \mathbf{M}_i T T^{-1} P^j$$

Il problema consiste di fatto nel recuperare tale proiettività, di calcolare cioè la matrice $T(4 \times 4)$ non singolare. Per poterlo fare è necessario che siano disponibili un numero sufficiente di fotogrammi e che i parametri di orientamento interno delle camere anche se incogniti, siano costanti (si ipotizza quindi che si stia facendo uso della stessa fotocamera). Dalla ricostruzione proiettiva otteniamo un insieme di MPP tali per cui:

$$M_0^p = [I|0]; \quad M_i^p = [Q_i|q_i]. \quad (5.40)$$

Se la struttura euclidea che stiamo cercando è data da M , T deve essere tale da:

$$M_i^e \cong M_i^p T \quad (5.41)$$

Fissando in maniera arbitraria la trasformazione rigida che lega le varie fotocamere impostando per la ricostruzione euclidea: $M_0^e = K[I|0]; M_i^e = K[R_i|t_i]$

Possiamo quindi scrivere la relazione tra le MPP utilizzando le prime matrice della ricostruzione:

$$M_0^e = M_0^p T \quad (5.42)$$

Da cui :

$$T = \begin{bmatrix} K & 0 \\ r^T & s \end{bmatrix} \quad (5.43)$$

Dove $r^T = (r_1, r_2, r_3)$ è un vettore arbitrario di 3 elementi. T è non singolare e fissando il fattore di scala $s=1$ dipende da 8 parametri: i 5 della matrice K e i 3 del vettore r. sostituendo le relazioni viste per le successive matrici vale dunque:

$$\begin{aligned} M_i^e \cong M_i^p T &\implies [KR_i | Kt_i] = [Q_i K + q_i r^T | q_i] \\ KR_i &\cong Q_i K + q_i r^T \end{aligned} \quad (5.44)$$

Questa equazione lega le incognite del problema (5 parametri di K e i 3 di r) ai dati disponibili Q_i, q_i . R non si conosce ma deve essere una matrice di rotazione. Tale relazione può essere scritta anche nella forma:

$$M_i^p \begin{bmatrix} K \\ r^T \end{bmatrix} \cong KR_i \quad (5.45)$$

Dato che $R_i R_i^T = I$ possiamo scrivere:

$$\begin{aligned} M_i^p \begin{bmatrix} K \\ r^T \end{bmatrix} \begin{bmatrix} K \\ r^T \end{bmatrix}^T M_i^{pT} &\cong KR_i R_i^T K^T = KK^T \\ M_i^p \begin{bmatrix} KK^T & Kr \\ r^T K^T & r^T r \end{bmatrix} M_i^{pT} &\cong KK^T \end{aligned} \quad (5.46)$$

L'espressione contiene 5 equazioni e il fattore di scala arbitrario ne elimina una. Pertanto ogni MPP, esclusa la prima, fornisce 4 equazioni in 8 incognite. Il fattore di scala può essere inserito in maniera esplicita:

$$0 = \lambda_i^2 KK^T - M_i^p \begin{bmatrix} KK^T & Kr \\ r^T K^T & r^T r \end{bmatrix} M_i^{pT} \quad (5.47)$$

Con tre fotocamere possiamo risolvere il sistema poiché otteniamo 12 equazioni in 10 incognite (8 parametri di T e i due fattori di scala). Di nuovo il sistema nel caso reale si risolve ai minimi quadrati.

5.5 TRIANGOLAZIONE (SPARSA)

Nei paragrafi 5.2 e 5.3 è stato osservato come ottenere le matrici di proiezione prospettica sia nel caso calibrato che nel caso non calibrato. Note tali matrici, il moto (*MOTION*) della nostra fotocamera (o la posizione di n fotocamere nello spazio) al momento dell'acquisizione delle immagini è stato correttamente ricostruito. Si deve ancora però ricostruire la struttura (*STRUCTURE*) della scena

tridimensionale. In questo capitolo verrà analizzato il metodo classico utilizzato nella fase di ricostruzione metrica, noto con il termine di *triangolazione*.

5.5.1 Caso semplificato: fotocamere parallele e allineate

In questo caso si ha una disparità puramente orizzontale (figura

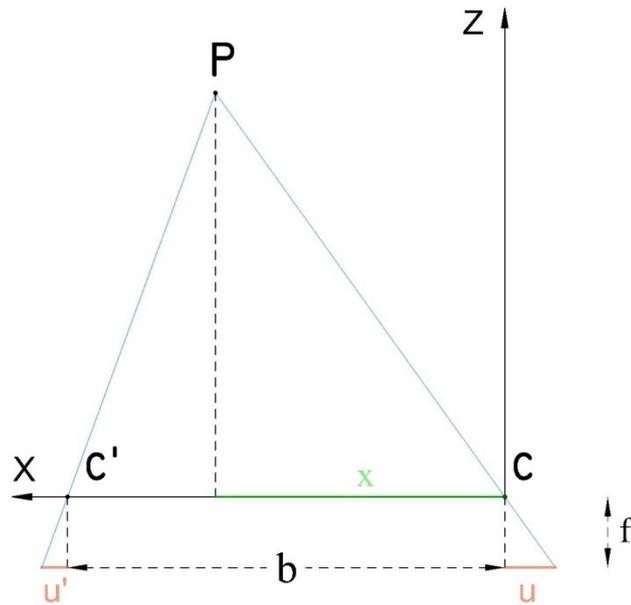


Figura 5.3. Triangolazione stereoscopica nel caso semplificato. Le camere sono parallele ed allineate. Il sistema di riferimento è centrato nella fotocamera di destra.

Fissato il sistema di riferimento mondo con la fotocamera di destra possiamo scrivere le seguenti relazioni:

$$\begin{cases} \frac{f}{Z} = \frac{-u}{X} \\ \frac{f}{Z} = \frac{-u'}{X-b} \end{cases} \quad (5.48)$$

da cui

$$\begin{cases} fX = -uZ \\ f(X-b) = -u'Z \end{cases} \quad (5.49)$$

Sostituendo fX nella seconda equazione si ottiene:

$$\begin{aligned} -uZ - fb &= -u'Z \\ Z(u - u') &= -fb \\ Z &= \frac{bf}{u' - u} \end{aligned} \quad (5.50)$$

Si definisce **disparità binoculare** ($u' - u$) la differenza tra le posizioni dei punti coniugati nelle rispettive immagini. Calcolata la disparità e nota la geometria del sistema composto dalle due immagini (b ed f) è possibile calcolare la coordinata Z . Come si può notare dalla figura la base di presa b si comporta come un afftore di scala. La triangolazione è possibile quindi anche se b è incognito, a meno di un fattore di scala.

5.5.2 Caso generalizzato

Osserviamo ora un caso più generale. Date le coordinate pixel di due punti coniugati e le due matrici di proiezione prospettica relative alle due fotocamere, si può facilmente ricostruire la posizione in coordinate assolute del punto di cui entrambi sono la proiezione:

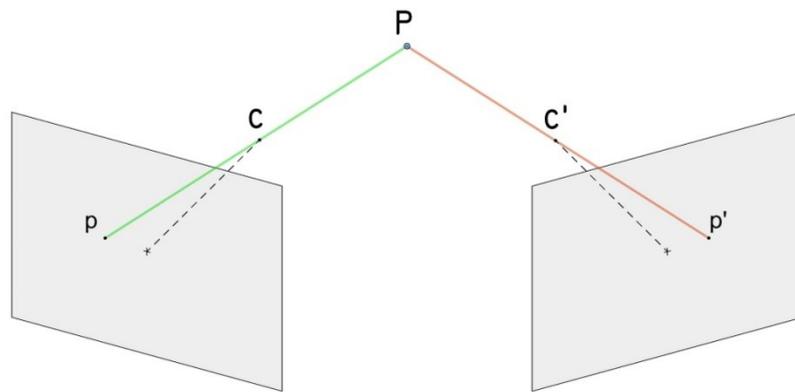


Figura 5.4. Due punti coniugati sulle immagini sono la proiezione del medesimo punto nello spazio oggetto.

Consideriamo dunque $p = [u, v, 1]^T$ e $p' = [u', v', 1]^T$ le proiezioni di P sulle due immagini, e la fotocamera di sinistra avente come MPP M . Dall'equazione della proiezione prospettica possiamo ottenere:

$$\begin{cases} u = \frac{m_1^T P}{m_3^T P} \\ v = \frac{m_2^T P}{m_3^T P} \end{cases} \quad (5.51)$$

Da cui, in forma matriciale

$$\begin{bmatrix} (m_1 - um_3)^T \\ (m_2 - vm_3)^T \end{bmatrix} P = 0 \quad (5.52)$$

Le solite due equazioni per ogni punto. Consideriamo ora il punto $p' = [u', v', 1]^T$ il punto coniugato di p sull'altra immagine di MPP pari a M' . Otteniamo le stesse equazioni viste sopra ma essendo p' la proiezione di uno stesso punto, tali equazioni si possono impilare ottenendo un sistema lineare omogeneo di quattro equazioni in quattro incognite:

$$\begin{bmatrix} (m_1 - um_3)^T \\ (m_2 - vm_3)^T \\ (m'_2 - u'm'_3)^T \\ (m'_2 - v'm'_3)^T \end{bmatrix} P = 0 \quad (5.53)$$

La soluzione va ricercata ancora una volta ai minimi quadrati con SVD come per la calibrazione. Quanto appena visto può essere generalizzato al caso di $n > 2$ fotocamere. Ogni fotocamera di fatto aggiunge 2 equazioni e si ottiene dunque un sistema omogeneo in $2n$ equazioni in 4 incognite. Come per la calibrazione anche in questo caso bisogna minimizzare un residuo di tipo geometrico del tipo:

$$\varepsilon(P) = \sum_{j=1}^m \left[\left(\frac{m_1^{jT} P}{m_3^{jT} P} - u \right)^2 + \left(\frac{m_2^{jT} P}{m_3^{jT} P} - v \right)^2 \right] \quad (5.54)$$

Ciascun termine della sommatoria è l'errore in pixel nel quadro della j -esima fotocamera. In questo esempio la triangolazione avviene per punti coniugati noti, ad esempio come nel caso dei punti caratteristici calcolati durante le fasi iniziali della elaborazione. Si chiama ricostruzione sparsa proprio perché non avviene per tutti i pixel dell'immagine ma soltanto per quelli noti. Nel prossimo paragrafo verrà preso in esame il caso della ricostruzione densa, o meglio del *dense matching*.

5.6 DENSE MATCHING

A questo punto della ricostruzione abbiamo a disposizione tutti gli elementi che ci permettono di eseguire una triangolazione dei punti pixel per pixel. Conosciamo di fatto la geometria del problema. L'orientamento esterno è stato risolto, resta da determinare la struttura della scena. Il processo di triangolazione è del tutto simile a quello visto per la ricostruzione sparsa, con la differenza che il processo deve essere applicato a ciascun pixel dell'immagine. Per poterlo fare è necessario calcolare nel modo più preciso possibile l'accoppiamento dei pixel tra due immagini, ovvero il calcolo della loro disparità (Figura 5.5).

Si ricorda che con il termine disparità binoculare si intende la differenza (intesa come vettore) tra le posizioni di due punti coniugati, immaginando di sovrapporre le due immagini.



Figura 5.5. Coppia di immagini stereo e mappa di disparità (immagini da: http://www.ces.clemson.edu/~stb/research/stereo_p2p/)

Il calcolo della disparità è dunque strettamente dipendente dal calcolo delle corrispondenze. In questa fase bisogna fare molta attenzione agli errori derivanti da accoppiamenti sbagliati. Onde evitare aberrazioni causate da outliers, è necessario introdurre dei vincoli, così da rendere il calcolo delle corrispondenze più accurato. Il vincolo più importante è stato già trattato, ed è proprio il vincolo epipolare. Oltre a questo ne esistono ovviamente altri che possono essere utilizzati durante la fase di calcolo. Qualsiasi vincolo si decida di utilizzare tipicamente ad ogni punto vengono computate più di una corrispondenza generando il problema delle false corrispondenze.

Inoltre una corretta stima dei punti coniugati deve tenere conto anche del fatto che spesso le due scene sono acquisite da punti di vista anche molto diversi tra loro. Bisogna infatti tener conto di occlusioni, e distorsioni, sia fotometriche che prospettiche. Più le camere sono distanti più ovviamente questi problemi diventano importanti. Allo stesso modo per avere un risultato significativo le due immagini devono conservare comunque una base di presa sufficientemente grande da renderle effettivamente separate tra loro.

Tipicamente gli algoritmi di matching vengono suddivisi in base alle loro caratteristiche principali. Si distinguono i **metodi locali** (considerano soltanto una finestra attorno al pixel) dai **metodi globali** (considerano l'intera linea di scansione o addirittura l'intera immagine).

DENSE MATCHING	METODI LOCALI	Area Based Matching (o Block matching)
		Gradients methods
		Feature Based Matching
	METODI GLOBALI	Decomposizione in sottoproblemi

Tabella 5.1: Suddivisione dei principali algoritmi di dense matching

5.6.1 Metodi locali

I metodi locali di correlazione possono essere molto efficienti, tuttavia essi sono notevolmente sensibili alla presenza di regioni delle immagini la cui disparità è ambigua (regioni con presenza di occlusioni o regioni con texture uniforme). E' possibile suddividere questi metodi in tre ampie classi: area based matching (o *block matching*), *gradient methods* e *feature based matching*.

5.6.1.1 Area Based Matching

I metodi area-based, cercano di stimare la disparità di un punto p nella prima immagine comparando una piccola regione presa attorno a p (frames di riferimento) con una serie di altre regioni, delle stesse dimensioni, estratte dalla seconda immagine (frames di ricerca). La finestra di ricerca è scorrevole e viene spostata sulla seconda immagine, attraverso incrementi interi, lungo la linea epipolare di p . In base al vincolo epipolare, i frames di ricerca avranno per centro un punto con coordinate di ascissa variabile, ma con ordinata fissa ed uguale al valore di ordinata per p . Il vincolo epipolare, poiché fissa il valore di ordinata, riduce la complessità della ricerca da bidimensionale a monodimensionale.

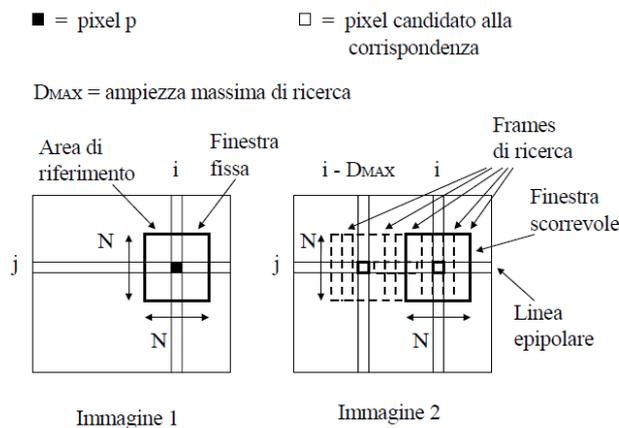


Figura 5.6. Metodi Area-Based per il calcolo delle corrispondenze

Una volta trovato il frame di ricerca che più assomiglia alla finestra di riferimento, il baricentro p' di tale frame sarà il pixel corrispondente a p . E' fondamentale stabilire in che modo si intende attribuire un grado di somiglianza per ogni singolo frame che si processa. In altre parole, deve essere definita una qualche *misura di similarità*, una funzione che riesca a stimare il grado di corrispondenza tra i

due pixel. Tipicamente vengono assunte per assolvere a tale scopo alcune funzioni dei livelli di intensità luminosa dei pixel presenti all'interno della finestra. Trovata la corrispondenza corretta si calcola la disparità tra i punti coniugati e si crea una mappa densa.

I criteri di somiglianza che tipicamente sfruttano questi algoritmi possono essere suddivisi in tre categorie:

- basati sulle differenze di intensità (SSD,SAD)
- basati su correlazione (NCC,ZNCC)
- basati su operatori di rango

Uno dei più comuni è la **SSD (Sum of Squared Differences)**

$$SSD(u, v, d) = \sum_{k,l} [I_1(u + k, v + l) - I_2(u + k + d, v + l)]^2 \quad (5.55)$$

Supponendo di avere un pixel di coordinate (u, v) ed una finestra di riferimento di dimensioni $(n \times m)$, si definisce con $I_i(u, v)$ il livello di grigio del pixel nell'immagine i -esima e con k ed l due incrementi tali che: $k \in [-n; n]; l \in [-m; m]$. Più piccolo sarà il valore della (5.55) più la similarità tra le regioni delle due immagini sarà significativa. La disparità calcolata è l'argomento del minimo della funzione errore:

$$d_0(u, v) = argmin SSD(u, v, d) \quad (5.56)$$

Vediamo un semplice esempio sul funzionamento del SSD. Si considerino due ipotetiche matrici che devono essere correlate (figura 5.7). Ogni cella (i, j) di una matrice contiene un intero che rappresenta il valore di colore per il punto (i, j) dell'immagine a cui si riferisce la matrice. Si supponga di voler trovare per il punto p , di coordinate riga-colonna $(4, 5)$, il proprio coniugato p' . Inoltre, si suppongano le dimensioni della finestra di correlazione pari a 3×3 e la massima ampiezza di ricerca pari a 2.

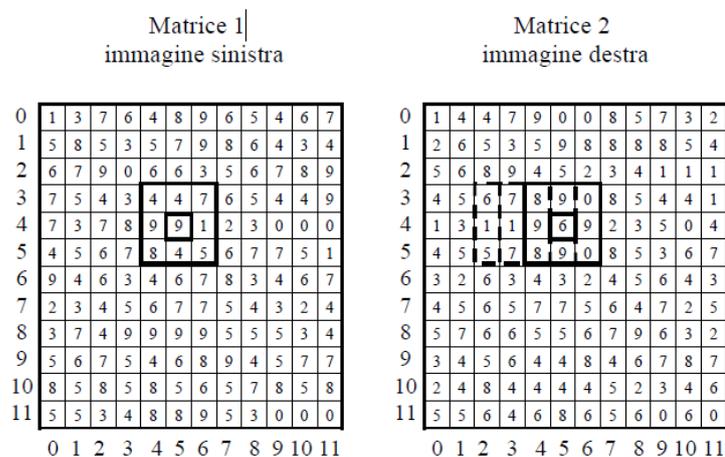


Figura 5.7. Immagini da correlare

In figura (5.8) si osservano i passaggi principali del calcolo.

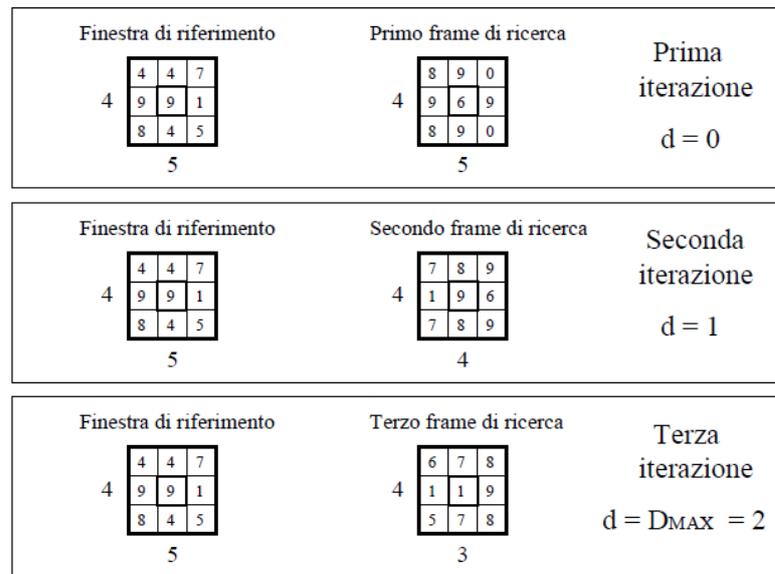


Figura 5.8. Iterazioni dell' algoritmo SSD

Il vincolo epipolare garantisce che p' abbia come ordinata il valore 4, mentre per il valore di ascissa devono essere prese in considerazione diverse ipotesi. Poiché il valore fissato per l'ampiezza di ricerca è 2, la finestra di riferimento dovrà essere confrontata con tre frames di ricerca.

Prima iterazione: (d=0)

$$S1 = (4-8)^2 + (4-9)^2 + (7-0)^2 + (9-9)^2 + (9-6)^2 + (1-9)^2 + (8-8)^2 + (4-9)^2 + (5-0)^2 = 213$$

Seconda iterazione: (d=1)

$$S2 = (4-7)^2 + (4-8)^2 + (7-9)^2 + (9-1)^2 + (9-9)^2 + (1-6)^2 + (8-7)^2 + (4-8)^2 + (5-9)^2 = 151$$

Terza iterazione: (d=2)

$$S3 = (4-6)^2 + (4-7)^2 + (7-8)^2 + (9-1)^2 + (9-1)^2 + (1-9)^2 + (8-5)^2 + (4-7)^2 + (5-8)^2 = 233$$

La disparità per p è 1. Infatti, adoperando il metodo SSD, un valore di somma nullo indica la corrispondenza perfetta, il miglior match è quello con il valore di somma più basso.

Simile al SSD è il **SAD (Sum of Absolute Differences)**, dove il quadrato viene sostituito con il valore assoluto.

$$SAD(u, v, d) = \sum_{k,l} |I_1(u+k, v+l) - I_2(u+k+d, v+l)| \quad (5.57)$$

Allo stesso modo vediamo i calcoli sulla base dell'esempio riportato in figura (5.7).

Prima iterazione: (d=0)

$$S1 = |4-8| + |4-9| + |7-0| + |9-9| + |9-6| + |1-9| + |8-8| + |4-9| + |5-0| = 37$$

Seconda iterazione: (d=1)

$$S2 = |4-7| + |4-8| + |7-9| + |9-1| + |9-9| + |1-6| + |8-7| + |4-8| + |5-9| = 31$$

Terza iterazione: (d=2)

$$S3 = |4 - 6| + |4 - 7| + |7 - 8| + |9 - 1| + |9 - 1| + |1 - 9| + |8 - 5| + |4 - 7| + |5 - 8| = 39$$

Anche adoperando il SAD il secondo frames risulta essere il più somigliante alla finestra di riferimento; si può concludere affermando che il coniugato di $p \equiv (4, 5)$ è $p' \equiv (4, 4)$.

Un altro algoritmo utilizzato frequentemente è la **NCC (Normalized Cross Correlation)**. è una misura di similarità (è dunque da massimizzare). La si può vedere come il prodotto scalare delle due finestre diviso per il prodotto scalare delle norme:

$$NCC(u, v, d) = \frac{\sum_{(k,l)} I_1(u+k, v+l) I_2(u+k+d, v+l)}{\sqrt{\sum_{(k,l)} I_1(u+k, v+l)^2} \sqrt{\sum_{(k,l)} I_2(u+k+d, v+l)^2}} \quad (5.58)$$

Per ottenere invarianza a cambi di luminosità si può sottrarre a ciascun pixel la media della finestra, ottenendo un altro algoritmo noto con il nome di **ZNCC (Zero-mean Normalized Cross Correlation)**.

Uno dei problemi principali di questi algoritmi è rappresentato dal fatto che se le finestre sono ampie e coprono regioni dell'immagine in cui la profondità dell'oggetto varia significativamente, la disparità calcolata sarà inevitabilmente affetta da errori, sarà poco accurata anche se la stima della miglior somiglianza è molto affidabile. Al contrario finestre molto piccole, danno origine a valori di disparità molto accurati ma poco affidabili, in quanto le variazioni di intensità non sono sufficienti a valutare con accuratezza la miglior somiglianza. In sostanza, valori di disparità accurati si ottengono da finestre piccole, valori affidabili da finestre grandi. Il problema può essere risolto applicando metodi che operano a diverse risoluzioni. L'idea che sta alla base è che finestre grandi forniscono un risultato inaccurato ma affidabile, mentre finestre più piccole ne migliorano l'accuratezza. La disparità ottenuta ad una risoluzione minore verrà usata come centro per l'intervallo al livello di risoluzione superiore (*coarse to fine*).

In alternativa alla correlazione, sono stati proposti metodi per il calcolo della corrispondenza che sfruttano l'applicazione alle immagini, prima del matching, di trasformazioni locali non parametriche (*operatori di rango*). Nel tentativo di eliminare la sensibilità al rumore, una trasformata legata al rango (rank transform) è localmente applicata alle regioni di entrambe le immagini. La rank transform per una regione R che circonda un pixel p è definita come il numero di pixels in R per i quali l'intensità è minore di quella di p . I valori risultanti sono basati sull'ordine relativo dei pixels in R anziché sul valore assoluto delle intensità.

Regione	89	63	72	p = 55 Rank Transform = 2
R	67	55	64	
	58	51	49	

Regione	22	24	34	p = 30 Rank Transform = 4
R	25	30	35	
	45	39	29	

Figura 5.9. Esempi di trasformazioni legate al rango.

Dopo aver applicato la rank transform, si procede adoperando la correlazione, tipicamente facendo uso della SAD. Se è vero che l'uso della trasformata legata al rango riduce la sensibilità ai rumori, è anche vero che essa riduce un po' la potenza computazionale dell'algoritmo di corrispondenza poiché fa perdere parte dell'informazione. Per limitare la perdita d'informazione, è stata proposta una trasformata che possa fornire dati anche sulla posizione relativa dei pixels nella regione di interesse, una sorta di censimento di tali pixels. La trasformata che effettua questo censimento (detta proprio *census transform*) (Zabih e Woodfill, 1994) conserva l'informazione sulla distribuzione spaziale dei pixels producendo una stringa di numeri binari. L'operatore di confronto è il seguente:

$$\xi(I, p, p') = \begin{cases} 1 & \text{se } I(p) > I(p') \\ 0 & \text{altrimenti} \end{cases} \quad (5.59)$$

Regione	89	63	72	p = 55 Census Transform = 0000011
R	67	55	64	
	58	51	49	

Regione	22	24	34	p = 30 Census Transform = 11010001
R	25	30	35	
	45	39	29	

Figura 5.10. Esempi di trasformazione di census.

La trasformata di census riassume la struttura spaziale locale. La correlazione è calcolata valutando la differenza tra stringhe di bit, utilizzando la *distanza di hamming*. Usando la stessa notazione dei casi precedenti si scrive *SCH (Sum of Census hamming distances)* come:

$$SCH(u, v, d) = \sum_{k,l} C[I_1(u+k, v+l)] \ominus C[I_2(u+k+d, v+l)] \quad (5.60)$$

Il simbolo \ominus indica la distanza di hamming tra le stringhe di bit.

5.6.1.2 Gradient method

I metodi basati sul calcolo di un gradiente cercano di determinare le disparità tra due immagini formulando una equazione differenziale che collega la disparità alla variazione della luminosità dell'immagine. La luminosità di un punto della scena viene assunta essere costante in tutte e due le immagini, in questo modo si può calcolare la traslazione orizzontale (disparità) di un punto da una immagine all'altra risolvendo l'equazione differenziale:

$$(\nabla_x E) v + Et = 0 \quad (5.61)$$

dove $\nabla_x E$ denota la componente orizzontale del gradiente dell'immagine, Et la derivata temporale (qui si riferisce alla differenza di intensità tra l'immagine sinistra e destra) e v la traslazione tra le due immagini.

5.6.1.3 Feature Based Method

Come accennato in precedenza i metodi *area based e gradient* sono molto sensibili alle discontinuità di profondità, giacché una finestra di ricerca può contenere punti con diversi valori di profondità. Questi metodi, inoltre, sono molto sensibili alle regioni delle immagini in cui la texture sia molto uniforme. I metodi feature-based cercano di oltrepassare questi problemi limitando il calcolo della corrispondenza solo ad alcune zone della scena, in cui il calcolo della disparità sia molto attendibile.

Queste zone (*features*) sono gli angoli, gli spigoli ed i bordi degli oggetti presenti nella scena. Certamente, questo modo di procedere non permette di realizzare mappe di disparità dense, perché solo una parte dei pixels delle due immagini è coinvolta nel processo di individuazione delle disparità; d'altro canto, il feature matching è estremamente efficiente sotto l'aspetto del tempo di elaborazione.

Come si può intuire, alcuni degli algoritmi che eseguono questo tipo di correlazioni sono quelli presentati nel capitolo 4 e pertanto non verranno trattati in questa sezione.

5.6.2 Metodi globali

I metodi globali sono molto meno sensibili rispetto ai metodi locali alle regioni che presentano disparità ambigua (occlusioni e regioni a texture uniforme) poiché i vincoli globali che essi sfruttano garantiscono un supporto addizionale per la gestione delle regioni difficilmente trattabili con metodi

locali. L'uso di questi vincoli globali, però, fa diventare la complessità computazionale dei metodi globali significativamente più elevata rispetto a quella dei metodi locali.

I metodi globali possono essere compresi facendo riferimento alla **DSI (Disparity Space Image)**, un'immagine tridimensionale in cui sono riportati i valori della metrica di accoppiamento tra i pixel delle due singole immagini. La mappa di disparità prodotta può essere vista come una superficie.

Il metodo globale migliore è quello proposto da Roy e Cox, 1998 e Kolmogorov e Zabih, 2001, basato sul taglio di un grafo. Per una miglior comprensione si rimanda alla lettura degli articoli di riferimento. Dal momento che per la seguente trattazione non verranno utilizzati metodi di questo tipo non saranno ulteriormente approfonditi.

6 RICOSTRUZIONE in PRATICA

Nei capitoli precedenti ho cercato di raccogliere le nozioni basilari per poter meglio comprendere il caso più generale possibile, ovvero la ricostruzione multivista non calibrata (ricostruzione proiettiva). Le formule analitiche riportate nel capitolo 5, fanno uso di una tecnica di fattorizzazione di matrici. Tale metodologia seppur molto valida, risulta complicata nella pratica, e spesso ad essa si preferisce un approccio sequenziale o gerarchico. I moderni software di restituzione operano proprio in questa direzione, basti pensare al fatto che spesso si trovano a dover lavorare con moltissime immagini, magari prese con fotocamere differenti e da punti di vista molto diversi tra loro. Ci ritroviamo dunque nel caso in cui le uniche informazioni che siamo in grado di reperire dalle immagini siano le corrispondenze dei punti caratteristici. Come si procede per la ricostruzione tridimensionale della scena? Di seguito verranno trattate le fasi principali di entrambi gli approcci. Di nuovo non sarà una trattazione esaustiva, ma utile a capirne il funzionamento generale.

6.1 APPROCCIO SEQUENZIALE

Partiamo dall'assunzione che per ciascuna coppia di immagini sia stato trovato un insieme sufficientemente ampio di features corrispondenti. L'idea di questo tipo di approccio è quella di partire da una ricostruzione iniziale di due immagini, per poi aggiungere via via un'immagine e la ricostruzione di punti che essa comporta.

Consideriamo per generalizzare il tutto il solo caso non calibrato. Per il caso calibrato sono comunque valide tutte le assunzioni che saranno prese in considerazione.

Il primo passo consiste nel calcolare le relazioni agenti tra le immagini accoppiate, ovvero nel recuperare la matrice fondamentale F partendo da un insieme di punti corrispondenti. Ovviamente non tutte le corrispondenze trovate sono di fatto corrette e prima di procedere con la stima della geometria di presa è importante eliminare le anomalie. Per fare ciò bisogna effettuare una stima robusta della geometria epipolare e un affinamento non lineare per ottenere una stima ottimale.

6.1.1 Filtraggio outliers (RANSAC)

Le procedure di calcolo indicate nel capitolo 7 sono valide se e soltanto se le corrispondenze trovate sono corrette. Se nel sistema lineare è presente una corrispondenza sbagliata, la matrice fondamentale calcolata sarà distorta e pertanto inutilizzabile. È quindi importante implementare nel processo di ricostruzione un algoritmo che sia in grado di individuare ed eliminare questo tipo di dato. L'algoritmo più utilizzato si chiama RANSAC (RANdom SAmple Consensus)(Fisher e Bolles, 1981; Meer et al., 1991). Gli algoritmi di questo vengono definiti stimatori, ovviamente ne esistono

numerosi ma l'idea che sta alla base del RANSAC è quella che tipicamente viene utilizzata all'interno dei processi di visione computazionale.

È un metodo iterativo per la stima dei parametri di un modello matematico a partire da un insieme di dati di input contenente una grande percentuale di valori outliers. È un algoritmo non deterministico, cioè produce un risultato corretto solo con una data probabilità, che aumenta al crescere delle iterazioni. Come sottolineato da Fischler e Bolles, la differenza sostanziale rispetto alle convenzionali tecniche di stima è che RANSAC genera delle soluzioni candidate utilizzando il minimo numero di osservazioni (dati di input) necessarie per stimare i parametri del modello. Infatti, mentre le classiche tecniche utilizzano la maggior quantità di dati possibile per ottenere una soluzione e solo successivamente procedono a eliminare i valori anomali, RANSAC utilizza il più piccolo insieme di dati possibile per determinare il modello e procede nell'ingrandire questo insieme con i punti di dati coerenti con il modello stimato. Di seguito si osserva a grandi linee il suo funzionamento.

Le corrispondenze trovate saranno dunque l'insieme N dei dati iniziali. Sarà un insieme del tipo:

$$D = (d_1, \dots, d_n) \text{ con } d_i \in \mathbb{R}^d \quad (6.1)$$

Si definisce **MSS (Minimal Sample Set)** l'insieme minimo s dei k campioni iniziali che servono a costruire un modello M (nel nostro caso che servono a calcolare la matrice fondamentale. Utilizzando il metodo lineare s sarà un insieme di 8 valori).

$$s = (d_{i_1}, \dots, d_{i_k}) \text{ con } d_{ii} \in D \quad (6.2)$$

Da questo insieme è possibile calcolare tutti i parametri del modello prescelto. Ad esempio per quanto riguarda la regressione lineare in \mathbb{R}^2 , il modello stimato sarà un modello lineare e quindi la dimensione dell'MSS sarà pari a 2. Infatti sono necessari e sufficienti due punti distinti per poter tracciare un'unica retta di \mathbb{R}^2 che passa per questi due punti.

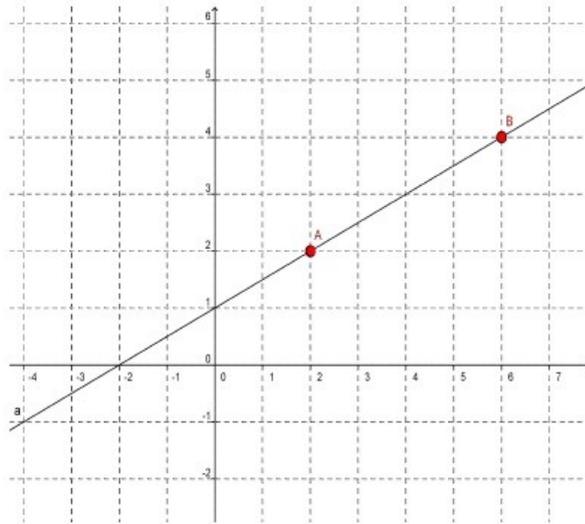


Figura 6.1. Regressione lineare in \mathbb{R}^2

Sia $\theta\{(d_{i1}, \dots, d_{ik})\}$ il vettore di parametri calcolati a partire dai punti appartenenti all'MSS. Il **modello M** è definito come:

$$M(\theta) \stackrel{\text{def}}{=} \{d \in \mathbb{R}^d : f_M(d; \theta) = 0\} \quad (6.3)$$

dove θ è il vettore dei parametri e f_M è una funzione che si azzera in tutti i punti che verificano il modello.

Si definisce **errore** associato ai punti d rispetto al modello assunto come la distanza tra i punti d e il modello $M(\theta)$:

$$e_{(M)}(d; \theta) = \min_{d^* \in M(\theta)} (\text{dist}(d, d^*) = \text{dist}(d, d^\perp)) \quad (6.4)$$

Dove d^* sono i punti che fanno parte dello spazio modello, e d^\perp è la proiezione ortogonale di d sullo spazio del modello. Infine si definisce **CS (Consensus Set)** l'insieme dei dati iniziali consistenti al modello calcolato secondo l'insieme MSS. Il CS di fatto contiene i dati inliers rispetto la regressione calcolata entro una certa soglia δ ed è:

$$s(\theta) \stackrel{\text{def}}{=} \{d \in \mathbb{R}^d : e_M(d; \theta) < \delta\} \quad (6.5)$$

Il numero di iterazioni deve essere scelto abbastanza alto da garantire che il modello calcolato alla fine delle iterazioni abbia probabilità p di non contenere alcun outlier. Se u rappresenta la probabilità che un dato sia un inlier, e quindi $v = 1 - u$ sia la probabilità che sia un outlier, allora il numero di iterazioni richieste h per il numero minimo di dati m da considerare nel MSS che risolve l'equazione:

$$1 - p = (1 - u^m)^h \quad (6.6)$$

quantità che tende a zero se h tende all'infinito. Se ipotizziamo che la probabilità di avere almeno un outlier $1 - p$ sia minore o uguale ad una soglia $1 - p \leq \varepsilon$ allora il numero di iterazioni deve garantire che, invertendo l'equazione di cui sopra:

$$h \geq \frac{\log(\varepsilon)}{\log(1 - (1 - v)^m)} \quad (6.7)$$

In pratica si definisce un insieme di osservazioni:

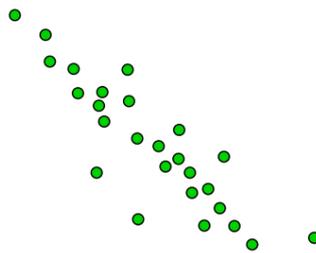


Figura 6.2. Insieme di osservazioni

di definisce un modello come ad esempio la regressione lineare in \mathbb{R}^2 e si inizia con le iterazioni prendendo a caso un campione di osservazioni che soddisfi il modello e calcolando per ogni iterazione il numero di osservazioni che verificano il modello scelto. Si calcola cioè il consenso dei dati.

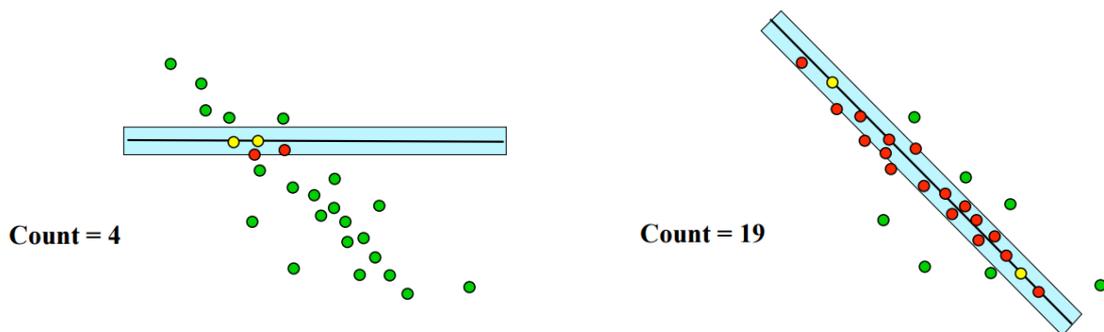


Figura 6.3. esempi di regressioni lineari passanti per un campione di punti che soddisfano un determinato modello. A sinistra il modello scelto ha un consenso pari a 4. A destra il consenso è pari a 19. Il modello selezionato sarà il secondo.

Al termine delle iterazioni si sceglie il modello che ha generato il maggior numero di consensi come quello corretto e si eliminano tutti i dati che non rispettano il modello scelto, cioè gli outliers.

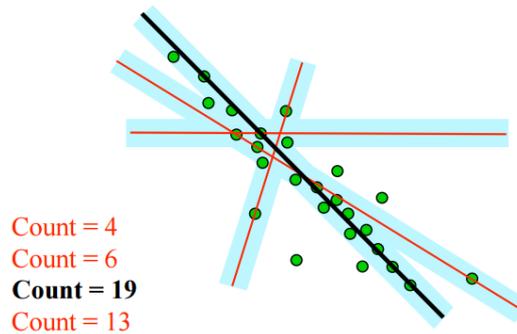


Figura 6.4: Confronto tra i modelli. Si sceglie il modello che ha generato il maggior numero di consensi.

Questa metodologia può essere implementata nel processo di ricostruzione nel seguente modo. In ogni iterazione l'algoritmo calcola una matrice fondamentale su un campione di 8 corrispondenze (caso lineare). Viene calcolato il consenso sulla base di quanti punti tra i rimanenti verificano il vincolo epipolare dato dalla matrice appena calcolata. La soluzione che presenta il consenso maggiore sarà quella che verifica meglio il vincolo epipolare. Le corrispondenze che non verificano tale vincolo saranno scartate.

Così come il RANSAC anche il vincolo epipolare è soltanto un esempio, di modelli da utilizzare nell'algoritmo ce ne sono diversi ed eventualmente anche più rigidi come il vincolo trifocale (Hartley, 1997).

Al termine della procedura sono stati ottenuti due risultati importanti. Le corrispondenze errate sono state eliminate e le matrici fondamentali calcolate sulla base delle corrispondenze corrette sono affidabili. Si può procedere con la ricostruzione del MOTION.

6.1.2 Ricostruzione incrementale

Come accennato nei paragrafi precedenti la ricostruzione procede in maniera sequenziale per differenti step. Al passo iniziale vengono scelte due immagini con un elevato numero di corrispondenze e con un'adeguata base di presa (immagini troppo vicine danno origine a problemi mal condizionati). Si calcolano le MPP delle due immagini e si fissa un sistema di riferimento arbitrario centrato rispetto il centro di presa di una delle due. Il caso è quello già osservato dell'orientamento relativo: Le due MPP saranno del tipo:

$$\mathbf{M} = \mathbf{K}[\mathbf{I}|\mathbf{0}]; \quad \mathbf{M}' = \mathbf{K}[\mathbf{I}|\mathbf{0}]\mathbf{G} = \mathbf{K}[\mathbf{R}|\mathbf{t}] \quad (6.8)$$

A questo punto per le due immagini scelte non resta altro che procedere alla triangolazione delle corrispondenze in quanto anche l'orientamento è stato risolto. Si conclude così il passo iniziale con il calcolo delle coordinate 3D dei punti associati alle features (ricostruzione sparsa) e si entra nel ciclo. Al passo successivo viene scelta un'altra immagine da aggiungere alla ricostruzione appena portata a

termine. Sulla base delle corrispondenze dei punti già ricostruiti viene stimata la *MPP* rispetto al sistema di riferimento esistente (è la procedura di orientamento esterno, svolta utilizzando le coordinate 3D dei punti nel sistema di riferimento relativo). Nota la *MPP* nuova viene ri-triangolata la posizione dei punti corrispondenti già calcolati al passo 1 utilizzando come vincoli anche le proiezione degli stessi punti nell'immagine nuova. Vengono poi aggiunti sempre per triangolazione anche i punti nuovi. Il ciclo procede in questo modo finché non sono state utilizzate tutte le immagini a disposizione.

Si osservi come l'uso delle matrici fondamentali (o essenziali se il caso è calibrato) avviene con questo approccio solo per le due immagini iniziali. Per tutte le altre coppie le *E* ed *F* calcolate sono utilizzate solo con lo scopo di validarne gli accoppiamenti, ma la stima delle *MPP* è fatta invertendo l'equazione di proiezione prospettica in quanto si utilizzano le coordinate triangolate dei punti stimate nei passaggi precedenti.

6.1.3 Bundle adjustment

Se il numero delle fotocamere è elevato, l'approccio sequenziale può accumulare errori che determinano una deriva del risultato. Basti pensare che errori sui parametri stimati nelle prime fotocamere si ripercuotono ed influenzano anche i punti che vengono aggiunti successivamente. Per ovviare a tale problema si deve minimizzare l'errore globale di riproiezione di punti, con una procedura nota con il termine di bundle adjustment. Si cerca di aggiustare sia la posizione delle *N* fotocamere che gli *n* punti 3D affinché la somma delle distanze al quadrato tra il punto *j*-esimo riproiettato tramite la fotocamera *i*-esima $\mathbf{M}_i P^j$ e il punto misurato m_i^j sia la più piccola possibile.

$$\min_{R_i, t_i, M^j} \sum_{i=1}^N \sum_{j=1}^n d(\mathbf{K}_i [\mathbf{R}_i | \mathbf{t}_i] \mathbf{M}^j, m_i^j)^2 \quad (6.9)$$

Questa minimizzazione equivale alla soluzione ai minimi quadrati di un sistema di equazioni non lineare. In questo modo eseguendo il bundle adjustment in ogni step della ricostruzione ci assicuriamo che le immagini aggiunte via via nei passi successivi siano posizionate con la massima precisione. La procedura può essere implementata anche dopo ogni singolo step. Di fatto conosciamo la posizione 3D del punto calcolata al passo precedente. Conosciamo la sua proiezione sull'immagine che abbiamo aggiunto, possiamo calcolare la sua riproiezione sulla base della *MPP* stimata e delle coordinate 3D calcolate. Stiamo quindi confrontando la distanza tra due proiezioni in un sistema di riferimento relativo centrato nell'immagine iniziale.

6.1.4 Riepilogo

La procedura di ricostruzione vista può essere quindi riassunta secondo il seguente schema:

DATI INIZIALI	
INPUT	Coordinate dei punti corrispondenti nelle singole immagini (m_i)
OUTPUT	Structure (coordinate 3D dei punti (M^i) nel sistema di riferimento relativo) from Motion (posizione delle camere ($K_i[R_i t_i]$) rispetto il sistema di riferimento relativo)
PRINCIPALI PASSAGGI DEL PROCESSO DI RICOSTRUZIONE	
PASSO 1	Calcolo delle corrispondenze con opportuni algoritmi di feature detection
PASSO 2	Validazione delle corrispondenze e stima robusta (RANSAC) delle matrici fondamentali (essenziali).
PASSO 3	Inizializzazione del processo di ricostruzione: <ol style="list-style-type: none"> 1. Selezione di 2 immagini iniziali; 2. Calcolo dell'orientamento relativo tra le due immagini (calcolo MPP); 3. Calcolo della struttura attraverso triangolazione delle corrispondenze.
PASSO 4	Per ciascuna immagine successiva: <ol style="list-style-type: none"> 1. Stima delle corrispondenze con i punti 3D ricostruiti; 2. Calcolo della MPP risolvendo il processo di orientamento esterno sulla base dei punti 3D già triangolati al passaggio precedente; 3. Ri-triangolazione dei punti e aggiunta di punti nuovi; 4. Raffinamento della costruzione attraverso bundle adjustment.

Tabella 6.1:i principali passaggi del processo di ricostruzione.

Il processo di ricostruzione metrico non è però completo. Ci troviamo ancora in un sistema di riferimento relativo del tutto arbitrario e la ricostruzione della struttura è svolta soltanto per le corrispondenze trovate non per tutti i pixel dell'immagine. Inoltre se non conosciamo i parametri interni della nostra camera la ricostruzione eseguita è di tipo proiettivo. Per poter scalare correttamente il nostro problema e ottenere la ricostruzione metrica (euclidea) della struttura della scena dobbiamo eseguire la procedura di autocalibrazione (se necessaria) e risolvere l'orientamento esterno secondo il sistema di riferimento assoluto, utilizzando alcuni punti di coordinate note direttamente misurati in campagna (Ground Control Points). Il processo può essere svolto direttamente su l'intero blocco fotogrammetrico sempre attraverso procedure di bundle adjustment, in cui verrà valutata la distanza tra i punti misurati sulle immagini e la loro riproiezione questa volta secondo le coordinate 3D ricavate dai GCP.

