

POLITECNICO DI MILANO  
Facoltà di Ingegneria dell'Informazione  
*Corso di Laurea Specialistica in  
Ingegneria dell'Automazione*



CALIBRAZIONE DI TELECAMERE PER SISTEMI DI  
VISIONE STEREOSCOPICA: CONFRONTO TRA  
ALGORITMI GENETICI E TECNICHE TRADIZIONALI

Relatore: Prof. Alfredo CIGADA

Correlatori: Ing. Alessandro BASSO

Ing. Mario GALIMBERTI

Tesi di Laurea Specialistica di:

Giacomo MAINETTI Matr. 735692

Anno Accademico 2010-2011

*Ai Miei Genitori*



# Ringraziamenti

Ringrazio il prof. Alfredo Cigada e l'ing. Remo Sala che mi hanno dato la preziosa possibilità di svolgere la tesi presso il VBLab, dove ho avuto l'occasione di maturare soprattutto a livello professionale. Grazie a Alessandro Basso e Mario Galimberti, che mi hanno aiutato, supportato e sopportato in questi mesi di lavoro e mi hanno trasmesso numerosissime conoscenze nel campo della visione artificiale e non solo. Ringrazio anche il resto dei colleghi con cui mi sono confrontato durante l'intero trascorso presso il VBLab: Giovanni Bassi, Alessandro Perego, Davide Malchiodi, Stefano Manzoni e il mitico Alfredo Valtolina, per i loro consigli e per aver reso le giornate di lavoro in laboratorio più piacevoli.

Il più sentito ringraziamento è rivolto ai miei genitori, ai quali dedico questo lavoro. È soprattutto grazie a loro che ho avuto la possibilità di percorrere il cammino che mi ha portato a questo traguardo importantissimo. Grazie per il sostegno, l'affetto e il calore che mi avete sempre trasmesso.

Inoltre ringrazio le persone che più mi sono state vicine in tutti i momenti importanti della vita: i nonni, zii, cugini, mio fratello Teo e soprattutto Ludovica.

Un ultimo ringraziamento va agli amici di oggi e di ieri con cui ho condiviso momenti di svago indimenticabili.

# Abstract

In numerosi ambiti applicativi di ricerca e industriali esiste la necessità di effettuare misure di posizione e spostamento di punti caratteristici del misurando, ponendo particolare attenzione alla minimizzazione degli effetti di carico. Tale problematica è stata affrontata sviluppando un sistema di misura ottico stereoscopico che consente di effettuare misure statiche e dinamiche delle coordinate tridimensionali di alcuni punti notevoli del misurando. Nella definizione delle specifiche è stata data particolare attenzione alla necessità della Galleria del Vento del Politecnico di Milano di effettuare misure dinamiche senza contatto, in modo da non disturbare la vena fluida del vento. Le specifiche sono quindi dettate da questa particolare applicazione, anche se in realtà si è voluto realizzare un sistema di misura che potesse essere utilizzato in un'ampia varietà di ambiti applicativi, con diverse condizioni e volumi di lavoro e diverse situazioni d'illuminazione. Al fine di minimizzare l'incertezza di misura sono state affrontate in dettaglio le problematiche legate alla taratura dello strumento. Essa si compone di due parti fondamentali: la calibrazione delle telecamere e la loro registrazione. Una parte rilevante di questo lavoro è stata dedicata all'analisi e confronto di alcuni metodi basati sugli algoritmi genetici e di una tecnica contenuta nella libreria commerciale per la visione artificiale HALCON, sviluppata da MVTec, con lo scopo di verificare se essi potessero rappresentare un approccio migliore al problema della calibrazione rispetto al metodo di Zhang, ottimizzato con una particolare formulazione dell'algoritmo di Levenberg-Marquardt.

Un'adeguata calibrazione delle telecamere non rappresenta la sola condizione per il corretto funzionamento del sistema di misura. Un altro aspetto rilevante è costituito dalla procedura di registrazione di più moduli d'acquisizione. Infatti, in alcune applicazioni, in cui ad esempio il misurando è grande o presenta elevate curvature, può essere necessario l'utilizzo di più strumenti. È importante che la registrazione di più moduli di acquisizione non introduca incertezze sensibili nelle misure, diminuendo l'accuratezza dell'intero sistema. È stato quindi sviluppato un

algoritmo di registrazione basato sulla decomposizione ai valori singolari (SVD), che è stato validato da una serie di prove atte a quantificare l'incertezza introdotta da tale tecnica.

Nella seconda parte della tesi invece si è sviluppato lo strumento di misura stereoscopico. Esso è gestito mediante un pacchetto software interamente implementato in questo lavoro di tesi. L'intero sistema è stato sottoposto a verifica metrologica ed è stato utilizzato per una campagna di prove sperimentali presso la Galleria del Vento del Politecnico di Milano, in cui è stato studiato il comportamento dinamico del modello di una struttura studiata per aumentare la stabilità di uno dei palazzi della nuova area CityLife di Milano, soggetto a condizioni di vento controllate.

# Indice

<b>1. Introduzione</b>	<b>1</b>
<b>2. Calibrazione di telecamere</b>	<b>6</b>
2.1 Introduzione	7
2.1.1 Stato dell'arte delle metodologie di calibrazione	8
2.1.2 Il modello della telecamera	11
2.2 Calibrazione mediante il metodo di Zhang e di Levenberg-Marquardt	22
2.2.1 Acquisizione dei dati di calibrazione	23
2.2.2 Calcolo del modello lineare	26
2.2.3 Ottimizzazione non lineare	30
2.2.4 Implementazione	37
2.3 Algoritmi genetici (GA)	39
2.3.1 Introduzione ai GA	40
2.3.2 Principi base dei GA	42
2.3.3 Implementazione di tecniche di calibrazione basate sugli algoritmi genetici	45
2.4 Confronto tra differenti metodologie	50
2.4.1 Zhang-LM	51
2.4.2 Tecnica di Yang	53
2.4.3 Tecniche di bundle adjustment	60
2.4.4 Tecniche basate sull'errore di triangolazione	66
2.4.5 Halcon	73
2.5 Conclusioni	81
<b>3. Metodologia di misura</b>	<b>84</b>
3.1 Tecniche di misura tridimensionali	84
3.2 Stereoscopia	89
3.3 Geometria epipolare	91

---

<b>4. Sistema di misura</b>	<b>98</b>
4.1 Setup del sistema	99
4.1.1 Parametri delle telecamere	99
4.1.2 Scelta delle telecamere	104
4.1.3 Sistema d'illuminazione	106
4.1.4 Altri componenti hardware	108
4.2 Descrizione del software	110
4.2.1 Acquisizione	111
4.2.2 Acquisizione immagini termiche	113
4.3 Taratura dello strumento	119
4.4 Ricostruzione stereoscopica	121
4.4.1 Implementazione	121
4.5 Registrazione di diversi moduli di acquisizione	124
4.5.1 Registrazione di moduli di acquisizione mediante decomposizione ai valori singolari	125
4.5.2 Implementazione	126
<b>5. Attività sperimentale</b>	<b>128</b>
5.1 Verifiche metrologiche	128
5.1.1 Ricostruzione del calibratore	129
5.1.2 Verifica della procedura di registrazione	132
5.1.3 Verifica dell'accuratezza di registrazione al variare del numero di punti utilizzati	137
5.2 Prove sperimentali in Galleria del Vento	142
5.2.1 Ricostruzione del calibratore a marker catarifrangenti	152
5.2.2 Prova con il modello fermo	155
5.2.3 Verifica della rigidità del modello nel tempo	157
5.2.4 Verifica con calibratore verticale per il calcolo dell'incertezza su z	158
5.2.5 Confronto tra dati di Galleria e misure del sistema stereoscopico	161

<b>6. Conclusioni</b>	<b>165</b>
6.1 Sviluppi futuri	168
6.1.1 Calibrazione di telecamere	168
6.1.2 Sistema di misura	169
<b>Bibliografia</b>	<b>171</b>

# 1. Introduzione

In numerosi ambiti applicativi di ricerca e industriali esiste la necessità di effettuare misure di posizione e spostamento di punti caratteristici del misurando, ponendo particolare attenzione alla minimizzazione degli effetti di carico. Tali obiettivi possono essere ottenuti mediante l'utilizzo di metodologie di misura 3D senza contatto. La scelta della tecnica e della tecnologia più idonee dipendono dalle applicazioni in cui le si intende utilizzare e dalle specifiche richieste. Questo particolare lavoro nasce dalle necessità della Galleria del Vento del Politecnico di Milano di effettuare misure dinamiche senza contatto, in modo da non disturbare la vena fluida del vento. Le specifiche sono quindi dettate da questa particolare applicazione, anche se in realtà si è voluto realizzare un sistema di misura che potesse essere utilizzato in un'ampia varietà di ambiti applicativi, con diverse condizioni e volumi di lavoro e diverse situazioni d'illuminazione. Sono quindi richieste caratteristiche di flessibilità e scalabilità del sistema di misura. Come sarà illustrato nella tesi, le particolari specifiche considerate in termini di volume di lavoro e incertezza di misura hanno delimitato la scelta a tecniche di misura basate sulla visione artificiale. In particolare questa è ricaduta sulla stereoscopia. Inoltre è utile che lo strumento possa elaborare i dati in linea con l'acquisizione, cioè possa ricavare le misure durante l'acquisizione stessa e non a monte di essa. In questo modo è possibile verificare immediatamente la bontà dell'acquisizione

e, soprattutto, velocizzare le procedure di elaborazione delle misure, poiché i dati acquisiti possono così essere analizzati in tempo mascherato durante la campagna sperimentale e non al suo termine. Inoltre, è possibile diminuire fortemente la richiesta di memoria fisica per il salvataggio dei dati, poiché non vengono in tal modo memorizzate le immagini grezze, che contengono informazioni sovrabbondanti, ma solamente le misure di interesse.

In questo lavoro è stato sviluppato un sistema di misura ottico stereoscopico che consente di effettuare misure statiche e dinamiche delle coordinate tridimensionali di alcuni punti notevoli nella scena, con frequenze di acquisizione fino a 17 Hz. Per effettuare misure stereoscopiche sono necessarie almeno due telecamere, tuttavia in questa tesi si è preferito utilizzare tre telecamere per irrobustire la procedura di accoppiamento automatico dei punti omologhi delle diverse immagini. Lo strumento composto da una tripletta di telecamere costituisce un singolo modulo di acquisizione. Al fine di minimizzare l'incertezza di misura sono state affrontate in dettaglio le problematiche legate alla procedura di taratura dello strumento. Essa si compone di due parti fondamentali: la calibrazione delle telecamere e la loro registrazione. La prima è la vera e propria taratura delle singole telecamere, vale a dire il processo che consente di stimare i parametri intrinseci ed estrinseci del modello del dispositivo. La registrazione, invece, consiste nel calcolare la posa dei dispositivi rispetto a un sistema di riferimento comune. Essendo la calibrazione delle telecamere la parte più critica dell'intera procedura di misura, ad essa è stata dedicata un'attenzione particolare. In letteratura esistono numerosi metodi di calibrazioni e non esiste una norma che stabilisca una procedura standard. Per questo motivo una parte rilevante di questo lavoro è stata dedicata al confronto tra alcune tecniche con lo scopo di individuare la più adatta al caso in esame. In particolare si è deciso di confrontare la tecnica di Zhang [2], ottimizzata con una particolare formulazione dell'algoritmo di Levenberg-Marquardt ottimizzato per il problema della calibrazione [1],[3],[4],[5], con una serie di metodi basati sugli algoritmi genetici e con una tecnica contenuta nella libreria commerciale per la visione artificiale HALCON [29], sviluppata da MVTEC. La prima è una tecnica già ampiamente utilizzata e

validata, anche presso il laboratorio VBLab del Dipartimento di Meccanica del Politecnico di Milano [7], in cui è stato svolto questo lavoro. Per quanto riguarda gli algoritmi genetici, partendo da tecniche presenti in letteratura, sono stati implementati altri metodi, tentando di migliorare gli aspetti più critici dell'utilizzo di questi algoritmi. I loro risultati sono stati confrontati con quelli ottenuti mediante la tecnica di Zhang e Levenberg-Marquardt e con la tecnica di HALCON.

Un'adeguata calibrazione delle telecamere non rappresenta la sola condizione per il corretto funzionamento del sistema di misura. Un altro aspetto rilevante è costituito dalla procedura di registrazione di più moduli d'acquisizione. Infatti, in alcune applicazioni, in cui ad esempio il misurando è grande o presenta elevate curvature, può essere necessario l'utilizzo di più strumenti, e dunque più triplette di telecamere. Ogni modulo esprime le misure nel sistema di riferimento della specifica taratura, ma è necessario che essi riferiscano le misure a un sistema di riferimento comune e siano quindi registrati. È indispensabile che la procedura di registrazione non introduca incertezze sensibili nelle misure, diminuendo l'accuratezza dell'intero sistema. È stato dunque sviluppato un algoritmo basato sulla decomposizione ai valori singolari (SVD), che consente di trovare la posa relativa di due strumenti confrontando le coordinate tridimensionali di una serie di punti, riferite ai sistemi di riferimento delle due triplette. Il metodo è stato validato da una serie di prove atte a quantificare l'incertezza introdotta da tale tecnica e verificare la dipendenza di essa dal numero di punti considerato.

Per quanto concerne lo sviluppo del sistema di misura, particolare attenzione è stata rivolta alla scalabilità e alla flessibilità dello strumento. Esso deve infatti poter essere applicato in diversi ambiti e avere un campo di misura variabile a seconda delle necessità. Per questo motivo le telecamere sono dotate di ottiche a lunghezza focale variabile, che consentono di avere un campo di misura più o meno ampio a seconda delle necessità. Inoltre il posizionamento relativo dei dispositivi può essere regolato a piacere prima di tarare e iniziare le misurazioni, così da aumentare la flessibilità del sistema. Nelle applicazioni in cui è richiesta un'accuratezza di misura elevata è possibile utilizzare lo strumento con condizioni

di luce controllata, grazie a un sistema di illuminatori all'infrarosso coassiali ad ogni telecamera. Il loro utilizzo è subordinato alla possibilità di montare sul misurando alcuni punti noti facilmente distinguibili dal resto della scena, detti marker. Mediante filtri interferenziali centrati sulla lunghezza d'onda dell'illuminatore, è possibile limitare il rumore, limitando sensibilmente le lunghezze d'onda indesiderate.

Lo strumento di misura è gestito mediante un pacchetto software interamente sviluppato in questo lavoro di tesi. Il software è molto flessibile, in quanto può essere utilizzato con numerosi modelli di telecamere commerciali e con diverse tipologie di marker. Esso, soddisfacendo le specifiche prima menzionate, consente di acquisire immagini in diverse modalità, tarare lo strumento, registrare più strumenti e calcolare le coordinate tridimensionali di punti noti in maniera dinamica. Una delle caratteristiche principali è la capacità, grazie a una programmazione multithreading, di effettuare misure in linea con la acquisizione. In questo modo è possibile diminuire notevolmente i tempi necessari per il calcolo delle coordinate tridimensionali a valle della campagna sperimentale e la quantità di dati immagazzinati. La ricostruzione dell'oggetto è, inoltre, visualizzata a schermo, così da poter avere una prima verifica della bontà delle misure che si stanno effettuando. Il software sviluppato consente anche a utenti meno esperti di poter utilizzare lo strumento di misura correttamente.

L'intero sistema è stato sottoposto a verifica metrologica mediante una serie di prove su un calibratore planare. Esso è stato infine utilizzato per una campagna di prove sperimentali presso la Galleria del Vento, in cui è stato studiato il comportamento dinamico del modello di una struttura studiata per aumentare la stabilità di uno dei palazzi della nuova area CityLife di Milano, soggetto a condizioni di vento controllate. Lo strumento sviluppato in questa tesi risulta particolarmente utile nelle prove in Galleria del Vento, dove è spesso necessario misurare lo spostamento di modelli soggetti alla forza del vento, senza però poter utilizzare strumenti a contatto con esso, poiché introdurrebbero effetti di carico e disturberebbero la vena fluida del vento.

La tesi è così strutturata:

- il capitolo 2 è interamente dedicato alla procedura di calibrazione. Viene inizialmente riportato lo stato dell'arte delle tecniche di calibrazione e il modello di telecamera utilizzato in questo lavoro. Viene poi descritto nei particolari il metodo di Zhang-LM e, dopo un'introduzione agli algoritmi genetici, vengono descritte alcune tecniche di calibrazione basate su questi. Ognuna di queste è messa a confronto con l'algoritmo di Zhang-LM. Viene poi aggiunta al confronto la tecnica di HALCON e vengono infine riassunte tutte le considerazioni fatte in questo capitolo.
- Nel capitolo 3 viene riportato lo stato dell'arte delle tecniche di misura tridimensionale e viene descritta nei particolari la stereoscopia. Infine sono riportati alcuni cenni teorici sulla geometria epolare che sta alla base della ricostruzione tridimensionale stereoscopica.
- Nel capitolo 4 viene descritto l'intero sistema di misura sviluppato in questo lavoro. In particolare viene descritto il setup del sistema, vale a dire l'hardware che lo compone, giustificando le decisioni prese in merito alla scelta dei componenti del sistema. Viene poi descritto in dettaglio il pacchetto software che è stato implementato e che gestisce l'intero sistema di misura. Viene mostrato come il software si interfaccia ai dispositivi e come viene eseguita l'acquisizione di immagini. Si descrivono gli algoritmi utilizzati per la ricostruzione tridimensionale e la procedura di taratura di un sistema stereoscopico, che si compone della calibrazione dei singoli dispositivi e la registrazione delle telecamere. Infine viene mostrato l'algoritmo che permette di registrare più triplette di telecamere tra di loro.
- Nel capitolo 5 sono riportati gli esperimenti effettuati in laboratorio per validare metrologicamente lo strumento. Infine si descrive la campagna sperimentale svolta presso la Galleria del Vento e i risultati ottenuti da essa.
- Nel capitolo 6 invece sono riassunte le principali considerazioni che sono state effettuate durante l'intera tesi.

## 2. Calibrazione di telecamere

*Come qualsiasi strumento di misura, anche il sistema stereoscopico sviluppato in questo lavoro necessita di essere tarato prima di eseguire qualsiasi misura. La taratura di uno strumento ottico stereoscopico consiste nella determinazione dei parametri costituenti il modello matematico per ogni telecamera e nella determinazione delle loro pose rispetto a un sistema di riferimento comune. La calibrazione delle telecamere consiste nella stima dei parametri che riguardano direttamente il dispositivo (intrinseci) e di quelli che ne descrivono la posa (estrinseci). È quindi possibile calibrare le singole telecamere che compongono la tripletta e poi esprimerne le pose in un sistema di riferimento comune a tutti i dispositivi. La calibrazione delle singole telecamere risulta dunque essere la procedura più critica dell'intero processo di taratura del sistema, perciò si vuole dedicare un capitolo solamente ad essa. In letteratura non esiste un metodo standard da utilizzare per calibrare. La tecnica più utilizzata è l'algoritmo di Zhang. Si è voluto però provare ad implementare alcune tecniche basate sugli algoritmi genetici, partendo prima da metodi già presenti in letteratura e implementandone poi di nuovi. I risultati restituiti da tali tecniche sono poi stati confrontati con quelli trovati con il metodo di Zhang. In questo capitolo verrà inizialmente riassunto lo stato dell'arte delle varie tecniche di calibrazione e*

*verrà descritto nei particolari il modello di telecamera utilizzato in questo lavoro. In seguito si approfondirà la procedura di calibrazione secondo il metodo di Zhang, ottimizzato mediante l'algoritmo di Levenberg-Marquardt. Dopo una breve introduzione agli algoritmi genetici verranno descritte le tecniche di calibrazione basate su tali algoritmi e i risultati verranno confrontati con quelli trovati con l'algoritmo di Zhang. Al confronto poi verrà aggiunta una tecnica di calibrazione contenuta in librerie commerciali per la visione artificiale.*

## **2.1 Introduzione**

Con il termine *calibrazione* di telecamere s'intende il processo di determinazione dei parametri che definiscono il modello della telecamera, descritto in dettaglio nel capitolo 2.1.2. La calibrazione è un processo necessario per ricavare informazioni metriche dall'immagine e potere così utilizzare una telecamera come strumento di misura. A oggi non esiste una norma che impone una particolare metodologia di calibrazione. In base all'applicazione in cui le telecamere sono utilizzate, si può ricorrere a una determinata tecnica piuttosto che un'altra. I principali metodi presenti ad oggi sono descritti nel capitolo 2.1.1.

Data l'importanza di questa procedura per il poter ottenere misurazioni accurate, è stato necessario dedicare una consistente parte della tesi all'analisi e al confronto di diverse procedure di calibrazione con lo scopo di trovare la miglior soluzione da poter utilizzare. Prima di tutto si descriverà nel dettaglio un metodo di calibrazione già ampiamente validato come quello di Zhang e si confronterà con una diversa tecnica basata sugli algoritmi genetici. Inoltre verrà considerato nel confronto anche un tool di calibrazione commerciale (Halcon). Questo lavoro di analisi e confronto ci consentirà di migliorare e irrobustire la procedura, di individuarne i pregi e i difetti e di capire quali sono gli errori che si possono fare durante la calibrazione e limitarli al minimo.

## 2.1.1 Stato dell'arte delle metodologie di calibrazione

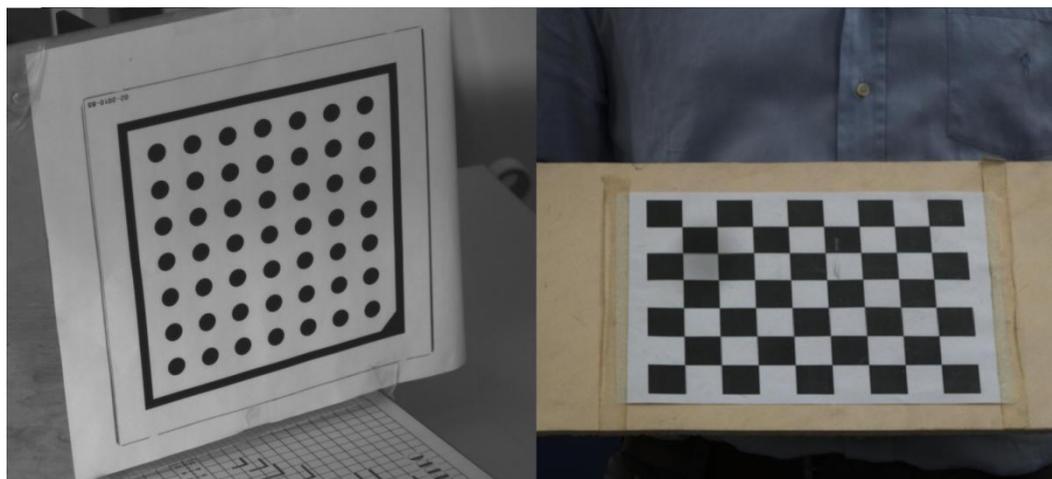
Ad oggi le tecniche di calibrazione presenti si possono suddividere in due principali categorie: i metodi basati sull'uso di un calibratore e le tecniche di auto calibrazione (*self-calibration*). I primi consistono nell'utilizzo di un calibratore, cioè un oggetto di cui si conosce con elevata accuratezza la disposizione nello spazio di alcuni punti noti, come riferimento nello spazio. Attraverso le corrispondenze che vengono individuate tra i punti nella scena e la loro proiezione nell'immagine, vengono determinati i parametri del modello della telecamera. Questi metodi consentono di raggiungere ottime prestazioni in termini di accuratezza [8],[9],[10]. I più famosi metodi tradizionali sono: il “*Direct Line Transformation*” (DLT) method, l'*ottimizzazione non lineare*, il *Tsai two-step method* e il *metodo di Zhang*.

Il DLT fu proposto per la prima volta da Abdel-Aziz e Karara nel 1971 [1]. Esso si basa sul calcolo dei parametri della matrice di proiezione, vale a dire la matrice che descrive la corrispondenza tra i punti nella scena e nell'immagine. Il grande vantaggio di questo metodo consiste nel fatto che i parametri intrinseci ed estrinseci sono calcolati risolvendo equazioni lineari senza l'utilizzo di algoritmi iterativi. Nonostante la sua velocità computazionale e la semplicità derivante dall'utilizzo di un numero ristretto di parametri, esso non è ad oggi una delle tecniche più utilizzate, poiché non prende in considerazione le distorsioni non lineari e quindi non riesce a raggiungere accuratezze elevate. Esso è stato poi perfezionato per includere i fattori non lineari, aumentando la complessità del problema. Queste tecniche di calibrazione non lineare hanno un'elevata accuratezza, ma sono computazionalmente lente e complesse; inoltre richiedono nella maggior parte dei casi una buona stima dei parametri iniziali.

Nel 1987 Tsai [8] sviluppò un metodo chiamato *two-step*, che consiste in due passaggi principali: prima di tutto viene risolto il problema lineare basandosi sui vincoli che legano i parametri della matrice di proiezione, poi si procede con un'ottimizzazione non lineare, inizializzata con i risultati trovati al primo punto,

che tiene conto delle distorsioni. Questo metodo può raggiungere elevate accuratezze pur operando in uno spazio di parametri meno esteso rispetto al primo [9]. Il metodo two-step proposto da Tsai considera solamente le distorsioni al secondo ordine, cioè tiene conto solo delle distorsioni radiali senza considerare quelle tangenziali. In generale è possibile raggiungere buoni risultati di calibrazione anche trascurando le distorsioni tangenziali, ma per una maggiore accuratezza è conveniente considerare il modello sviluppato da Brown, che considera entrambe le tipologie di non linearità [11]. Il modello non lineare di Brown è stato considerato durante le prove in questa tesi ed è descritto in seguito nel capitolo 2.1.2.

Nel 1998 Zhengyou Zhang propose un nuovo metodo [2], che richiede solamente l'osservazione di un *pattern planare* (figura 2.1) in diverse pose casuali nello spazio (almeno tre). Si può decidere di muovere il calibratore piuttosto che la telecamera, senza alterare i risultati di calibrazione e soprattutto senza il bisogno di misurare la posizione nello spazio del pattern rispetto alla telecamera a priori. Accoppiando i punti del calibratore nelle immagini con un modello di esso, è possibile calcolare la matrice di omografia e da questa ricavare i parametri della telecamera. L'algoritmo di Zhang per la stima dei parametri lineari può essere raffinato implementando un problema di minimizzazione non lineare, che quindi tiene conto delle distorsioni della lente. Il vantaggio di questa tecnica consiste il *setup* semplice ed economico avendo comunque risultati molto accurati in fase di calibrazione. Questo metodo è stato preso in considerazione in questo lavoro di tesi e sarà quindi spiegato più in dettaglio nel capitolo 2.2.



**Figura 2.1:** Esempi di pattern planari utilizzati durante calibrazioni col metodo di Zhang.

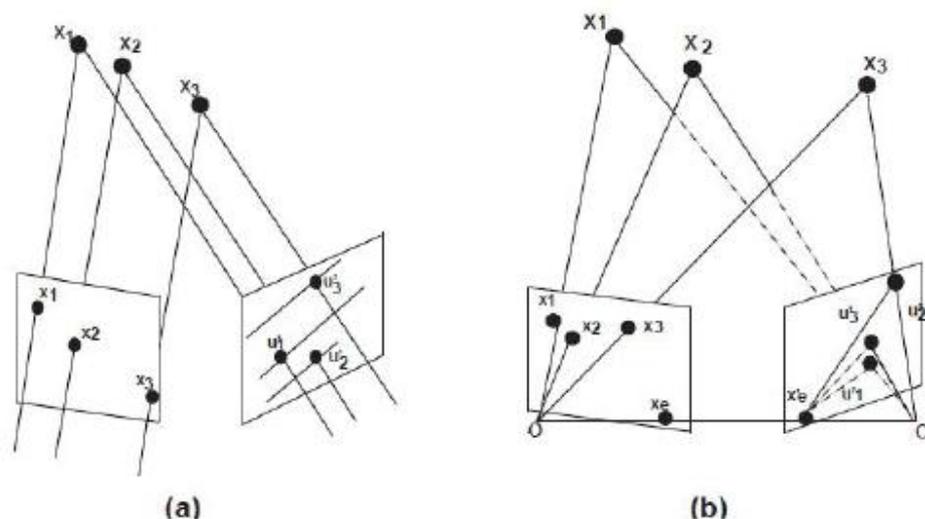
In alcune applicazioni non è possibile calibrare la telecamera utilizzando un calibratore e così sono state sviluppate negli ultimi anni numerose tecniche di *self-calibration*. Esse infatti non richiedono l'utilizzo di un calibratore particolare, ma, attraverso la movimentazione della telecamera in una scena statica si riescono a ricavare i vincoli che legano i parametri di calibrazione. Questo metodo è adatto a quelle applicazioni in cui i parametri della telecamera non possono essere mantenuti fissi ed è necessario dunque ricalibrare spesso e velocemente il sistema. A differenza dei metodi tradizionali questa tecnica è molto flessibile, tuttavia è anche più complessa e non permette di raggiungere accuratezze altrettanto elevate. Tra queste tecniche riscontrano grande interesse quelle basate sulla *visione attiva*, vale a dire le tecniche in cui i parametri di calibrazione vengono determinati attraverso il controllo attivo dei movimenti della telecamera. Esse hanno il vantaggio di basarsi su algoritmi semplici ma lo svantaggio di richiedere un sistema talvolta complesso di controllo.

Il costante sviluppo di alcune metodologie per l'ottimizzazione non lineare, quali reti neurali e algoritmi genetici (GA), ha permesso lo sviluppo di alcuni metodi di calibrazione in cui il problema non lineare viene risolto attraverso l'utilizzo di queste tecniche. Gli algoritmi ad apprendimento sono però ancora troppo lenti e spesso convergono a un minimo locale portando errori considerevoli in fase di calibrazione. Per questo si è deciso di utilizzare soltanto gli algoritmi

genetici e analizzare alcune tecniche basate su essi, anche se, come si vedrà nel capitolo 2.4, non hanno portato a risultati soddisfacenti. Un'innovativa tecnica che permette di combinare la capacità di fitting non lineare delle reti neurali con la buona capacità di ricerca globale degli algoritmi genetici di calibrazione è descritta in [18]. L'analisi delle reti neurali, e in particolare di quest'ultimo metodo, potrebbe rappresentare uno dei possibili sviluppi del lavoro svolto in questa tesi.

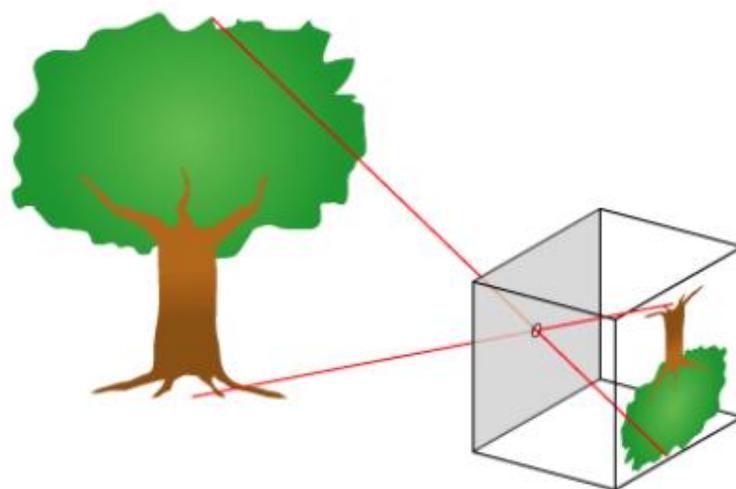
## 2.1.2 Il modello della telecamera

La telecamera è uno strumento in grado di acquisire immagini dal mondo circostante, immagazzinando le informazioni ottenute dai fotoni riflessi dagli oggetti che compongono la scena, in un determinato range di frequenze. Questo avviene tramite una conversione degli stessi fotoni in un'unità di misura elaborabile digitalmente, solitamente elettronici. Tale strumento effettua quindi una mappatura tra il mondo tridimensionale e un'immagine bidimensionale. Esistono diversi modelli geometrici [1] che descrivono tale processo e ognuno di questi caratterizza un diverso tipo di telecamera. Questi modelli possono essere distinti in base ai parametri che li compongono; ad esempio è possibile trovare modelli che descrivono telecamere con centro in posizione finita, e telecamere con centro "all'infinito", generalizzazione naturale di una proiezione di linee parallele ("*telecamera affine*"). Quest'ultimo modello viene in genere utilizzato quando il sistema non è riconducibile a un sistema a proiezione centrale, ovvero per ottiche di lunghezza focale molto elevata e non di interesse per questo lavoro. La differenza tra questi due modelli è ben visibile in figura 2.2: in (a) i punti vengono proiettati parallelamente tra loro nell'immagine, in (b) invece esiste un centro di proiezione nel quale i punti sono proiettati.

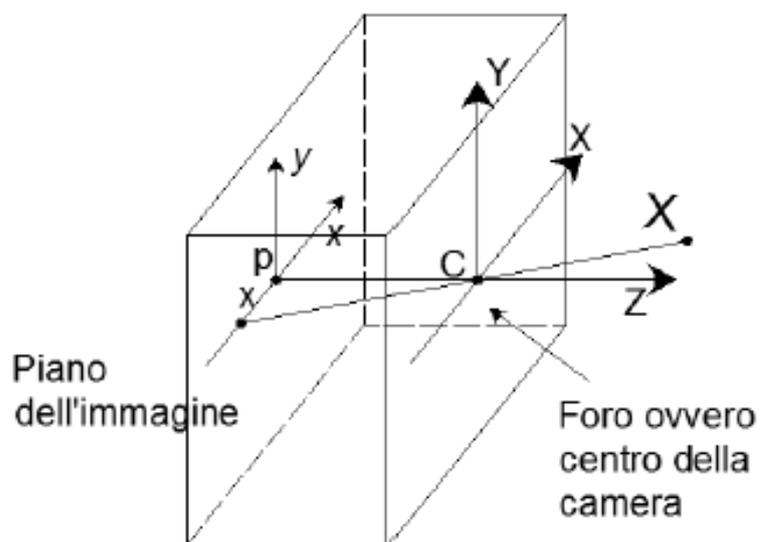


**Figura 2.2:** Differenza tra modello di telecamera affine (a) e telecamera proiettiva (b)

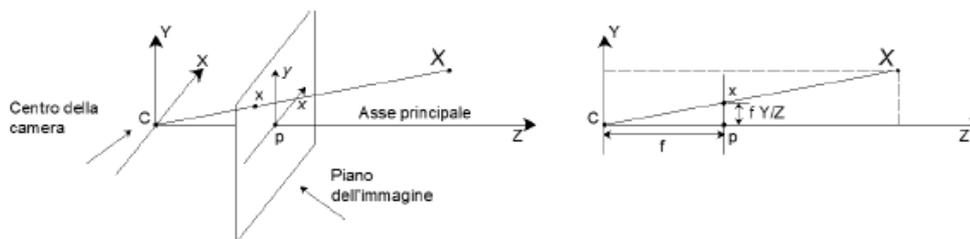
In questo lavoro viene utilizzato un modello di telecamera proiettiva, poiché le lunghezze focali qui utilizzate non richiedono l'utilizzo del modello di telecamera affine. In particolare verrà sempre fatto riferimento al modello più comunemente utilizzato, ovvero il *pinhole camera* [1],[6] (figura 2.3). Tale modello descrive la telecamera come un sistema a proiezione centrale, ovvero come una telecamera senza lenti con una singola apertura molto piccola. Si tratta di una scatola chiusa con un piccolo foro attraverso il quale passa la luce proveniente dalla scena e viene quindi proiettata un'immagine capovolta sull'altro lato della scatola. Tale rappresentazione sfrutta il concetto di *foro stenopeico* che descrive in maniera accurata l'effettiva formazione delle immagini attraverso una lente. Il foro stenopeico è usato per rappresentare quello che viene chiamato obiettivo della telecamera. Le telecamere che utilizzano tale struttura producono immagini tanto più nitide quanto più il diametro del foro è piccolo. Minori dimensioni del foro comportano però la diminuzione della luce incidente sul sensore posto dall'altro lato, aumentando i tempi di esposizione necessari. In teoria la dimensione del diaframma dovrebbe essere 1/100 della distanza tra essa e il sensore.



**Figura 2.3:** Principio base del modello pinhole camera.



**Figura 2.4:** Geometria del modello pinhole camera.



**Figura 2.5:** Geometria del modello pinhole camera frontale, vale a dire con centro di proiezione posto dietro al sensore.

La geometria che descrive tale modello è mostrata in figura 2.4. Il centro di proiezione del sistema è chiamato *centro della telecamera* o *centro ottico*. Una visione alternativa consiste nel mettere il centro della telecamera dietro il piano immagine, considerando quindi i punti immagine come se fossero già stati invertiti; così facendo si ottiene una formulazione geometrica più facile e comprensibile (figura 2.5). Tale modello viene chiamato *pinhole frontale* ed è quello che sarà preso in considerazione durante tutto il lavoro svolto. La linea dal centro della telecamera perpendicolare al piano immagine viene chiamata *asse principale* o *raggio principale* della telecamera e il punto in cui interseca il piano immagine viene chiamato punto principale. Infine, il piano su cui giace il centro telecamera, parallelo al piano dell'immagine è chiamato *piano principale*. È intuitivo come tale modello descriva il modo in cui il punto mondo  $X$  viene mappato in un punto sul piano immagine  $x$ . Esprimendo tali punti nelle tre coordinate spaziali che li compongono è possibile indicarli con dei vettori colonna:  $(X, Y, Z)^T$  esprimeranno le coordinate nel mondo, mentre  $(f \frac{X}{Z}, f \frac{Y}{Z})^T$  esprimeranno i punti nell'immagine. Ignorando l'ultima coordinata immagine si ottiene una mappatura da uno spazio euclideo in  $\mathbb{R}^3$  in uno spazio euclideo in  $\mathbb{R}^2$ :

$$(X, Y, Z)^T \rightarrow \left( f \frac{X}{Z}, f \frac{Y}{Z} \right)^T \quad (2.1)$$

Se il mondo e i punti immagine sono espressi mediante vettori omogenei, la proiezione centrale usata dal modello può essere facilmente espressa come mappatura lineare tra le loro coordinate omogenee; in particolare la 2.1 può essere

espressa come:

$$\begin{pmatrix} fX \\ fY \\ Z \end{pmatrix} = \begin{bmatrix} f & & 0 \\ & f & 0 \\ & & 1 & 0 \end{bmatrix} \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix} \quad (2.2)$$

La matrice  $3 \times 4$  che si ottiene è chiamata *matrice di proiezione* o matrice della telecamera (indicata con la lettera  $P$ ). Usando una notazione più compatta:

$$x = PX \quad (2.3)$$

Per esigenze di chiarezza è opportuno definire i sistemi di riferimento che vengono comunemente utilizzati nel modello della telecamera. Il primo è il SDR globale dell'intero sistema, che è detto SDR mondo ed è nelle nostre applicazioni spesso solidale con la prima telecamera o, in calibrazione, coincidente con la posizione del calibratore. Un successivo SDR è quello che individua il centro geometrico della telecamera, che viene utilizzato per individuare la posizione delle telecamere rispetto al SDR mondo. Come si osserverà in seguito, il centro geometrico del sensore non sempre coincide con il punto principale e la posizione di tale punto non è comunemente utilizzata per indicare le coordinate dei pixel; infatti, per questioni di comodità, essi sono espressi in riferimento all'angolo in alto a sinistra del sensore, in modo da avere sempre coordinate positive. Questo si traduce in un ulteriore SDR di telecamera che permette la definizione dei punti nell'immagine direttamente in coordinate pixel.

Il modello fin qui descritto però è molto semplice e non abbastanza preciso per essere utilizzato in molte applicazioni che richiedono un modello che tenga in considerazione le non idealità della telecamera. È necessario dunque complicare il modello per far sì che esso rispecchi più fedelmente la realtà. In primo luogo si deve, infatti, notare che la posizione dell'origine delle coordinate non sempre coincide con il punto principale; in tale caso l'espressione diventa:

$$\begin{pmatrix} fX + Zp_x \\ fY + Zp_y \\ Z \end{pmatrix} = \begin{bmatrix} f & p_x & 0 \\ & f & p_y & 0 \\ & & 1 & 0 \end{bmatrix} \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix} \quad (2.4)$$

Dove  $(p_x; p_y)$  sono le coordinate del punto principale.

Introducendo:

$$K = \begin{bmatrix} f & & p_x \\ & f & p_y \\ & & 1 \end{bmatrix} \quad (2.5)$$

possiamo scrivere in forma concisa:

$$x = K[I | 0]X_{cam} \quad (2.6)$$

La matrice  $K$  è chiamata *matrice di calibrazione* della telecamera.  $X_{cam}$  è scritto in questa forma solo per enfatizzare che la telecamera viene posta nell'origine degli assi cartesiani con l'asse principale diretto lungo l'asse  $Z$ . Nel caso in cui il SDR mondo coincida con il SDR di telecamera allora la matrice di calibrazione coincide con la matrice di proiezione. In generale invece i punti nello spazio sono espressi in un diverso sistema di riferimento Euclideo, ovvero il SDR mondo. I due sistemi di coordinate sono collegati tra loro da una rototraslazione. Esprimendo quindi con  $R$  la matrice di rotazione tra i due SDR e con  $\tilde{C}$  le coordinate del centro della telecamera in coordinate mondo si ottiene:

$$x = KR[I | -\tilde{C}]X \quad (2.7)$$

Dove  $X$  è ora espresso in coordinate mondo. Questa è la generica mappatura descritta dal modello pinhole camera. La matrice di calibrazione ha dunque 9 GDL: tre dalla matrice  $K$  (gli elementi  $f$ ;  $p_x$ ;  $p_y$ ), tre dalla matrice  $R$  (esprimibile come rotazione su 3 assi), e tre da  $\tilde{C}$  (tre coordinate spaziali del centro della telecamera). La matrice  $K$  contiene le informazioni dei parametri che dipendono unicamente dalla telecamera e che quindi vengono chiamati *intrinseci*, mentre  $R$  e  $\tilde{C}$  sono dipendono dalla posizione e orientamento della telecamera rispetto al mondo esterno, per tale ragione i parametri che li compongono sono chiamati *estrinseci*.

Spesso risulta conveniente non esprimere direttamente il centro della telecamera ma rappresentarlo in termini di trasformazione tra coordinate mondo e immagine. Si può quindi indicare:

$$P = K[R|t] \quad \text{con} \quad t = -R\tilde{C} \quad (2.8)$$

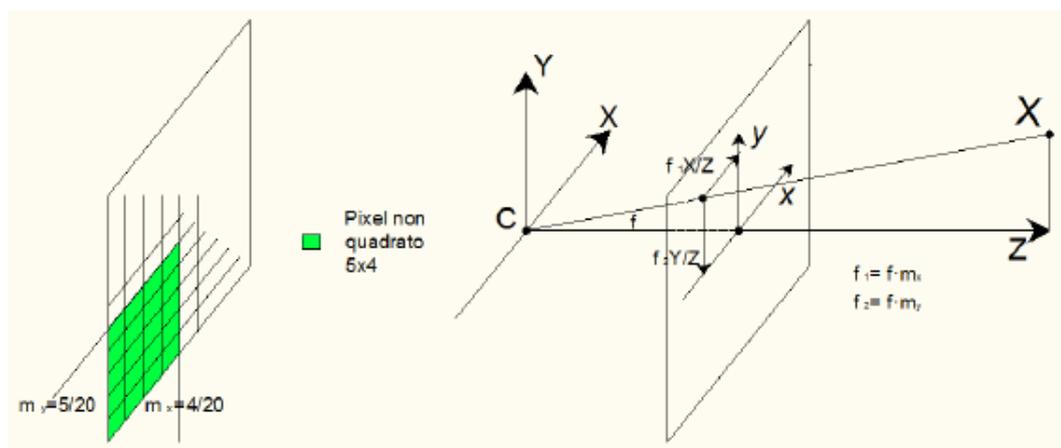
In questa tesi sono state utilizzate soltanto *telecamere digitali*; il sensore di tali telecamere è una matrice di semiconduttori chiamati *pixel*. Ognuno di essi compie la conversione da fotoni a elettroni della luce incidente nel suo settore. Un'immagine è formata quindi dalle informazioni coordinate di tutti i pixel che compongono il sensore. Tali componenti sono dotati di diverse sezioni: la prima effettua la conversione fotoelettrica, una seconda, detta di immagazzinamento e separata da quella di rilevamento, immagazzina temporaneamente le cariche acquisite da una precedente conversione, ed infine la terza sezione solitamente formata da un circuito di controllo integrato che gestisce le immagini, permettendone un immagazzinamento ulteriore in una memoria flash e gestisce l'invio in forma normale o compressa delle stesse ad altri dispositivi. La capacità di avere le immagini già in formato digitale è un forte contributo alla possibilità di eseguire successivamente operazioni automatiche su di esse.

I pixel delle telecamere non sono sempre perfettamente quadrati. Siccome normalmente le immagini vengono misurate in pixel, ciò rende diverso il fattore di scala orizzontale e verticale del sensore. Tale costruzione si vede in maniera molto accentuata in figura 2.6: dimensioni non quadrate dei pixel si traducono in un numero diverso di pixel per unità di lunghezza che comporta una scalatura diversa nella proiezione del punto lungo le due direzioni.

Per tener conto di questo fenomeno la forma generale della matrice di calibrazione delle telecamere digitali diventa quindi :

$$K = \begin{bmatrix} f_x & x_0 \\ & f_y & y_0 \\ & & 1 \end{bmatrix} \quad (2.9)$$

con  $f_x = f \cdot m_x$  e  $f_y = f \cdot m_y$ , dove  $m_x$  e  $m_y$  rappresentano il numero di pixel per unità di distanza rispettivamente lungo x e y.



**Figura 2.6:** Matrice dei pixel di una telecamera digitale con distanze focali diverse lungo x e y e effetto di questo fattore sulle coordinate immagine.

E' necessario introdurre un ulteriore parametro per descrivere interamente il modello di una telecamera digitale, ovvero lo *skew*. Esso descrive la possibilità che i pixel abbiano lati non perpendicolari tra loro ed è espresso da:

$$s = \cos(\alpha) * f \tag{2.10}$$

dove  $\alpha$  è l'angolo formato dai lati del pixel. Tale parametro è generalmente uguale a 0 per la maggior parte delle telecamere.

In definitiva la matrice di calibrazione viene espressa come:

$$K = \begin{bmatrix} f_x & s & x_0 \\ & f_y & y_0 \\ & & 1 \end{bmatrix} \tag{2.11}$$

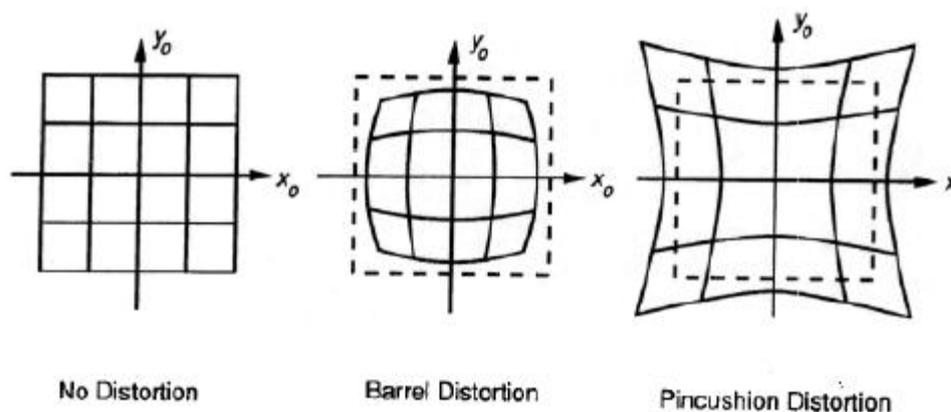
Tale modello di telecamera ha ora 11 GDL, due in più rispetto al precedente per l'aggiunta del parametro di skew e per la possibile diversità tra le due lunghezze focali.

Il principale vantaggio del pinhole camera è che descrive solamente relazioni lineari tra i parametri di telecamera. In molti casi tale modello offre una buona approssimazione della realtà. Non risulta però sufficientemente accurato nei casi in cui sono presenti delle distorsioni ottiche considerevoli. Tali distorsioni possono presentarsi in diversi modi, ma generalmente sono individuabili come delle non-linearità del sistema di visione. Il modello viene quindi modificato ulteriormente per avvicinarsi maggiormente alla geometria reale di una

telecamera, nella quale i punti del mondo, nel piano immagine e il centro ottico non sono più collineari (ovvero collegati da una linea retta) come nel caso ideale. Tali effetti sono molto visibili in figura 2.7: le linee dell'immagine, che dovrebbero essere rette e parallele tra loro sono invece curvate in maniera tanto più accentuata quanto più si trovano ai bordi dell'immagine.



**Figura 2.7:** Differenza tra un'immagine soggetta a distorsioni radiali (a sinistra) e una non soggetta (a destra).



**Figura 2.8:** Rappresentazione delle tipologie di distorsioni radiali; da sinistra verso destra: indeformato, a botte, a puntaspilli.