A questo punto non resta altro che istanziare la procedura di dense matching vista nel capitolo 9. L'intera struttura del problema sarà dunque completa. L'intero processo di ricostruzione diventa:

DATI INIZIALI	
INPUT	Coordinate dei punti corrispondenti nelle singole immagini (m_i)
OUTPUT	Structure (coordinate 3D dei punti (M^i) nel sistema di riferimento assoluto) from Motion (posizione delle camere ($K_i[R_i t_i]$) rispetto il sistema di riferimento assoluto)
PRINCIPALI PASSAGGI DEL PROCESSO DI RICOSTRUZIONE	
PASSO 1	Calcolo delle corrispondenze con opportuni algoritmi di feature detection
PASSO 2	Validazione delle corrispondenze e stima robusta (RANSAC) delle matrici fondamentali (essenziali).
PASSO 3	Inizializzazione del processo di ricostruzione: <ul style="list-style-type: none"> 4. Selezione di 2 immagini iniziali; 5. Calcolo dell'orientamento relativo tra le due immagini (calcolo MPP); 6. Calcolo della struttura attraverso triangolazione delle corrispondenze.
PASSO 4	Per ciascuna immagine successiva: <ul style="list-style-type: none"> 5. Stima delle corrispondenze con i punti 3D ricostruiti; 6. Calcolo della MPP risolvendo il processo di orientamento esterno sulla base dei punti 3D già triangolati al passaggio precedente; 7. Ri-triangolazione dei punti e aggiunta di punti nuovi; 8. Raffinamento della costruzione attraverso bundle adjustment.
PASSO 5	Auto-calibrazione (se necessaria)
PASSO 6	Promozione euclidea (se necessaria). Ricostruzione metrica della struttura.
PASSO 7	Bundle adjustment sulla base delle coordinate 3D dei GCP
PASSO 8	Ricostruzione densa della struttura attraverso algoritmi di dense matching

Tabella 6.2. Processo di ricostruzione completo in tutte le sue fasi.

PARTE SECONDA

-CASI STUDIO-

7 CASI STUDIO

7.1 INTRODUZIONE

In questa parte della tesi si vogliono mettere in evidenza tutti quegli aspetti che coinvolgono la sfera esecutiva del rilievo fotogrammetrico e che risultano strettamente legati alla restituzione multivista non calibrata. Verranno presentati due casi di studio: il primo, *la precisione nel rilievo integrato dei beni culturali*, tratterà la tematica della restituzione multivista, multiscala e multisensore non calibrata; il secondo, *il potenziale di tecnologie low-cost applicato ai beni culturali*, cercherà di studiare l'applicazione di sistemi a basso costo per l'acquisizione e la restituzione nel rilievo fotogrammetrico dei beni culturali.

In particolare, si cerca di valutare la correttezza e l'efficacia di alcune scelte progettuali in funzione della precisione e dell'accuratezza del modello tridimensionale che proprio tali scelte hanno contribuito a generare. Che attraverso l'uso di immagini si possa oggi ricostruire con elevata confidenza un qualsiasi oggetto appartenente alla scena ripresa è ormai un fatto assodato, ma con quale affidabilità e soprattutto con quale precisione possiamo eseguire oggi rilievi di questo tipo? La domanda che sempre ricorre spontanea riguarda gli errori; il modello creato può essere considerato una buona approssimazione dell'oggetto reale? In caso affermativo come possiamo utilizzare il rilievo eseguito?

Saranno queste le domande principali a cui tenteremo di dare risposta attraverso la sperimentazione avviata. Ancora una volta non sarà una sperimentazione esaustiva, le tipologie di oggetto rilevabile, i sensori utilizzati per il rilievo, le tecniche di rilievo (rilievo terrestre, utilizzo di RPAS ...) sono talmente varie che una trattazione generalizzata risulta quasi impossibile. Tuttavia esistono alcune procedure ricorrenti durante la fase di rilievo che indipendentemente dai fattori sopraelencati contribuiscono alla buona riuscita del rilevamento e dunque al raggiungimento della precisione ricercata.

Tipicamente un rilievo fotogrammetrico standard si può suddividere nelle seguenti fasi:

- Pianificazione del rilievo
- Acquisizione del dato
- Rilievo dei punti fotografici di appoggio (PFA o GCP)
- Restituzione (creazione di nubi dense o modelli texturizzati)
- Elaborazione dei dati (confronti metrici o vettorializzazione prospetti)

Eventualmente può essere utile come fase preliminare prevedere:

- Calibrazione della fotocamera

I moderni software di restituzione come abbiamo visto nella prima parte della tesi eseguono una procedura di autocalibrazione sulla base delle corrispondenze trovate con gli algoritmi di correlazione, ma se si dispone di una self-calibration, questa può essere utilizzata come punto di partenza per un'ottimizzazione dell'intero processo. Sarà poi il software a determinare le correzioni necessarie ai parametri interni della camera in maniera tale da minimizzarne gli errori.

Nel proseguo osserveremo in dettaglio le diverse fasi con particolare attenzione alla fase di elaborazione del dato, durante la quale verrà stimata la precisione metrica del rilievo. Sulla base dei risultati ottenuti saranno poi elaborate alcune semplici linee guida che possono servire da indicazioni pratiche per tutti coloro che si apprestano ad eseguire rilievi di questo tipo, atte a garantire una determinata soglia di precisione.

Dal momento che entrambi i casi di studio si sono svolti sullo stesso test-site prima di procedere con la presentazione dei singoli casi è opportuno presentare adeguatamente l'oggetto dei nostri rilievi. Questa è una sorta di fase preliminare, necessaria se si vogliono operare confronti di tipo metrico.

7.2 PREDISPOSIZIONE DEL TEST-SITE

Entrambi i casi di studio che verranno proposti riguardano il rilievo del Castello del Verginese (fig. 7.1.), un edificio situato a Gambulaga, un piccolo paese del Comune di Portomaggiore (Ferrara). Riconosciuto patrimonio dell'Umanità da parte dell'Unesco fu una delle 19 residenze degli Estensi, note con il nome di Delizie.



Figura 7.1. Vista frontale della "Delizia Estense Del Verginese"

Ad oggi non vi sono notizie certe sulla cronologia esatta della fondazione della struttura, dato che le prime testimonianze documentarie si riferiscono all'ultimo trentennio del Quattrocento; tra il 1485 ed il 1493. Si tratta di una villa rettangolare a due piani delimitata ai vertici da quattro torri merlate.

Costruita in mattoni intonacati di bianco. È collegata tramite un portico settecentesco ad una cappella privata. Nella parte posteriore c'è un brolo (un giardino rinascimentale) e una torre colombaia (presente nel ferrarese, in diverse ville signorili). È stato scelto come test site per la sua tipologia edilizia e per la sua locazione: rispecchia a tutti gli effetti i canoni dell'edificio storico definibile "bene culturale" ed è situato in aperta campagna, aspetto importante considerato che tra le tecniche di acquisizione fotogrammetrica è previsto l'utilizzo di RPAS (Remotely Piloted Aircraft Systems).

Per ottenere un completo rilievo dell'edificio sono state acquisite dunque sia immagini terrestri che aeree. L'integrazione tra le diverse tecniche ha contribuito ad una restituzione completa anche della copertura dell'edificio. Le caratteristiche della strumentazione utilizzata verranno descritte in dettaglio nei capitoli successivi. Al momento si voleva soltanto offrire una panoramica generale sull'inquadramento del test site e sulle metodologie del rilievo.

7.2.1 Il modello di riferimento

Per poter avere una stima corretta dell'accuratezza dei rilievi eseguiti dovremmo teoricamente confrontare il modello ottenuto per via fotogrammetrica con l'oggetto reale, valutando per differenza di coordinate lo scostamento tra i due. Questo darebbe una stima in senso assoluto della precisione del nostro rilievo, ma siccome un confronto con la realtà è impossibile da realizzare, dovremmo accontentarci di un'approssimazione del caso reale. È dunque necessario creare un modello di riferimento, utilizzando le tradizionali tecniche di rilievo come scansioni laser e stazione totale.

Più che una stima della precisione in termini assoluti, sarà dunque un confronto tra le tecniche di rilevamento tradizionali e le nuove possibilità offerte dall'utilizzo della fotogrammetria.

7.2.1.1 Inquadramento

Prima di eseguire un qualsiasi tipo di rilievo è stato necessario inquadrare l'edificio ed il terreno che lo circonda in un opportuno sistema di riferimento. La rete di inquadramento è stata materializzata e rilevata nel 2013. La fase di materializzazione della rete è avvenuta posizionando diversi capisaldi su manufatti ritenuti stabili nel tempo, come ad esempio i tombini in calcestruzzo e successivamente è stato effettuato il rilievo tramite tecnologia GNSS per ottenere con precisione le coordinate planimetriche dei punti in un sistema di riferimento globale ed in maniera tale da avere un numero sufficiente di misure ridondanti, utili poi nella fase di compensazione. Nella figura seguente si osserva la posizione dei capisaldi.



Figura 7.2. Posizione dei capisaldi della rete di inquadramento

Allo scopo di definire per ciascun caposaldo la quota ortometrica più precisa rispetto alla quota geodetica, è stato effettuato un rilievo altimetrico mediante la *Livellazione Geometrica dal mezzo* di Alta Precisione, usufruendo dell'allacciamento al caposaldo 117060 di quota nota pari a 3,0470 m appartenente al consorzio di bonifica, situato in prossimità dell'edificio. La rete istituita per il rilievo è stata divisa in 5 anelli come illustrato in figura:



Figura 7.3. Rete di livellazione

Nel sistema di riferimento cartografico UTMED50 pertanto le coordinate compensate definitive sono riportate in tabella.

ID	NORD	EST	H (ortometrica)
	[m]	[m]	[m]
S1	4958051.929	723086.306	3.7293
S2	4958102.953	723173.734	3.7927
S3	4958305.682	723176.585	2.2066
S4	4958343.333	723153.588	2.3287
S5	4958316.980	723035.859	2.4483
S6	4958255.515	723022.282	2.8022
S7	4958216.423	723046.694	2.6171
S8	4958255.343	723099.278	2.3288
S9	4958293.478	723076.170	2.3856
S10	4958284.795	723040.695	2.9456
S11	4958340.444	723189.045	2.5069

Tabella 7.1: Coordinate compensate dei caposaldi di riferimento

7.2.1.2 Il rilievo tramite Scansione Laser

Il modello di riferimento, quello che più si avvicina al caso reale, è stato ottenuto mediante rilievo terrestre laser-scanner. Un rilievo di questo tipo unisce precisione a densità dei punti acquisiti, l'ideale per un confronto diretto con la nube di punti proveniente da un rilievo di tipo fotogrammetrico. Facendo stazione sui vari vertici della rete di inquadramento è stato possibile ottenere un modello tridimensionale dell'intero edificio. Il Laser Scanner utilizzato è il modello C10 della Leica. In figura 7.4. i risultati ottenuti:

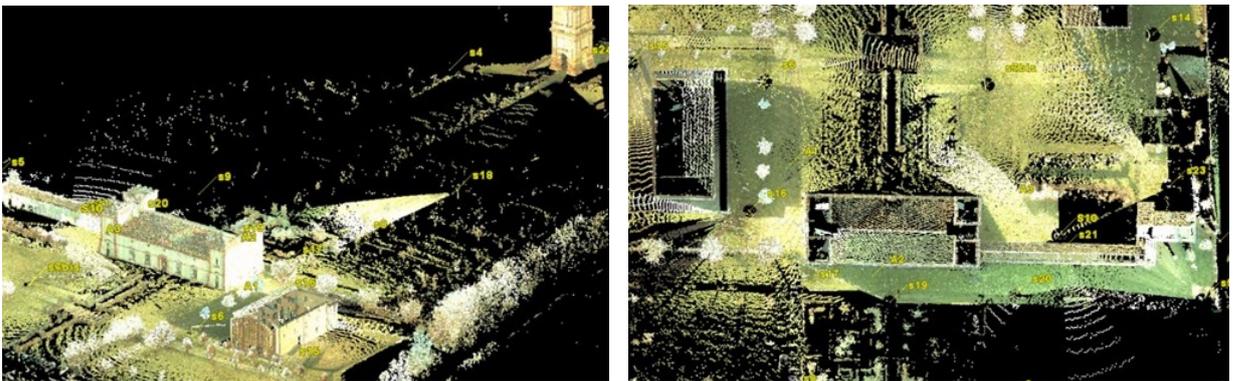


Figura 7.4. Nuvola di punti da Laser Scanner vista dall'alto e vista assonometrica

Oltre a questo, al fine di acquisire una serie di punti di coordinate note necessari sia per la fase di restituzione (saranno i nostri punti fotografici di appoggio), sia per le verifiche metriche (verranno utilizzati anche come punti di controllo), si è effettuato un rilievo di dettaglio del fabbricato con stazione totale Leica TS06.

In totale sono stati acquisiti 417 punti di cui 389 PFA sulle facciate dell'edificio e 28 GCP sul terreno circostante. A titolo di esempio si riporta in figura 7.5. e 7.6. la posizione dei PFA nel prospetto Ovest del Castello e i GCP rilevati.



Figura 7.5. PFA rilevati sul prospetto Ovest

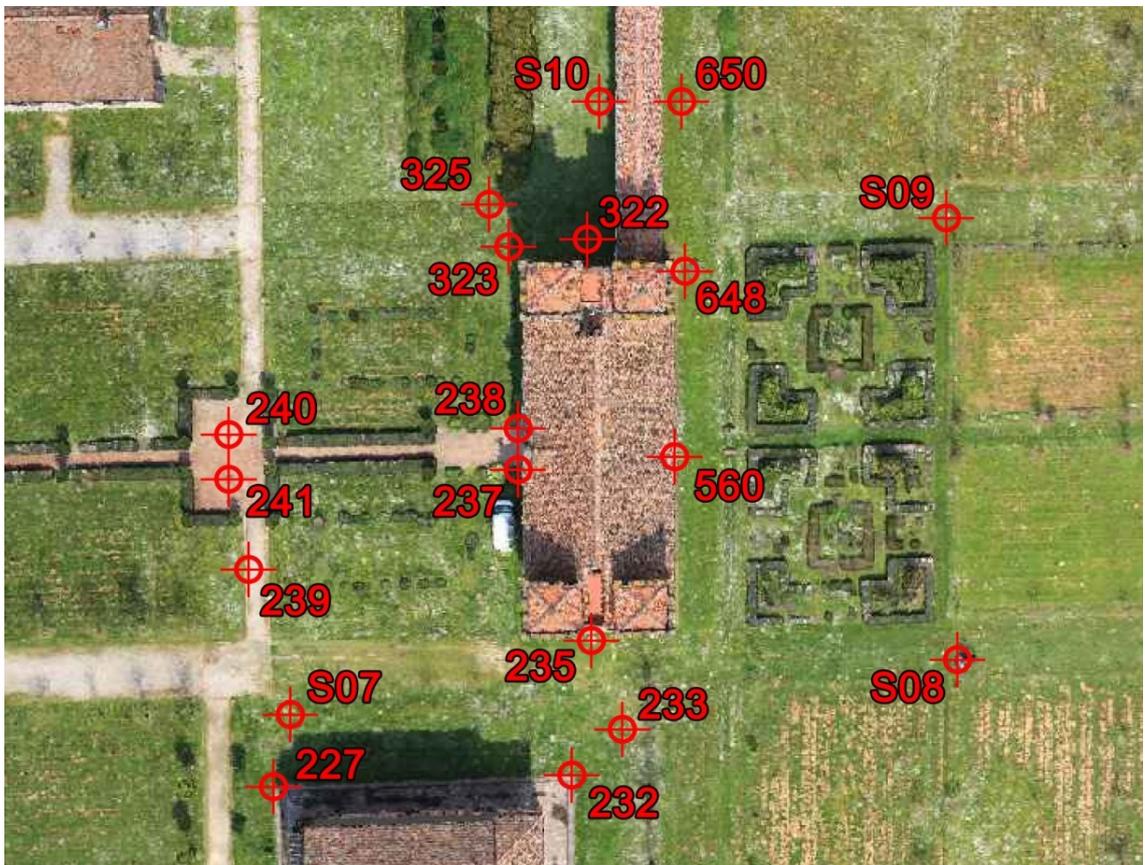


Figura 7.6. Ground Control Point (GCP) rilevati con Stazione Totale

Al fine di validare il modello di riferimento, che utilizzeremo come termine di confronto, è stata eseguita una semplice comparazione tra i punti fotografici di appoggio rilevati con la stazione totale e la nube di punti acquisita attraverso la scansione laser. Il confronto è stato eseguito con il software open source **CloudCompare**, secondo il metodo “nearest neighbor distance”. Viene direttamente computata la distanza tra il punto in esame ed il punto più vicino della nube di riferimento. Sono altresì possibili altre strategie di misura della distanza come la “least square plane”, ma è risultata troppo sensibile al rumore.

CONFRONTO TRA SCANSIONE LASER e PFA (Stazione Totale)



Figura 7.7. Scostamento tra i modelli osservati rappresentato in mappa di colore. Il modello confrontato è composto dall'insieme dei PFA

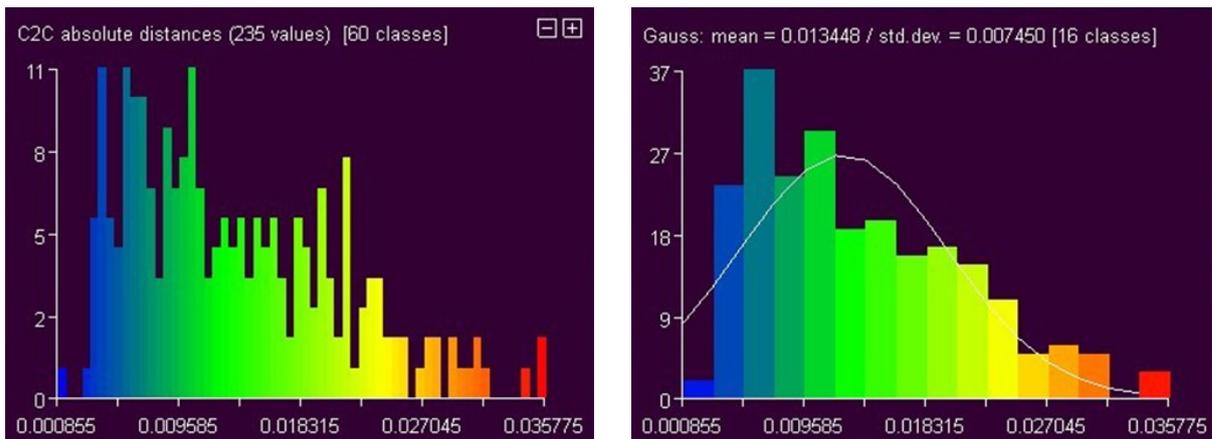


Figura 7.8. Istogramma di frequenza e distribuzione di probabilità.

Mediamente ho uno scostamento tra i dati di circa 1,3 cm. Il modello di riferimento è pertanto ritenuto affidabile. Nei capitoli successivi vedremo i diversi casi di studio affrontato seguendo passo-passo le varie fasi che li compongono.

8 LA PRECISIONE NEL RILIEVO INTEGRATO DEI BENI CULTURALI

È il primo caso di studio che andremo ad osservare, si tratta del rilievo integrato della “Delizia Estense del Verginese”. Lo scopo di questo primo caso di studio è quello di valutare la precisione di due tipologie standard di rilievo fotogrammetrico, quello definito terrestre, cioè con immagini realizzate da terra e quello aereo, immagini acquisite attraverso l’utilizzo di un RPAS (acronimo di Remotely Piloted Aircraft Systems). L’utilizzo di questa tecnologia nel rilievo architettonico è strettamente legato alla necessita di poter rilevare con precisione anche zone difficili da raggiungere, come la copertura o le parti più alte di una facciata. Come si può osservare dalle nubi di punti ottenute tramite scansioni laser il rilievo della copertura da terra risulta spesso problematico o completamente mancante. Saranno dunque utilizzati sensori diversi e scale differenti. Nell’ottica di una possibile restituzione integrata, si andranno a valutare attraverso procedure di confronto gli errori commessi, in maniera da poterne stimare la precisione attraverso un loro trattamento statistico. Di seguito si osservano passo dopo passo le varie fasi del rilievo.

8.1 STRUMENTAZIONE UTILIZZATA

La piattaforma utilizzata per il conseguimento del rilievo aerofotogrammetrico è un RPAS modello **Dji S800**, un esacottero del peso di circa 8 kg. In figura 8.1. lo si osserva in volo sopra il castello.



Figura 8.1. Drone impiegato per l’acquisizione del dato tramite rilievo aerofotogrammetrico.

Montata sul drone attraverso un apposito *gimbal* vi è la fotocamera digitale. Per questo tipo di rilievo è stata scelta una **CANON EOS M** con sensore CMOS di dimensioni 22,3 x 14,9 mm a 18 megapixel con lunghezza focale di circa 18 mm.



Figura 8.2. Fotocamera digitale Canon EOS M

SCHEDA TECNICA CANON EOS M	
Sensore immagine	
Tipo	CMOS da 22,3 x 14,9 mm
Numero effettivo pixel	Ca. 18 Megapixel
Numero totale pixel	Ca. 18,5 Megapixel
Rapporto immagine	3:2
Filtro low-pass	Integrato / Fisso
Pulizia sensore	EOS integrated deaning system
Filtro colore	Colori primari
Processo immagine	
Tipo	DIGIC 5
Obiettivo	
Lunghezza focale	Equivalente a 1,6 x la lunghezza focale nominale dell'obiettivo
Messa a fuoco	
Tipo	Sistema Hybrid AF con sensore CMOS dedicato
Sistema AF / punti	31 punti AF max
Intervallo di utilizzo AF	EV 1 – 18 (a 23°C e ISO 100)
Modalità AF	One Shot, Servo
Selezione del punto AF	Automatica, Manuale
Raggio Ausiliario AF	Led
Controllo esposizione	
Compensazione esposizione	±3 EV con incrementi di 1/3 di stop o ½ stop
Sensibilità ISO	AUTO (400-6.400), 100 – 12800 con incrementi di 1 stop
Otturatore	
Velocità	30-1/4.000 sec (con incrementi di 1/3 o ½ stop) + Bulb (la gamma completa dei tempi di scatto varia secondo la modalità di scatto)
Tipo	Otturatore con prima tendina elettronica e seconda meccanica
Scheda memoria	
Tipo scheda	SD, SDHC, SDXC (UHS-I)
Alimentazione	
Batterie	Batteria ricaricabile Li-ion LP-E12 (batteria e caricabatteria a corredo)
Dimensioni e peso	
Materiale corpo fotocamera	Acciaio inox, lega di magnesio, resina di policarbonato con fibra di vetro
Peso	298 g ca.

Tabella 8.1: Scheda tecnica Canon EOS M

La macchina fotografica impiegata per le foto terrestri è invece una **CANON EOS 7D**, obiettivo con distanza focale fissa di 35 mm; si tratta di una fotocamera digitale reflex di elevatissima qualità, dotata di sensore CMOS da 18 megapixel di dimensioni 22,3 x 14,9 mm, gamma ISO fino a 12800.



Figura 8.3: Fotocamera digitale Canon EOS 7D

SCHEDA TECNICA CANON EOS 7D	
Sensore immagine	
Tipo	CMOS da 22,3 mm x 14,9 mm
Pixel effettivi	Ca. 18,00 Megapixel
Pixel totali	Ca. 19,00 Megapixel
Rapporto	3:2
Filtro passa basso	Integrato/fisso con rivestimento al fluoro
Pulizia sensore	Sistema integrato di pulizia EOS
Tipo filtro colore	Colori primari
Processore	
Tipo	Doppio "DIGIC 4"
Obiettivo	
Lunghezza focale	Equivalente a 1,6 volte la lunghezza focale dell'obiettivo
Messa a fuoco	
Tipo	Sensore CMOS con TIL-CT-SIR
Otturatore	
Tipo	Otturatore con scorrimento verticale sul piano focale controllato elettronicamente
Alimentazione	
Batterie	Batteria ricaricabile agli ioni di litio LP-E6 (in dotazione), 1xCR1616 per data e impostazioni
Durata batteria	Ca. 800 (a 23°C, AE 50%, FE 50%) Ca. 750 (a 0°C, AE 50%, FE 50%)
Caratteristiche fisiche	
Materiali corpo fotocamera	Lega di magnesio
Dimensioni (L x A x P)	148,2 x 110,7 x 73,5 mm
Peso	820 gr ca.

Tabella 8.2: Scheda tecnica Canon EOS 7D

Per la restituzione è stato utilizzato il software commerciale **Photoscan di Agisoft**, con i seguenti settaggi:

Workflow	
Align photo	
Accuracy	Medium
Pair preselection	Disabled
Point Limit	40000
Build preliminary mesh	
Surface type	Arbitrary
Source data	Sparse
Interpolation	Enabled
Polygon count	Custom (200000)
Point classes	All
Import gcp (ground control settings)	
Camera accuracy (m)	10
Marker accuracy (m)	0.005
Scale bar accuracy (m)	0.001
Projection accuracy (m)	0.1
Tie point accuracy (pix)	4
Align and merge chunks	
Method	Point Based
Accuracy	Medium
Preselection images	None
Build dense cloud	
Quality	Medium
Depth filtering	Aggressive

Tabella 8.3: Parametri di Photoscan utilizzati per la fase di restituzione

Per i confronti invece è stato utilizzato il software **CloudCompare** esattamente come già visto nel paragrafo §7.2.1.2.

8.2 PIANIFICAZIONE DEL RILIEVO ED ACQUISIZIONE DEL DATO

Considerate le esigenze del rilievo in fase di pianificazione del volo si è deciso di operare in due direzioni, eseguendo due tipologie di volo. Un primo ad assi della camera nadirali, con lo scopo di acquisire le immagini necessarie alla restituzione del coperto, ed un secondo ad assi inclinati volando attorno al fabbricato per completarne il rilievo inserendo nell'elaborazione anche le immagini delle facciate, altrimenti mancanti.



Figura 8.4. Immagini aeree del castello del Verginense acquisite tramite RPAS

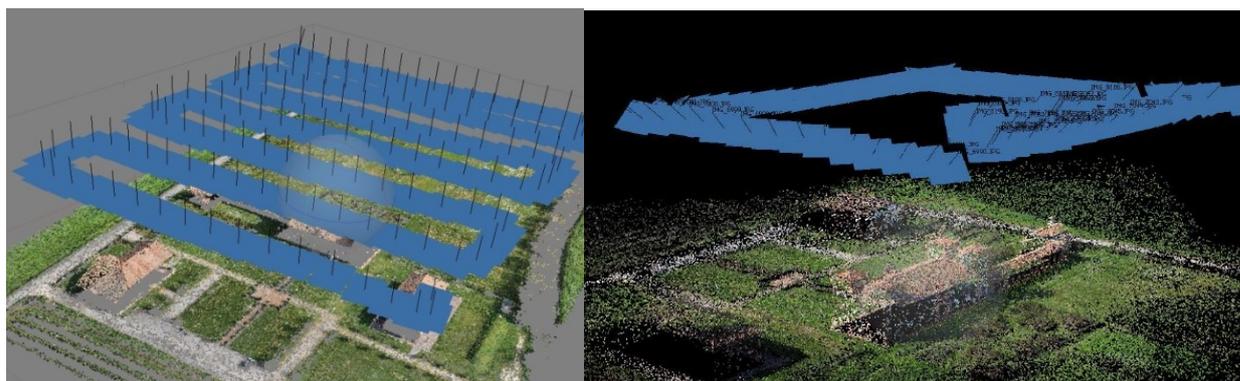


Figura 8.5. Schemi di presa: a sinistra una vista della posizione delle camere durante il volo ad assi nadirali. A destra il volo ad assi inclinati.

Viste le dimensioni dell'edificio e l'estensione della zona da rilevare in fase di pianificazione è stata scelta una quota di volo pari a circa 50 metri. A questa quota, con la camera scelta, si ottiene un GSD pari a 11 mm, ritenuto compatibile con il passo di campionamento del laser scanner e sufficientemente piccolo da descrivere in maniera completa la geometria della copertura. La stessa distanza dall'oggetto sarà mantenuta anche durante il volo ad assi inclinati.

La geometria di presa per il volo ad assi nadirali è la classica aerofotogrammetrica, composta da strisciate ad assi paralleli. È stata settata la camera per scattare foto ad intervalli regolari, secondo un principio di sovrapposizione dell'80% in direzione longitudinale e del 60% in direzione trasversale. Gli scatti totali sono risultati 130 per una superficie di circa 3300 m². La sovrapposizione è stata scelta

volutamente abbondante, in questo modo è possibile ottenere una volta tornati in laboratorio diversi dataset fotografici con cui effettuare le elaborazioni.

Anche per il rilievo fotogrammetrico terrestre sono state acquisite una sequenza di immagini con sovrapposizione media dell'80% sia longitudinale che trasversale, collocandosi a una distanza di circa 7 metri dai prospetti del castello. Gli scatti sono avvenuti ponendosi di fronte alle facciate e facendo circa 5- 6 fotografie lungo tutta l'altezza del castello rimanendo nella medesima posizione ed inclinando la macchina fotografica, poi ci si è spostati di qualche passo e di nuovo, posizionandosi frontalmente si sono eseguite le foto per tutta l'altezza del fabbricato dalla nuova ubicazione e così via per tutta la lunghezza del prospetto.

Tale procedimento è stato adottato per tutte le facciate, avendo cura di fare negli spigoli una serie di scatti d'insieme a distanza di presa maggiore da utilizzare durante la fase di orientamento dei singoli fotogrammi in un'ottica di elaborazione integrata, in modo da avere immagini di collegamento tra i prospetti.

Dunque gli schemi di acquisizione delle immagini seguiti, sono illustrati di seguito.

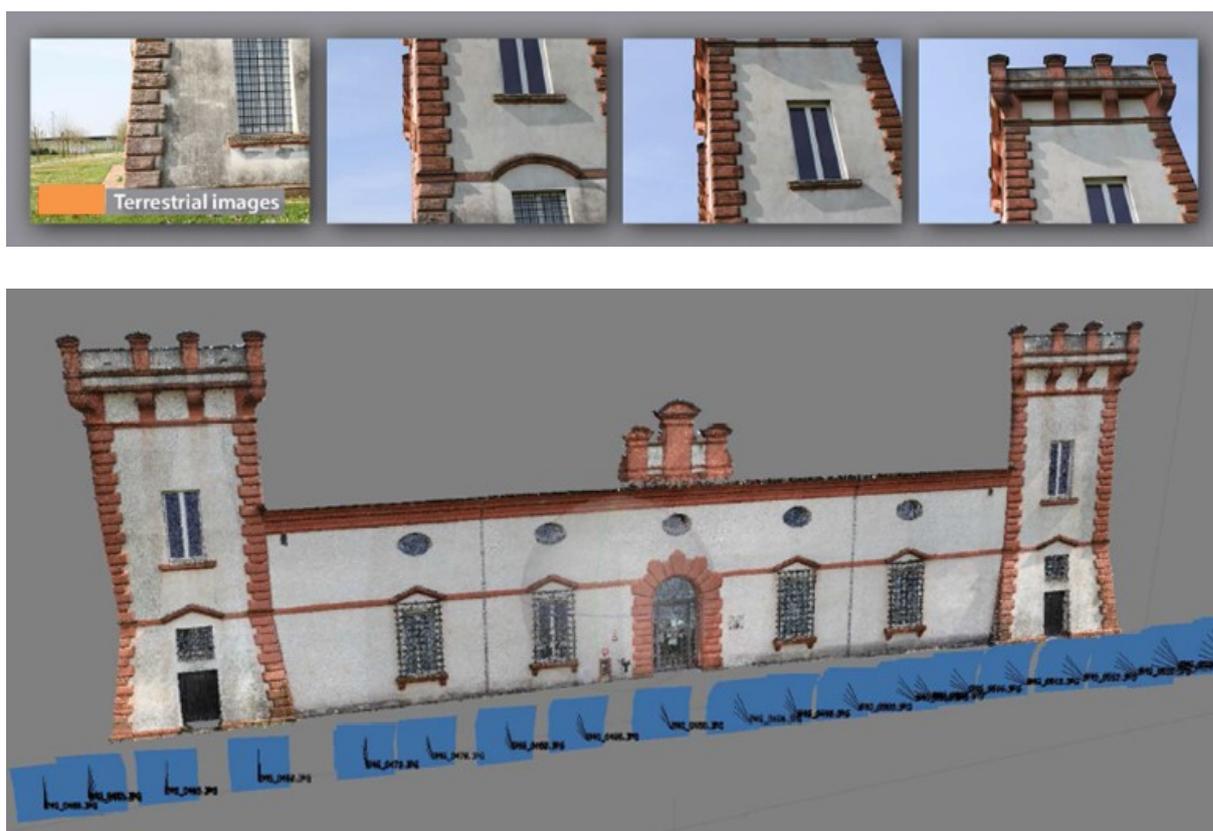


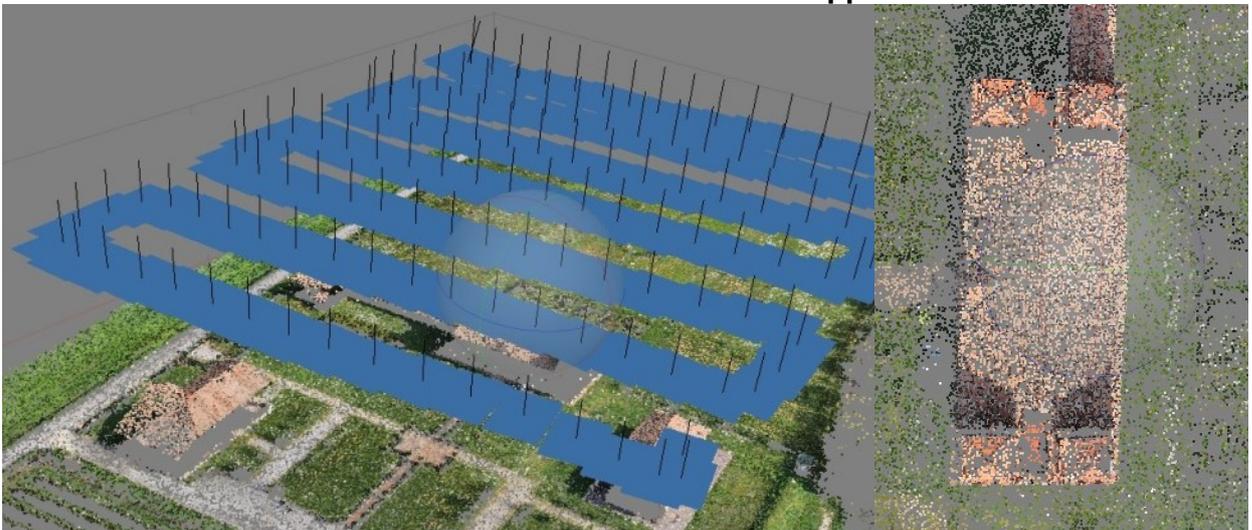
Figura 8.6. In alto una sequenza di immagini acquisite da terra; sotto lo schema tipico di presa terrestre.

In totale sono state acquisite circa 700 immagini poi opportunamente selezionate per creare i 4 dataset utili ai fini della restituzione. Un dataset per ciascuna facciata dell'edificio.

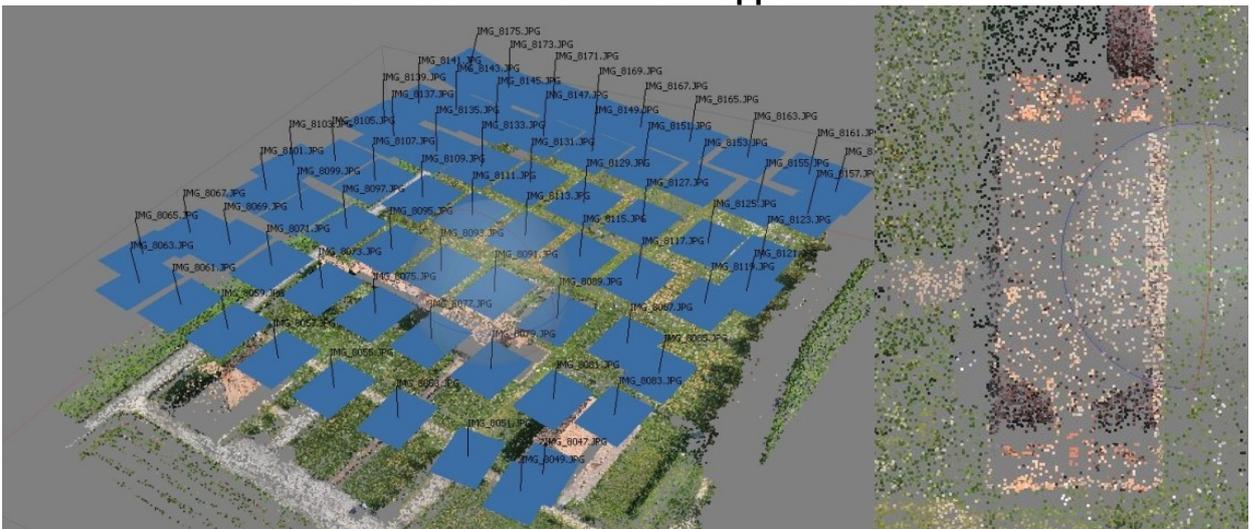
8.3 ELABORAZIONE DEL DATO

Come accennato in fase di pianificazione per l'elaborazione del dato non sono state utilizzate tutte le immagini ma da ciascuna tipologia di presa è stato estratto il dataset ritenuto più idoneo. Non conoscendo a priori la sovrapposizione ideale da utilizzare in questi casi è stato eseguito un piccolo test di precisione per verificare quale fosse l'overlap ideale tra le sequenze di immagini. Il controllo è stato eseguito sul dataset acquisito con RPAS e camera con asse di presa nadirale. Nel dettaglio sono stati ricavati 3 dataset: ricoprimento pari all'80% (tutte le foto), il 60% (una foto ogni due) ed il 40% (una foto ogni tre). Le immagini di ciascun dataset sono poi state elaborate con il software Photoscan di Agisoft.

DATASET COMPLETO 130 IMMAGINI – Sovrapposizione 80%



DATASET 650 IMMAGINI – Sovrapposizione 60%



DATASET 43 IMMAGINI – Sovrapposizione 40%

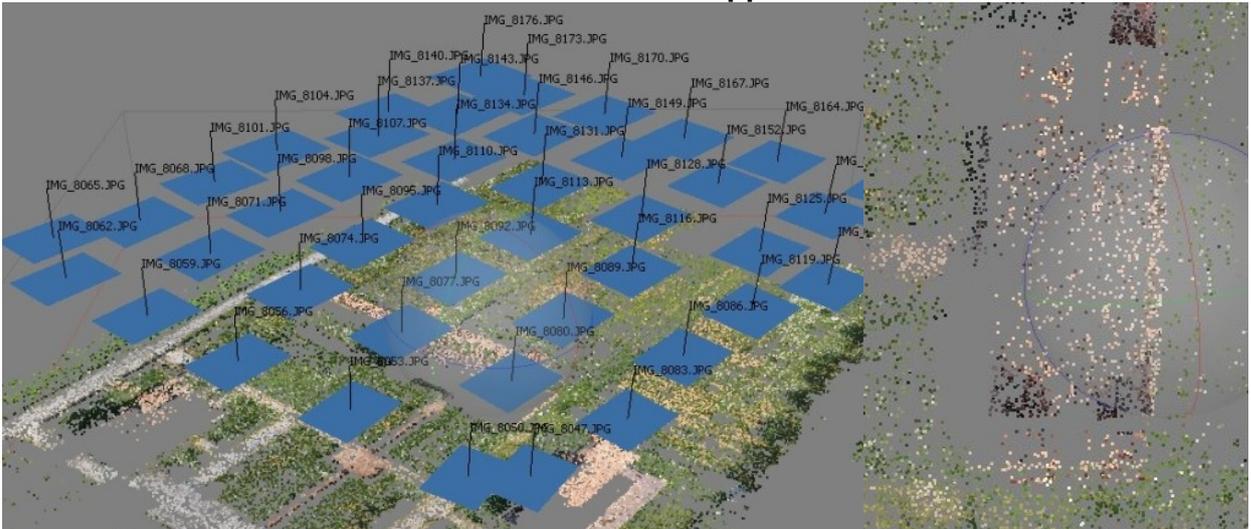
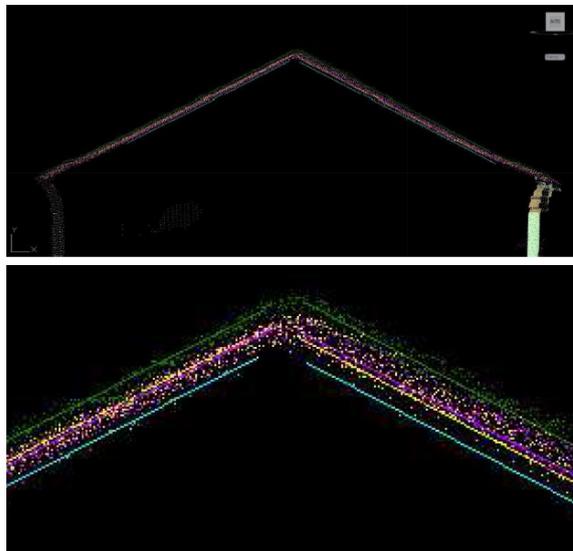


Figura 8.7. Orientamento ricostruzione sparsa dei dataset utilizzati.

In questa fase preliminare è stata utilizzata la nube sparsa, cioè quella creata sulla base delle corrispondenze trovate, in quanto ritenuta sufficiente per la valutazione della precisione. Sono dunque state esportate le nubi ed inserite in ambiente cad. Sono poi state ricavate manualmente 5 sezioni ad intervalli regolari su tutto il coperto ed è stata misurata la distanza tra i punti di ciascuna nube e la nube di riferimento.



SEZIONI	SCOSTAMENTI DAL MODELLO DI RIFERIMENTO		
	80% (m)	60% (m)	40% (m)
1	0.046	0.057	0.180
2	0.091	0.092	0.219
3	0.061	0.079	0.173
4	0.058	0.073	0.155
5	0.06	0.08	0.179

Figura 8.8. Nell'immagine a sinistra si osservano le sezioni delle nubi di punti esportate in corrispondenza della copertura. In tabella i risultati in termini di differenze tra le sezioni estratte dalle nubi sparse fotogrammetriche e la scansione laser.

Osservando i valori ottenuti si deduce che un ricoprimento del 40% non è sufficiente ad ottenere una precisione elevata, mentre già considerando ricoprimenti dell'ordine del 60% o dell'80% si ottengono scostamenti minori (pari alla metà circa). Considerato che tra i due dataset (60% e 80%) le differenze sono minime (anche se risulta leggermente più preciso il rilievo ottenuto con ricoprimento pari

all'80%) è ragionevole continuare nell'elaborazione con un ricoprimento del 60% in quanto un numero minore (pari alla metà) di fotogrammi comporta tempi computazionali molto più ridotti. Dal momento che i test da effettuare sono numerosi si è preferito perseguire tale strada.

Si procede dunque con la creazione di un modello completo utilizzando per il momento le sole foto ad asse nadirale con un ricoprimento pari al 60%. Non verranno riportate nel dettaglio tutte le fasi dell'elaborazione, ma soltanto un breve resoconto finale, in quanto si preferisce dare spazio alle considerazioni successive in termini di confronti metrici ed alle considerazioni finali; per maggiori approfondimenti si rimanda all'articolo *"Accuracy of Cultural Heritage 3D models by RPAS and terrestrial photogrammetry"*, Bolognesi et al. (2014).

L'orientamento interno dei parametri è stato completamente affidato al software mentre per l'orientamento esterno sono stati utilizzati 19 ground control points. In totale sono state processate 65 immagini. I risultati della procedura di orientamento sono riportati nella tabella seguente:

MODELLO	IMMAGINI	GCP/PFA	ERROR [m]	ERROR [pix]
AEREO (NADIRALI)	65	19	0.027	0.439

Tabella 8.4: Risultati della procedura di orientamento delle immagini aeree ad assi nadirali.

Ritenendo accettabili gli errori si è proceduto alla creazione della nube densa, che poi sarà oggetto di confronto con la nube di riferimento.

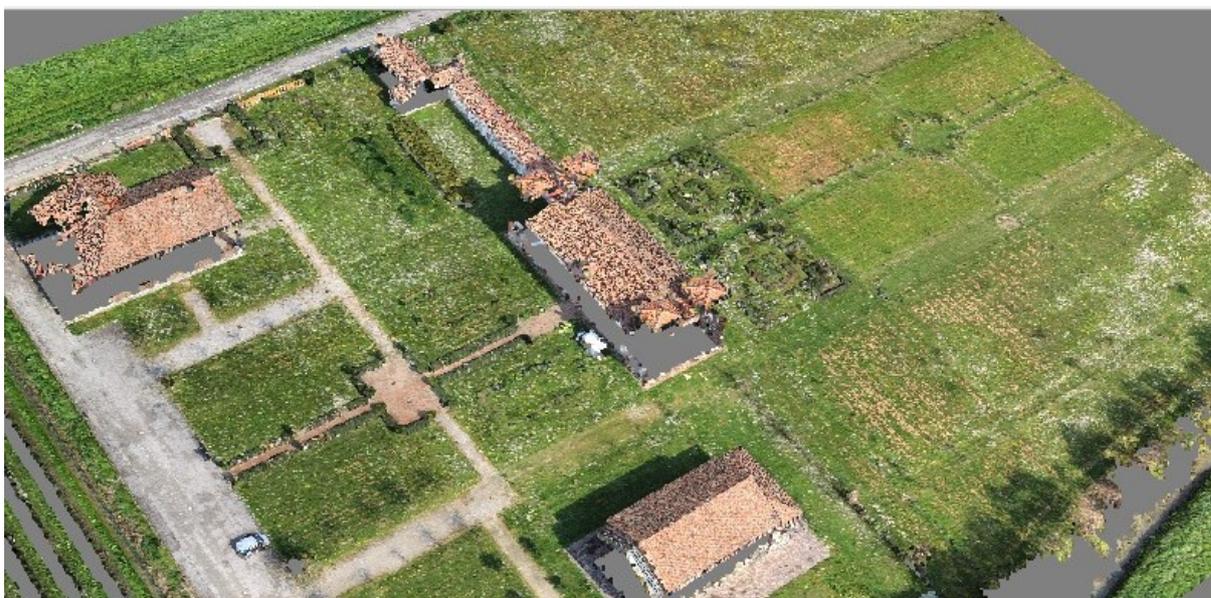


Figura 8.9. Nube densa di punti ottenuta con le immagini aeree ad assi nadirali

Come ci si aspettava, le facciate dell'edificio non sono state ricostruite questo perché dalle sole foto ad asse nadirale non risultano visibili. Per ottenere un modello completo è necessario integrare tale elaborazione con le immagini acquisite ad assi inclinati. Sono state aggiunte 134 immagini. Il report sulla fase di orientamento delle sole immagini ad asse nadirali è il seguente:

MODELLO	FOTOGRAFIE	GCP/PFA	ERROR [m]	ERROR [pix]
AEREO (INCLINATI)	134	19	0.027	0.938

Tabella 8.5: Risultati della procedura di orientamento delle immagini aeree ad assi inclinati.

Come si può osservare l'errore è simile a quello ottenuto con le sole immagini nadirali. Il modello con l'aggiunta di tali immagini risulta nel complesso completa, come dimostrano le immagini successive:



Figura 8.10. Nube densa di punti ottenuta con le immagini aeree ad assi inclinati.

Prima di valutare la precisione di tali ricostruzioni è bene completare la fase di rilievo ed elaborazione considerando anche il rilievo da terra. Le immagini acquisite da terra sono state suddivise in 4 dataset distinti, uno per ciascuna facciata del fabbricato. Allo stesso modo si riportano i risultati della procedura di orientamento:

MODELLO	FOTOGRAFIE	GCP/PFA	ERRORE [m]	ERRORE [pix]
Prospetto OVEST	102	15	0.008	0.312
Prospetto SUD	89	15	0.005	1.428
Prospetto EST	91	15	0.006	0.210
Prospetto NORD	127	15	0.004	0.903

Tabella 8.6: Risultati della procedura di orientamento delle immagini aeree ad assi inclinati.

La precisione è molto più alta rispetto ai modelli creati con le immagini aeree. Questo aspetto è da imputare semplicemente alla minor distanza di presa che determina a parità di risoluzione un dettaglio maggiore nelle immagini. Il posizionamento sulle stesse dei GCP risulta più semplice e preciso, così come la qualità del modello ottenuto, come dimostrano le immagini seguenti:

PROSPETTO OVEST



Figura 8.11. Nube densa di punti del prospetto ovest ottenuto dal rilievo fotogrammetrico terrestre.

PROSPETTO EST



Figura 8.12. Nube densa di punti del prospetto est ottenuto dal rilievo fotogrammetrico terrestre.

PROSPETTO NORD

PROSPETTO SUD



Figura 8.13. Nubi dense di punti dei prospetti nord e sud ottenuto dal rilievo fotogrammetrico terrestre.

Le nubi ottenute sono state opportunamente pulite e filtrate per poter poi eseguire un corretto confronto con la nube di riferimento.

8.3.1 Il modello Integrato

Come accennato in precedenza a questo punto si vuole creare un modello unico che utilizzi le foto terrestri per la costruzione delle facciate e quelle da RPAS per il coperto. In un'ottica di ricostruzione multivista e multiscala il software sarebbe in grado di elaborare tutte le immagini contemporaneamente ma si andrebbe incontro a tempi computazionali lunghissimi, in quanto si tratta di elaborare circa 608 immagini contemporaneamente. Prima di procedere inoltre è bene notare che utilizzando le immagini terrestri, le foto aeree ad assi inclinati non sono più necessarie, in quanto il dato relativo alla geometria delle facciate è facilmente reperibile attraverso le prime. Essendo tra l'altro molto più dettagliate, per questa elaborazione si decide di non utilizzare le immagini aeree ad assi inclinati. Si passa da 608 immagini a 494 immagini. Sono un numero ancora elevato.

Per velocizzare il processo di restituzione si opta per una strategia che ci consente in tempi relativamente brevi di costruire il nostro modello integrato in maniera sequenziale. Photoscan ci consente infatti di realizzare all'interno di uno stesso lavoro modelli separati e poi di unirli in funzione dei GCP o eseguendo una nuova triangolazione di punti. Pertanto saranno realizzati separatamente i modelli relativi alle facciate e quello relativo al coperto. Verranno poi uniti in un secondo momento per sovrapposizione dei GCP. Del nuovo modello così creato verrà poi eseguito un nuovo bundle adjustment e successivamente verrà creata una nube di punti densa che andremo a confrontare con il modello di riferimento.

Seguendo questo processo il risultato della procedura di orientamento del modello integrato è la seguente:

MODELLO	FOTOGRAFIE	GCP/PFA	ERRORE [m]	ERRORE [pix]
MODELLO INTEGRATO	494	28	0.018	0.689

Tabella 8.7. Risultato della procedura di orientamento del modello integrato.

Mediamente otteniamo un errore sull'orientamento quasi stimabile a priori avendo osservato i relativi errori dei singoli modelli. Ricordiamo come il modello realizzato utilizzando le sole foto ad assi nadirali presentava dei residui di 2,7 cm mentre mediamente quelli relativi alle singole facciate erano affetti da errori compresi tra gli 8 ed i 4 mm. Avendo eseguito un orientamento soddisfacente si procede con la creazione della nube densa sempre utilizzando le stesse impostazioni dei precedenti modelli.



Figura 8.14. Vista Sud-Ovest della nube densa di punti del modello integrato



Figura 8.15. Vista Nord-Est della nube densa di punti del modello integrato

Le nubi di punti ottenute a seguito della procedura di dense matching sono poi state opportunamente meshate con lo scopo di ottenere un modello texturizzato dell'intero castello. Le mesh di forma triangolare vanno a costituire una superficie solida alla quale viene applicata la texture direttamente dalle immagini disponibili. Di seguito a titolo di esempio si riportano i risultati ottenuti:



Figura 8.16. Vista Nord-Est del modello tridimensionale texturizzato integrato del castello del Verginese



Figura 8.17. Vista Sud-Ovest modello tridimensionale texturizzato integrato del castello del Verginese

Il risultato finale è indubbiamente di buona fattura, ma lo scopo della modellazione non è la qualità del prodotto finito, ma la sua precisione in termini metrici. Ora che sono state create tutte le nubi ed il modello texturizzato finale è possibile eseguire i confronti con il modello di riferimento ottenuto attraverso scansione laser e stazione totale.

8.4 VALUTAZIONE DELLA PRECISIONE

In questa sezione verranno presentati i confronti eseguiti tra i modelli creati per via fotogrammetrica e quelli di riferimento. In particolare saranno presentati tre tipologie di confronto. Il primo caso che andremo ad esporre sarà il confronto diretto tra le nubi fotogrammetriche e la nube di riferimento rilevata tramite scansione laser. Il secondo caso che andremo ad osservare sarà invece il confronto tra i modelli tridimensionali meshati ottenuti per via fotogrammetrica e la nube di riferimento della scansione laser. Il terzo ed ultimo confronto sarà tra le nubi dense fotogrammetriche ed i PFA. Per i primi due confronti verranno analizzate sia le singole nubi (i singoli prospetti ottenuti dalle immagini

terrestri e il modello ottenuto dalle riprese aeree) che il modello integrato. Per il terzo confronto si osservano soltanto i modelli realizzati delle facciate. Di seguito si riportano i risultati di tutti i confronti.

8.4.1 Confronto automatico tra nube densa di punti fotogrammetrica e scansione laser.

PROSPETTO OVEST: Confronto tra nube densa e scansione laser

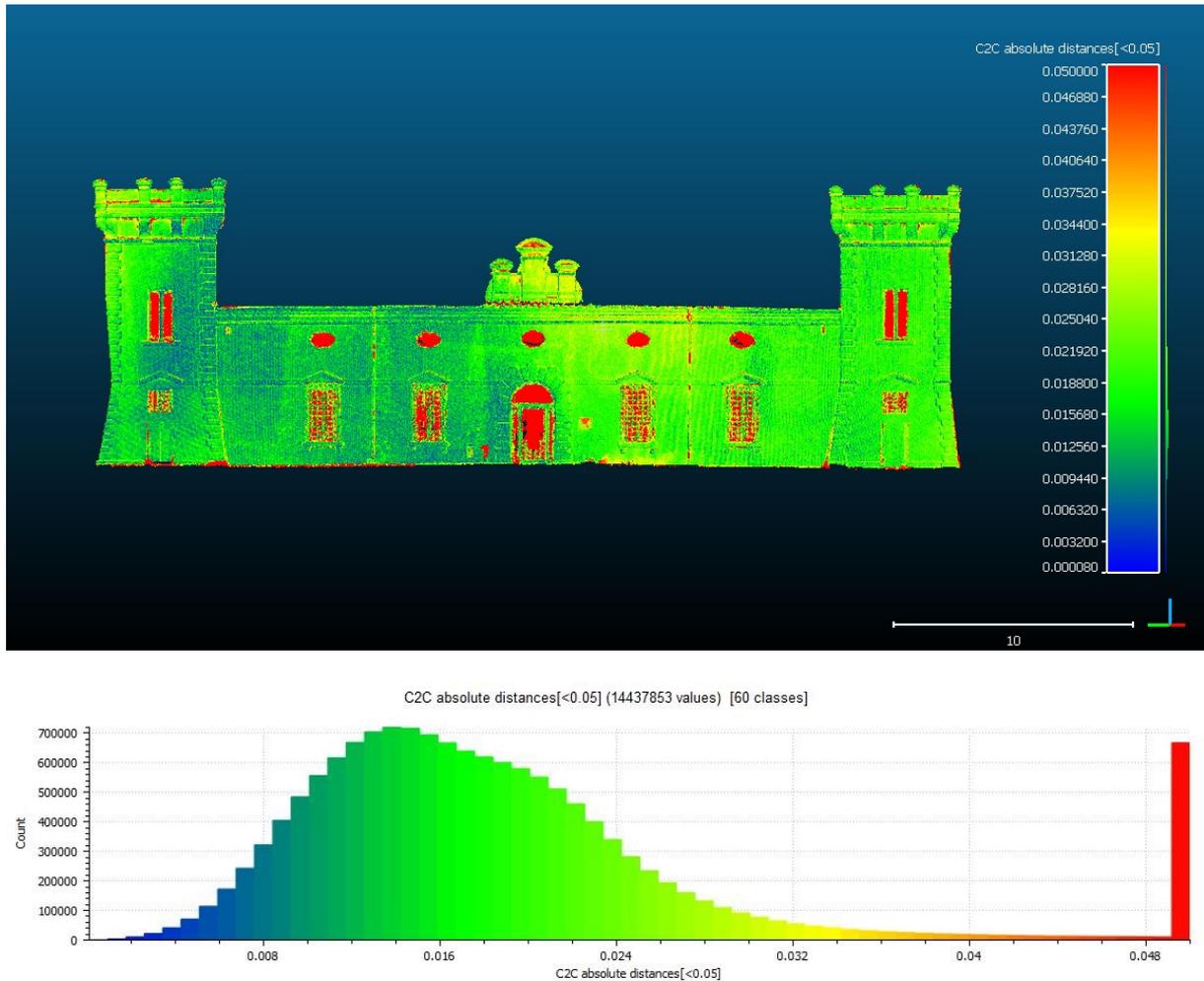


Figura 8.18. In alto si osservano le differenze in distanze assolute tra i modelli in esame attraverso una mappa di colore. In rosso le zone che presentano maggiori scostamenti. Nell'immagine in basso si riporta l'istogramma di frequenza relativo alle differenze misurate.

PROSPETTO EST: Confronto tra nube densa e scansione laser

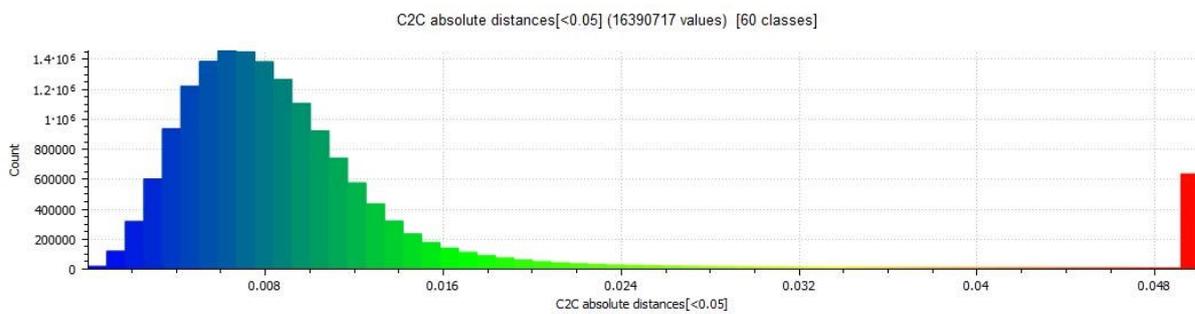
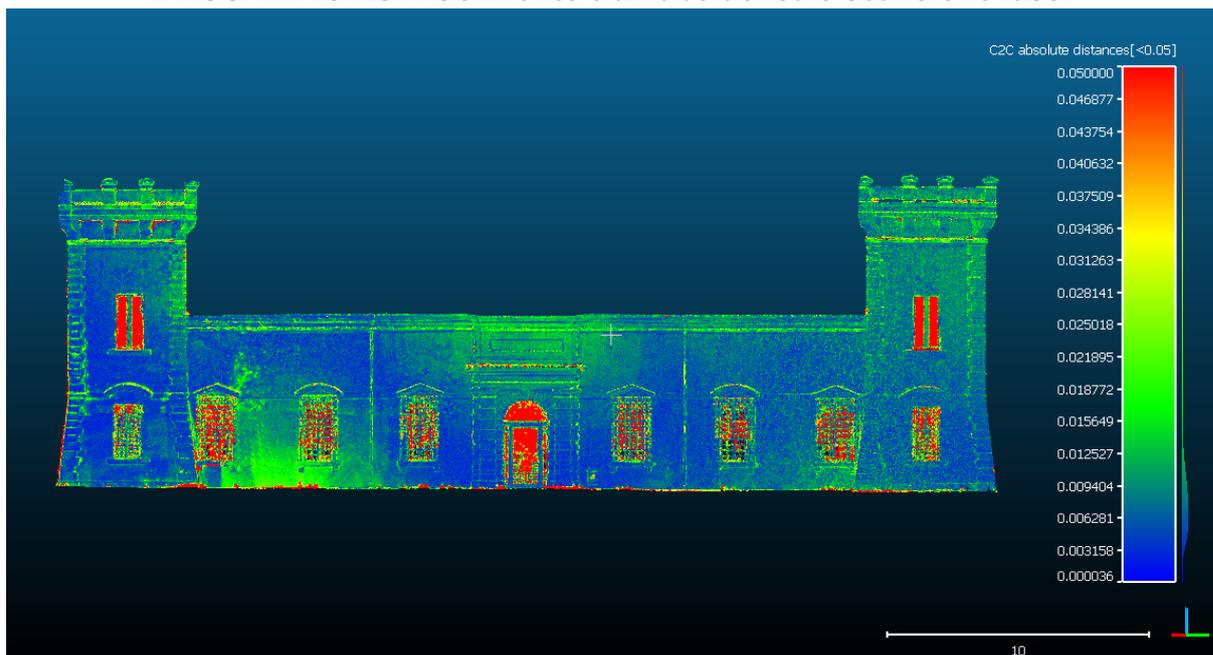


Figura 8.19. In alto si osservano le differenze in distanze assolute tra i modelli in esame attraverso una mappa di colore. In rosso le zone che presentano maggiori scostamenti. Nell'immagine in basso si riporta l'istogramma di frequenza relativo alle differenze misurate.

PROSPETTO NORD: Confronto tra nube densa e scansione laser

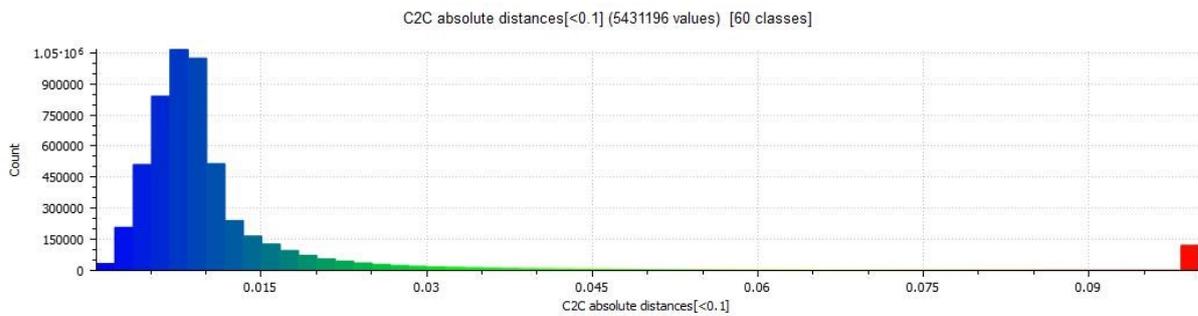
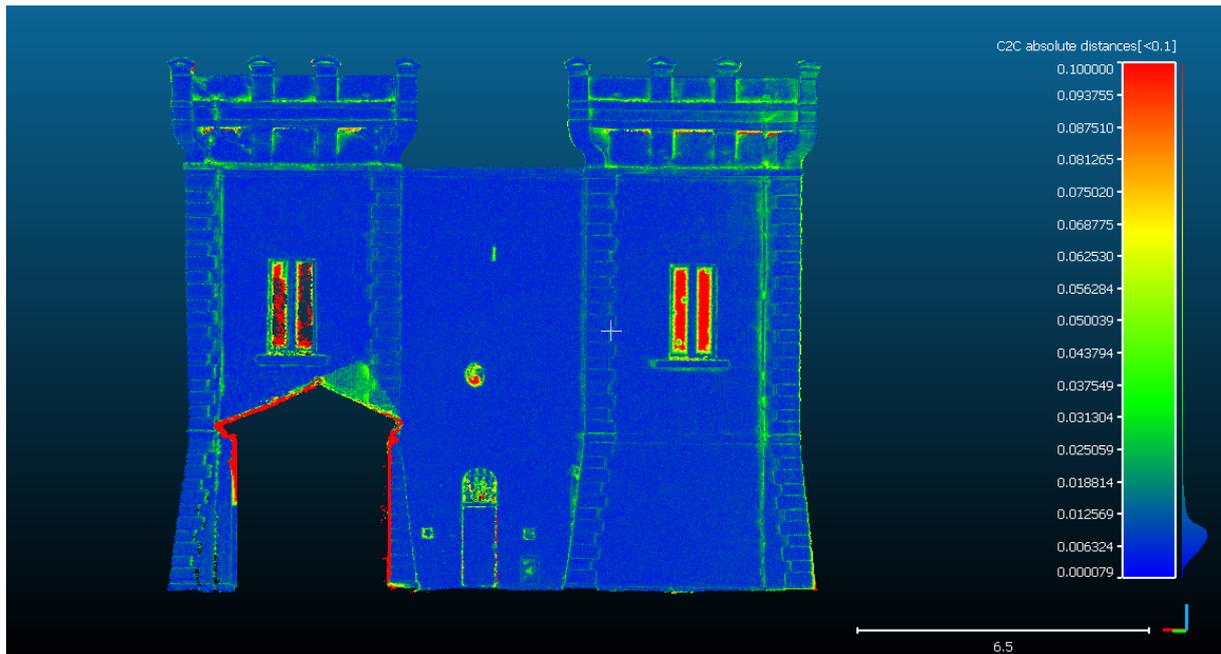


Figura 8.20. In alto si osservano le differenze in distanze assolute tra i modelli in esame attraverso una mappa di colore. In rosso le zone che presentano maggiori scostamenti. Nell'immagine in basso si riporta l'istogramma di frequenza relativo alle differenze misurate.

PROSPETTO SUD: Confronto tra nube densa e scansione laser

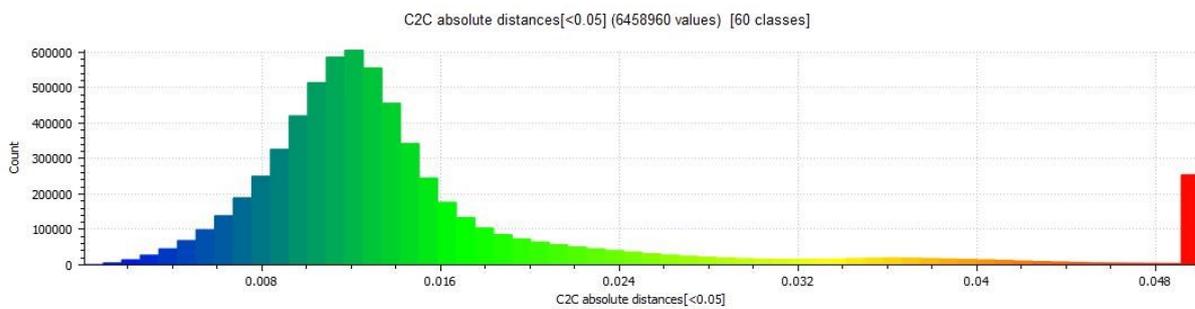
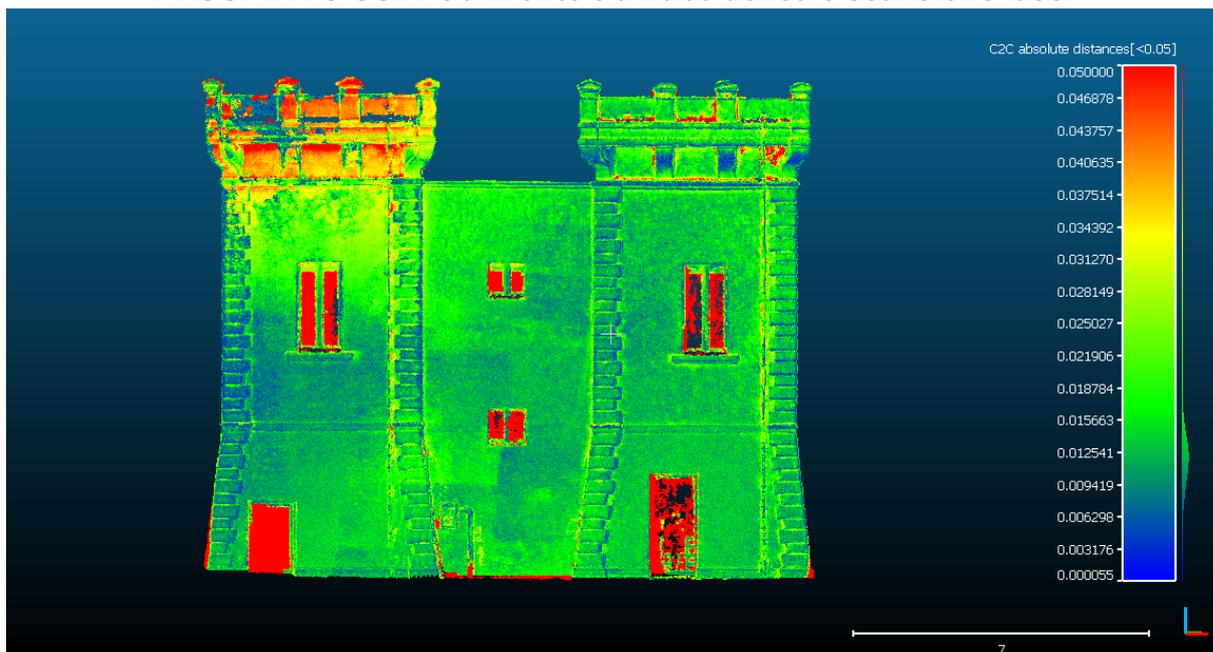


Figura 8.21. In alto si osservano le differenze in distanze assolute tra i modelli in esame attraverso una mappa di colore. In rosso le zone che presentano maggiori scostamenti. Nell'immagine in basso si riporta l'istogramma di frequenza relativo alle differenze misurate.

MODELLO INTEGRATO: Confronto tra nube densa e scansione laser

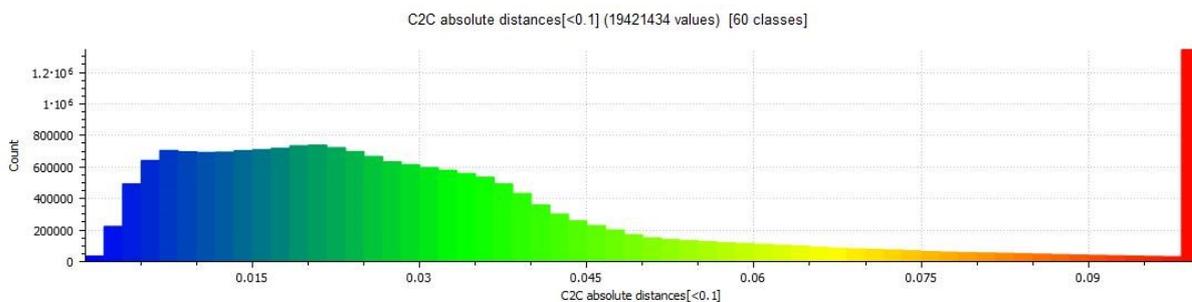
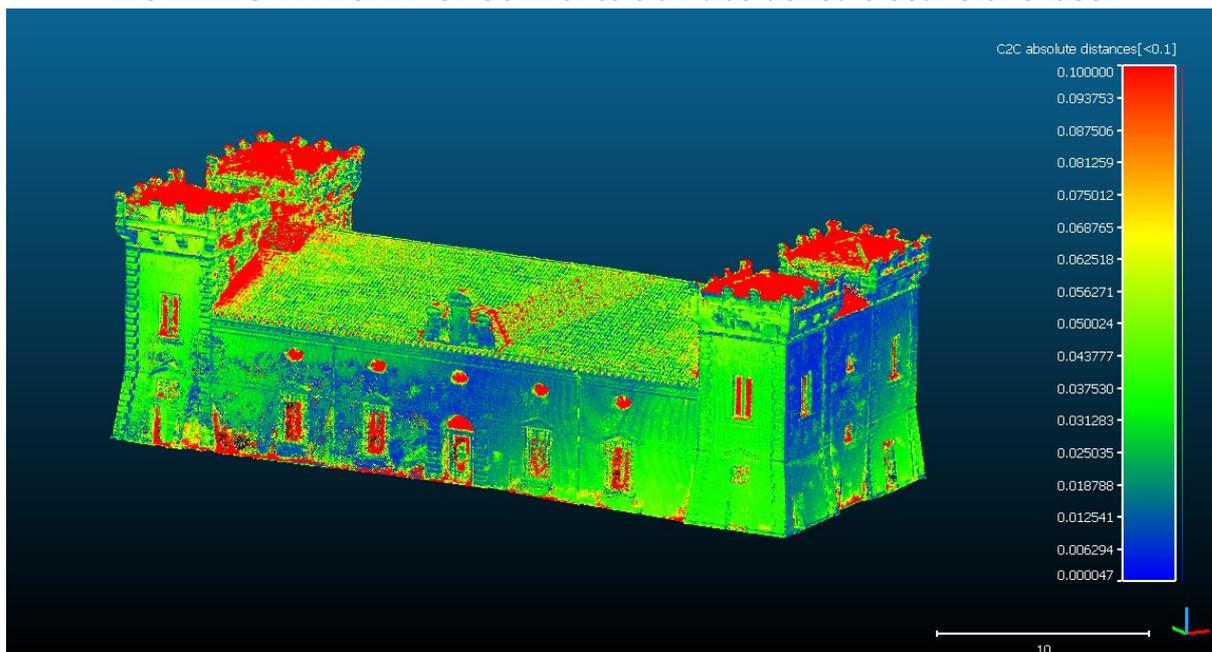


Figura 8.22. In alto si osservano le differenze in distanze assolute tra i modelli in esame attraverso una mappa di colore. In rosso le zone che presentano maggiori scostamenti. Nell'immagine in basso si riporta l'istogramma di frequenza relativo alle differenze misurate.

Nella seguente tabella si riportano i principali risultati ottenuti:

CONFRONTO TRA NUBI DENSE DI PUNTI (distanze assolute)		
MODELLO	DIFFERENZE IN VALORE ASSOLUTO NELLO SPAZIO	
	Media (m)	Deviazione standard (m)
Prospetto OVEST	0.019	0.010
Prospetto EST	0.010	0.010
Prospetto NORD	0.011	0.010
Prospetto SUD	0.015	0.010
MODELLO INTEGRATO	0.033	0.026

Tabella 8.8. Medie e deviazioni standard dei confronti metrici tra scansione laser e nubi dense di punti.

I risultati ottenibili con questo tipo di confronto sono in termini di valore assoluto. Sebbene gli istogrammi estratti ci consentono di stabilire con accuratezza la differenza presente tra i modelli, la

presenza di outliers (classe rossa presente in ogni istogramma all'estrema destra) e la mancanza del segno dello scostamento rendono il problema di analisi statistica mal condizionato.

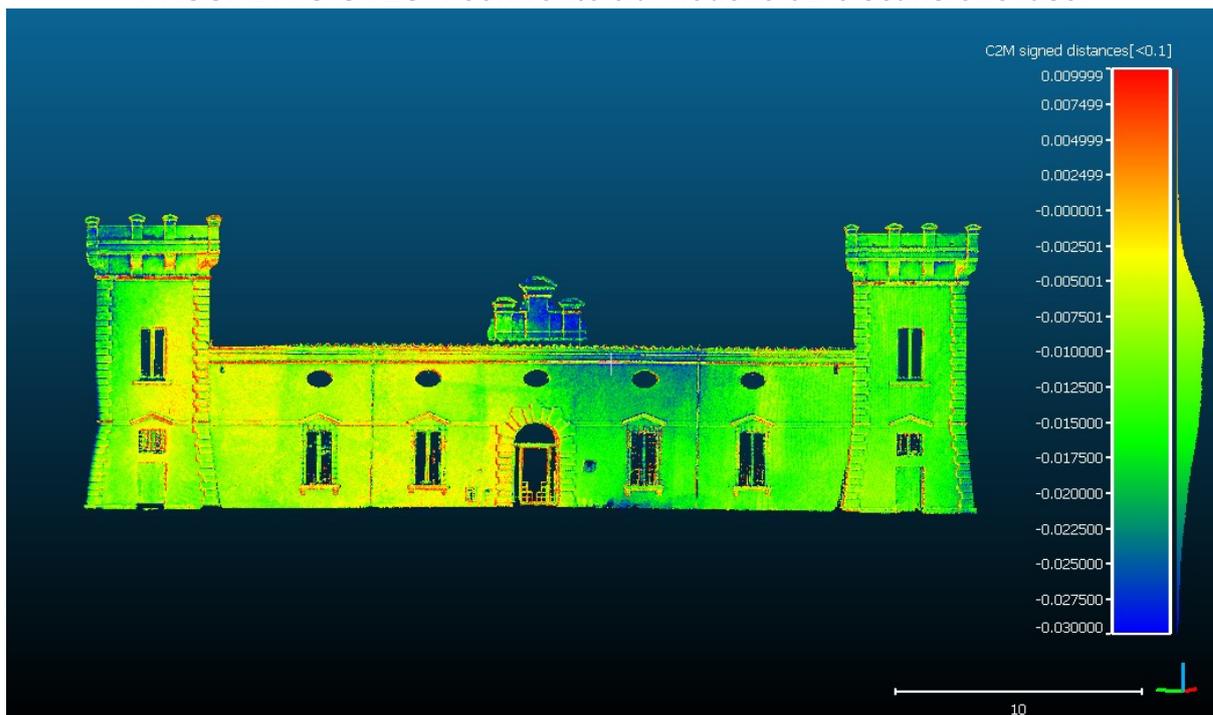
Dalle figure possiamo notare che gli outliers si trovano principalmente in corrispondenza delle parti vetrate del castello e, come si osserva nel confronto del modello integrato, nelle parti di coperto che tramite scansione laser non sono state rilevate (coperto delle torri). Le parti vetrate sono per definizione zone affette da rumore in quanto tramite fotogrammetria il riflesso e la trasparenza del vetro comportano da sempre problemi durante la fase di ricostruzione. Per quanto riguarda il coperto invece la questione è molto semplice: lo scostamento rilevato dal software dipende soltanto dal fatto che tramite scansione laser la copertura risulta a tratti incompleta. In quelle zone la densità dei punti del modello di riferimento non è compatibile con quella della nube fotogrammetrica pertanto con il criterio di confronto "nearest neighbor distance" il computo delle differenze risulta molto elevato. Per un'analisi più accurata tali punti devono essere rimossi dalla nube, oppure riempiti con un processo di ricostruzione solida tridimensionale. Come osservato nel paragrafo precedente in ambiente Photoscan è possibile costruire modelli tridimensionali composti da superfici solide (mesh). CloudCompare esegue allo stesso modo anche comparazioni tra mesh e nubi di punti.

Questa procedura ha inoltre un altro vantaggio, e cioè che la stima delle distanze viene svolta tenendo conto del segno dello scostamento. Disponendo di tale informazione sarà poi possibile condurre un'analisi statistica con lo scopo di valutare anche la presenza di errori di tipo sistematico. Pertanto si procede con la creazione dei modelli 3D e con i relativi confronti.

8.4.2 Confronto automatico tra modelli tridimensionali e scansione laser.

I risultati delle analisi sono riportati nelle immagini seguenti.

PROSPETTO OVEST: confronto tra modello 3D e scansione laser



Gauss: mean = -0.012044 / std.dev. = 0.006772 [845 classes]

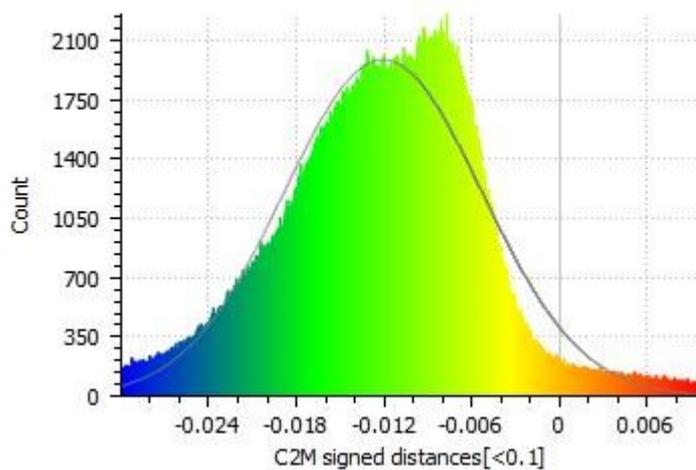
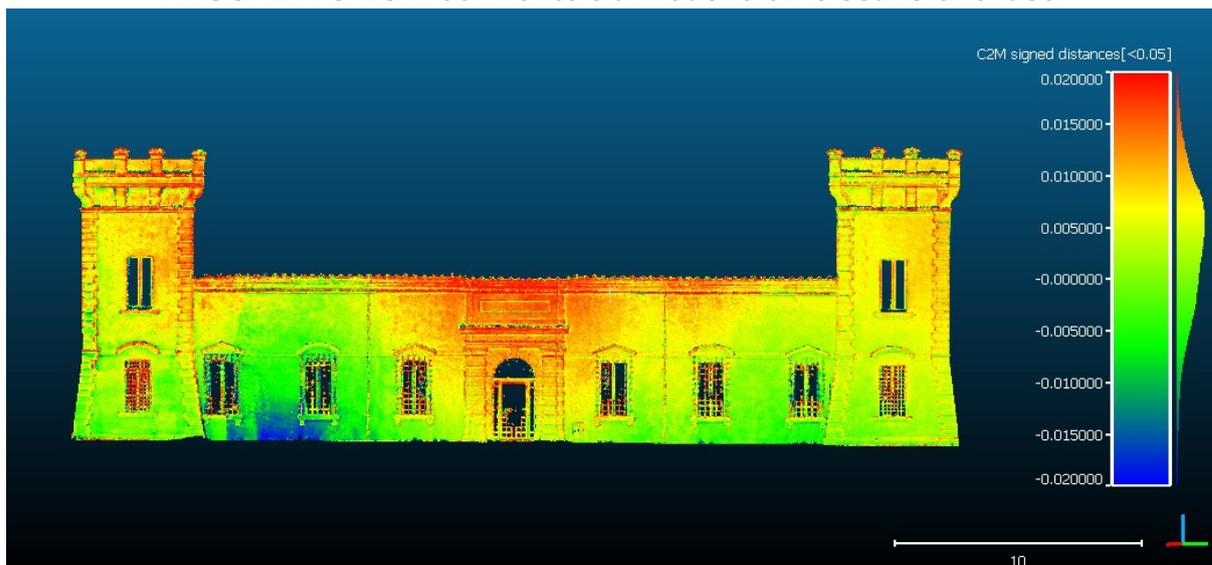


Figura 8.23. In alto si osservano le differenze tra i modelli in esame attraverso una mappa di colore. In rosso le zone che presentano maggiori scostamenti. Nell'immagine in basso si riporta l'istogramma di frequenza relativo alle differenze misurate e la distribuzione di probabilità calcolata.

PROSPETTO EST: confronto tra modello 3D e scansione laser



Gauss: mean = 0.002476 / std.dev. = 0.007087 [1610 classes]

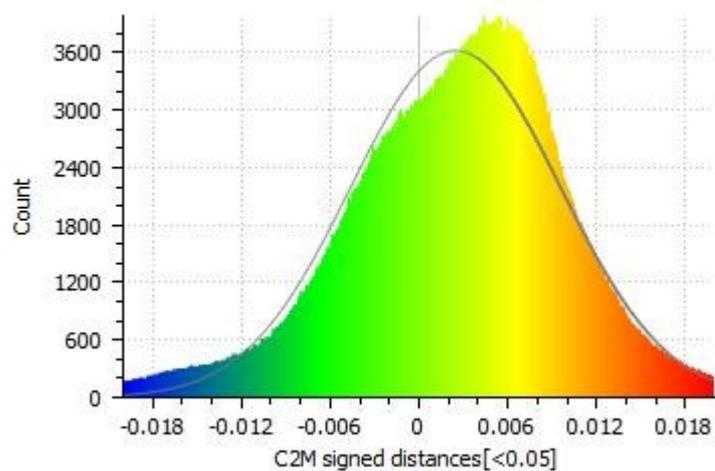
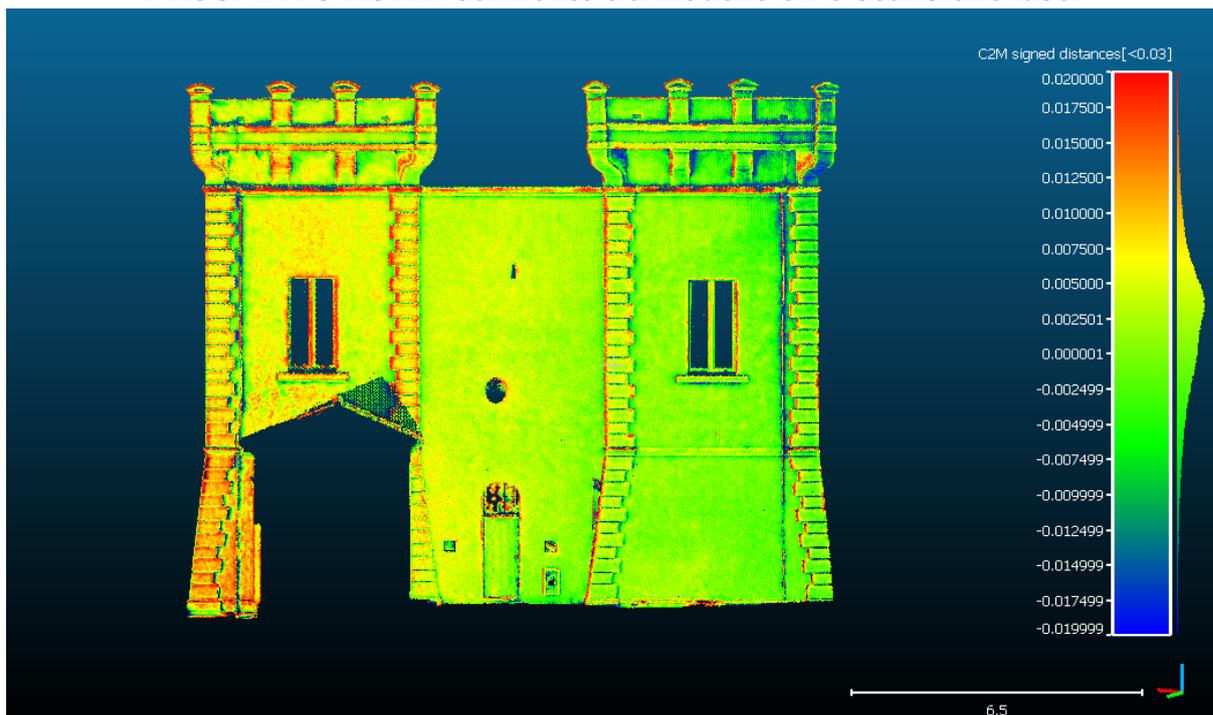


Figura 8.24. In alto si osservano le differenze tra i modelli in esame attraverso una mappa di colore. In rosso le zone che presentano maggiori scostamenti. Nell'immagine in basso si riporta l'istogramma di frequenza relativo alle differenze misurate e la distribuzione di probabilità calcolata.

PROSPETTO NORD: confronto tra modello 3D e scansione laser



Gauss: mean = 0.001439 / std.dev. = 0.006679 [824 classes]

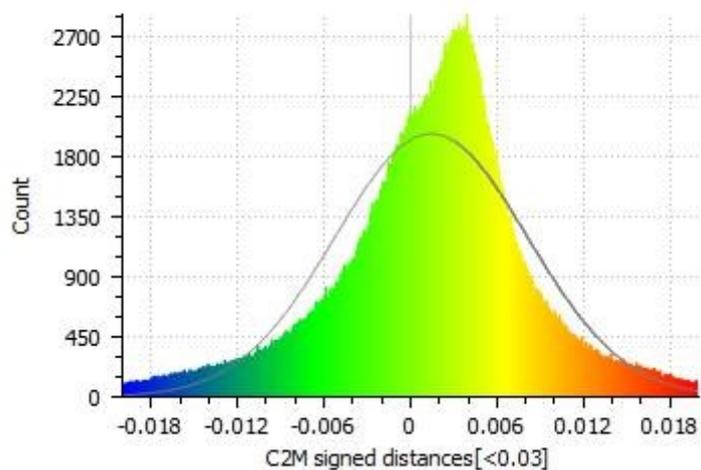
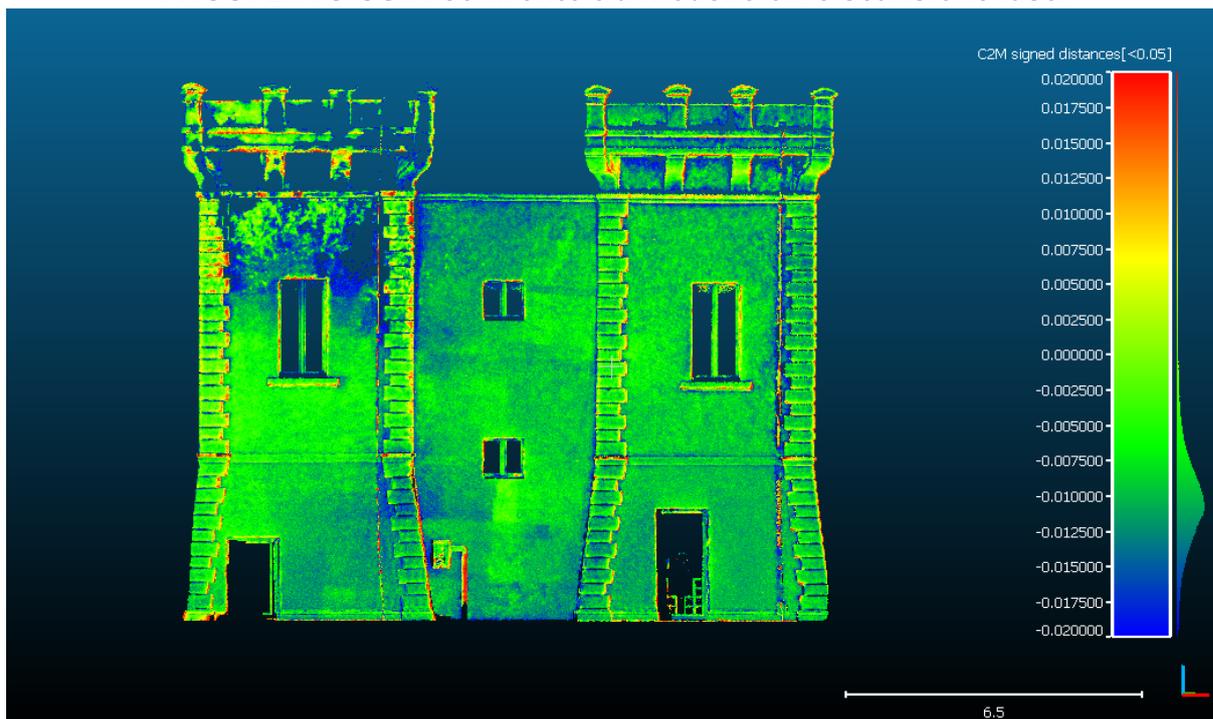


Figura 8.25. In alto si osservano le differenze tra i modelli in esame attraverso una mappa di colore. In rosso le zone che presentano maggiori scostamenti. Nell'immagine in basso si riporta l'istogramma di frequenza relativo alle differenze misurate e la distribuzione di probabilità calcolata.

PROSPETTO SUD: confronto tra modello 3D e scansione laser



Gauss: mean = -0.008658 / std.dev. = 0.006620 [892 classes]

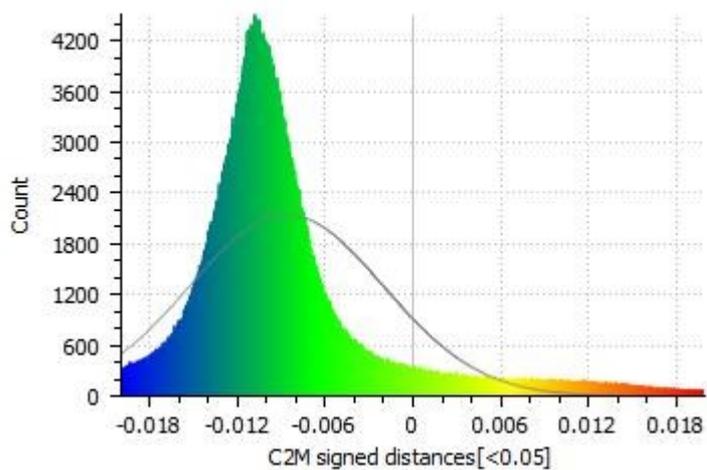
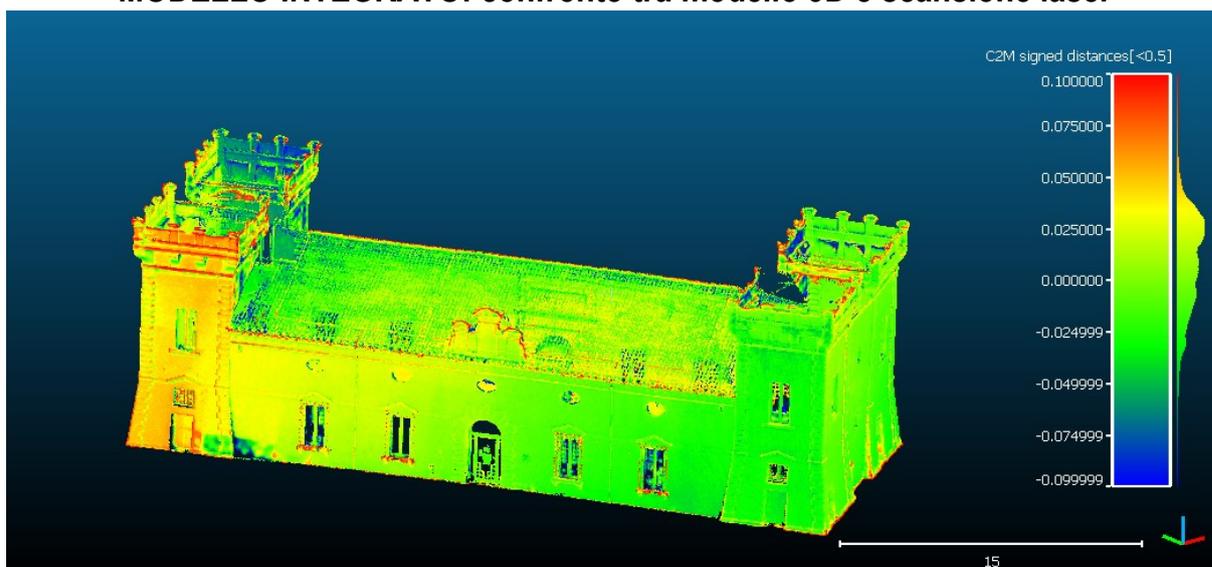


Figura 8.26. In alto si osservano le differenze tra i modelli in esame attraverso una mappa di colore. In rosso le zone che presentano maggiori scostamenti. Nell'immagine in basso si riporta l'istogramma di frequenza relativo alle differenze misurate e la distribuzione di probabilità calcolata.

MODELLO INTEGRATO: confronto tra modello 3D e scansione laser



Gauss: mean = 0.007339 / std.dev. = 0.028501 [2323 classes]

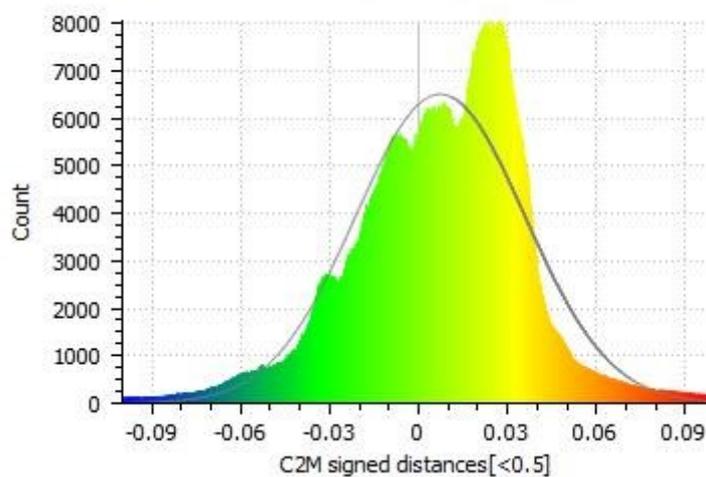


Figura 8.27. In alto si osservano le differenze tra i modelli in esame attraverso una mappa di colore. In rosso le zone che presentano maggiori scostamenti. Nell'immagine in basso si riporta l'istogramma di frequenza relativo alle differenze misurate e la distribuzione di probabilità calcolata.