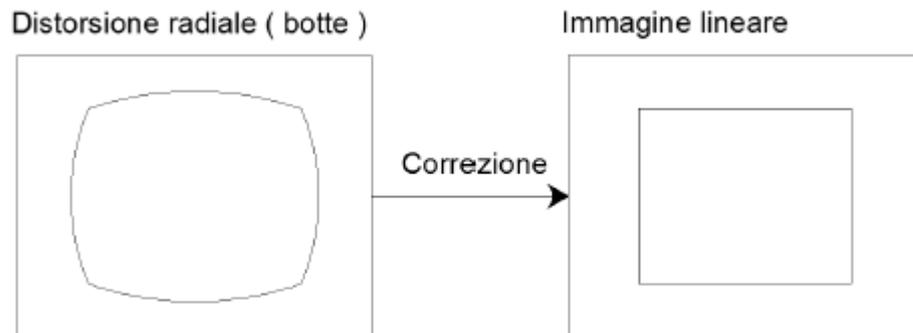
Tale distorsione viene chiamata *radiale* e può verificarsi in due modalità, a seconda del segno della distorsione; a botte oppure a puntaspilli (figura 2.8). Questa distorsione solitamente diventa più accentuata con il diminuire della distanza focale.

Una volta corrette le distorsioni, il fenomeno di formazione delle immagini può essere nuovamente ricondotto al caso lineare. Il processo per effettuare questa correzione viene mostrato in figura 2.10; in coordinate pixel risulta:

$$\hat{x} = x_c + L(r)(x - x_c) \quad \hat{y} = y_c + L(r)(y - y_c) \quad (2.12)$$

$$\text{con } r^2 = (x - x_c)^2 + (y - y_c)^2$$

dove  $(x; y)$  sono le coordinate immagine misurate,  $(\hat{x}; \hat{y})$  sono le coordinate corrette, mentre  $(x_c; y_c)$  è il centro delle distorsioni radiali. La funzione  $L(r)$  è definita solo per valori positivi di  $r$  e viene solitamente risolta come uno sviluppo di un'equazione di Taylor  $L(r) = 1 + k_1 r + k_2 r^2 + k_3 r^3 + \dots$



**Figura 2.9:** Correzione di un'immagine soggetta a distorsione radiale.

I coefficienti di distorsione  $k_1, k_2, k_3, \dots$  sono considerati parte della calibrazione interna della telecamera.

Un modello migliore per trattare tali aberrazioni ottiche venne introdotto da Brown nel 1966 [11], chiamato modello “*Plumb Bob*” (piombino). Lo spostamento del centro delle distorsioni viene esplicitato come una distorsione tangenziale, dovuta solitamente a un problema di montaggio della lente o a un difetto costruttivo. Indicando con  $x_n$  le coordinate normalizzate di telecamera  $(x, y)$  la correzione delle coordinate diventa:

$$x_d = (1 + k_1 r^2 + k_2 r^4 + k_3 r^6)x_n + dx \quad (2.13)$$

Con  $dx$  termine che descrive lo spostamento introdotto dalle distorsioni tangenziali:

$$dx = \begin{bmatrix} 2p_1xy + p_2(r^2 + 2x^2) \\ p_1(r^2 + 2y^2) + 2p_2xy \end{bmatrix} \quad (2.14)$$

Si è infine giunti al modello completo per descrivere la telecamera. Questo è caratterizzato da 16 GDL, separabili in intrinseci ed estrinseci. La calibrazione è l'insieme delle operazioni finalizzate a determinare tali parametri, che a scopo riassuntivo vengono qui indicati nella loro notazione completa e suddivisi in parametri intrinseci, estrinseci e di distorsione:

Modello pinhole camera completo:

- 5 parametri intrinseci:
  - $f_1, f_2$  ovvero la distanza focale espressa nelle due direzioni
  - $c_1, c_2$  le due coordinate del centro ottico
  - $s$  ovvero lo skew
- 6 parametri estrinseci:
  - $r_1, r_2, r_3$  tre parametri di rotazione esprimibili in angoli di Eulero, Rodrigues etc.
  - $t_1, t_2, t_3$  tre parametri di traslazione
- 5 parametri di distorsione:
  - $k_1, k_2, k_3$  tre distorsioni radiali
  - $p_1, p_2$  due distorsioni tangenziali

Non sempre però è necessario ricorrere all'utilizzo del modello completo; in alcuni casi, infatti, la complessità introdotta da questo modello è eccessiva per le applicazioni in cui esso è utilizzato. Spesso le distorsioni sono trascurabili; in particolare è abbastanza frequente che non si presentino le prime due distorsioni radiali, mentre è ancora più frequente che non si presentino quelle tangenziali e l'ultima radiale [6],[7]. Per motivi legati alla convergenza dei metodi di minimizzazione, considerare nel modello anche parametri che non sono presenti in realtà, conduce spesso a risultati di calibrazione sbagliati.

## **2.2 Calibrazione mediante il metodo di Zhang e di Levenberg-Marquardt**

La calibrazione è un processo essenziale per poter effettuare misure tramite telecamere. Nel capitolo 2.1.1 sono state descritte le principali tecniche presenti in letteratura per affrontare questo problema. L'applicazione nella quale si utilizzeranno i dispositivi studiati in questo lavoro di tesi, vale a dire misurazioni, richiede che i risultati di calibrazione siano molto accurati. Altra specifica importante è che il setup del processo di calibrazione possa essere semplice ed economico. L'intera procedura deve essere svolta nell'arco di poche ore al massimo, quindi si può dire che il tempo non è una specifica critica; la calibrazione può essere infatti effettuata prima di eseguire le misure, una volta impostato l'apertura del diaframma e la focale. In caso venga variato uno di questi parametri l'intera procedura deve necessariamente essere rieseguita. Se invece viene modificata solamente la posa sarà necessario solo registrar la telecamera, cioè calcolarne la posa rispetto a un sistema di riferimento fissato; andranno quindi ricalcolati solamente i parametri estrinseci, senza ripetere l'intera calibrazione.

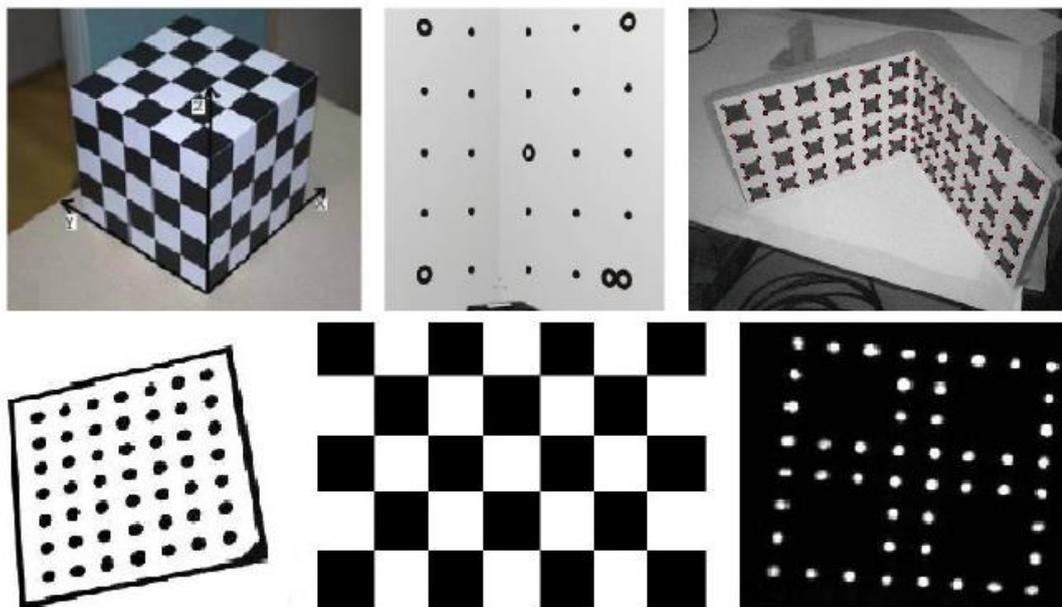
Date queste specifiche è stato deciso di utilizzare il metodo di Zhang per risolvere il modello lineare; i risultati di questo poi possono essere utilizzati per inizializzare un problema di ottimizzazione del modello non lineare completo. Questo è stato risolto utilizzando l'algoritmo di Levenberg-Marquardt (LM) come vedremo in seguito. La tecnica di Zhang-LM è già stata validata nel corso degli anni da molti studiosi [9] ed è stata già implementata dallo staff del VBLab. Inoltre in [7] si sono eseguite numerose prove con lo scopo di studiare più a fondo questo approccio alla calibrazione. Partendo allora dalle conoscenze già acquisite su tale tecnica dal VBLab, si vuole quindi confrontarla con altre metodologie. Si passa ora a descrivere più in dettaglio la procedura di calibrazione utilizzando questi algoritmi.

## 2.2.1 Acquisizione dei dati di calibrazione

L'acquisizione delle immagini del calibratore in pose differenti si può considerare la parte più critica dell'intera procedura di calibrazione. Infatti se non svolta in maniera corretta, si rischia di introdurre numerosi errori che limitano le prestazioni degli algoritmi di calibrazione. Prima di tutto bisogna scegliere un calibratore adeguato per l'applicazione, scelta essenziale per poter ottenere una buona accuratezza di calibrazione. Di norma esso deve occupare da  $1/8$  a  $1/4$  dell'immagine [7]. Dall'esperienza sviluppata in laboratorio si è osservato che con calibratori più piccoli si ottengono risultati migliori. Questo perché punti vicini sono in grado di caratterizzare più in dettaglio le zone del sensore. Una volta scelta la focale e l'apertura del diaframma, sarà possibile acquisire immagini di oggetti solamente a una data *profondità di campo*, cioè l'insieme di distanze in cui l'oggetto appare a fuoco. È, infatti, necessario che tutti i punti noti del calibratore siano a fuoco. Questo fattore limita gli spostamenti dell'oggetto sull'asse principale della telecamera; le dimensioni del calibratore allora dipendono dalla profondità di campo in cui si vuole acquisire misurazioni.

Una volta individuate le dimensioni del calibratore da utilizzare, bisogna determinarne le caratteristiche; esso, infatti, deve necessariamente possedere alcune proprietà fondamentali:

- Punti noti facilmente distinguibili (almeno 4).
- Distanze tra i punti noti conosciute con accuratezza.
- Caratteristiche morfologiche accurate (tolleranze di planarità e parallelità molto elevate).
- Osservabile con la telecamera utilizzata (ad esempio costruito con marker rifrangenti se vengono utilizzati filtri ad infrarossi)



**Figura 2.10:** Esempi di calibratori 3D e planari.

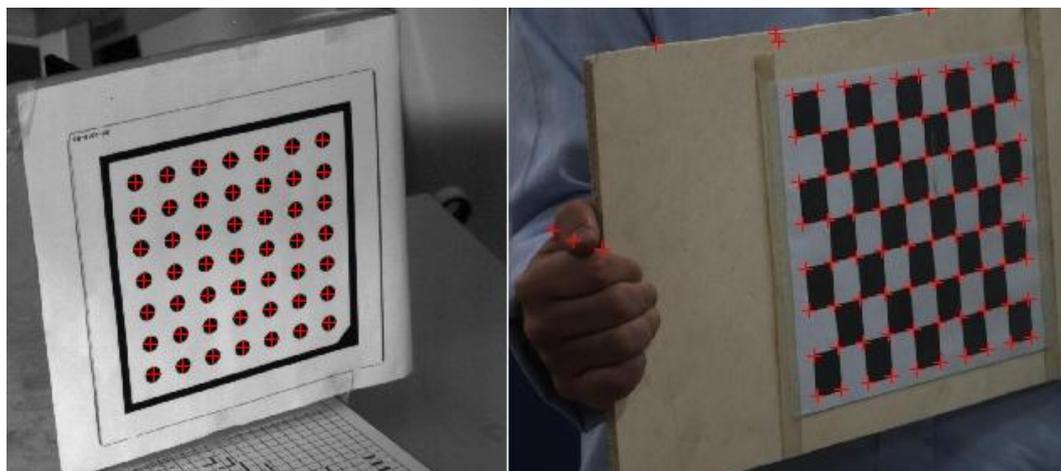
In virtù delle caratteristiche sopra elencate, generalmente i punti notevoli sono rappresentati da *blob* (regioni nell'immagine che sono più chiare o scure rispetto alle circostanti) o *corner*. Gli spigoli sono invece dei punti che corrispondono a brusche variazioni d'intensità di scala di grigi. Nella figura 2.10 sono rappresentati diversi esempi di calibratore. Le figure sotto rappresentano calibratori planari, mentre sopra sono raffigurati esempi di calibratori tridimensionali, che non sono altro che un insieme di calibratori planari messi in pose differenti nello spazio. Si nota poi come alcuni di essi presentano alcuni cerchi (*blob*) facilmente distinguibili dal resto della scena, mentre altri sono caratterizzati da zone quadrate che presentano *corners* facilmente determinabili.

Dopo aver scelto il calibratore più adatto, si procede all'acquisizione delle immagini. Esse devono essere acquisite in modo da caratterizzare al meglio tutto il sensore e con una qualità elevata, così da poter determinare con accuratezza la posizione dei punti d'interesse. Più si varia la posizione del calibratore, in modo da coprire tutto lo spazio in cui avverranno le misure, e più la calibrazione sarà accurata. Anche le pose devono essere più varie possibili; come si vedrà nel seguito, è indispensabile per risolvere il problema di calibrazione che nelle

immagini acquisite il calibratore giaccia su almeno tre piani differenti. Chiaramente la procedura di acquisizione delle immagini si può eseguire sia muovendo il calibratore che spostando le telecamere in diverse posizioni.

Anche l'illuminazione della scena rappresenta un importante parametro per la riuscita della calibrazione. I punti noti del calibratore, infatti, devono ricevere un'illuminazione omogenea e tale da creare un forte contrasto tra essi e il resto della scena (senza però saturare l'immagine), in modo da renderli ben visibili e riconoscibili.

Le tecniche più utilizzate per l'estrazione dalle coordinate nelle immagini dei punti noti sono l'analisi dei *blob* [12] e l'algoritmo di *corner detection* di Harris [25]. L'analisi dei *blob* individua i punti d'interesse tramite la binarizzazione dell'immagine e dalla successiva individuazione delle regioni adiacenti caratterizzate dalla medesima polarità. Per fare questo occorre solitamente definire una soglia di taglio (*threshold*), che viene scelta per discriminare in chiaro o scuro i 255 livelli dell'immagine, separando in tal modo quello che viene chiamato sfondo dalle regioni di interesse. L'immagine viene così binarizzata e può successivamente essere analizzata con gli strumenti tipici della *blob analysis*, ovvero tutto quell'insieme di algoritmi che di tali regioni calcolano area e forme.



**Figura 2.11:** Esempio di applicazione dell'algoritmo di *blob detection* (a sinistra) e di *Harris corner detection* (a destra).

Il metodo di Harris permette di determinare le coordinate degli spigoli (o *corners*) presenti nell'immagine. Questo, proposto da C. Harris e Stephens [25], tratta il problema dell'individuazione dei corners attraverso funzioni locali di autocorrelazione dell'intensità d'immagine. Vengono create tante piccole finestre di dimensioni fisse all'interno della quale si esegue la ricerca dei corners. Per ognuna di queste si ricava una funzione che esprime la variazione di intensità nelle due direzioni. Tale intensità viene utilizzata per individuare se la finestra corrente contiene un angolo. La finestatura viene eseguita per tutta l'immagine, individuando così una matrice di punti con la posizione di tutti gli angoli individuati.

Sono disponibili software che permettono a un utente dotato di competenze non specializzate di impostare direttamente le soglie corrette per effettuare le analisi sopra descritte. In questa tesi, ad esempio, è stato utilizzato un software sviluppato presso il laboratorio VBLab, chiamato "FeatureLinker". Esso consente di trovare in maniera manuale le corrispondenze tra i punti del modello che descrive il calibratore e i punti nell'immagine. Si è deciso di utilizzare un approccio manuale, più lento e laborioso di un approccio completamente automatico, perché con un riscontro visivo immediato, si riduce la possibilità di avere errori in fase di correlazione dei punti. Si aggiunge infine che per ottenere informazioni sull'orientamento di calibratori quadrati o rettangolari è bene inserire un'asimmetria nel calibratore.

## 2.2.2 Calcolo del modello lineare

Se l'applicazione per cui viene eseguita la calibrazione non richiede accuratèzze elevate o se le distorsioni delle lenti sono trascurabili, il calcolo del modello lineare rappresenta l'intero processo di calibrazione. Se invece è importante utilizzare un più completo modello non lineare, allora i risultati provenienti dal problema lineare possono essere utilizzati per inizializzare un algoritmo di ottimizzazione non lineare. Alcune tecniche di calibrazioni non utilizzano l'algoritmo di Zhang, ma si limitano a inizializzare i parametri con valori sensati

(dimensioni del sensore, dimensioni del pixel, ottica montata), lasciando alla sola ottimizzazione finale il compito di convergere ai valori veri. Chiaramente non sempre è possibile conoscere questi dati con precisione e dunque risulta più utile partire dalla tecnica di Zhang.

L'algoritmo di Zhang prende in ingresso le coordinate che descrivono il calibratore planare e i corrispondenti punti nelle varie immagini e consente di calcolare i parametri intrinseci lineari ed estrinseci. Dato che il sistema di riferimento è considerato solidale al calibratore, i parametri estrinseci della telecamera saranno diversi per ogni immagine acquisita. Inoltre è possibile considerare nulla la coordinata  $Z$  dei punti nello spazio, cioè la coordinata ortogonale al piano del calibratore. Questo ci consente di limitare il problema al semplice calcolo di più omografie. L'*omografia* è la relazione tra punti in due spazi tali per cui ogni punto di uno spazio corrisponde ad uno e un solo punto del secondo spazio.

Per eseguire il calcolo dell'omografia si procede come segue:

1. Si normalizzano i punti  $X$  ovvero i punti mondo bidimensionali (poiché  $Z=0$ ).
2. Si normalizzano i punti  $x$  ovvero i punti ricavati dalle immagini.
3. Si applica l'omografia  $H$  fra i punti del modello del calibratore e i punti trovati nell'immagine.
4. Si denormalizza la matrice  $H$ .

La normalizzazione è necessaria per poter avere un migliore condizionamento del problema. Così facendo si riferiscono i punti del calibratore ad una posizione fissa nello spazio. Solitamente tale posizione è scelta in modo da posizionare il centro di massa dei punti nell'origine  $(0; 0)^T$ , scalati in modo che la loro distanza media dall'origine sia  $\sqrt{2}$ .

La normalizzazione si riconduce nell'individuare una matrice di trasformazione che la effettua, la cui equazione può essere ricavata da libri di testo (ad esempio [1]) e risulta essere:

$$T = \begin{bmatrix} \frac{\sqrt{2}}{std(x)} & 0 & -\left(\frac{\sqrt{2}}{std(x)}\right)m_x \\ 0 & \frac{\sqrt{2}}{std(y)} & -\left(\frac{\sqrt{2}}{std(y)}\right)m_y \\ 0 & 0 & 1 \end{bmatrix} \quad (2.15)$$

In cui  $m_x$  e  $m_y$  sono le medie dei punti nelle due direzioni e  $std(x)$  e  $std(y)$  sono le deviazioni standard. I punti normalizzati mediante la matrice di trasformazione T saranno  $\tilde{X} = TX$ . Per i punti immagine si applica la medesima procedura ottenendo anche qui i loro corrispondenti normalizzati  $\tilde{x} = Tx$ . Eseguita la normalizzazione si può applicare l'algoritmo per il calcolo delle omografie che, partendo dalle corrispondenze  $\tilde{X} \leftrightarrow \tilde{x}_i$ , determina la matrice H per cui  $\tilde{x}_i = H\tilde{X}_i$ . La procedura viene descritta in dettaglio in [1]. La matrice H così individuata è una 3x3 con 8 gradi di libertà. Definendo con  $h_i$  le colonne di H e con  $r_i$  le colonne della matrice di rotazione è possibile esprimere, per un calibratore planare, la relazione che lega i parametri intrinseci all'omografia:

$$H = [h_1 \ h_2 \ h_3] = K [r_1 \ r_2 \ t] \quad (2.16)$$

Questa espressione è possibile se il calibratore è planare ovvero  $Z=0$ ; infatti la terza colonna della matrice di rotazione, essendo moltiplicata per zero, non viene indicata.

La matrice H è dipendente dunque sia dai parametri intrinseci che da quelli estrinseci. Ricordando che per ogni immagine di calibrazione viene calcolata una omografia, degli 8 GDL di H ne vengono utilizzati 6 per definire la posa della telecamera, ovvero la matrice di rotazione e il vettore di traslazione, quindi ne rimangono solo 2 per definire i parametri intrinseci. Questo vincolo a volte non è sufficiente poiché, a seconda del modello utilizzato, per rappresentare la telecamera, si deve fare ricorso ad un numero di parametri intrinseci maggiore di 2. Per far fronte a questo problema bisogna disporre di corrispondenze tra punti in

più immagini in cui il calibratore è posto in pose differenti tra loro. Per definire 5 parametri intrinseci si dovranno acquisire almeno 3 immagini. A questo punto, per determinare i parametri intrinseci, è necessario imporre un vincolo costante in tutte le immagini che viene utilizzato per correlarle tra loro; tale vincolo è quello di ortonormalità tra  $r_1$  e  $r_2$ , che equivale ad imporre che gli assi  $x$  e  $y$  siano ortogonali tra loro in tutte le immagini:

$$\begin{cases} h_1^T K^{-T} K^{-1} h_2 = 0 \\ h_1^T K^{-T} K^{-1} h_1 = h_2^T K^{-T} K^{-1} h_2 \end{cases} \quad (2.17)$$

La soluzione viene trovata, come mostrato da Zhang in [2], indicando con  $B = K^{-T} K^{-1}$  e con  $b = [B_{11} \ B_{12} \ B_{22} \ B_{13} \ B_{23} \ B_{33}]^T$ . L'equazione 2.17 diventa:

$$h_i^T B h_i = v_{ij} b \quad (2.18)$$

Con  $v_{ij}$  costruito come:

$$v_{ij} = [h_{i1} h_{j1}; h_{i1} h_{j2} + h_{i2} h_{j1}; h_{i2} h_{j2}; h_{i3} h_{j1} + h_{i1} h_{j3}; h_{i3} h_{j2} + h_{i2} h_{j3}; h_{i3} h_{j3}]$$

È possibile allora risolvere l'equazione formulando il problema come:

$$\begin{bmatrix} v_{12}^T \\ (v_{11} - v_{22})^T \end{bmatrix} b = 0 \quad (2.19)$$

ovvero nella forma  $Vb = 0$  risolvibile tramite la tecnica della decomposizione ai valori singolari SVD. I parametri intrinseci si estraggono direttamente dalla matrice  $B$  tramite le equazioni:

$$\begin{aligned} x_0 &= (B_{12} B_{13} - B_{11} B_{23}) / (B_{11} B_{22} - B_{21}^2) \\ \lambda &= B_{33} - [B_{13}^2 + x_0 (B_{12} B_{13} - B_{11} B_{23})] / B_{11} \\ f_1 &= \sqrt{\lambda / B_{11}} \\ f_2 &= \sqrt{\lambda B_{11} / (B_{11} B_{22} - B_{21}^2)} \\ y_0 &= s x_0 / f_2 - B_{13} f_1^2 / \lambda \end{aligned} \quad (2.20)$$

dove  $\lambda$  indica il coefficiente di scala.

Per quanto riguarda i parametri estrinseci essi possono essere ricavati, una volta determinata la matrice  $K$ , come:

$$\begin{aligned} r_1 &= \lambda K^{-1} h_1 \\ r_2 &= \lambda K^{-1} h_2 \\ r_3 &= r_1 \times r_2 \\ t &= \lambda K^{-1} h_3 \end{aligned} \quad (2.21)$$

In generale, la matrice di rotazione ottenuta potrebbe non rispettare il vincolo di ortonormalità ( $R^T R = I$ ) a causa di errori sui punti; si riporta la matrice di rotazione ad una che rispetta il vincolo risolvendo il problema di minimizzazione:

$$\min_R \|R - Q\|_F^2 \quad \text{soggetto a } R^T R = I \quad (2.22)$$

Questo problema può essere risolto più semplicemente mediante la scomposizione ai valori singolari, esprimendo  $Q = USV^T$  dove  $S$  è una matrice diagonale contenente gli autovalori di  $Q$ , mentre  $U$  e  $V$  sono matrici unitarie. Si ricava  $R$  come:

$$R = UV^T \quad (2.23)$$

La matrice qui trovata rappresenta la soluzione del problema ai minimi quadrati [1].

### 2.2.3 Ottimizzazione non lineare

É ora possibile utilizzare i risultati ottenuti mediante la procedura lineare di Zhang fin qui descritta come soluzione di primo tentativo per la risoluzione del problema non lineare di calibrazione. Per fare ciò si utilizza uno dei metodi iterativi per la risoluzione del problema non lineare di minimizzazione; tali tecniche presentano solitamente alcuni svantaggi:

- la convergenza può essere lenta;
- richiedono valori iniziali prossimi alla soluzione;
- esiste la possibilità di non-convergenza, o convergenza ad un minimo locale anziché globale;
- la scelta dei criteri di terminazione dell'iterazione non è sempre banale;

- il problema numerico fa perdere ai parametri il loro rigido significato fisico, generando soluzioni inaccettabili fisicamente e difficili da verificare.

Di conseguenza, tali tecniche richiedono generalmente una più attenta implementazione, valutazione dei risultati ottenuti. Le tecniche di minimizzazione iterativa consistono generalmente in cinque passaggi:

1. Funzione di costo: una funzione di costo è scelta come nucleo della minimizzazione. Per il problema corrente è possibile impostare diverse funzioni.
2. Parametrizzazione: le trasformazioni devono essere espresse in funzione di un finito numero di parametri. Non sempre è necessario considerare il set minimo di parametri, ma in alcuni casi una sovrparametrizzazione può portare dei vantaggi.
3. Scelta della funzione di costo.
4. Inizializzazione: il calcolo dei parametri iniziali deve essere eseguito con accuratezza; nel caso corrente si può utilizzare il metodo di Zhang descritto sopra.
5. Iterazioni: iniziando dalla soluzione di primo tentativo, i parametri sono iterativamente migliorati minimizzando la funzione di costo.

È possibile utilizzare differenti funzioni di costo; in letteratura [1],[3] sono utilizzati ad esempio:

- Distanza algebrica
- Distanza geometrica in una o due immagini
- Errore di retroproiezione
- Errore di Sampson
- Funzioni di costo statistiche (un'immagine, due immagini, distanza di Mahalanobis)

In questa tesi, come nella maggior parte della letteratura [1],[9],[10], è stato scelto di utilizzare, come funzione obiettivo per l'ottimizzazione non lineare, l'*errore di retroproiezione*, cioè la distanza tra i punti individuati nelle immagini (i dati di partenza) e i punti del calibratore trasformati sulla base dei parametri di

telecamera, ovvero retroproiettati. Questa scelta è stata fatta poiché l'errore di retroproiezione mantiene fortemente un significato fisico ed è quindi immediato da verificare con riscontri visivi, quali grafici o immagini. Inoltre questa funzione di costo è relativamente semplice da implementare ed equivale a cercare il migliore set di parametri che minimizzi la funzione:

$$\sum_i d(x'_i, H_i x_i)^2 \quad (2.24)$$

Dove  $d(x, y)$  rappresenta la distanza euclidea tra i punti individuati nell'immagine  $x'_i$  e i punti del calibratore retroproiettati  $H_i x_i$ .

Per ogni funzione di costo ci sono differenti scelte di parametrizzazione. In generale nessun effetto negativo è stato riscontrato se il problema di minimizzazione viene sovra-parametrizzato. In particolare algoritmi come quello di Gill-Murray [24] (che è una modifica di quello di Gauss-Newton) o quello di Levenberg-Marquardt [1],[3],[4],[5] gestiscono parametri ridondanti facilmente.

Per semplificare il processo di parametrizzazione si osserva che  $H_i = K[R_i t_i]$ , vale a dire che la matrice  $H$  può essere scomposta in due parti separate: un insieme di parametri dipendente solamente dai parametri intrinseci (e dunque comuni a tutte le immagini) e un insieme di parametri estrinseci dipendente dalla particolare immagine. L'equazione complessiva potrà essere ottenuta utilizzando un insieme di parametri creato in modo da tenere conto solamente una volta della matrice  $K$  e separatamente, per ogni immagine, della sua rototraslazione relativa. La parametrizzazione del problema diventa:

$$f : (K, R_1, t_1, \dots, R_n, t_n) \quad (2.25)$$

L'equazione da minimizzare risulta quindi:

$$\sum_{i=1}^n \sum_{j=1}^m \|x_{ij} - \hat{x}_i(K, R_i, t_i, X_j)\| \quad (2.26)$$

Dove  $X_j$  sono i punti del calibratore e  $\hat{x}_i(K, R_i, t_i, X_j)$  la loro retroproiezione sull'immagine  $i$  secondo l'equazione  $\hat{x}_i = H X_j$ . Si indica con  $n$  il numero di punti in un'immagine, e con  $m$  il numero di immagini considerate. Inoltre, per ridurre il numero dei parametri, solitamente la matrice  $R_{3 \times 3}$  viene parametrizzata con un

vettore di 3 elementi corrispondenti a angoli di rotazioni rappresentati in una forma standard quale ad esempio gli angoli di Rodrigues o di Eulero.

Per completare il modello della telecamera bisogna considerare anche i parametri non lineari che si riferiscono alle distorsioni delle lenti. Come descritto in precedenza, tali distorsioni dipendono dalla conformazione della lente e non variano da un'immagine all'altra. Tali parametri possono essere considerati intrinseci. Il problema finale diventa quindi la minimizzazione della funzione:

$$\sum_{i=1}^n \sum_{j=1}^m \|x_{ij} - \hat{x}_i(K, k_1, k_2, k_3, p_1, p_2, R_i, t_i, X_j)\| \quad (2.27)$$

Non sempre è utile inserire tutti i parametri che rappresentano le distorsioni nel modello; alle volte, anzi, è meglio trascurarne alcune, come ad esempio l'ultima radiale, che di solito è meno influente delle altre.

Un volta scelto il set di parametri e la funzione di costo, si passa alla selezione del metodo di minimizzazione. Esistono vari metodi iterativi [1],[3] ed i più utilizzati, descritti in seguito, sono:

- Metodo di Gauss-Newton;
- Metodo a discesa di gradiente;
- Metodo di Levenberg-Marquardt;

Supponendo di avere una relazione funzionale  $X = f(P)$ , dove  $X$  è il vettore delle misure e  $P$  è il vettore dei parametri rispettivamente in  $\mathbb{R}^n$  e in  $\mathbb{R}^m$ , si desidera trovare il vettore  $\hat{P}$  che meglio soddisfa tale relazione. Più precisamente si cerca il vettore  $\hat{P}$  che soddisfa  $X = f(P) - \epsilon$ , nel quale viene minimizzato  $\|\epsilon\|$ . Nel caso  $f$  sia una funzione lineare, tale sistema è risolto semplicemente con l'equazione DLT. Se invece è una funzione non-lineare, si deve partire da un valore iniziale dei parametri  $P_0$  per poi procedere al raffinamento della stima sotto l'assunzione che la funzione  $f$  sia localmente lineare. Sia  $\epsilon_0 = f(P_0) - X$  e si assuma che la funzione  $F$  sia approssimata in  $P_0$  da  $f(P_0 + \Delta) = f(P_0) + J\Delta$ , dove  $J$  è una mappatura lineare rappresentata dalla matrice Jacobiana  $J = \frac{df}{dP}$ .

La ricerca del punto successivo d'iterazione  $f(P_1)$ , avviene con la funzione  $P_1 = P_0 + \Delta$ , che minimizza  $f(P_1) - X = f(P_0) + J\Delta = \epsilon_0 + J\Delta$ . Ovvero si riduce al calcolo di un problema di minimizzazione lineare descritto da  $\| \epsilon_0 + J\Delta \|^2$  su  $\Delta$ . Il vettore  $\Delta$  è ottenuto risolvendo l'equazione normale  $\Delta = -J^+ \epsilon_0$ . Dove  $J^+$  è la pseudo inversa di  $J$ . Quindi il vettore  $\hat{P}$  è ottenuto iniziando da una  $P_0$  stimato e calcolato ricorsivamente in accordo alla formula:

$$P_{i+1} = P_i + \Delta_i \quad (2.28)$$

dove  $\Delta_i$  è la soluzione di un problema lineare ai minimi quadrati:

$$J\Delta_i = -\epsilon_i \quad (2.29)$$

La matrice  $J$  è lo Jacobiano di  $df/dP$  valutato in  $P_i$  e  $\epsilon_i = f(P_i) - X$ . Si ipotizza che tale algoritmo converga alla soluzione dei minimi quadrati richiesta; in alcuni casi è però possibile che le iterazioni convergano ad un minimo locale, o non convergano affatto. Il comportamento dell'algoritmo dipende fortemente dalla stima iniziale  $P_0$ . Si voglia ora cercare il minimo di una funzione di molte variabili. Si considera quindi una generica funzione  $g(P)$ ; l'obiettivo diventa ora la minimizzazione di tale funzione rispetto tutte le variabili che compongono  $P$ . Si fanno però due assunzioni: che  $g(P)$  abbia un valore minimo ben definito e che sia conosciuto un punto  $P_0$  ragionevolmente vicino al minimo. Espandendo in serie di Taylor la funzione rispetto  $P_0$  si ottiene:

$$g(P_0 + \Delta) = g + g'\Delta + \Delta^T g'' \frac{\Delta}{2} + \dots \quad (2.30)$$

Si indicano con  $g'$  le derivate prime della funzione e con  $g''$  le derivate seconde. Si desidera minimizzare questa quantità in rispetto a  $\Delta$ . A questo fine, si può derivare rispetto a  $\Delta$  e imporre la derivata a 0, arrivando quindi all'equazione  $g + g'\Delta = 0$  oppure:

$$g''\Delta = -g' \quad (2.31)$$

In questa equazione  $g''$  è la matrice delle derivate seconde, chiamata Hessiano di  $g$ , essa è composta da termini del tipo  $d^2g/dp_i dp_j$ , con  $i$  e  $j$  corrispondenti ai  $i$ -esimo e  $j$ -esimo componente del vettore dei parametri. Il vettore  $g'$  è il gradiente di  $g$ . L'equazione 2.31 è quella utilizzata dal *metodo di Newton* come parametro d'iterazione. Il problema principale di tale procedura sta nella difficoltà del calcolo dell'Hessiano; una soluzione a tale problema deriva dalla semplificazione introdotta dal *metodo di Gauss-Newton* [1],[3], nel quale l'Hessiano viene calcolato in maniera approssimata come:

$$g'' = J^T J \quad (2.32)$$

ovvero riscrivendo la (2.31):

$$J^T J \Delta = -J^T \epsilon \quad (2.33)$$

Una variante a tale metodo deriva dalla possibilità di utilizzare un passo di incremento variabile. Se si considera, infatti, che  $-g' = -J^T \epsilon$  definisce la direzione di più rapida discesa della funzione costo, una strategia di minimizzazione consiste nel muoversi iterativamente nella direzione del gradiente. La lunghezza del passo può essere calcolata ricercando lungo la linea definita dalla direzione negativa del gradiente. In questo caso il parametro di incremento  $\Delta$  viene calcolato dall'equazione  $\lambda \Delta = -g'$ , dove  $\lambda$  controlla la lunghezza dello step. Tale metodologia è chiamata *a discesa di gradiente* [1],[3], e può essere interpretata come un metodo di Newton nel quale l'Hessiano è stato approssimato da una matrice scalare  $\lambda I$ . La discesa di gradiente solitamente non è una buona strategia di minimizzazione, poichè tipicamente caratterizzata da una bassa convergenza dovuta a spostamenti “a zigzag”.

Il *metodo di Levenberg-Marquardt (LM)* [1],[3],[4],[5] può essere considerato come un'interpolazione tra il metodo di Gauss-Newton e la discesa di gradiente. L'equazione 2.33 viene sostituita da un'equazione normale aumentata :

$$(J^T J + \lambda I) \Delta = -J^T \epsilon \quad (2.34)$$

Il valore di  $\lambda$  varia tra un'iterazione e l'altra e assume tipicamente valori di circa un millesimo della media della diagonale di  $J^T J$ . Il valore d'incremento  $\Delta$  ottenuto risolvendo l'equazione aumentata viene ridotto in caso di aumento dell'errore, altrimenti esso viene accettato e  $\lambda$  viene diviso per un fattore (solitamente 10) prima della successiva iterazione. In altre parole, se  $\Delta$  porta a un incremento dell'errore, allora  $\lambda$  è incrementato dello stesso fattore e l'equazione aumentata viene risolta nuovamente; tale processo continua finché viene individuato un  $\Delta$  in grado di diminuire l'errore. Per meglio comprendere tale metodo si può osservare come per valori molto piccoli di  $\lambda$  questo si riduce essenzialmente al metodo di Gauss-Newton, in grado di convergere velocemente alla soluzione. Viceversa per valori grandi di  $\lambda$  l'equazione può essere approssimata da  $\lambda \Delta = -J^T \epsilon$ , corrispondente al metodo a discesa di gradiente, che garantisce una migliore convergenza in situazioni difficili. Tra tutti questi metodi, il più indicato per la risoluzione del problema di ottimizzazione non lineare dei parametri di calibrazione risulta quindi essere il metodo di Levenberg-Marquardt. Esso, pur essendo poco più lento rispetto agli altri due, è sicuramente più robusto; infatti, rispetto agli altri, anche partendo da parametri iniziali leggermente più lontani dal minimo globale, esso ha più possibilità di convergere alla soluzione corretta [1],[7].

## 2.2.4 Implementazione

L'implementazione del metodo di Zhang-LM è stata interamente svolta in MatLab dallo staff del VBLab. È stato sviluppato un software che risolve il problema di calibrazione utilizzando prima il metodo di Zhang e poi ottimizzando i risultati con l'algoritmo di Levenberg-Marquardt. L'interfaccia principale del software, mostrata in figura 2.12 consente all'utente di selezionare i dati su cui eseguire la calibrazione, cioè la sequenza di punti del calibratore e le corrispondenti posizioni nelle immagini. Tali dati sono stati creati a priori utilizzando il software "FeatureLinker" creato presso il VBLab (figura 2.13).

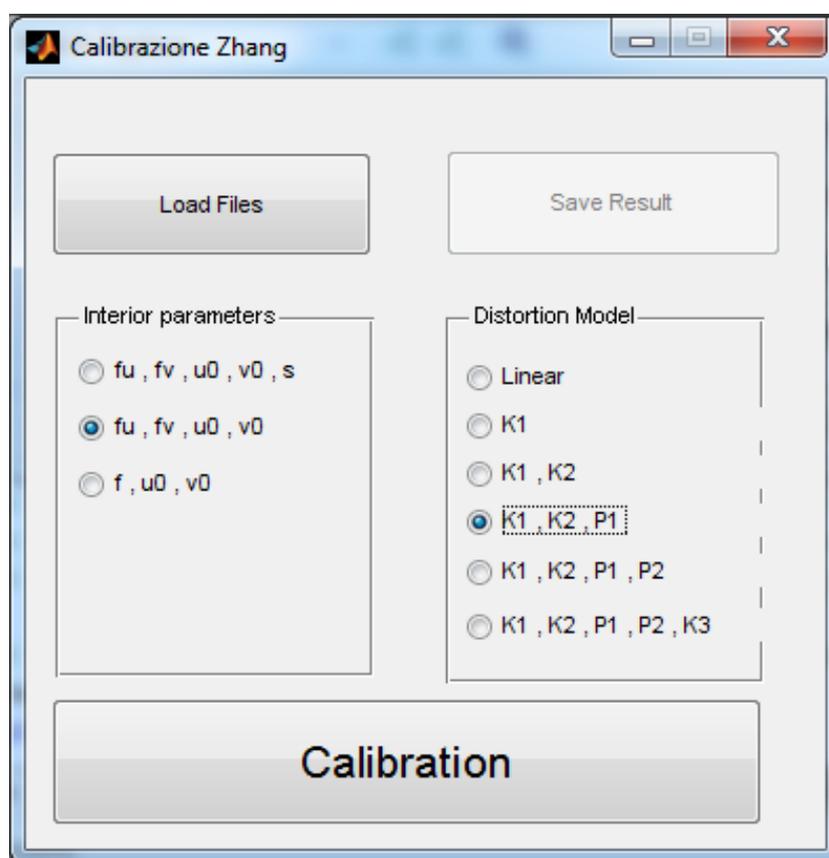
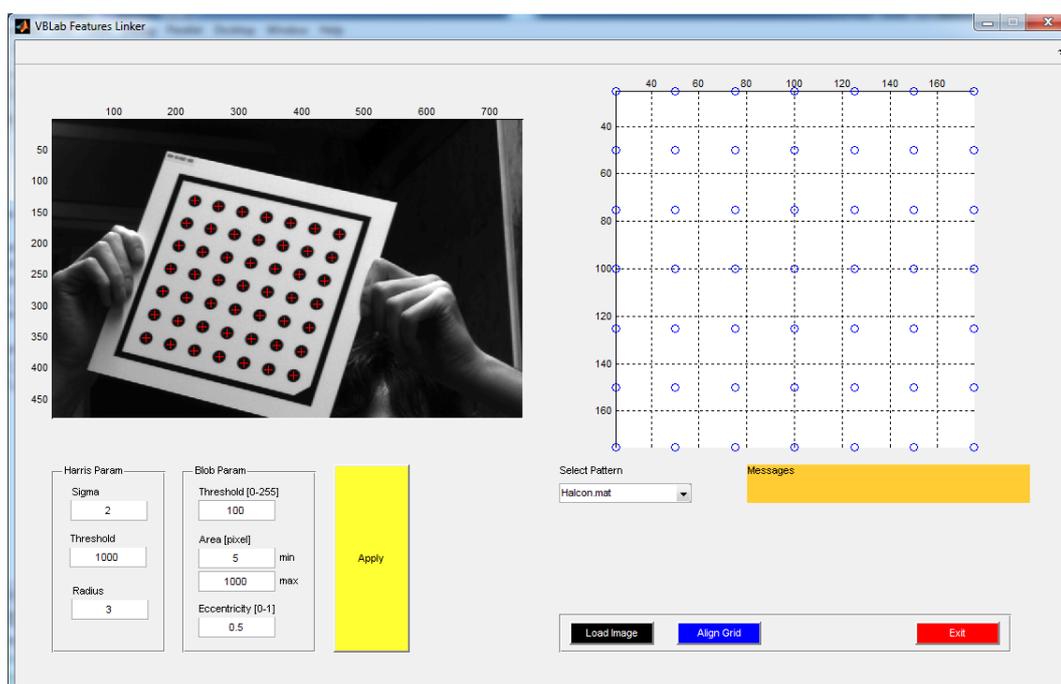


Figura 2.12: Interfaccia principale del software di calibrazione mediante metodo di Zhang-LM.



**Figura 2.13:** Interfaccia principale del software “Feature Linker” sviluppato presso il VBLab.

Esso consente di trovare manualmente le corrispondenze tra i punti del modello del calibratore e i loro rispettivi nelle immagini e salvarne le coordinate in file di testo. Dall’interfaccia principale inoltre l’utente ha la possibilità di selezionare il modello che sarà utilizzato in calibrazione. In particolare vi è la possibilità di utilizzare un modello lineare completo (5 parametri:  $skew$ , coordinate del centro ottico  $u_0$  e  $v_0$ , e lunghezze focali nelle due direzioni  $f_u$  e  $f_v$ ), oppure a 4 parametri (con  $skew=0$ ) o a 3 parametri (con  $f_u = f_v = f$ ). Inoltre si può aggiungere al modello lineare i parametri non lineari legati alle distorsioni (radiali K1, K2, K3 e tangenziali P1 e P2), e scegliere quali di questi utilizzare per la calibrazione.

## 2.3 Algoritmi genetici (GA)

Il costante sviluppo di alcune metodologie per l'ottimizzazione non lineare, quali gli algoritmi genetici, ha permesso lo sviluppo di alcuni metodi di calibrazione basati sull'utilizzo di queste tecniche. Queste metodologie non sono molto diffuse nel campo della visione artificiale in quanto viene criticato il fatto che esse si distaccano dal significato fisico dei vari parametri, evolvendo in modo casuale sino all'ottimo, e sono quindi difficili da gestire. Da qui si è deciso di provare a implementare e analizzare questi algoritmi applicandoli al problema della calibrazione e confrontarli con il metodo Zhang-LM sopra descritto. Gli algoritmi genetici sono descritti in dettaglio in seguito in questo capitolo.

In letteratura [14],[15],[16],[17],[18] vi sono svariati articoli che descrivono l'utilizzo di questi algoritmi per il problema di calibrazione. In [16], ad esempio, vengono utilizzati per risolvere solamente il modello lineare, in [14] invece viene implementata una procedura che consente la risoluzione del modello completo: partendo dalla conoscenza delle coordinate nello spazio di alcuni punti, viene ricavato il modello completo della telecamera. Con questo approccio, però, non viene sfruttato uno dei maggiori vantaggi dell'algoritmo di Zhang, ovvero un setup semplice ed economico, senza necessità di misurare la posa esatta del calibratore. Oltre ai lunghi tempi di preparazione della procedura di calibrazione, lo svantaggio di questo metodo è la complessità del setup, che potrebbe anche introdurre incertezze in fase di acquisizione dei dati. Per completezza anche questa procedura è stata comunque implementata e verificata; da essa poi sono state qui sviluppate diverse metodologie, frutto di svariate osservazioni e considerazioni.

Altre tecniche "ibride" basate sugli algoritmi genetici sono presenti in letteratura. In [17] è stato implementato un metodo che utilizza i genetici per inizializzare l'algoritmo di Levenberg-Marquardt. Anche questo metodo è stato implementato e verificato (capitolo 2.4.2). In [18] invece i genetici sono stati utilizzati per migliorare i risultati dopo che il problema di calibrazione è stato risolto utilizzando le reti neurali. Queste però non sono state prese in

considerazione in questa tesi, ma comunque si pensa che possano rivestire un ruolo importante, soprattutto se ibridizzate con gli algoritmi genetici. Un possibile sviluppo di questo lavoro potrebbe appunto essere lo studio di queste tecniche e la loro applicazione al problema di calibrazione.

Si vuole quindi studiare gli algoritmi genetici per poi utilizzarli nella risoluzione del modello non lineare della telecamera; esso viene inizializzato con i parametri ricavati dalla tecnica di Zhang, poiché non si vuole perdere la semplicità del setup che la caratterizza. Dopo alcune analisi e sperimentazioni degli algoritmi presenti in letteratura, sono stati implementati diversi approcci per tentare di risolvere le problematiche associate a essi. Sono dunque state sviluppate diverse nuove tecniche, frutto di sperimentazioni e considerazioni, delle quali non è stato trovato alcun riscontro in letteratura. Esse sono sempre state confrontate alla tecnica descritta nel capitolo 2.2 basata sull'algoritmo di Zhang e Levenberg-Marquardt.

### **2.3.1 Introduzione ai GA**

Gli algoritmi genetici (GA) [19],[20],[21],[22] sono metodi adattativi che possono essere usati per risolvere problemi di ricerca e ottimizzazione. Essi si ispirano ai processi genetici degli organismi biologici. Dopo molte generazioni, le popolazioni si evolvono secondo i principi della selezione naturale e della sopravvivenza del migliore, come teorizzato per la prima volta da Charles Darwin nella sua opera "L'origine delle specie". Imitando questi processi, gli algoritmi genetici sono in grado di evolvere soluzioni per problemi del mondo reale, se sono stati codificati opportunamente. I principi di base dei GA sono stati definiti per la prima volta da John Holland tra la fine degli anni '60 e l'inizio degli anni '70 del secolo scorso: essi simulano quei processi che nelle popolazioni naturali sono essenziali per l'evoluzione. In Natura, gli individui di una popolazione competono uno con l'altro per risorse come cibo, acqua e territorio. Inoltre membri della stessa specie spesso competono per attrarre un compagno. Gli individui che hanno più successo nella sopravvivenza e nella riproduzione avranno un numero

relativamente grande di discendenti. Gli individui che si mostreranno essere meno adatti produrranno poca o nessuna prole. Questo significa che i geni degli individui più adattati (fit individuals) saranno trasmessi a un crescente numero di individui in ciascuna delle generazioni successive. La combinazione delle buone caratteristiche di diversi antenati può a volte produrre una discendenza molto adattata (superfit), la cui qualità è superiore a quella di ciascun genitore. In questo modo le specie si evolvono e diventano sempre più adattate al loro ambiente.

I GA usano una diretta analogia con il comportamento della natura. Lavorano con una popolazione di individui, ciascuno dei quali rappresenta una possibile soluzione del problema posto. A ogni individuo è associato un punteggio di adattamento "fitness score" a seconda di quanto sia buona la soluzione al problema. In natura è equivalente a stabilire quanto un individuo riesce a competere per le risorse. Gli individui migliori hanno maggiore possibilità di riprodursi producendo nuovi discendenti che condividono alcune caratteristiche di ciascun genitore. Gli individui meno adattati hanno meno probabilità di riprodursi e quindi tendono a estinguersi. Così viene a generarsi una nuova intera popolazione di possibili soluzioni, che contiene una proporzione più alta di caratteristiche possedute dagli individui buoni della precedente generazione. In questo modo dopo molte generazioni le buone caratteristiche vengono propagate a tutta la popolazione, essendo mischiate e scambiate con altre buone caratteristiche. Favorendo l'accoppiamento tra gli individui più adatti, vengono esplorate le aree più promettenti dello spazio di ricerca. Se il GA è stato costruito bene, la popolazione convergere a una soluzione ottima del problema.