Nella seguente tabella si riportano i principali risultati ottenuti:

CONFRONTO TRA SCANSIONE LASER E MODELLI TRIDIMENSIONALI		
MODELLO	DIFFERENZE NELLO SPAZIO	
	Media	Deviazione standard
	(m)	(m)
Prospetto OVEST	-0.012	0.007
Prospetto EST	0.002	0.007
Prospetto NORD	0.001	0.007
Prospetto SUD	-0.008	0.007
MODELLO INTEGRATO	0.007	0.028

Tabella 8.9. Medie e deviazioni standard dei confronti metrici tra scansione laser e modelli tridimensionali.

I risultati ottenuti filtrando gli outliers e considerando il giusto segno dello scostamento sono ritenuti più attendibili e significativi dei precedenti. Ad ogni modo la tendenza risulta essere la stessa, sono ovviamente diminuite le deviazioni standard, indice dell'accuratezza del rilievo. Per quanto riguarda la presenza di errori sistematici si osserva una disposizione dei valori medi molto vicina allo zero escluso il caso del prospetto Ovest che sembra avere una tendenza positiva di circa 1 cm.

8.4.3 Confronto automatico tra modelli tridimensionali e punti fotografici di appoggio.

PROSPETTO OVEST: Confronto tra modello 3D e PFA



Gauss: mean = -0.002553 / std.dev. = 0.024441 [12 classes]

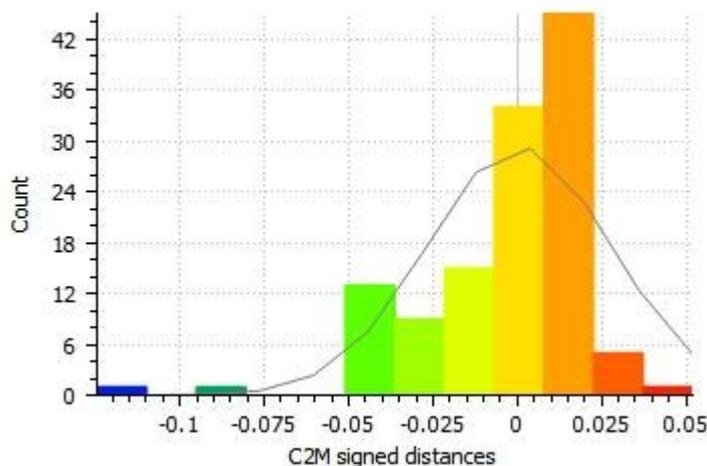
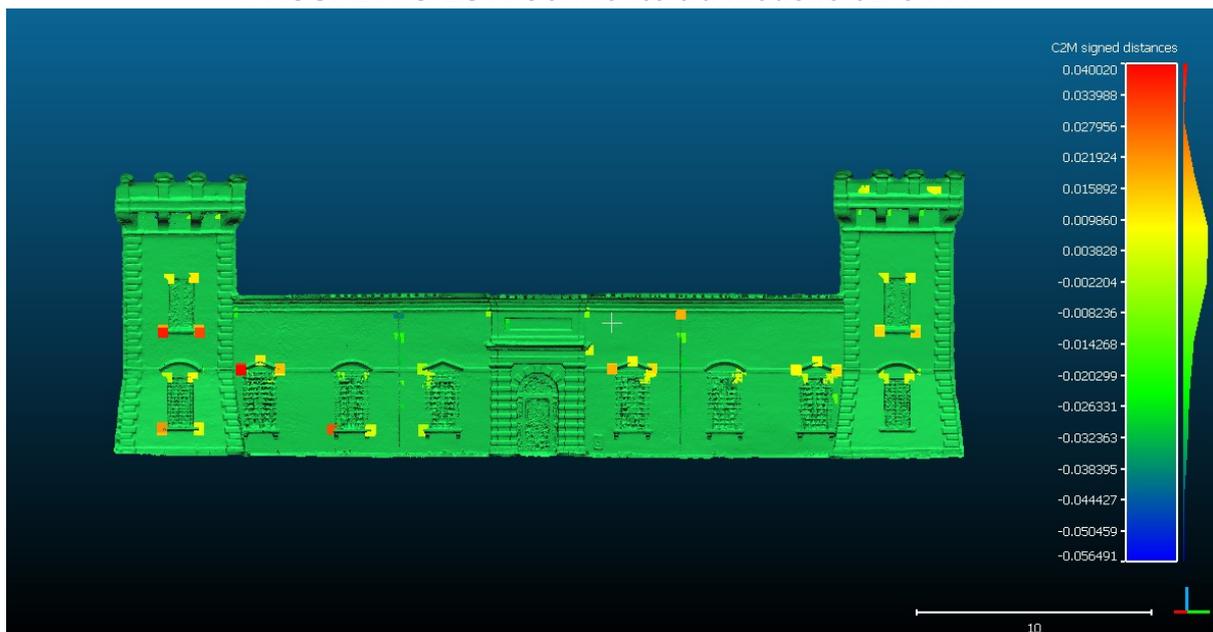


Figura 8.28. In alto si osservano le differenze tra il modello tridimensionale e i punti rilevati con stazione totale. In rosso i punti che presentano maggiori scostamento. Nell'immagine in basso si riporta l'istogramma di frequenza relativo alle differenze misurate e la distribuzione di probabilità calcolata.

PROSPETTO EST: Confronto tra modello 3D e PFA



Gauss: mean = -0.001135 / std.dev. = 0.016414 [10 classes]

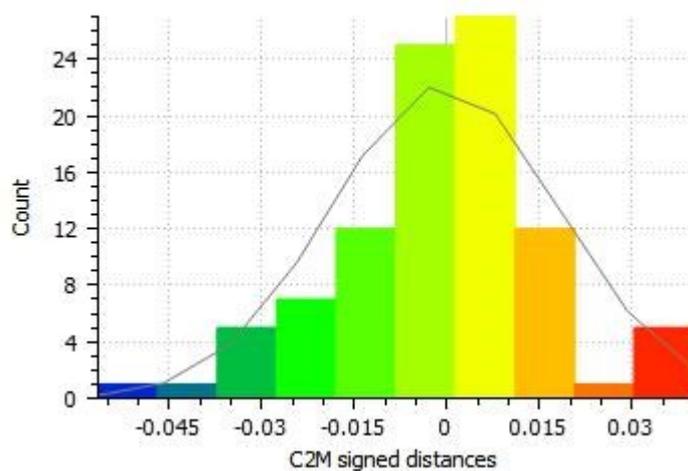
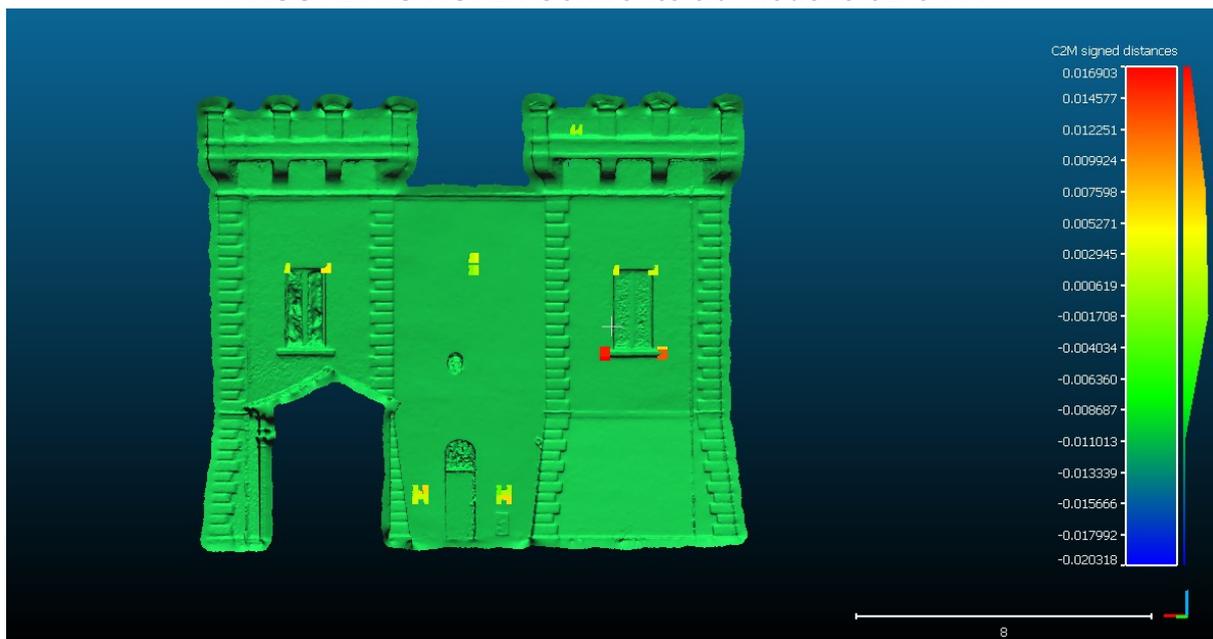


Figura 8.29. In alto si osservano le differenze tra il modello tridimensionale e i punti rilevati con stazione totale. In rosso i punti che presentano maggiori scostamento. Nell'immagine in basso si riporta l'istogramma di frequenza relativo alle differenze misurate e la distribuzione di probabilità calcolata.

PROSPETTO NORD: Confronto tra modello 3D e PFA



Gauss: mean = 0.002284 / std.dev. = 0.007269 [5 classes]

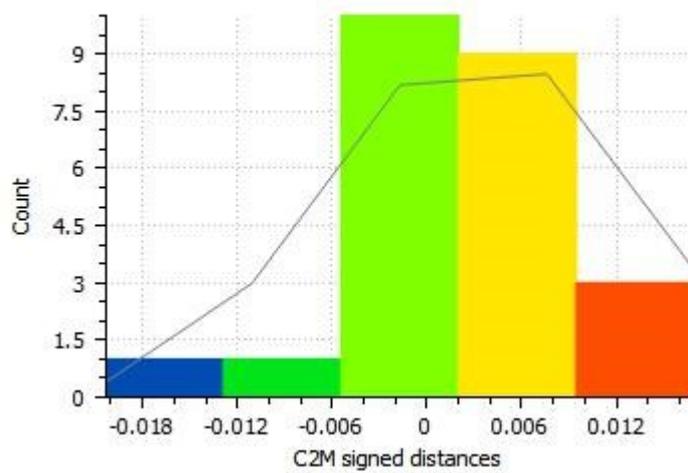
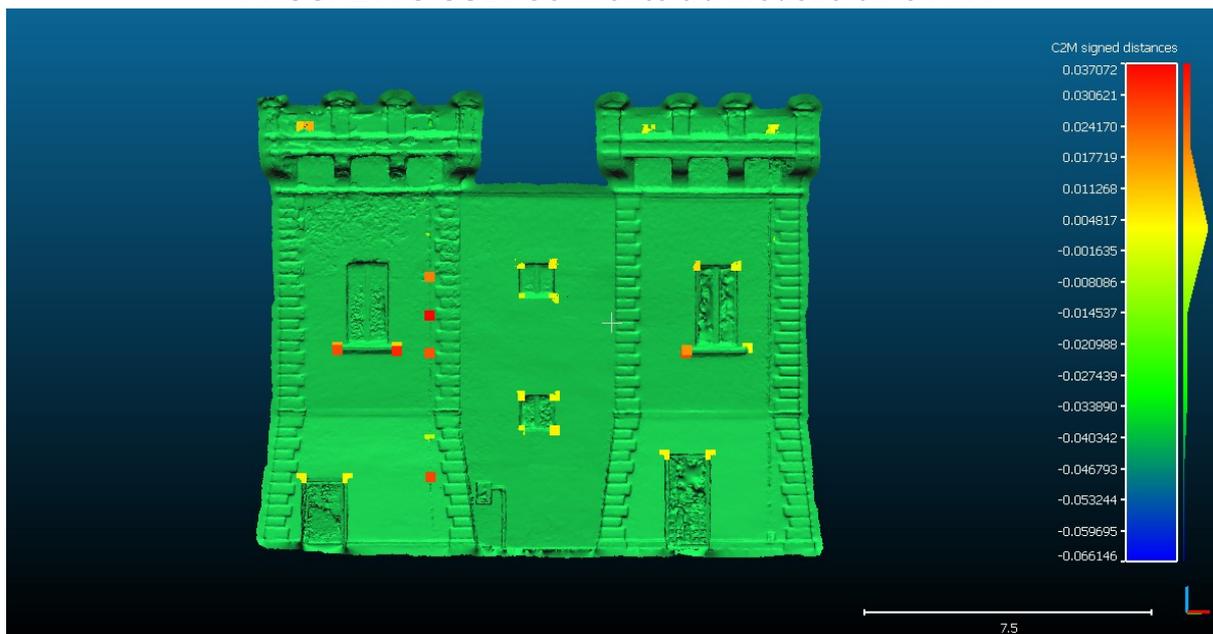


Figura 8.30. In alto si osservano le differenze tra il modello tridimensionale e i punti rilevati con stazione totale. In rosso i punti che presentano maggiori scostamento. Nell'immagine in basso si riporta l'istogramma di frequenza relativo alle differenze misurate e la distribuzione di probabilità calcolata.

PROSPETTO SUD: Confronto tra modello 3D e PFA



Gauss: mean = 0.000168 / std.dev. = 0.019852 [7 classes]

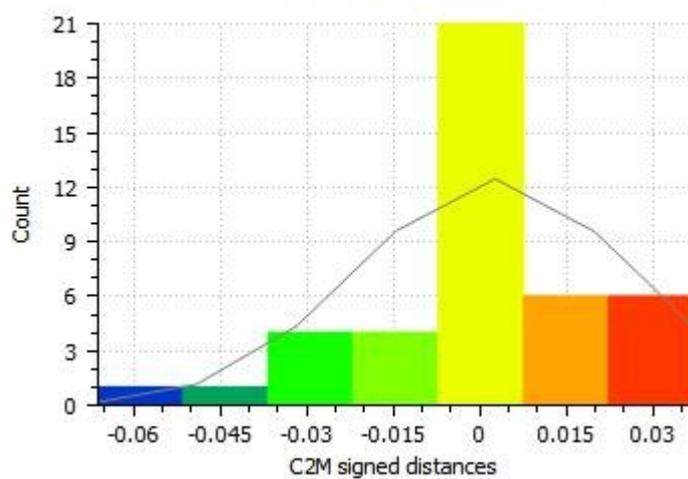
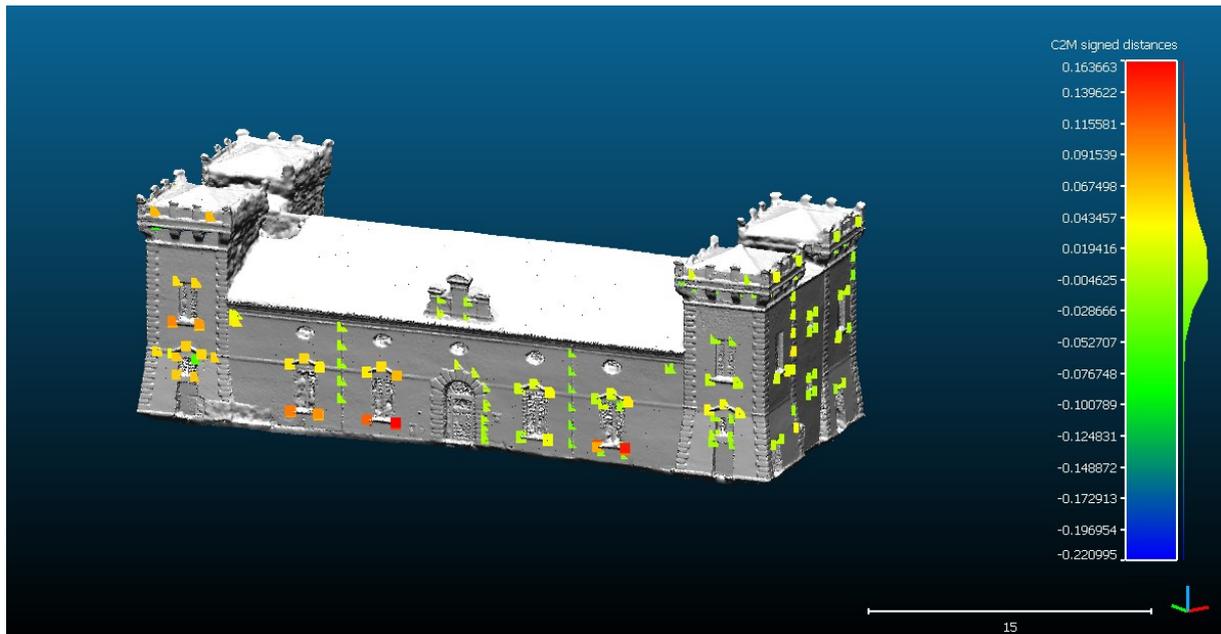


Figura 8.31. In alto si osservano le differenze tra il modello tridimensionale e i punti rilevati con stazione totale. In rosso i punti che presentano maggiori scostamento. Nell'immagine in basso si riporta l'istogramma di frequenza relativo alle differenze misurate e la distribuzione di probabilità calcolata.

MODELLO INTEGRATO: Confronto tra modello 3D e PFA



Gauss: mean = 0.014288 / std.dev. = 0.043403 [16 classes]

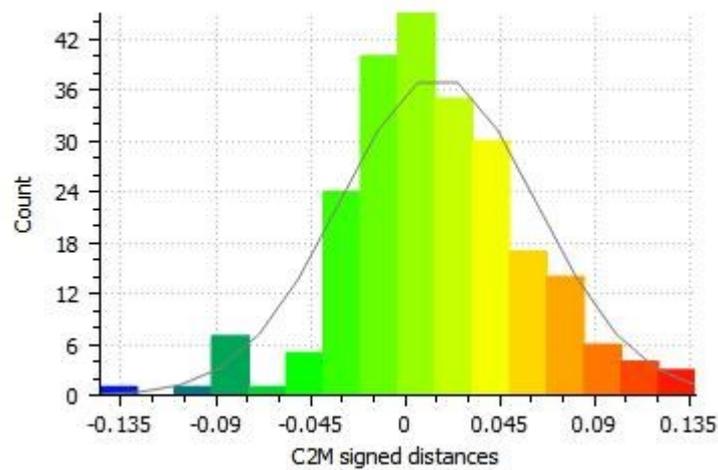


Figura 8.32. In alto si osservano le differenze tra il modello tridimensionale e i punti rilevati con stazione totale. In rosso i punti che presentano maggiori scostamento. Nell'immagine in basso si riporta l'istogramma di frequenza relativo alle differenze misurate e la distribuzione di probabilità calcolata.

Nella seguente tabella si riportano i principali risultati ottenuti:

CONFRONTO TRA MODELLI TRIDIMENSIONALI E PFA		
MODELLO	DIFFERENZE NELLO SPAZIO	
	Media	Deviazione standard
	(m)	(m)
Prospetto OVEST	0.002	0.024
Prospetto EST	-0.001	0.016
Prospetto NORD	0.002	0.007
Prospetto SUD	0.000	0.020
MODELLO INTEGRATO	0.014	0.043

Tabella 8.10. Medie e deviazioni standard dei confronti metrici tra PFA e modelli tridimensionali.

8.4.4 Confronto manuale tra nube di punti fotogrammetrica e punti di controllo.

Per validare quello che è stato fatto con il software CloudCompare sono stati realizzati gli stessi confronti tra nubi dense di punti e PFA in maniera manuale, cioè riconoscendo sulla nube fotogrammetrica il particolare punto e calcolando la differenza con le coordinate rilevate con la stazione totale. La procedura è stata implementata all'interno di un foglio excel, con cui sono state condotte anche le analisi di tipo statistico. Essendo questa volta un confronto per coordinate è possibile separare le tre componenti di errore nelle tre direzioni principali (Est, Nord e Quota). Si riportano i risultati ottenuti ed i grafici delle analisi statistiche.

PROSPETTO OVEST: Confronto manuale tra PFA e nube di densa di punti

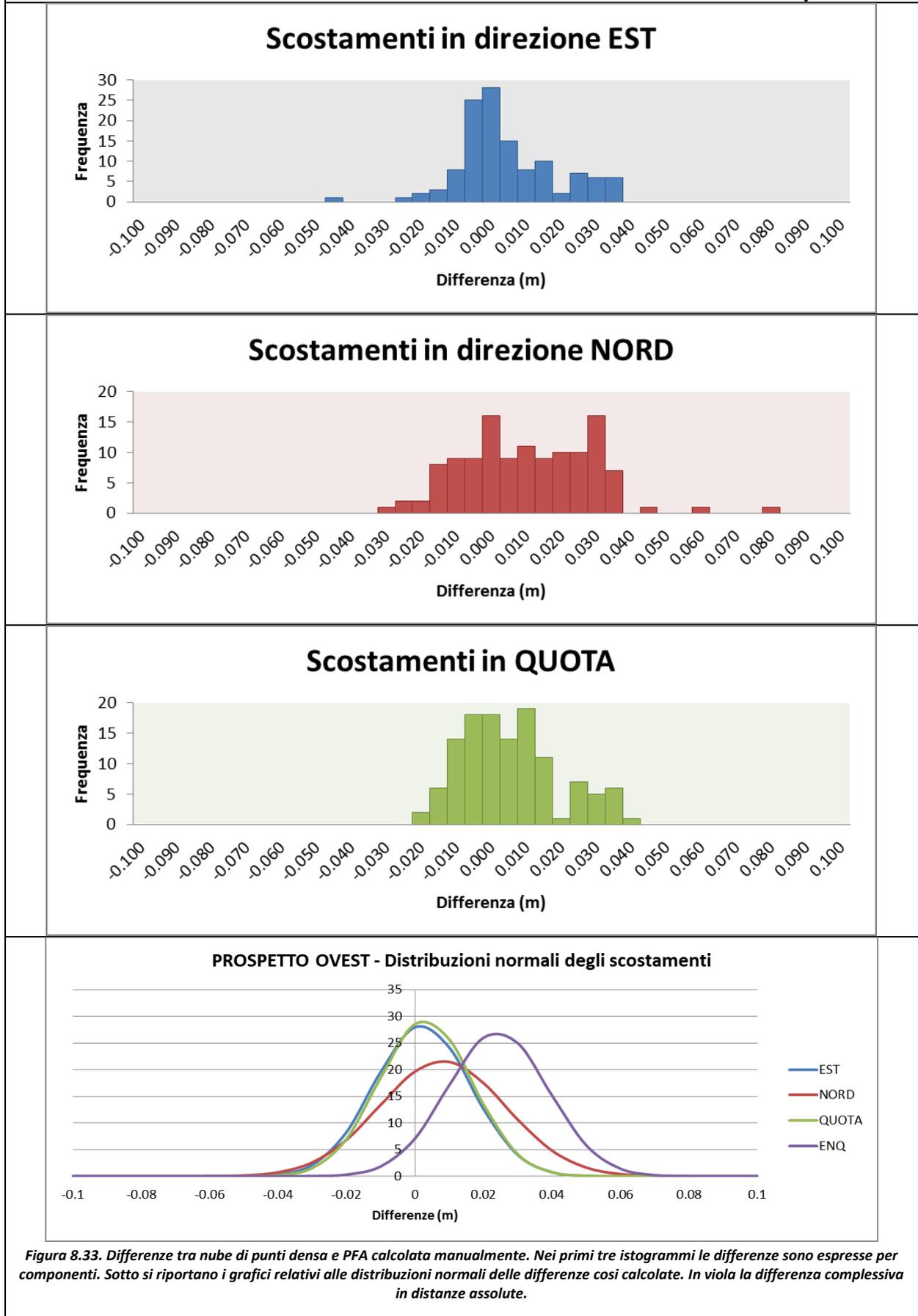


Figura 8.33. Differenze tra nube di punti densa e PFA calcolata manualmente. Nei primi tre istogrammi le differenze sono espresse per componenti. Sotto si riportano i grafici relativi alle distribuzioni normali delle differenze così calcolate. In viola la differenza complessiva in distanze assolute.

PROSPETTO EST: Confronto manuale tra PFA e nube di densa di punti

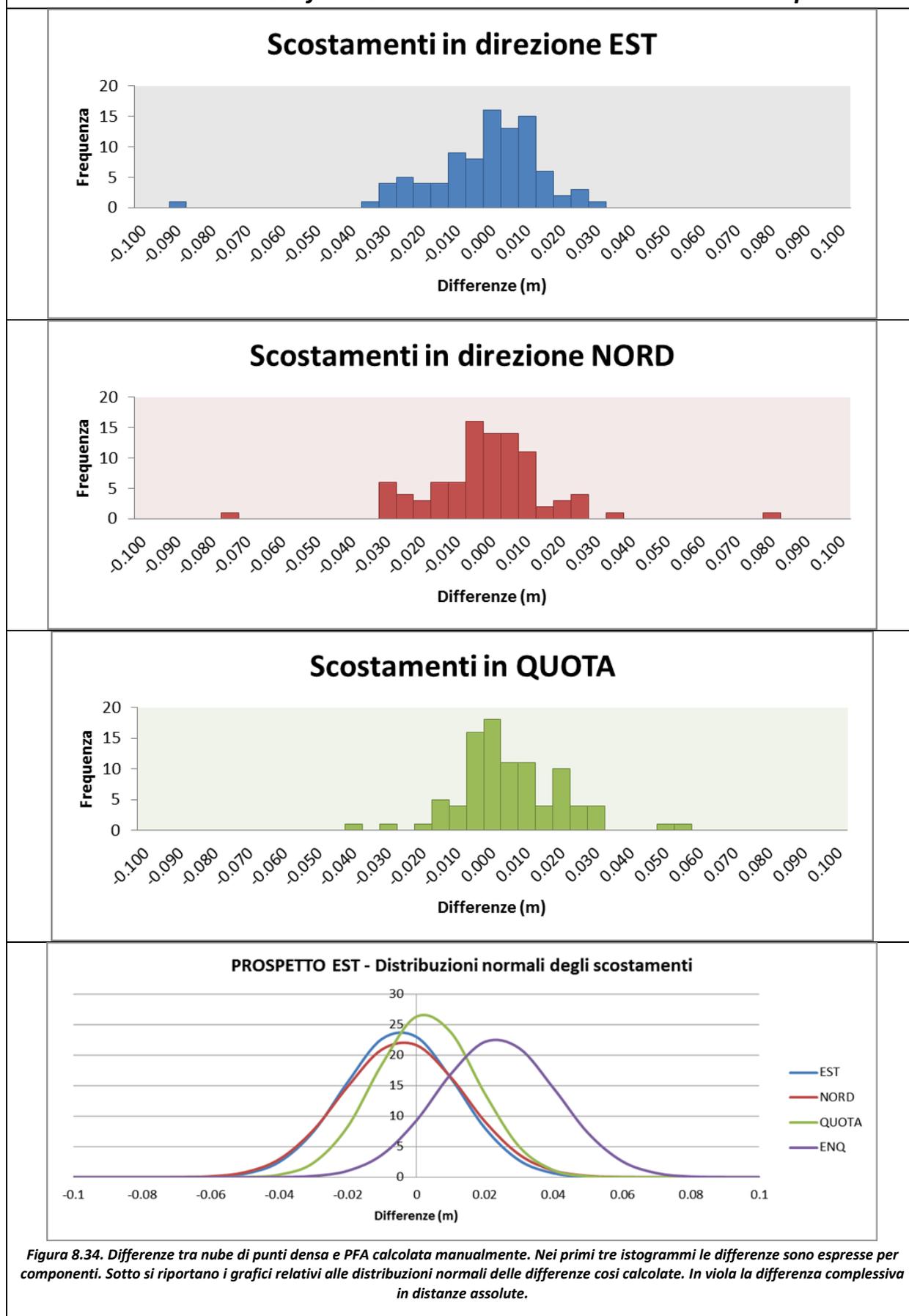


Figura 8.34. Differenze tra nube di punti densa e PFA calcolata manualmente. Nei primi tre istogrammi le differenze sono espresse per componenti. Sotto si riportano i grafici relativi alle distribuzioni normali delle differenze così calcolate. In viola la differenza complessiva in distanze assolute.

PROSPETTO NORD: Confronto manuale tra PFA e nube di densa di punti

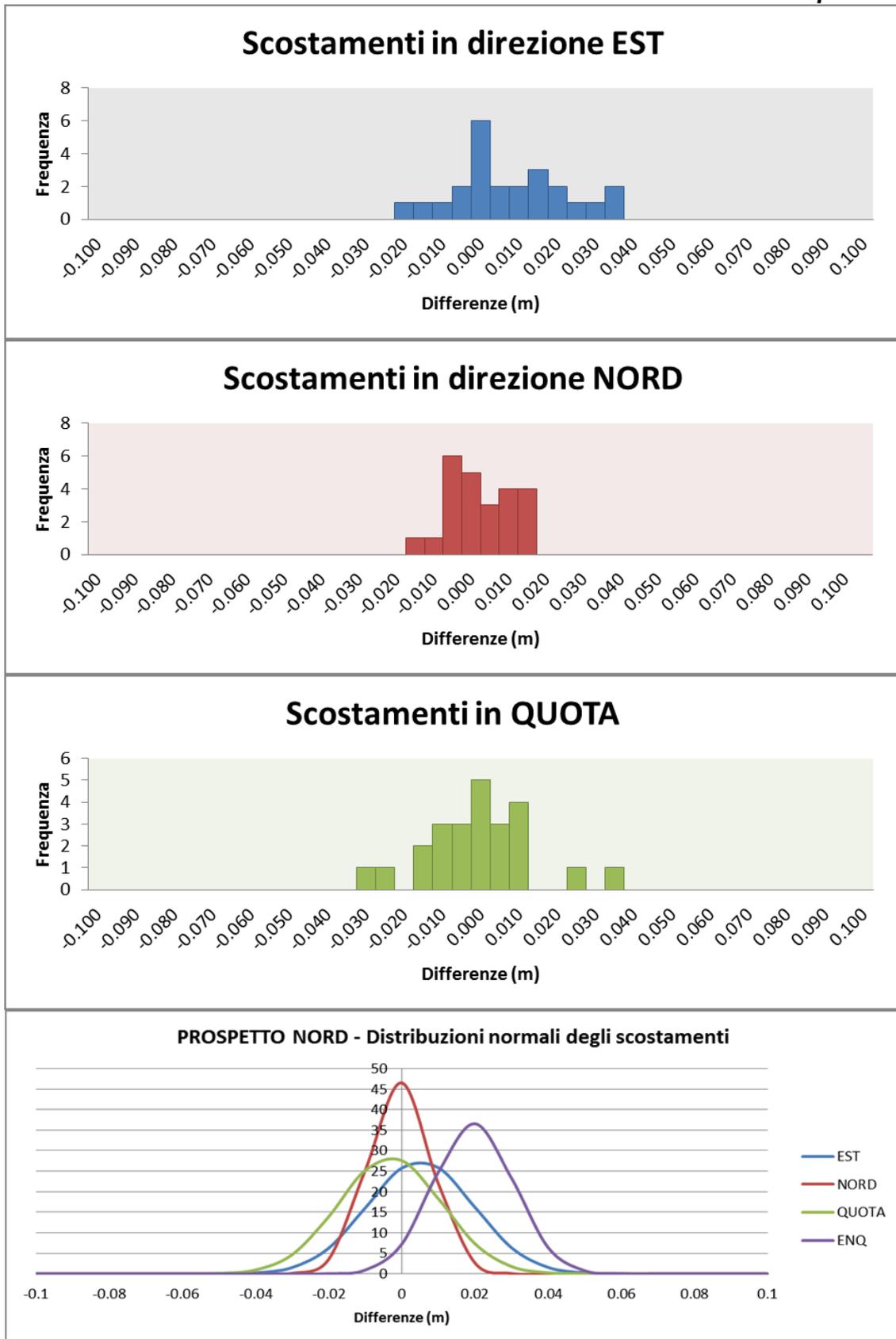


Figura 8.35. Differenze tra nube di punti densa e PFA calcolata manualmente. Nei primi tre istogrammi le differenze sono espresse per componenti. Sotto si riportano i grafici relativi alle distribuzioni normali delle differenze così calcolate. In viola la differenza complessiva in distanze assolute.

PROSPETTO SUD: Confronto manuale tra PFA e nube di densa di punti

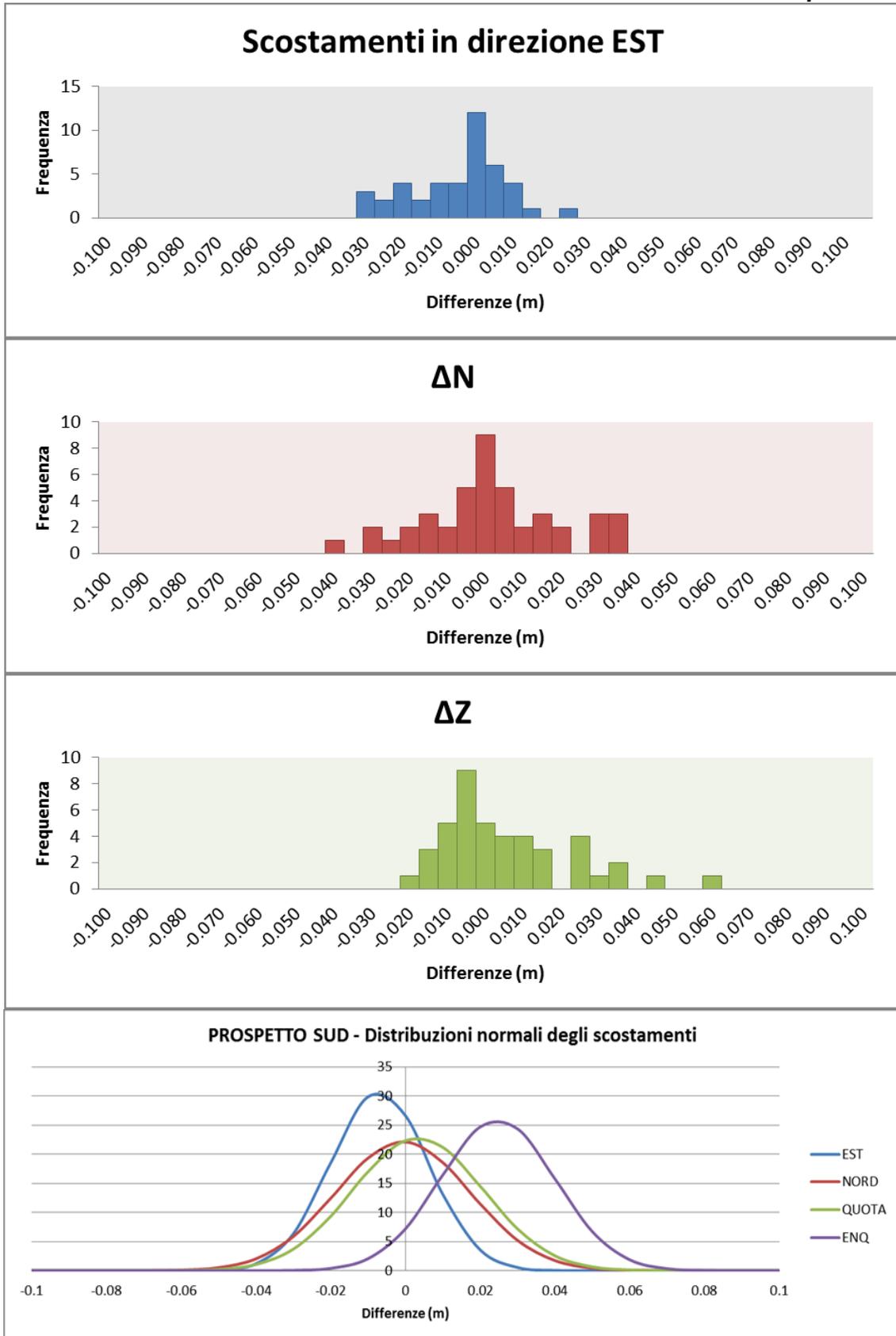


Figura 8.36. Differenze tra nube di punti densa e PFA calcolata manualmente. Nei primi tre istogrammi le differenze sono espresse per componenti. Sotto si riportano i grafici relativi alle distribuzioni normali delle differenze così calcolate. In viola la differenza complessiva in distanze assolute.

TOTALE: Confronto manuale tra PFA e nube di densa di punti

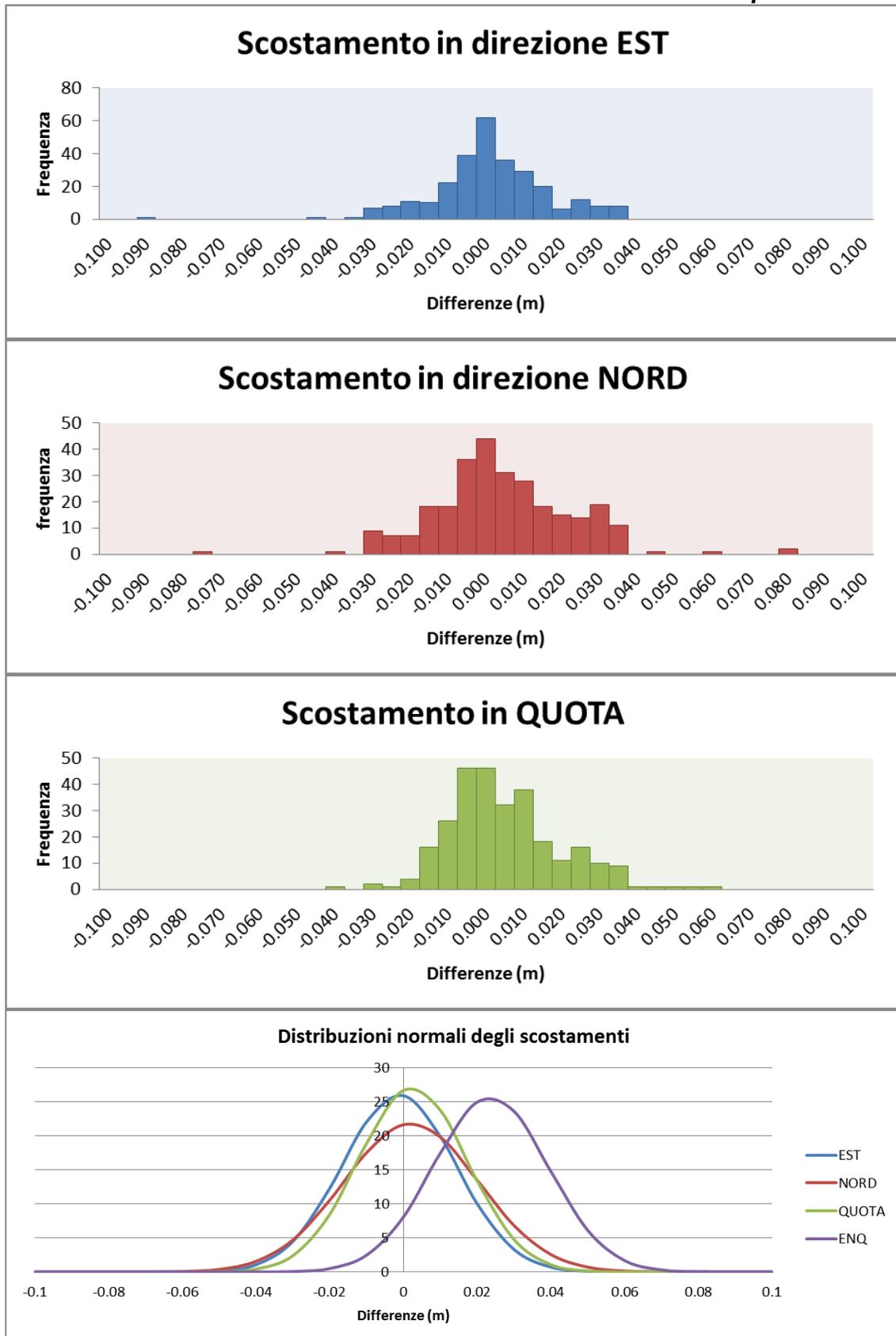


Figura 8.37. Differenze tra nube di punti densa e PFA calcolata manualmente. Nei primi tre istogrammi le differenze sono espresse per componenti. Sotto si riportano i grafici relativi alle distribuzioni normali delle differenze così calcolate. In viola la differenza complessiva in distanze assolute.

Si riportano in tabella 15 i risultati ottenuti.

PROSPETTO OVEST				
	EST	NORD	QUOTA	DIST.ASSOLUTA
MEDIA	0.002	0.008	0.003	0.024
DEV.STANDARD	0.014	0.018	0.014	0.014
PROSPETTO EST				
	EST	NORD	QUOTA	DIST.ASSOLUTA
MEDIA	-0.005	-0.004	0.003	0.023
DEV.STANDARD	0.017	0.018	0.015	0.018
PROSPETTO NORD				
	EST	NORD	QUOTA	DIST.ASSOLUTA
MEDIA	0.005	0.000	0.003	0.020
DEV.STANDARD	0.015	0.008	0.014	0.011
PROSPETTO SUD				
	EST	NORD	QUOTA	DIST.ASSOLUTA
MEDIA	-0.007	0.000	0.003	0.025
DEV.STANDARD	0.013	0.018	0.017	0.015
TOTALE				
	EST	NORD	QUOTA	DIST.ASSOLUTA
MEDIA	-0.001	0.002	0.002	0.024
DEV.STANDARD	0.015	0.018	0.015	0.016

Tabella 8.11. Differenze tra nubi dense fotogrammetriche e PFA calcolati tramite stazione totale, divisi per componenti.

8.5 CONSIDERAZIONI FINALI

Si riportano di seguito tutti i risultati ottenuti attraverso i confronti tra i modelli.

CONFRONTO TRA NUBI DENSE DI PUNTI E SCANSIONI LASER (distanze assolute)		
MODELLO	DIFFERENZE IN VALORE ASSOLUTO NELLO SPAZIO	
	Media	Deviazione standard
	(m)	(m)
Prospetto OVEST	0.019	0.010
Prospetto EST	0.010	0.010
Prospetto NORD	0.011	0.010
Prospetto SUD	0.015	0.010
MODELLO INTEGRATO	0.033	0.026

Tabella 8.12. Medie e deviazioni standard dei confronti metrici tra nubi dense di punti e scansione laser

CONFRONTO TRA SCANSIONE LASER E MODELLI TRIDIMENSIONALI		
MODELLO	DIFFERENZE NELLO SPAZIO	
	Media	Deviazione standard
	(m)	(m)
Prospetto OVEST	-0.012	0.007
Prospetto EST	0.002	0.007
Prospetto NORD	0.001	0.007
Prospetto SUD	-0.008	0.007
MODELLO INTEGRATO	0.007	0.028

Tabella 8.13. Medie e deviazioni standard dei confronti metrici tra modelli 3D e scansione laser

CONFRONTO TRA MODELLI TRIDIMENSIONALI E PFA		
MODELLO	DIFFERENZE NELLO SPAZIO	
	Media	Deviazione standard
	(m)	(m)
Prospetto OVEST	0.002	0.024
Prospetto EST	-0.001	0.016
Prospetto NORD	0.002	0.007
Prospetto SUD	0.000	0.020
MODELLO INTEGRATO	0.014	0.043

Tabella 8.14. Medie e deviazioni standard dei confronti metrici tra modelli 3D e PFA rilevati con stazione totale

CONFRONTO MANUALE TRA MODELLI TRIDIMENSIONALI E PFA								
MODELLO	EST		NORD		QUOTA		ASSOLUTO	
	Media	Dev.St.	Media	Dev.St.	Media	Dev.St.	Media	Dev.St.
	(m)	(m)	(m)	(m)	(m)	(m)	(m)	(m)
Prospetto OVEST	0.002	0.014	0.008	0.018	0.003	0.014	0.024	0.014
Prospetto EST	-0.005	-0.017	-0.004	0.018	0.003	0.015	0.023	0.018
Prospetto NORD	0.005	0.015	0.000	0.008	0.003	0.014	0.020	0.011
Prospetto SUD	-0.007	0.013	0.000	0.018	0.003	0.018	0.025	0.015
TOTALE	-0.001	0.015	0.002	0.018	0.002	0.018	0.024	0.016

Tabella 8.15. Medie e deviazioni standard dei confronti metrici tra modelli 3D e PFA rilevati con stazione totale. Confronto manuale

Le prime due tabelle formalmente esprimono lo stesso risultato, considereremo pertanto solo i risultati emersi dal confronto tra modello tridimensionale e scansione laser, in quanto ritenuto più significativo. Le stime in valore assoluto servono soltanto a definire un ordine di grandezza dello scostamento. In tabella (8.13.) si nota immediatamente come le deviazioni standard dei singoli modelli siano sensibilmente minori rispetto a quella del modello integrato. Questa tendenza conferma anche quanto riscontrato durante la fase di orientamento. I singoli modelli sono stati ottenuti utilizzando soltanto le immagini terrestri, pertanto la miglior precisione è da imputare esclusivamente alla minor distanza di presa ed alle migliori caratteristiche della fotocamera utilizzata. Tra le due probabilmente influisce maggiormente la minor distanza di presa, dal momento che i due sensori sono uguali. Vi sono differenze in termini di ottica e di distorsione che tuttavia potrebbero aver influenzato il risultato delle due modellazioni. Questo aspetto se da un lato è un handicap per il rilievo fotogrammetrico terrestre dall'altro rappresenta un notevole vantaggio per il rilievo aerofotogrammetrico. Il modello integrato presenta infatti una precisione maggiore rispetto ad un modello ottenuto con le sole immagini aeree, come si osserva nell'immagine successiva.

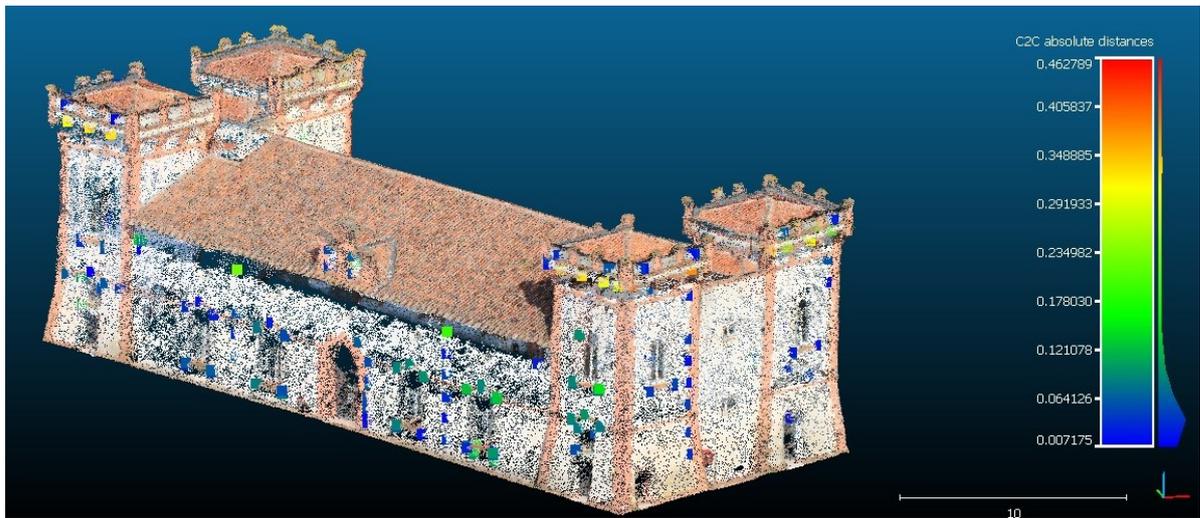


Figura 8.38. differenze tra la nube di punti densa ricavata dalle sole immagini aeree e i punti rilevati con stazione totale. In rosso i punti che presentano maggiori scostamento

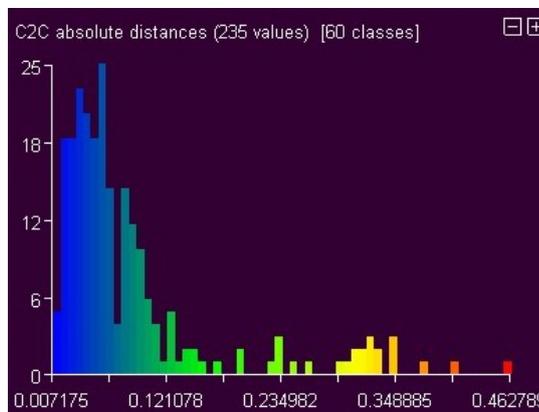


Figura 8.39. Istogramma di frequenza relativo alle differenze misurate.