La potenza degli algoritmi genetici viene dalla loro robustezza e dal fatto che possono essere usati con successo in molti campi differenti (vengono ad esempio impiegati per la pianificazione di traiettorie navali piuttosto che di robot, per progettare circuiti elettrici o sistemi informativi, per simulare dinamiche della concorrenza in economia o per stabilire strategie di controllo, ecc...). I GA non garantiscono di trovare una soluzione ottima per un problema, ma generalmente trovano una soluzione sufficientemente buona e in tempi sufficientemente rapidi. Dove esistono tecniche specializzate per risolvere particolari problemi, queste

hanno spesso prestazioni migliori dei GA sia in termini di accuratezza che di velocità. Il terreno migliore dei GA sono dunque le aree dove non esistono tecniche specializzate; dove invece esistono tecniche che funzionano bene, si possono avere miglioramenti "ibridizzandole" con i GA.

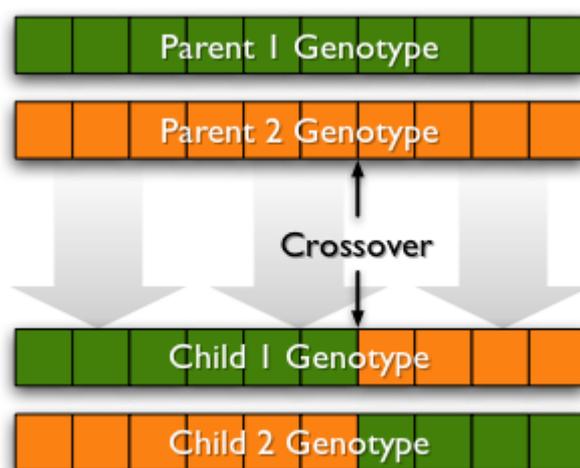
## 2.3.2 Principi base dei GA

Per ciascun problema da risolvere deve essere costruita una specifica funzione fitness. Dato un particolare cromosoma (cioè una variabile del problema), la funzione fitness restituisce un singolo valore numerico, detto "*fitness*" o una "figura di merito", che si suppone sia funzione dell'utilità o abilità dell'individuo che il cromosoma rappresenta. Nel nostro caso, avendo a che fare con un problema di ottimizzazione, è ovvio che la funzione fitness debba misurare il valore stesso della funzione. In termini genetici, l'insieme dei parametri rappresentati da un particolare cromosoma è chiamato *genotipo*. Il genotipo contiene le informazioni richieste per costruire un organismo che è chiamato *fenotipo*. Gli stessi termini sono usati nei GA. Il fitness di un individuo dipende dunque dalle performance del fenotipo.

Durante la fase di riproduzione di un GA gli individui sono selezionati tra la popolazione e ricombinati, producendo così la discendenza che sarà compresa nella generazione successiva. I genitori sono selezionati casualmente usando uno schema che favorisce gli individui migliori. Gli individui buoni saranno probabilmente selezionati più volte per la riproduzione, mentre quelli peggiori potrebbero non essere mai scelti. Avendo selezionato due individui, i loro cromosomi sono ricombinati, tipicamente usando il meccanismo del *crossover* e la *mutazione*.

Durante il *crossover* vengono selezionati due individui (genitori), e le stringhe dei loro due cromosomi vengono divise in due in una posizione scelta a caso, così da produrre due segmenti "testa" (head) e due segmenti "coda" (tail). I segmenti coda sono scambiati per produrre due nuovi cromosomi di lunghezza completa. Ciascuno dei figli eredita dunque alcuni geni da ogni genitore. Questo tipo di

crossover è conosciuto come *single point* ed è quello più comunemente utilizzato. Esso si differenzia dal *two-point*, dove invece il cromosoma viene diviso in tre segmenti differenti che vengono scambiati tra i genitori in maniera casuale. Il crossover non è abitualmente applicato a tutte le coppie di individui selezionati per l'accoppiamento. Si ha infatti la possibilità, scegliendo un valore per il parametro *elite count*, di selezionare un numero di figli che vengono generati duplicando i genitori con valore di fitness migliore. Questo dà la possibilità a ogni individuo di propagare i propri geni nella generazione successiva senza la scissione operata dal crossover. Inoltre si sceglie un parametro detto *crossover fraction* (tipicamente compreso tra 0.6 e 1) che indica la percentuale di figli che sarà generata dal crossover. Tutti i figli che non duplicano i genitori o che non sono frutto di un crossover saranno creati con una mutazione.



**Figura 2.14:** Esempio di *single-point crossover*.

Durante la *mutazione* viene alterato a caso ogni gene. La teoria tradizionale ritiene che il crossover sia più importante della mutazione per quanto riguarda la rapidità nell'esplorare lo spazio di ricerca. La mutazione introduce casualità nella ricerca e aiuta ad assicurare che nessun punto nello spazio abbia probabilità nulla di essere esaminato, cioè riduce la probabilità di incorrere in un minimo locale.

Se il GA è correttamente implementato, la popolazione evolverà in molte generazioni in modo che il fitness del miglior individuo e la media in ogni

generazione converga verso l'ottimo globale. A livello operativo possiamo riassumere l'evoluzione di un algoritmo genetico nei seguenti punti:

- generazione, in maniera casuale, di una popolazione iniziale;
- creazione di una sequenza di nuove popolazioni (generazioni). In ciascuna iterazione, gli individui della popolazione corrente sono usati per creare la generazione successiva, e a questo scopo si compiono degli ulteriori passi:
  - ciascun membro della popolazione corrente è valutato calcolandone il rispettivo valore di fitness (idoneità);
  - si determina un opportuno ordinamento di tali individui sulla base dei valori di fitness;
  - gli individui più promettenti sono selezionati come genitori;
  - a partire da tali individui si genera un pari numero di individui della generazione successiva. Alcuni genitori sono duplicati nella generazione successiva mentre gli altri figli sono generati secondo due modalità distinte, vale a dire effettuando cambiamenti casuali su un singolo genitore (mutazione) oppure combinando opportunamente le caratteristiche di una coppia di genitori (crossover);
  - gli individui così generati vanno a sostituire i genitori consentendo la formazione della generazione successiva;
- infine, l'algoritmo s'interrompe quando uno dei criteri d'arresto è soddisfatto.

Solitamente per rappresentare un cromosoma si utilizzano sequenze di cifre binarie, ma in alcuni casi, tra cui quello analizzato, può essere più comodo usare cromosomi dove ciascun simbolo rappresenta un intero o un floating point. Infatti i parametri del problema sono spesso numerici e la loro rappresentazione in numeri, anziché in cifre binarie, può essere un vantaggio. Così è possibile definire facilmente degli operatori di crossover e mutazione specifici. Tra gli operatori di combinazione più utilizzati si ricordano:

- Media: effettua la media aritmetica dei geni dei due genitori.
- Media Geometrica: effettua la radice quadrata del prodotto dei due valori.

- Extension: effettua la differenza tra due valori e la aggiunge al più alto o la sottrae al più basso.

Mentre tra gli operatori di mutazione i più comuni sono:

- Rimpiazzamento casuale: rimpiazza un valore con uno casuale.
- Creep: aggiunge o sottrae un piccolo numero generato casualmente.
- Geometric Creep: moltiplica per un valore prossimo a 1.

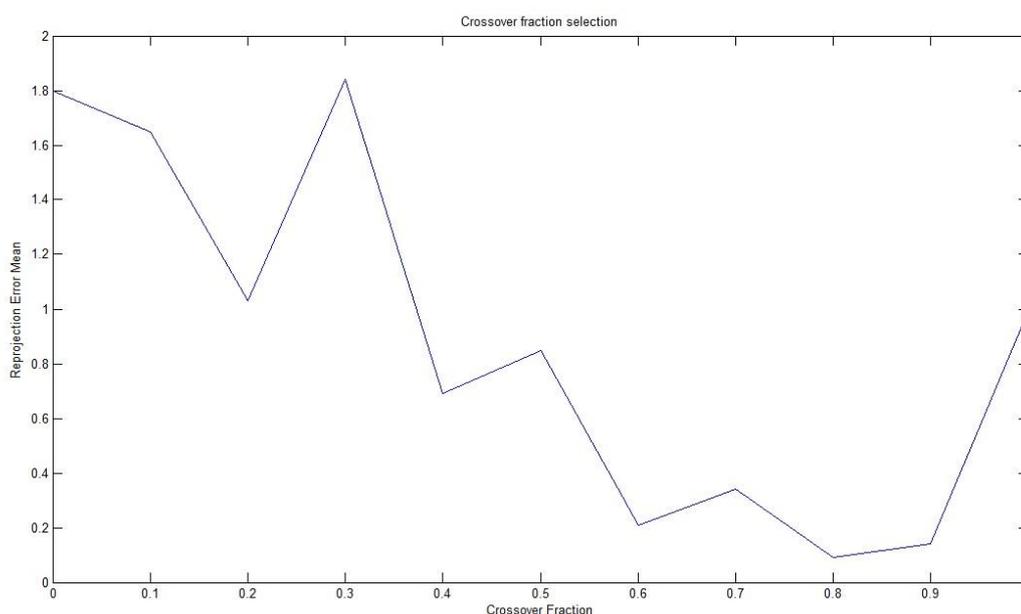
Per entrambi gli operatori creep il numero generato casualmente può avere diverse distribuzioni: uniforme dentro un dato range, esponenziale, gaussiana, binomiale etc.

### **2.3.3 Implementazione di tecniche di calibrazione basate sugli algoritmi genetici**

Anche per l'implementazione di tecniche di calibrazione basate su algoritmi genetici è stato deciso di utilizzare MatLab, data la sua potenza di calcolo, il linguaggio di programmazione ad alto livello non complesso e le capacità grafiche semplici e potenti. Grazie a queste caratteristiche MatLab risulta essere un comodo strumento per la ricerca scientifica e un potente metodo di prototipazione. Come descritto nel capitolo 2.2.4, anche l'algoritmo di Zhang e quello di Levenberg-Marquardt sono stati implementati in MatLab, e l'utilizzo dello stesso software sicuramente semplifica le procedure di scrittura dell'algoritmo e rende immediato il confronto tra le due tecniche. Inoltre MatLab contiene un'ottima libreria dedicata agli algoritmi genetici, oltre che un tool dedicato all'ottimizzazione ("MatLab Optimization Tool") in cui è possibile risolvere problemi di minimizzazione anche usando i GA. Il software di calibrazione descritto nel capitolo 2.2.4 è stato quindi ampliato. Oltre all'algoritmo di Zhang-LM, viene risolto il problema di calibrazione con l'utilizzo di diverse tecniche, descritte in seguito, basate sugli algoritmi genetici. Il software permette anche di confrontare i risultati tra i vari metodi. Anche in questo caso viene data la possibilità all'utente di scegliere quale modello di telecamera utilizzare e i dati sono creati ancora una volta grazie al programma "FeatureLinker" del VBLab.

L'algoritmo genetico implementato prende in ingresso una serie di parametri di configurazione. Prima di tutto bisogna stabilire la dimensione della popolazione, ovvero quanti individui sono contenuti in ogni generazione. Più è ampia la popolazione e minore sarà la possibilità di convergere in un minimo locale, in quanto il GA andrà a esplorare tutto lo spazio delle variabili. Ciò però ridurrà considerevolmente la velocità di esecuzione. È necessario poi passare in ingresso all'algoritmo un individuo iniziale, che verrà utilizzato per la creazione della prima generazione. L'individuo iniziale può essere costituito da una stima delle variabili di minimizzazione, oppure creato in maniera del tutto casuale. Chiaramente un set di variabili iniziali sensato può diminuire considerevolmente la velocità con la quale l'algoritmo converge al minimo globale. Per ogni cromosoma che compone l'individuo, quindi per ogni variabile, si può configurare un range in cui vengono scelti i valori di quel dato cromosoma per la creazione della popolazione iniziale. Non è necessario che tale range contenga la soluzione ottima del problema, poiché l'algoritmo genetico evolve esplorando soluzioni all'infuori di esso. Chiaramente, se la soluzione è compresa in esso, più il limite superiore e inferiore si avvicinano ad essa e più l'algoritmo ha possibilità di convergere in fretta alla soluzione ottima. Bisogna inoltre definire i criteri d'arresto dell'algoritmo; in particolare è stato deciso di terminare l'esecuzione del GA quando viene raggiunto un certo numero di generazioni massimo, che assume valori differenti a seconda delle tecniche utilizzate. Altro parametro è il *crossover fraction*, definito nel capitolo 2.3.2, che in tutte le simulazioni è stato posto a 0.8. Questo valore, come spesso accade nella scelta dei parametri degli algoritmi genetici, è frutto di sperimentazioni e tentativi. Essendo i GA algoritmi che progrediscono in maniera casuale, non è sempre facile capire quale sia la configurazione dei parametri migliore per far sì che converga più velocemente possibile alla soluzione ottima. Spesso l'unico metodo valido per capire quale sia la configurazione migliore è la sperimentazione di vari tentativi con diverse simulazioni. Una volta implementato ogni metodo la prima verifica da eseguire è dunque verificare quale sia il valore migliore del parametro crossover fraction. Per fare ciò si lanciano più esecuzioni dell'algoritmo implementato, tenendo tutti i

parametri fissi e variando solamente il crossover fraction tra esse. I GA evolvono in maniera casuale e spesso non si ottengono gli stessi valori lanciando due volte lo stesso GA. Per questo sono state fatte più prove di questo tipo, anche con differenti sottoinsiemi di immagini, e i risultati sono stati mediati per ogni valore di crossover fraction. Ad esempio in figura 2.15 sono stati rappresentati i risultati trovati durante una di queste analisi. Dal grafico si nota che un valore pari a 0.8 del parametro crossover fraction porta a risultati migliori in termini di errore di retroproiezione finale. Questo valore è stato verificato essere il migliore per tutti i metodi implementati.



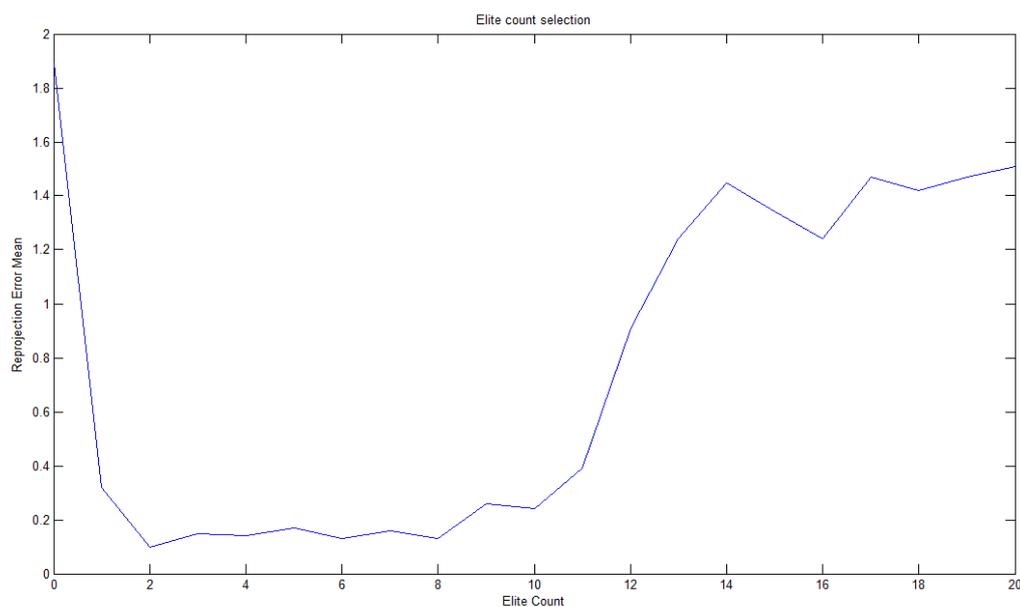
**Figura 2.15:** Risultati degli esperimenti per l'individuazione del miglior valore da assegnare al parametro crossover fraction.

Lo stesso tipo di esperimento è stato svolto per verificare quale tipo di crossover fosse il migliore. Non è stata però trovata alcuna differenza tra i risultati ricavati utilizzando un crossover a *single-point* e a *two-point* (tabella 2.1). Si è dunque deciso di utilizzare il primo.

CROSSOVER TYPE	Reprojection Error Mean
Single-point	0.117947538
Two-point	0.118354563

**Tabella 2.1:** Risultati degli esperimenti per l'individuazione del miglior tipo di crossover.

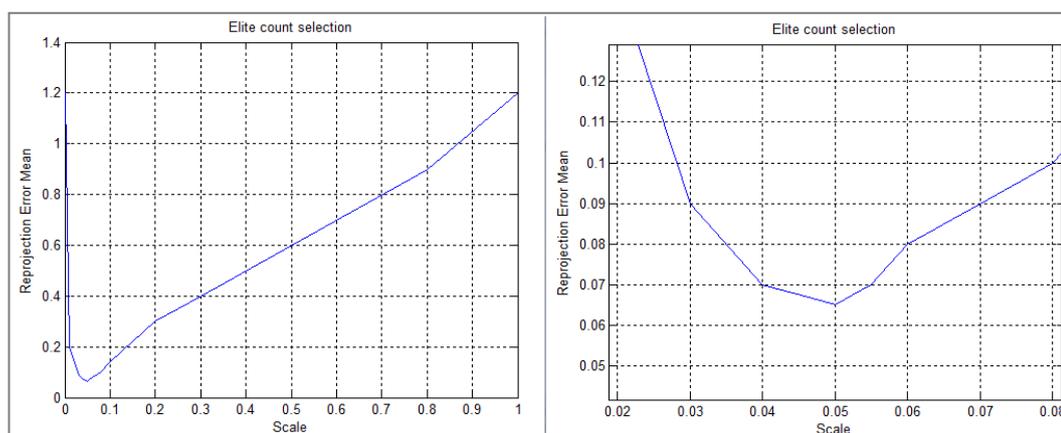
Il parametro *elite count*, ovvero il numero degli individui che vengono duplicati nella generazione seguente senza essere modificati, è stato impostato a 2 in tutte le simulazioni, poiché analisi simili a quelle effettuate per trovare il miglior valore di *crossover fraction* hanno mostrato che valori tra 2 e 8 portano nella maggior parte dei casi a risultati migliori (figura 2.16).



**Figura 2.16:** Risultati degli esperimenti per l'individuazione del miglior valore di *elite count*.

Infine le mutazioni sono state eseguite selezionando in modo casuale un valore per ogni cromosoma, scelto in una distribuzione gaussiana che ha come valore medio il valore della variabile nella generazione precedente e deviazione standard definita da un parametro chiamato *Scale*. Inoltre l'ampiezza delle mutazioni in ogni generazione è definito dal parametro *Shrink*. Infatti la deviazione standard della distribuzione gaussiana diminuisce linearmente in ogni generazione in modo tale da avere alla fine della simulazione un valore pari a  $1 - \text{Shrink}$ . Quindi, se per esempio il valore di *Shrink* è settato pari a 1, l'ammontare delle mutazioni decresce in ogni generazione fino ad arrivare a 0 nell'ultima generazione. Questo parametro è molto utile per consentire all' algoritmo di convergere più velocemente all'ottimo. Infatti dopo numerose generazioni si può pensare che gli individui della popolazione assumano valori vicini alla soluzione finale ed è

quindi inutile compiere mutazioni eccessive che creerebbero individui con fitness più alto. Al contrario se non si eseguissero mutazioni l'algoritmo non esplorerebbe tutto lo spazio dei parametri e quindi rischierebbe di convergere in un minimo locale. Anche in questo caso per ogni tecnica implementata sono stati eseguiti numerosi esperimenti e tentativi per stabilire quale fosse il miglior valore da assegnare ai parametri *scale* e *shrink*. In tutte queste prove si è evidenziato che mutazioni con una distribuzione gaussiana di deviazione standard 0.005 (*scale*) e *shrink* 1 portano ai risultati migliori. A testimonianza di ciò si riportano nella figura 2.17 alcuni dei risultati trovati dagli esperimenti per trovare il miglior valore di deviazione standard, con il parametro di *shrink* fisso a 1. Con errore di retroproiezione medio si intende il valore medio risultante da differenti prove con diversi set di immagini. Si nota che si hanno valori più bassi di errore di retroproiezione medio con *scale* intorno a 0.05.



**Figura 2.17:** Risultati degli esperimenti per l'individuazione del miglior valore di *scale*, con *shrink* fisso a 1.

## 2.4 Confronto tra differenti metodologie

Si analizzano i risultati di diverse calibrazioni per poter confrontare il metodo Zhang-LM, descritto nel capitolo 2.2, con varie tecniche basate sugli algoritmi genetici. Lo scopo di tale confronto è la comprensione dei limiti e dei vantaggi di un approccio al problema della calibrazione basato sui GA, al fine di verificare se esistono soluzioni alternative al metodo Zhang-LM, attualmente utilizzato in laboratorio, altrettanto robuste e accurate. Partendo dagli algoritmi presenti in letteratura sono state effettuate varie considerazioni, che hanno portato all'implementazione di diversi metodi di calibrazione basati sui GA. Questi si possono classificare come segue:

- *tecnica di Yang*, dalla letteratura [14];
- *tecniche di bundle adjustment*, cioè tecniche basate sulla minimizzazione dell'errore di retroproiezione;
- *tecniche basate sulla triangolazione*, cioè nel quale la funzione di fitness è l'errore di ricostruzione tridimensionale.

Ognuna di queste tecniche e le considerazioni che hanno portato ad esse sono spiegate nel seguito del capitolo. Inoltre è stato aggiunto al confronto un tool di calibrazione contenuto in Halcon, una libreria commerciale per la visione artificiale.

Come modello di telecamera si utilizza sempre il modello *pinhole camera*. La parola “modello” nel seguito non indicherà i differenti modelli di telecamera (*pinhole camera*, *affine camera*, etc...), bensì i vari tipi di modelli *pinhole camera* a diverso numero di parametri (ad esempio lineare a 4 parametri o non lineare a 8 parametri, etc...). Per confrontare i vari metodi si è deciso di utilizzare un numero di parametri del modello della telecamera comune a tutte le prove. In particolare si è deciso di utilizzare un modello che fosse appropriato al caso considerato e il più semplice possibile. Infatti, avendo svolto prove con un'ottica compresa tra 8 e 10mm, che quindi presenta distorsioni non trascurabili, sono stati considerati i parametri relativi alle prime due distorsioni radiali e alla prima distorsione tangenziale, trascurando però la terza distorsione radiale e la seconda tangenziale,

che sono in questo caso trascurabili. Dal datasheet della telecamera si può verificare che i pixel sono quadrati e quindi si è imposto il parametro di skew nullo e le lunghezze focali nelle due direzioni del sensore uguali. Ne risulta che il modello complessivo considerato è composto da tre parametri intrinseci (coordinate del centro ottico  $u_0$  e  $v_0$  e focale  $f$ ) e tre parametri di distorsione (due radiali  $K_1$  e  $K_2$  e uno tangenziale  $P_1$ ).

Per eseguire le prove è stata utilizzata una telecamera Dalsa GENIE serie hm1400, con ottica modificabile tra 8 e 48 mm. Le dimensioni dei pixel sono  $7.4 \times 7.4 \mu m$  e la risoluzione  $1400 \times 1024 \text{ pixel}$ . L'ottica è tenuta fissa 13 mm. Dai risultati ci si aspetta quindi che le coordinate del centro ottico siano vicine al centro del sensore  $u_0 \simeq 700 \text{ pixel}$  e  $v_0 \simeq 512 \text{ pixel}$ . Inoltre deve risultare che  $f_{pix} \simeq \frac{13 \text{ mm}}{7.4 \mu m} = 1756.75$ .

### 2.4.1 Zhang-LM

Ogni tecnica implementata viene confrontata con l'algoritmo di Zhang-LM, ampiamente descritto e commentato al capitolo 2.2. Si vogliono qui riassumere velocemente i pregi e i difetti di questo approccio, che saranno poi alla base delle considerazioni ottenute dal confronto con le varie tecniche che utilizzano gli algoritmi genetici e con la tecnica di HALCON. Per prima cosa l'algoritmo di Zhang permette di calibrare senza dover conoscere la posizione del calibratore, poiché il sistema di riferimento mondo è solidale con esso. Questo semplifica notevolmente la procedura di setup di calibrazione, complicando però il problema, poiché i parametri estrinseci variano per ogni immagine acquisita. Si ricorda che bisogna prestare molta attenzione alla fase di acquisizione delle immagini, poiché è necessario che il calibratore venga posto in pose più differenziate possibile, coprendo tutto il sensore. Il metodo di Zhang richiede almeno l'uso di 3 immagini, anche se si con LM si ottengono risultati accurati, specialmente se il modello utilizzato è complesso, con almeno 15 immagini. L'algoritmo di Levenberg-Marquardt converge velocemente a una soluzione ottima solo se inizializzato correttamente, quindi solo se i parametri trovati con la tecnica di

Zhang sono vicini alla soluzione ottima. In questo caso l'algoritmo raggiunge accuratezze elevate. Infine è stato deciso di utilizzare un metodo manuale per trovare le correlazioni fra i punti nelle immagini, utilizzando il software "Feature Linker". In questo modo si può ottenere un riscontro visivo immediato dei risultati di correlazione, limitando gli errori che possono accadere con procedure automatiche. Chiaramente però un approccio manuale è più lungo e laborioso. Nelle seguenti tabelle sono riportati i risultati di calibrazione trovati con il metodo di Zhang lineare e il metodo di Zhang ottimizzato con Levenberg-Marquardt. In entrambi i casi il modello di telecamera utilizzato è quello a 6 parametri ( $u_0$ ,  $v_0$ ,  $f$ ,  $K_1$ ,  $K_2$ ,  $P_1$ ) e il dispositivo ha risoluzione 1400x1024 pixel, dimensione dei pixel  $7.4 \mu m$ , e viene utilizzata un'ottica da 13 mm. Sono state utilizzate 16 immagini per la calibrazione. Si nota come ancora una volta i risultati di Zhang-LM ottimizzano quelli di Zhang, restituendo un errore di retroproiezione notevolmente inferiore.

<b>ZHANG LINEARE</b>	
<b>Errore di retroproiezione [pixel]</b>	1.27
<b>Tempo impiegato</b>	0 sec
<b>Focale (f)</b>	1733.1505
<b>Coordinate centro ottico (<math>u_0 \times v_0</math>) [pixel]</b>	685.45142 x 532.98135
<b>Distorsione K1</b> $\left[ \frac{1}{m^2} \right]$	0
<b>Distorsione K2</b> $\left[ \frac{1}{m^4} \right]$	0
<b>Distorsione P1</b> $\left[ \frac{1}{m^2} \right]$	0

**Tabella 2.2:** Risultati di calibrazione con il metodo di Zhang lineare, 16 immagini, modello a 6 parametri ( $u_0$ ,  $v_0$ ,  $f$ ,  $K_1$ ,  $K_2$ ,  $P_1$ ), dimensione dei pixel  $7.4 \mu m$ , risoluzione 1400x1024 pixel e ottica da 13 mm.

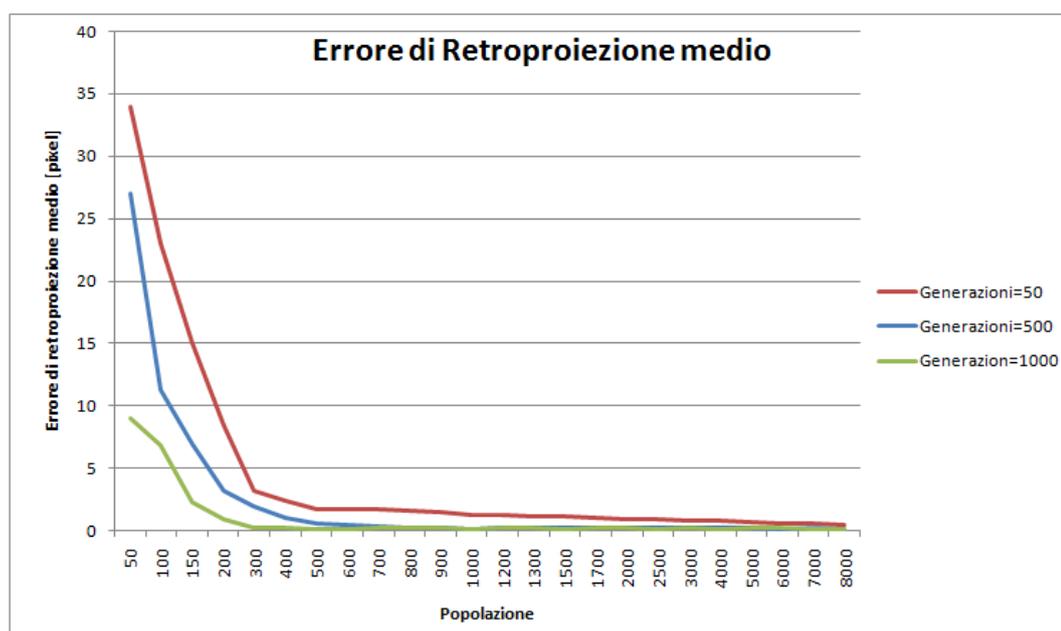
<b>ZHANG -LM</b>	
<b>Errore di retroproiezione [pixel]</b>	0.0663
<b>Tempo impiegato</b>	2 sec
<b>Focale (f)</b>	1752.5306
<b>Coordinate centro ottico (<math>u_0 \times v_0</math>) [pixel]</b>	698.79611 x 527.04831
<b>Distorsione K1</b> $\left[ \frac{1}{m^2} \right]$	-0.12174
<b>Distorsione K2</b> $\left[ \frac{1}{m^4} \right]$	2.4538
<b>Distorsione P1</b> $\left[ \frac{1}{m^2} \right]$	0.00402

**Tabella 2.3:** Risultati di calibrazione con il metodo di Zhang-LM non lineare, 16 immagini, modello a 6 parametri ( $u_0$ ,  $v_0$ ,  $f$ ,  $K1$ ,  $K2$ ,  $P1$ ), dimensione dei pixel  $7.4 \mu m$ , risoluzione  $1400 \times 1024$  pixel e ottica da 13 mm.

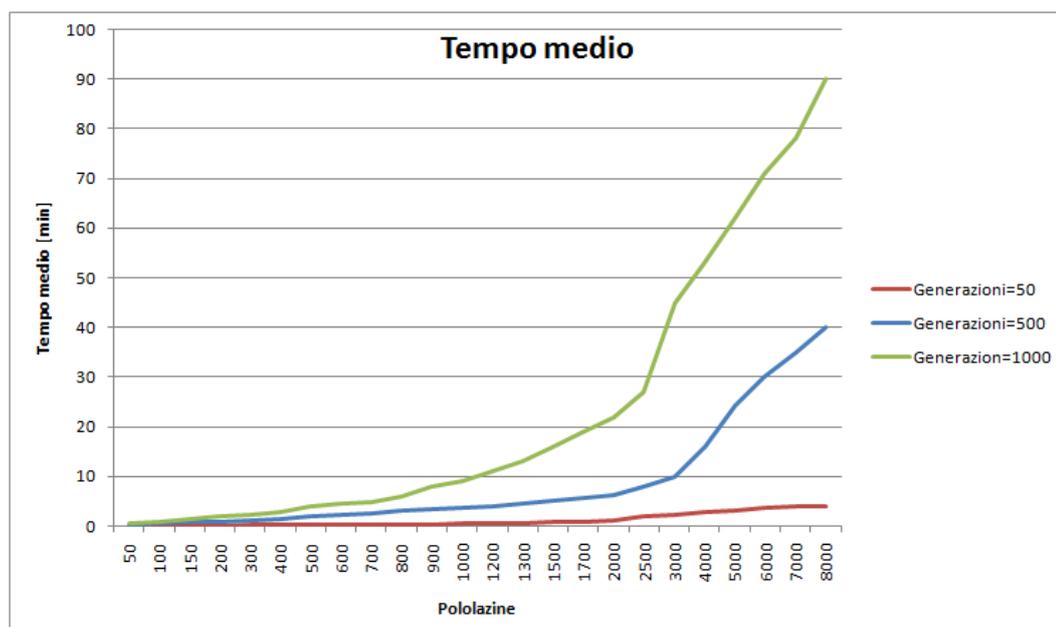
## 2.4.2 Tecnica di Yang

Per implementare il problema della calibrazione con algoritmi genetici è stato preso in considerazione, prima di tutto, un approccio basato sulle varie metodologie che si trovano in letteratura. In particolare è stata implementato l'algoritmo descritto in [14]. Questa tecnica si basa sulla conoscenza delle coordinate di punti noti nello spazio (e quindi la necessità di misurare a monte le coordinate tridimensionali di essi) e delle loro correlazioni con i rispettivi punti nell'immagine. Da queste corrispondenze viene calcolata la funzione di costo, che permette di minimizzare l'errore di retroproiezione. Essa viene quindi ottimizzata usando un algoritmo genetico. In questa procedura dunque è necessario misurare la posizione del calibratore rispetto alla telecamera. Il metodo di Yang consente, se i punti noti sul calibratore sono tanti, di risolvere il problema della calibrazione anche utilizzando una sola immagine. Per prima cosa è stata dunque acquisita una sola immagine del calibratore posto in una posizione nota. I risultati ottenuti con questo metodo sono mostrati nella tabella 2.4. In particolare essi sono stati ottenuti una popolazione di 8000 individui, per 500 generazioni, iniziando le variabili a valori sensati contenuti in range di valori plausibili. Come si nota nella

figura 2.18, che mostra l'andamento dell'errore di proiezione in funzione delle dimensioni della popolazione e del numero di generazioni, i risultati migliori si ottengono con una vasta popolazione e un numero di generazioni elevato. Ciò comporta che per avere risultati soddisfacenti l'algoritmo impiega molto tempo (figura 2.19). Nella scelta delle dimensioni della popolazione e del numero di generazioni si predilige una maggiore accuratezza rispetto a un tempo di esecuzione basso.



**Figura 2.18:** Grafico che descrive l'errore di retroproiezione di un GA in funzione delle dimensioni della popolazione e al variare del numero di generazioni. Risultati ottenuti mediando i valori calcolati in diverse prove utilizzando la tecnica di Yang e l'immagine.



**Figura 2.19:** Grafico che descrive l'andamento del tempo di esecuzione di un GA in funzione delle dimensioni della popolazione e al variare del numero di generazioni. Risultati ottenuti mediando i valori calcolati in diverse prove utilizzando la tecnica di Yang e 1 immagine.

Come si vede nella tabella 2.4 l'errore di retroproiezione risultante dalla tecnica di Yang è più alto rispetto a quello trovato in Zhang-LM; inoltre il tempo di esecuzione di questo metodo è di molto superiore. Quindi questo approccio sarebbe già da scartare, poiché se paragonato a Zhang-LM presenta numerosi svantaggi. È chiaro che in questo caso il confronto tra GA e Zhang-LM non è stato fatto con lo stesso set di immagini iniziali poiché l'algoritmo di Zhang richiede in ingresso i dati trovati con almeno tre immagini acquisite e con il calibratore posizionato in pose il più varie possibile. Solo in questo caso quindi le immagini acquisite per calibrare nei due metodi sono differenti.

Nel tentativo di migliorare l'accuratezza di tale metodo, si è provato a eseguire la tecnica di Yang acquisendo dieci immagini differenti. Per rendere più veloce il setup, il calibratore è stato posizionato parallelamente al piano del sensore, a una distanza nota dalla telecamera, su una slitta che gli permettesse di traslare solamente lungo l'asse Z, ovvero lungo l'asse principale del dispositivo (perpendicolare al sensore). Sono state poi acquisite 10 immagini traslando in ognuna il calibratore di una distanza nota lungo Z. In questo modo le coordinate

nello spazio dei punti noti su X e Y non variano in ogni immagine e vanno misurate soltanto nella prima acquisizione. Mentre le coordinate in Z sono date dalla misura effettuata alla prima acquisizione più lo spostamento noto del calibratore. I risultati di tale tecnica sono rappresentati in tabella 2.5. Questa tecnica presenta un tempo di esecuzione di molto superiore rispetto le altre; ciò la rende di per sé inutilizzabile. In più si nota che l'errore di retroproiezione è aumentato e dunque sembrerebbe che la soluzione trovata sia meno accurata. In realtà i parametri relativi alla lunghezza focale e alla posizione del centro ottico sono più vicini alla soluzione che ci aspetteremmo di trovare rispetto ai risultati trovati con una sola immagine. Questo perché la tecnica con più immagini cerca di minimizzare lo stesso spazio dei parametri con un numero maggiore di vincoli, dati da un numero maggiore di corrispondenze di punti noti. Quindi l'algoritmo ha maggiori difficoltà a trovare una soluzione con errore di retroproiezione molto basso, trovando una soluzione che si addica di più a tutte le immagini. Se utilizzata con una sola immagine invece questa tecnica converge a una soluzione che si addice solamente all'immagine acquisita; essa ha un errore di retroproiezione finale minore, ma i parametri trovati valgono solamente per quell'immagine. Quindi si potrebbe dire che nonostante la soluzione con 10 immagini presenta un errore di retroproiezione maggiore, essa è più robusta.

In generale dunque se confrontato con Zhang-LM questo metodo risulta più lento e meno accurato in termini di errore di retroproiezione. Da sottolineare però il fatto che l'algoritmo di Yang non necessita di inizializzazioni dei parametri con valori sensati, sfruttando a pieno le caratteristiche che differenziano i GA dalle altre tecniche di ottimizzazione. Infatti una delle caratteristiche principali dei GA è proprio che essi convergono al minimo globale anche se le variabili non sono inizializzate con valori vicini alla soluzione ottima. Questo fattore però fa sì che per convergere a una buona soluzione sia necessario aumentare molto la dimensione della popolazione e il numero delle generazioni, per consentire all'algoritmo di esplorare uno spazio delle variabili molto vasto e non convergere a un minimo locale. Questo comporta inevitabilmente un aumento spropositato del tempo necessario all'algoritmo per convergere alla soluzione finale, come è

stato mostrato in tabella 2.4 e 2.5. Infatti se si vuole che l'algoritmo di Yang raggiunga prestazioni vicine a quelle di Zhang-LM è necessario che esso venga eseguito per circa 1 ora (1 immagine, popolazione di 8000 individui per 500 generazioni). Probabilmente aumentando considerevolmente il numero di generazioni e la popolazione si potrebbe ottenere una soluzione ancor più accurata, ottenendo però un algoritmo che deve essere eseguito per un tempo eccessivamente alto e quindi di poco interesse per questo lavoro. È chiaro che, nonostante il processo di calibrazione non presenti specifiche di tempo molto restrittive, la velocità di esecuzione di tale algoritmo lo rende inutilizzabile, soprattutto considerando che la tecnica di Zhang-LM ha raggiunto una soluzione di migliore in un tempo notevolmente inferiore (qualche secondo di esecuzione).

<i>GA (Tecnica di Yang – 1 immagine)</i>	
<b>Errore di retroproiezione [pixel]</b>	0.17
<b>Tempo impiegato</b>	1 h
<b>Focale (f)</b>	16867.41
<b>Coordinate centro ottico (<math>u_0 \times v_0</math>) [pixel]</b>	680.59 x 496.83
<b>Distorsione K1</b> $\left[ \frac{1}{m^2} \right]$	-0.0764
<b>Distorsione K2</b> $\left[ \frac{1}{m^4} \right]$	0.26
<b>Distorsione P1</b> $\left[ \frac{1}{m^2} \right]$	0.0086

**Tabella 2.4:** Risultati di calibrazione con il metodo di Yang, 1 immagine, modello a 6 parametri ( $u_0$ ,  $v_0$ ,  $f$ ,  $K1$ ,  $K2$ ,  $P1$ ), dimensione dei pixel  $7.4 \mu m$ , risoluzione  $1400 \times 1024$  pixel, ottica da 13 mm, popolazione di 8000 individui per 500 generazioni.

<b>GA (Tecnica di Yang – 10 immagini)</b>	
<b>Errore di retroproiezione [pixel]</b>	1.98
<b>Tempo impiegato</b>	5 h
<b>Focale (f)</b>	1777.552
<b>Coordinate centro ottico (<math>u_0 \times v_0</math>) [pixel]</b>	678.47 x 478.36
<b>Distorsione K1</b> $\left[ \frac{1}{m^2} \right]$	-0.3136
<b>Distorsione K2</b> $\left[ \frac{1}{m^4} \right]$	-0.3501
<b>Distorsione P1</b> $\left[ \frac{1}{m^2} \right]$	-0.014

**Tabella 2.5:** Risultati di calibrazione con il metodo di Yang, 10 immagini, modello a 6 parametri ( $u_0$ ,  $v_0$ ,  $f$ ,  $K1$ ,  $K2$ ,  $P1$ ), dimensione dei pixel  $7.4 \mu m$ , risoluzione  $1400 \times 1024$  pixel, ottica da 13 mm, popolazione di 8000 individui per 500 generazioni.

Per risolvere questo problema, prendendo spunto da [17], questa tecnica è stata raffinata ottimizzando i risultati del GA con l'algoritmo di Levenberg-Marquardt. In questo modo è possibile implementare un GA più veloce rispetto al caso precedente (con popolazione e numero delle generazioni inferiori al caso precedente), che partendo da valori dei parametri casuali, converge a una soluzione che non è quella ottima ma si avvicina a essa. I risultati di questo sono poi ottimizzati con LM per trovare la soluzione ottima. Questa tecnica permette di utilizzare il metodo di Yang, diminuendo il tempo di esecuzione dell'algoritmo genetico e aumentandone l'accuratezza. I risultati sono mostrati in tabella 2.6, e sono stati trovati eseguendo il GA con una popolazione di 500 individui per 100 generazioni, utilizzando 10 immagini. È stato verificato che in questo caso l'utilizzo di una sola immagine porta a risultati peggiori poiché LM necessita di un elevato numero di punti per convergere all'ottimo globale. Dai risultati si nota che l'errore di retroproiezione è migliorato rispetto al caso precedente ed è uguale a quello trovato con Zhang-LM. Il tempo di esecuzione in questo caso assume valori accettabili, anche se ancora maggiori di quelli di Zhang-LM. Inoltre rimangono ancora alcune problematiche. Prima di tutto LM richiede che la soluzione iniziale sia vicina a quella ottima per convergere a essa. Diminuendo la dimensione delle popolazioni nel GA, si aumentano le probabilità di convergere

ad un ottimo locale che non si trova vicino alla soluzione globale. In questo caso si inizierebbe in modo sbagliato LM, trovando una soluzione di calibrazione del tutto errata. Inoltre questa tecnica, come quella di Yang, non sfrutta una delle caratteristiche principali del metodo di Zhang, ovvero la possibilità di utilizzare un setup di calibrazione veloce ed economico, senza la necessità di misurare a priori la posizione del calibratore. È evidente che questa caratteristica consista in un vantaggio enorme; si pensi ad esempio a misurazioni in galleria del vento, dove muovere un calibratore nello spazio casualmente è molto meno complicato che costruire una struttura che consenta di misurare la posizione e la posa del calibratore a ogni acquisizione. Inoltre se le misurazioni della posizione del calibratore non sono fatte con adeguata accuratezza si rischia di introdurre incertezze nel problema di calibrazione. Riassumendo si può concludere che la tecnica di Yang necessita di un'ottimizzazione tramite algoritmo di Levenberg-Marquardt per convergere a una soluzione ottima in tempi ragionevoli. Comunque la tecnica di Zhang-LM raggiunge gli stessi risultati in tempi inferiori, presentando anche vantaggi in termini di facilità del setup di calibrazione.

<i>GA (Yang – LM)</i>	
<b>Errore di retroproiezione [pixel]</b>	0.0604
<b>Tempo impiegato</b>	4 min
<b>Focale (f)</b>	16837.571
<b>Coordinate centro ottico (<math>u_0 \times v_0</math>) [pixel]</b>	689.87 x 530.43
<b>Distorsione K1</b> $\left[ \frac{1}{m^2} \right]$	0
<b>Distorsione K2</b> $\left[ \frac{1}{m^4} \right]$	-0.0586
<b>Distorsione P1</b> $\left[ \frac{1}{m^2} \right]$	-0.0061

**Tabella 2.6:** Risultati di calibrazione con il metodo di Yang ottimizzato con LM, 10 immagini, modello a 6 parametri ( $u_0$ ,  $v_0$ ,  $f$ ,  $K1$ ,  $K2$ ,  $P1$ ), dimensione dei pixel  $7.4 \mu m$ , risoluzione  $1400 \times 1024$  pixel, ottica da 13 mm, popolazione di 500 individui per 100 generazioni.

### 2.4.3 Tecniche di bundle adjustment

Per risolvere le problematiche che caratterizzano la tecnica di Yang, ovvero setup complesso e tempi lunghi di esecuzione dell'algoritmo, si è pensato di inizializzare l'algoritmo genetico con la tecnica di Zhang. Questa è quindi utilizzata per risolvere il modello lineare della telecamera e i risultati di essa sono poi ottimizzati da un algoritmo genetico. Questa procedura consente di utilizzare un setup di calibrazione veloce ed economico, senza la necessità di misurare a priori la posizione del calibratore. Inoltre un'inizializzazione del GA dovrebbe portare enormi benefici in termini di velocità di convergenza dei risultati. Per l'implementazione di questa tecnica bisogna ancora una volta scegliere quale funzione di costo minimizzare, vale a dire scegliere la funzione di fitness. Anche in questo caso è stato deciso di utilizzare l'errore di retroproiezione, cioè la media delle distanze tra i punti individuati nelle immagini (i dati di partenza) e i punti del calibratore trasformati sulla base dei parametri di telecamera, ovvero retroproiettati. Questa funzione è già stata implementata per l'algoritmo di Levenberg-Marquardt e la tecnica di Yang, e le motivazioni che hanno dettato questa scelta sono descritte nel capitolo 2.2.3. A queste si può aggiungere che utilizzando la stessa funzione obiettivo per le due tecniche il confronto risulta molto più semplice ed immediato.

È bene ricordare che nell'algoritmo di Zhang i parametri estrinseci della telecamera sono riferiti ad un sistema di riferimento solidale con il calibratore e dunque variano per ogni immagine acquisita. I parametri intrinseci invece sono invece comuni a tutte le immagini. Un altro problema da affrontare è dunque la scelta del set di parametri più indicativo per la risoluzione del problema mediante algoritmo genetico. È chiaro che se si sceglie di ottimizzare tutti i parametri del modello, lo spazio dei parametri diviene enorme e l'algoritmo risulta molto lento. Se si pensa, infatti, a una calibrazione in cui si utilizzano  $n$  immagini,  $i$  parametri intrinseci,  $e$  parametri estrinseci e  $d$  distorsioni, il numero totale dei parametri  $P$  che l'algoritmo deve ottimizzare risulta:

$$P = i + n \cdot e + d \quad (2.35)$$

Il modello completo della telecamera è composto da 6 parametri estrinseci (3 di rotazione e 3 di traslazione), 6 parametri intrinseci, 5 distorsioni; considerando che di solito per calibrare vengono acquisite al massimo 30 immagini, è immediato calcolare che il numero totale di parametri che il GA deve ottimizzare è, nel peggiore dei casi:  $P = 6 + 30 \cdot 6 + 5 = 191$ .

Sapendo che la velocità computazionale degli algoritmi genetici decresce esponenzialmente con l'aumentare del numero di parametri, una parametrizzazione di questo genere renderebbe l'algoritmo eccessivamente lento.

Un primo tentativo è stato dunque fatto mantenendo i parametri estrinseci fissi nel GA, calcolati con il metodo di Zhang. Le uniche variabili rimangono dunque i parametri intrinseci e le distorsioni. Questo ha permesso di velocizzare molto l'algoritmo ma, andando a confrontare i risultati (tabella 2.7) con quelli calcolati col metodo Zhang-LM, si nota che l'errore di retroproiezione è rimasto molto più alto e che i parametri si sono allontanati pochissimo da quelli lineari calcolati con Zhang. Tali risultati sono stati calcolati con 16 immagini e una popolazione di 800 individui per 100 generazioni. È stato verificato che anche aumentando ulteriormente il numero di generazioni o le dimensioni della popolazione il GA converge sempre alla stessa soluzione, che quindi è la soluzione migliore che può raggiungere.

<i>GA (Bouundle Adjustment – parametri estrinseci fissi da Zhang)</i>	
<b>Errore di retroproiezione [pixel]</b>	0.792
<b>Tempo impiegato</b>	7 min
<b>Focale (f)</b>	1733.425
<b>Coordinate centro ottico (<math>u_0 \times v_0</math>) [pixel]</b>	685.637 x 532.43
<b>Distorsione K1</b> $\left[ \frac{1}{m^2} \right]$	0.0133
<b>Distorsione K2</b> $\left[ \frac{1}{m^4} \right]$	1.1260
<b>Distorsione P1</b> $\left[ \frac{1}{m^2} \right]$	0.004

**Tabella 2.7:** Risultati di calibrazione con il metodo di Bouundle Adjustment, 16 immagini, modello a 6 parametri ( $u_0$ ,  $v_0$ ,  $f$ ,  $K1$ ,  $K2$ ,  $P1$ ), dimensione dei pixel  $7.4 \mu m$ , risoluzione  $1400 \times 1024$  pixel, ottica da 13 mm, popolazione di 800 individui per 100 generazioni.

PARAMETRI ESTRINSECI		
	ZHANG	ZHANG-LM
IMMAGINE 1		
	$\alpha$ 0,0026	0,0116
	$\beta$ 0,1652	0,1598
	$\gamma$ 0,0141	0,0137
	x -51,1003	-59,4062
	y -119,1614	-115,4484
	z 1083,8549	1093,0107
IMMAGINE 2		
	$\alpha$ 0,00741	0,013110
	$\beta$ 0,6759	0,66039
	$\gamma$ 0,02757	0,02629
	x 80,4191	72,0139
	y -119,50456	-116,1376
	z 1096,1207	1107,8411
IMMAGINE 3		
	$\alpha$ -0,000438	0,0018
	$\beta$ -0,309123	-0,28576
	$\gamma$ -0,000673	0,000149
	x -229,82862	-239,258
	y -114,04446	-111,75586
	z 966,7392	985,4939
...	...	...

**Tabella 2.8:** Parametri estrinseci relativi alle prime tre immagini, trovati col metodo di Zhang e con il metodo di Zhang-LM.  $\alpha$ ,  $\beta$ ,  $\gamma$  sono gli angoli di posa rappresentati in angoli di Eulero ed espressi in radianti, mentre  $x$ ,  $y$  e  $z$  sono le traslazioni lungo i rispettivi assi espresse in millimetri. Si noti come vi sono differenze considerevoli tra i risultati dei due metodi, specialmente nei parametri di traslazione che variano fino a 20 mm.

Analizzando più a fondo le motivazioni di questo consistente errore, ci si rende conto che nella maggior parte dei casi i parametri estrinseci calcolati con la tecnica di Zhang sono considerevolmente diversi da quelli trovati ottimizzando con LM. I parametri estrinseci trovati con queste due tecniche sono rappresentati nella tabella 2.8. Da questa si nota come, soprattutto nei parametri estrinseci di traslazione, vi sono differenze sostanziali nei risultati delle due differenti tecniche. L'algoritmo genetico allora non riesce a convergere alla soluzione ottima trovata con Zhang-LM, poiché converge a una soluzione che è ottima solamente per i parametri estrinseci fissi di Zhang, che sono sbagliati, e dunque a una soluzione molto meno accurata. In pratica si sta ottimizzando con un GA una funzione di costo che ha i parametri estrinseci bloccati a un valore errato. Per verificare tali considerazioni si è provato a fissare gli estrinseci ai valori calcolati con LM. In questa condizione l'algoritmo genetico converge alla stessa soluzione di LM, con lo stesso valore di errore di retroproiezione (tabella 2.9). Questa prova non ha alcun senso ai fini dell'implementazione di tecniche di calibrazione, ma è comunque un'ulteriore conferma al fatto che l'algoritmo genetico qui descritto, se implementato in modo corretto, converge alla stessa soluzione di LM.

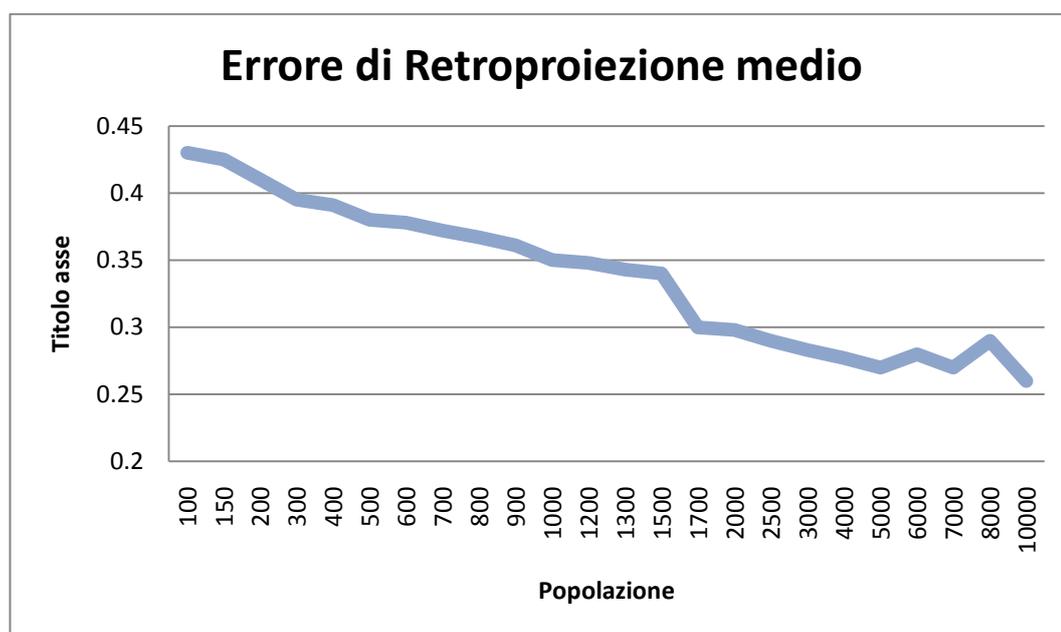
<i>GA (Bouundle Adjustment – parametri estrinseci fissi da LM)</i>	
<b>Errore di retroproiezione [pixel]</b>	0.0664
<b>Tempo impiegato [h]</b>	6 min
<b>Focale (f)</b>	1752.53
<b>Coordinate centro ottico (<math>u_0 \times v_0</math>) [pixel]</b>	698.796 x 527.048
<b>Distorsione K1 <math>\left[\frac{1}{m^2}\right]</math></b>	-0.1198
<b>Distorsione K2 <math>\left[\frac{1}{m^4}\right]</math></b>	2.4262
<b>Distorsione P1 <math>\left[\frac{1}{m^2}\right]</math></b>	0.004

**Tabella 2.9:** Risultati di calibrazione con il metodo di Bouundle Adjustment con parametri estrinseci fissi e calcolati, 16 immagini, modello a 6 parametri ( $u_0$ ,  $v_0$ ,  $f$ ,  $K1$ ,  $K2$ ,  $P1$ ), dimensione dei pixel  $7.4 \mu m$ , risoluzione  $1400 \times 1024$  pixel, ottica da 13 mm, popolazione di 800 individui per 100 generazioni.

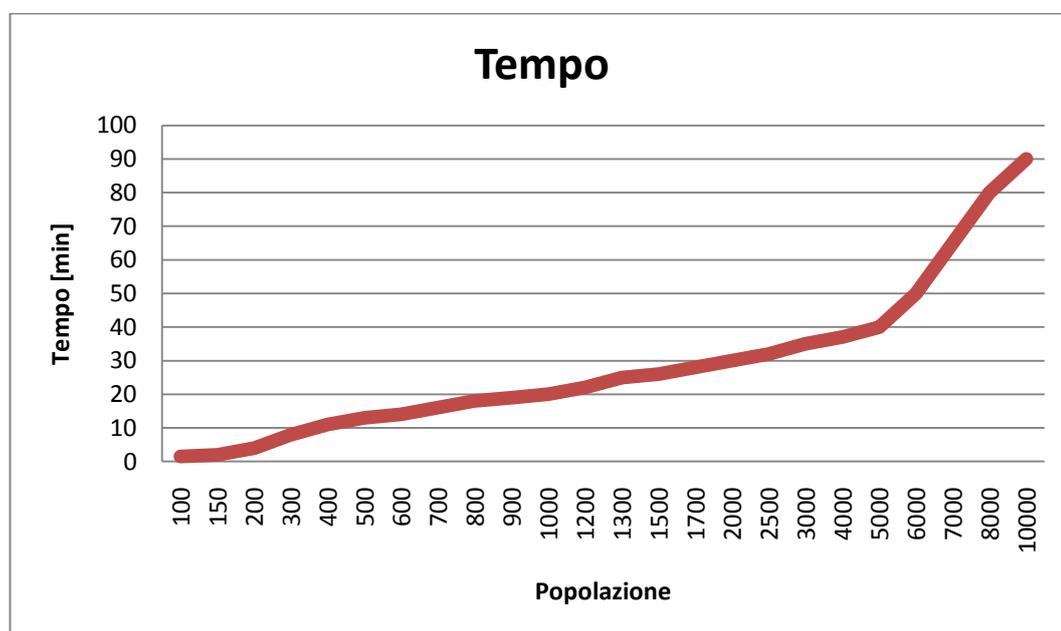
Per provare a risolvere il problema di parametrizzazione del GA il passo successivo è stato quello di lasciare anche i parametri estrinseci variabili, aumentando considerevolmente la complessità computazionale dell'algoritmo e di conseguenza i tempi d'esecuzione. Così facendo, anche in questo caso, come si vede dalla tabella 2.10, si è arrivati ad una soluzione molto vicina a quella di LM, ma ancora una volta il tempo di esecuzione dell'algoritmo necessario per fare ciò è stato circa 13 ore. Come nel caso della tecnica di Yang, anche con questo metodo l'errore di retroproiezione è inferiore all'aumentare delle dimensioni della popolazione e il numero delle generazioni, mentre il tempo che impiega l'algoritmo aumenta considerevolmente. Ad esempio in figura 2.20 e 2.21 è rappresentato l'andamento dell'errore di retroproiezione e del tempo in funzione del numero di individui che compongono la popolazione, con numero di generazioni fissa a 100. Si nota come aumentando le dimensioni della popolazione si ottiene una soluzione sempre più vicina a quella trovata con l'algoritmo di Levenberg-Marquardt, aumentando notevolmente il tempo computazionale.

<i>GA (Boundle Adjustment – parametri estrinseci liberi)</i>	
<b>Errore di retroproiezione [pixel]</b>	0.0968
<b>Tempo impiegato [h]</b>	13 h
<b>Focale (f)</b>	1737.6212
<b>Coordinate centro ottico (<math>u_0 \times v_0</math>) [pixel]</b>	685.92 x 532.84
<b>Distorsione K1 <math>\left[\frac{1}{m^2}\right]</math></b>	-0.017
<b>Distorsione K2 <math>\left[\frac{1}{m^4}\right]</math></b>	0.999
<b>Distorsione P1 <math>\left[\frac{1}{m^2}\right]</math></b>	0.0097

**Tabella 2.10:** Risultati di calibrazione con il metodo di Boundle Adjustment con parametri estrinseci liberi, 16 immagini, modello a 6 parametri ( $u_0$ ,  $v_0$ ,  $f$ ,  $K1$ ,  $K2$ ,  $P1$ ), dimensione dei pixel  $7.4 \mu m$ , risoluzione  $1400 \times 1024$  pixel, ottica da 13 mm, popolazione di 10000 individui per 1000 generazioni.



**Figura 2.20:** Grafico che descrive l'errore di retroproiezione risultante utilizzando la tecnica basata su GA che minimizzano l'errore di ricostruzione tridimensionale in funzione delle dimensioni della popolazione. Risultati ottenuti con numero delle generazioni pari a 100 e 6 immagini.



**Figura 2.21:** Grafico che descrive il tempo impiegato per eseguire la calibrazione con la tecnica basata su GA che minimizzano l'errore di ricostruzione tridimensionale in funzione delle dimensioni della popolazione. Risultati ottenuti con numero delle generazioni pari a 100 e 6 immagini.

Le tecniche di bundle adjustment quindi non portano a risultati soddisfacenti. Se, infatti, utilizziamo un set di parametri che comprende intrinseci, estrinseci (6 per ogni immagine) e distorsioni, il problema è troppo complesso e il GA impiega troppo tempo per convergere a una soluzione ottima, pur partendo dai risultati dell'algoritmo di Zhang. Se invece si bloccano gli estrinseci e si impongono uguali a quelli calcolati con la tecnica di Zhang, la soluzione non converge a quella ottima e i valori dei parametri variabili rimangono molto vicini a quelli trovati con la risoluzione del problema lineare.

## **2.4.4 Tecniche basate sull'errore di triangolazione**

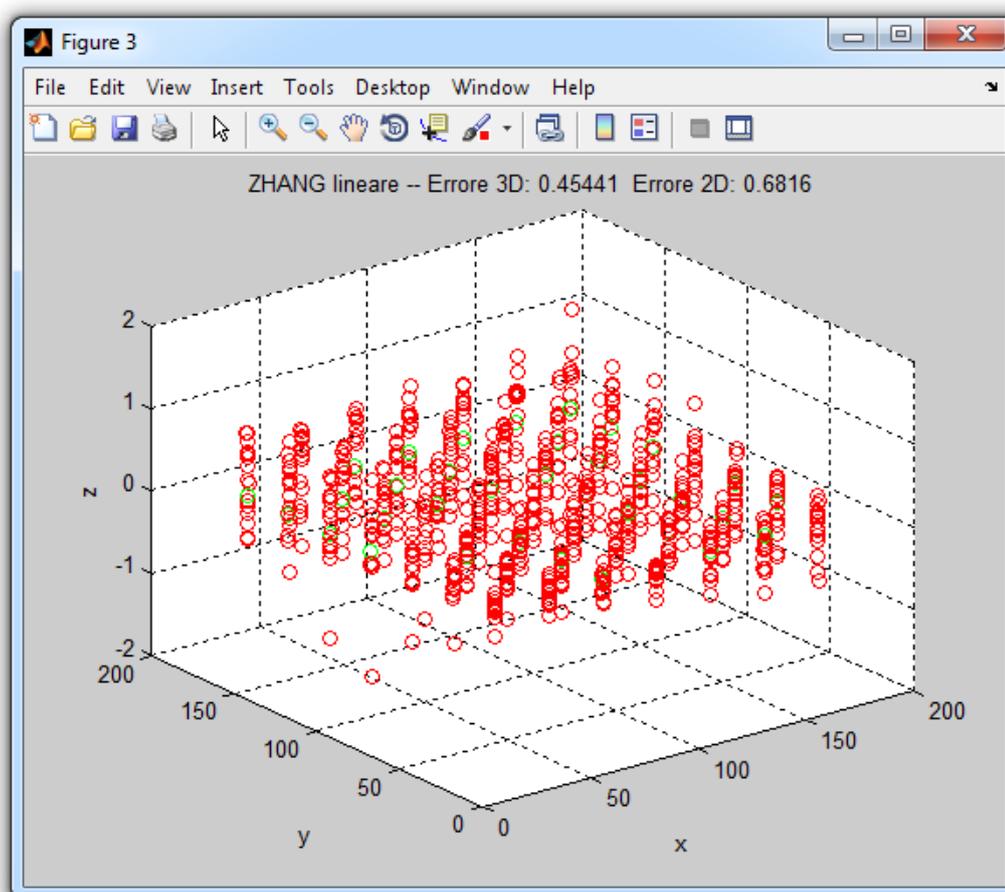
Per tentare di migliorare le tecniche di bundle adjustment si è deciso di sperimentare una tecnica che prevede di calibrare ottimizzando l'errore di ricostruzione tridimensionale. Per fare ciò è stato implementato un algoritmo genetico che ha come funzione di fitness la media delle distanze tra i punti noti del calibratore e i punti ricostruiti tramite un algoritmo di triangolazione. Le coordinate dei punti noti nello spazio sono conosciute in quanto, essendo il sistema di riferimento sempre solidale con il calibratore, la coordinata lungo Z sarà uguale a zero, mentre (X,Y) saranno pari alle coordinate del modello del calibratore 2D. L'algoritmo di triangolazione invece si basa sulla teoria descritta nel capitolo 3.3. Esso necessita delle informazioni provenienti da due diverse immagini tra quelle acquisite. Infatti è possibile calcolare le coordinate tridimensionali dei punti noti nello spazio, conoscendone la posizione in due immagini e conoscendo i parametri estrinseci relativi alle due immagini e i parametri intrinseci del dispositivo. Quindi l'algoritmo genetico minimizza l'errore di triangolazione modificando i parametri estrinseci relativi alle due immagini. Bisognerà allora eseguire un GA per ogni coppia di immagini tra quelle acquisite. Sono stati verificati diversi metodi di accoppiamento delle immagini e quello che ha restituito risultati migliori è così schematizzato:

- Si prende la prima immagine.

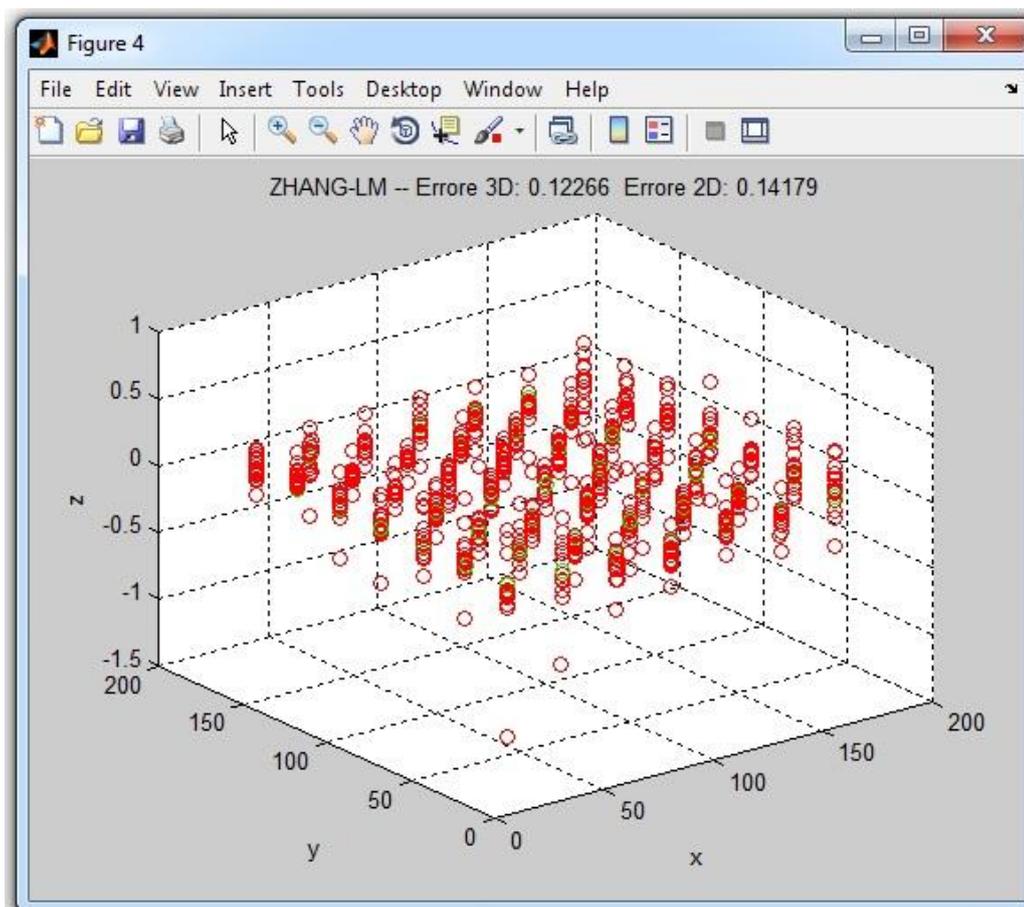
- Si accoppia con tutte le altre e per ogni accoppiamento viene eseguito un algoritmo genetico che minimizza l'errore di ricostruzione, utilizzando come variabili i parametri della prima immagine e mantenendo fissi i parametri delle altre.
- Eseguiti tutti gli accoppiamenti per la prima immagine si avrà una serie di parametri leggermente diversi tra loro relativi a essa; questi vengono mediati tra loro trovando un solo set di parametri.
- Si passa alla seconda immagine e la si accoppia con tutte le altre. Si lasciano fissi i parametri delle altre immagini (che nel caso della prima immagine saranno quelli trovati al punto precedente) e si fanno variare solamente i parametri della seconda immagine.
- Si mediano i risultati trovando un nuovo set di parametri per la seconda immagine.
- Si passa alla terza immagine e così via per tutte le altre ...

Si è notato che ripetere l'intera procedura qui descritta più volte porta a risultati più accurati. In particolare la sequenza è stata eseguita 3 volte. Chiaramente tutto ciò fa sì che non è possibile utilizzare un numero elevato di immagini, altrimenti l'intero algoritmo sarebbe eccessivamente lento anche utilizzando GA con basso numero di generazioni e popolazioni ristrette. Infatti ogni GA va rieseguito per ogni coppia di immagini, per tre volte. Quindi se ad esempio si utilizzassero 16 immagini come con Zhang-LM (capitolo 2.4.1), si dovrebbe eseguire l'algoritmo genetico  $((16 - 1) \times 16) \times 3 = 720$  volte. Si è dunque deciso di ridurre il numero delle immagini acquisite a 6. Inoltre l'algoritmo di triangolazione non tiene conto delle distorsioni ed è necessario dunque utilizzare un modello di telecamera lineare a 3 parametri ( $u_0$ ,  $v_0$  e  $f$ ). In questo modo non è più possibile confrontare i risultati delle tecniche basate sull'errore di triangolazione con quelli mostrati nel capitolo 2.4.1. Sono stati quindi eseguiti nuovamente gli algoritmi di Zhang lineare e Zhang-LM utilizzando solamente 6 immagini e un modello a 3 parametri, in modo da poter confrontare i risultati con i GA basati sull'errore di ricostruzione. Si è inoltre implementato una funzione che dato il set di immagini e i parametri di calibrazione, restituisce l'errore di retroproiezione medio tra tutte

le immagini e l'errore di triangolazione medio tra tutte le coppie di immagini possibili. Essa mostra poi in un grafico i punti ricostruiti come ad esempio in figura 2.22 e 2.23. In esse sono mostrati in verde i punti del modello del calibratore e in rosso i punti ricostruiti. Questi sono stati trovati accoppiando tutte le immagini acquisite e triangolando per ogni accoppiamento. In questo modo per ogni tecnica è possibile avere un riscontro visivo dell'errore di ricostruzione e un valore numerico per quanto riguarda l'errore di retroproiezione medio e di ricostruzione medio.



**Figura 2.22:** Risultati di calibrazione con la tecnica di Zhang, 6 immagini, modello lineare a 3 parametri ( $u_0$ ,  $v_0$ ,  $f$ ), dimensione dei pixel  $7.4 \mu\text{m}$ , risoluzione  $1400 \times 1024$  pixel, ottica da 13 mm. Viene mostrato l'errore di retroproiezione medio (errore 2D) e l'errore di ricostruzione medio (errore 3D). Nel grafico sono rappresentati in verde i punti del modello del calibratore e in rosso i punti ricostruiti per ogni coppia di immagini.



**Figura 2.23:** Risultati di calibrazione con la tecnica di Zhang-LM, 6 immagini, modello lineare a 3 3 parametri ( $u_0, v_0, f$ ), dimensione dei pixel  $7.4 \mu\text{m}$ , risoluzione  $1400 \times 1024$  pixel, ottica da 13 mm. Viene mostrato l'errore di retroproiezione medio (errore 2D) e l'errore di ricostruzione medio (errore 3D). Nel grafico sono rappresentati in verde i punti del modello del calibratore e in rosso i punti ricostruiti per ogni coppia di immagini.

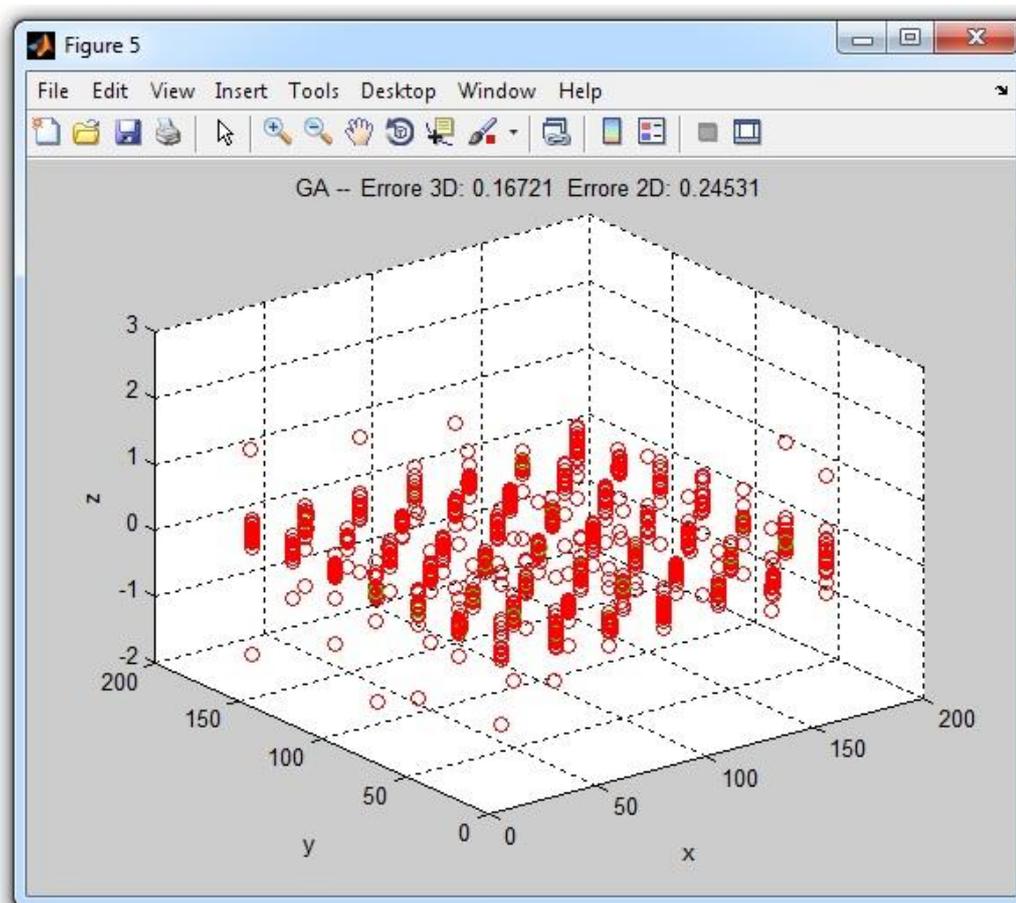
Ad esempio in figura 2.22 sono rappresentati i risultati della tecnica di Zhang lineare utilizzando solamente 6 immagini. In questo caso l'errore di retroproiezione medio è di 0.454 pixel, mentre quello di triangolazione medio è 0.682 millimetri. La figura 2.23 invece rappresenta i risultati di Zhang-LM applicato a 6 immagini. Si nota che i risultati in termini di errore di retroproiezione sono diversi da quelli mostrati nel capitolo 2.4.1. Utilizzando meno immagini e trascurando le distorsioni, la differenza tra l'errore di retroproiezione trovato in Zhang e in Zhang-LM non è più così marcata, e la

soluzione calcolata con LM restituisce un valore più alto di errore di retroproiezione rispetto a quello in tabella 2.3, ed è dunque meno accurata.

Inizialmente si è pensato di utilizzare i GA basati sull'errore di ricostruzione solamente per la stima dei parametri estrinseci. Il problema delle tecniche di bundle adjustment, come è stato mostrato nel capitolo 2.4.3, è che per convergere a una soluzione ottima impiegano un tempo eccessivamente alto se i parametri estrinseci sono lasciati liberi, mentre non converge alla soluzione ottima se i valori degli estrinseci sono fissati a quelli calcolati con Zhang. Inizialmente si è implementata un algoritmo ibrido per tentare di risolvere questi problemi. Esso consiste nell'eseguire l'algoritmo di Zhang, inizializzare una serie di GA che minimizzano l'errore di retroproiezione con la procedura di accoppiamento delle immagini descritta in precedenza e con gli estrinseci fissati a quelli calcolati in Zhang. Gli intrinseci vengono poi calcolati con un altro GA che ottimizza l'errore di retroproiezione. Fin dalle primi risultati si è subito capito che questa tecnica non risolveva alcun problema, in quanto i valori dei parametri risultanti rimanevano molto vicini a quelli trovati con Zhang. Questo fatto è causato dal fatto che i GA che stimano gli estrinseci ottimizzano l'errore di ricostruzione utilizzando parametri intrinseci fissi e uguali a quelli trovati in Zhang. Quindi la soluzione che si viene a trovare rimane molto vicina a quella di Zhang. Basandosi su questa soluzione viene eseguito il GA che stima gli intrinseci, che ancora una volta non si sposteranno dalla soluzione di Zhang. Anche implementando in loop più volte questa serie di GA si ottiene sempre un risultato molto vicino a quello trovato con la tecnica di Zhang.

Ancora una volta allora l'unica soluzione possibile è quella di utilizzare un problema più complesso in cui tutti i parametri sono variabili. Si è voluto allora implementare una tecnica basata sui genetici che minimizzasse solamente l'errore di ricostruzione, con tutti i parametri del modello (intrinseci ed estrinseci) variabili. Questa volta lo scopo non è più quello di trovare una soluzione alternativa a Zhang-LM, poiché a priori si sa che la soluzione trovata non sarà accurata come quella trovata nel capitolo 2.4.1 e i tempi di esecuzione saranno molto maggiori. Infatti con questa tecnica si è dovuto utilizzare un numero

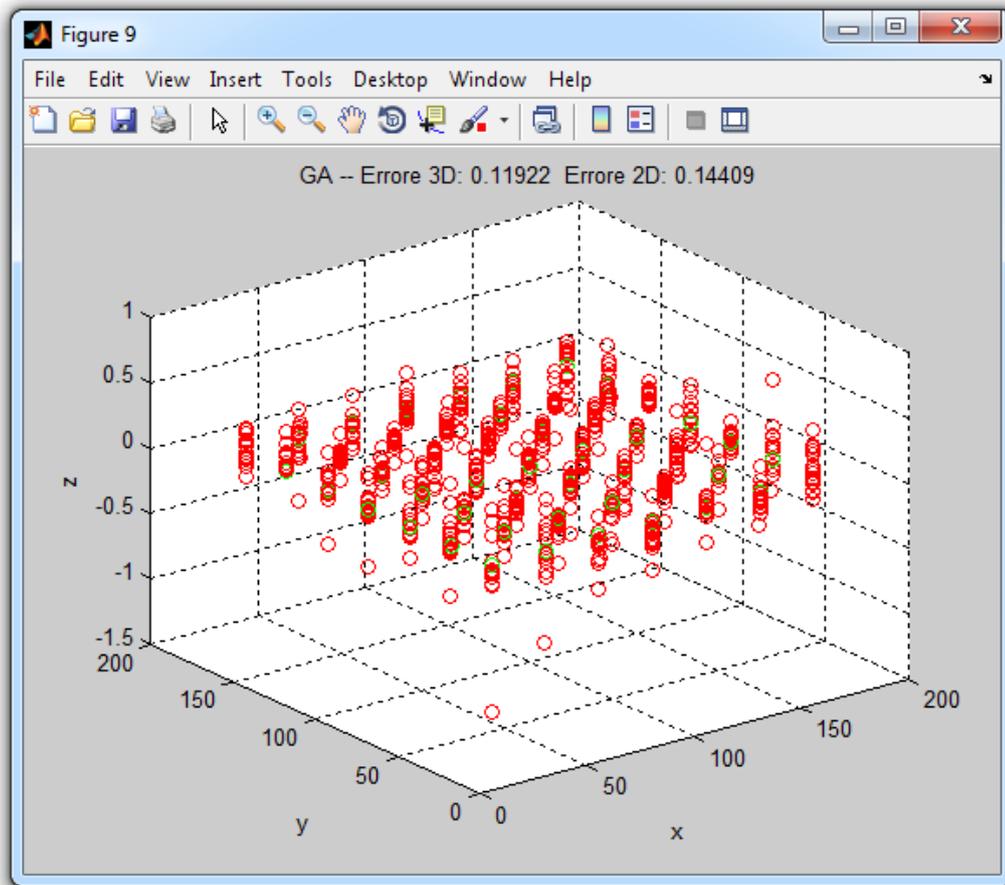
ristretto di immagini e un modello a 3 parametri lineare, limitazioni che influenzano l'accuratezza dei risultati di calibrazione. Rimane comunque la curiosità di voler implementare una tecnica innovativa e confrontarne i risultati con un metodo comunemente utilizzato come quello di Zhang-LM.



**Figura 2.24:** Risultati di calibrazione con la tecnica basata sui GA che minimizzano l'errore di ricostruzione tridimensionale, 6 immagini, modello lineare a 3 parametri ( $u_0$ ,  $v_0$ ,  $f$ ), dimensione dei pixel  $7.4 \mu\text{m}$ , risoluzione  $1400 \times 1024$  pixel, ottica da 13 mm. Viene mostrato l'errore di retroproiezione medio (errore 2D) e l'errore di ricostruzione medio (errore 3D). Nel grafico sono rappresentati in verde i punti del modello del calibratore e in rosso i punti ricostruiti per ogni coppia di immagini.

Come nel caso precedente la procedura di accoppiamento delle immagini è quella descritta sopra. Una volta calcolati i risultati di calibrazione finali viene comunque calcolato l'errore di retroproiezione medio e l'errore di ricostruzione medio, così da poter confrontare i risultati con quelli di figura 2.23. Anche in questo caso sono stati fatti numerosi tentativi variando il numero di individui per popolazione e in numero di generazione, ma per ottenere risultati vicini a quelli trovati con Zhang-LM, è stato necessario utilizzare una popolazione di 800 individui per 500 generazioni, e dunque eseguire l'algoritmo per 36 ore. Si sono dunque ottenuti i risultati in figura 2.24. Ancora una volta il tempo di esecuzione è eccessivamente alto, soprattutto considerando che sono state utilizzate solamente 6 immagini (invece che 16 come nelle tecniche precedenti).

Anche se la tecnica qui descritta non ha portato a risultati soddisfacenti, il confronto tra essa e Zhang-LM ha portato a considerazioni molto interessanti. Si noti come, anche se minimizzando due funzioni diverse, i due metodi restituiscano risultati simili in termini di errore di retroproiezione e di ricostruzione. Questo dimostra che la tecnica di Zhang-LM, pur minimizzando l'errore di retroproiezione, ottiene ottimi risultati anche in termini di errore di ricostruzione. Ad ulteriore verifica di questo fatto si è voluto eseguire l'algoritmo basato su una serie di algoritmi genetici sopra descritto inizializzando i parametri con quelli trovati in Zhang-LM. Lasciando eseguire l'algoritmo anche per diverse ore i risultati si spostano di pochissimo da quelli di Zhang-LM. In particolare viene leggermente migliorato l'errore di ricostruzione, mentre l'errore di retroproiezione varia di pochissimo, come dimostra la figura 2.25. Da ciò si può concludere che, anche se Zhang-LM minimizza solamente l'errore di retroproiezione, la soluzione trovata è ottima sia per quanto riguarda l'errore di retroproiezione, che per l'errore di ricostruzione tridimensionale.



**Figura 2.25:** Risultati di calibrazione con la tecnica basata su GA che minimizzano l'errore di ricostruzione tridimensionale, inizializzato con i parametri trovati con il metodo di Zhang-LM, 6 immagini, modello lineare a 3 parametri ( $u_0$ ,  $v_0$ ,  $f$ ), dimensione dei pixel  $7.4 \mu\text{m}$ , risoluzione  $1400 \times 1024$  pixel, ottica da 13 mm. Viene mostrato l'errore di retroproiezione medio (errore 2D) e l'errore di ricostruzione medio (errore 3D). Nel grafico sono rappresentati in verde i punti del modello del calibratore e in rosso i punti ricostruiti per ogni coppia di immagini.

## 2.4.5 Halcon

Al confronto tra varie tecniche di calibrazione si è voluto aggiungere quella contenuta in una libreria commerciale specifica per la visione artificiale, chiamata HALCON e sviluppata da MVTec [29]. Questo applicativo consente di eseguire calcoli e manipolazioni su immagini, utilizzando un proprio linguaggio nell'ambiente di sviluppo HDevelop. In esso vi sono moltissime funzioni, che consentono di effettuare l'intera procedura di calibrazione. In alternativa, è

possibile utilizzare un tool molto pratico e immediato, chiamato *Calibration Assistant*. In esso si ha la possibilità di effettuare una calibrazione completa, vale a dire con tutti i parametri da stimare, oppure fissarne alcuni a un valore prestabilito e stimare gli altri. Per ogni parametro che viene stimato durante la calibrazione si lascia la possibilità di inicializzarlo o meno a un valore definito dall'utente. In caso di calibrazione completa è comunque necessario impostare un valore iniziale di lunghezza focale e dimensioni della cella. Mentre la focale e la componente orizzontale del pixel sono solamente valori da cui partire per effettuare la stima, il valore che esprime la componente verticale del pixel rimane fisso durante la calibrazione e non viene stimato. Questo è uno dei principali svantaggi della tecnica di HALCON, ovvero dover conoscere con esattezza le dimensioni del pixel. Questa informazione è facilmente ricavabile da datasheet nel caso in cui si sta utilizzando telecamere industriali, mentre è più difficile da ricavare nel caso si utilizzi ad esempio una webcam. Se non si conosce con esattezza le dimensioni dei pixel che compongono il sensore la tecnica di HALCON non può essere utilizzata.

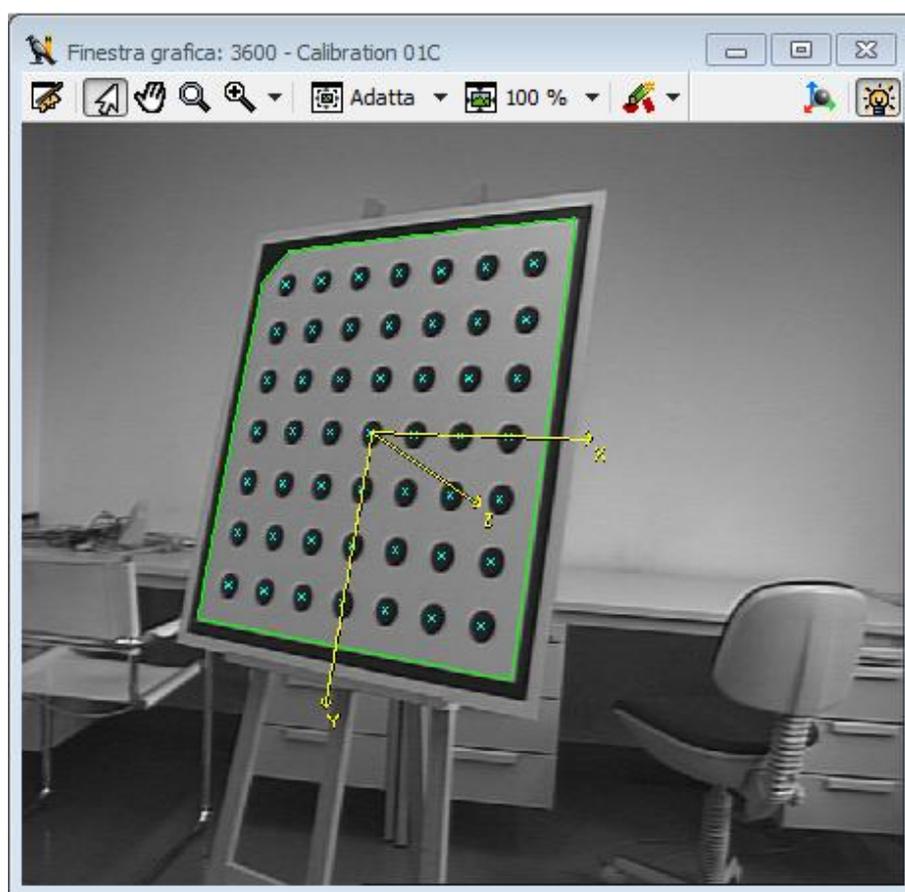
Selezionando una calibrazione completa, è possibile scegliere due modelli di telecamera: uno chiamato *divisionale* e uno *polinomiale*. Il primo ha come parametri l'ampiezza dei pixel nelle due direzioni (impostato dall'utente), la lunghezza focale, la posizione del centro ottico (inicializzata nel centro del sensore) e un parametro che tiene conto delle distorsioni radiali *Kappa*. Il modello polinomiale è invece più complesso e utilizza cinque parametri che rappresentano le distorsioni (invece di uno solo, Kappa), tre per le distorsioni radiali (K1, K2, K3) e due per le tangenziale (P1 e P2). Da numerose prove ed esperimenti si è notato che l'utilizzo del secondo modello porta sempre a errori di retroproiezioni minore e spesso a calibrazioni più accurate. Ad esempio in figura 2.28 e 2.29 sono rappresentati rispettivamente i risultati di calibrazione trovati utilizzando un modello divisionale e polinomiale. Si nota che l'errore medio (vale a dire l'errore di retroproiezione) nella seconda figura è minore rispetto a quello nella prima.

Una volta selezionato il modello e inizializzati correttamente i parametri, si passa all'acquisizione delle immagini. HALCON richiede che le immagini siano prese in maniera adeguata seguendo una serie di regole:

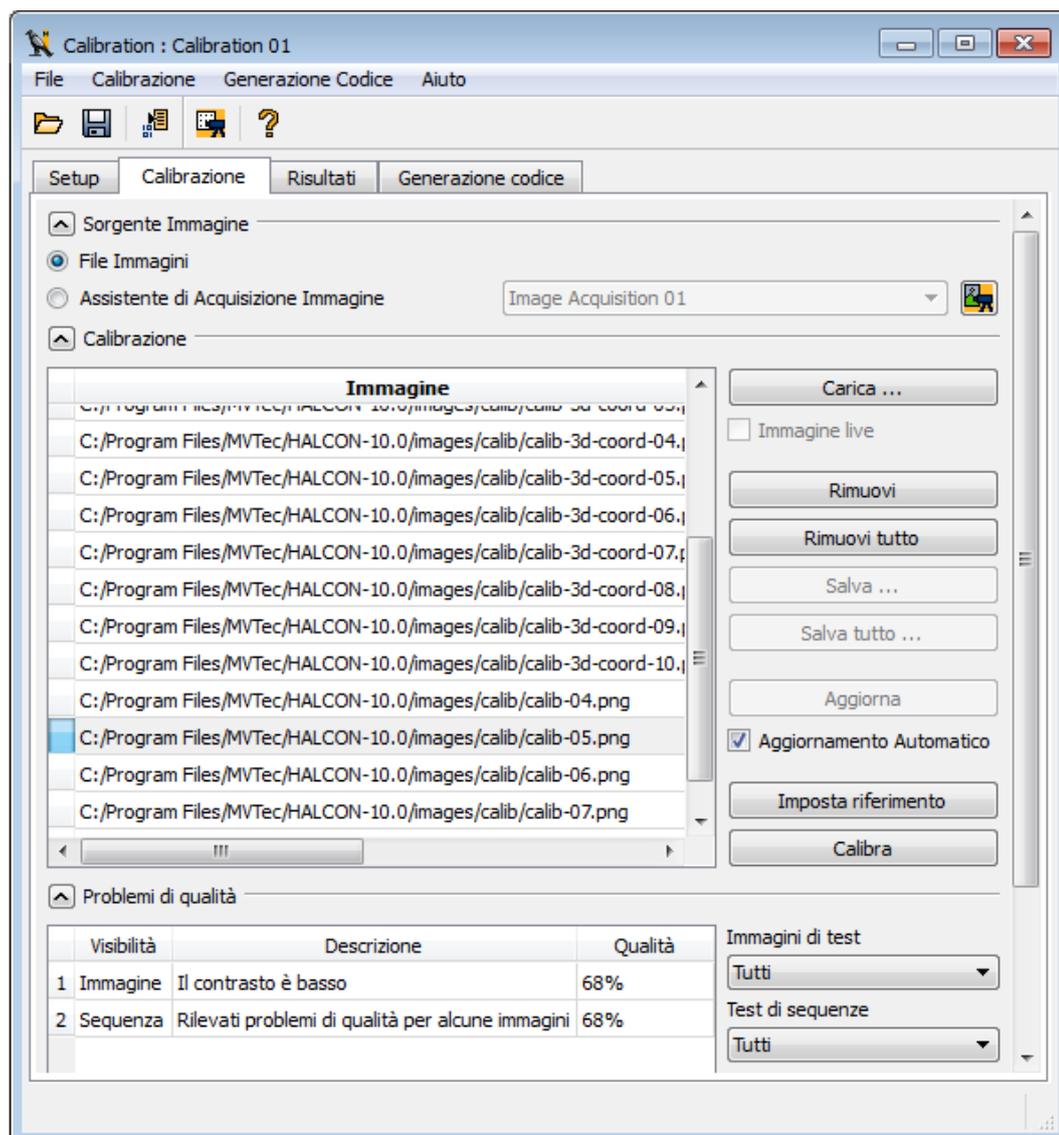
1. Il calibratore deve essere abbastanza grande da coprire almeno un quarto dell'immagine.
2. Il diametro minimo dei blob deve essere almeno 10 pixel.
3. Usare un'illuminazione che consenta di lasciare lo sfondo molto più scuro rispetto al calibratore (il contrasto tra uno e l'altro deve essere almeno 100 valori di grigi).
4. Il valore di grigio della parte bianca del calibratore deve essere almeno 100.
5. L'illuminazione deve assicurare che il calibratore sia rappresentato da valori di grigi omogenei.
6. L'immagine non deve essere sovraesposta, vale a dire non devono essere presenti valori di grigi maggiori di 255.
7. Il set di immagini deve coprire l'intero campo di visione.
8. Bisogna utilizzare varie pose del calibratore.
9. Per una buona calibrazione è necessario acquisire un numero di immagini compreso tra 15 e 20; per poter calibrare sono richieste comunque almeno 10 immagini.

L'acquisizione delle immagini può essere fatta in due modi. Il primo consiste nell'acquisire una serie di immagini (di solito almeno venti o trenta) a monte della calibrazione, utilizzando HALCON o qualsiasi programma che consente di eseguire un'istantanea dal dispositivo. Le immagini vengono poi caricate nel Calibration Assistant, che effettua un controllo qualità per ognuna di esse e in maniera automatica trova le correlazioni tra i punti nelle immagini e il modello del calibratore. Grazie a questo è possibile controllare che le immagini siano state acquisite in modo adeguato, così da poter eliminare quelle meno adatte o aggiungerne alcune acquisite in modo più accurato. Un esempio di controllo qualità delle immagini è mostrato in figura 2.26 e 2.27. In particolare nella figura 2.27 è rappresentata l'interfaccia del Calibration Assistant che effettua il controllo

qualità sulle immagini caricate. In questo caso è selezionata quella mostrata in figura 2.26 e il controllo è eseguito su di essa. Grazie alla verifica della qualità possiamo vedere che l'immagine presenta un contrasto basso e dunque non è un'immagine buona per la calibrazione. Il secondo metodo di acquisizione consiste nel visualizzare il calibratore in modalità live tramite il dispositivo; in real-time HALCON effettua il controllo della qualità dell'immagine e trova le correlazioni tra i punti e il modello del calibratore. In questo modo è possibile acquisire un'istantanea ogni volta che il controllo qualità è passato ed è stata trovata la corrispondenza tra il modello del calibratore e i punti nell'immagine. Entrambi i metodi dunque permettono anche a un utente poco esperto in calibrazione di non commettere errori in fase di acquisizione delle immagini, che possono compromettere la calibrazione e dunque l'intera misurazione.



**Figura 2.26:** Esempio di immagine di calibrazione nel quale è stata trovata la corrispondenza tra il modello di calibratore e i punti nell'immagine. Come si vede dal controllo qualità (figura 2.27) l'immagine ha però il contrasto basso e dunque non è un'immagine buona per calibrare.



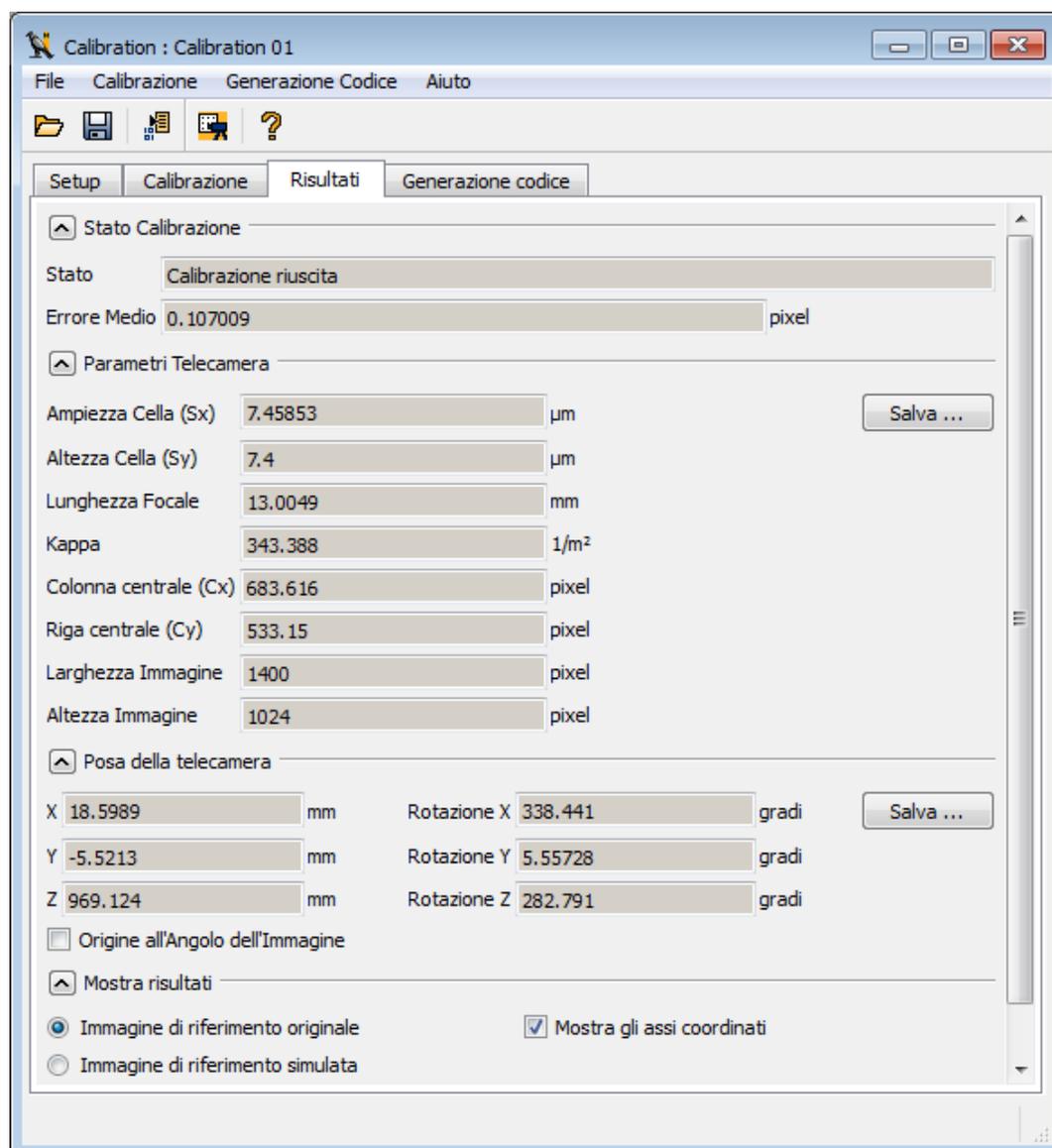
**Figura 2.27:** Interfaccia di controllo della qualità delle immagini. In questo caso è selezionata l'immagine di figura 2.26, in cui il contrasto risulta essere basso.