Confrontando una nube di punti densa realizzata con le sole immagini aeree ed i PFA si ottengono errori dell'ordine di 4-6 cm (sono gli errori che si presentano con maggiore frequenza). Pertanto ***l'integrazione delle immagini terrestri aumenta la precisione del rilievo aerofotogrammetrico.*** Pertanto è possibile affermare che ***minore è la distanza di presa, maggiore sarà il livello di dettaglio e di conseguenza la precisione del rilievo.*** Lo stesso comportamento è osservabile anche per gli altri confronti effettuati. Osservando i risultati del confronto automatico con i PFA si nota un aumento dell'errore. Dai 7 mm di deviazione standard del confronto con la scansione si passa in un intervallo che va dai 7 mm ai 24 mm. Lo stesso avviene per il modello integrato, che mostra uno scostamento dai PFA di 43 mm mentre in precedenza aveva mostrato un errore di circa 28 mm. Questo fatto può essere interpretato osservando proprio i risultati in termini di distanza assoluta ottenuti nella prima elaborazione. Si riporta l'immagine per una miglior comprensione.

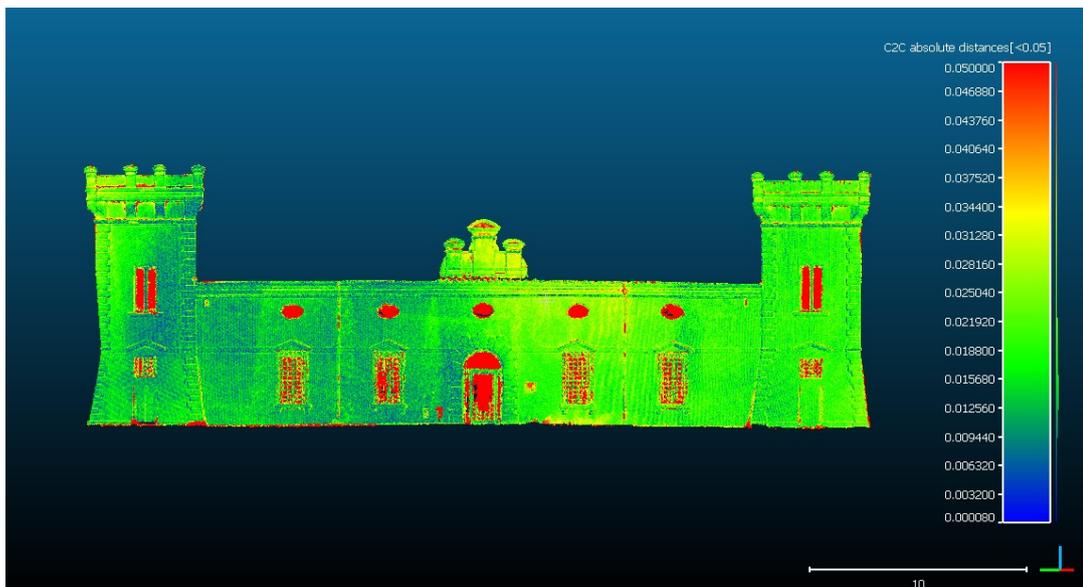


Figura 8.40. Differenze tra nube densa di punti e scansione laser. In rosso i punti che presentano maggiori scostamento. Nell'immagine in basso si riporta l'istogramma di frequenza relativo alle differenze misurate e la distribuzione di probabilità calcolata.

Gli errori maggiori sono sempre in corrispondenza degli elementi vetrati o di repentini cambi di profondità. Merlature, tegole, spigoli sono alcuni degli elementi più soggetti a rumore. Tutto il resto dell'edificio, le zone flat per intenderci sono invece in ottimo accordo con il rilievo laser scanner. I PFA rilevati per definizione devono essere punti ben riconoscibili sulle immagini pertanto in campagna si tende a misurare proprio, spigoli, angoli, finestre, porte, tutti punti soggetti questo tipo di rumore. Questo spiega perché rispetto ai PFA gli errori sono di ordine maggiore.

Elementi sporgenti o piccole rientranze sono rilevate in maniera impeccabile dallo scanner mentre risultano molto più arrotondate se si utilizzano tecniche fotogrammetriche. Nelle immagini seguenti sono state estratte due sezioni in corrispondenza di una torre proprio per osservare questo comportamento. Come si può osservare il rilievo eseguito con lo scanner è complessivamente più pulito.

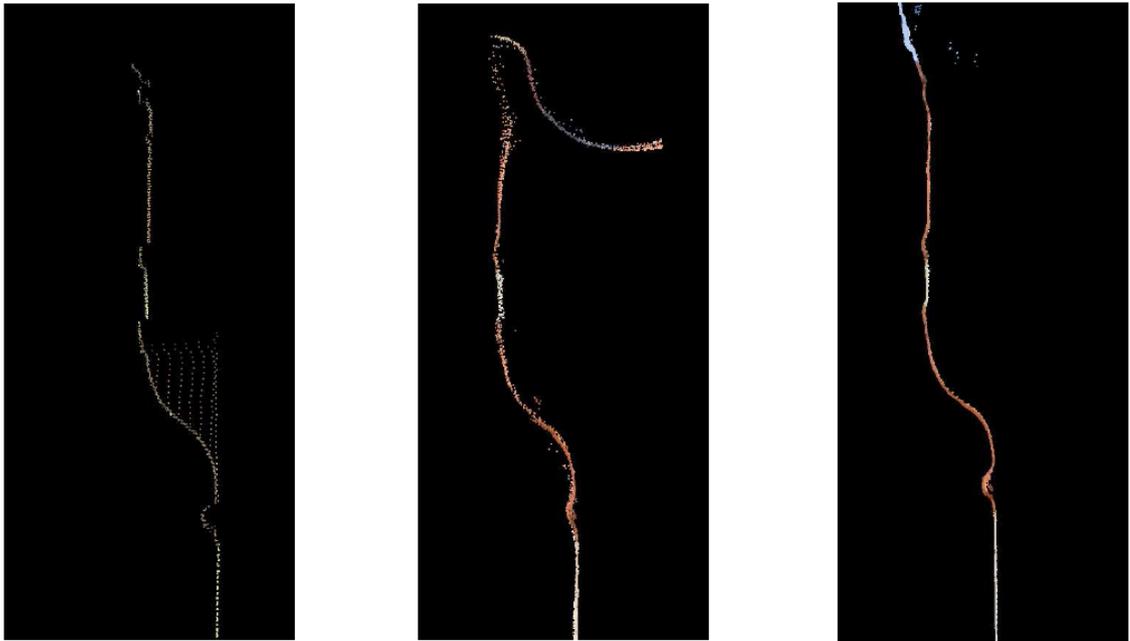


Figura 8.41. Sezioni di una porzione di torre (merlatura presente in sommità) estratta dalla nube rilevata con scansione laser (a sinistra), dalla nube estratta dal modello integrato (al centro) e dalla nube estratta dal modello della singola facciata (a destra).

L'immagine di sinistra ci mostra la sezione estratta dalla scansione laser. La stessa sezione è stata individuata sulla nube ottenuta dal modello integrato e la nube del singolo prospetto. Ancora una volta si ha la dimostrazione di quanto affermato in precedenza. Nonostante sia evidente che la scansione offra risultati migliori, diminuendo la distanza di presa ottengo via via risultati sempre più puliti.

Pertanto **è bene utilizzare le immagini aeree solo dove strettamente necessario in quanto una loro integrazione in un unico modello ne riduce la precisione.** Non sono state rilevate altre evidenti peculiarità da cui poter trarre interessanti considerazioni. Non sono stati evidenziati errori di orientamento o errori di tipo sistematico. Le medie ottenute tramite il confronto diretto con i modelli tridimensionali mostrano infatti una buona distribuzione attorno al valore zero.

Non si dimentichi inoltre che anche tra scansione laser e PFA erano state trovate alcune differenze. Qualsiasi tipo di oggetto se è il risultato di un processo di misurazione ammette un errore, che non può essere eliminato, ma può essere ritenuto accettabile ai fini del rilievo. Le differenze misurate sono da intendersi pertanto come semplici scostamenti tra due tecniche di misura e non come errori in senso assoluto.

Con tali esperienze è stato dimostrato che tra le diverse tecniche vi sono differenze dell'ordine di qualche centimetro, dai 2,6 cm tra fotogrammetria e scansione e 4,3 cm tra fotogrammetria e stazione totale. Diminuendo la distanza di presa tali differenze si riducono, scendendo al di sotto dei 2,4 cm (si hanno 7 mm costanti nei confronti tra fotogrammetria e scansione per i soli rilievi

terrestri). Qualora la geometria o le caratteristiche dell'edificio lo richiedano è possibile variare la distanza di presa così da coglierne ogni piccolo dettaglio.

Piuttosto che realizzare un modello integrato utilizzando tutte le immagini è consigliabile utilizzare modelli diversi , così da mantenere le precisioni raggiunte. Eventualmente si possono unire i risultati sovrapponendo le nubi di punti ottenute singolarmente.

In definitiva sembra ***lecito affermare che la tecnica fotogrammetrica, applicata secondo le modalità illustrate in questa tesi, è notevolmente accurata*** poiché non mostra errori sistematici significativi mentre presenta una precisione stimabile mediamente intorno in 2-3 cm, come risulta dalla deviazione standard dei risultati. Precisioni assolutamente compatibili con le normali esigenze del rilevamento dei beni culturali.

Naturalmente l'esperienza non è esaustiva ed è necessario condurre altri test, variando le modalità di acquisizione del dato e utilizzando diversi parametri nel software di restituzione.

9 IL POTENZIALE DI TECNOLOGIE LOW-COST APPLICATE AL RILIEVO DEI BENI CULTURALI

9.1 INTRODUZIONE

I sistemi low-cost destano da sempre un grande interesse in tutte le discipline e spesso costituiscono una vera e propria rivoluzione in un particolare settore tecnologico. Nell'ambito del rilievo fotogrammetrico questo ruolo negli ultimi anni è ricoperto dai sistemi RPAS (Remotely Piloted Aircraft Systems) abbinati a software commerciali di fotogrammetria digitale. La loro accessibilità e relativa facilità d'uso li rende uno strumento di misura molto competitivo ed utilizzabile anche da professionisti e tecnici non strettamente appartenenti al mondo del rilevamento fotogrammetrico classico. Così come nel primo caso di studio osservata rimane il solito problema di fondo: "che risultati si possono raggiungere con RPAS di tipo low-cost?". Di nuovo per poter analizzare le potenzialità ed i limiti dei sistemi low-cost nel rilievo architettonico dei Beni Culturali sono stati condotti diversi test avvalendosi di un piccolo RPAS equipaggiato con una camera ultraleggera, acquisendo numerose immagini sia con assi di presa nadirali che inclinati. Oltre alle immagini sono stati acquisiti anche diversi filmati da cui sono stati estratti numerosi frames per poter poi processare anche quelli come dato fotogrammetrico.

Il test site su cui è stata condotta la ricerca è di nuovo la "Delizia Estense del Verginese". Utilizzando sempre il software commerciale Photoscan di Agisoft sono stati creati alcuni modelli ciascuno dei quali è stato confrontato con l'ormai classico modello di riferimento (§7.2.1.). Ancora si sono valutate le differenze esistenti tra i modelli e avvalendosi di un'analisi di tipo statistico è stato possibile valutare la precisione del rilievo. Si osservano di seguito le fasi principali del lavoro di ricerca condotto.

9.2 STRUMENTAZIONE UTILIZZATA

Il sistema low-cost utilizzato per la sperimentazione eseguita si compone di un piccolo RPAS ready to fly di tipo commerciale equipaggiato con una fotocamera ultra leggera di nuova generazione. Si tratta nel dettaglio del *Dji Phantom 2* (Fig.9.1.) equipaggiato con una *GoPro Hero 3+ Black Edition* con la quale sono stati acquisiti sia i filmati che le immagini digitali.



Figura 9.1. Remotely Piloted Aircraft Systems (RPAS) utilizzato nella sperimentazione. Dji Phantom 2 equipaggiato con una fotocamera ultraleggera GoPro Hero 3+ Black Edition.

Tramite la GoPro è stato possibile effettuare video di qualità pari a 1080p, da cui sono stati estratti i fotogrammi per l'elaborazione, e prese fotografiche "continue", ovvero opportunamente intervallate con regolazione tramite la funzione "Time Lapse" che consente di impostare il tempo di scatto tra una foto e la successiva. La scelta della fotocamera è stata dettata dalla sua recente popolarità e dal fatto che è perfettamente compatibile sia con le caratteristiche del drone che con l'obiettivo della sperimentazione. Si tratta infatti di una fotocamera ultraleggera che può tranquillamente essere montata sul piccolo velivolo attraverso un apposito gimbal. Il suo peso contenuto non comporta di fatto alcun problema riguardo la capacità di carico e il suo basso costo la rende accessibile a tutti. Se da un lato quindi fotocamere di questo tipo presentano notevoli vantaggi, da un punto di vista ottico non sono ideali per elaborazioni fotogrammetriche. Presentano infatti un obiettivo tipo "fisheye" che comporta distorsioni molto elevate, che devono essere opportunamente corrette prima di iniziare la fase di restituzione.

9.3 PIANIFICAZIONE DEL RILIEVO ed ACQUISIZIONE DEL DATO

Per le riprese tramite RPAS, sia video che fotografiche sono stati adottati gli stessi accorgimenti utilizzati nel primo caso di studio. È stato eseguito un volo con assi di presa nadirali per il rilievo del coperto, mentre per il rilievo delle singole facciate, sono state effettuate una serie di strisciate con assi di presa perpendicolari alle pareti dell'edificio. Considerando l'ampio campo di vista della camera è stato deciso di volare ad una quota di volo relativamente più bassa rispetto a quella utilizzata nel primo caso di studio anche per ottenere un GSD compatibile con l'oggetto da rilevare. Va osservato infatti che la risoluzione della GoPro (4000x3000) è minore rispetto a quella della Canon EOS 7D (12

megapixel contro i 18 della Canon). Ancora una volta è stata scelta una sovrapposizione longitudinale dell'80% ed una trasversale del 60%. Nelle figure seguenti un esempio della strategia di presa.

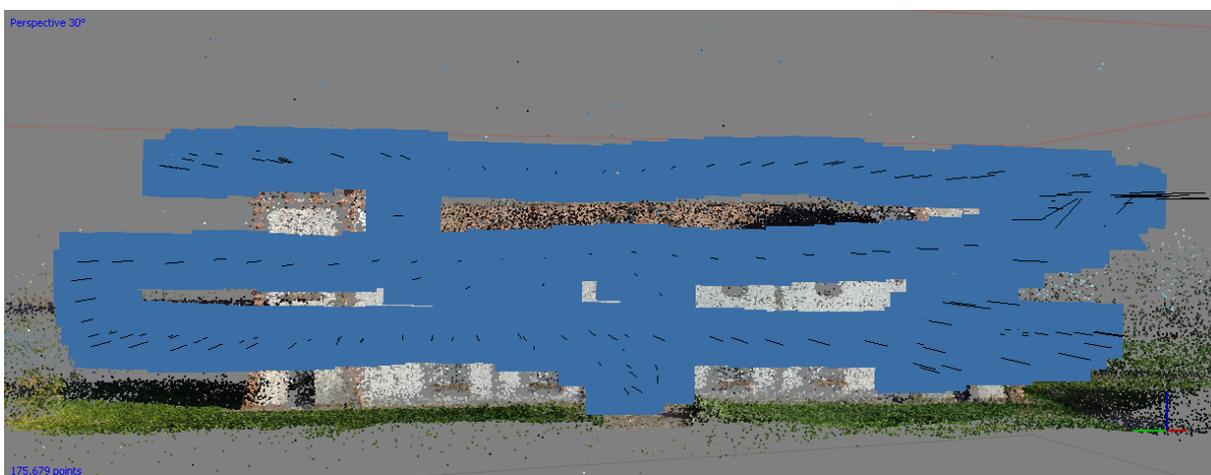
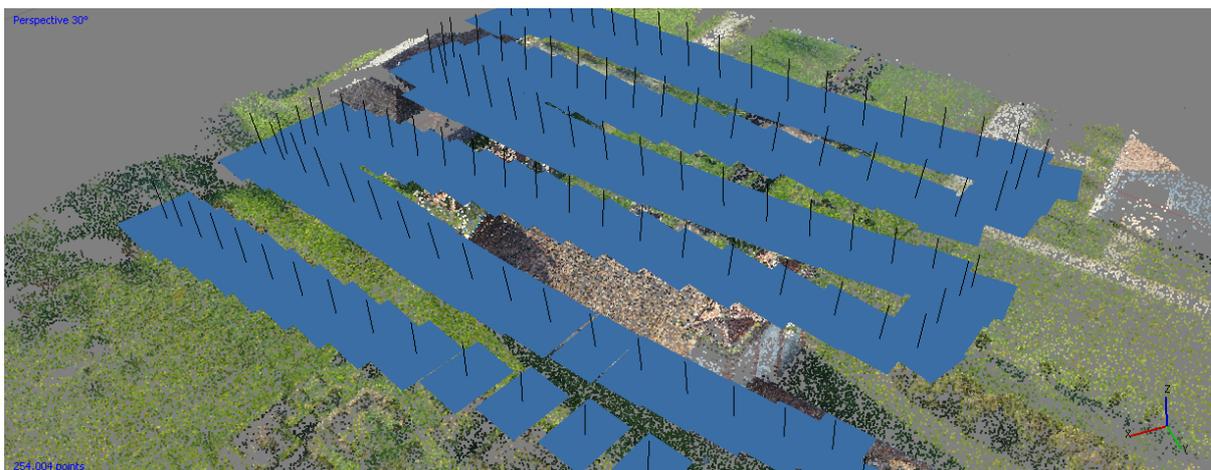


Figura 9.2. Esempi di presa fotogrammetrica. In alto la strategia di presa per immagini acquisite ad assi nadirali. Sotto La strategia di presa per immagini acquisite ad assi orizzontali in corrispondenza delle facciate del fabbricato.

Di nuovo presso gli spigoli del fabbricato sono stati eseguiti spostando il drone sulla sua verticale alcuni voli con lo scopo di acquisire prese d'insieme per facilitare l'orientamento complessivo dei fotogrammi nelle zone d'angolo. In totale sono state acquisite 802 immagini di cui 162 con asse di presa nadirale e 640 con assi di presa perpendicolare alle facciate.

TIPOLOGIA DI VOLO	IMMAGINI ACQUISITE	IMMAGINI UTILIZZATE	GSD (mm)
Assi Nadirali	162	114	12,8
Prospetto OVEST	191	175	4,0
Prospetto EST	168	152	3,2
Prospetto NORD	141	124	3,2
Prospetto SUD	140	105	2,7

Tabella 9.1. Report del rilievo eseguito.



Figura 9.3. Immagini acquisite con GoPro. Da notare l'elevata distorsione.

Anche l'acquisizione dei video ha seguito il medesimo piano di volo. Sono poi stati estratti un totale di 1650 frames con una risoluzione di 1920x1080.

9.4 ELABORAZIONE DEL DATO (IMMAGINI FOTOGRAFICHE)

Osservando la fig.9.3. si nota che la distorsione generata dall'obiettivo della camera è molto elevata. Sorge dunque il dubbio che tale distorsione possa comportare errori durante la fase di elaborazione del dato, in particolare durante la fase di autocalibrazione da parte del software. Inserendo le immagini all'interno di Photoscan e lanciando una primissima procedura di orientamento si nota immediatamente la presenza degli effetti di tale distorsione (in questo primo approccio al problema l'orientamento delle immagini è avvenuto con le impostazioni standard del software). Si osserva in particolare il classico "bowl-effect".

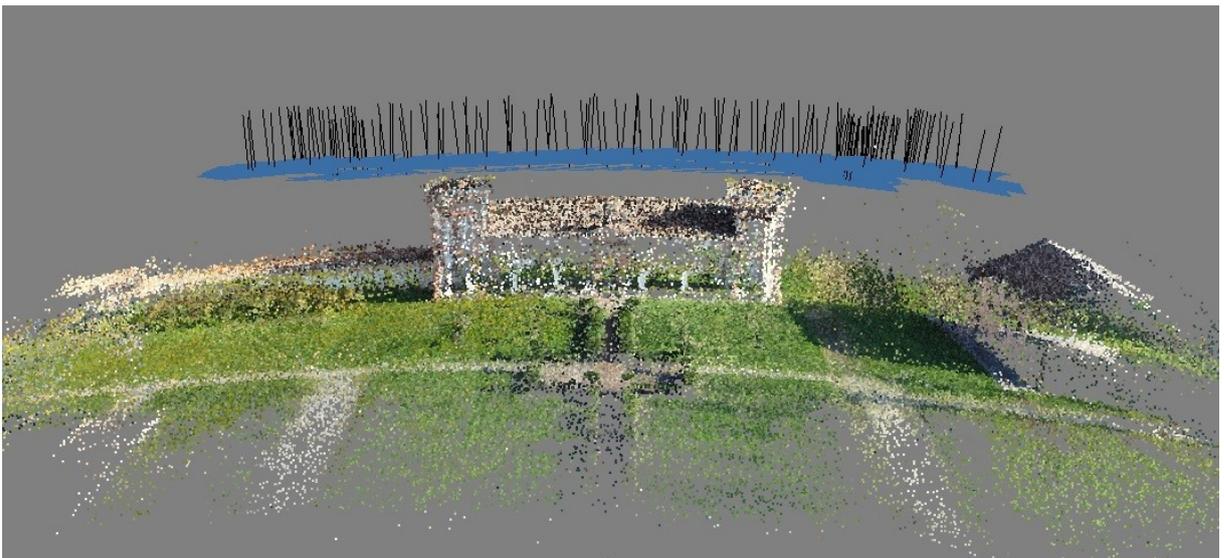


Figura 9.4. Primo orientamento realizzato con i parametri di default del software. La correzione della distorsione non avviene in maniera corretta. Sia nell'orientamento relativo che nella restituzione si apprezza il cosiddetto "bowl-effect".

Pertanto è opportuno cercare di correggere il più possibile questo aspetto anche con procedure esterne al processo di restituzione tramite Photoscan. Per valutare il miglior processo di restituzione sono stati esaminati tre differenti approcci alla calibrazione della camera.

- Calibrazione eseguita con il software Photomodeler ed inserimento in Photoscan dei parametri interni calibrati. Il software di Agisoft utilizzerà poi tali parametri come punto di partenza per una migliore autocalibrazione.
- Eliminazione della distorsione dalle immagini ed inserimento in Photoscan delle immagini già corrette.
- Procedura di autocalibrazione di Photoscan con l'utilizzo della procedura implementata per gli obiettivi tipo "fisheye".

9.4.1 Modelli realizzati con self-calibration (Photomodeler)

Il primo caso che andremo a studiare è quello relativo alla procedura esterna di self-calibration. Utilizzando un pannello di calibrazione apposito è stato possibile stimare i parametri interni della fotocamera. Il risultato della calibrazione è riportato nella tabella seguente:

	PhotoModeler	Convertite per PhotoScan
Formato foto	4000x3000 [px]	4000x3000 [px]
Focal Length	2.742 [mm]	2.742 [mm]
Xp (principal point x)	3.154 [mm]	2014.55 [px]
Yp (principal point y)	2.325 [mm]	148.74 [px]
Fw (format width)	6.265 [mm]	1702.38 [px]
Fh (format height)	4.699 [mm]	1701.23 [px]
K1 (radial distortion 1)	0.0392300	-0.191453
K2 (radial distortion 2)	0.0006149	0.034744
K3 (radial distortion 3)	0.0001640	-0.003290
P1 (decentering distortion 1)	-0.0002634	0.000001
P2 (decentering distortion 2)	-0.0001167	0.000053
Pixel size	1.566 [µm]	1.566 [µm]

Tabella 9.2. Parametri di orientamento calcolati per la camera Go Pro Hero 3+

È stato necessario apportare una piccola correzione ai parametri sopraelencati in quanto Photoscan ammette l'inserimento di tali parametri espressi in pixel mentre in output da Photomodeler si ottengono in (mm). Una volta inseriti i parametri all'interno del software si può procedere con l'orientamento dei fotogrammi. In figura si osserva il corretto posizionamento delle camere e la sparizione del "bowl-effect".

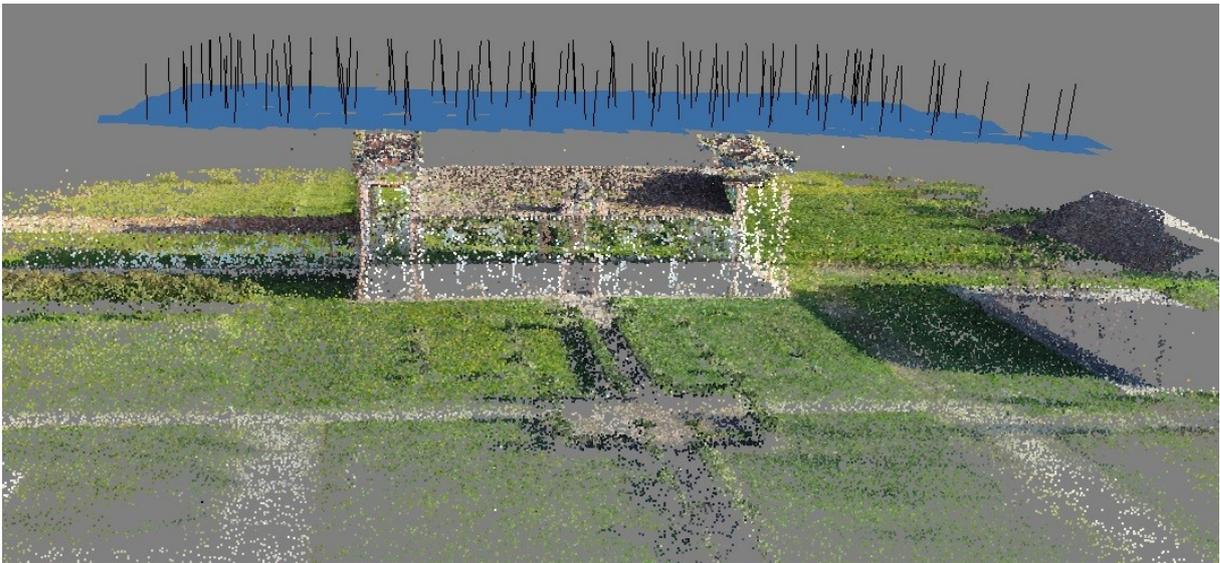


Figura 9.5. Orientamento ottenuto a partire dai dati di self-calibration inseriti all'interno del software. L'effetto "bowl-effect" non è più visibile.

Inserendo i GCP è possibile dare una valutazione metrica del risultato dell'orientamento. In tabella si riportano i residui ottenuti:

MODELLO	FOTOGRAFIE	GCP/PFA	ERROR [m]	ERROR [pix]
Assi Nadirali	114	8	0.057	0.778
Prospetto OVEST	175	8	0.021	7.546
Prospetto EST	152	8	0.022	11.814
Prospetto NORD	124	8	0.022	2.419
Prospetto SUD	105	8	0.023	6.942

Tabella 9.3. Report della procedura di orientamento.

In termini di posizionamento assoluto si ottengono residui non esageratamente elevati ma se osserviamo l'errore sulla collimazione dei punti (pixel) ci si accorge subito che c'è qualche problema. Probabilmente è ancora presente qualche effetto della distorsione e le proiezioni dei punti sulle varie immagini ne risentono.

Tuttavia si decide di procedere con l'elaborazione della nube di punti, così da poter quantificare meglio la portata di questi errori con il confronto diretto con il modello di riferimento. Al solito è stata creata una nube di punti per ciascun elemento ed un modello integrato. Riportiamo per semplicità solo alcune immagini relative ai modelli, per avere una idea qualitativa del prodotto ottenuto.



Figura 9.6. Nube di punti densa realizzata utilizzando le immagini acquisite delle singole facciate.



Figura 9.7. Nube di punti densa del modello integrato.

9.4.2 Modelli realizzati con immagini ricampionate senza distorsione

Il secondo approccio che verrà valutato è l'utilizzo di immagini ricampionate, in cui è stata eliminata la distorsione.



Figura 9.8. Immagine ricampionata in cui è stata corretta la distorsione.

Allo stesso modo si procede con la procedura di orientamento e la creazione delle nubi dense che successivamente saranno confrontate con la nube di riferimento.

Si riportano a titolo di esempio soltanto i risultati della procedura di orientamento e alcune immagini delle nubi.

MODELLO	FOTOGRAFIE	GCP/PFA	ERROR [m]	ERROR [pix]
Assi Nadirali	114	8	0.021	0.253
Prospetto OVEST	175	8	0.007	0.191
Prospetto EST	152	8	0.004	0.095
Prospetto NORD	124	8	0.004	0.200
Prospetto SUD	105	8	0.003	0.168

Tabella 9.4. Report della procedura di orientamento

Si nota subito che utilizzando tale procedura i residui sono sensibilmente diminuiti sia sul posizionamento 3D dei singoli GCP che sulle proiezioni nelle singole immagini. Da questo punto di vista questo approccio è da preferire al caso precedente.

La creazione delle nubi dense ha portato i seguenti risultati:



Tabella 9.5. Dettaglio della nube di punti ottenuta con le immagini ricampionate acquisite ad assi nadirali

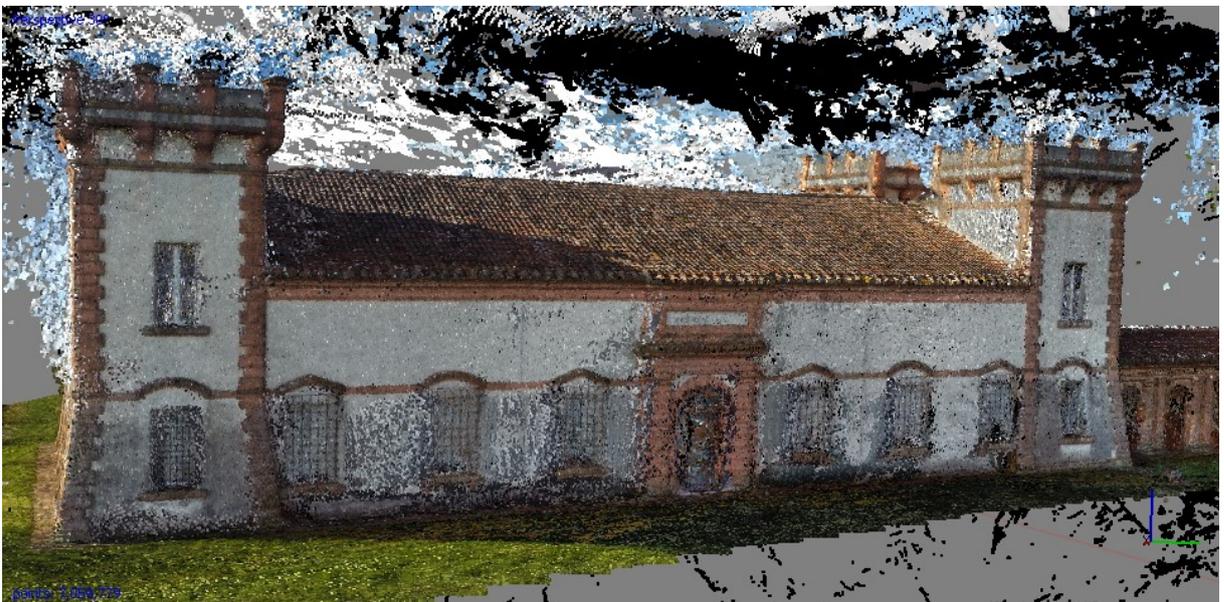


Tabella 9.6. Dettaglio della nube di punti ottenuta con le immagini ricampionate acquisite ad assi orizzontali in corrispondenza delle singole facciate

Nonostante l'orientamento dei fotogrammi abbia dato risultati migliori i modelli creati risultano qualitativamente peggiori. Si riscontra rumore, e zone non restituite. Questo probabilmente è dovuto al fatto che si stanno utilizzando immagini ricampionate.

9.4.3 Modelli realizzati con la procedura di autocalibrazione “fisheye” di Photoscan

Questo tipo di elaborazione sfrutta una delle funzionalità del software, in particolare la correzione della distorsione causata da obiettivi tipo fisheye. Tramite il menu a tendina del *Camera type* si cambia l'impostazione di base *Frame* (utilizzata per foto realizzate con obiettivi comuni) con quella *Fisheye*; grazie a questa scelta il programma sarà ora in grado di gestire i dati ricavati dalle fotografie tramite appositi algoritmi. Impostata la tipologia di camera è possibile finalmente far partire la procedura di orientamento dei fotogrammi (fig.9.9.).

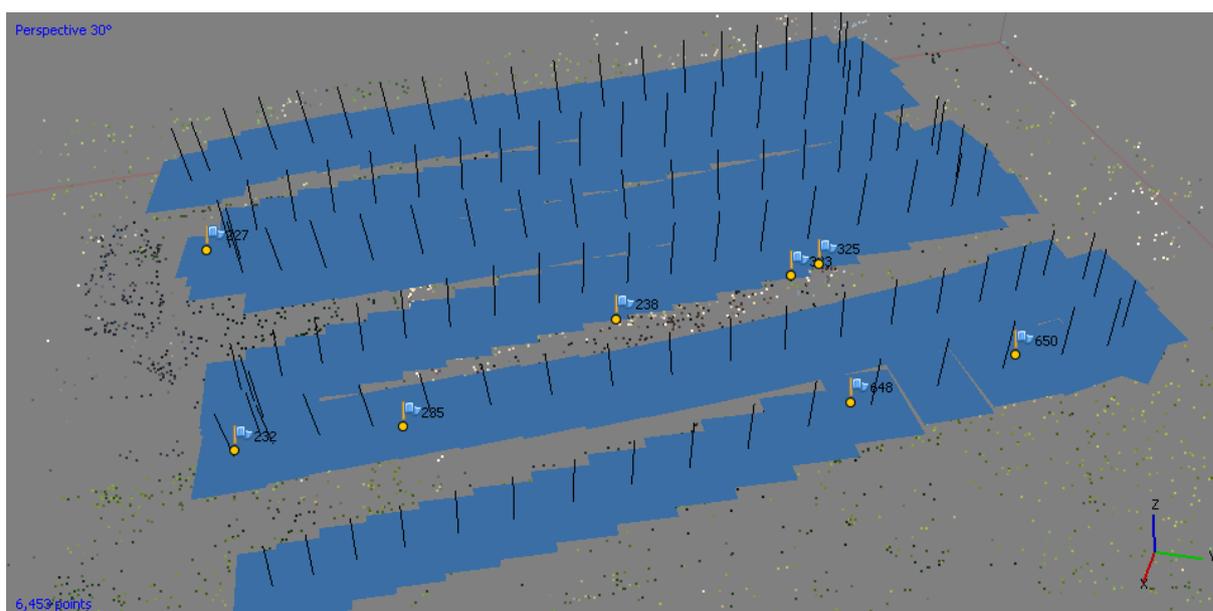


Figura 9.9. Distribuzione dei GCP sul modello in fase di orientamento

A causa dell'esiguo numero di punti utilizzati per l'allineamento, non è possibile valutare qualitativamente l'eliminazione dell'effetto di sfericità introdotto dalla distorsione fisheye; pertanto, anche in questo caso, sono stati introdotti i GCPs necessari a georeferenziare l'intero allineamento. Si nota immediatamente come l'errore risulti molto contenuto ed in particolare esso risulta assumere il valore più basso in assoluto, se confrontato con quello ottenuto nelle elaborazioni precedenti. Lo stesso procedimento è stato applicato agli altri blocchi di fotogrammi relativi alle facciate della delizia. In tabella si riportano i risultati delle procedure di orientamento.

MODELLO	FOTOGRAFIE	GCP/PFA	ERROR [m]	ERROR [pix]
Assi Nadirali	114	8	0.014	0.207
Prospetto OVEST	175	8	0.008	0.357
Prospetto SUD	105	8	0.005	0.475
Prospetto EST	152	8	0.007	0.222
Prospetto NORD	124	8	0.006	0.362

Tabella 9.7. Report della procedura di orientamento.

Allo stesso modo saranno creati un modello per ogni singola facciata e dal momento che con tale metodologia si sono ottenuti i risultati migliori si procede anche alla creazione di un modello integrato elaborando nuovamente tutte le immagini rilevate da RPAS.

MODELLO	FOTOGRAFIE	GCP/PFA	ERROR [m]	ERROR [pix]
Modello Integrato	670	40	0.014	0.364

Tabella 9.8. Report della procedura di orientamento del modello integrato



Figura 9.10. Immagine della nube densa di punti del modello integrato ottenuta a partire da 670 immagini.

9.5 VALUTAZIONE DELLA PRECISIONE (IMMAGINI FOTOGRAFICHE)

Nonostante le procedure di orientamento e creazione delle nubi dense ci indirizzino verso una conclusione scontata (il metodo implementato in Photoscan per il trattamento delle immagini fisheye sia il migliore), eseguiremo comunque i confronti tra tutti i modelli ottenuti ed il modello di riferimento. In particolare verranno eseguiti come nel caso di studio 1 i confronti diretti tra le nubi per avere un'idea dello scostamento assoluto e un confronto tra superfici solide e nubi di riferimento per poter trarre considerazioni riguardo la sistematicità o meno dell'errore. Dai primi

confronti verranno estratti gli istogrammi di frequenza, mentre dal secondo una distribuzione di probabilità di tipo Gaussiano.

Per i primi due casi si riportano le immagini relative ai confronti realizzati con il solo modello integrato titolo di esempio, mentre per la procedura di autocalibrazione "fisheye" saranno riportate tutte le elaborazioni studiate. Al termine i dati di tutti i confronti saranno raccolti in un'unica tabella.

9.5.1.1 Modelli realizzati con self-calibration (Photomodeler)

MODELLO INTEGRATO: Confronto tra nube densa di punti e scansione laser

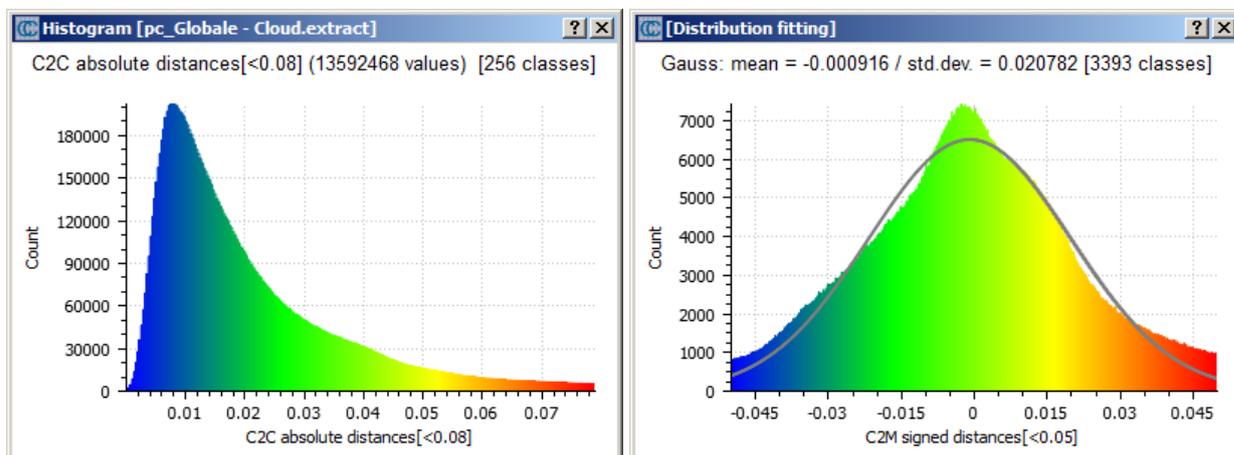
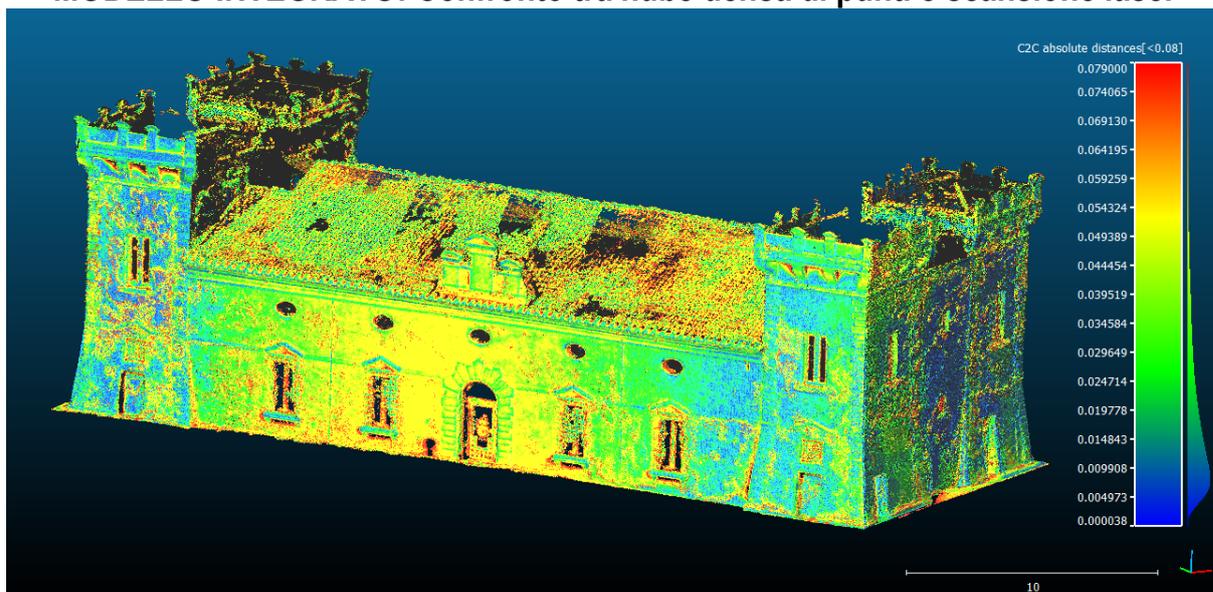


Figura 9.11. Differenze misurate tra nubi dense di punti e scansione laser raffigurate mediante mappa di colore. In rosso le zone di maggior scostamento. Sotto si riportano gli istogrammi di frequenza in distanze assolute (a sinistra) e non (a destra). Considerando i segni reali degli scostamenti è possibile costruirne la distribuzione normale di probabilità.

9.5.1.2 Modelli realizzati con immagini ricampionate senza distorsione

MODELLO INTEGRATO: Confronto tra nube densa di punti e scansione laser

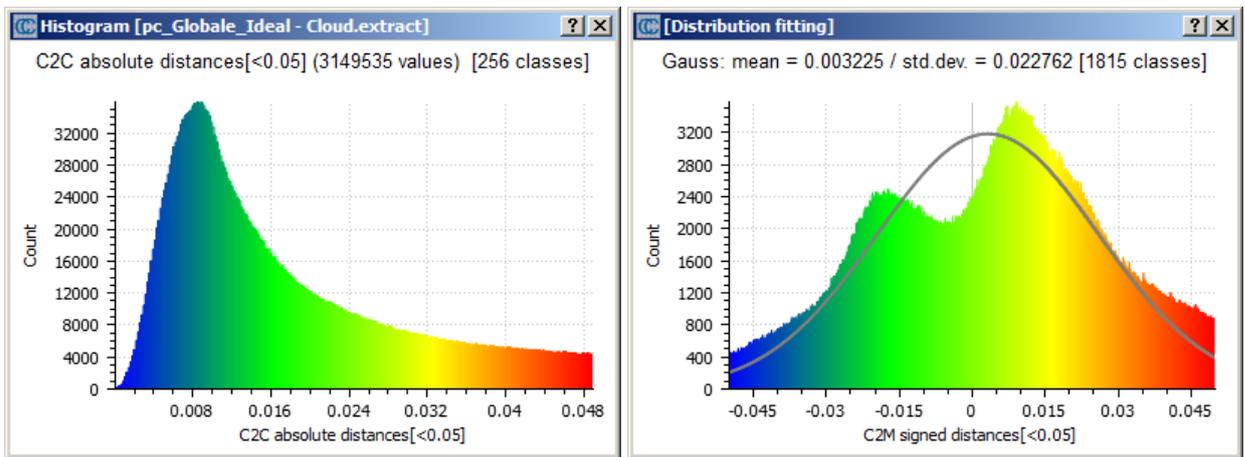
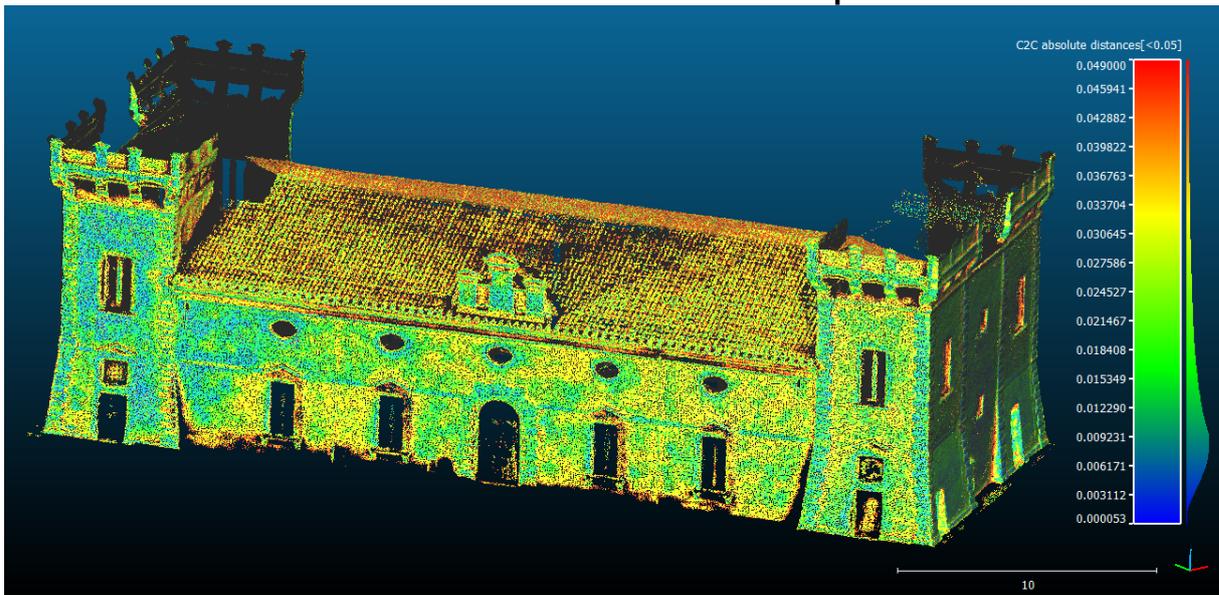


Figura 9.12. Differenze misurate tra nubi dense di punti e scansione laser raffigurate mediante mappa di colore. In rosso le zone di maggior scostamento. Sotto si riportano gli istogrammi di frequenza in distanze assolute (a sinistra) e non (a destra). Considerando i segni reali degli scostamenti è possibile costruirne la distribuzione normale di probabilità.