Una volta acquisito il set di immagini, è possibile procedere con la calibrazione. Bisogna solamente selezionare quale immagine usare come sistema di riferimento; esso viene posizionato nel centro del calibratore. I risultati sono poi mostrati in una schermata come quella di figura 2.28, in cui in alto è mostrato l'errore medio, vale a dire l'errore di retroproiezione medio, e l'elenco dei parametri intrinseci, delle distorsioni e infine i parametri estrinseci (“Posa della telecamera”).

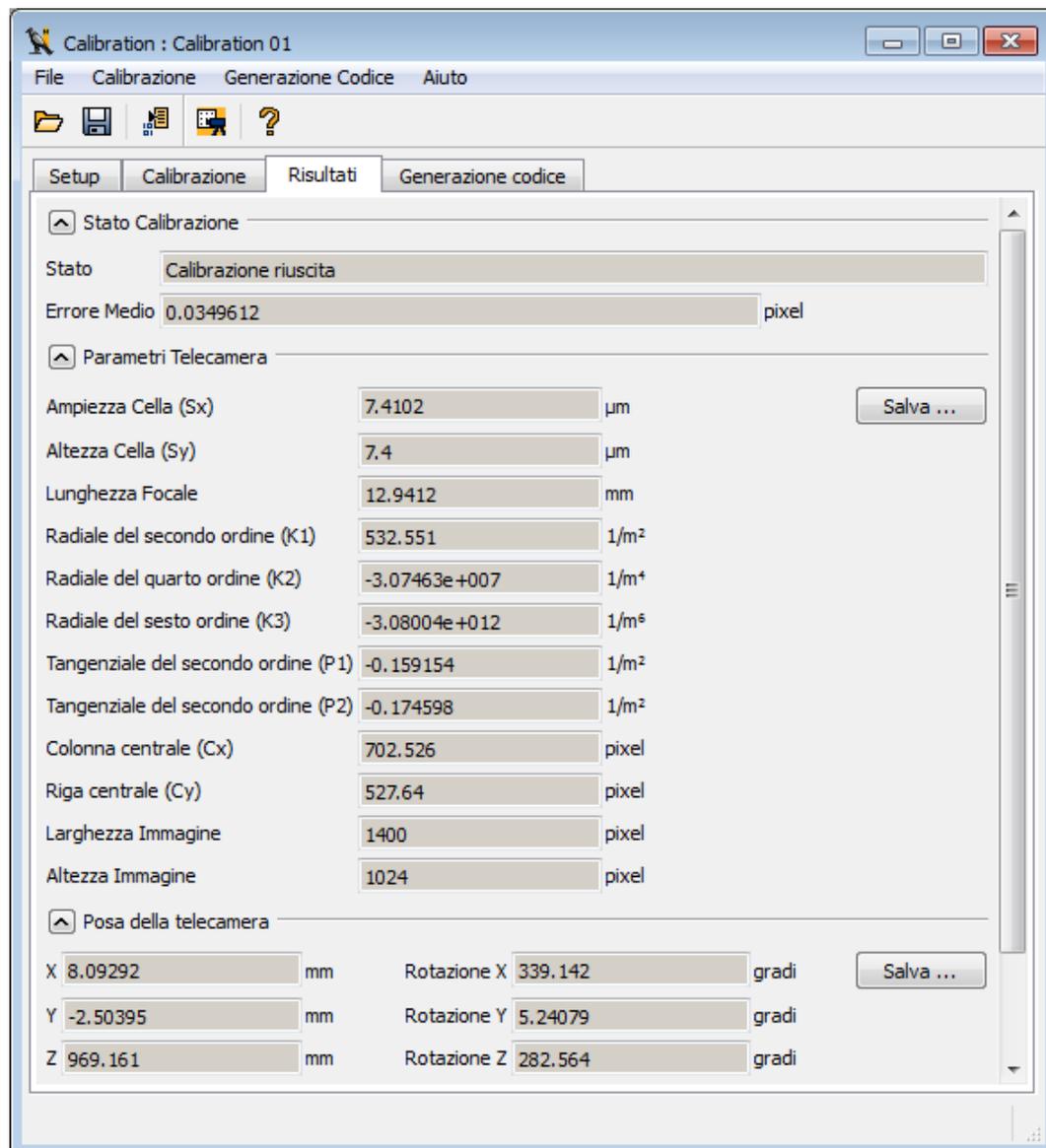
La tecnica di HALCON, oltre a essere estremamente *user friendly*, permette, come dimostrato in seguito, di ottenere risultati accurati. Grazie alla procedura automatica di controllo della qualità delle immagini acquisite infatti l'algoritmo riesce a raggiungere prestazioni elevate. Per il confronto con l'algoritmo di Zhang-LM non sono state utilizzate le stesse immagini poiché alcune non hanno superato il controllo qualità di HALCON ed è stato quindi necessario aggiungerne altre. Il dispositivo utilizzato è comunque lo stesso, con dimensione dei pixel  $7.4 \mu\text{m}$ , risoluzione  $1400 \times 1024$  e ottica da 13 mm. Chiaramente il modello utilizzato non può essere quello a sei parametri del capitolo 2.4.1, poiché HALCON prevede solamente la risoluzione di due tipi di modello: quello divisionale e quello polinomiale. Nella figura 2.28 e 2.29 sono rappresentati i risultati del problema di calibrazione rispettivamente nel caso di modello divisionale e polinomiale. Nel primo caso l'errore di retroproiezione è molto vicino a quello trovato con Zhang-LM e anche i parametri intrinseci sono circa uguali a quelli trovati nel capitolo 2.4.1 (tabella 2.3). Nel caso di modello polinomiale invece i risultati si discostano di pochissimo, trovando un errore di retroproiezione leggermente minore e dunque una soluzione poco più accurata. Per quanto riguarda i tempi di esecuzione dell'algoritmo di calibrazione di HALCON, essi sono compatibili con i tempi impiegati dall'algoritmo di Zhang-LM (pochi secondi) e non rappresentano dunque un problema per la calibrazione.

Riassumendo, HALCON presenta alcuni vincoli non indifferenti. Per prima cosa per poter calibrare è necessario conoscere la dimensione dei pixel che compongono il sensore con esattezza. Quindi tale tecnica non può sempre essere applicata. Secondariamente HALCON richiede che per calibrare si acquisisca un insieme di immagini con caratteristiche molto esigenti. Il processo di acquisizione delle immagini deve dunque essere fatto con molta attenzione. Questo fattore però può anche essere considerato un vantaggio, in quanto le limitazioni restringenti sulla bontà delle immagini fanno sì che gli errori in fase di acquisizione siano meno frequenti. Inoltre HALCON effettua la ricerca delle correlazioni tra il modello del calibratore e i punti nell'immagine in maniera semplice e automatica. Grazie al Calibration Assistant, HALCON si presta ad essere utilizzato anche da

utenti meno esperti. Infine è stato verificato che l'accuratezza dei risultati di calibrazione calcolati con HALCON è compatibile con quella trovata col metodo di Zhang-LM.



**Figura 2.28:** Risultati di calibrazione utilizzando il Calibration Assistant di HALCON con un modello divisionale ( $u_0$ ,  $v_0$ ,  $f$ ,  $K$ ), 15 immagini, dimensione dei pixel 7.4  $\mu\text{m}$ , risoluzione 1400x1024 pixel, ottica da 13 mm.



**Figura 2.29:** Risultati di calibrazione utilizzando il Calibration Assistant di HALCON con un modello polinomiale ( $u_0$ ,  $v_0$ ,  $f$ ,  $K_1$ ,  $K_2$ ,  $K_3$ ,  $P_1$ ,  $P_2$ ), 15 immagini, dimensione dei pixel  $7.4 \mu\text{m}$ , risoluzione  $1400 \times 1024$  pixel, ottica da 13 mm.

## 2.5 Conclusioni

In conclusione il metodo di Zhang-LM risulta essere sicuramente un approccio migliore dei GA. Inoltre riesce a convergere a una soluzione ottima in tempi ristretti, di pochi secondi. Tutte le tecniche che si basano sugli algoritmi genetici invece hanno lo svantaggio di essere eccessivamente lente e impiegano tempi esageratamente elevati per arrivare a risultati compatibili a quelli trovati con Zhang-LM., dunque per convergere alla soluzione ottima.

Grazie all'implementazione di una tecnica basata su algoritmi genetici che minimizzano l'errore di ricostruzione tridimensionale, al posto dell'errore di retroproiezione, è stato possibile dimostrare che la soluzione trovata con il metodo di Zhang-LM non solo è ottima per quanto riguarda l'errore di retroproiezione, ma è molto vicina all'ottimo globale anche per quanto concerne l'errore di ricostruzione tridimensionale. Nella tabella 2.11 e 2.12 sono riassunti alcuni dei risultati che sono stati utilizzati durante il capitolo 2.4 come termine di paragone tra le varie tecniche di calibrazione.

ALGORITMO	ERRORE DI RETROPROIEZIONE	TEMPO
Zhang lineare	1.27 (pixel)	0 s
Zhang-LM	0.066 (pixel)	2 s
Yang (GA)	0.17 (pixel)	1 h
Yang-LM (GA)	0.06 (pixel)	4 min
Bound. Adj. (GA)	0.12 (pixel)	13 h

**Tabella 2.11:** Confronto tra gli algoritmi di Zhang (lineare), di Zhang-LM (non lineare) e di Yang (GA), di Yang-LM (GA) e di bundle adjustment (GA). Si nota come Zhang-LM consente di trovare soluzioni più accurate in tempi notevolmente inferiori alle tecniche basate sui GA.

ALGORITMO	ERRORE DI RETROPROIEZIONE	ERRORE DI RICOSTRUZIONE	TEMPO
<b>Zhang lineare</b>	0.454 (pixel)	0.682 mm	0 s
<b>Zhang-LM</b>	0.123 (pixel)	0.142 mm	2 s
<b>GA 3D</b>	0.135 (pixel)	0.122 mm	3 h

**Tabella 2.12:** Confronto tra gli algoritmi di Zhang (lineare), di Zhang-LM (non lineare) e GA basato sull'ottimizzazione dell'errore di ricostruzione tridimensionale. Si nota come ancora una volta Zhang-LM consente di trovare soluzioni più accurate in tempi notevolmente inferiori alle tecniche basate sui GA.

Per quanto riguarda la tecnica di HALCON, l'unico svantaggio riscontrato è che essa richiede di conoscere la dimensione dei pixel del sensore e che raggiunge risultati accurati solamente se le immagini di calibrazione sono acquisite rispettando più possibile i vincoli imposti dal controllo qualità. Dato che sono state utilizzate telecamere industriali in cui la dimensione dei pixel è ricavabile dal datasheet, questo non ha comportato grossi problemi durante il lavoro svolto. Vantaggi della procedura di HALCON invece sono che essa permette di controllare la qualità delle immagini acquisite e risulta veloce e facile da utilizzare, anche da parte di un utente meno esperto in calibrazione. Da alcuni esperimenti è stato provato che i risultati della tecnica di HALCON sono compatibili a quelli ricavati con il metodo di Zhang-LM (si veda ad esempio i risultati espressi al capitolo 2.4.5 e riassunti nella tabella 2.13). Per queste ed altre ragioni che verranno indicate nel capitolo 4.3, si è deciso di utilizzare come metodo di calibrazione per il sistema di misura in esame la tecnica di HALCON. È da notare però che per avere risultati così ottimali è necessario acquisire immagini che rispettano i vincoli imposti da HALCON. Non sempre è però possibile fare ciò e in alcuni casi particolari dunque l'accuratezza della tecnica di HALCON risulta peggiore di quella di Zhang-LM. È stato infatti verificato che in alcuni casi in cui non si rispettano alcuni vincoli, come ad esempio nel caso in cui vi è un'illuminazione molto discontinua, le prestazioni di Zhang-LM sono superiori a quelle di HALCON. Si è comunque deciso di utilizzare la tecnica di HALCON,

ma si sottolinea che se è richiesta un'accuratezza particolarmente elevata è consigliabile controllare i risultati di essa con la tecnica di Zhang-LM.

ALGORITMO	ERRORE DI RETROPROIEZIONE	TEMPO
<b>Zhang-LM</b>	0.066 (pixel)	2 s
<b>HALCON (modello divisionale)</b>	0.107 (pixel)	2 s
<b>HALCON (modello polinomiale)</b>	0.035 (pixel)	4 s

**Tabella 2.13:** Confronto tra gli algoritmi di Zhang-LM (non lineare) e la tecnica di calibrazione di HALCON, rispettivamente utilizzando un modello divisionale e polinomiale. Si nota come la tecnica di HALCON si avvicini molto ai risultati trovati con Zhang-LM, addirittura ottenendo una soluzione di poco migliore con il modello polinomiale.

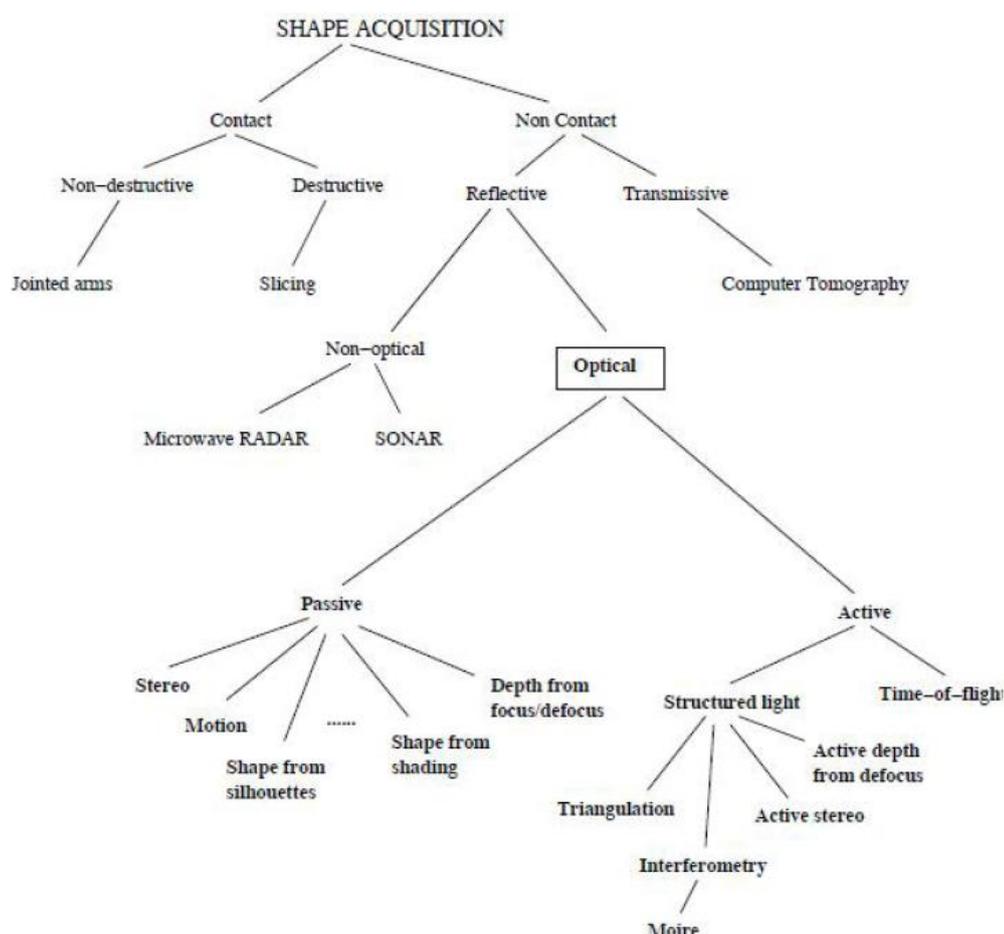
## 3. Metodologia di misura

*Lo strumento di misura in esame consente di ricostruire le coordinate tridimensionali di alcuni punti noti mediante la stereoscopia. In questo capitolo si vuole dare un'adeguata introduzione al problema della misurazione con tecniche stereoscopiche. Dopo un breve riassunto delle principali tecniche di misura tridimensionale si procede in questo capitolo alla descrizione della stereoscopia e delle motivazioni che hanno portato a sviluppare uno strumento di misura ottico stereoscopico. In seguito viene introdotta la geometria che sta alla base di essa, chiamata geometria epipolare e vengono spiegate alcune problematiche legate alle misure mediante stereoscopia e le motivazioni che hanno portato all'utilizzo di una terza telecamera.*

### 3.1 Tecniche di misura tridimensionali

Le tecniche di misura 3D coprono un'ampia serie di tecnologie data la vastità dei campi applicativi nei quali sono utilizzate. Uno schema riassuntivo di tutte le tecniche di misura tridimensionale [23],[24] è espresso in figura 3.1. Tra questi, i *metodi ottici* sono i più compatibili con le applicazioni in cui verrà utilizzato il sistema di misure in esame (ad esempio per misure in galleria del vento). Queste

tecniche, basate sull'analisi della scena mediante telecamere, non richiedono contatto con l'oggetto da misurare e consentono un'analisi piuttosto veloce e a bassi costi. Esse poi raggiungono accuratèzze elevate e possono essere applicate a campi di misura di differenti vastità. Le limitazioni includono il fatto di poter acquisire solo la parte di superficie visibile e la sensibilità ad alcune proprietà dei materiali quali trasparenza, brillantezza e colore.



**Figura 3.1:** Schema riassuntivo delle varie tecniche di ricostruzione tridimensionale.

La distinzione fondamentale tra le tecniche ottiche riguarda l'impiego o meno di sorgenti d'illuminazione strutturate. In particolare, si può fare una distinzione tra i *metodi attivi*, che irradiano la scena con radiazioni elettromagnetiche opportune (pattern luminosi, luce laser, radiazioni infrarosse, ecc...) e i *metodi passivi*, che si basano sull'analisi d'immagini della scena in condizioni variabili di luce naturale. I primi hanno il vantaggio di raggiungere risoluzioni elevate, ma sono più costosi

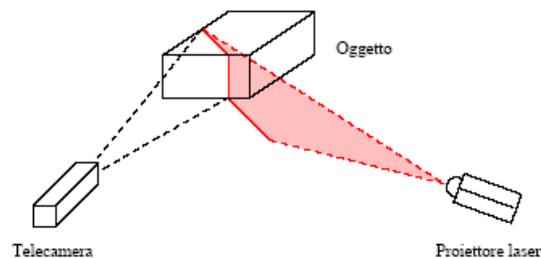
e non sempre applicabili. I secondi sono economici e pongono meno vincoli per l'applicabilità; la procedura di riconoscimento dei punti noti è generalmente meno affidabile e rischia di introdurre incertezze nel problema. Per questo si è preferito soffermarsi sui metodi attivi. In realtà la tecnica da noi utilizzata, ovvero marker catarifrangenti soggetti a illuminatori all'infrarosso, è considerata una tecnica semiattiva. Infatti la scena viene irraggiata da un fascio di luce che viene riflessa dai marker e quindi i marker non sono veramente attivi. Si utilizzano, però, illuminatori infrarossi e quindi non luce naturale. Il confine tra tecnica attiva e semiattiva è molto sottile e quindi esse vengono solitamente entrambe classificate come tecniche attive.

Le tecniche attive si dividono in *metodi a tempo di volo* e *tecniche a luce strutturata*. I primi sfruttano il medesimo principio di funzionamento del RADAR, ma utilizzano come sorgente delle onde elettromagnetiche il laser: uno scanner rileva punto a punto il ritardo fra l'emissione e la ricezione del raggio riflesso. La ricostruzione è possibile conoscendo la geometria del sistema di visione e la velocità di propagazione delle onde luminose. Tale metodo presenta il difetto di essere lento, in quanto richiede una scansione punto a punto. Esso non è dunque di interesse per l'applicazione in esame. I sistemi a luce strutturata più utilizzati sono: lo *sfocamento attivo*, la *triangolazione*, l'*interferometria* e la *stereoscopia attiva*.

Con lo *sfocamento attivo* (*Depth from defocus*), utilizzando un'illuminazione strutturata, da varie immagini caratterizzate da diversi parametri di fuoco si determina la profondità di tutti i punti della scena, grazie alle relazioni fra profondità, sfocamento e parametri dell'ottica. Le problematiche più evidenti sono la misurazione dello sfocamento e la calibrazione della relazione profondità – sfocamento, oltre alla necessità di acquisire diverse immagini. E' evidente come questo metodo risulti lento e richieda ottiche costose. Di conseguenza lo sfocamento non è utilizzabile nel caso in esame.

La *triangolazione attiva* consente di utilizzare una sola telecamera, proiettando una lama laser sul misurando, resa però solidale al proiettore. In un certo senso, quest'ultimo e la telecamera vedono la scena da posizioni diverse ed è

quindi possibile determinare la terza dimensione intersecando il raggio ottico di un punto dell'immagine con il corrispondente piano di luce prodotto dal proiettore. Per un corretto funzionamento della triangolazione è necessario realizzare un modello geometrico del proiettore e la calibrazione di quest'ultimo. Il grande vantaggio di questo metodo è quello di non richiedere la valutazione di alcuna corrispondenza, l'aspetto negativo è la richiesta, di un elevato numero di immagini e la possibilità di eseguire solamente misure statiche. Le applicazioni in cui viene utilizzato il sistema in esame richiedono la possibilità di effettuare misure dinamiche.



**Figura 3.2:** Applicazione della triangolazione attiva.

Con il termine *Interferometria* viene essenzialmente considerato il metodo delle *frange di interferenza di Moiré*. La teoria che è alla sua base non è banale e in realtà esistono diverse tecniche di acquisizione che si rifanno ad essa. L'idea di fondo è quella di proiettare una griglia su un oggetto e di acquisirne l'immagine attraverso una seconda griglia, analoga ma leggermente ruotata: le due griglie interferiscono e si creano delle frange di interferenza, dette appunto di moiré (*moiré fringe patterns*). Esse risultano essere fasce alternate di luce ed ombra: un'opportuna analisi di questi pattern permette di risalire alle variazioni di profondità nella scena e quindi alla ricostruzione 3D. Esistono diverse tecniche per generare le griglie, fra le altre esse possono essere proiettate tramite diapositiva oppure essere create tramite software. Il metodo moiré è veloce, estremamente accurato e non costoso, tuttavia presenta alcuni inconvenienti: innanzitutto è computazionalmente oneroso, inoltre non è immediato ricostruire la

scena laddove esistano nella scena dei salti bruschi di profondità e le superfici troppo angolate talvolta non sono misurabili, poiché presentano una densità delle bande troppo elevata. Dato che il sistema di misura in esame deve poter misurare diversi oggetti, indipendentemente dalla loro forma, non è consigliabile utilizzare un metodo che abbia queste limitazioni. Per questo e per la complessità computazionale del metodo, esso non è stato considerato in questo lavoro di tesi.



**Figura 3.3:** Applicazione del metodo di Moirè.

La stereoscopia attiva (*Active stereo*) consente, utilizzando almeno due telecamere, di ricostruire la terza dimensione analizzando le corrispondenze tra i punti nelle immagini acquisite dai diversi dispositivi. La luce strutturata facilita la risoluzione del problema dell'identificazione delle corrispondenze. Sono utilizzate diverse tipologie di luce strutturata (infrarossi, laser, etc...). Nel caso in esame, ad esempio, sono stati posizionati sull'oggetto da misurare marker catarifrangenti, illuminati da led all'infrarosso. Grazie a un filtro è possibile acquisire solamente la radiazioni infrarosso riflesse dai marker, riducendo notevolmente il campo del visibile. Così la determinazione dei punti noti, e dunque l'identificazione delle corrispondenze, risulta più facile. Questa tecnica permette di effettuare misure in maniera dinamica, specifica indispensabile per le applicazioni di interesse in questo lavoro. Uno svantaggio di questo metodo è la possibilità di misurare soltanto un numero limitato di punti, ovvero soltanto i punti noti di cui si conosce

le corrispondenze. Nelle applicazioni di interesse in questo lavoro però non è richiesta la misura di un numero di punti elevato (generalmente non più di 20 punti). Quindi è stato deciso di utilizzare come metodo di misura tridimensionale la stereoscopia attiva, descritta più in dettaglio nel capitolo 3.2.

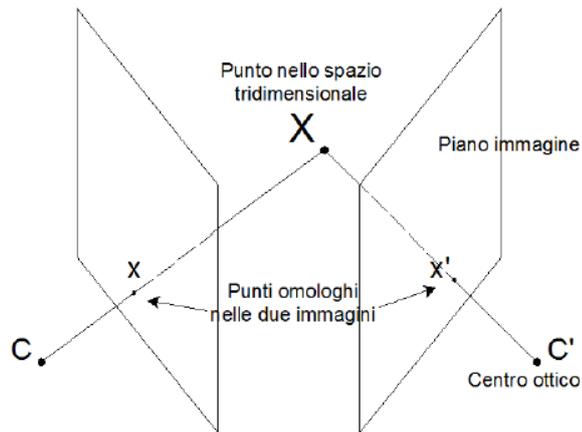
## 3.2 Stereoscopia

La *stereoscopia*, o *stereopsi*, è la capacità percettiva che consente di unire le immagini provenienti dai due punti di vista (ad esempio gli occhi nel caso di esseri umani), che grazie alla loro disposizione strutturale, presentano uno spostamento laterale. Questa disparità è sfruttata dal cervello per trarre informazioni sulla profondità e sulla posizione spaziale dell'oggetto mirato. Di conseguenza la stereopsi permette di generare la visione tridimensionale. Tale metodo naturale è stato studiato e applicato alla visione artificiale dando vita alla stereoscopia (o stereopsi computazionale), ovvero l'insieme di tecniche per ricavare informazioni sulla struttura tridimensionale della scena da almeno una coppia di immagini acquisite in sincronia da posizioni spaziali diverse e note.

Questo problema può essere risolto scomponendolo in due sottoproblemi: l'individuazione nelle immagini delle corrispondenze, cioè di punti immagine riferiti allo stesso punto fisico, e la ricostruzione tridimensionale degli stessi date le posizioni relative tra le due telecamere (processo chiamato triangolazione).

I punti immagine corrispondenti sono chiamati punti omologhi. La loro individuazione può risultare però difficile per la presenza di errori dovuti a occlusioni, distorsioni di luminosità e distorsioni proiettive. Le occlusioni sono i punti di cui non si ha corrispondenza in una delle immagini tipicamente perché essi non sono visibili da entrambe le posizioni per effetto di coperture. Le distorsioni di luminosità si hanno quando lo stesso punto, nelle due immagini, presenta un valore di luminosità differente. Questo accade tipicamente a causa di ombre o riflessi di luce sulla superficie osservata. La distorsione proiettiva invece

si manifesta quando un oggetto ripreso da due punti di vista differenti si proietta sulle due immagini in maniera differente.



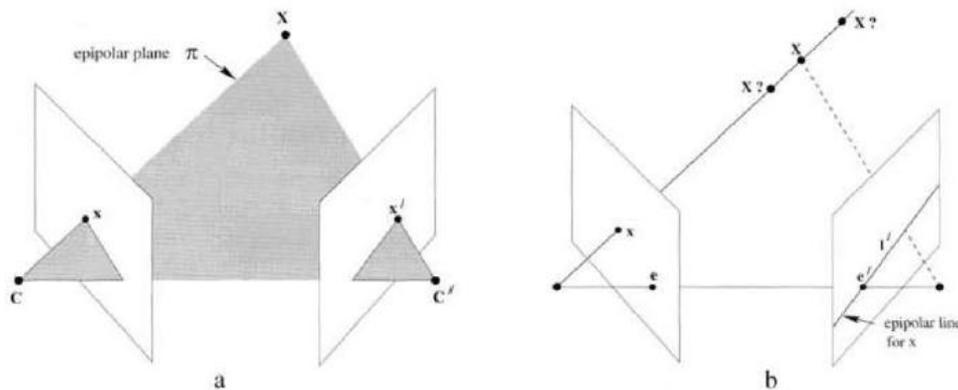
**Figura 3.4:** I due punti  $x$  e  $x'$  sono proiezione dello stesso punto della scena e sono dunque detti punti omologhi (o coniugati).

Per agevolare l'individuazione dei punti omologhi si impongono vincoli che semplificano la ricerca. Il più importante è il vincolo epolare che permette di affermare che il corrispondente di un punto in un'immagine può trovarsi solo su una retta (retta epolare) nell'altra immagine, spiegato più in dettaglio nel seguito. Una volta noti gli accoppiamenti tra i punti è possibile ottenere la posizione del punto tridimensionale che li ha generati; questo processo viene chiamato triangolazione.

Il grosso svantaggio di questo metodo di ricostruzione tridimensionale è la necessità di avere un sistema di misura calibrato e la possibilità di essere applicato solo a oggetti che presentino caratteristiche facilmente distinguibili in maniera automatica come blob, forme geometriche particolari, fori, spigoli, e altre discontinuità.

### 3.3 Geometria epipolare

Come descritto in dettaglio nel capitolo 3.2, per eseguire la triangolazione stereoscopica è necessario individuare le corrispondenze tra i punti nelle due immagini, detti *omologhi*, che sono la proiezione del medesimo punto nello spazio (figura 3.5). Questa procedura non è banale e per fare ciò bisogna ricorrere alla *geometria epipolare*. Essa descrive le relazioni geometriche tra due viste prospettiche della medesima scena 3D.



**Figura 3.5:** Geometria dei punti omologhi. (a) Il piano univocamente definito dalle coordinate dei centri delle telecamere e il punto  $X$  è detto piano epipolare. (b) La proiezione della linea congiungente il centro della telecamera di sinistra al punto  $X$  (che incontra il piano della telecamera nel punto  $x$ ) sul piano immagine destro è detta retta epipolare; il punto  $x'$  omologo di  $x$  giacerà idealmente su tale retta.

Date le coordinate dei due centri telecamera e le coordinate di un punto  $X$  nello spazio è possibile definire univocamente un piano chiamato piano epipolare. Collegando i centri delle telecamere tra loro si ottiene una linea chiamata *baseline* (o linea base), che intersecata con i piani dei sensori genera due punti chiamati *epipoli* ( $e; e'$ ). Questi punti possono essere visti come la proiezione nella seconda immagine del centro ottico della prima e viceversa. Se si proietta il raggio di proiezione (ovvero la retta che unisce il centro della telecamera  $C$  al punto  $X$ ) sul piano della seconda immagine, si trova la cosiddetta *retta epipolare*, che passa per il secondo epipolo  $e'$  e la proiezione del punto  $X$ , cioè  $x'$ . È dunque facile intuire

come dato un punto in un'immagine, il suo corrispondente nella seconda immagine possa trovarsi solamente lungo una linea passante per l'epipolo (la retta epipolare). Questa proprietà riduce notevolmente la complessità computazionale della ricerca dei punti omologhi, anche se in realtà, a causa di errori di misura, il punto coniugato non viene sempre a trovarsi esattamente sulla retta epipolare bensì molto vicino ad essa.

Per ricavare le rette epipolari è necessario conoscere la posa relativa tra le telecamere che riprendono la scena, informazione ricavabile grazie al processo di calibrazione ed espressa algebricamente tramite la *matrice fondamentale di telecamera*. Tale matrice, omogenea 3x3, di rango 2 e con 7 gradi di libertà, descrive la corrispondenza tra i punti nelle due immagini, e per definizione risulta che:

$$x' F x = 0 \quad (3.1)$$

Dove  $x$  e  $x'$  sono le coppie di punti omologhi.

Esistono due tipi di algoritmi che permettono di calcolarla: uno basato direttamente sulle corrispondenze tra i punti delle due immagini (almeno 8), il secondo basato sulla conoscenza della matrici di calibrazione. Il primo metodo è caratterizzato da un errore maggiore di ricostruzione sui punti rispetto il secondo.

Se le matrici di calibrazione delle due telecamere sono ottenute da immagini comuni espresse nel medesimo sistema di riferimento, è possibile ricavare direttamente la matrice fondamentale. Come visto nel capitolo 2.1.2 le matrici di proiezione di telecamera sono così definite:

$$P = K[R \ t] \quad P' = K' [R' \ t'] \quad (3.2)$$

Dove  $K$  e  $K'$  sono le matrici dei parametri intrinseci delle due telecamere e  $[R \ t]$ ,  $[R' \ t']$  la posizione relativa delle due telecamere rispetto ad un SDR comune. Dalla conoscenza di queste matrici di proiezione è possibile ricavare la posizione dei centri delle due telecamere (nello stesso sistema di riferimento) tramite scomposizione ai valori singolari della matrice  $P$ . Il centro della telecamera corrisponde, infatti, al vettore associato più piccolo valore singolare della matrice di proiezione:

$$[U; S; V] = SVD(P) \longrightarrow C = v_{end} \quad (3.3)$$

Dove  $v_{end}$  indica l'ultima colonna della matrice  $V$ . Una volta nota la posizione dei centri è possibile ricavare la posizione degli epipoli come:

$$e' = P'C \quad (3.4)$$

Come in precedenza gli apici indicano la seconda telecamera, mentre i valori senza apice sono riferiti alla prima telecamera. Dagli epipoli è poi ricavabile la matrice fondamentale come:

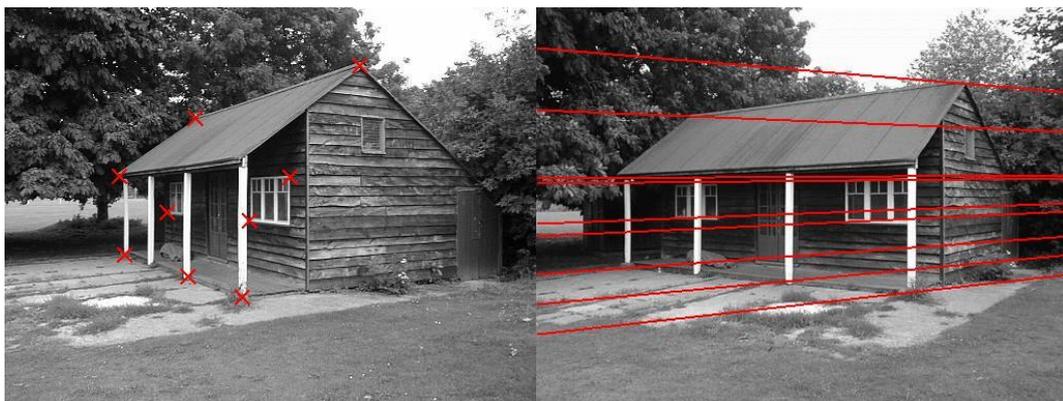
$$F = e'P'P^+ \quad (3.5)$$

Dove  $P^+$  è la pseudo-inversa della matrice di telecamera.

Una volta ricavata la matrice fondamentale e sapendo che essa rappresenta una correlazione, ovvero una mappatura proiettiva tra i punti nelle immagini e la linea epipolare che essi generano sull'altra immagine, si può scrivere che:

$$l' = Fx \quad l = F^T x' \quad (3.6)$$

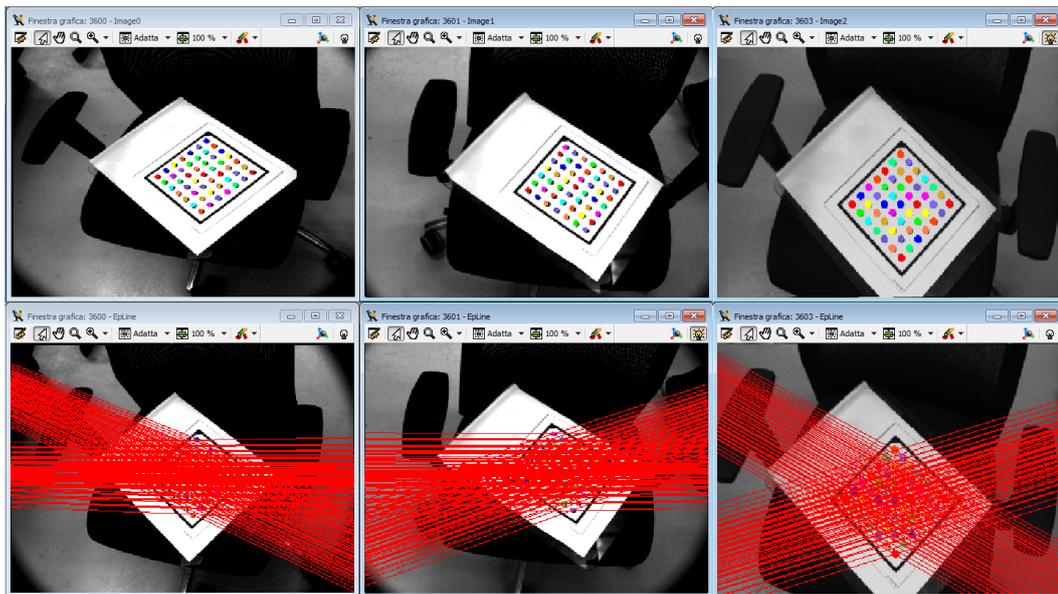
Le coordinate della linea che si ottengono sono espresse in un vettore  $3 \times 1$ , che descrive l'equazione  $l_1x + l_2y + l_3 = 0$ . Può essere riportato all'equazione esplicita  $y = mx + q$  con  $m = -l_1/l_2$  e  $q = -l_3/l_2$ . Tale calcolo può essere eseguito per tutti i punti dell'immagine e per ognuna delle due immagini in relazione all'altra. Si possono quindi creare su ogni immagine una griglia di rette corrispondenti alle epipolari ottenute rispetto ai punti nell'altra immagine e utilizzare tali rette non solo per individuare le corrispondenze tra punti ma anche per stimare le incertezze di ricostruzione sui punti stessi (figura 3.6).



**Figura 3.6:** Esempio di geometria epipolare. Per ogni punto selezionato nella prima immagine sono state tracciate le rette epipolari nella seconda immagine.

Anche se in posizioni totalmente diverse all'interno delle immagini diventa facile tramite questo metodo ricavare la posizione dei punti omologhi ricercandoli solamente nell'intorno delle rette epipolari generate dalla prima immagine. Una volta eseguita l'associazione è possibile eseguire la triangolazione dei punti.

Un limite importante di questa tecnologia proviene dal fatto che, nel caso in cui più marker sono disposti lungo linee parallele alla baseline, s'incorre in casi di ambiguità. Per risolvere questo problema si può ricorrere all'aggiunta di una terza telecamera, aumentando la complessità del sistema di misura ma aggiungendo stabilità al metodo d'individuazione dei punti omologhi basato sulle rette epipolari. Tale costruzione potrebbe essere affrontata introducendo il tensore trifocale, ovvero la matrice duale della fondamentale ma valida per una tripletta di telecamere. Esso non viene però calcolato in questo lavoro di tesi poiché aumenta la complessità del modello del sistema di acquisizione senza portare benefici considerevoli in termini di accuratezza; è ben più semplice e utile vedere il sistema come costituito da tre coppie di telecamere, esprimendo per ognuna di esse la matrice fondamentale rispetto alle altre due singolarmente. Se vengono tracciate le rette epipolari corrispondenti a due punti omologhi sulla terza immagine esse si incontreranno in un punto, che corrisponde al punto che rispetta i vincoli epipolari per entrambe le immagini. Essendo l'obiettivo corrente quello di individuare i punti omologhi senza ambiguità, si può limitare la ricerca dei punti omologhi solamente nei dintorni di tale punto ideale e non lungo un'intera linea epipolare (figura 3.7).



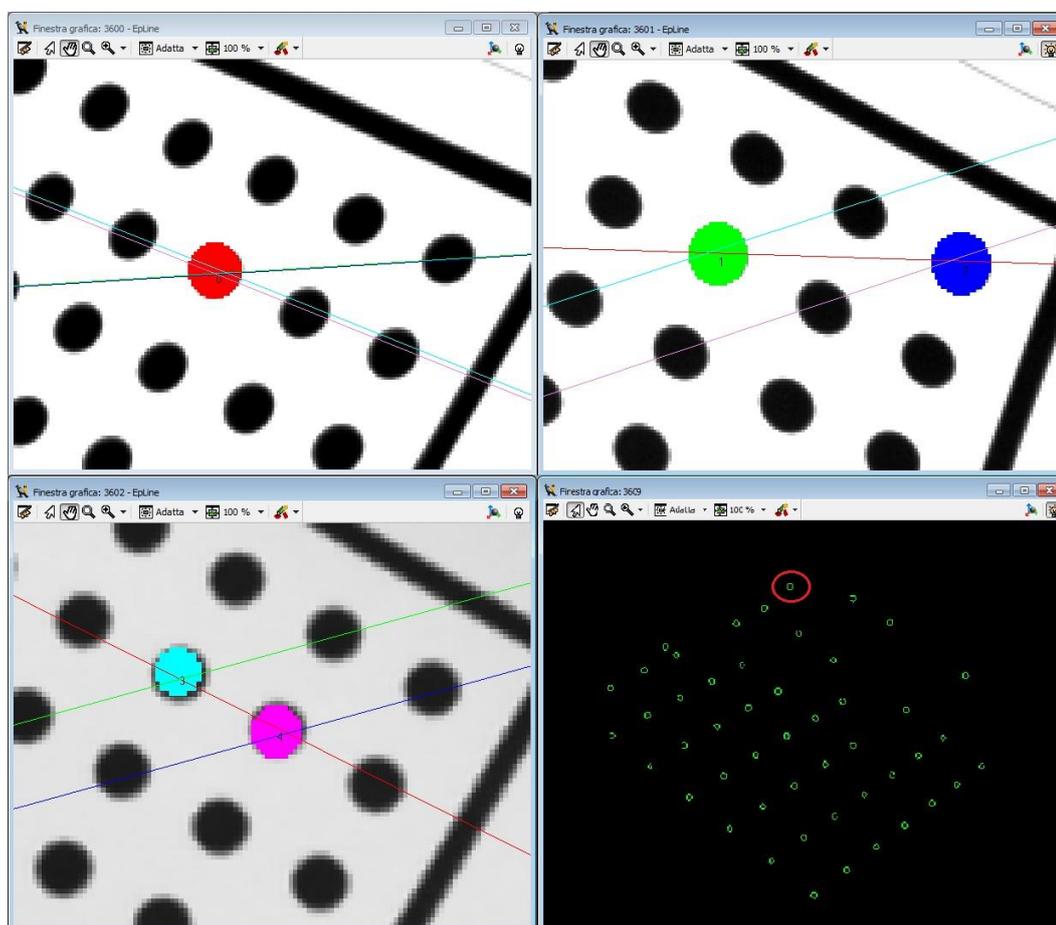
**Figura 3.7:** Esempio di individuazione delle rette epipolari per un calibratore planare mediante l'uso di una tripletta di telecamere. I punti del calibratore sono stati trovati utilizzando l'algoritmo di blob detection.

La procedura completa per eseguire questa individuazione diventa:

1. Per ogni punto nella prima immagine si traccia la retta epipolare nella seconda.
2. Per tutti i punti con distanza da tale retta entro una determinata soglia (*tolleranza*) si traccia la retta epipolare nella terza immagine.
3. Per tutti i punti con distanza da tale retta epipolare entro una determinata soglia si traccia la retta epipolare nella prima immagine.
4. Se la retta epipolare tracciata incontra il punto dal quale è cominciata l'analisi, allora si seleziona la terna di punti.
5. Per ogni punto della terna si calcola la distanza dal punto d'incontro delle due rette epipolari nell'immagine.
6. Viene calcolata la media delle distanze.
7. Per ogni terna si seleziona quella che ha media delle distanze minore e si passa al punto successivo.

Questa procedura che consente di eseguire una ricerca di corrispondenze nei dintorni di un punto, risulta maggiormente mirata, fornisce cioè un numero minore di false associazioni rispetto alla tecnica con due telecamere.

In realtà anche nel caso in cui vengono utilizzate tre telecamere è possibile, anche se meno di frequente, incorrere in circostanze particolarmente sfortunate in cui ci si ha ambiguità. A dimostrazione di questa affermazione viene riportato l'esempio di figura 3.8.



**Figura 3.8:** Esempio di ambiguità nella ricerca dei punti corrispondenti in un sistema stereoscopico con tre telecamere.

La figura mostra un tentativo di ricostruzione di un calibratore planare utilizzando le immagini acquisite con una tripletta di telecamere. Come si nota dall'ultima figura (d) in fase di triangolazione alcuni punti sono stati ricostruiti in maniera errata e risultano infatti esterni al piano del calibratore. Con un'attenta analisi si è notato che, in realtà, l'utilizzo di un sistema stereoscopico a tre telecamere non risolve tutte le ambiguità. Si prenda ad esempio il punto cerchiato in rosso nella figura (d): esso corrisponde rispettivamente ai punto rosso, verde e azzurro nelle tre immagini. Partendo dunque dal punto rosso, e applicando l'algoritmo per la

determinazione delle corrispondenze, viene disegnata la retta epipolare nella seconda immagine (retta in rosso). Come si può vedere, essa incontra il blob verde e quello blu; in realtà ne incontrerà molti di più ma ai fini di questa analisi ci interessano solo questi due. Sappiamo già che il punto verde è esatto e quindi la terna composta dai punti [rosso, verde, azzurro] verrà selezionata. Vogliamo ora analizzare cosa succede se l'algoritmo seleziona l'altro punto (blu) come omologo. In tal caso viene disegnata nella terza immagine la retta epipolare corrispondente, che è rappresentata in blu e incontra il punto in viola. Da questo poi viene tracciata la retta epipolare nella prima immagine che incontra proprio il punto da cui è cominciata l'analisi, ovvero quello rosso. Quindi anche la terna [rosso, blu, viola] verrà considerata. In realtà l'algoritmo funziona analizzando contemporaneamente il caso in cui, dalla prima immagine si disegni la retta epipolare sulla terza e poi sulla seconda, per ritornare alla prima. Si è verificato che anche in questo caso non cambia nulla e entrambe le terne vengono selezionate. A questo punto l'algoritmo seleziona la terna che ha media delle distanze tra il centro dei blob e il punto di incontro delle rette epipolari in quell'immagine. A causa di piccole incertezze di calibrazione, in questo caso la media più piccola è proprio quella della terna errata. Selezionando una corrispondenza di punti errata, la ricostruzione tridimensionale avviene in modo errato. Per questo in figura (d) alcuni punti si trovano abbondantemente fuori dal piano del calibratore. È chiaro che questa circostanza particolarmente critica avviene molto raramente se pensiamo ad esempio all'acquisizione dei marker posizionati su una vela. In questo caso, infatti, si utilizzerà un numero molto inferiore di punti, che saranno più distanziati tra loro e soprattutto non saranno collineari e avranno posizioni meno ordinate e più casuali. Da notare che più il parametro di tolleranza dell'algoritmo di selezione delle corrispondenze è alto e maggiore sarà la probabilità di incorrere in ambiguità. Viceversa, con una tolleranza eccessivamente bassa, molte corrispondenze non vengono trovate e quindi alcuni punti non vengono ricostruiti.

## 4. Sistema di misura

*In questo capitolo viene descritto il sistema di misura realizzato in questa tesi. Prima di tutto è approfondito il funzionamento dei dispositivi che lo compongono e le motivazioni che hanno portato alla loro scelta. Il sistema di misura è costituito da tre telecamere ad alta risoluzione e da tre illuminatori coassiali a led infrarossi. Lo strumento è interamente gestito da un pacchetto software implementato appositamente durante questo lavoro, che consente di acquisire immagini, effettuare l'intera procedura di taratura dello strumento (calibrazione e registrazione delle telecamere) e misurare le coordinate tridimensionali di alcuni punti notevoli nella scena. Inoltre è possibile registrare diverse triplette di telecamere, in modo da poter utilizzare più strumenti di misura di questo tipo contemporaneamente. Nel corso del capitolo viene quindi descritta in dettaglio ognuna di queste procedure e la modalità in cui esse sono state implementate nel pacchetto software.*

## 4.1 Setup del sistema

In questa tesi si sviluppato un sistema di misura ottico stereoscopico, vale a dire basato su telecamere. In commercio esistono numerose tipologie di telecamere digitali. Alcuni parametri e caratteristiche comuni tra i vari modelli sono descritti nel capitolo 4.1.1. La scelta del dispositivo da utilizzare non è banale e dipende fortemente dalle applicazioni. Il criterio di scelta delle telecamere con caratteristiche più adatte al caso in esame è descritta nel capitolo 4.1.2. Oltre al sistema stereoscopico vero e proprio lo strumento di misura è composto da altri dispositivi, quali ad esempio gli illuminatori infrarossi e il loro controller, che sono descritti nei capitoli 4.1.3 e 4.1.4.

### 4.1.1 Parametri delle telecamere

Le telecamere digitali sono sistemi complessi e, nella maggior parte di esse, è possibile agire direttamente su determinate caratteristiche, fra cui le più importanti sono:

- Frequenza di acquisizione (o *frame rate*)
- Modalità di acquisizione (sincrona/asincrona/triggered)
- Definizione di ROI direttamente in fase di acquisizione
- Tempo di esposizione (o *shutter*)
- Formato e compressione delle immagini

Tutti questi parametri devono essere presi in considerazione per la corretta scelta della telecamera da utilizzare nell'applicazione di interesse. Altre caratteristiche importanti per tale scelta sono la tecnologia del sensore (CCD o CMOS), le sue dimensioni e risoluzione e la modalità del trigger. In un sistema di misura con telecamere bisogna ricordare che la frequenza di acquisizione è data quasi esclusivamente dal tempo di acquisizione (più un ritardo dovuto all'elettronica interna del dispositivo). Telecamere con risoluzioni basse, oltre che costare meno, portano solitamente dei vantaggi in termini di tempi di acquisizione. Chiaramente con basse risoluzioni diminuisce l'accuratezza in fase di misura; è dunque necessario trovare un compromesso tra le due caratteristiche.

L'utilizzo di un sistema triggerato permette di acquisire immagini sincrone da diversi dispositivi, *on demand* o con una precisa frequenza di acquisizione, evitando che esse acquisiscano in istanti non sincroni. In caso non si possa utilizzare un trigger per l'acquisizione, è possibile lasciare il dispositivo in modalità *free run* e impostare la frequenza di acquisizione mediante il parametro *frame rate*. Così facendo però non è possibile sincronizzare più dispositivi tra loro. Nel caso in esame il sistema è composto da più telecamere e da illuminatori, ed è quindi molto importante utilizzare un sistema triggerato, poiché le misurazioni vanno eseguite su immagini acquisite in maniera perfettamente sincrona. Le telecamere in esame possono essere comandate sia da un trigger software che da un trigger esterno. Mentre il vantaggio di un segnale generato via software è chiaramente la semplicità, i vantaggi di un segnale esterno generato da hardware sono una maggiore accuratezza e la possibilità di sincronizzare altri dispositivi, come ad esempio illuminatori. Nel nostro caso il segnale di trigger viene generato da una scheda Pokeys 55 (capitolo 4.1.4) e viene inviato a tutti i dispositivi connessi, ovvero telecamere e illuminatori. Anche nel caso in cui il sistema sia triggerato, questo tipo di telecamere possono lavorare in maniera sincrona o asincrona. Nel caso di acquisizione triggerata sincrona (il cui diagramma di acquisizione è rappresentato in figura 4.1) l'arrivo del trigger non fa scattare l'acquisizione che è invece legata al successivo trigger interno alla telecamera. Si osserva inoltre come nuovi impulsi di trigger non abbiano effetto se giungono durante il tempo di esposizione. Il caso asincrono è solitamente quello più usato poiché permette di lasciare al trigger il compito di coordinare l'intero processo di acquisizione. In figura 4.2 è mostrato il diagramma temporale di tale schema: l'esposizione può iniziare dopo un ritardo impostabile dall'arrivo del trigger per permettere eventualmente di sincronizzare l'acquisizione con illuminatori o altri dispositivi.

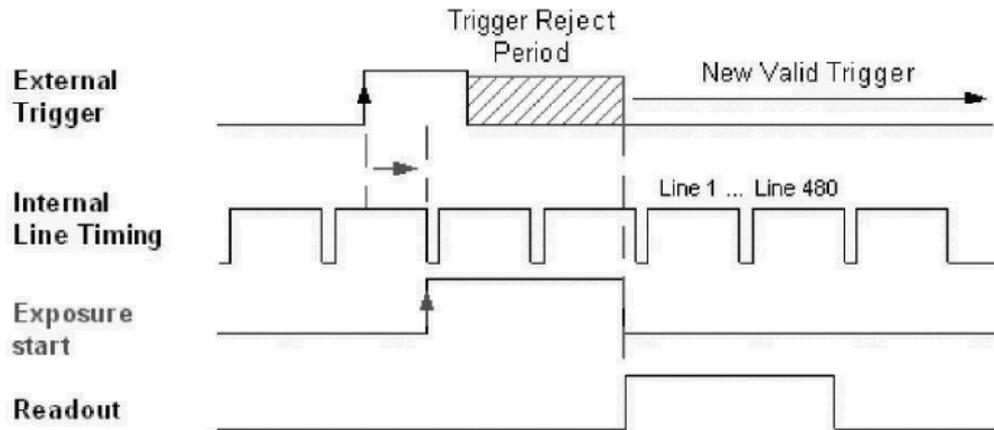


Figura 4.1: Diagramma temporale di un'acquisizione triggerata sincrona.

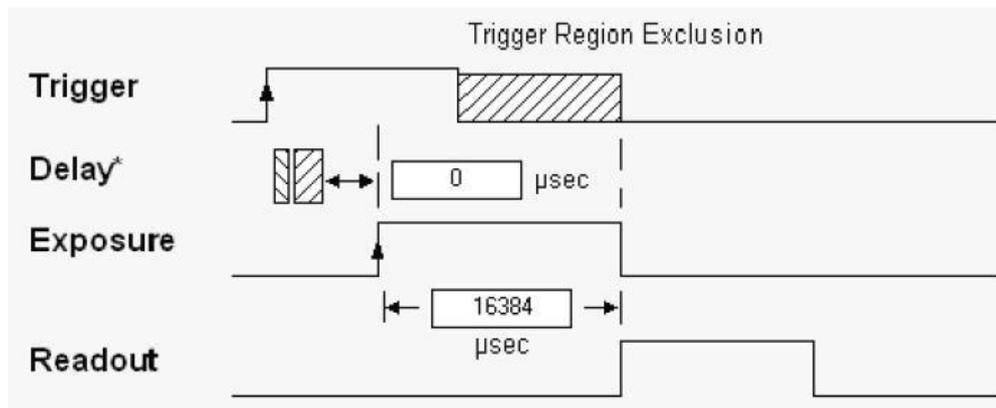


Figura 4.2: Diagramma temporale di un'acquisizione triggerata asincrona.

Il *tempo di esposizione* (o *shutter*), vale a dire è il tempo durante il quale l'otturatore della telecamera rimane aperto per permettere alla luce di raggiungere il sensore, è il principale fattore che limita superiormente la frequenza di acquisizione della telecamera. Questo parametro influenza la luminosità dell'immagine acquisita; aumentando il tempo di esposizione l'immagine è più luminosa. Tale caratteristica non è contemplata dal modello *pinhole* (capitolo 2.1.2); in esso infatti il passaggio della luce attraverso un foro puntiforme le garantisce una corretta messa a fuoco. Nella realtà invece per ogni impostazione dell'obiettivo (lunghezza focale e apertura del diaframma fissati), esiste un'unica

distanza a cui gli oggetti appaiono perfettamente a fuoco; la nitidezza diminuisce gradualmente spostandosi in avanti (verso il dispositivo) e indietro. La *profondità di campo* è l'intervallo di distanze davanti e dietro al soggetto in cui la sfocatura è impercettibile o comunque tollerabile.

Un buon metodo per aumentare la frequenza di acquisizione della telecamera è la definizione di una *ROI* (Region of interest) che limita la regione di sensore effettivamente acquisita. In tale modo sia l'acquisizione che le operazioni successive risultano molto più rapide. Chiaramente però si perde volume di misura, riducendo la quantità di dati che possono essere analizzati. Tale possibilità viene fornita solitamente per i soli dispositivi digitali basati sui CMOS, nei quali ogni pixel è dotato dell'elettronica di condizionamento del segnale. Contrariamente, i dispositivi a CCD eseguono la scansione sequenziale di tutte le linee che compongono il sensore. In realtà anche con quest'ultimo tipo di sensore è possibile ricavare una ROI ma con vantaggi notevolmente inferiori in termini di diminuzione di informazioni acquisite.

La frequenza di acquisizione effettiva imposta al dispositivo tiene in considerazione i tempi di acquisizione, quelli di elaborazione e il trasferimento delle immagini. E' quindi necessario che la scheda di acquisizione sia in grado di sopportare il carico di informazioni inviate dalla telecamere senza rallentamenti o perdita di informazioni. Questo obiettivo non è sempre facilmente raggiungibile, soprattutto nel caso in cui la risoluzione delle immagini o la frequenza di acquisizione sono molto elevate.

Il tempo di ritardo totale per un sistema di acquisizione comandato da trigger potrà quindi essere calcolato come:

$$t_{delay\ telec.} = t_{delay\ trigger} + t_{esposiz.} + t_{preprocessing} + t_{trasferimento} \quad (4.1)$$

È chiaro che in un sistema stereoscopico è di fondamentale importanza che le telecamere acquisiscano in maniera sincrona. Per ottenere ciò, però, non è sufficiente utilizzare un trigger esterno per tutti i dispositivi. Infatti bisogna prestare molta attenzione anche in fase di sviluppo del software per l'acquisizione ed è necessario aiutarsi con la programmazione multithread. Il trigger è infatti un

segnale fisico esterno a una frequenza fissa e comune a tutti i dispositivi; può accadere che il comando di acquisizione sia dato via software in istanti differenti. In questo modo, anche se il sistema è triggerato, si rischierebbe di acquisire immagini in istanti diversi per ogni dispositivo. Ogni telecamera è dunque gestita da un thread separato e sincronizzato con i thread relativi agli altri dispositivi, in modo tale che il comando di acquisizione sia inviato nello stesso istante a tutti i dispositivi. Se non si utilizzasse la programmazione multithread, infatti, il comando sarebbe inviato in maniera sequenziale ai vari dispositivi, che quindi acquisirebbero in istanti diversi tra loro. Per testare il sincronismo in fase di acquisizione è stato sperimentato il sistema acquisendo immagini di un cronometro. Così facendo si è verificato che le telecamere acquisiscono le rispettive immagini nello stesso istante e alla frequenza stabilita.



**Figura 4.3:** Alcune immagini acquisite durante il test del cronometro. Si nota come le immagini delle due telecamere sono acquisite in maniera sincrona.

## 4.1.2 Scelta delle telecamere

Il criterio di scelta del dispositivo più adatto dipende fortemente dalle applicazioni in cui esso viene utilizzato. Nel sistema stereoscopico in esame si è deciso di utilizzare telecamere Basler Scout scA1390-17gm (figura 4.5), le quali specifiche principali sono riassunte in figura 4.4.

Basic Information	
Vendor	Basler
Model	scA1390-17gm
Type of data presented	Typical
Number of samples	100
Sensor	ICX267AL
Sensor type	CCD
Sensor diagonal	Diagonal 8 mm (Type 1/2)
Indication of lens category to be used	C-Mount
Resolution	1392 x 1040 pixel
Pixel width	4.65 $\mu\text{m}$
Pixel height	4.65 $\mu\text{m}$
Readout type	Progressive scan
Transfer type	Interline transfer
Shutter type	-
Overlap capabilities	
Maximum frame rate	17 frames/second
General conventions	
Interface type	Gigabit Ethernet

**Figura 4.4:** Specifiche principali della telecamera Basler Scout scA1390-17gm utilizzata in questo lavoro.



**Figura 4.5:** Telecamera Basler Scout scA1390-17gm, il modello utilizzato nel sistema stereoscopico in esame.

Per rispettare le specifiche dello strumento di misura è necessario utilizzare una risoluzione elevata per quanto riguarda le telecamere industriali. Il dispositivo scelto, infatti, presenta un sensore di dimensioni 1392x1040 pixel. Chiaramente immagini ad alta risoluzione necessitano più tempo per essere trasferite e occupano più spazio fisico in memoria. Per il formato immagine utilizzato dal dispositivo in esame ogni pixel occupa 8 bit (1 byte) ; un'intera immagine dunque occuperà circa 1.45 MB. Scegliendo una risoluzione così elevata si è dunque data meno importanza alla frequenza di acquisizione. Nel dispositivo in esame il *frame rate* massimo è infatti 17 Hz se non viene impostata alcuna ROI, ma può aumentare a seconda delle dimensioni della ROI utilizzata. Questo valore è più che sufficiente, poiché per le applicazioni in cui verrà usato lo strumento si richiede una frequenza di acquisizione massima intorno ai 10 Hz. Nel caso in cui è necessario avere un frame rate superiore, è sempre possibile diminuire la ROI aumentando la frequenza di acquisizione.

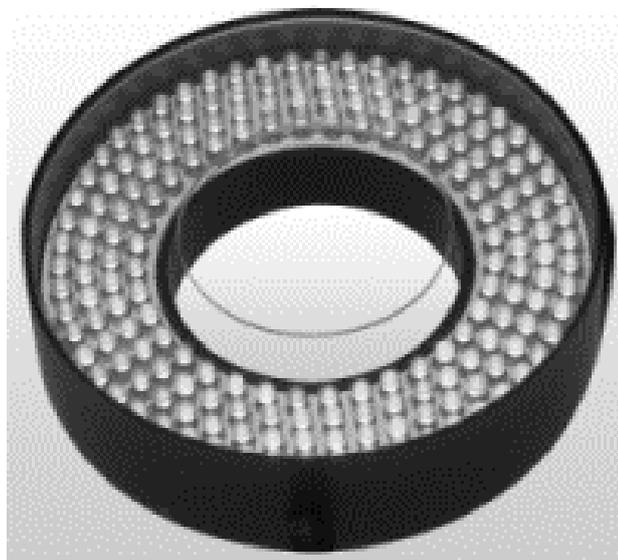
La telecamera permette di acquisire immagini in *freerun* o in modalità triggerata, permettendo di controllare l'acquisizione con un trigger software o esterno. Nella scelta della telecamera è indispensabile verificare che essa presenti la possibilità di utilizzare un trigger hardware esterno, poiché questo è il metodo più semplice e sicuro per poter sincronizzare più dispositivi differenti (telecamere e illuminatori). La telecamera può essere comandata da un segnale di trigger

esterno di ampiezza compresa tra 2 e 12 V. Inoltre la telecamera consente di acquisire in modalità sincrona o asincrona come descritto nel capitolo 4.1.1.

La scelta di utilizzare un dispositivo che comunica con il computer attraverso una porta Gigabit Ethernet è dettata principalmente dal fatto che essa consente di utilizzare cavi molto più lunghi rispetto ad altre tecnologie. Questa caratteristica è molto importante soprattutto nelle applicazioni in Galleria del Vento, dove lo strumento viene posizionato in sala prove mentre il computer è collocato in sala controlli. In questo caso è dunque necessario utilizzare un cavo lungo (10-15 m) per la comunicazione tra telecamere e computer.

### **4.1.3 Sistema d'illuminazione**

Più l'ambiente è controllato nelle sue caratteristiche più le prove e le misure effettuate saranno ripetibili e meno soggette a errori. Anche se il sistema di misura può essere usato in condizioni di luce naturale, in alcune applicazioni può essere richiesta un'accuratezza molto elevata ed è dunque opportuno utilizzare un sistema di illuminazione controllata. La progettazione di tali sistemi è in genere complessa e deve essere studiata per la specifica applicazione. Nel caso in esame è però richiesto che il sistema di illuminazione possa essere utilizzato in svariate applicazioni differenti. Si è dunque deciso di utilizzare tre illuminatori infrarossi coassiali rispetto l'ottica delle telecamere e di posizionare marker catarifrangenti sull'oggetto da misurare. Viene utilizzato un filtro interferenziale centrato sulla lunghezza d'onda dell'illuminatore in modo tale da tagliare le lunghezze d'onda indesiderate. Esso viene posto sull'ottica delle telecamere così da escludere il campo del visibile mettendo in evidenza solamente i marker posti sul misurando. Questa tecnica consente prima di tutto di limitare notevolmente i disturbi causati dalla luce naturale che possono rendere più difficile l'individuazione dei marker. Inoltre questo metodo può essere utilizzato in qualunque applicazione, applicando al misurando un numero di marker opportuno. Il numero e la posizione di marker deve essere scelta in modo intelligente e dipende fortemente dalla struttura del misurando e da ciò che si vuole misurare.



**Figura 4.6:** *Illuminatore circolare BMT.*



**Figura 4.7:** *Controllore Gardasoft PP420, utilizzato per la gestione degli illuminatori.*

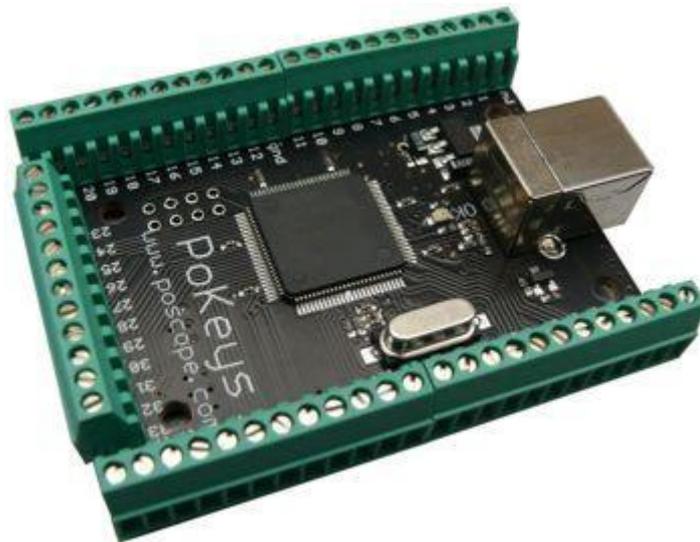
Il sistema di misura è dunque composto da tre illuminatori coassiali BMT modello BDRL-IR110/58 a lunghezza d'onda di 850 nm (figura 4.6), controllati da Gardasoft PP420 (figura 4.7) che consente di gestire separatamente i tre illuminatori. In particolare il controllore, comandato via Ethernet, permette di utilizzare gli illuminatori in maniera continua, con un trigger software o con un trigger esterno. Utilizzando i tre illuminatori in modalità trigger esterno, con lo stesso trigger che comanda anche l'acquisizione delle telecamere si possono

sincronizzare i sei dispositivi. Inoltre il controllore consente di impostare il livello di luminosità e l'intervallo in cui l'illuminatore rimane acceso dopo l'arrivo del segnale di trigger.

Inoltre vengono utilizzati sulle ottiche filtri interferenziali passa banda con banda passante intorno a 850 nm. In questo modo si limitano i disturbi, eliminando la luce a lunghezza d'onda differente da quella degli illuminatori.

#### 4.1.4 Altri componenti hardware

Per sincronizzare le tre telecamere e i rispettivi illuminatori è stata utilizzata una scheda Pokeys 55, rappresentata in figura 4.8. Essa infatti permette di generare un segnale PWM a 3.3 V. Variando l'ampiezza del periodo e impostando il *duty cycle* al 50% di esso è quindi possibile generare un segnale di trigger alla frequenza desiderata. Questo segnale viene inviato alle telecamere e agli illuminatori, così che essi siano tutti comandati dallo stesso segnale e dunque perfettamente sincronizzati tra di loro. La scheda comunica con il computer attraverso l'interfaccia USB.



**Figura 4.8:** Scheda Pokeys55 utilizzata per la generazione del segnale di trigger.

Il computer a cui lo strumento è stato collegato è fornito di due differenti schede Gigabit Ethernet, perciò si è deciso di utilizzarne una per la comunicazione con il controllore degli illuminatori e una per la comunicazione con le tre telecamere.

Per fare ciò si è dovuti ricorrere a uno switch Gigabit Ethernet, in particolare è stato utilizzato e verificato il NETGEAR ProSafe Plus GS108 a 8 porte (figura 4.9). È stato calcolato che il fatto di utilizzare uno switch Gigabit Ethernet non compromette la velocità di acquisizione. Infatti la Gigabit Ethernet trasferisce dati alla velocità di 1 Gigabit ( $10^9$  bit) al secondo, che corrispondono a 125 MB al secondo. Come visto nel capitolo 4.1.2 le immagini acquisite dai dispositivi sono di circa 1.45 MB ed utilizzando tre telecamere l'informazione che si trasmette al computer ad ogni singola acquisizione è 4.35 MB. La frequenza massima a cui si può acquisire con queste condizioni sarà dunque:

$$\frac{125 \frac{MB}{s}}{4.35 MB} \cong 28 Hz \quad (4.2)$$

Essendo il *frame rate* massimo dei dispositivi pari a 17 Hz, è evidente che anche utilizzando uno switch Gigabit Ethernet non viene saturata la capacità di trasmissione dati e dunque non vengono introdotti ritardi o errori in fase di comunicazione tra il computer e i dispositivi.



**Figura 4.9:** Switch Gigabit Ethernet NETGEAR ProSafe Plus GS108 a 8 porte utilizzato per la comunicazione tra il computer e le tre telecamere.

## 4.2 Descrizione del software

L'applicativo software è stato sviluppato nell'ambiente di sviluppo Visual Studio 2010 della Microsoft. In particolar modo è stato scelto il linguaggio C# con il framework .Net 4.0, soprattutto perché permette di utilizzare la programmazione ad oggetti che per sua natura è molto versatile, flessibile e consente di gestire ottimamente gli eventi che si susseguono all'interno del programma. Inoltre è da sottolineare la facile mantenibilità del software, la modularità, che consente di scomporre il problema in più semplici sotto-problemi, e la possibilità di riutilizzare facilmente il codice in altri ambiti. Nella potente piattaforma .Net vi sono moltissime librerie che consentono di effettuare parte delle applicazioni ordinarie. Inoltre è un'ottima piattaforma di programmazione poiché consente di integrare facilmente librerie prodotte da terze parti.

Si è deciso di integrare la programmazione in C# con la libreria di image processing HALCON, realizzata da MVTEC. Questa possiede un proprio ambiente di sviluppo, chiamato HDevelop, ed è integrabile con i più comuni linguaggi di programmazione: C, C++, C#, Visual Basic .Net e Delphi. Alcune applicazioni sono quindi state implementate in HDevelop e poi esportate in linguaggio C#, mentre altre più semplici sono state scritte direttamente in C# utilizzando comandi provenienti dalla libreria HALCON.

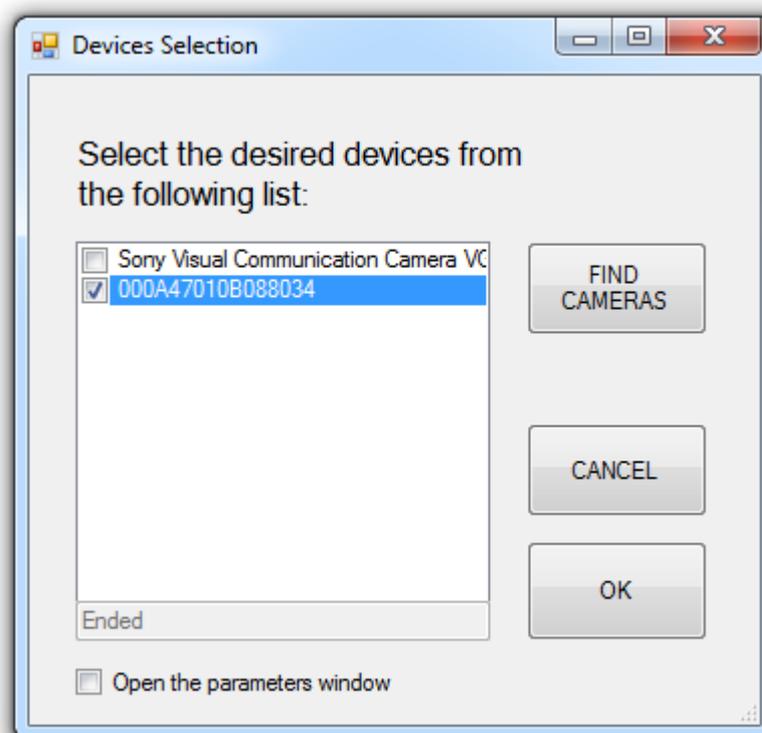
Il problema della creazione di un software complesso da associare al sistema hardware descritto nel capitolo 4.1 è stato approcciato, come spesso viene fatto in questi casi, scomponendolo in sottoproblemi:

- Acquisizione di immagini da diversi dispositivi e con diverse modalità
- Calibrazione delle triplette
- Registrazione di più triplette
- Triangolazione online

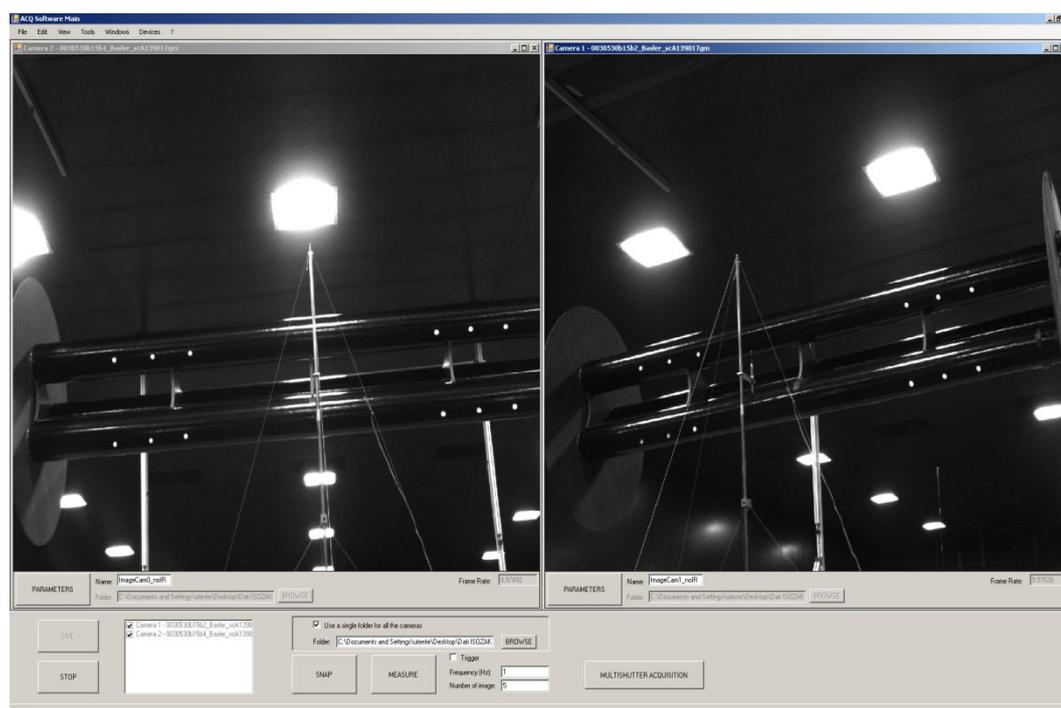
Di seguito si analizzerà più in dettaglio la risoluzione di questi problemi dal punto di vista teorico e pratico.

## 4.2.1 Acquisizione

Il software deve prima di tutto consentire alle diverse telecamere di acquisire e salvare immagini. Una delle principali caratteristiche che si è voluto dare al software è un'elevata flessibilità e scalabilità, vale a dire che esso possa funzionare con diverse telecamere. Perciò quando viene avviato il software, esso deve controllare se vi sono dispositivi collegati a qualunque porta del computer e visualizzarli in un elenco. Come mostrato in figura 4.10, l'utente ha poi la possibilità di selezionare quali telecamere utilizzare tra quelle trovate. Per ogni dispositivo selezionato viene quindi aperta una finestra nel programma principale.



**Figura 4.10:** *Interfaccia per la selezione dei dispositivi che si vogliono utilizzare nel software.*



**Figura 4.11:** *Interfaccia principale del software di acquisizione sviluppato in questa tesi.*

Nella figura 4.10, ad esempio, è rappresentata l'interfaccia principale, o *main*, nel caso in cui sono selezionate due telecamere. Si nota che essa è composta dunque da due sottofinestre che contengono entrambe l'immagine acquisita dal singolo dispositivo, un tasto "Parameters" che consente di aprire una finestra per impostare i parametri della telecamera, e due caselle per definire il percorso in cui verranno salvate le immagini relative al dato dispositivo e il prefisso che assumeranno, che verrà seguito dal numero dell'immagine acquisita. Infine, in basso a destra viene mostrata la frequenza di acquisizione relativa alla telecamera. In basso, sotto alle due sottofinestre, è presente il pannello dei comandi. Con esso è possibile selezionare con quale tra i dispositivi aperti nel main si vuole effettuare l'acquisizione, cominciare e concludere la visualizzazione a schermo delle immagini dai dispositivi selezionati, attivare un trigger esterno, effettuare un fermo immagine ("Snap") o salvare una serie di immagini a una frequenza stabilita (impostabile via software se il trigger esterno è disattivato) con il tasto "Measure". È anche possibile selezionare un percorso comune a tutti i dispositivi in cui verranno salvate le immagini acquisite. Infine è presente il tasto

“Multishutter acquisition”, che consente di impostare un’acquisizione multishutter, utile per un’altro progetto sviluppato presso il VBLab e descritto nel capitolo 4.2.2. Dal menu a tendina è possibile anche modificare l’interfaccia principale in modo da avere un pannello di comandi per ogni dispositivo e non più uno comune a tutti, così da poter gestire ogni telecamera singolarmente. Inoltre è possibile aprire l’interfaccia per la calibrazione delle telecamere descritta nel capitolo 4.3 e quella per la ricostruzione tridimensionale (capitolo 4.4).

## **4.2.2 Acquisizione immagini termiche**

Il software di acquisizione descritto nel capitolo 4.2.1 è stato utilizzato, oltre che per lo sviluppo di uno strumento di misura ottico stereoscopico, anche per un altro progetto realizzato presso il VBLab [28]. Questo riguarda la creazione di uno strumento di misura che consente di acquisire immagini termiche attraverso telecamere dotate di filtro infrarossi.

Tale strumento consente dunque di ricavare immagine termografiche di oggetti con temperature comprese tra i 500 e i 1130 °C. Con il termine termografia s’intende la capacità di ricavare la temperatura di un oggetto tramite la misura della radiazione da esso irraggiata. L’energia termica caratterizzata dall’infrarosso consiste in radiazione elettromagnetica la cui lunghezza d’onda ha un valore troppo elevato per poter essere apprezzato dall’occhio umano: essa occupa infatti quella parte di spettro elettromagnetico da noi percepita come calore. Essistono in commercio strumenti sensibili anche nell’infrarosso che rappresentano un accurato strumento di misurazione senza contatto delle temperature, vale a dire le termocamere. Esse rilevano la radiazione infrarossa incidente e la convertono in un segnale elettronico, che viene in seguito elaborato al fine di valutare il valore di temperatura. In questo progetto si vuole però creare uno strumento che possa eliminare i principali svantaggi legati all’acquisizione con termocamere, pur avendo una riduzione dell’accuratezza di misura. Le termocamere presentano infatti alcuni svantaggi, primo tra tutti un costo elevato dello strumento. Molte volte, quindi, le aziende vengono scoraggiate dai costi che

talvolta risultano esagerati in relazione al lavoro che deve essere svolto. Naturalmente se un'azienda ha la necessità di compiere misurazioni che richiedono ampi range di temperatura e alta sensibilità, l'utilizzo di una termocamera è fondamentale e il costo dell'apparecchiatura è commisurato all'elevata qualità dello strumento. Non in tutti i casi, però, la termocamera potrebbe risultare la scelta più adeguata. Si pensi ad esempio ad un'azienda che ha la necessità di effettuare misurazioni su organi meccanici sottoposti a temperature comprese tra 900°C e 1000°C e a cui interessa conoscere in quale punto di un componente meccanico sia la temperatura massima e il suo valore, senza però necessità di avere risoluzione termica e sensibilità elevate. Per avere una mappatura termica del componente meccanico essa sarebbe costretta a comprare una termo camera, con un investimento eccessivo rispetto al campo applicativo in cui essa viene utilizzata.

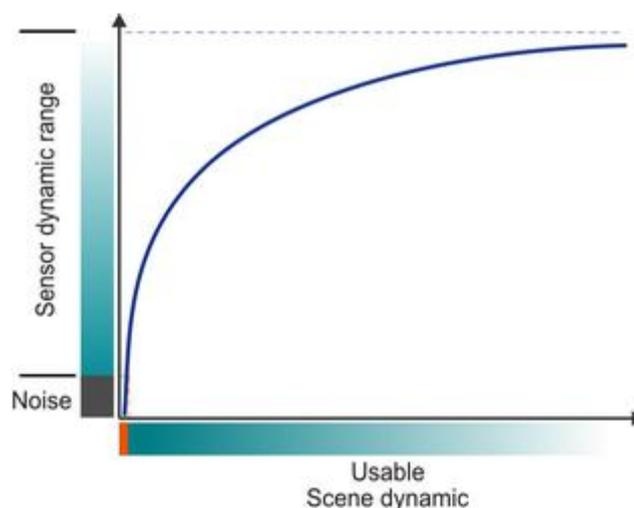
Si è voluto dunque creare uno strumento di misura di temperature che si basasse sull'utilizzo di telecamere industriali. In questo modo lo strumento ha un costo sicuramente più basso rispetto alle termo camere, dovuto alla diversa tecnologia di cui dispongono. Le telecamere infatti non sono provviste di sensori termici o fotonici per la determinazione del calore e quindi risultano essere più semplici e meno delicate. L'immagine fornita da una telecamera dipende da valori di luminosità acquisita, che vengono tradotti, a video, in scala di grigio in relazione al valore stesso di luminosità. Inoltre lo strumento realizzato presenta una risoluzione spaziale molto elevata (500-1130 °C) rispetto alle normali termocamere. Una termo camera con una risoluzione spaziale paragonabile avrebbe un costo sul mercato almeno dieci volte superiore.

Chiaramente l'utilizzo di telecamere industriali presenta alcuni svantaggi dovuti al fatto che la loro tecnologia non è stata sviluppata per questo particolare tipo di misurazioni. In particolare il più grosso limite è sicuramente un limitato campo di misura, poiché le telecamere presentano sensori sensibili solo a lunghezze d'onda che ricadono nello spettro del vicino infrarosso e del visibile. Le immagini di una telecamera vengono composte dai pixel, ciascuno dei quali caratterizzato da un valore dipendente dal numero di bit caratteristico (per

esempio tra 0 e 255 per immagini a 8 bit o tra 0 e 4095 per immagini a 12 bit) e che ne indica l'intensità luminosa in scala di grigi. La variazione di intensità luminosa visibile causata da un innalzamento o abbassamento di temperatura è apprezzabile solo per valori superiori a circa 750°C. Le termocamere, invece, hanno un sensore sensibile all'infrarosso (non solo al vicino infrarosso) e al visibile. Per questo motivo è possibile effettuare misurazioni di temperature che nel visibile non produrrebbero variazioni di intensità.

Vengono utilizzati due differenti tipi di telecamera. La prima è una AVT MARLIN (Allied Vision Technologies) 131-F con sensore CMOS NIR: I sensori CMOS della telecamera sono di tipo NIR (Near Infrared). Questi sensori permettono l'acquisizione anche di lunghezze d'onda nel vicino infrarosso, tipicamente per valori compresi tra 0.8  $\mu\text{m}$  e 2.5  $\mu\text{m}$ . Questo permette di misurare temperatura comprese fra i 500°C e i 750°C. La radiazione elettromagnetica incidente infatti è in relazione lineare con la carica elettrica sviluppata in risposta dalla telecamera. La tecnologia di acquisizione si basa su un processo integrativo nel tempo, per il quale maggiore è il tempo di esposizione, maggiore è la carica elettrica (e quindi l'intensità luminosa del pixel) ottenuta in risposta dalla telecamera. Un corpo alle temperature del range di interesse è caratterizzato dall'aver valori di intensità luminosa molto differenti. La linearità dello strumento comporta il fatto che per diversi intervalli di intensità luminosa (e quindi di radiazione incidente), occorrono tempi di esposizione differenti. In particolare, per valori bassi di intensità luminosa (caratteristici di temperature più basse) occorreranno tempi di esposizione del sensore più elevati in quanto servirà una quantità di radiazione incidente superiore. Al tempo stesso, un tempo elevato di esposizione del sensore penalizza i valori di intensità luminosa più alti, fino al verificarsi del fenomeno della saturazione (il valore reale di intensità luminosa esce dal fondo scala dello strumento e viene, quindi, "perso"). Per questo l'acquisizione delle immagini viene effettuata mediante una procedura a *multi-shutter*. Per la stessa scena acquisita a shutter differenti si hanno diversi valori di intensità luminosa, in quanto l'obiettivo della telecamera rimane aperto per tempi differenti. In conseguenza a questo si ottiene, per ogni shutter, un differente

diagramma di taratura. La seconda telecamera utilizzata invece è una IDS GigE uEye SE con sensore CMOS HDR: la telecamera è caratterizzata dall'aver un diagramma di risposta logaritmico, come quello mostrato in figura xxx. Il comportamento del sensore CMOS HDR montato sulla telecamera ha una tecnologia differente rispetto a quello della AVT Marlin: generalmente, mentre il fotodiode genera una corrente proporzionale alla radiazione elettromagnetica (Marlin), il sensore HDR genera una differenza di potenziale con andamento logaritmico in relazione alla radiazione elettromagnetica "istantanea", proprio come se fosse una piccola cella solare. Non essendoci un processo integrativo nel tempo l'acquisizione delle immagini *non* viene effettuata mediante un sistema multi-shutter, in quanto lo strumento è in grado di elaborare allo stesso tempo aree con valori di luminosità molto differenti. Si ha così il vantaggio di poter sfruttare un solo diagramma di taratura, a scapito di un maggior costo dello strumento.



**Figura 4.12:** Caratteristica logaritmica di un sensore HDR

Partendo dal software di acquisizione descritto nel capitolo 4.2.1 è stato sviluppato, durante questa tesi, un software che consente di ricavare l'immagine termica per entrambi i dispositivi sopra descritti. L'acquisizione multi-shutter è dunque necessaria per effettuare la procedura di taratura della telecamera AVT Marlin. L'interfaccia che consente di impostare e far partire l'acquisizione multishutter è mostrata in figura 4.12.

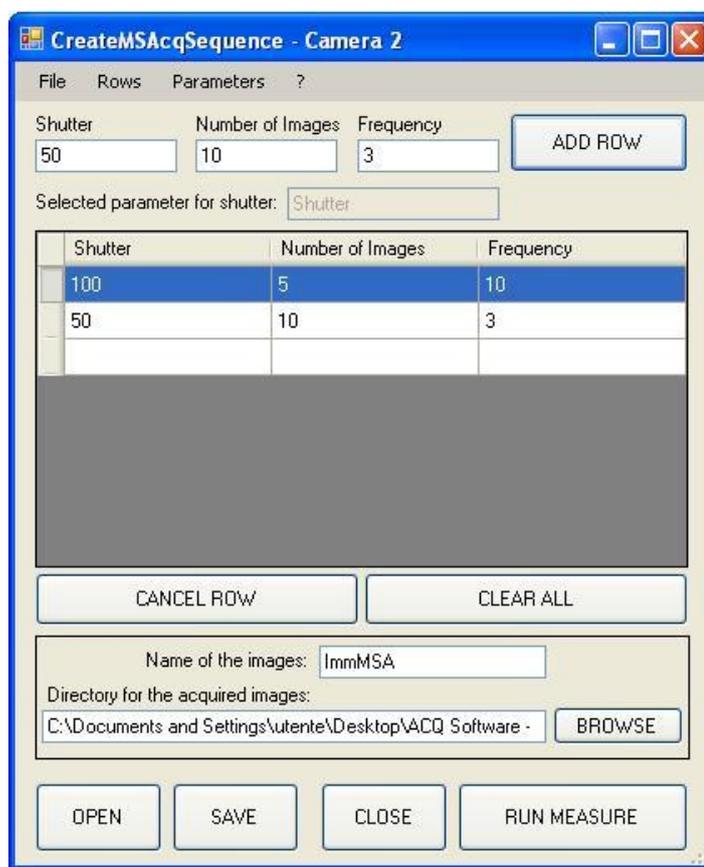


Figura 4.13: Interfaccia per l'impostazione dell'acquisizione multishutter.

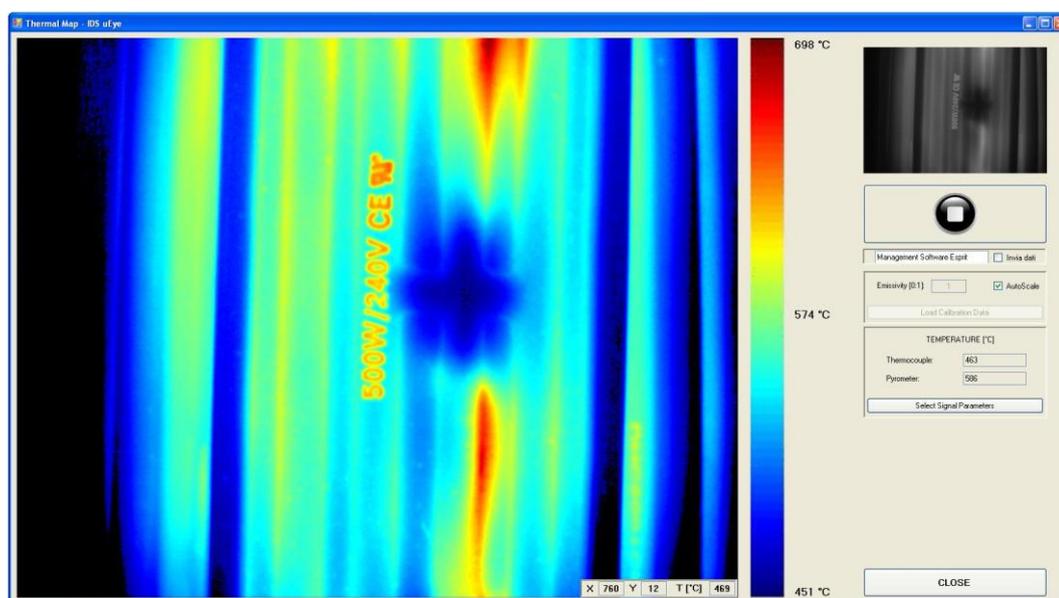


Figura 4.14: Interfaccia di acquisizione della mappa termica.

Dall'interfaccia principale è possibile aprire un'ulteriore finestra che consente di mostrare a schermo la mappa termica acquisita da un singolo dispositivo alla volta. Essa è mostrata in figura 4.13. Solo nel caso in cui viene utilizzata la telecamera Marlin è anche possibile impostare i valori di shutter che sono utilizzati nell'acquisizione.

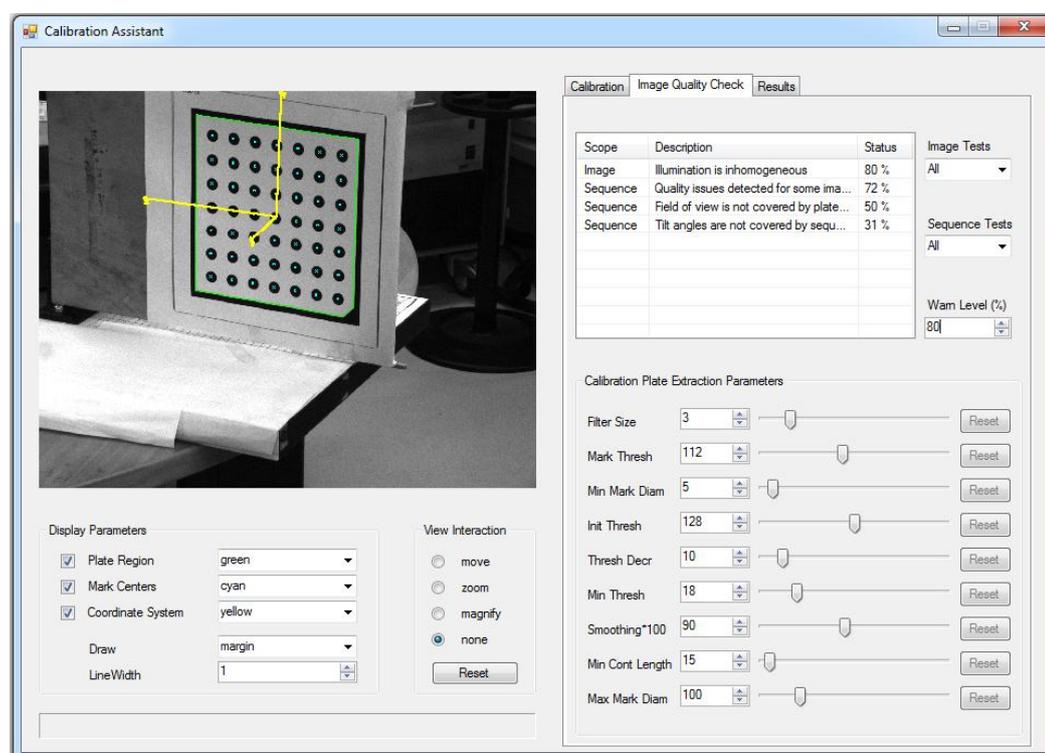
Il software permette di acquisire immagini da una delle due telecamere, selezionabile dall'interfaccia principale. L'immagine acquisita (o la somma delle immagini nel caso di acquisizione multishutter) viene trasformata in una matrice che contiene i valori di grigi per ogni pixel. Tale matrice viene poi convertita, secondo la mappatura trovata in fase di taratura in una matrice che contiene il valore di temperatura corrispondente per ogni pixel. La matrice dei dati di temperatura è poi mostrata a schermo come immagine a colori secondo due diverse modalità selezionabili dall'utente. Nella modalità normale viene associato il colore più in basso nella scala cromatica (blu scuro) alla temperatura di 500 °C e quello più alto (rosso scuro) alla temperatura di 1130°C. Nella modalità "autoscale" invece viene associato il colore blu scuro alla temperatura più bassa della matrice delle temperature, mentre quello rosso scuro è associato al valore più alto di essa. In ogni caso vengono rappresentati pixel neri laddove il range di temperatura è esterno al quello dello strumento (500-1130°C). Oltre alle telecamere, sono stati aggiunti allo strumento un pirometro e una termocoppia che consentono di verificare che le misure che si stanno effettuando sono coerenti. Il software mostra in realtime il valore di temperatura misurato dai due strumenti di misura. L'interfaccia grafica di tale software è mostrata in figura 4.14.

## 4.3 Taratura dello strumento

Un requisito necessario per effettuare misurazioni tridimensionali è che il sistema di misura sia tarato. Per la taratura della tripletta di telecamere si è deciso di sviluppare un tool in C#, utilizzando la libreria HALCON. Il metodo di calibrazione utilizzato è stato dunque quello proposto da HALCON, che è stato validato e testato con numerosi esperimenti (capitolo 2.4.5), con i quali si è verificato che l'accuratezza di tale metodo è compatibile con quella di Zhang-LM, anche se esso presenta alcune limitazioni. Nel capitolo 2.5 è stato sottolineato come tali limitazioni siano ininfluenti per le applicazioni in cui si colloca il sistema di misura in esame. La scelta di utilizzare l'algoritmo di HALCON è stata dettata da due principali motivi. Prima di tutto si è voluto creare un software utilizzabile anche da un utente che non conosca nel dettaglio il processo di calibrazione. HALCON risulta semplice e immediato da utilizzare, essendo un software commerciale creato anche per utenti meno esperti. Inoltre si vuole inserire la procedura di taratura in un pacchetto software unico, utilizzando il linguaggio di programmazione C# e la libreria HALCON. È vero che la taratura della tripletta viene fatta offline a monte delle acquisizioni, ma risulta comunque più comodo e ordinato utilizzare lo stesso linguaggio di programmazione per tutto il pacchetto software. In casi in cui è richiesta un'accuratezza molto elevata è possibile verificare i risultati di calibrazione del metodo di HALCON, calibrando ogni singola telecamera con l'algoritmo di Zhang-LM, implementato in MatLab. Come è stato dimostrato nel capitolo 2.4.5, esso può essere considerato un valido metodo di calibrazione, che porta generalmente a risultati molto accurati. Un'ulteriore motivazione che ha influenzato tale scelta è che il calcolo delle corrispondenze tra i punti del modello del calibratore e i punti immagine in HALCON è effettuata in maniera automatica. Infine HALCON è particolarmente comodo per la taratura di un sistema stereoscopico composto da più dispositivi. Infatti, oltre alla calibrazione di telecamere singole descritta nel capitolo 2.4.5, è possibile eseguire facilmente la taratura di un sistema stereoscopico formato da più dispositivi. È sufficiente selezionare il numero di telecamere che lo

compongono e per ognuna di esse acquisire un set di immagini in maniera sincrona con le altre. L'algoritmo di taratura di un sistema stereoscopico di HALCON consente di trovare i parametri intrinseci ed estrinseci di tutte le telecamere del sistema. I parametri estrinseci sono riferiti a uno dei dispositivi del sistema stereoscopico, selezionato dall'utente. La procedura di taratura di una tripletta restituisce un valore di errore di retroproiezione che non sarà più riferito alla singola telecamera, bensì all'intera tripletta.