9.5.1.3 Modelli realizzati con la procedura di autocalibrazione "fisheye" di Photoscan

PROSPETTO OVEST: Confronto tra nube densa di punti e scansione laser

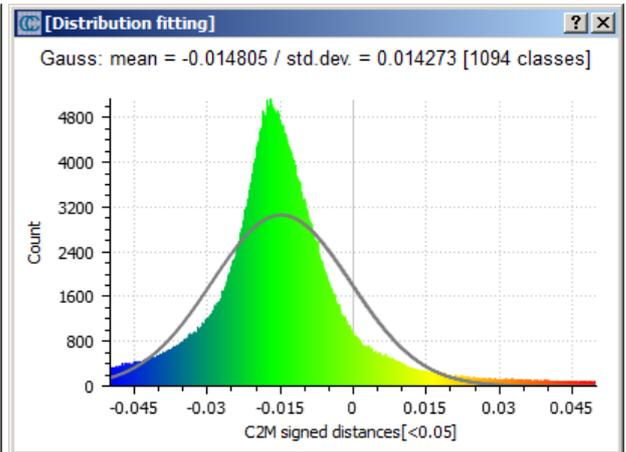
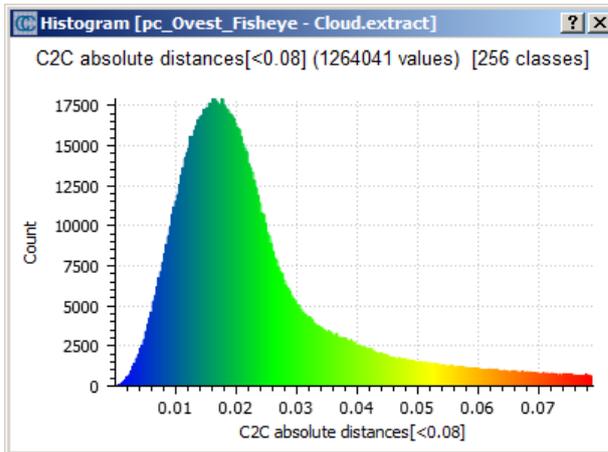
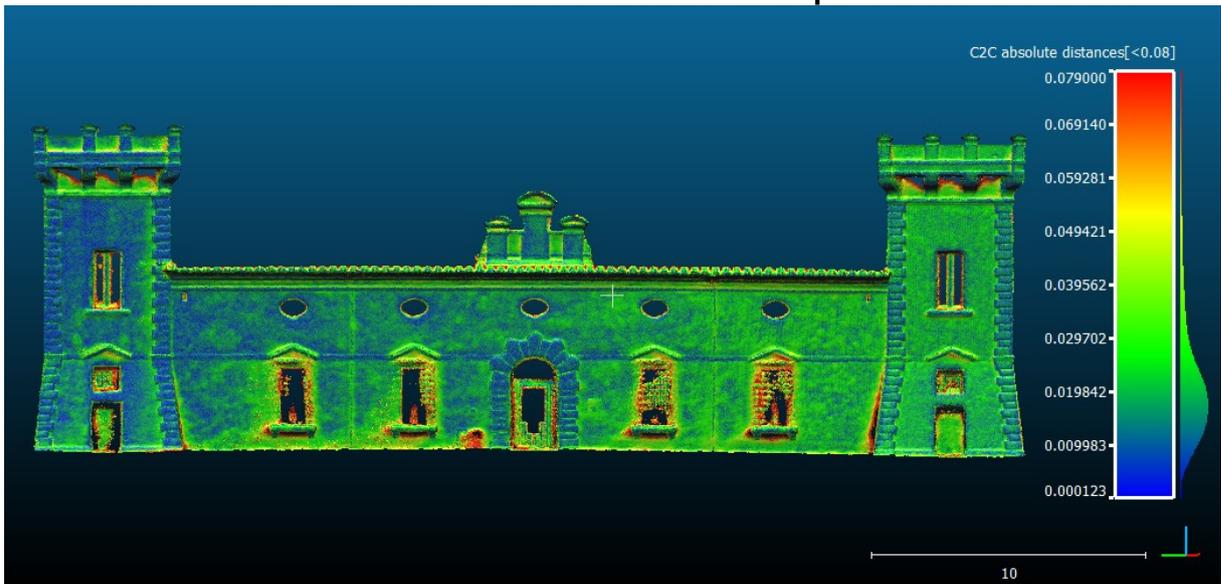


Figura 9.13. Differenze misurate tra nubi dense di punti e scansione laser raffigurate mediante mappa di colore. In rosso le zone di maggior scostamento. Sotto si riportano gli istogrammi di frequenza in distanze assolute (a sinistra) e non (a destra). Considerando i segni reali degli scostamenti è possibile costruirne la distribuzione normale di probabilità.

PROSPETTO EST: Confronto tra nube densa di punti e scansione laser

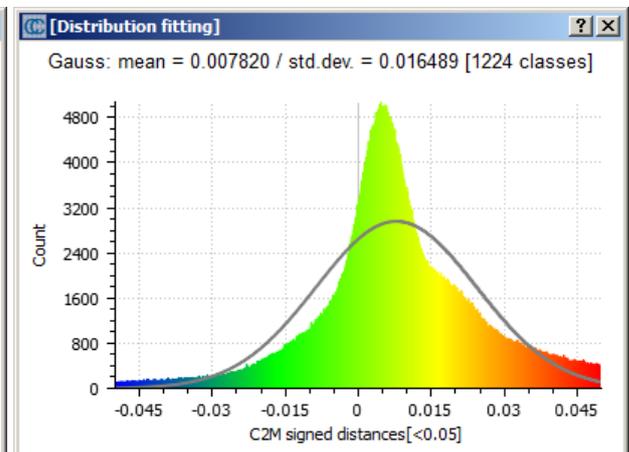
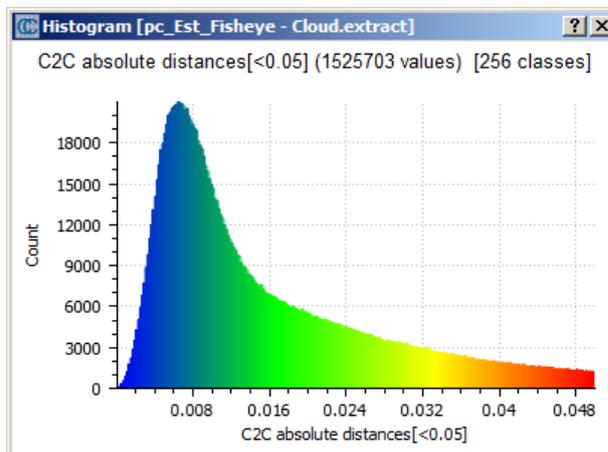
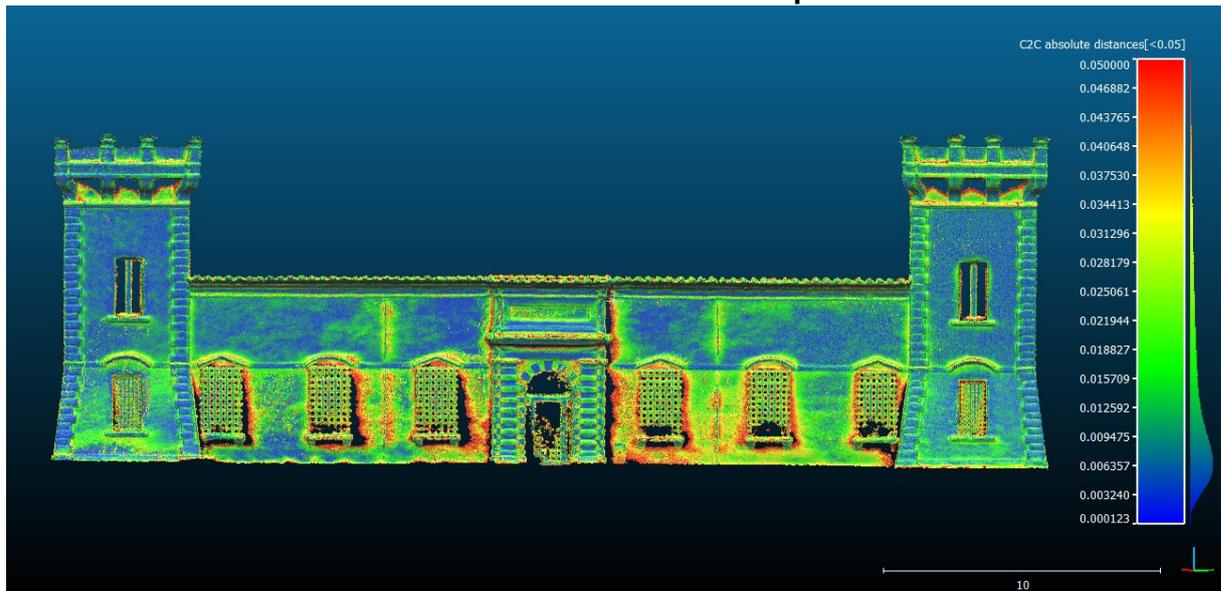


Figura 9.14. Differenze misurate tra nubi dense di punti e scansione laser raffigurate mediante mappa di colore. In rosso le zone di maggior scostamento. Sotto si riportano gli istogrammi di frequenza in distanze assolute (a sinistra) e non (a destra). Considerando i segni reali degli scostamenti è possibile costruirne la distribuzione normale di probabilità.

PROSPETTO NORD: Confronto tra nube densa di punti e scansione laser

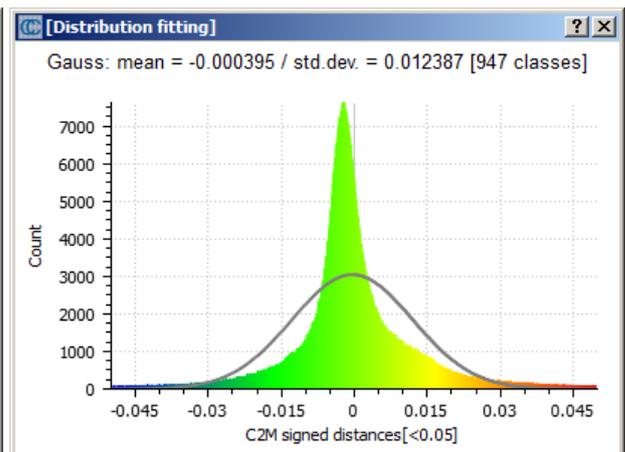
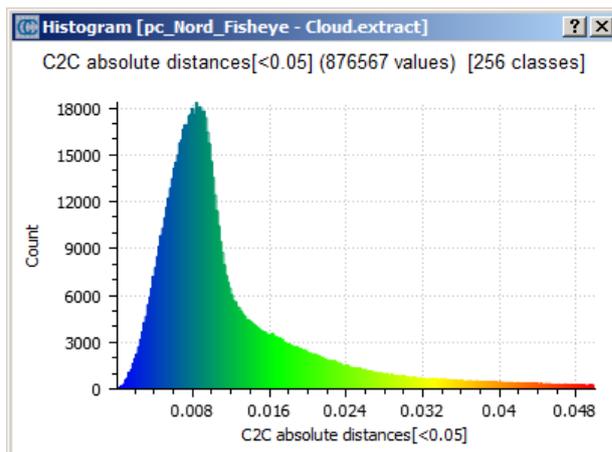
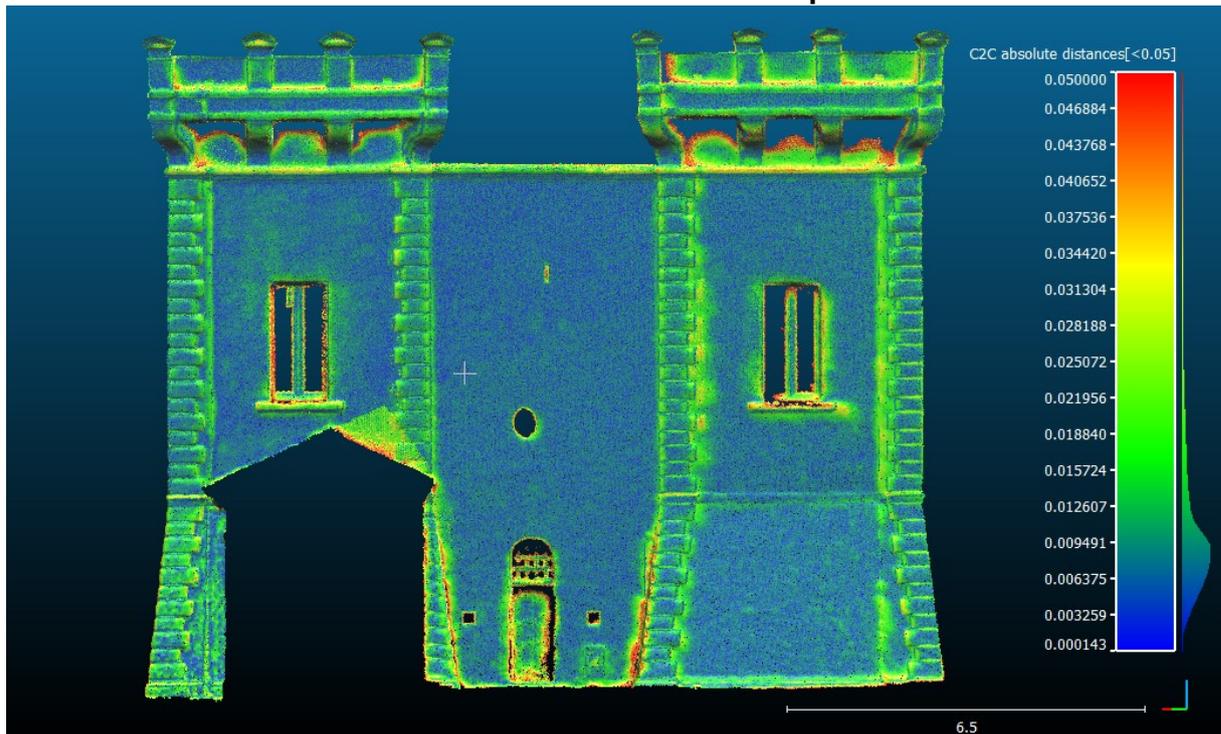


Figura 9.15. Differenze misurate tra nubi dense di punti e scansione laser raffigurate mediante mappa di colore. In rosso le zone di maggior scostamento. Sotto si riportano gli istogrammi di frequenza in distanze assolute (a sinistra) e non (a destra). Considerando i segni reali degli scostamenti è possibile costruirne la distribuzione normale di probabilità.

PROSPETTO SUD: Confronto tra nube densa di punti e scansione laser

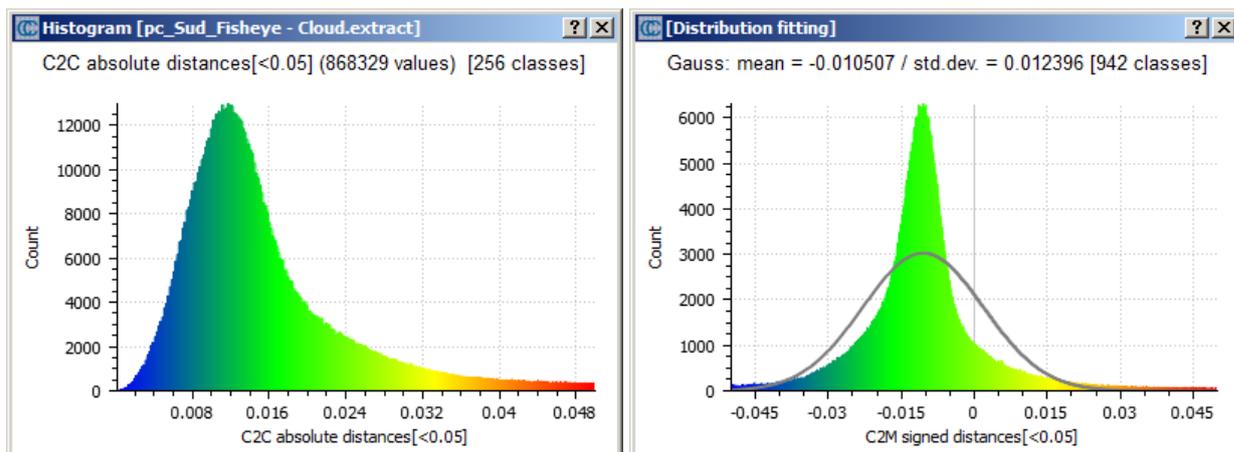
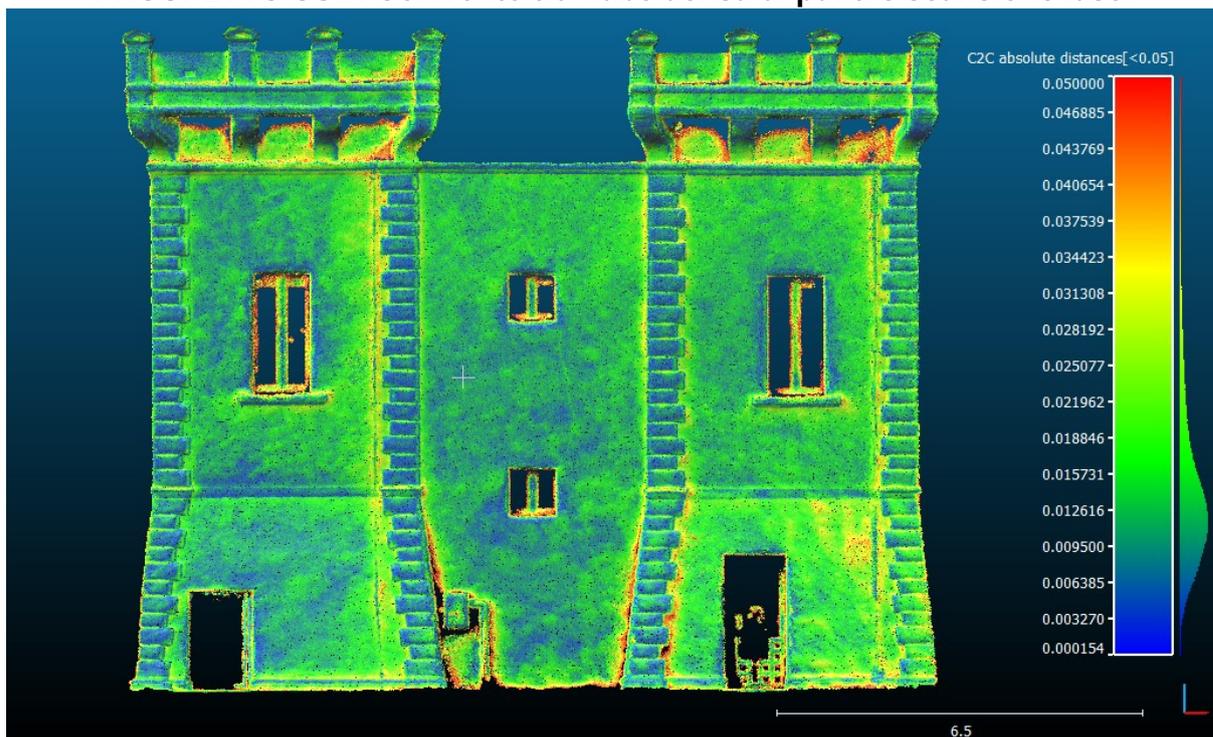


Figura 9.16. Differenze misurate tra nubi dense di punti e scansione laser raffigurate mediante mappa di colore. In rosso le zone di maggior scostamento. Sotto si riportano gli istogrammi di frequenza in distanze assolute (a sinistra) e non (a destra). Considerando i segni reali degli scostamenti è possibile costruirne la distribuzione normale di probabilità.

MODELLO INTEGRATO: Confronto tra nube densa di punti e scansione laser

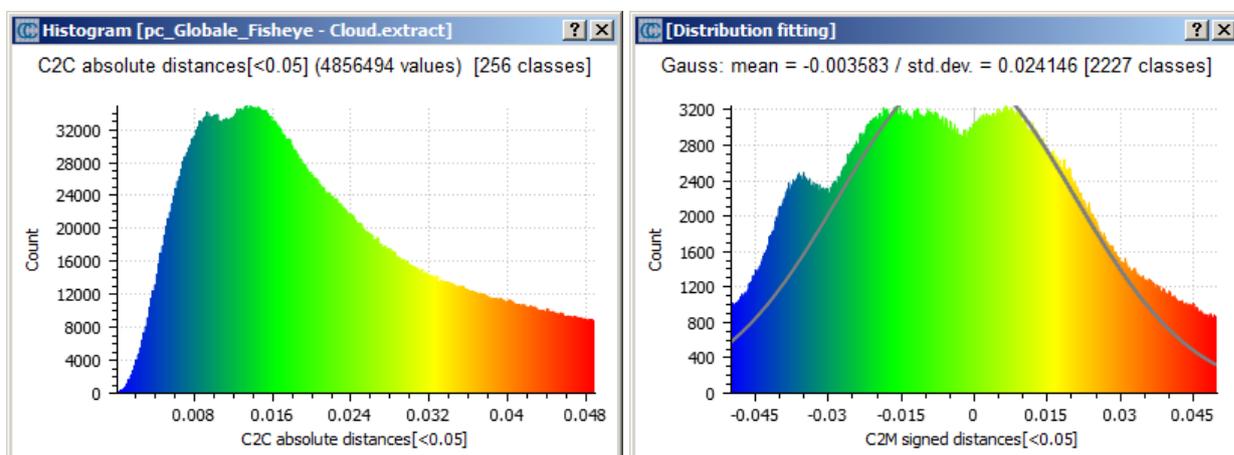
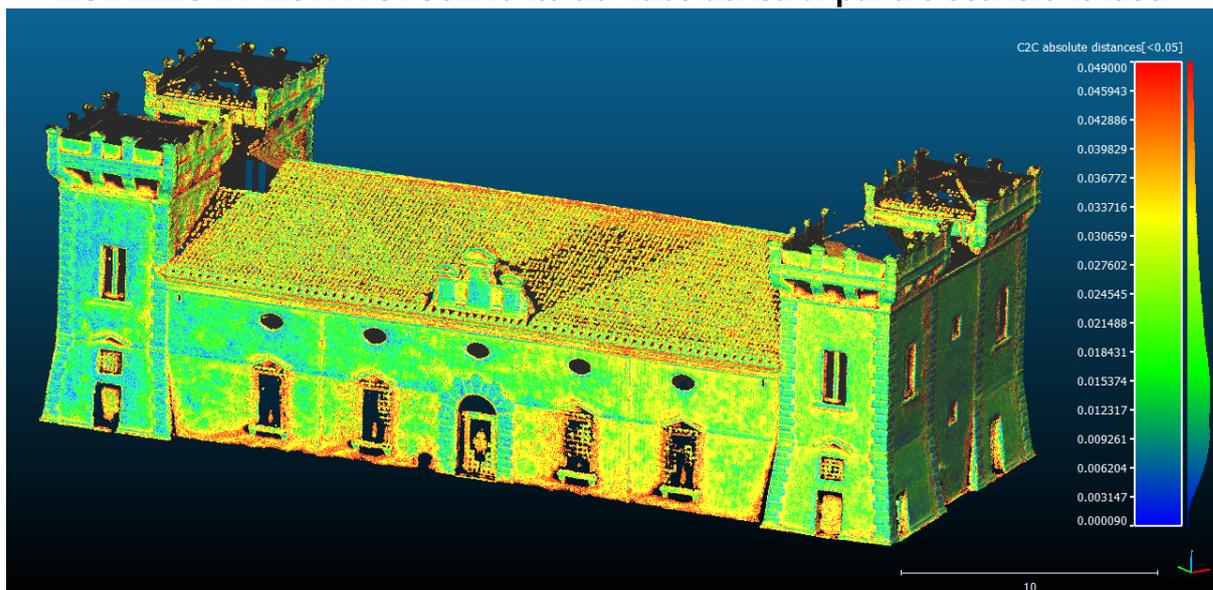


Figura 9.17. Differenze misurate tra nubi dense di punti e scansione laser raffigurate mediante mappa di colore. In rosso le zone di maggior scostamento. Sotto si riportano gli istogrammi di frequenza in distanze assolute (a sinistra) e non (a destra). Considerando i segni reali degli scostamenti è possibile costruirne la distribuzione normale di probabilità.

In tabella i principali risultati di tutte le elaborazioni:

CONFRONTO: NUBI DENSE DI PUNTI E SCANSIONE LASER		
MODELLI REALIZZATI CON AUTOCALIBRAZIONE (fisheye) DI PHOTOSCAN	DIFFERENZE	
	Media (m)	Deviazione standard (m)
Prospetto OVEST	-0.015	0.014
Prospetto EST	0.008	0.016
Prospetto NORD	0.001	0.012
Prospetto SUD	-0.010	0.012
MODELLO INTEGRATO	-0.003	0.024

Tabella 9.9. Report del confronto tra nube densa di punti ottenuta con autocalibrazione di PhotoScan e scansione laser

CONFRONTO: NUBI DENSE DI PUNTI E SCANSIONE LASER
--

MODELLI REALIZZATI CON CALIBRAZIONE DI PHOTODELER	DIFFERENZE	
	Media	Deviazione standard
	(m)	(m)
Prospetto OVEST	-0.016	0.029
Prospetto EST	-0.001	0.020
Prospetto NORD	-0.001	0.016
Prospetto SUD	-0.004	0.017
MODELLO INTEGRATO	-0.001	0.021

Tabella 9.10. Report del confronto tra nube densa di punti ottenuta con self-calibration di Photomodeler e scansione laser

CONFRONTO: NUBI DENSE DI PUNTI E SCANSIONE LASER		
MODELLI REALIZZATI CON IMMAGINI CORRETTE DALLA DISTORSIONE	DIFFERENZE	
	Media	Deviazione standard
	(m)	(m)
Prospetto OVEST	-0.014	0.019
Prospetto EST	0.008	0.017
Prospetto NORD	0.001	0.014
Prospetto SUD	-0.010	0.013
MODELLO INTEGRATO	0.003	0.023

Tabella 9.11. Report del confronto tra nube densa di punti ottenuta da immagini ricampionate e scansione laser

Complessivamente si può affermare che la metodologia che ha riscontrato la precisione migliore è stata quella implementata nel software di restituzione per il trattamento delle immagini tipo “fisheye”. Unica eccezione è quella del modello realizzato utilizzando la calibrazione esterna di Photomodeler che ha mostrato una deviazione standard leggermente più piccola rispetto agli altri. Nonostante ciò visto la piccola differenza tra i casi osservati si preferisce comunque il metodo fisheye in quanto anche da un punto di vista qualitativo è stato quello che ha portato alla realizzazione delle nubi più complete e pulite.

9.6 ELABORAZIONE DEL DATO (VIDEO)

La seconda parte della sperimentazione si è incentrata sullo studio dei risultati ottenibili da riprese video effettuate tramite la medesima camera utilizzata in precedenza, ovvero la GoPro Hero 3+ Black Edition, e lo stesso APR, il Phantom 2. Grazie all’utilizzo dei video non è stato necessario porsi il problema relativo al ricoprimento in quanto estraendo i fotogrammi ad intervalli di tempo molto ravvicinati è possibile ottenere ricoprimenti ben oltre l’80%.

La ripresa video, a differenza delle prese fotografiche, presenta tuttavia una risoluzione decisamente inferiore. Infatti, nonostante la ripresa ad alta definizione utilizzata sia pari a 1920×1080 pixel, la

risoluzione dei fotogrammi è di 4000×3000 pixel. Questa differenza di risoluzione porta ad avere immagini meno nitide con una più difficile identificazione dei particolari da parte dell'operatore in fase di introduzione dei punti di controllo.

I fotogrammi estratti dal video presentano un ulteriore problema rispetto alle foto, ovvero non portano con sé le informazioni riguardanti le caratteristiche di presa, come pixel-size e focale, necessarie a PhotoScan per eseguire la calibrazione. A questo proposito sono state eseguite due diverse prove: la prima utilizzando le foto affette da distorsione e fatte elaborare al programma in modalità fisheye, nonostante non sia in grado di garantire un'ottimale correzione; nella seconda prova il video è stato invece preventivamente corretto all'interno del programma *GoPro Studio*, attraverso il quale è stata eliminata la distorsione, e sono stati utilizzati frame non distorti.

I casi di studio per l'elaborazione video sono in numero minore a quanto visto in precedenza per le foto. Nello specifico sono state prese in esame solamente quelle tipologie di elaborazione che hanno dato i risultati migliori nella modellazione precedente.

Pertanto gli approcci sviluppati sono stati solamente due, ovvero:

- L'utilizzo di frame affetti da distorsione Fisheye e la successiva elaborazione tramite la relativa impostazione di gestione.
- L'utilizzo di frame precedentemente corretti e gestiti dal programma come immagini normali.

Questi due approcci sono stati sviluppati entrambi fino alla realizzazione della nube densa di punti e confrontati tramite la comparazione tra nubi utilizzata precedentemente per le fotografie.



Figura 9.18. Vista della nube densa di punti ottenuta con frames estratti da video.



Figura 9.19. Nube densa di punti ripulita per il confronto con la nube di riferimento

9.7 VALUTAZIONE DELLA PRECISIONE (VIDEO)

9.7.1 Modelli realizzati con autocalibrazione "fisheye" di Photoscan

PROSPETTO OVEST: Confronto tra nube densa di punti e scansione laser

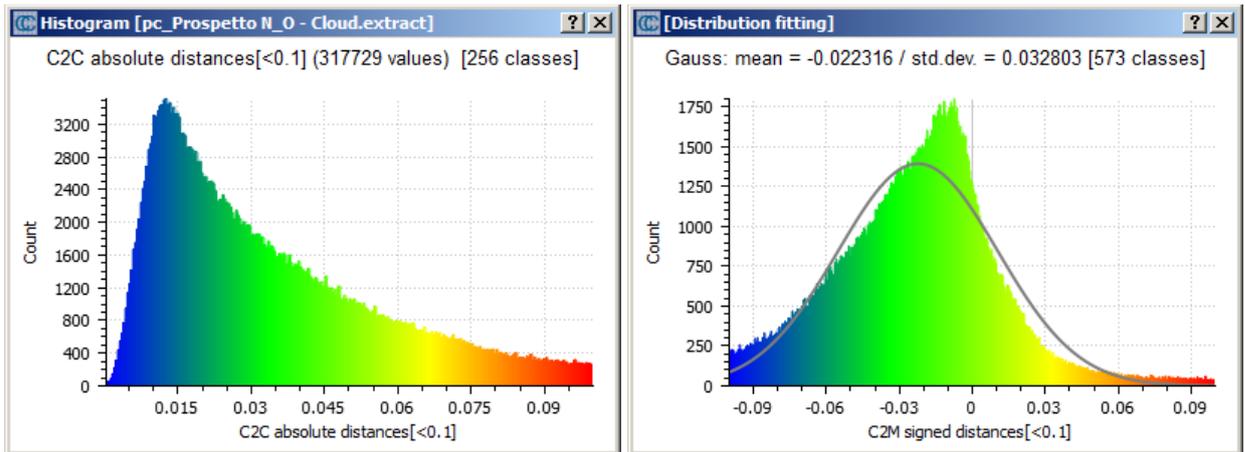
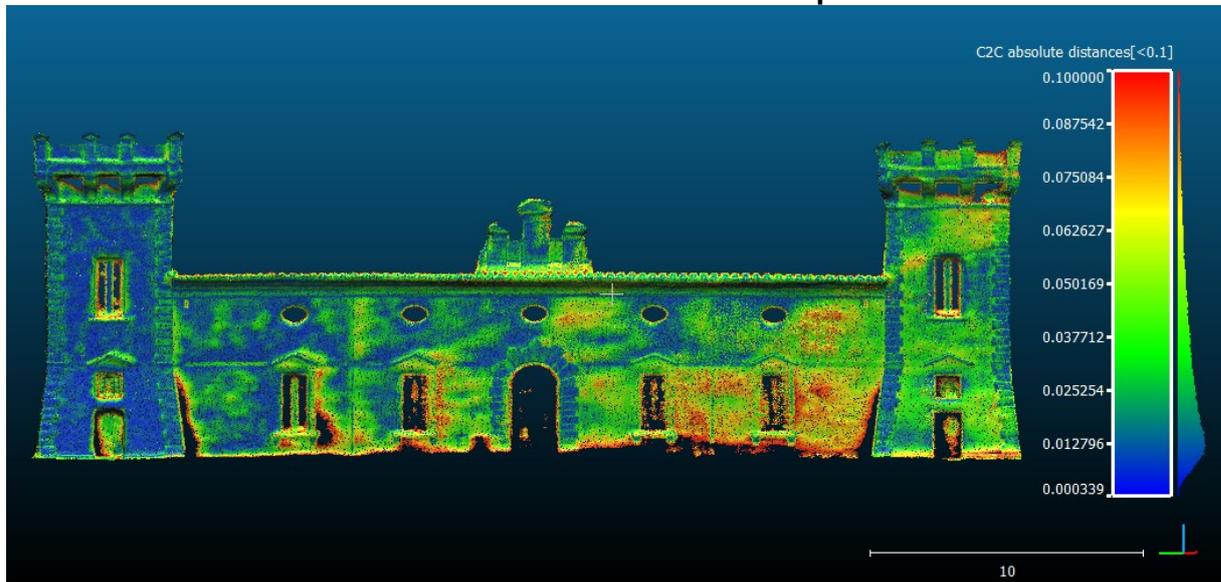


Figura 9.20. Differenze misurate tra nubi dense di punti e scansione laser raffigurate mediante mappa di colore. In rosso le zone di maggior scostamento. Sotto si riportano gli istogrammi di frequenza in distanze assolute (a sinistra) e non (a destra). Considerando i segni reali degli scostamenti è possibile costruirne la distribuzione normale di probabilità.

PROSPETTO EST: Confronto tra nube densa di punti e scansione laser

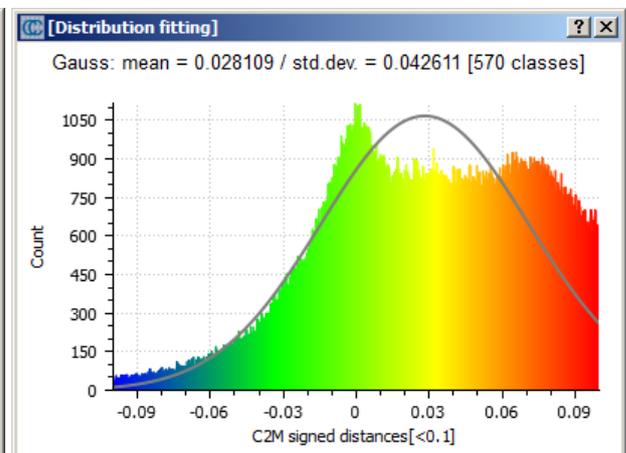
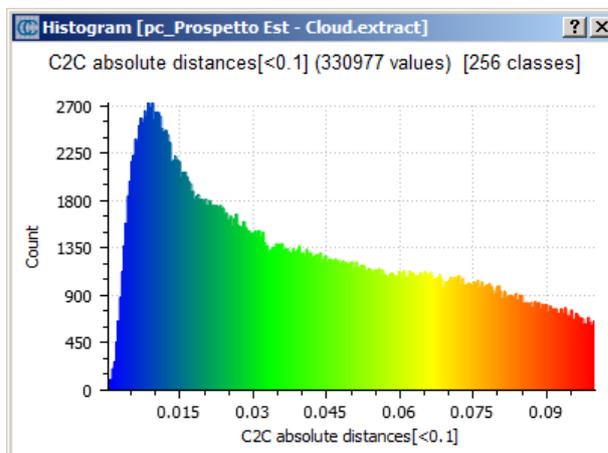
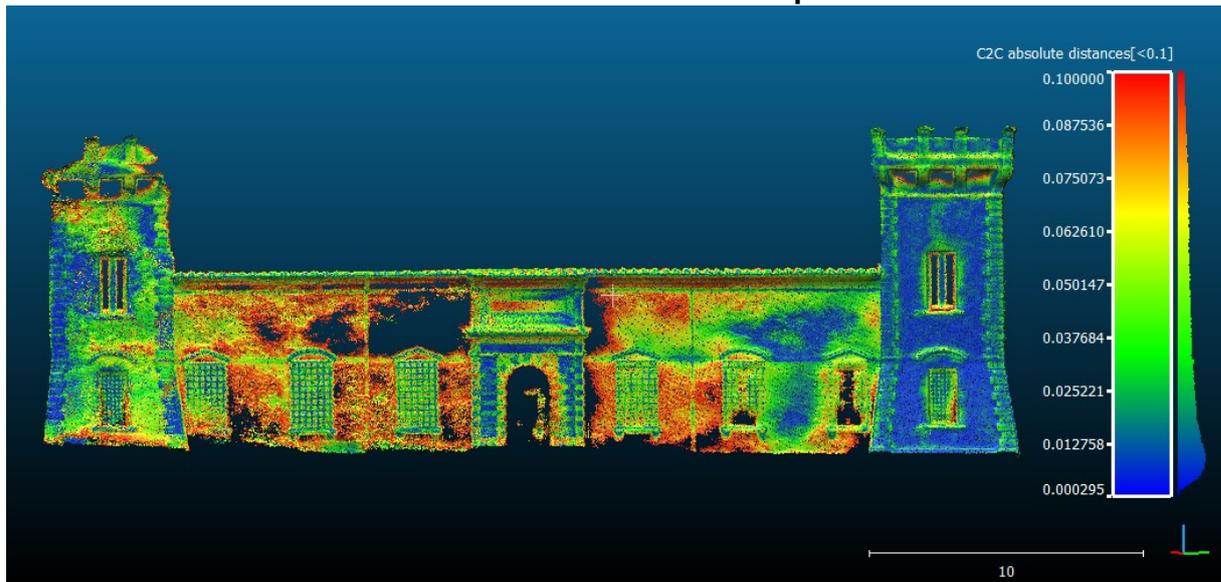


Figura 9.21. Differenze misurate tra nubi dense di punti e scansione laser raffigurate mediante mappa di colore. In rosso le zone di maggior scostamento. Sotto si riportano gli istogrammi di frequenza in distanze assolute (a sinistra) e non (a destra). Considerando i segni reali degli scostamenti è possibile costruirne la distribuzione normale di probabilità.

PROSPETTO NORD: Confronto tra nube densa di punti e scansione laser

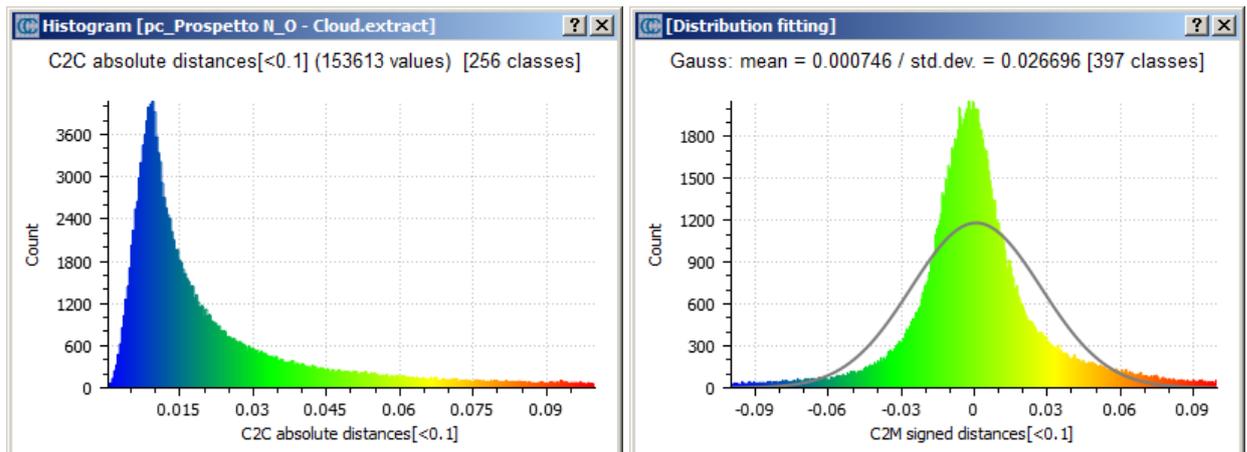
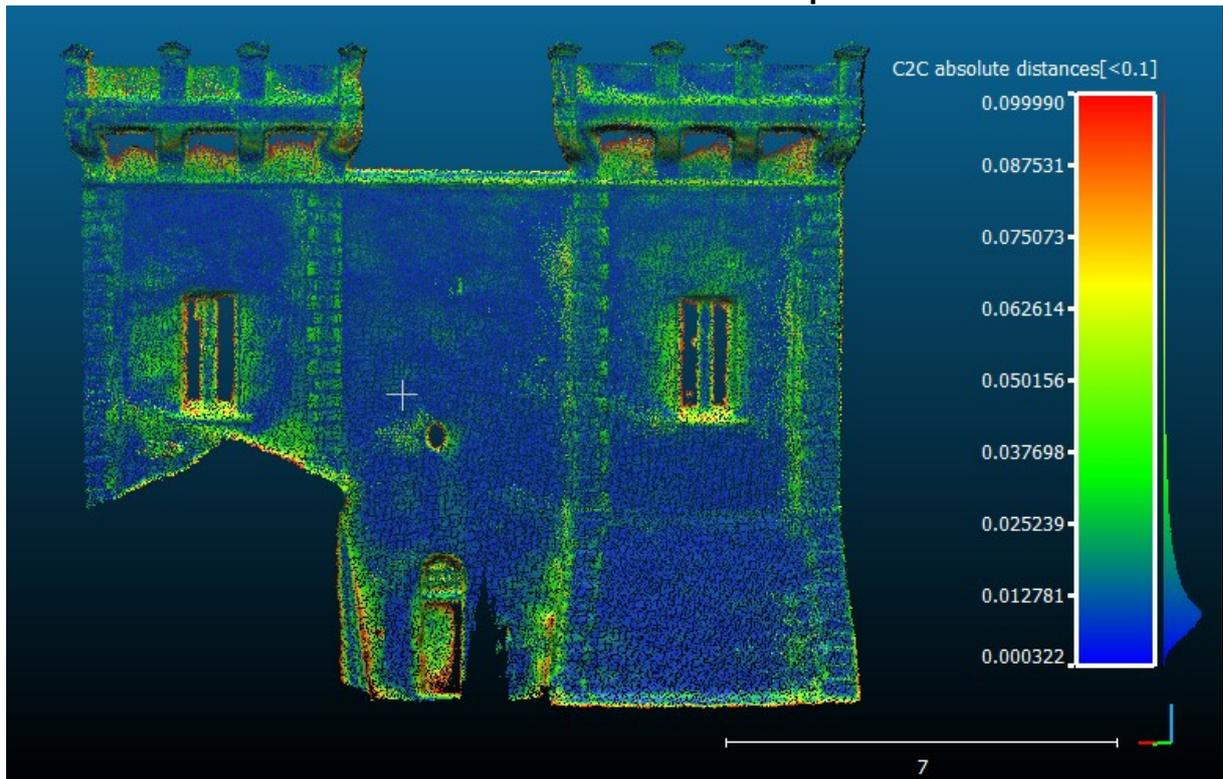


Figura 9.22. Differenze misurate tra nubi dense di punti e scansione laser raffigurate mediante mappa di colore. In rosso le zone di maggior scostamento. Sotto si riportano gli istogrammi di frequenza in distanze assolute (a sinistra) e non (a destra). Considerando i segni reali degli scostamenti è possibile costruirne la distribuzione normale di probabilità.

PROSPETTO SUD: Confronto tra nube densa di punti e scansione laser

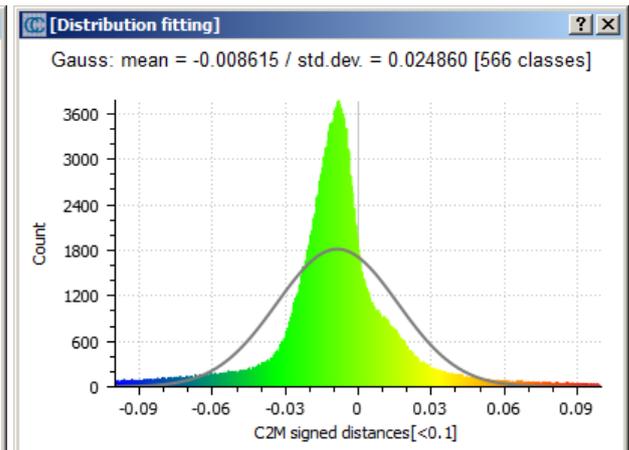
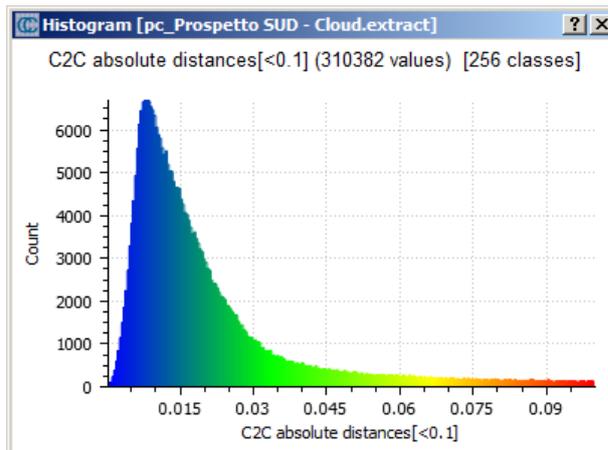
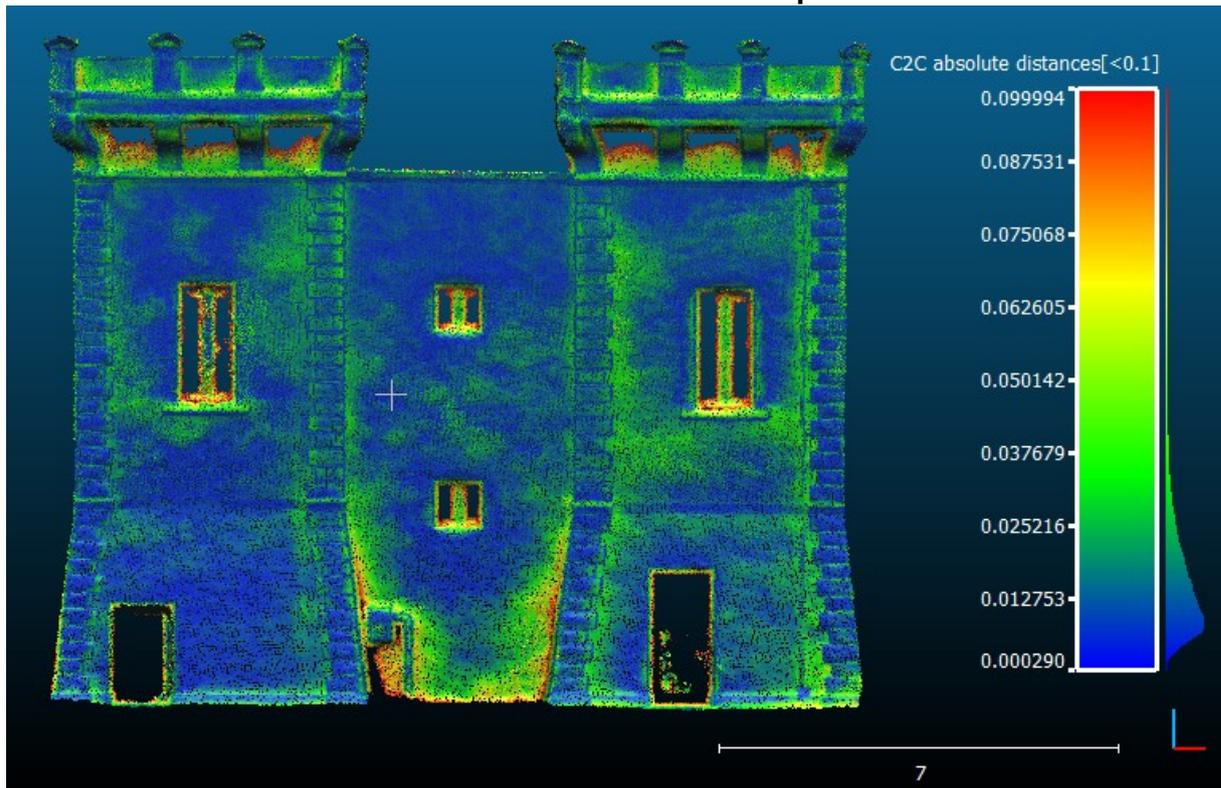


Figura 9.23. Differenze misurate tra nubi dense di punti e scansione laser raffigurate mediante mappa di colore. In rosso le zone di maggior scostamento. Sotto si riportano gli istogrammi di frequenza in distanze assolute (a sinistra) e non (a destra). Considerando i segni reali degli scostamenti è possibile costruirne la distribuzione normale di probabilità.

MODELLO INTEGRATO: Confronto tra nube densa di punti e scansione laser

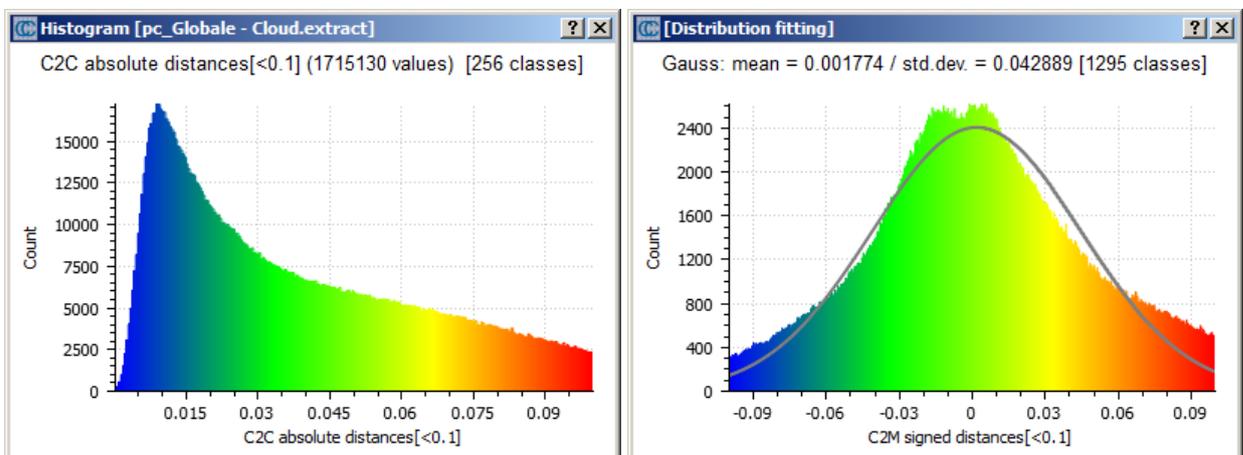
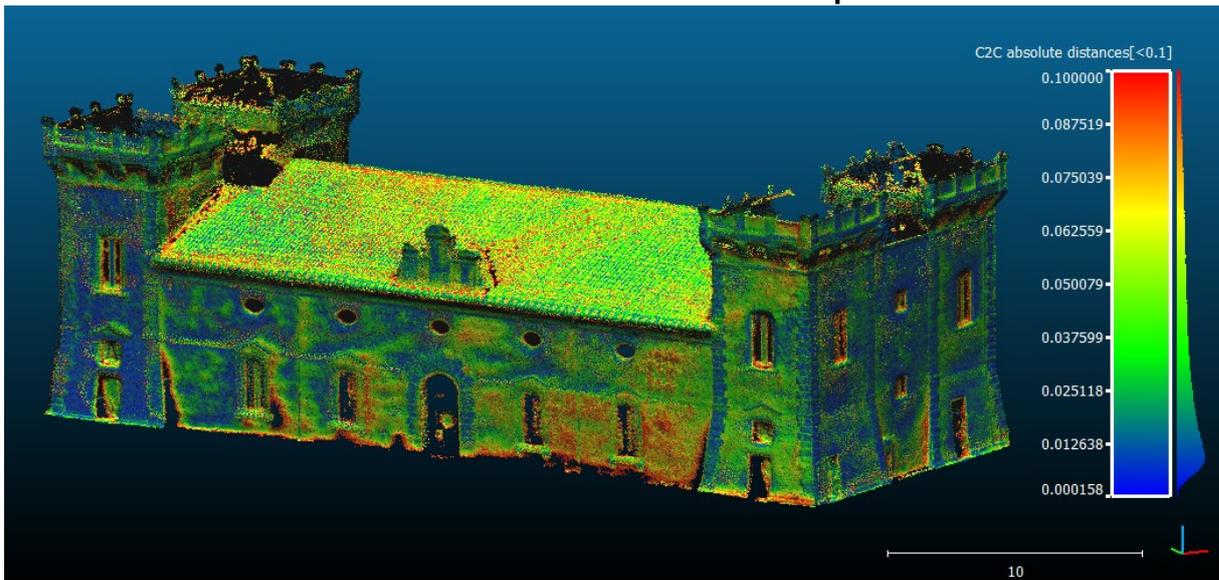


Figura 9.24. Differenze misurate tra nubi dense di punti e scansione laser raffigurate mediante mappa di colore. In rosso le zone di maggior scostamento. Sotto si riportano gli istogrammi di frequenza in distanze assolute (a sinistra) e non (a destra). Considerando i segni reali degli scostamenti è possibile costruirne la distribuzione normale di probabilità.

9.7.2 Modelli realizzati a partire da frames corretti con GoPro Studio

PROSPETTO OVEST: Confronto tra nube densa di punti e scansione laser

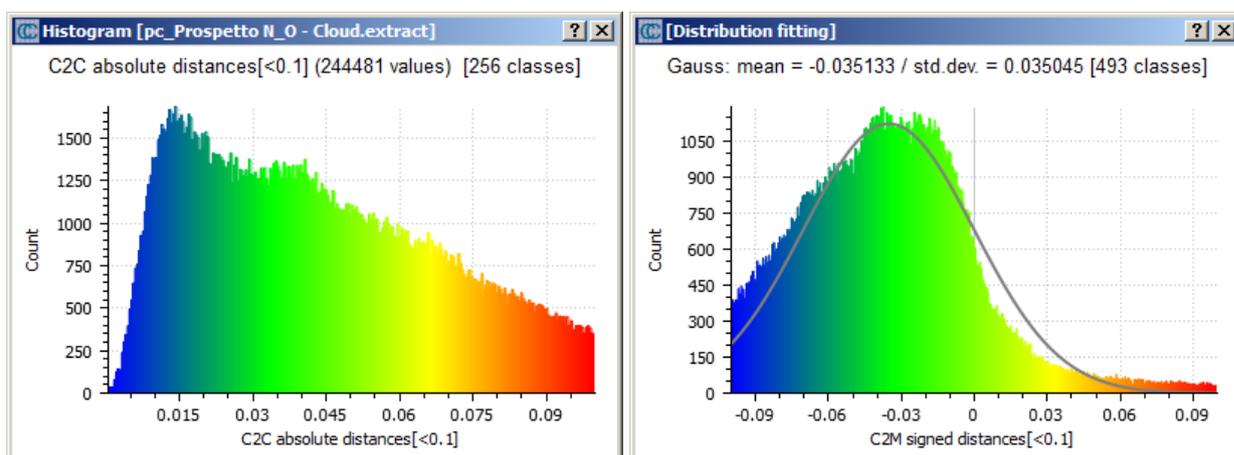
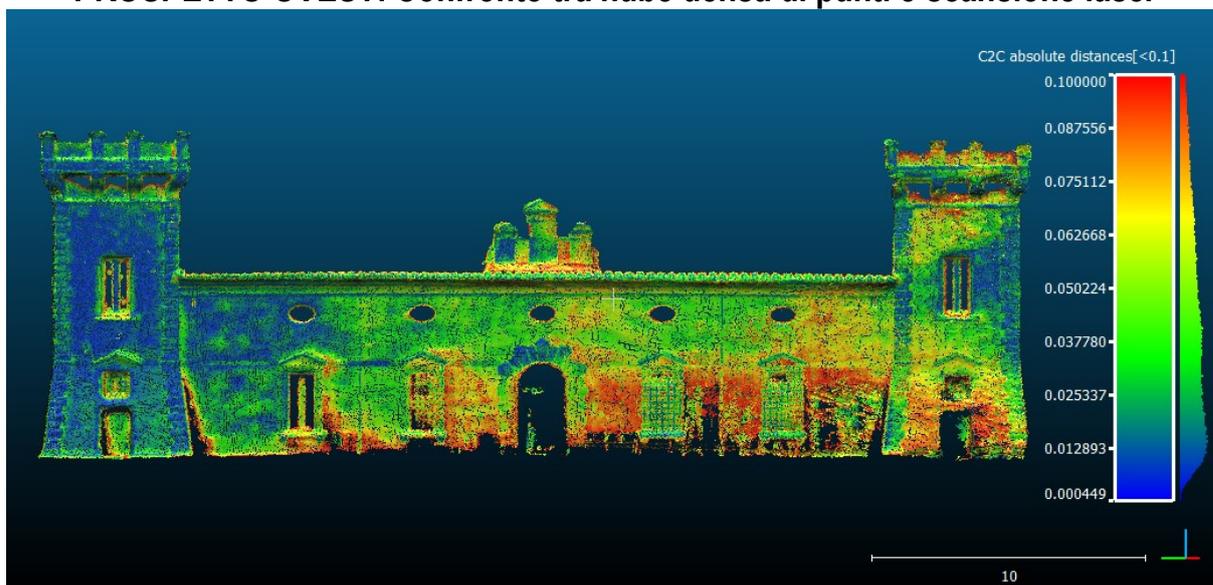


Figura 9.25. Differenze misurate tra nubi dense di punti e scansione laser raffigurate mediante mappa di colore. In rosso le zone di maggior scostamento. Sotto si riportano gli istogrammi di frequenza in distanze assolute (a sinistra) e non (a destra). Considerando i segni reali degli scostamenti è possibile costruirne la distribuzione normale di probabilità.

PROSPETTO EST: Confronto tra nube densa di punti e scansione laser

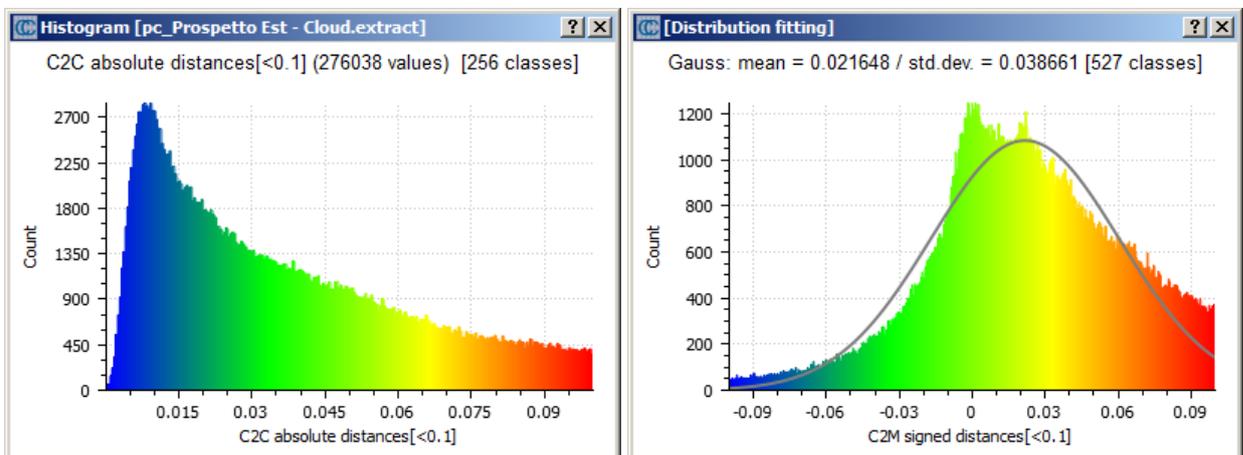
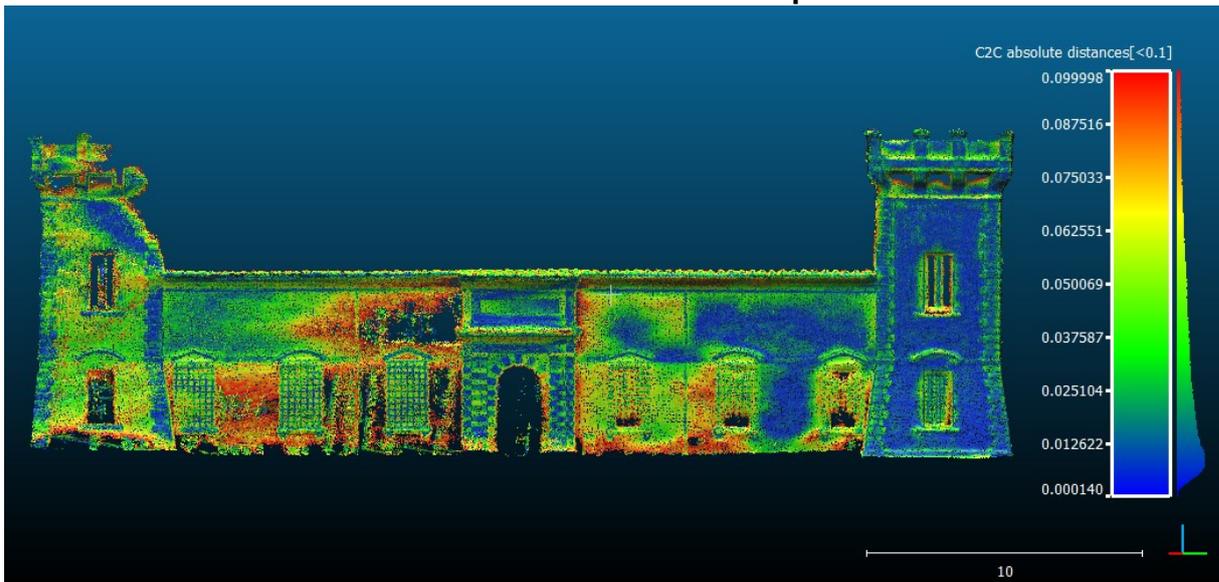


Figura 9.26. Differenze misurate tra nubi dense di punti e scansione laser raffigurate mediante mappa di colore. In rosso le zone di maggior scostamento. Sotto si riportano gli istogrammi di frequenza in distanze assolute (a sinistra) e non (a destra). Considerando i segni reali degli scostamenti è possibile costruirne la distribuzione normale di probabilità.

PROSPETTO NORD: Confronto tra nube densa di punti e scansione laser

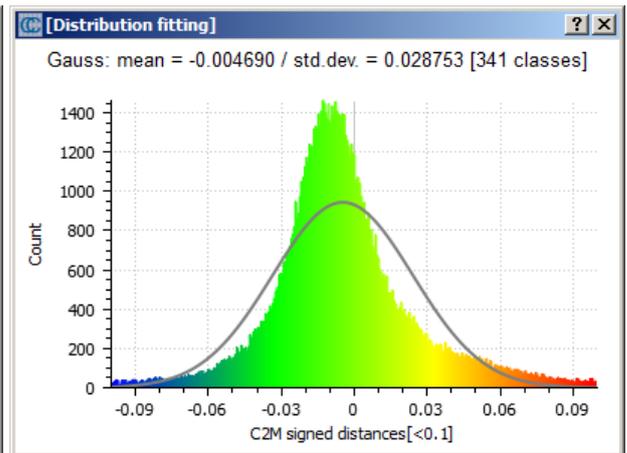
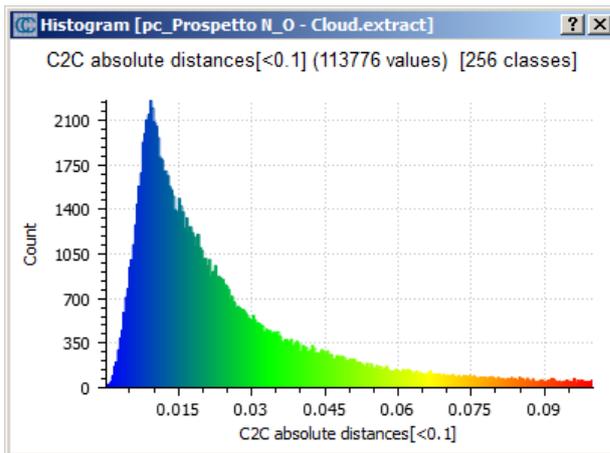
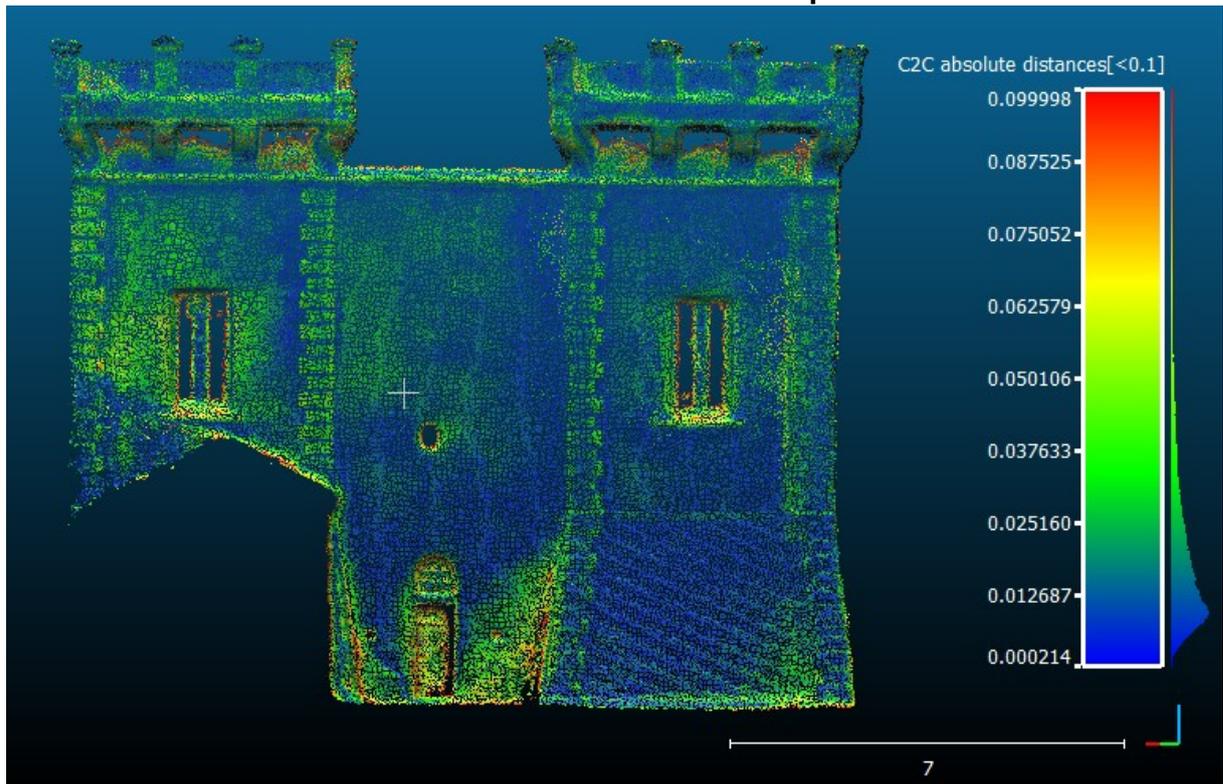


Figura 9.27. Differenze misurate tra nubi dense di punti e scansione laser raffigurate mediante mappa di colore. In rosso le zone di maggior scostamento. Sotto si riportano gli istogrammi di frequenza in distanze assolute (a sinistra) e non (a destra). Considerando i segni reali degli scostamenti è possibile costruirne la distribuzione normale di probabilità.

PROSPETTO SUD: Confronto tra nube densa di punti e scansione laser

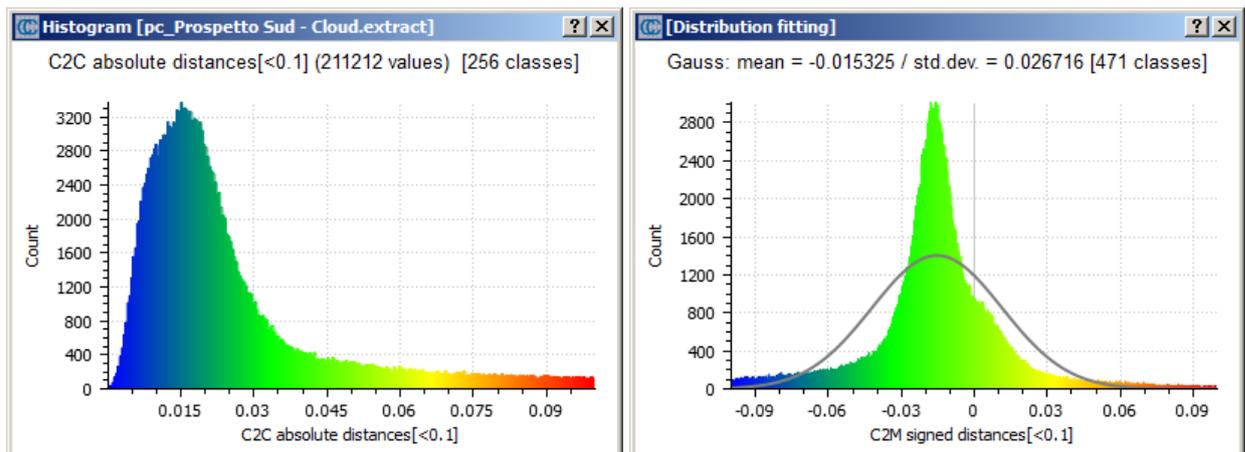
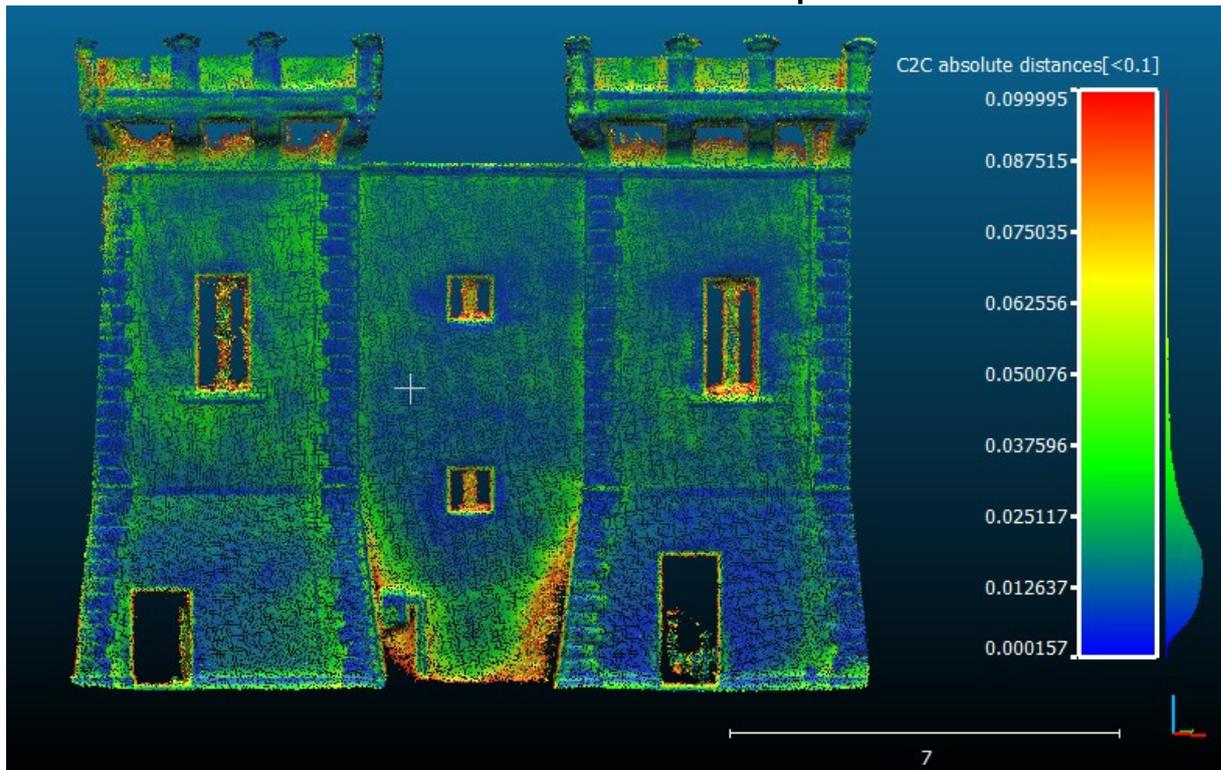


Figura 9.28. Differenze misurate tra nubi dense di punti e scansione laser raffigurate mediante mappa di colore. In rosso le zone di maggior scostamento. Sotto si riportano gli istogrammi di frequenza in distanze assolute (a sinistra) e non (a destra). Considerando i segni reali degli scostamenti è possibile costruirne la distribuzione normale di probabilità.

PROSPETTO SUD: Confronto tra nube densa di punti e scansione laser

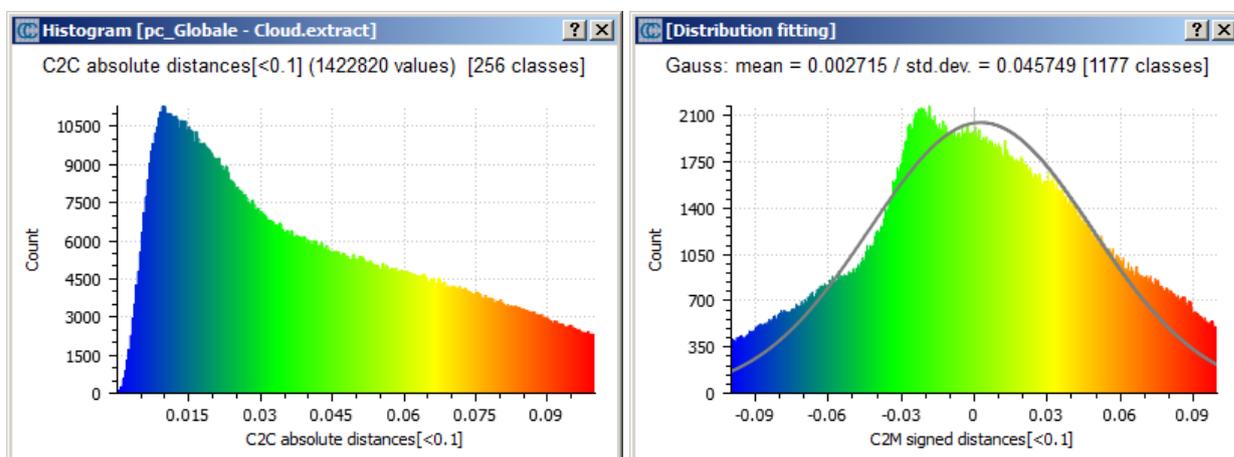
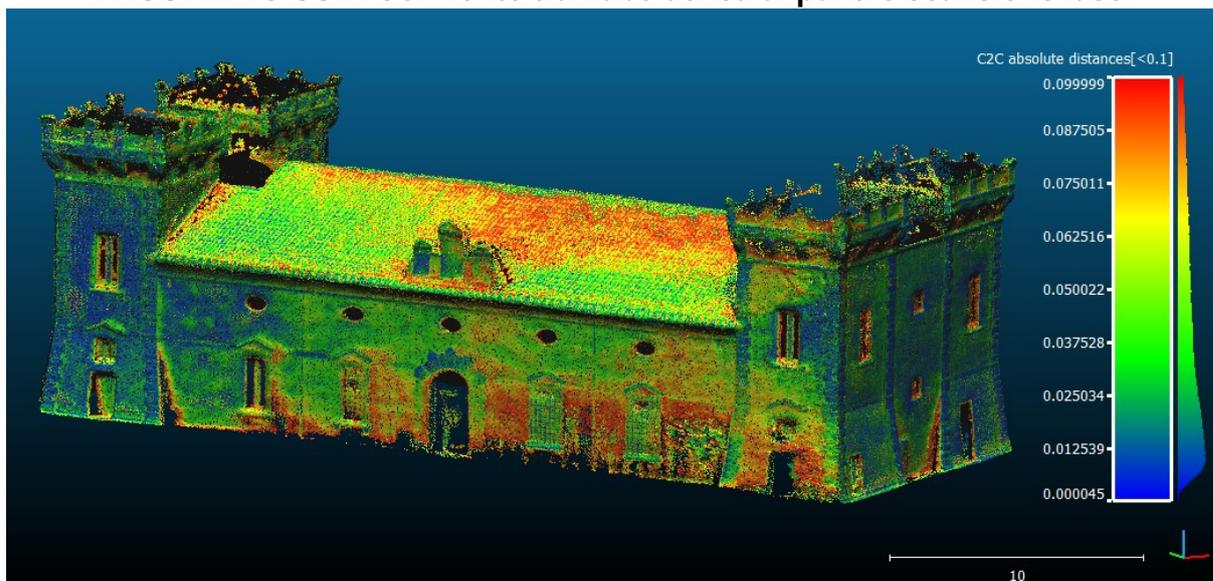


Figura 9.29. Differenze misurate tra nubi dense di punti e scansione laser raffigurate mediante mappa di colore. In rosso le zone di maggior scostamento. Sotto si riportano gli istogrammi di frequenza in distanze assolute (a sinistra) e non (a destra). Considerando i segni reali degli scostamenti è possibile costruirne la distribuzione normale di probabilità.

Nella tabella successiva sono stati riportati per semplicità tutti i risultati ottenuti:

CONFRONTO: MODELLI FOTOGRAMMETRICI DA VIDEO – MODELLO DI RIFERIMENTO		
MODELLI REALIZZATI CON AUTOCALIBRAZIONE DI PHOTOSCAN	DIFFERENZE	
	Media (m)	Deviazione standard (m)
Prospetto OVEST	-0.022	0.033
Prospetto EST	0.028	0.043
Prospetto NORD	0.001	0.027
Prospetto SUD	-0.009	0.025
MODELLO INTEGRATO	0.002	0.043

Tabella 9.12. Report del confronto tra nube densa di punti ottenuta con autocalibrazione di PhotoScan e scansione laser

CONFRONTO: MODELLI FOTOGRAMMETRICI DA VIDEO – MODELLO DI RIFERIMENTO
--

MODELLI REALIZZATI CON FRAMES SENZA DISTORSIONE (GoPro Studio)	DIFFERENZE	
	Media	Deviazione standard
	(m)	(m)
Prospetto OVEST	-0.035	0.035
Prospetto EST	0.021	0.039
Prospetto NORD	-0.005	0.029
Prospetto SUD	-0.015	0.027
MODELLO INTEGRATO	0.003	0.046

Tabella 9.13. Report del confronto tra nube densa di punti ottenuta con correzione della distorsione dei video con Go Pro Studio e scansione laser

Le differenze riscontrate tra le due metodologie sono molto piccole, potrebbero essere utilizzate indifferentemente una o l'altra. Rispetto all'utilizzo delle immagini fotografiche si osserva invece un forte incremento della deviazione standard, sintomatico della scarsa precisione del rilievo. Il comportamento è da attribuire alla minor risoluzione dei filmati che inevitabilmente lavorano con GSD maggiori. Nel passaggio tra le immagini e i video si riscontra dunque un calo di risoluzione che si traduce in un calo di accuratezza.

9.8 CONSIDERAZIONI FINALI

Sono state sviluppate, quali metodologie di rilievo, due diverse soluzioni, una mediante fotogrammi ed un'altra attraverso l'impiego di frame estratti dai video. Entrambi gli approcci hanno richiesto la medesima quantità di tempo per l'acquisizione del dato e si sono basati sugli stessi percorsi di volo; da questo punto di vista essi sono quindi del tutto equivalenti.

Tra i vari approcci utilizzati durante l'elaborazione, sia delle foto che dei video, la metodologia di intervento migliore è risultata essere, in entrambi i casi, quella che prevede l'utilizzo delle impostazioni per la correzione della distorsione *fisheye* già implementate all'interno del programma *PhotoScan*. Tramite l'utilizzo di questa procedura, le fasi di elaborazione risultano essere molto più rapide rispetto al procedimento che prevede l'idealizzazione preliminare dei fotogrammi (approccio che restituisce comunque risultati molto soddisfacenti), inoltre il modello di correzione per obiettivi *fisheye* già disponibile all'interno del software permette una migliore gestione delle distorsioni.

Nell'analisi della precisione del rilievo stesso, si hanno tuttavia risultati assai diversi (Tabelle 9.12. e 9.13.). Dalla comparazione con la nube di riferimento, si evidenziano errori di gran lunga maggiori nell'elaborazione dei frame estratti dai video, che risultano essere circa doppi rispetto all'approccio fotografico. La sostanziale differenza tra i risultati ottenuti è chiaramente visibile in Figura 9.30.

Confronto tra le gaussiane del modello Globale derivante da Foto e Video

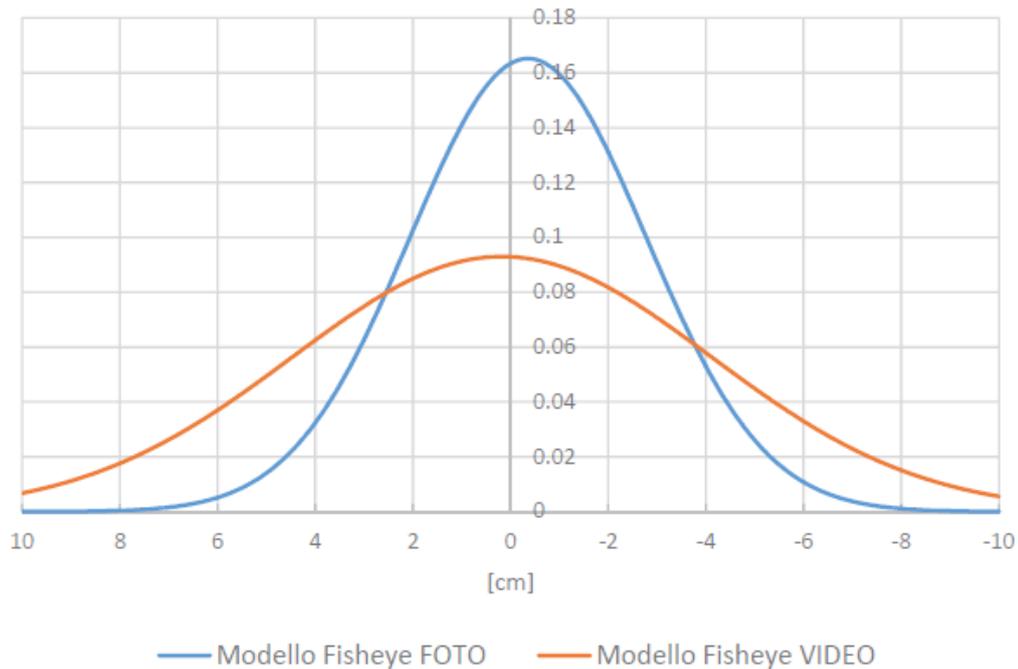


Figura 9.30. Distribuzioni normali ottenute dai confronti metrici tra i modelli realizzati da immagini fotografiche e modelli realizzati da video. Il confronto mostra la maggior precisione riscontrata nell'utilizzo delle immagini piuttosto che i frames estratti dai video.

Questo risultato è dovuto ad una qualità inferiore del dato video (1920×1080) rispetto a quello fotografico (4000×3000), nonché alla mancanza, nel dato video, delle informazioni base (focale e pixel size) necessarie a *PhotoScan* per eseguire una calibrazione attendibile.

Anche per questo caso di studio **è lecito affermare che, mediante l'impiego della tecnica fotogrammetrica da aeromobile a pilotaggio remoto ed attraverso l'utilizzo di camere fisheye come la GoPro, sia possibile ottenere modelli con un grado di precisione stimabile in 2÷3 cm**, come è possibile osservare analizzando i risultati ottenuti in questa tesi. Tali valori risultano essere ampiamente in linea con la precisione richiesta per il rilievo architettonico dei beni culturali.

10 CONCLUSIONI FINALI

In questo capitolo si vogliono mettere in evidenza alcune semplici indicazioni progettuali per l'esecuzione e la restituzione di rilievi fotogrammetrici applicati ai Beni Culturali. Si potrebbero definire come una sorta di *"Linee Guida"* anche se limitate alle modalità esecutive presentate in questa tesi. Infatti le indicazioni che verranno fornite avranno comunque una valenza di carattere generale, ma i valori numerici di precisione e di accuratezza attesi dal rispetto di tali metodologie saranno da associare esclusivamente alle modalità di acquisizione ed elaborazione del dato presentate. Analizzando fase dopo fase il processo di rilievo verranno dunque messi in luce gli aspetti ritenuti più rilevanti e significativi, da tener sempre in considerazione se si vuole eseguire un rilievo di questo tipo. Queste indicazioni sono rivolte a tutti coloro che si avvicinano ad un rilievo architettonico non conoscendo a fondo i processi che stanno dietro alla fase di restituzione. Sono indicazioni di carattere pratico che possono essere utilizzate come semplice guida in fase di rilievo fotogrammetrico.

10.1 PIANIFICAZIONE DEL RILIEVO

La fase di pianificazione, a volte sottovalutata o eseguita in maniera frettolosa è spesso una tra le cause principali di errore. Una corretta valutazione della strumentazione da utilizzare e della strategia di presa da adottare sono sempre fondamentali in quanto direttamente responsabili del buon esito della restituzione. Tali caratteristiche sono infatti strettamente collegate allo scopo finale del rilievo.

Capire in fase di pianificazione qual è il vero obiettivo del rilievo (ad esempio la realizzazione di una planimetria, di prospetti, di sezioni o un semplice inquadramento, può far risparmiare tempo e denaro oltre che portare ad una corretta ed accurata restituzione. La precisione del rilievo richiesta e la sua accuratezza dipendono esclusivamente dalle sue finalità.

10.1.1 Strumentazione

Ciò che governa la scelta della strumentazione da utilizzare, in questo caso la scelta della fotocamera, è sempre funzione del dettaglio che si vuole raggiungere con l'immagine o il filmato che si acquisisce. Più la risoluzione del nostro sensore è grande più a parità di distanza di presa sarà elevato il dettaglio dell'immagine. Più è elevato il dettaglio, più sarà precisa la fase di orientamento e di autocalibrazione della stessa. Tipicamente si cerca di capire in questa fase quale potrebbe essere il GSD necessario a descrivere in maniera accurata la scena del nostro rilievo.

Oggi in commercio si trovano fotocamere di tutti i tipi e dalle caratteristiche piuttosto simili, una netta differenza la fanno ancora i modelli tipo reflex (come quello utilizzato nel primo caso di studio)

rispetto le normali digitali compatte. Eventualmente una dotazione composta da una fotocamera tipo reflex per gli elementi di maggior dettaglio ed una ultraleggera da poter anche utilizzare accoppiata ad un piccolo drone potrebbero risolvere qualsiasi tipo di casistica. Con una coppia di camere di questo tipo, lavorando poi sulla distanza di presa è possibile ottenere GSD molto variabili ed adattabili a seconda delle varie esigenze.

La possibilità di disporre di un piccolo RPAS anche di tipo low-cost come quello utilizzato nel secondo caso di studio è sicuramente un vantaggio in quanto ci consente di raggiungere anche quelle zone che a causa di occlusioni o distanze elevate risultano di difficile acquisizione da parte di un operatore a terra.

È sempre necessario inoltre disporre almeno di una stazione totale, (o di un GPS) per poter acquisire con precisione i punti fotografici di appoggio (o i ground control point) necessari per l'orientamento assoluto delle nostre immagini. Non si può trascurare tale fase in quanto oltre ad inquadrare e scalare correttamente il modello creato serve ad ottimizzare, attraverso una procedura di bundle adjustment, l'orientamento delle immagini e di conseguenza la triangolazione densa dei punti che ne consegue. A questo scopo è bene anche dotarsi di opportuni target ben visibili nelle immagini, da utilizzare quando non siano disponibili punti naturali da utilizzare come punti fotografici di appoggio.

10.1.2 Pianificazione della strategia di presa

Scelta la strumentazione più idonea e fissati quindi alcuni parametri come pixel size, focale e GSD si può pianificare la strategia di presa (o piano di volo se si lavora con un drone). Si ricorda come le dimensioni del GSD siano direttamente proporzionali alla qualità del modello creato, sia in termini di precisione che in termini di definizione degli oggetti da rilevare. La scelta del GSD può anche essere variabile in funzione degli elementi presenti nella scena. In particolare si possono distinguere due casistiche: fotogrammetria terrestre (close range photogrammetry) e fotogrammetria aerea (tramite l'utilizzo di RPAS).

10.1.2.1 Fotogrammetria terrestre

I casi che prevedono l'utilizzo di questa metodologia sono molteplici, dal rilievo di interi edifici, a quello dettagliato di singoli prospetti, o porzioni di edifici, come bassorilievi, o altri elementi architettonici significativi. In questo caso è consigliabile eseguire una serie di riprese con assi perpendicolari alle facciate. Il ricoprimento sufficiente può essere dell'ordine del 60% in entrambe le direzioni anche se resta consigliabile un ricoprimento maggiore così da poter gestire in autonomia le diverse immagini acquisite, creando diversi dataset a seconda delle esigenze. Un approccio di questo tipo scongiura problematiche causate da immagini sfocate, mosse o da zone con scarso ricoprimento o mancanti.

Dove possibile variare sempre la base di presa tra due immagini successive. Se necessario in caso di rilievo di facciate ad elevato sviluppo verticale è possibile inclinare leggermente la camera verso l'alto per riprendere le zone più alte dell'edificio; in questo caso otterremo una serie di immagini con base di presa nulla. Onde evitare problemi in fase di triangolazione dei punti è consigliabile di variare almeno la base di presa in direzione orizzontale.

Qualora lo si ritenga necessario è preferibile avvicinarsi all'oggetto da rilevare per acquisire immagini ad elevata risoluzione se si rileva la presenza di particolari architettonici significativi.

Oltre alle immagini acquisite per una completa copertura dell'oggetto secondo la percentuale di ricoprimento stabilita, acquisire sempre anche immagini d'insieme, soprattutto nelle zone d'angolo o di unione di più pareti. Saranno immagini che verranno utilizzate soltanto in fase di orientamento per agevolare il riconoscimento automatico delle corrispondenze. Terminata la fase di orientamento, quindi nota la posizione del centro di presa delle varie camere, è consigliabile eliminare tali immagini prima di iniziare il processo di ricostruzione densa, in quanto come dimostrato dalle esperienze, immagini con GSD maggiori tendono a ridurre la precisione e la qualità del rilievo.

10.1.2.2 Fotogrammetria aerea

Applicabile in tutti i casi in cui vi siano zone non visibili da terra, questa particolare metodologia si esplica tramite l'utilizzo di un RPAS in grado di trasportare una qualsiasi fotocamera di tipo digitale. Come per le riprese terrestri anche in questo caso è opportuno operare con immagini ad assi perpendicolari rispetto l'oggetto che si intende rilevare (assi nadirali se si rilevano terreno o coperture, assi orizzontali se l'oggetto del rilievo è una facciata). Anche in questo caso il ricoprimento necessario è intorno al 60%, mentre quello consigliato rimane sull'80%. Ancora prese ad assi inclinati possono favorire il corretto orientamento delle immagini oltre che completare il rilievo.

La loro applicazione al modello va però utilizzata con cura, poiché alte precisioni si ottengono con piccole distanze di presa e solitamente con i droni è difficile oltre che sconsigliato avvicinarsi troppo all'oggetto da riprendere.

10.1.3 Calibrazione della camera

Questa fase può essere svolta sia in laboratorio come fase preliminare oppure può essere affidata completamente al software di restituzione. Le nostre esperienze non mostrano particolari problemi durante l'esecuzione di autocalibrazioni, anche se una buona self-calibration aiuta lo stesso software alla risoluzione dell'orientamento interno, soprattutto con fotocamere affette da elevate distorsioni. Si ricorda come la procedura di autocalibrazione passi attraverso una stima a convergenza di matrici, pertanto partendo da una calibrazione disponibile, anche se non perfetta riduce i passaggi necessari a portare l'algoritmo a convergenza.

10.1.4 Rilievo dei Punti Fotografici di Appoggio

È una fase importante e necessaria per una corretta restituzione. I punti fotografici di appoggio devono essere rilevati con un'elevata precisione, o almeno con precisione maggiore delle dimensioni del GSD dell'immagine. Questo poiché il riconoscimento di tali punti sui fotogrammi è funzione delle dimensioni del pixel. Possono essere punti naturali o target appositamente realizzati per una migliore collimazione e riconoscimento sulle immagini.

È meglio disporre di un elevato numero di punti rilevati, poiché non sempre sulle immagini si ottengono un numero sufficiente di proiezioni per una corretta triangolazione. Non dovranno necessariamente essere utilizzati tutti, in quanto con i moderni software di restituzione ne bastano pochi ma ben distribuiti sull'intera scena. I punti non utilizzati per l'orientamento possono essere utilizzati come punti di controllo finale per valutare la bontà del lavoro svolto.

10.2 AQUISIZIONE DEL DATO

Questa è la fase di presa di fotogrammi. Non ci sono indicazioni particolarmente esaustive o significative. Le immagini devono essere nitide e ferme. Se si lavora con un RPAS è bene valutare attentamente la velocità del drone durante il rilievo così da scongiurare il pericolo di immagini mosse. Allo stesso modo settare la fotocamera con un tempo di esposizione compatibile con l'intervallo di scatto tra una foto e la successiva, così da poter mantenere costante la velocità del drone durante il rilievo.

10.3 ELABORAZIONE DEL DATO

L'ultima fase del rilievo consiste nell'elaborazione delle immagini acquisite. Questa fase è strettamente legata al software che si utilizza per la gestione del progetto. È probabilmente la parte meno gestibile del rilievo, ma sicuramente la più importante. L'automazione raggiunta dai software commerciali offre oggi la possibilità a chiunque di produrre elaborati fotogrammetrici spesso con risultati all'apparenza discreti. Esistono anche software open source completamente editabili e personalizzabili in ogni loro fase per le esigenze degli utenti più esperti.

In questo caso l'esperienza svolta in entrambi i casi di studio non si traduce in indicazioni precise in quanto si è voluto testare appositamente le impostazioni di default di uno dei più famosi software commerciali, PhotoScan di Agisoft. PhotoScan, per la realizzazione di modelli tridimensionali opera utilizzando il principio della ricostruzione multivista non calibrata osservato nella prima parte della tesi. L'allineamento delle immagini e la ricostruzione del modello 3D avvengono entrambe in maniera del tutto automatica.

Generalmente l'obiettivo finale del processo fotogrammetrico con PhotoScan è proprio la ricostruzione del modello tridimensionale texturizzato e le procedure impiegate a tal fine comprendono alcune fasi ricorrenti che andremo ad elencare:

1. **Allineamento delle immagini (align photos):** in questa fase il software risolve l'orientamento relativo del blocco fotogrammetrico. Per inizializzare tale procedura viene dapprima svolta la ricerca delle corrispondenze e la loro correlazione (processo di feature detection).



Figura 10.1. Immagine originale (sinistra). Punti caratteristici individuati sull'immagine (destra)

Successivamente filtra ed elimina gli outliers ed esegue l'accoppiamento delle corrispondenze trovate durante la prima fase.