È stato quindi implementato un tool, la cui interfaccia è mostrata in figura 4.15, che consente di acquisire, tramite i tre dispositivi, una serie di immagini di un calibratore in diverse pose. Le immagini sono acquisite in maniera sincrona per ogni dispositivo e, se il calibratore viene trovato con successo in tutte e tre le immagini, esse vengono salvate. Così è possibile creare l'insieme di immagini da cui l'algoritmo di HALCON ricava i dati per eseguire la taratura delle tre telecamere. I risultati di taratura sono poi salvati in un file e serviranno per eseguire la ricostruzione stereoscopica, oggetto del capitolo 4.4.



**Figura 4.15:** Interfaccia del tool di calibrazione inserito nel pacchetto software.

## 4.4 Ricostruzione stereoscopica

Si vuole quindi procedere alla misurazione delle coordinate tridimensionali dei punti noti (*marker*). Per fare ciò è stato implementato un software che consente, di individuare nelle immagini i punti noti e ricavarne la posizione 3D. Per ricavare questa informazione si ricorre alla *ricostruzione stereoscopica*, descritta in dettaglio nel capitolo 3.2.

### 4.4.1 Implementazione

Fino ad oggi, presso la galleria del vento del Politecnico di Milano, è stato utilizzato un sistema di misura che consente di acquisire e salvare immagini durante la prova sperimentale, mentre le misure sono ricostruite *offline*, cioè una volta terminata l'acquisizione. In questo modo l'ammontare dei dati che bisogna immagazzinare e poi analizzare per il calcolo delle coordinate 3D risulta molto elevato. A maggior ragione se si considera che la maggior parte delle applicazioni in galleria del vento richiedono frequenze di acquisizione elevate (20 Hz) e immagini ad alta risoluzione (1392 x 1040 pixel). Questa procedura risulta inoltre molto lenta poiché non è possibile lavorare a tempo mascherato, vale a dire non è possibile rielaborare i dati durante l'acquisizione ma solamente a monte di essa.

Con lo scopo di rendere l'intera procedura di triangolazione notevolmente più veloce, si è implementato un software che permette l'estrazione dei markers, il calcolo delle correlazioni e la ricostruzione tridimensionale in linea (figura 4.16). In questo modo vengono salvati direttamente i dati contenenti informazioni riguardanti la posizione dei markers nell'immagine e le loro coordinate tridimensionali, riducendo notevolmente i tempi di ricostruzione e la quantità di dati da immagazzinare. Il tutto va eseguito senza diminuire la frequenza di acquisizione, che deve comunque essere non inferiore ai 20 Hz. Per risolvere questo problema si è fatto ricorso alla programmazione multithread, scomponendo il software in 3 thread differenti.

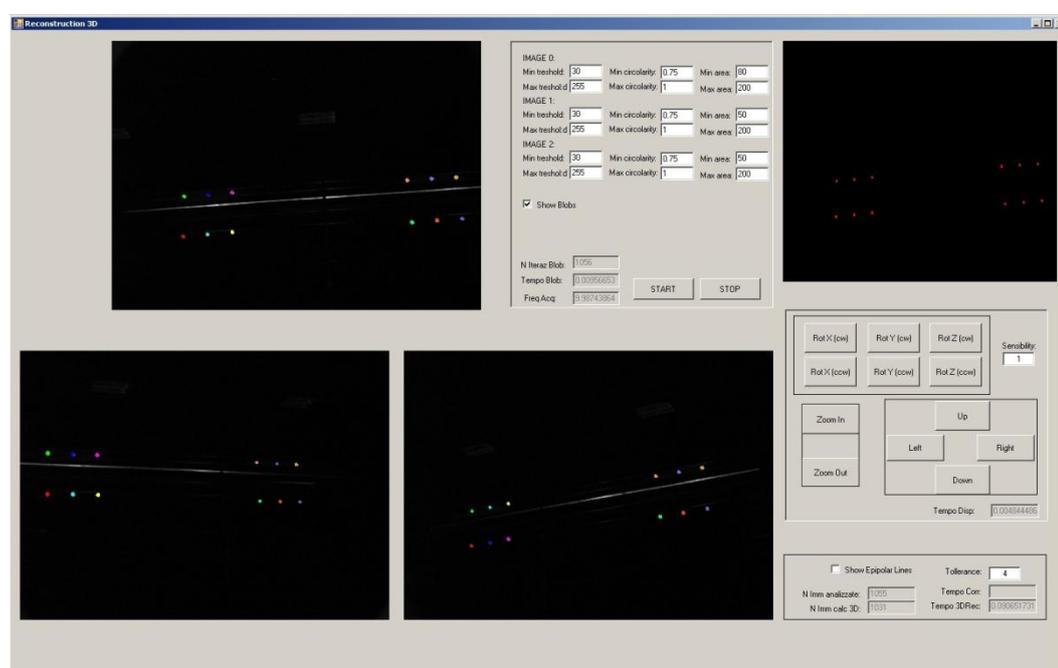
Il primo thread acquisisce continuamente immagini e, per ognuna, esegue la ricerca dei blob (o *markers*). L'algoritmo di ricerca dei blob effettua il *threshold*

(ovvero una soglia) dell'immagine, andando a selezionare solamente i valori di pixel compresi tra un minimo e un massimo, definiti nell'interfaccia dall'utente. Vengono poi selezionate le regioni nell'immagine che hanno un'area e una "circolarità" (somiglianza con una circonferenza) comprese tra valori minimi e massimi, anch'essi modificabili dall'interfaccia utente. La possibilità di selezionare i parametri relativi alla ricerca dei blob rende il software più flessibile ai cambiamenti di luce che possono verificarsi tra diverse acquisizioni e alle differenti condizioni di illuminazione a cui sono soggette le immagini acquisite dalle diverse telecamere. Il primo thread salva le coordinate del centro dei blob in una coda.

Nel secondo thread, per ogni blob trovato nella coda (FIFO), viene disegnata la retta epipolare nelle altre due immagini. Si lascia all'utente la possibilità di visualizzare le rette epipolari, diminuendo però la velocità computazionale dell'algoritmo. Viene poi applicato l'algoritmo descritto in dettaglio nella sezione precedente per trovare le corrispondenze dei punti omologhi. L'utente può inserire un valore arbitrariamente scelto di tolleranza, ovvero il parametro che definisce la soglia di distanze dalla retta epipolare entro il quale un punto viene selezionato (si veda la sezione precedente per maggiori chiarimenti). Scegliendo una tolleranza troppo bassa si rischia di non trovare alcune corrispondenze tra i punti omologhi; dall'altra parte, aumentando molto questo parametro, si diminuisce drasticamente la velocità dell'algoritmo. Per ogni punto di cui vengono trovate corrispondenze nelle altre due immagini si ricostruiscono le coordinate tridimensionali. Queste vengono salvate in un'ulteriore coda. Essendo spesso il processo di triangolazione più lento dell'acquisizione, esso continuerà per qualche secondo una volta terminata l'acquisizione, dopodiché viene salvato un file di testo per ogni immagine nel quale sono state trovate coordinate 3D dei punti, contenente la posizione nelle immagini dei punti analizzati e le rispettive coordinate nello spazio.

Infine un ultimo thread consente di visualizzare i risultati della triangolazione. Esso prende gli ultimi dati acquisiti nel primo thread, esegue la ricostruzione tridimensionale come nel secondo thread e visualizza i risultati grazie a un

visualizzatore 3D implementato in laboratorio. Esso consente di ruotare o traslare nelle tre direzioni la ricostruzione dei markers. Essendo il processo di triangolazione e visualizzazione più lento dell'acquisizione, non verrà mostrato il 3D per ogni immagine acquisita. Verrà invece mostrato il maggior numero di ricostruzioni possibile, che dipende dalla velocità computazionale del thread di visualizzazione. Questo è stato aggiunto con il solo scopo di fornire all'utente un riscontro visivo immediato così da poter controllare durante l'acquisizione che l'intero processo di triangolazione stia funzionando adeguatamente.



**Figura 4.16:** Interfaccia del software che permette di ricostruzione tridimensionale durante un'applicazione dello strumento presso la Galleria del Vento.

## 4.5 Registrazione di diversi moduli di acquisizione

In applicazioni complesse, come ad esempio misurazioni in galleria del vento, può essere necessario utilizzare più moduli di acquisizione (triplette) contemporaneamente. Per evitare, infatti, che alcune parti dell'oggetto da misurare non siano visibili al sistema di acquisizione, è spesso necessario utilizzare più di tre telecamere. Si pensi ad esempio alle misurazioni della forma di una vela di superficie molto estesa e ad elevata curvatura, come può essere uno *spinnaker* (vela simmetrica di grandi dimensioni che viene issata nelle andature di lasco e poppa); in tale caso bisogna utilizzare almeno due moduli di acquisizione per riuscire ad analizzare tutta la superficie della vela ed effettuare misure su di essa. Ogni tripletta di telecamere viene tarata separatamente (e *offline*) con il metodo descritto nel capitolo 4.3 e durante la taratura vengono calcolati, oltre ai parametri intrinseci, anche i parametri estrinseci delle tre telecamere. Viene cioè calcolata la posa di ogni dispositivo rispetto a un sistema di riferimento mondo fissato. Nel caso di due triplette vicine, esse potrebbero ad esempio essere tarate rispetto a una posa comune del calibratore. Chiaramente questo non è possibile quando le triplette sono lontane tra loro o nel caso in cui si utilizzino numerosi strumenti di misura. In tal caso infatti una posa del calibratore visibile da alcuni dispositivi potrebbe non esserlo da altri. Perciò si è deciso di non utilizzare un sistema di riferimento comune a tutte le triplette, ma di tarare ogni singolo strumento rispetto a un proprio sistema di riferimento (di solito si utilizza la posizione della prima telecamera). In questo modo, andando a ricostruire le coordinate tridimensionali dei marker, si troveranno coordinate differenti per ogni tripletta utilizzata, poiché riferite a diversi sistemi di riferimento. Per effettuare la misura è necessario che esse siano riferite ad un unico sistema di riferimento mondo (di solito la posa della prima telecamera della prima tripletta). Si vuole quindi calcolare la matrice di roto-traslazione che trasformi le coordinate dei punti dai sistemi di riferimento dei diversi strumenti a un sistema di riferimento mondo comune. La procedura che permette di fare ciò

viene chiamata *registrazione*, ed equivale a trovare la posa delle triplette rispetto al SDR mondo. Ogni volta che una modulo di acquisizione viene spostato non sarà quindi necessario ricalibrare tutto il sistema, ma basterà registrare la tripletta nuovamente. Uno dei metodi più conosciuti per eseguire questa procedura, descritto nel capitolo 4.5.1, è basato sulla scomposizione ai valori singolari (*SVD: Singular Value Decomposition*) [27].

### 4.5.1 Registrazione di moduli di acquisizione mediante decomposizione ai valori singolari

Si supponga di avere due insiemi di punti 3D, ad esempio le coordinate tridimensionali dei markers riferite a due sistemi di riferimento tripletta differenti. Chiameremo uno di questi insiemi  $X$  e l'altro  $Y$ . Si assuma che per ogni punto di  $Y$  sia noto il punto corrispondente in  $X$ . Il problema dell'orientazione assoluta (o problema della registrazione di insiemi di punti) consiste nel trovare la trasformazione rigida 3D (rotazione e traslazione) da applicare ad  $Y$  che lo porti a coincidere con  $X$ , oppure, in presenza di rumore, che renda minima la distanza tra i due insiemi di punti. Si consideri il problema più generale dell'orientazione assoluta con scala, nel quale la relazione tra i due insiemi è la seguente:

$$X_i = s(RY_i + t) \quad \text{per ogni } i = 1 \dots N \quad (4.3)$$

dove  $R$  è una matrice di rotazione  $3 \times 3$ ,  $t$  è un vettore di traslazione  $3 \times 1$ ,  $s$  è uno scalare e  $X_i, Y_i$  sono punti corrispondenti (in coordinate cartesiane) negli insiemi  $X$  ed  $Y$ . L'obiettivo della registrazione è risolvere:

$$\min_{R,t} \sum_{i=1}^N \|X_i - s(RY_i + t)\|^2 \quad (4.4)$$

Facendo la media su  $i$  di entrambi i membri della (4.3) si ottiene immediatamente una espressione per la traslazione:

$$t = \frac{1}{s} \left( \frac{1}{N} \sum_{i=1}^N X_i \right) - R \left( \frac{1}{N} \sum_{i=1}^N Y_i \right) \quad (4.5)$$

Sostituendo questa nella (4.3) è possibile eliminare la traslazione dal problema:

$$X_i = sRY_i \quad (4.6)$$

dove  $\bar{X}_i = X_i - \frac{1}{N} \sum_{i=1}^N X_i$  e  $\bar{Y}_i = Y_i - \frac{1}{N} \sum_{i=1}^N Y_i$  sono gli insiemi “centralizzati”, ottenuti sottraendo i rispettivi centroidi (o centri di massa). Poiché le matrici ortogonali applicate ad un vettore non ne alterano il modulo, si può immediatamente ricavare il fattore di scala  $s$  prendendo la norma della (4.6):

$$\|\bar{X}_i\| = s\|\bar{Y}_i\| \quad (4.7)$$

Rimane il problema di stimare la rotazione nella (4.6). Sia  $\bar{X}$  la matrice  $3 \times N$  ottenuta accostando uno accanto all'altro i vettori colonna  $\bar{X}_i$  ed  $\bar{Y}$  la matrice ottenuta allo stesso modo con i vettori  $s\bar{Y}_i$ . Si verifica facilmente che la funzione obiettivo (o residuo) si riscrive:

$$\varepsilon = \sum_{i=1}^N \|X_i - sRY_i\|^2 = \|\bar{X} - sR\bar{Y}\|_F^2 \quad (4.8)$$

Ricordando la formula di Lagrange, vale a dire che dati quattro vettori  $a, b, c, d \in \mathbb{R}^3$  si ha:

$$(a \times b)^T (c \times d) = (c^T a)(b^T d) - (d^T a)(b^T c) \quad (4.9)$$

Dalla (4.8), tenendo conto della (4.9), è possibile ricavare la soluzione finale, che è data da:

$$R = V \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & \det(VU^T) \end{bmatrix} U^T \quad (4.10)$$

dove  $UDV^T = \bar{Y}\bar{X}^T$  è la SVD della matrice  $\bar{Y}\bar{X}^T$ . La matrice  $R$  rappresenta la soluzione ai minimi quadrati del problema [1]. Infine si sostituisce l'identità tra  $V$  e  $U$  con una matrice che serve a garantire che il risultato sia una matrice di rotazione, cioè abbia determinante positivo [1].

## 4.5.2 Implementazione

In questo lavoro si è sempre utilizzata una sola tripletta di telecamere e mai più strumenti stereoscopici contemporaneamente. In alcune applicazioni però è necessario utilizzare più strumenti contemporaneamente ed è dunque necessario prevedere un algoritmo che possa registrare più triplette contemporaneamente. È stato dunque sviluppato un programma che permette, date le coordinate tridimensionali in sistema di riferimento tripletta, di trovare la posa di una tripletta

rispetto all'altra. Questo algoritmo non è stato inserito nel pacchetto software completo che gestisce lo strumento, ma è stato lasciato a parte e sono stati implementati alcuni programmi in HDevelop che consentono di verificare l'algoritmo di registrazione anche solamente utilizzando una sola tripletta. Alcuni di queste prove sono state descritte nel capitolo 5.1. Si è quindi implementata una funzione utilizzabile da qualsiasi programma in HALCON che prende in ingresso due insiemi di coordinate tridimensionali di punti ordinati e calcola la matrice di rotazione che permette di passare dal primo insieme di punti al secondo. In questo modo passando alla funzione un insieme di punti riferiti a due triplette differenti, si può quindi trovare la posa di una rispetto all'altra. È chiaramente indispensabile che i punti siano ordinati nello stesso modo, cioè il primo punto di un insieme deve essere lo stesso punto posto in prima posizione nell'altro insieme e così via.

## 5. Attività sperimentale

*In questo capitolo si descrivono una serie di esperimenti che sono stati effettuati per la validazione metrologica dello strumento. Per prima cosa sono riportati alcuni esperimenti realizzati presso il VBLab con lo scopo di verificare il corretto funzionamento del sistema e validarlo metrologicamente. In seguito lo strumento è stato utilizzato durante una campagna sperimentale presso la Galleria del Vento e si riportano i dati raccolti in essa.*

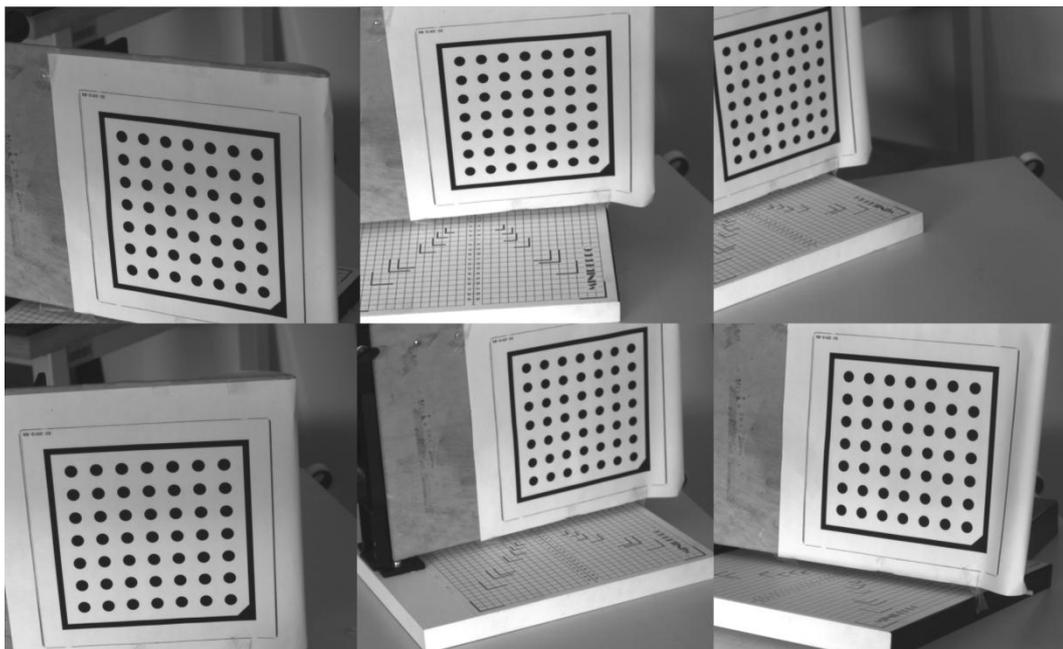
### 5.1 Verifiche metrologiche

Prima di applicare lo strumento a una vera e propria campagna sperimentale si è voluto verificare il corretto funzionamento dello stesso e studiare il comportamento di diverse configurazioni per capire come vari l'accuratezza modificando la posizione relativa delle telecamere o la distanza alla quale si effettuano le misure. In questo capitolo dunque si descrivono una serie di esperimenti atti a validare metrologicamente lo strumento di misura prima di utilizzarlo in una vera e propria campagna sperimentale presso la Galleria del Vento.

### 5.1.1 Ricostruzione del calibratore

Prima di tutto si verifica il corretto funzionamento del processo di calibrazione e ricostruzione. L'esperimento classico più significativo è la ricostruzione di un insieme di punti di cui si conosce con elevata accuratezza la distanza relativa. Così facendo è possibile verificare lo scostamento tra le distanze dei punti reali e le distanze calcolate con la ricostruzione.

Per prima cosa si vuole ricostruire il calibratore HALCON mostrato in figura 5.1 in diverse pose. HALCON presenta alcune funzioni già implementate che consentono di trovare in un'immagine questo tipo di calibratore ed effettuare automaticamente la correlazione tra i punti nelle immagini acquisite dalle tre diverse telecamere. In questo modo non è necessario utilizzare la geometria epipolare per correlare i punti e dunque si può verificare solamente il processo di ricostruzione tridimensionale.



**Figura 5.1:** Alcune immagini del calibratore HALCON che si vuole ricostruire in diverse pose.

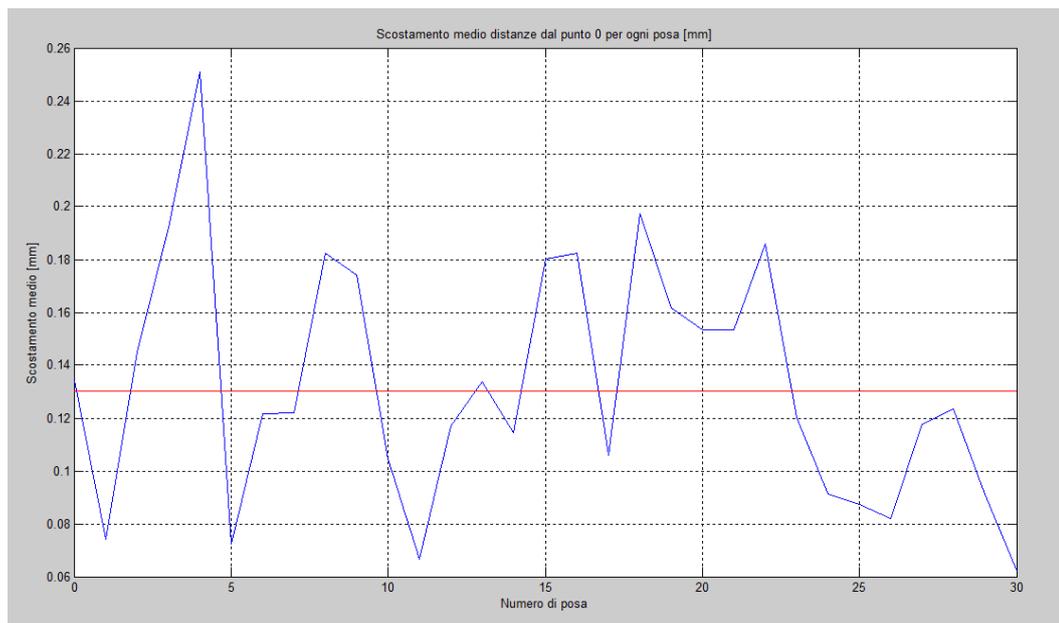
Con il sistema stereoscopico sono state dunque acquisite 31 pose differenti e per ognuna è stato ricostruito il calibratore, trovando le coordinate tridimensionali di ogni punto. Da esse poi sono state calcolate le distanze di ogni blob dal primo punto, vale a dire quello nell'angolo in cui è presente l'asimmetria (in pratica

quello in basso a destra nel calibratore delle immagini di figura 5.1). Viene quindi trovato un vettore delle distanze misurate di ogni punto dal primo, di 49 elementi, per ogni posa del calibratore. Per come è costruito il calibratore si sa che la distanza tra i blob è 250 mm e il vettore delle distanze dal primo punto reali sarà dunque quello della tabella 5.1. Per motivi di spazio e leggibilità non vengono riportati tutti i 49 valori del vettore ma solamente alcuni, solo per dare un'idea su come è costruito tale vettore.

<b>Punto:</b>	0	1	2	3	...	7	8	9	...
<b>Distanza dal punto 0 [m]:</b>	0	0.025	0.05	0.075	...	0.025	0.03535	0.0559	...

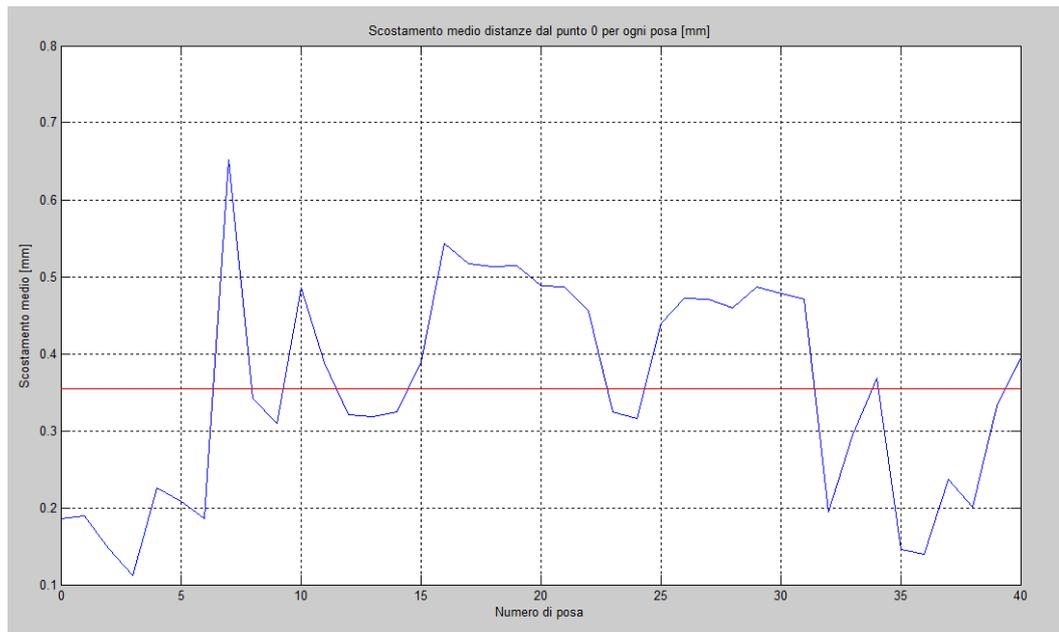
**Tabella 5.1:** Vettore delle distanze reali dei punti dal punto 0, vale a dire il punto nell'angolo in cui è presente l'asimmetria.

Dalla differenza in modulo tra il vettore delle distanze reali e ogni vettore delle distanze misurate si trova un vettore degli scostamenti tra le distanze per ogni posa. Viene poi trovata la media tra gli elementi di ognuno dei vettori degli scostamenti. Così facendo si ha un valore di scostamento medio per ogni posa. Si riporta dunque in figura 5.2 il grafico che mostra l'andamento del valore dello scostamento medio per ogni posa del calibratore. Da esso si nota come esso non superi mai 0.25 mm ed è in media pari a 0.13 mm.



**Figura 5.2:** Scostamento medio dal valore reale delle distanze dei punti dal punto 0, per ogni posa del calibratore.

È da notare che i risultati ottenuti dipendono fortemente dalla bontà della calibrazione. Infatti in questo caso la calibrazione della tripletta ha restituito un errore di retroproiezione di 0.67 pixel, quindi relativamente basso. L'esperimento sopra descritto è stato ripetuto calibrando il sistema volontariamente in modo meno adeguato. In particolare è bastato acquisire le immagini di calibrazione con il calibratore posto a maggior distanza e dunque più piccolo nell'immagine rispetto al caso precedente. Così facendo la calibrazione ha restituito un valore di retroproiezione maggiore (1.12 pixel). Vengono poi acquisite nuove immagini del calibratore ancora una volta in 31 pose differenti e si trova il grafico dello scostamento medio per ogni posa del calibratore mostrato in figura 5.3. Da esso si nota che il valor medio dello scostamento è raddoppiato rispetto al caso precedente, assumendo un valore di 0.35 mm.



**Figura 5.3:** *Scostamento medio dal valore reale delle distanze dei punti dal punto 0 per ogni posa del calibratore nel caso in cui la calibrazione ha errore di retroproiezione più alto.*

## 5.1.2 Verifica della procedura di registrazione

In questo lavoro si è sviluppato un algoritmo che permetta di registrare più strumenti stereoscopici tra loro, cioè ricavare un sistema di riferimento comune ed in particolare si è implementata la funzione descritta nel capitolo 4.5.2. In questo capitolo si descrivono le prove finalizzate a verificare che la registrazione sia eseguita correttamente e non introduca incertezze eccessive nelle misure. Esse consistono nella registrazione di due triplette poste a distanze diverse dal calibratore. Quello che si è fatto in pratica, non avendo a disposizione due triplette di telecamere, è stato acquisire due pose differenti del calibratore. Chiaramente il fatto di spostare il calibratore è equivalente allo spostamento della tripletta, che a sua volta può essere considerato come un'acquisizione sincrona dello stesso calibratore da due strumenti in posizioni differenti tra loro. In particolare sono state eseguite diverse prove in cui il calibratore è stato spostato in pose differenti. Nel seguito si descrivono a titolo esemplificativo i passaggi per la realizzazione di una di queste prove, per poi riportare solamente i risultati ricavati da alcune prove.

Nella prima prova effettuata si è spostato un calibratore, appoggiato su un tavolo, di 268 mm lungo l'asse perpendicolare al piano del sensore (asse z). Ciò coincide con un'acquisizione sincrona dello stesso calibratore da due differenti triplette, una posta a 268 mm dall'altra. Le pose delle due triplette rispetto al calibratore (che in realtà sarebbero le pose del calibratore rispetto al sistema di riferimento tripletta) sono espresse nella tabella 5.2. Esse sono composte dalle tre componenti di traslazione ( $t_x$ ,  $t_y$  e  $t_z$ ) in millimetri e tre di rotazione ( $a$ ,  $b$  e  $g$ ) in gradi espresse secondo la notazione di Eulero [13].

	Posa tripletta 1	Posa tripletta 2
$t_x$ [mm]	212.5	210.8
$t_y$ [mm]	20.7	81.7
$t_z$ [mm]	1345.2	1606.2
$a$ [°]	13.7	13.7
$b$ [°]	358.4	358.2
$g$ [°]	175.7	175.8

**Tabella 5.2:** Pose delle due triplette rispetto al calibratore.

Prima di tutto si è provveduto ad effettuare una veloce verifica di calibratura, verificando che la distanza tra le pose dei due calibratori sia equivalente a quella misurata, e quindi 268 mm. La distanza è data da:

$$D_{pose} = \sqrt{(t_{x_1} - t_{x_2})^2 + (t_{y_1} - t_{y_2})^2 + (t_{z_1} - t_{z_2})^2} = 268 \text{ mm}$$

Vengono ricostruiti i calibratori dalle due diverse triplette e, dalle coordinate dei punti (ordinati), l'algoritmo implementato ricava la matrice di rotazione che permette di passare da una tripletta all'altra, utilizzando la decomposizione ai valori singolari (*svd*). Tale matrice può essere convertita in vettore posa, vale a dire il vettore che esprime la posa della seconda tripletta rispetto alla prima, mostrato nella tabella 5.3.

	Posa tripletta 1	Posa tripletta 2
tx [mm]	0	4.9
ty [mm]	0	55.8
tz [mm]	0	259.9
a [°]	0	359.8
b [°]	0	0.1
g [°]	0	359.9

**Tabella 5.3:** Pose delle due triplette rispetto alla prima tripletta. La posa della tripletta 2 è stata calcolata con l'algoritmo per la registrazione con svd.

Si imposta quindi alla seconda tripletta lo stesso sistema di riferimento della prima, semplicemente imponendo che essa si trova nella posa calcolata con l'algoritmo di registrazione e mostrata in tabella 5.3. Fatto ciò si ricalcolano le coordinate tridimensionali del calibratore, che questa volta non saranno più riferite al sistema di riferimento seconda tripletta, bensì alla prima. In questo modo le nuove coordinate dei punti dovrebbero essere identiche a quelle trovate con la prima tripletta, a meno di un'incertezza introdotta dalla procedura di registrazione. Per verificare l'accuratezza della procedura vengono dunque ricavati tre vettori degli scostamenti dal valore reale delle coordinate dei punti su x, y e z. Essi sono calcolati come la differenza in modulo delle coordinate dei punti trovate con la prima tripletta e le coordinate misurate con la seconda tripletta registrata. Viene poi calcolato il valore medio di ognuno dei tre vettori e il valore medio di tutti e tre i vettori, riportati nella tabella 5.4.

Scostamento medio su x [mm]	Scostamento medio su y [mm]	Scostamento medio su z [mm]	Scostamento medio totale [mm]
0.05	0.03	0.06	0.05

**Tabella 5.4:** Scostamento medio dal valore reale delle coordinate dei punti sui singoli assi x, y e z e su tutti e tre gli assi. Esperimento in cui è stata modificata la coordinata z del calibratore.

Dalla tabella 5.4 si nota lo scostamento medio calcolato è di 0.05 mm, e quindi in questo caso la procedura di registrazione ha accuratezza elevata e non introduce nelle misure incertezze sensibili. La stessa prova è stata effettuata posizionando il

calibratore in varie pose differenti. Alcuni risultati significativi sono mostrati nella tabella 5.5. Da tali prove si è notato che aumentando la distanza relativa tra le due triplette lo scostamento medio totale aumenta e dunque l'incertezza introdotta nelle misure dalla procedura di registrazione è maggiore. Ad esempio nella prova 1 il calibratore è stato spostato su  $z$  di 1000 mm e lo scostamento medio ha raggiunto un valore di 0.1 mm. Analizzando lo scostamento percentuale, cioè lo scostamento medio totale diviso la distanza del calibratore dalla tripletta, si è notato che esso aumenta con l'aumentare della coordinata  $z$ . A titolo di esempio si riportano alcuni risultati in tabella 5.6 di prove in cui il calibratore è stato mosso solamente lungo l'asse  $z$ . Nella tabella 5.6 è mostrata la coordinata  $z$  del calibratore, lo scostamento totale medio e lo scostamento percentuale in [mm/m]. Il fatto che anche lo scostamento percentuale aumenti con la distanza lungo  $z$  dimostra che l'accuratezza di registrazione peggiora incrementando la coordinata  $z$ . Questo aumento di incertezza è dovuto al fatto che, posizionando il calibratore più lontano, esso risulta più piccolo nell'immagine. Ciò comporta una maggiore incertezza nell'individuazione della posizione dei marker.

Inoltre si è osservato che anche variando l'angolazione del calibratore lo scostamento totale medio aumenta. Nella prova 2 della tabella 5.5 ad esempio, oltre a spostare di 1000 mm il calibratore, esso è stato ruotato di  $45^\circ$  intorno all'asse verticale (asse  $y$ ). Si ritiene che ciò sia dovuto al fatto che, ruotando il calibratore, i marker circolari piani diventano ellittici nell'immagine e dunque la stima della posizione del centroide è meno accurata. Si è notato che spostamenti sugli assi paralleli al sensore e rotazioni intorno all'asse perpendicolare ad esso non introducono incertezze trascurabili. Ad esempio nella prova 3 il calibratore è stato traslato di 200 mm lungo gli assi paralleli al sensore ( $x$  e  $y$ ) e ruotato intorno all'asse verticale ( $y$ ) di  $45^\circ$ . Un aumento dello scostamento medio totale è stato riscontrato anche dalla composizione di traslazioni e rotazioni. Ad esempio nella prova 4 il calibratore è stato traslato su  $z$  di 1000 mm, su  $y$  di 350 mm, e ruotato rispetto a  $x$  e  $y$  di  $30^\circ$  e lo scostamento totale è risultato essere di 0.19 mm. Nella prova 5 il calibratore è stato anche ruotato di  $45^\circ$  rispetto all'asse verticale e ne è risultato uno scostamento totale di 0.22 mm.

Per una stima finale dell'incertezza introdotta dalla procedura di ricostruzione si è posizionato il calibratore in modo da comporre rotazioni e traslazioni su tutti e tre gli assi. Chiaramente la posa del calibratore è vincolata al campo di visione dello strumento. Si è deciso di posizionare il secondo calibratore al limite del campo visivo. Ad esempio lungo z è stato possibile posizionare il secondo calibratore ad un massimo di 2500 mm dal primo, altrimenti esso sarebbe risultato troppo piccolo nell'immagine e dunque non identificato correttamente. Allo stesso modo non è stato possibile modificare l'angolazione intorno all'asse verticale di più di  $45^\circ$  e intorno all'asse x di più di  $30^\circ$ , altrimenti il calibratore non sarebbe stato visibile a una delle telecamere. Inoltre le traslazioni lungo x e y sono limitate dal fatto che il calibratore deve essere interamente visibile a tutte le tre telecamere, e il loro valore massimo dipende dalla distanza a cui è posizionato il calibratore su z. Non vi sono limitazioni per quanto riguarda la rotazione intorno all'asse perpendicolare al sensore (asse z). Nella prova 6 il secondo calibratore è stato spostato lungo z di 2500 mm e lungo x e y di 350 mm. Esso è stato ruotato di  $30^\circ$  intorno a x,  $45^\circ$  intorno a y e a z. Si nota che in questo caso lo scostamento medio totale è di quasi 0.3 mm e dunque l'incertezza introdotta assume valori bassi ma non trascurabili.

N	Scost. medio x [mm]	Scost. medio y [mm]	Scost. medio z [mm]	Scost. medio tot. [mm]
1	0.11	0.07	0.12	0.1
2	0.17	0.16	0.11	0.15
3	0.024	0.02	0.04	0.03
4	0.2	0.17	0.2	0.19
5	0.2	0.24	0.19	0.22
6	0.31	0.27	0.28	0.3

**Tabella 5.5:** Risultati di alcune delle prove effettuate: scostamento medio dal valore reale delle coordinate dei punti sui singoli assi x, y e z e su tutti e tre gli assi.

Distanza del calibratore lungo z [mm]	Scostamento medio totale [mm]	Scostamento percentuale [mm/m]
2200	0.04	0.018
2600	0.07	0.027
3000	0.1	0.033

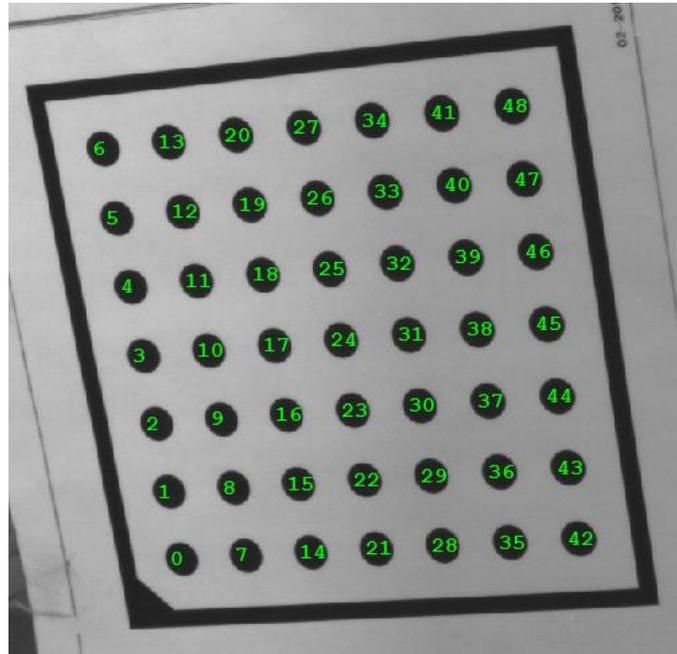
**Tabella 5.6:** Risultati di scostamento percentuale in alcune prove in cui il calibratore è stato mosso solamente lungo l'asse z. Lo scostamento percentuale è calcolato come il rapporto tra lo scostamento totale medio (seconda colonna) e la distanza del calibratore dallo strumento di misura (prima colonna) ed è rappresentato in [mm/m].

In conclusione è stato mostrato che l'incertezza introdotta dalla procedura di calibrazione è trascurabile se il calibratore è ben visibile da entrambe le triplette. Se invece esso è posto al limite del campo visivo per uno dei due moduli di acquisizione, è possibile introdurre nelle misure incertezze basse ma comunque non trascurabili. Ciò è causato principalmente dalla scarsa accuratezza con cui è possibile trovare il centroide dei marker se il calibratore è in una posa particolarmente sfavorevole. È dunque conveniente registrare i moduli di acquisizione utilizzando un calibratore ben visibile a entrambi e se ciò non fosse possibile di verificare la procedura con una prova simile alle precedenti.

### 5.1.3 Verifica dell'accuratezza di registrazione al variare del numero di punti utilizzati

Le prove descritte nel capitolo 5.1.2 sono state eseguite utilizzando tutti i punti del calibratore HALCON. La procedura di registrazione basata sulla decomposizione ai valori singolari (*svd*) dipende fortemente dal numero di punti che si utilizzano. Nel caso precedente sono dunque stati utilizzati 49 punti, mentre nelle applicazioni comuni si utilizza un numero di punti compresi tra 10 e 20. Quindi i risultati ricavati al capitolo 5.1.2 sono anche frutto di un numero di dati maggiore rispetto a quello necessario. Si vuole mediante alcune prove capire come varia

l'accuratezza di registrazione modificando il numero di punti utilizzati per l'algoritmo basato sul *svd*. Per fare ciò si sono numerati i blob del calibratore da 0 a 48, come mostrato in figura 5.4. Il primo punto è sempre quello nell'angolo in cui è presente l'asimmetria.

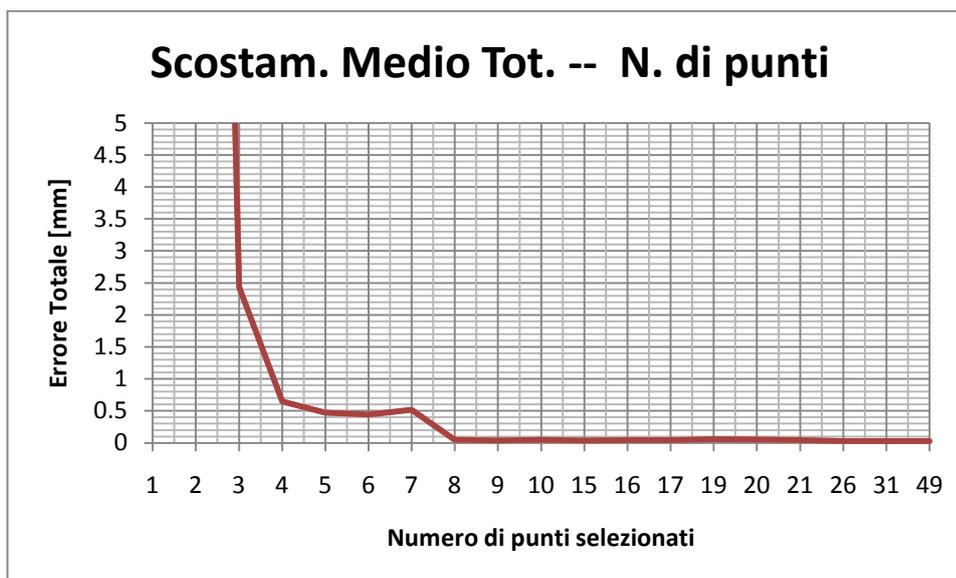


**Figura 5.4:** Numerazione assegnata al calibratore HALCON; il primo blob è sempre quello nell'angolo in cui è presente l'asimmetria.

Una volta numerati i vari blob si è eseguito l'algoritmo di registrazione con due triplete simulate come nelle prove mostrate al capitolo 5.1.2, trovando lo scostamento medio tra tutti gli assi. Questa volta però si passano all'algoritmo di registrazione soltanto alcuni dei punti, variando anche la sequenza di blob selezionati, in modo da verificare oltre alla dipendenza dell'accuratezza dal numero di punti utilizzati anche se essa può essere influenzata dalla loro posizione. I risultati di queste prove sono espressi nella tabella 5.7, che mostra quanti e quali punti sono stati selezionati e lo scostamento medio totale trovato utilizzando tali punti per l'algoritmo di registrazione.

NUMERO DI PUNTI	PUNTI	SCOST. TOTALE MEDIO [mm]	NUMERO DI PUNTI	PUNTI	SCOST. TOTALE MEDIO [mm]
49	0 - 48	0.028	8	41 - 48	0.067
31	10 - 40	0.029	7	0 - 6	0.17
26	12 - 35	0.03	7	17 - 23	0.054
21	0 - 20	0.044	7	42 - 48	1.33
20	0 - 19	0.051	6	0 - 5	0.052
19	0 - 18	0.057	6	17 - 22	0.11
17	0 - 16	0.054	6	43 - 48	1.17
17	17 - 33	0.033	5	0 - 4	0.16
17	32 - 48	0.036	5	17 - 21	0.16
16	0 - 15	0.05	5	44 - 48	1.1
16	17 - 32	0.033	4	0 - 3	0.356
15	0 - 14	0.053	4	17 - 20	0.07
15	17 - 31	0.033	4	45 - 48	1.52
15	34 - 48	0.035	3	0 - 2	3.5
10	0 - 9	0.055	3	17 - 19	0.54
10	17 - 26	0.036	3	46 - 48	5.65
9	0 - 8	0.043	3	0,6,48	0.053
9	40 - 48	0.044	2	0 - 1	48
9	0,6,14,20,24, 28,34,42,48	0.032	2	17 - 18	32
8	0 - 7	0.04	2	0, 48	8.4
8	17 - 24	0.038	1	17	59

**Tabella 5.7:** Tabella che mostra i valori di scostamento medio totale per diversi punti selezionati.

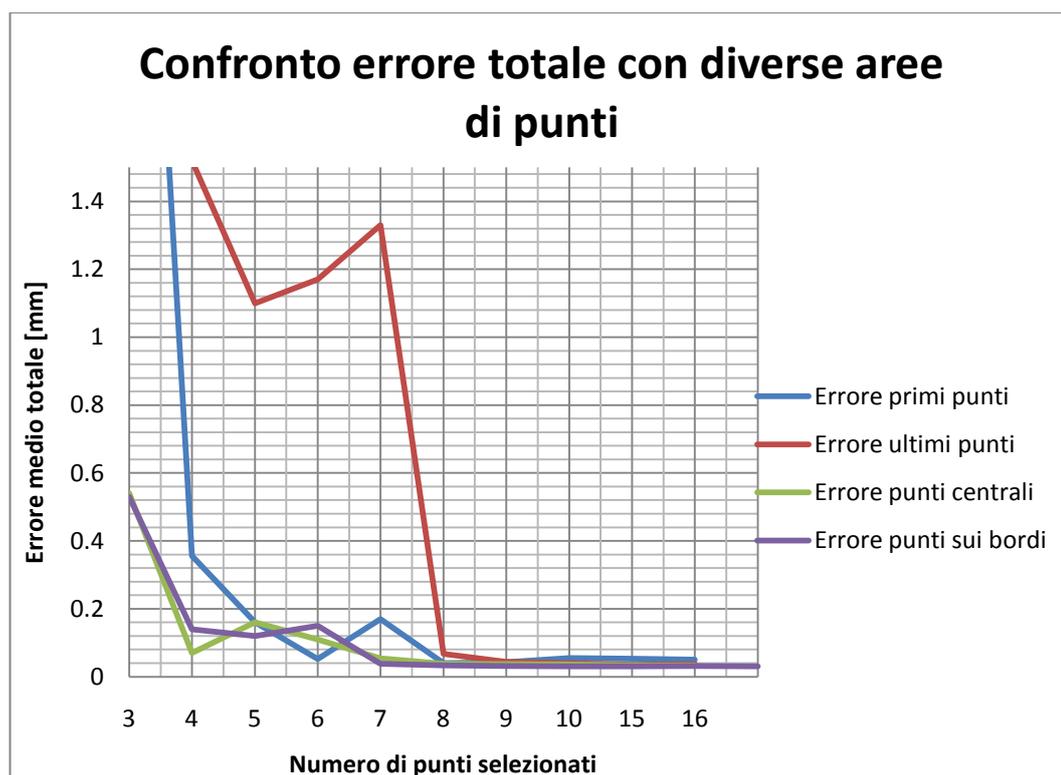


**Figura 5.5:** Grafico dell'andamento dello scostamento medio totale al variare del numero di punti selezionati per l'algoritmo di registrazione.

Dalla tabella 5.7 si nota subito che lo scostamento medio rimane contenuto se si utilizzano molti punti, ma cresce molto con pochi punti. Per un'analisi più dettagliata viene rappresentato l'andamento dello scostamento medio totale al variare del numero di punti selezionati, mostrato nel grafico di figura 5.5. Da esso si intuisce che utilizzare meno di 4 punti restituisce uno scostamento elevato (superiore al millimetro). Utilizzando invece un numero di punti compreso tra 4 e 7 è possibile avere un'incertezza inferiore al millimetro. Se si vuole avere un'incertezza inferiore al decimo di millimetro invece è necessario utilizzare almeno 8 punti.

Volendo invece verificare come cambia l'accuratezza della procedura di registrazione selezionando punti in aree diverse del calibratore si è creato il grafico di figura 5.6. In esso è mostrato ancora una volta l'andamento dello scostamento medio totale al variare del numero di punti selezionati. Questi punti però sono stati selezionati in zone differenti del calibratore. In particolare si è scelto di selezionare i primi punti (quindi nella figura 5.4 quelli più a sinistra), i punti centrali, gli ultimi punti (più a destra) e punti sparsi sui bordi del calibratore, presi in modo tale da coprire il più possibile tutta la superficie di esso. Dai risultati mostrati in figura 5.6 si nota che se si selezionano più di 8 punti è indifferente in

che posizione essi siano presi, poiché si avrà sempre un'accuratezza inferiore al decimo di millimetro, anche se la scelta di punti centrali e punti sui bordi restituisce scostamenti di poco inferiori. Se invece si utilizzano meno di 8 punti è più conveniente utilizzare punti centrali o sui bordi, evitando di utilizzare solamente i punti che si trovano su un lato solo del calibratore.