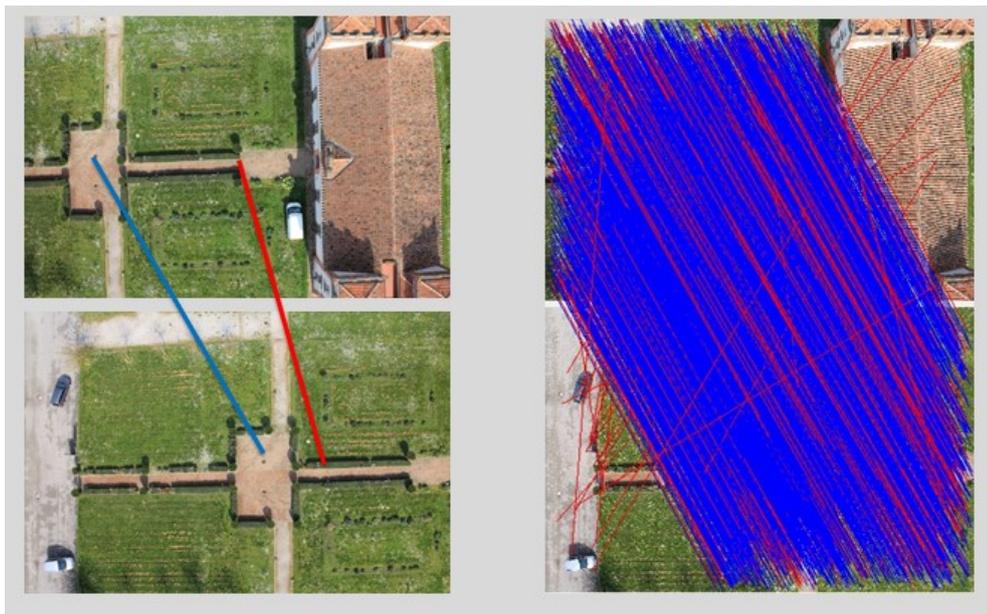


Figura 10.2. Stima delle corrispondenze tra le features trovate. La linea rossa indica la corrispondenza errata stimata tra due features in due immagini distinte. La linea blu indica la presenza di una corrispondenza esatta. A destra tutte le corrispondenze trovate tra due immagini campione

Infine viene ricostruita la nube sparsa della scena in cui si osservano le posizioni 3D delle features trovate e le posizioni delle fotocamere al momento dell'acquisizione delle immagini.

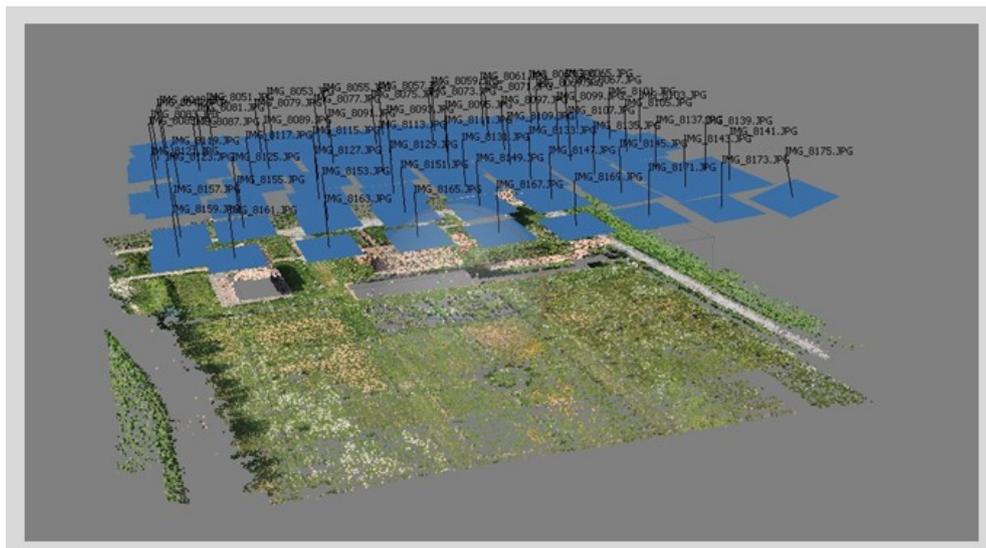


Figura 10.3. Esempio di orientamento relativo del blocco fotogrammetrico

Il tutto si trova in un sistema di riferimento completamente arbitrario. Il modello è realizzato a meno di un fattore di scala.

2. **Orientamento esterno (optimize alignment):** è la fase in cui vengono inseriti i GCP o i PFA e attraverso una procedura di bundle adjustment viene risolto l'orientamento esterno delle camere. Al termine della fase sono note le posizioni dei centri di presa delle fotocamere nel sistema di riferimento assoluto.

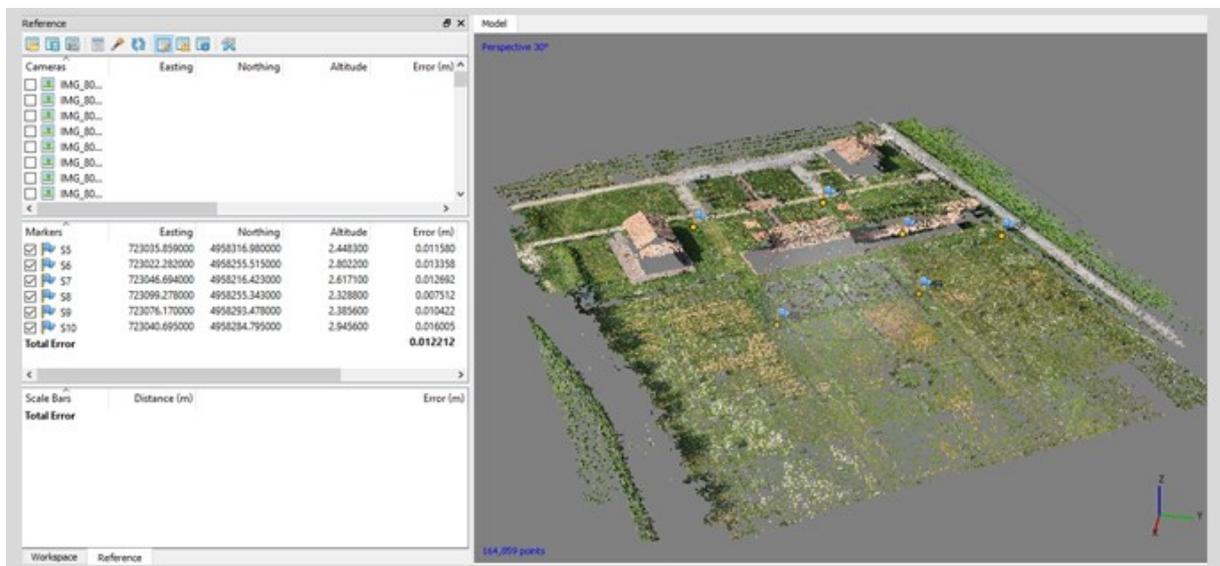


Figura 10.4. Inserimento dei ground control point

3. **Ricostruzione densa (build dense cloud):** Nota la geometria di presa, inizia il processo di correlazione automatica delle immagini attraverso algoritmi di dense matching. Pixel dopo pixel viene ricostruita l'intera geometria della scena. Ad ogni punto viene associato il corrispondente valore RGB mediato, ricavato dalle proiezioni immagine trovate di quel punto.



Figura 10.5. Nube densa di punti ottenuta da una sequenza di immagini digitali.

4. **Costruzione del modello tridimensionale (build mesh):** Viene ricostruita la superficie solida dell'oggetto attraverso meshatura a triangoli (T.I.N.) collegando i punti della nube con elementi solidi (mesh). Generalmente gli algoritmi di calcolo per generare le mesh usati da Photoscan sono due: il primo adatto per superfici planari, *Height Field* ed il secondo per qualsiasi tipo di oggetto, *Arbitrary*.

Costruita la mesh, può essere necessario editarla rimuovendo parti indesiderate.

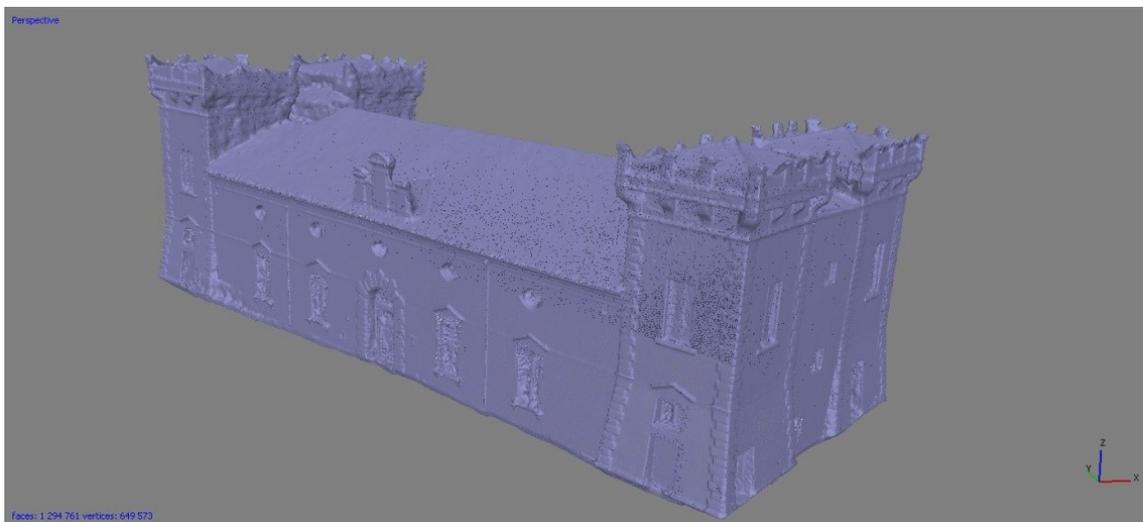


Figura 10.6. Modello tridimensionale non texturizzato creato partendo dalla nube di punti realizzata al passo precedente.

5. **Applicazione della texture (build texture):** L'ultima fase prevede vari modi possibili di texturizzazione della mesh. La texture da applicare al modello solido viene ricavata dalle immagini a disposizione.



Figura 10.7. Modello tridimensionale texturizzate del "Castello del Verginese"

6. **Estrazione dei risultati (export):** al termine della restituzione è possibile esportare i dati ottenuti in diversi formati. Si possono salvare i punti della nube semplicemente in .txt oppure esportare l'intero modello tridimensionale texturizzato. Tipicamente vengono anche create delle viste del modello tridimensionale, a discrezione dell'operatore che vengono poi esportate come ortofoto. Vi è la possibilità di georeferenziare tali immagini. Non vengono create delle vere e proprie ortofoto ma le viste in proiezione ortografica del modello creato.



Figura 10.8. Ortofoto della Facciata frontale (prospetto Ovest). È possibile vettorializzare l'immagine in ambiente CAD per restituire la tavola del prospetto del fabbricato in esame.

ALTRE APPLICAZIONI

11 FOTOGRAMMETRIA APPLICATA ALLO STUDIO DI MECCANISMI DI COLLASSO DI TRAVI IN C.A.

11.1 INTRODUZIONE

Tipicamente la fotogrammetria è associata alla topografia, o alla cartografia, ma può trovare numerose applicazioni anche in affiancamento ad altre discipline, come ad esempio la Scienza delle Costruzioni. La sperimentazione in cui si colloca il nostro rilievo è finalizzata allo studio dei meccanismi di collasso di particolari travi in cemento armato, sottoposte a diversi interventi di miglioramento sismico. In particolare si vogliono testare le potenzialità dei metodi fotogrammetrici applicati alle prove di laboratorio tipiche dell'Ingegneria Strutturale.

In particolare sono state analizzate le prove di rottura condotte su un campione di venti travi in conglomerato cementizio armato, all'interno di una ricerca condotta per la sperimentazione e lo sviluppo di una tecnica chiamata ETS, acronimo di *Embedded Trough Section*, consistente nell'inserimento di barre in acciaio o in fibra di carbonio (CFRP) in appositi fori realizzati sulle travi oggetto dell'intervento di rinforzo. La suddetta ricerca è coordinata dalla Prof.ssa Alessandra Aprile del Dipartimento di Ingegneria dell'Università degli Studi di Ferrara.

Attraverso l'acquisizione di numerosissimi fotogrammi ad alta precisione si vuole misurare, non solo l'evoluzione delle fessure durante la fase di carico e di rottura (sviluppo ed ampiezza), ma anche la deformazione della trave durante l'arco dell'intera prova. Interessante sarà osservare inoltre la ricostruzione del diagramma sforzi deformazioni, conoscendo il carico applicato e misurando la deformata per i diversi intervalli di carico. I risultati di questa sperimentazione verranno successivamente validati attraverso il confronto diretto con quanto misurato dai trasduttori presenti durante la prova.

11.2 ALLESTIMENTO DELLA PROVA

Le prove di carico sono state condotte su una serie di travi in cemento armato, di forma a T, con le dimensioni riportate in figura 11.1:

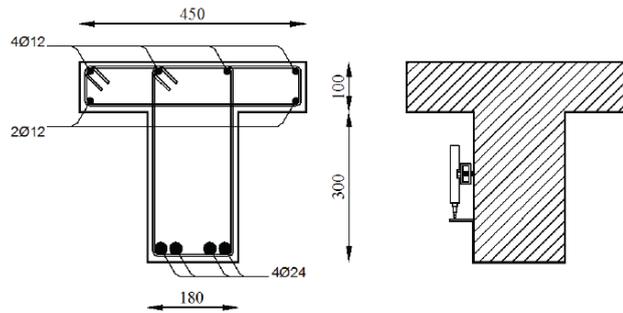


Figura 11.1. sezione trasversale dei provini in C.A.

La distanza tra gli appoggi, la posizione del martinetto e la posizione dei trasduttori sono indicate nella figura 11.2.

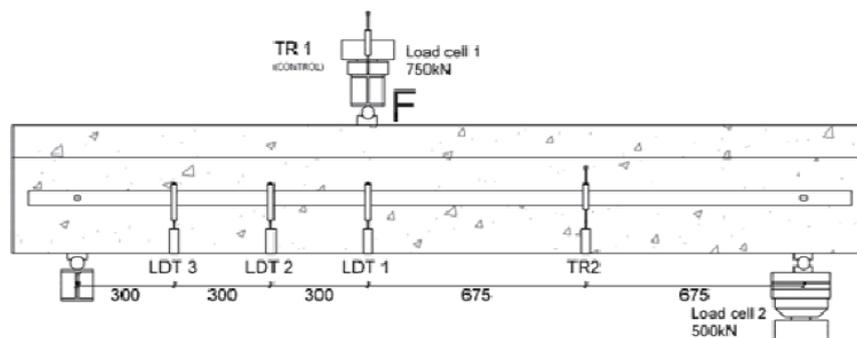


Figura 11.2. disposizione dei trasduttori di spostamento

Durante ciascuna prova sono stati acquisiti i dati derivanti dal carico applicato dal martinetto idraulico provenienti dalla cella di carico ed i dati relativi allo spostamento misurati in continuo dai tre trasduttori. Saranno questi i dati che verranno utilizzati per il confronto con le misure fotogrammetriche.

11.3 IL RILIEVO FOTOGRAMMETRICO

Il rilievo fotogrammetrico proposto si basa sull'utilizzo di fotogrammi opportunamente rettificati, acquisiti con frequenze diverse durante la prova. Si va da pochi scatti al minuto (istanti iniziali) fino a 4-5 immagini al secondo man mano che ci si avvicina all'istante della rottura. Questo poiché si vuole osservare, non solo la deformazione della trave durante i vari step di carico, ma anche l'evoluzione del quadro fessurativo (sviluppo e ampiezza) durante tutta la fase di carico. Si è deciso di utilizzare le immagini e non una ripresa video poiché la loro risoluzione è di gran lunga superiore rispetto al formato HD con cui sarebbe stato acquisito un filmato.

La strumentazione necessaria per fare quanto appena descritto consiste essenzialmente in una fotocamera digitale ad alta risoluzione e soprattutto in un obiettivo con piccole distorsioni ottiche

che verranno successivamente calibrate. È stata pertanto scelta una **CANON EOS 7D** con obiettivo **EF 35mm f/1.4L USM**.

Per la rettifica dei fotogrammi è stato necessario inserire nella scena alcuni punti di coordinate note, rilevabili tramite l'utilizzo di opportuni target e una stazione totale (nel nostro caso una **LEICA TS06**). Per rendere l'illuminazione della scena idonea alla individuazione anche delle più piccole lesioni sono stati utilizzati due farette da 500W, che eliminano qualsiasi tipo di ombra sull'anima della trave. Infine per poter associare alle immagini acquisite il carico applicato durante le varie fasi della prova è stato inserito nel campo visivo della camera un semplice timer sincronizzato con la cella di carico principale. Nelle immagini seguenti possiamo osservare l'allestimento della prova:

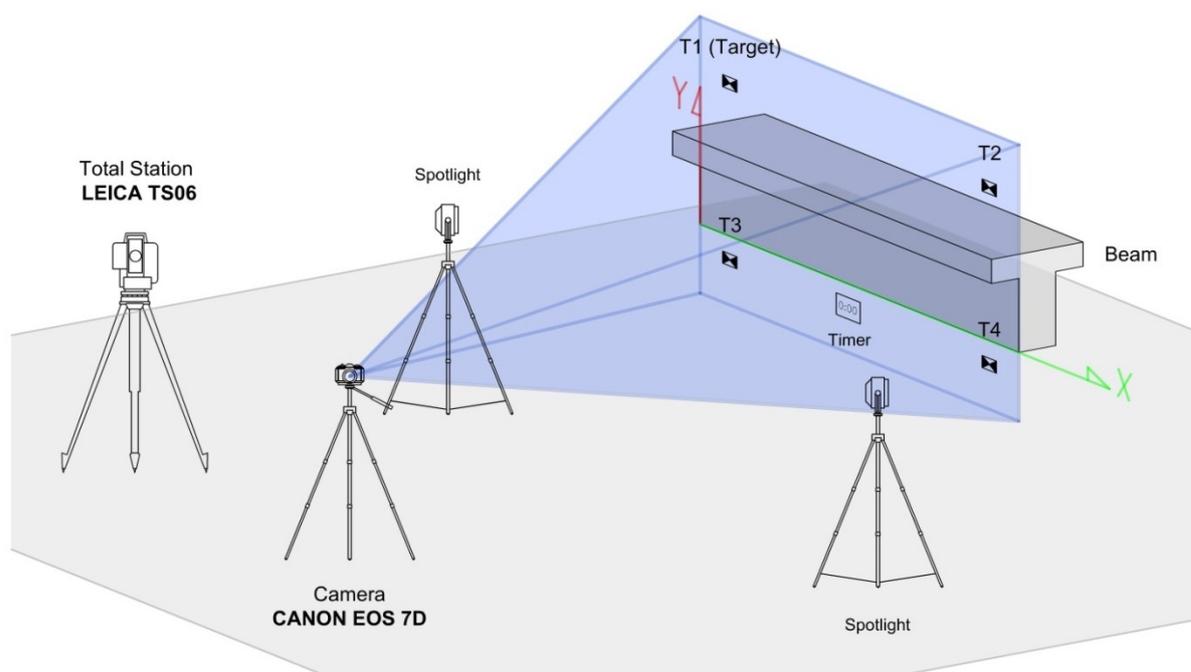


Figura 11.3. Allestimento della prova fotogrammetrica

La prima fase del rilievo consiste nella materializzazione del nostro sistema di riferimento, indispensabile se si vogliono eseguire misure sull'immagine acquisita. Le fasi di elaborazione dei fotogrammi che vorremmo applicare sono essenzialmente due: **correzione della distorsione, e raddrizzamento**.

Per la prima è fondamentale calibrare la fotocamera utilizzata e studiarne il polinomio di distorsione, mentre per la seconda è importante dover riconoscere un piano su cui poter effettuare il raddrizzamento.

Al termine della procedura di calibrazione, possiamo stabilire quale sia la distanza di presa più idonea per eseguire le fasi di lavoro successive. Per questioni operative si è deciso di posizionare la camera

ad una distanza di circa 5 metri dall'oggetto da rilevare. A questa distanza infatti otteniamo un GSD pari a circa 0,55 mm e possiamo osservare l'intera trave all'interno del campo visivo dell'immagine.

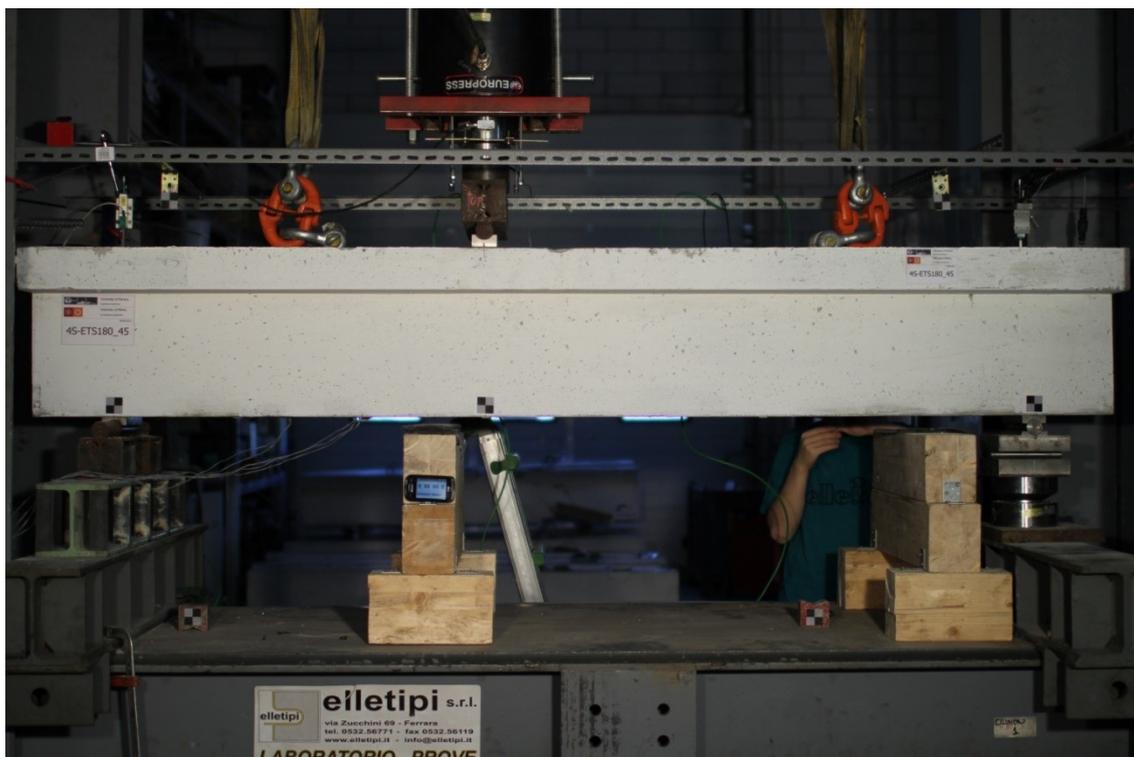


Figura 11.4. Esempio di immagine acquisita durante la prova di carico.

Prima di iniziare ciascuna prova è necessario rilevare con la stazione totale la posizione dei target inseriti, così da poter raddrizzare le immagini secondo il piano medio contenente l'anima della trave.

Tale procedura è fondamentale se si vogliono evitare errori di misura legati ad effetti prospettici delle immagini. È nostra intenzione lasciare la trave il più pulita possibile senza aggiungere altri target in prossimità dell'anima, onde evitare che le lesioni possano essere coperte e quindi risultare poco visibili. Per materializzare al meglio il piano di raddrizzamento sono stati adottati pochi target di tipo rimovibile che verranno posizionati con cura ai quattro angoli dell'immagine e rilevati con attenzione prima di ciascuna prova. Probabilmente sarebbe stato più semplice utilizzare un sistema di riferimento fisso, solidale al telaio, così da rilevarne le coordinate soltanto una volta ma questo tipo di soluzione andava incontro ad alcune problematiche relative alle procedure di sostituzione delle travi tra una prova e l'altra.

Per ciascuna trave sono state acquisite dalle 3000 alle 6000 fotografie così da formare un database completo di tutto ciò che accade durante l'arco dell'intera prova. Una volta in laboratorio da questo database sono state selezionate soltanto alcune immagini e soltanto su queste verranno eseguite le operazioni di rettifica necessarie per eseguire le misurazioni richieste. La distorsione è rimossa in maniera automatica sempre dal software Photomodeler Scanner mentre le immagini sono state raddrizzate con il software open source RDF. Il prodotto che si ottiene al termine di queste procedure

non è altro che un'immagine opportunamente ricampionata sulla quale, a meno di un fattore di scala noto, è possibile effettuare ogni tipo di misura si desideri eseguire. Associando dunque ad ogni intervallo di carico una immagine possiamo misurarne la deformata, studiarne il quadro fessurativo, misurare l'ampiezza delle fessure e ricostruire la curva sforzi deformazioni, ottenendo da una semplice immagine tutte le informazioni necessarie per lo studio della prova. Le misure vengono realizzate attraverso software di disegno automatico, come ad esempio AutoCAD e fogli excel per l'estrazione e l'interpretazione dei dati.

11.4 PRIMI CONFRONTI

Al termine della fase di sperimentazione in laboratorio iniziano i primi test di validazione del metodo. Come accennavo in precedenza, un confronto tra gli spostamenti rilevati dai trasduttori e quelli misurati sui fotogrammi è necessario per comprendere il grado di affidabilità della metodologia proposta. Tali misure saranno ripetute per ciascuna trave a determinati intervalli di carico.

Di seguito analizzeremo in dettaglio i risultati ottenuti per la trave 4S ETS 180-90. Le immagini sono state scattate a circa 4,89 metri di distanza dalla trave. Noti i parametri di orientamento interni della fotocamera possiamo calcolare il GSD dell'immagine, che si attesta intorno a 0,6 mm. Utilizzeremo tale valore come stima della precisione.

ID FOTO	MISURE TRASDUTTORI			CARICO	MISURE FOTOGAMMETRICHE			RESIDUI		
	LDT1 mm	LDT2 mm	LDT3 mm		LDT1 mm	LDT2 mm	LDT3 mm	LDT1 mm	LDT2 mm	LDT3 mm
5	0	0	0	0	0	0	0	0	0	0
21	1,233	0,891	0,426	129,375	0,9800	0,8867	0,3933	0,2530	0,0043	0,0327
24	1,647	1,293	0,711	150,12	1,8200	1,5800	0,7400	0,1730	0,2870	0,0290
29	2,235	1,779	1,047	196,605	2,2400	2,0267	1,0133	0,0050	0,2477	0,0337
35	3,084	2,514	1,5	252,72	3,2200	2,7800	1,4400	0,1360	0,2660	0,0600
166	3,759	3,141	1,932	300,105	3,9600	3,3733	1,7867	0,2010	0,2323	0,1453
278	4,014	3,366	2,073	314,01	4,1200	3,6800	1,8400	0,1060	0,3140	0,2330
587	4,752	4,098	2,52	350,145	4,8000	4,4000	2,2000	0,0480	0,3020	0,3200
1099	5,796	5,088	3,09	400,05	6,1400	5,3267	2,6133	0,3440	0,2387	0,4767
1287	6,324	5,676	3,39	413,19	6,6600	5,8733	3,0867	0,3360	0,1973	0,3033
1370	6,717	6,36	4,296	366,93	7,2200	6,8467	4,0733	0,5030	0,4867	0,2227

Tabella 11.1. Report del calcolo delle differenze tra gli spostamenti misurati tramite i trasduttori e quelli rilevati con tecniche fotogrammetriche. Tutti gli spostamenti trovati sono minori delle dimensioni del GSD assunte come tolleranza.

Come possiamo osservare i residui sono sempre minori del valore del GSD (<0,6 mm) pertanto possiamo considerare gli errori trascurabili. Questo ci consente di affermare con una discreta precisione che le misure effettuate sono coerenti con quanto realmente è stato monitorato dai trasduttori in fase di prova.

11.5 STIMA DELLE DEFORMAZIONI

Avendo accertato l'affidabilità del metodo possiamo passare alla stima delle deformazioni. Avendo a disposizione il carico applicato è lo spostamento misurato possiamo calcolare gli stress path relativi ai tre trasduttori. Ancora una volta confronteremo i risultati ottenuti con la fotogrammetria.

Di seguito mostreremo i grafici relativi ai diagrammi dei tre trasduttori.

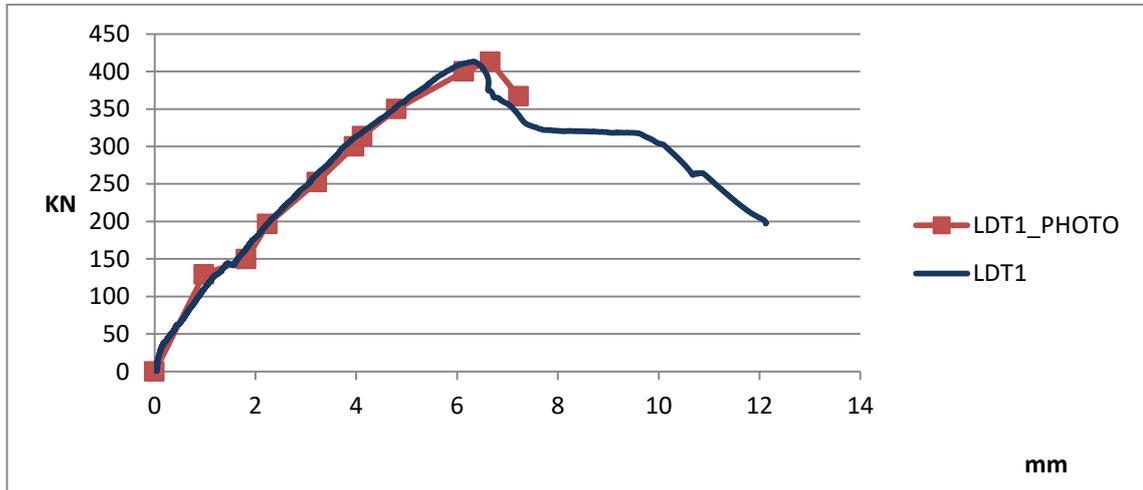


Figura 11.5. Diagramma CARICO-SPOSTAMENTO realizzato per via fotogrammetrica (linea rossa). La linea blu indica invece il diagramma carico-spostamento misurato con l'utilizzo dei trasduttori. (Il digramma si riferisce al primo trasduttore)

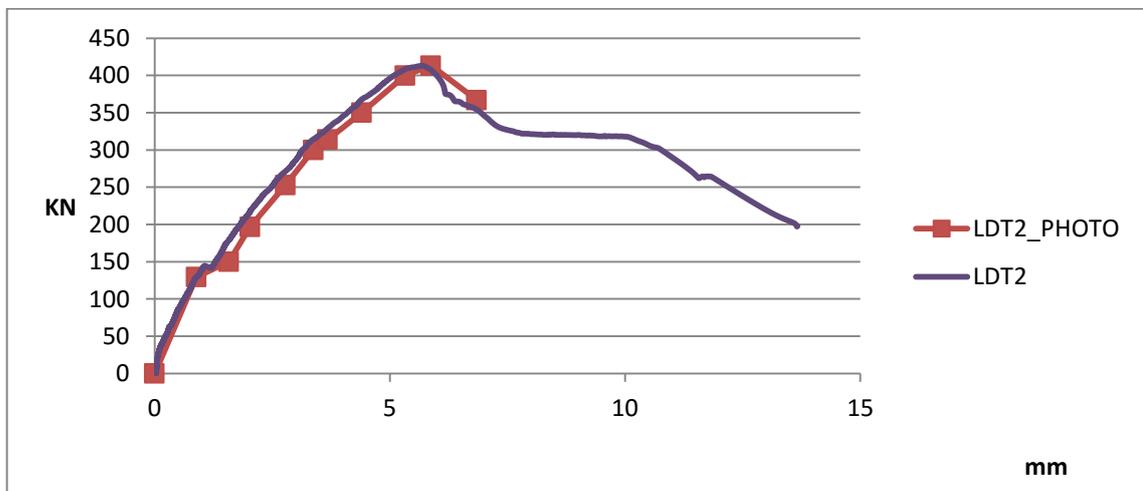


Figura 11.6. Diagramma CARICO-SPOSTAMENTO realizzato per via fotogrammetrica (linea rossa). La linea blu indica invece il diagramma carico-spostamento misurato con l'utilizzo dei trasduttori. (Il digramma si riferisce al secondo trasduttore)

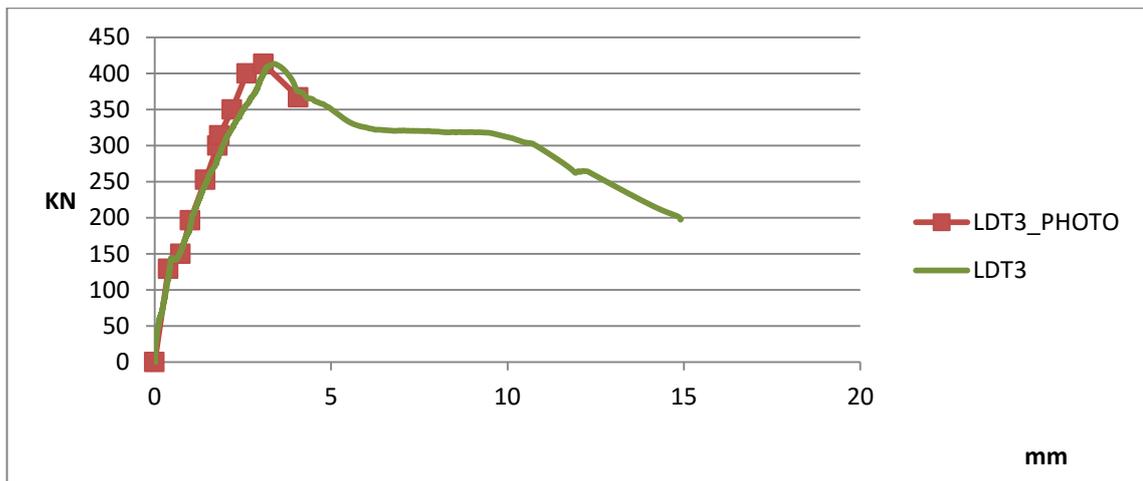


Figura 11.7. Diagramma CARICO-SPOSTAMENTO realizzato per via fotogrammetrica (linea rossa). La linea blu indica invece il diagramma carico-spostamento misurato con l'utilizzo dei trasduttori. (Il digramma si riferisce al terzo trasduttore)

Ancora una volta possiamo osservare una buona corrispondenza. In tabella si riportano i risultati relativi alle altre travi analizzate:

TRAVE 15

ID immagine	Carico [kN]	Valori veri			Valori rilevati (depurati da rototraslazioni)			Differenze		
		LDT1	LDT2	LDT3	LDT1	LDT2	LDT3	LDT1	LDT2	LDT3
		[mm]	[mm]	[mm]	[mm]	[mm]	[mm]	[mm]	[mm]	[mm]
1	0	0	0	0	0.000	0.000	0.000	0.000	0.000	0.000
19	172.67	1.602	1.302	0.702	2.020	1.347	0.873	0.418	0.045	0.171
21	192.69	1.836	1.497	0.825	2.200	1.500	1.000	0.364	0.003	0.175
23	213.44	2.091	1.695	0.933	2.540	1.793	1.247	0.449	0.098	0.314
356	250.02	2.706	2.148	1.212	3.240	2.460	1.580	0.534	0.312	0.368
797	300.15	3.474	2.724	1.476	4.040	2.693	1.747	0.566	-0.031	0.271
1193	350.15	4.230	3.291	1.788	4.800	3.167	2.133	0.570	-0.124	0.345
1530	388.98	4.869	3.753	2.139	5.440	3.593	2.547	0.571	-0.160	0.408
1635	400.01	5.097	3.951	2.265	5.640	3.793	2.647	0.543	-0.158	0.382
2084	450.09	6.069	4.734	2.784	6.640	4.560	3.080	0.571	-0.174	0.296
2572	500.04	7.149	5.601	3.339	7.640	5.227	3.513	0.491	-0.374	0.174
2613	503.87	7.248	5.682	3.396	7.800	5.500	3.700	0.552	-0.182	0.304
3267	550.08	8.661	6.981	4.254	9.100	7.300	4.300	0.439	0.319	0.046
5417	566.42	11.007	9.678	5.736	10.520	9.247	5.473	-0.487	-0.431	-0.263

Tabella 11.2. Rapporto di prova – trave 15

TRAVE 16

ID immagine	Carico	Valori veri			Valori rilevati (depurati da rototraslazioni)			Differenze		
		LDT1	LDT2	LDT3	LDT1	LDT2	LDT3	LDT1	LDT2	LDT3
	[kN]	[mm]	[mm]	[mm]	[mm]	[mm]	[mm]	[mm]	[mm]	[mm]
1	0.00	0	0	0	0.000	0.000	0.000	0.000	0.000	0.000
17	147.11	1.509	1.275	0.630	1.340	1.527	1.013	-0.169	0.252	0.383
23	202.86	2.307	1.980	1.080	2.240	2.227	1.513	-0.067	0.247	0.433
28	250.02	3.087	2.601	1.389	3.280	2.820	1.860	0.193	0.219	0.471
166	270.18	3.417	2.823	1.521	3.360	3.240	2.020	-0.057	0.417	0.499
409	300.11	3.936	3.270	1.713	3.940	3.527	2.213	0.004	0.257	0.500
455	305.78	4.038	3.363	1.755	4.060	3.873	2.287	0.022	0.510	0.532
634	325.04	4.500	3.768	2.001	4.660	4.240	2.520	0.160	0.472	0.519
934	350.06	5.211	4.485	2.274	5.340	4.993	2.747	0.129	0.508	0.473
1046	360.00	5.472	4.722	2.379	5.660	5.273	2.887	0.188	0.551	0.508
1189	370.49	5.886	5.157	2.592	6.300	5.867	3.033	0.414	0.710	0.441

Tabella 11.3. Rapporto di prova – trave 16

TRAVE 17

ID immagine	Carico	Valori veri			Valori rilevati (depurati da rototraslazioni)			Differenze		
		LDT1	LDT2	LDT3	LDT1	LDT2	LDT3	LDT1	LDT2	LDT3
	[kN]	[mm]	[mm]	[mm]	[mm]	[mm]	[mm]	[mm]	[mm]	[mm]
1	0.00	0	0	0	0.000	0.000	0.000	0.000	0.000	0.000
22	150.35	1.524	1.248	0.606	1.740	1.660	0.980	0.216	0.412	0.374
28	205.56	2.415	2.040	1.089	2.520	2.213	1.307	0.105	0.173	0.218
33	253.04	3.276	2.754	1.521	3.500	3.033	1.967	0.224	0.279	0.446
77	300.06	4.014	3.351	1.896	4.300	3.833	2.267	0.286	0.482	0.371
557	350.10	4.977	4.143	2.418	5.260	4.507	2.653	0.283	0.364	0.235
986	400.14	5.835	4.890	2.850	6.220	5.380	3.240	0.385	0.490	0.390
1007	402.35	5.877	4.938	2.871	6.380	5.453	3.327	0.503	0.515	0.456
1133	416.25	6.165	5.193	3.033	6.640	5.660	3.380	0.475	0.467	0.347
1483	450.45	6.990	6.081	3.477	7.500	6.533	3.667	0.510	0.452	0.190
1615	461.84	7.326	6.429	3.660	7.860	6.840	3.820	0.534	0.411	0.160
1931	483.35	8.079	7.302	4.068	8.520	7.713	4.107	0.441	0.411	0.039
2070	500.45	8.523	7.791	4.332	9.000	8.267	4.333	0.477	0.476	0.001
2078	512.42	8.733	7.989	4.464	9.380	8.387	4.593	0.647	0.398	0.129
2089	534.69	9.531	8.859	4.938	10.000	9.267	4.833	0.469	0.408	-0.105

Tabella 11.4. Rapporto di prova – trave 17

TRAVE 18

ID immagine	Carico [kN]	Valori veri			Valori rilevati attraverso raddrizzamento geometrico			Differenze		
		LDT1	LDT2	LDT3	LDT1	LDT2	LDT3	LDT1	LDT2	LDT3
		[mm]	[mm]	[mm]	[mm]	[mm]	[mm]	[mm]	[mm]	[mm]
1	-	0	0	0	0	0	0	0	0	0
15	103.91	1.068	0.699	0.360	1.0	0.7	0.5	-0.068	0.001	0.140
18	-	1.317	0.891	0.447	1.6	1.3	0.8	0.283	0.409	0.353
21	147.96	1.728	1.245	0.630	1.9	1.5	1.0	0.172	0.255	0.370
26	193.64	2.442	1.809	1.089	2.8	2.1	1.4	0.358	0.291	0.311
32	249.35	3.390	2.499	1.530	3.5	2.8	1.8	0.110	0.301	0.270
287	300.02	4.269	3.132	1.914	4.5	3.5	2.3	0.231	0.368	0.386
708	350.06	5.157	3.804	2.241	5.4	4.3	2.6	0.243	0.496	0.359
804	360.00	5.349	3.954	2.304	5.7	4.5	2.8	0.351	0.546	0.496
1097	376.88	6.201	4.944	2.571	6.4	5.6	3.1	0.199	0.656	0.529

Tabella 11.5. Rapporto di prova – trave 18

TRAVE 19

ID immagine	Carico [kN]	Valori veri			Valori rilevati attraverso raddrizzamento geometrico			Differenze		
		LDT1	LDT2	LDT3	LDT1	LDT2	LDT3	LDT1	LDT2	LDT3
		[mm]	[mm]	[mm]	[mm]	[mm]	[mm]	[mm]	[mm]	[mm]
40	303.17	3.417	2.706	1.491	3.417	2.706	1.491	0.000	0.000	0.000
405	350.06	4.128	3.345	1.824	4.4	3.1	2.1	0.272	-0.245	0.276
683	387.14	4.713	3.879	2.139	4.8	3.6	2.3	0.087	-0.279	0.161
787	400.10	4.935	4.080	2.232	5.1	3.9	2.6	0.165	-0.180	0.368
1195	450.09	5.772	4.824	2.613	5.7	4.3	2.8	-0.072	-0.524	0.187
1625	500.09	6.723	5.676	3.120	7.1	5.7	3.4	0.377	0.024	0.280
2522	550.04	7.974	6.873	3.888	8.0	6.4	4.0	0.026	-0.473	0.112
4363	600.08	10.266	9.225	5.580	10.2	8.7	5.1	-0.066	-0.525	-0.480
5600	616.86	11.979	11.157	7.356	12.2	10.7	6.4	0.221	-0.457	-0.956

Tabella 11.6. Rapporto di prova – trave 19

TRAVE 20

ID immagine	Carico [kN]	Valori veri			Valori rilevati attraverso raddrizzamento geometrico			Differenze		
		LDT1	LDT2	LDT3	LDT1	LDT2	LDT3	LDT1	LDT2	LDT3
		[mm]	[mm]	[mm]	[mm]	[mm]	[mm]	[mm]	[mm]	[mm]
16	427.70	0.891	0.657	0.357	0.891	0.657	0.357	0.000	0.000	0.000
21	549.90	1.521	1.188	0.654	2.000	1.700	1.000	0.479	0.512	0.346
25	733.20	2.181	1.713	1.017	2.600	2.100	1.500	0.419	0.387	0.483
135	855.40	2.925	2.358	1.518	3.400	2.800	2.000	0.475	0.442	0.482
346	916.50	3.345	2.808	1.701	3.900	3.300	2.200	0.555	0.492	0.499
421	977.60	3.534	3.006	1.785	4.000	3.400	2.300	0.466	0.394	0.515
716	1038.70	4.206	3.723	2.223	4.700	4.000	2.500	0.494	0.277	0.277
821	1099.80	4.461	3.978	2.415	5.000	4.300	2.700	0.539	0.322	0.285
900	1099.80	4.653	4.161	2.550	5.300	4.500	2.900	0.647	0.339	0.350
1161	1160.90	5.508	5.268	4.032	6.200	5.800	4.000	0.692	0.532	-0.032

Tabella 11.7. Rapporto di prova – trave 20

11.6 CONSIDERAZIONI FINALI

Tutte le travi analizzate hanno mostrato una buona corrispondenza con il dato dei trasduttori. Tra i vantaggi che questa tecnica offre ci sono sicuramente il basso costo e l'assenza di contatto diretto con l'oggetto da misurare, così come la possibilità di avere un numero pressoché infinito di punti di controllo. La misura della deformazione può avvenire in un qualunque punto all'intradosso della trave e non soltanto in corrispondenza dei trasduttori. Inoltre la preparazione e l'allestimento della prova ha tempistiche notevolmente ridotte in quanto il posizionamento della camera e l'acquisizione del dato possono avvenire durante la preparazione del provino e la prova stessa. Questi sono soltanto i primi risultati ottenuti, in un'ottica di maggiore efficienza, soprattutto nell'elaborazione del dato si potrebbe pensare di automatizzare il calcolo degli spostamenti attraverso l'uso di appositi target riconoscibili da particolari algoritmi.

In definitiva si può affermare che con una strumentazione come quella utilizzata nella prova è possibile eseguire le misurazioni necessarie in affiancamento o in sostituzione alle tradizionali misure eseguite per mezzo di trasduttori di spostamento.

BIBLIOGRAFIA

- BOLOGNESI M.,FURINI A.,RUSSO V.,PELLEGRINELLI A.,RUSSO P., (2014) *Accuracy of cultural Heritage 3D models by RPAS and terrestrial photogrammetry*. Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci., XL-5, 113-119.
- BOLOGNESI M.,FURINI A.,RUSSO V.,PELLEGRINELLI A.,RUSSO P., (2015). *Testing the low-cost RPAS potential in 3D Cultural Heritage reconstruction*. Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci., XL-5/W4, 229-235.
- BISCUOLA L. (2014). Operazioni topografiche per la predisposizione di un “test-site” destinato alla sperimentazione di tecniche di rilevamento 3D dei Beni Culturali. Tesi di Laurea triennale in Ingegneria Civile e Ambientale. Università degli Studi di Ferrara.
- FAUGERAS O. (1993). *Three-Dimensional Computer Vision: A Geometric Viewpoint*. The MIT Press, Cambridge, MA.
- FERRARI F., (2014). *Integrazione di tecniche fotogrammetriche per il rilevamento e la modellazione 3D dei Beni Culturali: il caso di studio della “Delizia del Verginese”*. Tesi di Laurea Magistrale in Ingegneria Civile, Università degli studi di Ferrara.
- FIORE P.D. (2001). *Efficient linear solution of exterior orientation*. IEEE Transactions on Pattern Analysis and machine intelligence, 23(2), 140-148.
- FORSYTH D., PONCE J. (2012). *Computer Vision: A modern approach*. Pearson, Boston.
- FUSIELLO A. (2013). *Visione computazionale, tecniche di ricostruzione tridimensionale*. Franco Angeli editore, Milano, Italia.
- HARRIS C., STEPHENS M. (1988). *A combined corner and edge detector*. Plessey Research Roke Manor, United Kingdom.
- HUANG T., FAUGERAS O. (1989). *Some properties of the E matrix in two-view motion estimation*. IEEE Transactions on Pattern Analysis and machine intelligence, 11(12), 1310-1312.
- LONGUET-HIGGINS H. C. (1981). *A computer algorithm for reconstructing a scene from two projections*. Nature, 293(10), 133-135.

LOWE D. G. (2004). *Distinctive image features from scale-invariant keypoints*. International journal of Computer Vision, 60(2), 91-110.

SHI J., TOMASI C. (1994). *Good features to track*. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 593-600.

SZELISKI R. (2011). *Computer vision: algorithms and applications*. Springer, New York.

TADDIA Y., (2014). *La fotogrammetria digitale per il rilievo delle deformazioni di elementi strutturali sottoposti a prove di carico*. Tesi di Laurea Magistrale in Ingegneria Civile, Università degli studi di Ferrara.

ZATTONI R., (2014). *Studio sperimentale della precisione del rilievo fotogrammetrico dei Beni Culturali mediante l'impiego dei Sistemi Aeromobili a Pilotaggio Remoto*. Tesi di Laurea Magistrale in Ingegneria Civile, Università degli studi di Ferrara.

Siti Consultati

<http://opencv-python-tutroals.readthedocs.org/>