**Figura 5.6:** Confronto tra lo scostamento medio totale risultante utilizzando punti presi in diverse zone del calibratore.

## 5.2 Prove sperimentali in Galleria del Vento

Il sistema, una volta verificato metrologicamente, è stato utilizzato in una campagna sperimentale effettuata presso la Galleria del Vento. In particolare si è voluto studiare il comportamento del modello raffigurato in figura 5.7 di una struttura che servirà per aumentare la stabilità di uno dei palazzi del nuovo complesso CityLife di Milano, soggetto alla forza del vento.

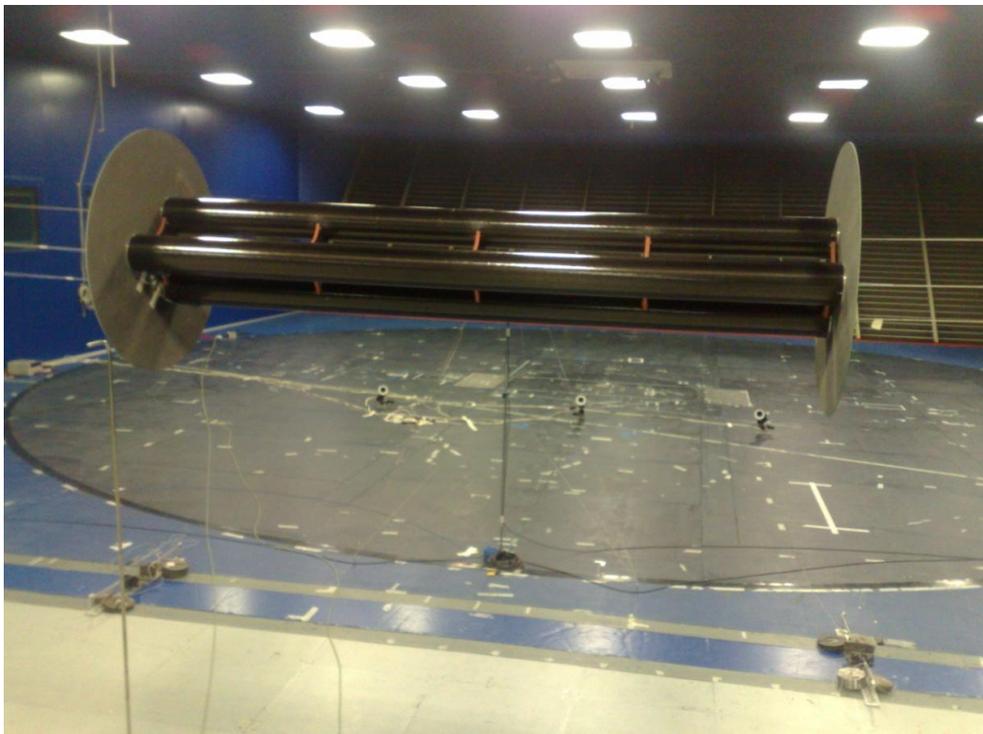


**Figura 5.7:** Modello della struttura utilizzato per gli esperimenti in Galleria del Vento.

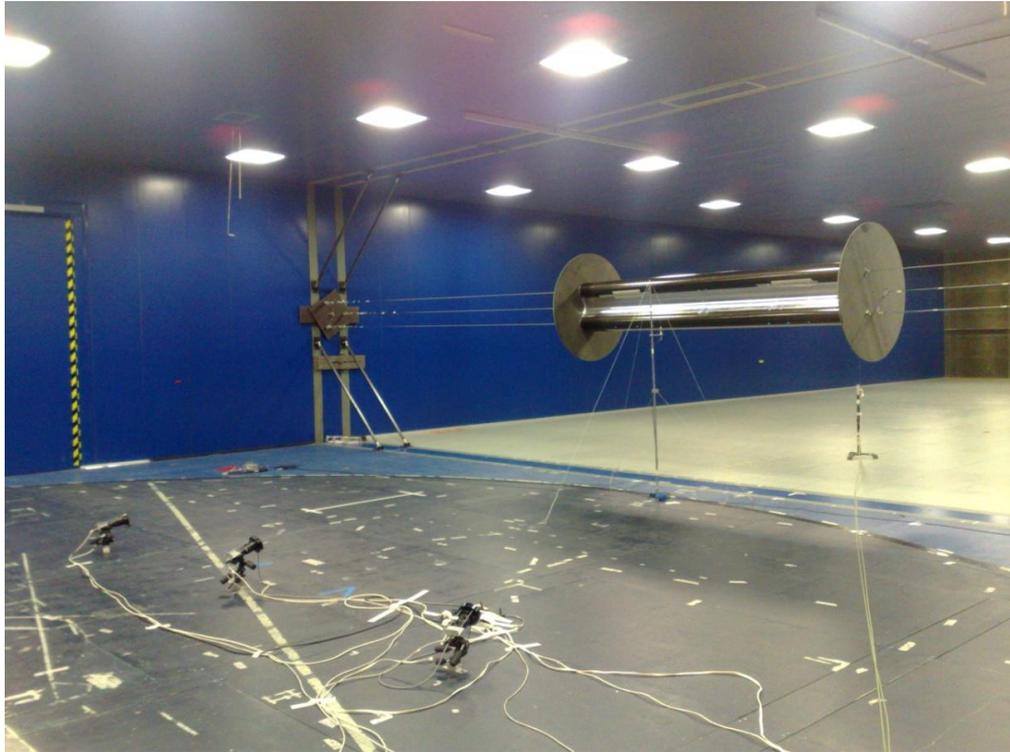
Il modello si compone di tre tubi lunghi 3.22 m ai quali estremi sono inseriti due piattelli circolari di diametro 1.2 m. L'intero modello viene strallato ai lati della sala di prove, così da risultare sospeso a 1.5 m di altezza. Lo spazio compreso tra i tre tubi è riempito da diversi materiali poiché si vuole studiare la stabilità del modello con diversi riempimenti.

La scelta del setup del sistema di misura rappresenta una parte essenziale dell'intero processo di misura. In questo caso, essendo il modello 3.22 m, il campo di misura è vasto. Quindi le tre telecamere sono posizionate piuttosto

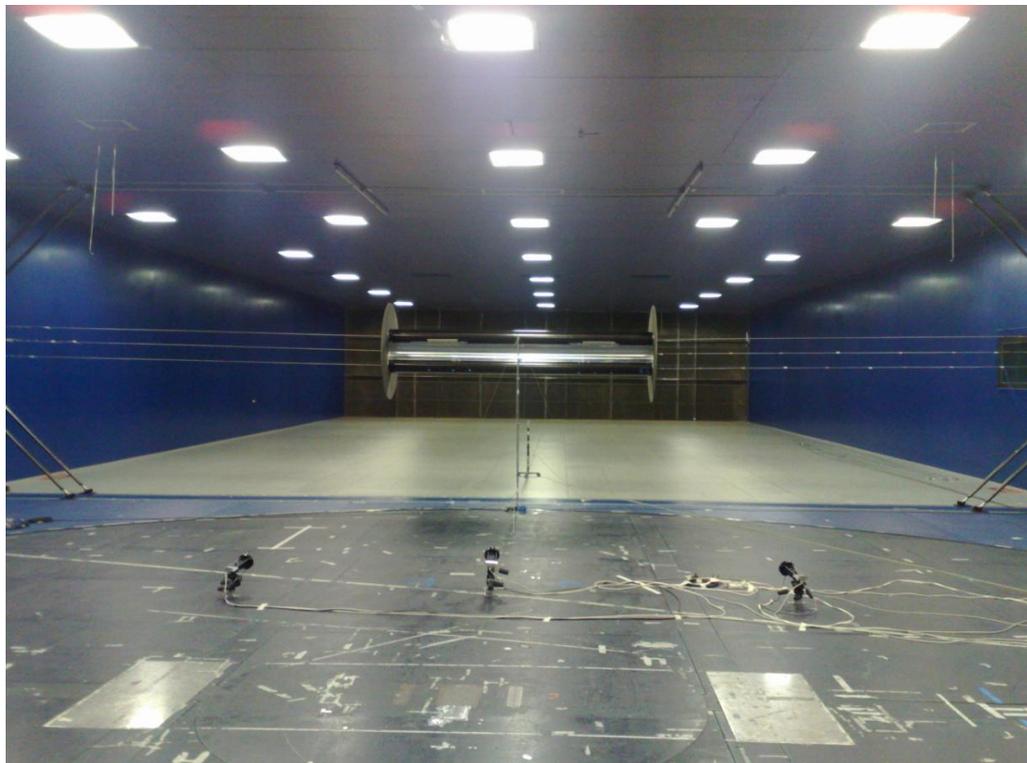
lontane dal modello e tra di loro, in modo da poterlo inquadrare interamente. Quindi non è possibile utilizzare la struttura di sostegno delle telecamere costruita per lo strumento in esame, poiché troppo piccola. Nella sala prove della Galleria del Vento le telecamere possono essere posizionate in alcuni supporti già predisposti, che però sono posti ai lati della sala e a una distanza eccessiva dal modello, quindi si è deciso di posizionare le telecamere in terra proprio dietro al modello. Questa scelta è stata presa poiché in questa zona è presente il banco di prova, ovvero una tavola rotante circolare con diametro 7 m, facilmente visibile nella figura 5.8, su cui solitamente (non in questo caso) sono posizionati i modelli da studiare. Questo può essere forato a piacere e dunque consente di sistemare le telecamere in qualunque posizione desiderata. Le telecamere sono state dunque poste a circa 5 m dal modello, con una distanza relativa tra loro di circa 2 m, come mostrato nelle figure 5.9 e 5.10. Per conoscere con maggiore accuratezza le distanze relative tra le telecamere si fa riferimento alla tabella 5.9, nella quale sono riportati i parametri estrinseci dei dispositivi, trovati dalla calibrazione con sistema di riferimento coincidente con quello della telecamera centrale.



**Figura 5.8:** Setup del sistema per le misure in Galleria del Vento. Le telecamere sono fissate a terra sul banco prova dietro al modello.



**Figura 5.9:** Setup del sistema per le misure in Galleria del Vento.

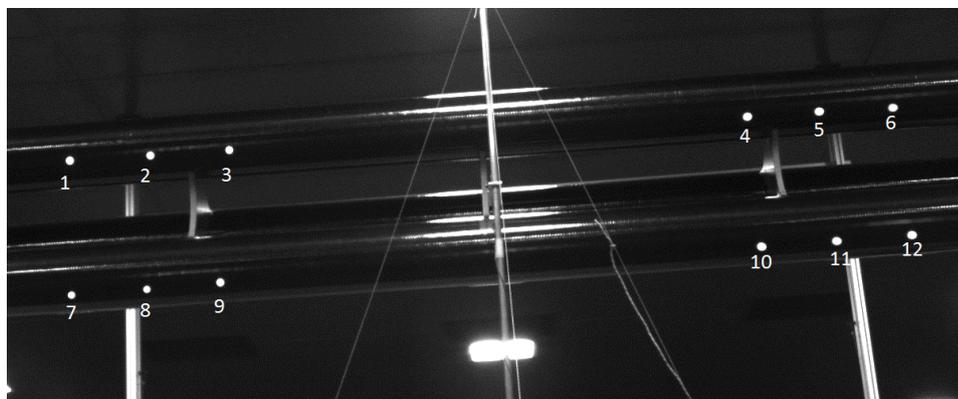


**Figura 5.10:** Setup del sistema per le misure in Galleria del Vento.

Si è deciso di posizionare sul modello 12 marker catarifrangenti. È necessario utilizzare un numero di marker sufficiente a verificare ogni spostamento possibile del modello, tentando però di utilizzarne un numero più basso possibile, in modo da non rendere la procedura di calcolo delle coordinate tridimensionali troppo lenta. I 12 marker sono stati posizionati come mostrato in figura 5.11 e numerati come mostrato in figura 5.12. Si noti che in questa campagna sperimentale sono stati eccitati solamente i moti rigidi, vale a dire quelli inerenti al primo modo di vibrare del modello.



**Figura 5.11:** Posizione dei marker sul modello.



**Figura 5.12:** Numerazione assegnata ai dodici marker.

La scelta del setup presa durante la campagna sperimentale descritta non è sicuramente la migliore. Non avendo una struttura su cui posizionare le telecamere esse sono state sistemate in terra, e quindi sullo stesso piano. Ciò ha fatto sì che esse abbiano tutte *baseline* parallele e quindi anche rette epipolari parallele. Così facendo il processo di individuazione delle corrispondenze con le rette epipolari risulta molto complesso e talvolta presenta degli errori che compromettono la ricostruzione. Essendo il modello rigido, i marker avranno sempre una posizione relativa costante e quindi è possibile inserire nel software un controllo che riordina i punti a seconda della loro posizione. In questo caso allora la procedura di ricerca delle corrispondenze non è più eseguita solamente da un algoritmo basato sulle rette epipolari come descritto nel capitolo 4.4, ma è stato aggiunto un controllo che riordina i punti a seconda della loro posizione. Questo algoritmo quindi divide i 12 punti in due insiemi, uno con i punti posti più in alto nell'immagine e l'altro con i punti più in basso. Ogni insieme viene a sua volta diviso in altri due insiemi di 3 punti ognuno; uno contenente i punti più a sinistra e uno quelli più a destra. Per ognuno di questi sottoinsiemi si trova il punto posto più a sinistra, quello più a destra e quello al centro. Così si identificano i 12 differenti punti in ognuna delle immagini acquisite dalle tre telecamere e è facile imporre le corrispondenze. In questo modo allora non è più necessario utilizzare le rette epipolari, che risultano quindi superflue. Chiaramente questa tecnica può essere utilizzata per questa campagna sperimentale, in cui il vi sono solamente moti rigidi ma non è applicabile a molti altri casi. Nelle applicazioni future in Galleria del Vento futuro allora bisognerà utilizzare una struttura costruita appositamente per la Galleria del Vento, che consente di posizionare le telecamere a distanze relative elevate (circa 3-5 metri) su piani differenti.

In questa applicazione sono stati utilizzati gli illuminatori all'infrarosso e i filtri interferenziali come si nota nella figura 5.13 e 5.14. Il controllore degli illuminatori è stato sistemato in camera di prova (figura 5.15), come anche l'alimentatore del sistema. In camera di controllo invece sono stati posizionati il pc e la scheda che genera il trigger (figura 5.16 e 5.17).



**Figura 5.13:** Particolare raffigurante una singola telecamera e il rispettivo illuminatore.



**Figura 5.14:** Particolare raffigurante illuminatore e filtro interferenziale di una singola telecamera.

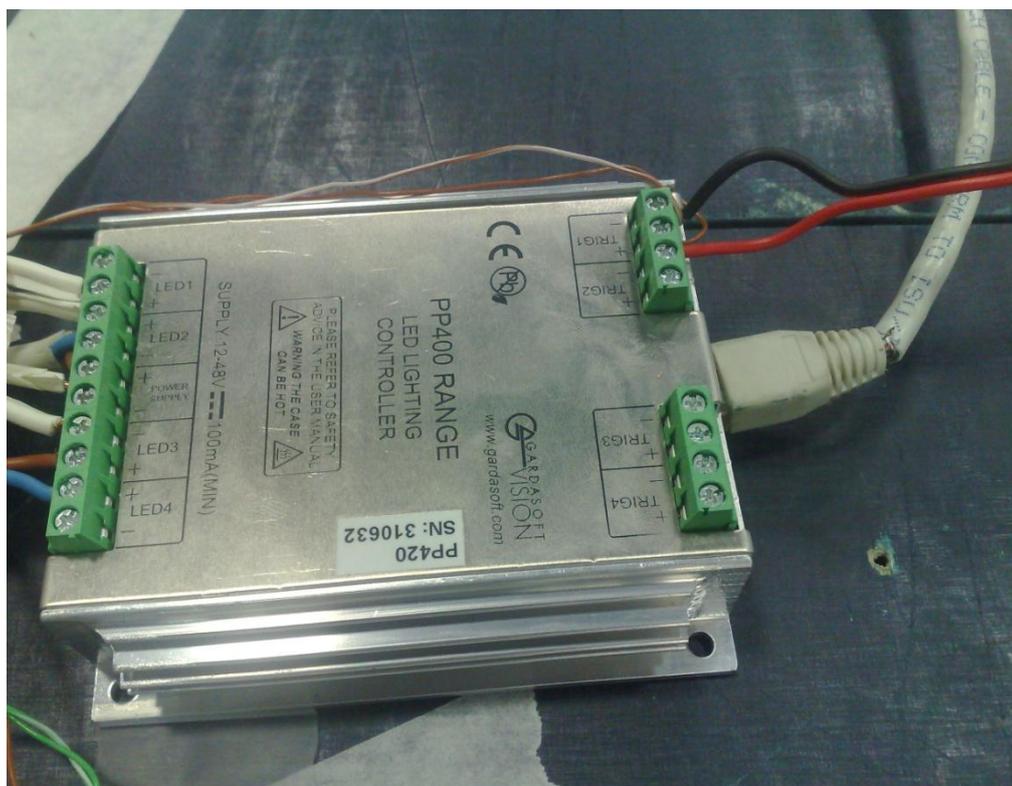


Figura 5.15: Particolare raffigurante il controllore degli illuminatori.

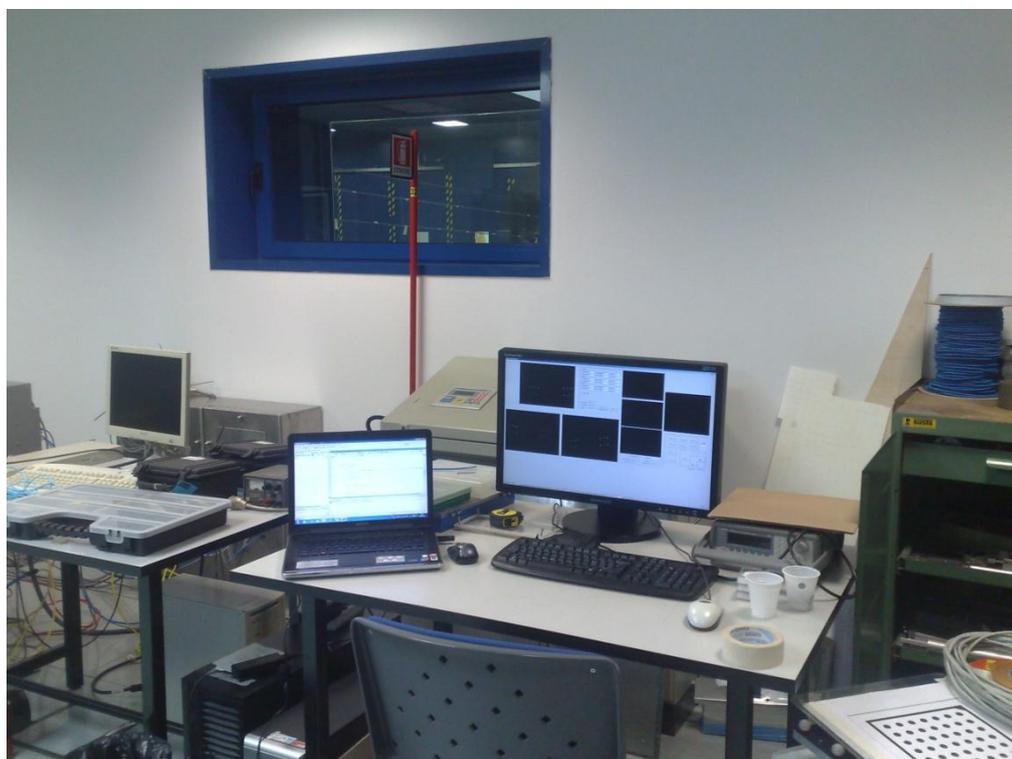
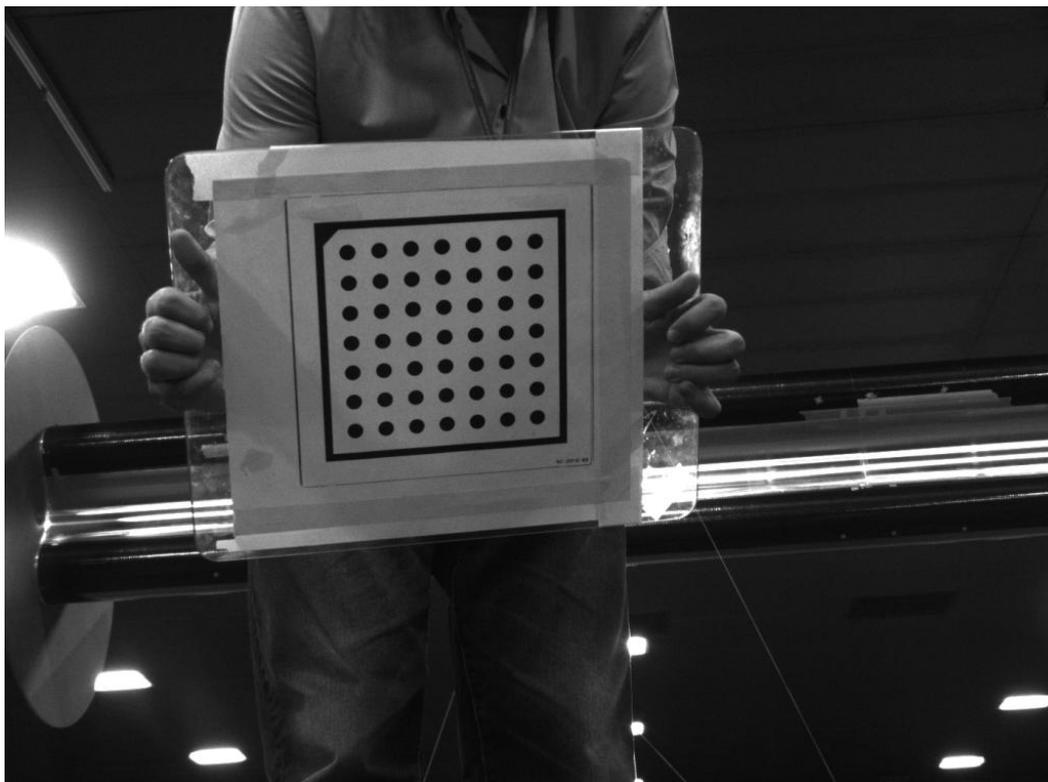


Figura 5.16: Sala controlli con il computer utilizzato per l'acquisizione.



**Figura 5.17:** Particolare raffigurante la scheda che genera il segnale di trigger.

La taratura del sistema è stata effettuata con il software sviluppato in questo lavoro acquisendo una serie di immagini del calibratore HALCON (figura 5.18). Essendo le telecamere molto distanti tra di loro e il calibratore molto piccolo rispetto al campo di misura, è stato necessario calibrare ogni singolo dispositivo separatamente, trovando i parametri intrinseci di ogni telecamera, per poi registrarle utilizzando alcune immagini acquisite in comune a tutte e tre i dispositivi. Quindi per ogni dispositivo si sono acquisite una serie di immagini del calibratore, come quella in figura 5.18. Si sono poi acquisite alcune immagini come quella di figura 5.19, in cui invece il calibratore è posto in modo da essere visibile alle tre telecamere. Chiaramente sarebbe stato possibile calibrare l'intero sistema stereoscopico direttamente dall'insieme di immagini utilizzate per la registrazione, ma come si nota, il calibratore è molto piccolo nell'immagine e quindi si sarebbe rischiato di introdurre incertezza in fase di calibrazione.



**Figura 5.18:** Esempio di immagine acquisita per la calibrazione dei singoli dispositivi.



**Figura 5.19:** Esempio di immagini acquisite per la registrazione dei singoli dispositivi.

I parametri intrinseci dei singoli dispositivi e i parametri estrinseci ricavati dalla registrazione sono riportati di seguito. L'errore di retroproiezione trovato è stato di 1.025 pixel.

PARAMETRI INTRINSECI	CAMERA 0	CAMERA 1	CAMERA 2
Focus [m]	0.0104649	0.0102559	0.0108421
K1 [1/m <sup>2</sup> ]	-2593.22	-2264.22	1908.84
K2 [1/m <sup>4</sup> ]	7.74857e+008	4.37801e+008	-3.90651e+007
K3 [1/m <sup>6</sup> ]	-5.83146e+013	-2.0161e+013	-4.6645e+012
P1 [1/m <sup>2</sup> ]	-1.14636	0.162211	-1.22307
P2 [1/m <sup>2</sup> ]	-1.06627	0.803395	0.0901016
Sx [m]	4.65964e-006	4.6174e-006	4.63374e-006
Sy [m]	4.65e-006	4.65e-006	4.65e-006
Cx [pix]	781.958	720.8	664.613
Cy [pix]	591.639	488.137	520.065

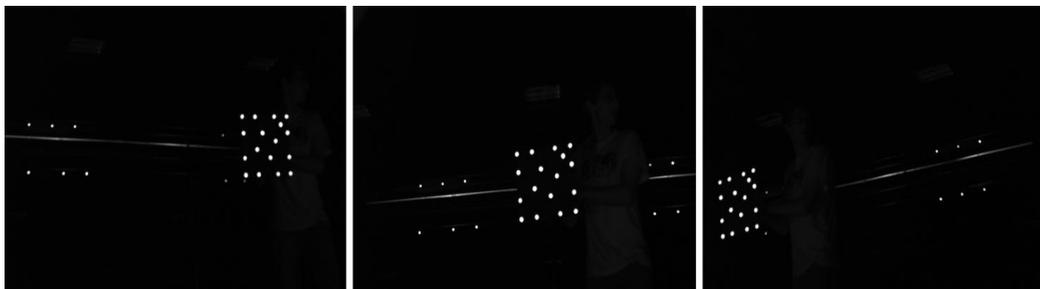
**Tabella 5.8:** Parametri intrinseci trovati calibrando i singoli dispositivi: lunghezza focale, distorsioni, dimensioni del pixel, centro ottico.

PARAMETRI ESTRINSECI	CAMERA 0	CAMERA 1	CAMERA 2
Tx [m]	0	2.02595	-1.71295
Ty [m]	0	-0.13164	0.128058
Tz [m]	0	0.0426514	-0.102726
a [grad]	0	0.44508	359.468
b [grad]	0	339.246	20.5695
g [grad]	0	7.01731	353.832

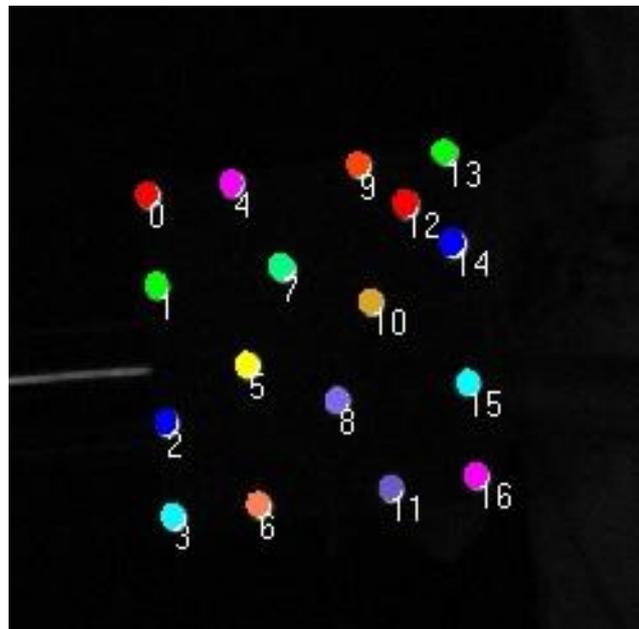
**Tabella 5.9:** Parametri estrinseci trovati dalla registrazione delle tre telecamere. Il sistema di riferimento è stato posto sulla telecamera 0, ovvero quella centrale.

## 5.2.1 Ricostruzione del calibratore a marker catarifrangenti

Prima di tutto è stata verificata la calibrazione ricostruendo il calibratore a marker catarifrangenti di figura 5.20, di cui si conoscono le distanze tra i marker con accuratezza elevata. In figura 5.20 sono mostrate le vedute di tale calibratore dalle tre diverse telecamere. Con la blob detection vengono trovati i marker del calibratore e ad ognuno di essi viene assegnato un numero, in modo da identificarlo rispetto agli altri, come mostrato in figura 5.21.



**Figura 5.20:** Esempio di immagini acquisite per la ricostruzione del calibratore infrarossi.



**Figura 5.21:** Numerazione assegnata ai marker del calibratore infrarossi dall'algoritmo di blob detection.

Sono state quindi calcolate le coordinate tridimensionali dei blob del calibratore. Il sistema di riferimento utilizzato in questo esperimento è quello della telecamera centrale. Le coordinate dei punti misurate sono riportate nella tabella 5.10. Un metodo per stimare la bontà della ricostruzione (e dunque anche della calibrazione), è il confronto fra le distanze tra i blob e le rispettive distanze reali. Alcuni risultati sono riportati nella tabella 5.11.

PUNTO	X	Y	Z
0	0.436842	-0.224904	3.35861
1	0.442941	-0.145318	3.34532
2	0.448841	-0.027665	3.32595
3	0.453161	0.0515989	3.31348
4	0.514317	-0.22576	3.37913
5	0.523348	-0.0690321	3.35254
6	0.530395	0.0495428	3.33274
7	0.558204	-0.147995	3.37505
8	0.603296	-0.0309194	3.36561
9	0.632004	-0.228322	3.40703
10	0.632004	-0.110181	3.38747
11	0.647721	0.0469813	3.3621
12	0.67256	-0.190403	3.41029
13	0.71012	-0.230172	3.42582
14	0.71351	-0.151303	3.41439
15	0.720016	-0.0339675	3.39428
16	0.724369	0.0450675	3.38149

**Tabella 5.10:** Coordinate dei marker del calibratore infrarossi.

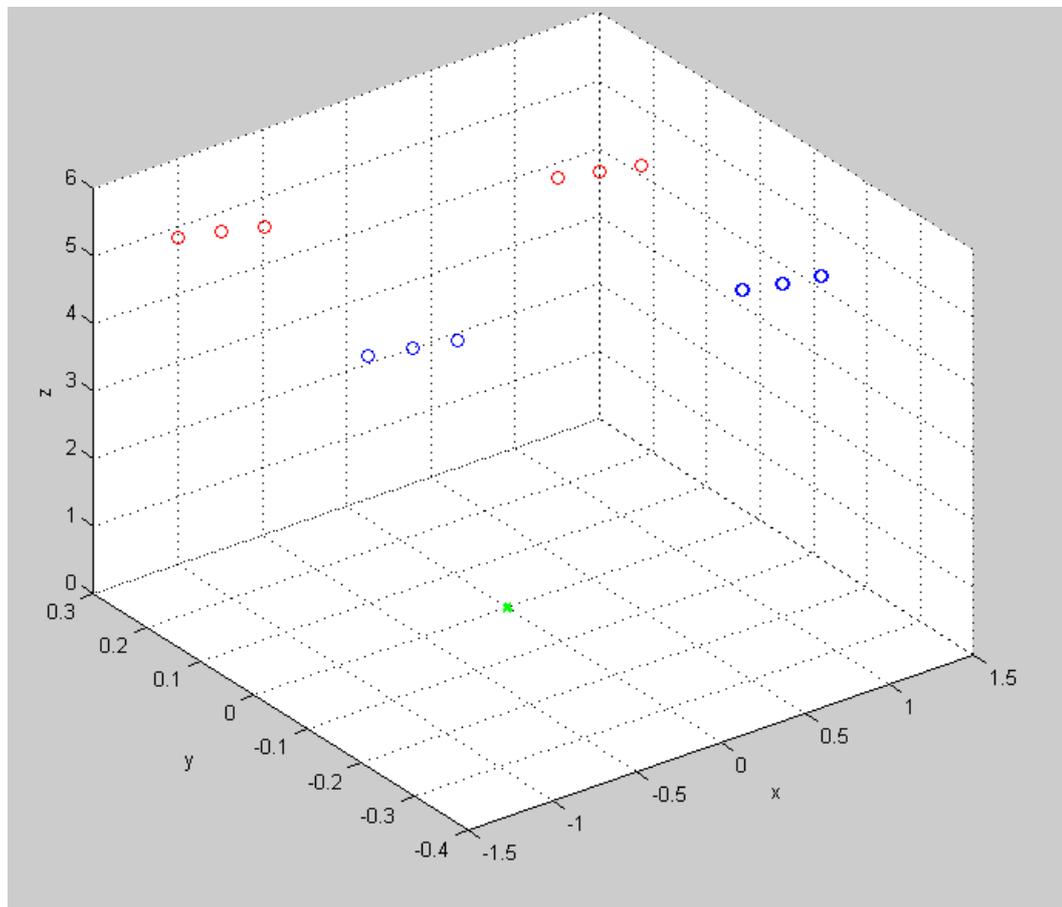
DISTANZA PUNTI	Valore Reale [mm]	Valore ricostruzione [mm]	da Scostamento Abs [mm]
0 - 4	80	80.1495	0.1495
4 - 9	120	120.978	0.978
9 -13	80	80.3641	0.3641
9 - 12	56.56854	55.6174	0.95114
9 -14	113.137	112.38	0.757
0 - 3	280	280.637	0.637
0 - 16	395.9798	395.069	0.9108
0 - 1	80	80.9197	0.9197
13 - 16	280	279.151	0.849
15 - 16	80	80.1821	0.1821

**Tabella 5.11:** Confronto tra le distanze tra alcuni marker trovate dalla ricostruzione e le distanze reali.

Dalla tabella 5.11 si nota che lo scostamento dalle distanze reali non supera mai il millimetro. Questa prova quindi mostra che lo strumento è ben calibrato, poiché l'accuratezza trovata ricostruendo il calibratore è dell'ordine di grandezza che ci si aspettava (ovvero inferiore al millimetro). Dall'esperienza maturata in laboratorio, si può anche aggiungere che con alcune accortezze è possibile migliorare l'accuratezza della calibrazione. Infatti il calibratore utilizzato per questa campagna di prove non era del tutto adatto. Per prima cosa il calibratore deve coprire almeno un quarto del campo di misura, mentre il pattern HALCON era ben più piccolo. Secondariamente l'utilizzo di un calibratore che non presenta marker catarifrangenti può introdurre errori di calibrazione dovuti a possibili movimenti dell'ottica quando si mette e toglie il filtro. Uno sviluppo possibile del software di calibrazione potrebbe essere quello di tener conto della possibilità di dover utilizzare un calibratore diverso da quello di HALCON e infrarossi, caratteristica che non è tuttora presente nel software sviluppato durante questo lavoro.

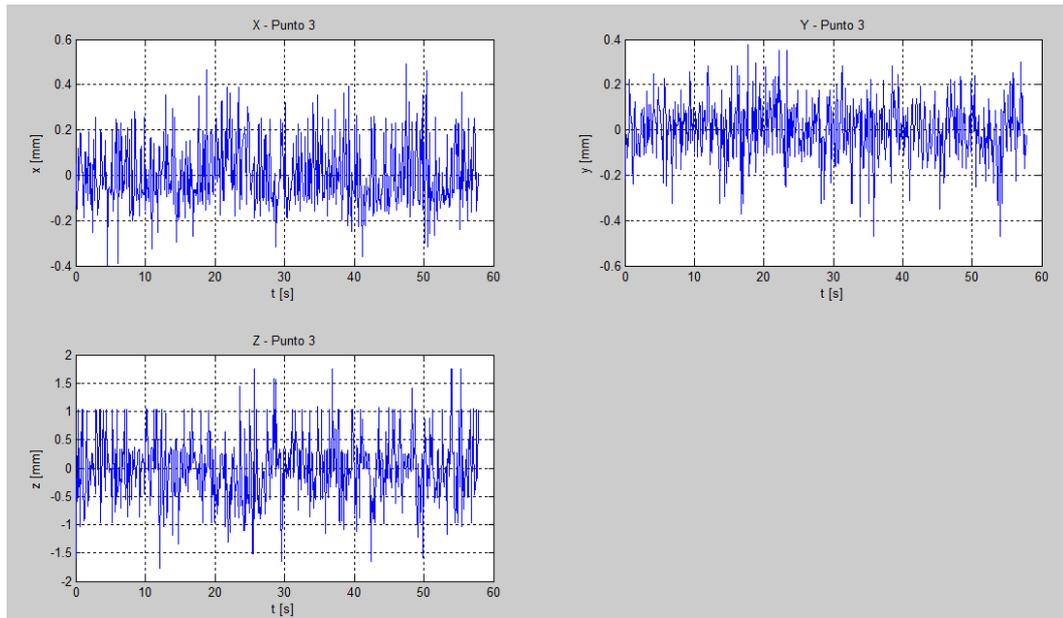
## 5.2.2 Prova con il modello fermo

Questa prova è stata eseguita come prima verifica del corretto funzionamento del sistema di misura. Viene semplicemente effettuata un'acquisizione con modello fermo e si verifica che non vi sia movimento tra i diversi istanti di acquisizione. Nella figura 5.22 sono rappresentati tutti i punti 3D in tutti gli istanti di acquisizione. In blu sono raffigurati i blob posti sul tubo superiore del modello e in rosso quelli sul tubo inferiore; in verde invece viene mostrato il centro del sistema di riferimento, ovvero la telecamera centrale. Già da questa rappresentazione tridimensionale si nota che non vi sono spostamenti a meno di piccolissime oscillazioni.

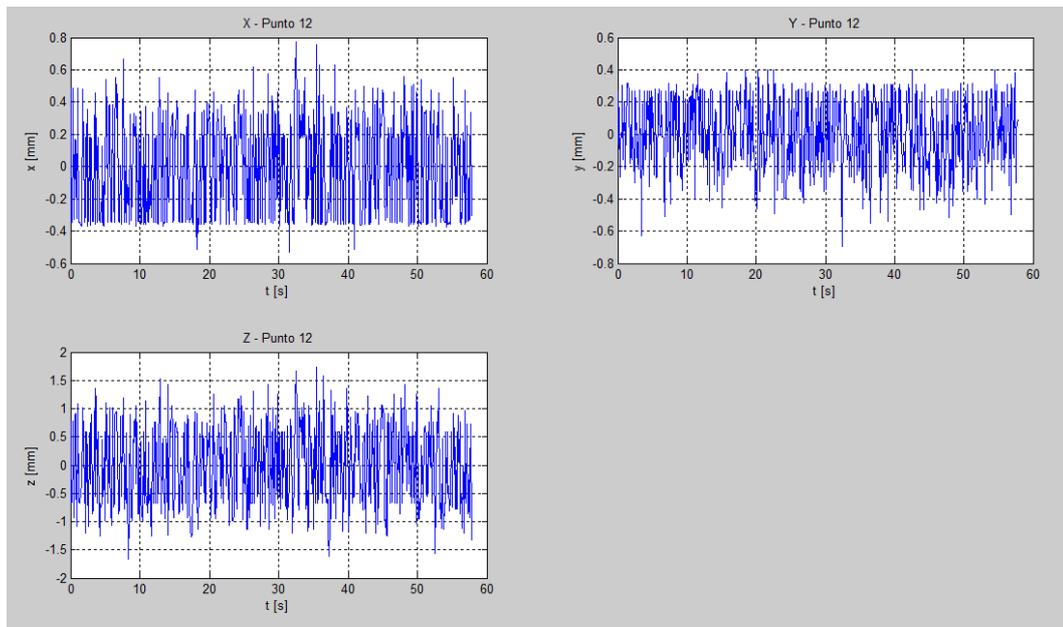


**Figura 5.22:** Rappresentazione delle posizioni dei marker nei diversi istanti di acquisizione. Si nota come non vi siano spostamenti a meno di piccolissime oscillazioni.

Per un'analisi più approfondita viene rappresentato lo spostamento di ogni punto sui tre assi del sistema di riferimento (solidale con la telecamera centrale) nel tempo. A titolo di esempio si riportano nelle figure 5.23 e 5.24 gli spostamenti su x, y e z dei punti 3 e 12.



**Figura 5.23:** Spostamento lungo i tre assi del punto 3.



**Figura 5.24:** Spostamento lungo i tre assi del punto 12.

Dall'analisi degli spostamenti si nota che le oscillazioni massime su x e y sono nell'ordine di  $\pm 0.5$  mm, mentre quelle su z, sul quale si hanno sempre scostamenti maggiori, sono dell'ordine di  $\pm 1.5$  mm. Lo stesso vale per tutti i dodici punti, anche quelli qui non rappresentati. L'incertezza maggiore quindi è presente sull'asse z ed è inferiore a  $\pm 1.5$  mm.

### 5.2.3 Verifica della rigidità del modello nel tempo

Un'altra prova utile a dimostrare che lo strumento di misura funziona correttamente è la verifica che le distanze tra i blob non cambino nel tempo, dato che il modello è rigido. A ogni istante quindi viene calcolata la matrice delle distanze tra tutti i punti (tabella 5.12) e viene sottratta alla matrice delle distanze dell'istante precedente, trovando una serie di matrici degli scostamenti. Una volta trovate tutte le matrici degli scostamenti, viene calcolata la matrice dello scostamento medio (tabella 5.13). Tale prova è stata ripetuta per diverse acquisizioni, portando sempre agli stessi risultati. Si riportano ad esempio i risultati relativi a una delle prove. Durante questa prova sono stati acquisiti 2923 campioni a 10 Hz.

	1	2	3	4	5	6	7	8	9	10	11	12
1	0	222.5581	439.4311	1.8449e+03	2.0389e+03	2.2375e+03	382.6024	436.1632	555.6322	1.8796e+03	2.0737e+03	2.2682e+03
2	222.5581	0	216.8798	1.6223e+03	1.8164e+03	2.0150e+03	437.4664	379.6804	418.0720	1.6617e+03	1.8548e+03	2.0486e+03
3	439.4311	216.8798	0	1.4055e+03	1.5995e+03	1.7981e+03	575.5803	441.9514	374.5895	1.4509e+03	1.6427e+03	1.8355e+03
4	1.8449e+03	1.6223e+03	1.4055e+03	0	194.0417	392.6844	1.8763e+03	1.6762e+03	1.4831e+03	369.1842	415.8815	536.4458
5	2.0389e+03	1.8164e+03	1.5995e+03	194.0417	0	198.6742	2.0669e+03	1.8660e+03	1.6719e+03	419.3627	368.3286	416.7992
6	2.2375e+03	2.0150e+03	1.7981e+03	392.6844	198.6742	0	2.2619e+03	2.0602e+03	1.8653e+03	539.1716	414.5438	362.4074
7	382.6024	437.4664	575.5803	1.8763e+03	2.0669e+03	2.2619e+03	0	204.6264	402.1189	1.8351e+03	2.0331e+03	2.2309e+03
8	436.1632	379.6804	441.9514	1.6762e+03	1.8660e+03	2.0602e+03	204.6264	0	197.5158	1.6304e+03	1.8285e+03	2.0263e+03
9	555.6322	418.0720	374.5895	1.4831e+03	1.6719e+03	1.8653e+03	402.1189	197.5158	0	1.4330e+03	1.6310e+03	1.8288e+03
10	1.8796e+03	1.6617e+03	1.4509e+03	369.1842	419.3627	539.1716	1.8351e+03	1.6304e+03	1.4330e+03	0	198.0514	395.8544
11	2.0737e+03	1.8548e+03	1.6427e+03	415.8815	368.3286	414.5438	2.0331e+03	1.8285e+03	1.6310e+03	198.0514	0	197.8152
12	2.2682e+03	2.0486e+03	1.8355e+03	536.4458	416.7992	362.4074	2.2309e+03	2.0263e+03	1.8288e+03	395.8544	197.8152	0

Tabella 5.12: Esempio di matrice delle distanze (in mm) all'ultimo istante 292,3 s.

	1	2	3	4	5	6	7	8	9	10	11	12
1	0	0.4363	0.3892	0.4185	0.3967	0.4493	0.6465	0.5896	0.5944	0.4141	0.4350	0.4340
2	0.4363	0	0.3504	0.3991	0.3822	0.4312	0.3465	0.5513	0.5857	0.4007	0.4164	0.4195
3	0.3892	0.3504	0	0.3569	0.3363	0.3811	0.2468	0.3072	0.3625	0.3112	0.3300	0.3310
4	0.4185	0.3991	0.3569	0	0.3664	0.3954	0.4300	0.4126	0.4180	0.6068	0.4457	0.3364
5	0.3967	0.3822	0.3363	0.3664	0	0.4242	0.3947	0.3653	0.3856	0.6375	0.5571	0.4515
6	0.4493	0.4312	0.3811	0.3954	0.4242	0	0.4841	0.4524	0.4530	0.6023	0.6327	0.5850
7	0.6465	0.3465	0.2468	0.4300	0.3947	0.4841	0	0.4109	0.3935	0.4255	0.4429	0.4424
8	0.5896	0.5513	0.3072	0.4126	0.3653	0.4524	0.4109	0	0.3594	0.3561	0.3796	0.3902
9	0.5944	0.5857	0.3625	0.4180	0.3856	0.4530	0.3935	0.3594	0	0.3399	0.3656	0.3878
10	0.4141	0.4007	0.3112	0.6068	0.6375	0.6023	0.4255	0.3561	0.3399	0	0.4040	0.3445
11	0.4350	0.4164	0.3300	0.4457	0.5571	0.6327	0.4429	0.3796	0.3656	0.4040	0	0.4008
12	0.4340	0.4195	0.3310	0.3364	0.4515	0.5850	0.4424	0.3902	0.3878	0.3445	0.4008	0

Tabella 5.13: Matrice dello scostamento medio tra le distanze (in mm).

Come si vede dalla matrice dello scostamento medio in tabella 5.13, i blob non si spostano tra un istante e l'altro a meno di uno scostamento medio di 0.35 mm e massimo di 0.65 mm. Inoltre viene confrontata la matrice delle distanze al primo istante con quella all'ultimo istante per verificare che non vi sia una sorta di deriva sulla posizione dei punti, trovando la matrice degli scostamenti in tabella 5.7.

	1	2	3	4	5	6	7	8	9	10	11	12
1	0	1.3586	0.7420	1.2731	0.7208	0.9225	0.9197	1.0563	1.1726	0.8334	0.8961	0.7662
2	1.3586	0	0.5765	0.0242	0.5831	0.3843	0.3224	0.8855	0.8954	0.7937	0.6960	0.7919
3	0.7420	0.5765	0	0.5456	0.0137	0.1894	0.2613	0.0919	0.3713	0.0108	0.0742	0.0215
4	1.2731	0.0242	0.5456	0	0.5295	0.3409	0.5823	0.3904	0.5561	0.7386	0.6104	0.3104
5	0.7208	0.5831	0.0137	0.5295	0	0.1970	0.0195	0.1715	0.0136	0.7937	0.5722	0.0624
6	0.9225	0.3843	0.1894	0.3409	0.1970	0	0.1440	0.0496	0.0987	0.7911	0.8358	0.2695
7	0.9197	0.3224	0.2613	0.5823	0.0195	0.1440	0	0.1569	0.0396	0.1418	0.1519	0.1036
8	1.0563	0.8855	0.0919	0.3904	0.1715	0.0496	0.1569	0	0.1206	0.0120	0.0034	0.2597
9	1.1726	0.8954	0.3713	0.5561	0.0136	0.0987	0.0396	0.1206	0	0.1157	0.1228	0.1341
10	0.8334	0.7937	0.0108	0.7386	0.7937	0.7911	0.1418	0.0120	0.1157	0	0.0146	0.2306
11	0.8961	0.6960	0.0742	0.6104	0.5722	0.8358	0.1519	0.0034	0.1228	0.0146	0	0.2472
12	0.7662	0.7919	0.0215	0.3104	0.0624	0.2695	0.1036	0.2597	0.1341	0.2306	0.2472	0

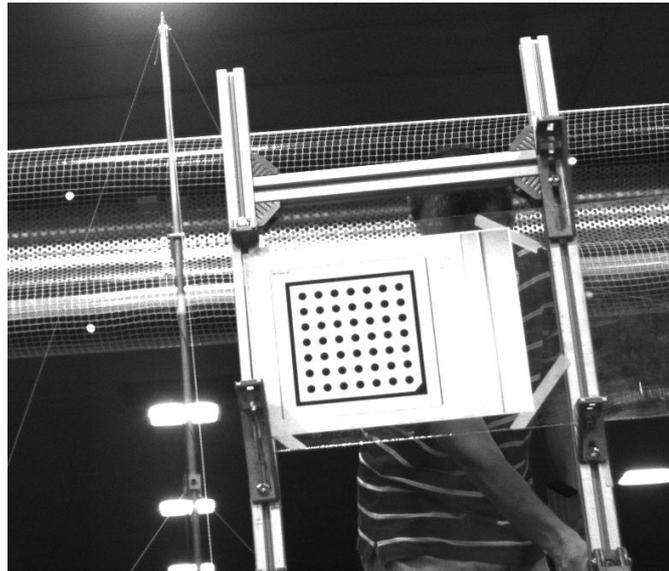
**Tabella 5.7:** Matrice degli scostamenti delle distanze all'ultimo istante e da quelle al primo istante (in mm).

Da questa si nota che lo scostamento massimo delle distanze relative tra i blob accumulato in circa 293 secondi è 1.5 mm. Gli stessi test sono stati eseguiti su gran parte delle prove effettuate, restituendo risultati simili ai precedenti. Da notare inoltre che nelle misure di distanze si combinano le incertezze relative a due diverse misure di posizione.

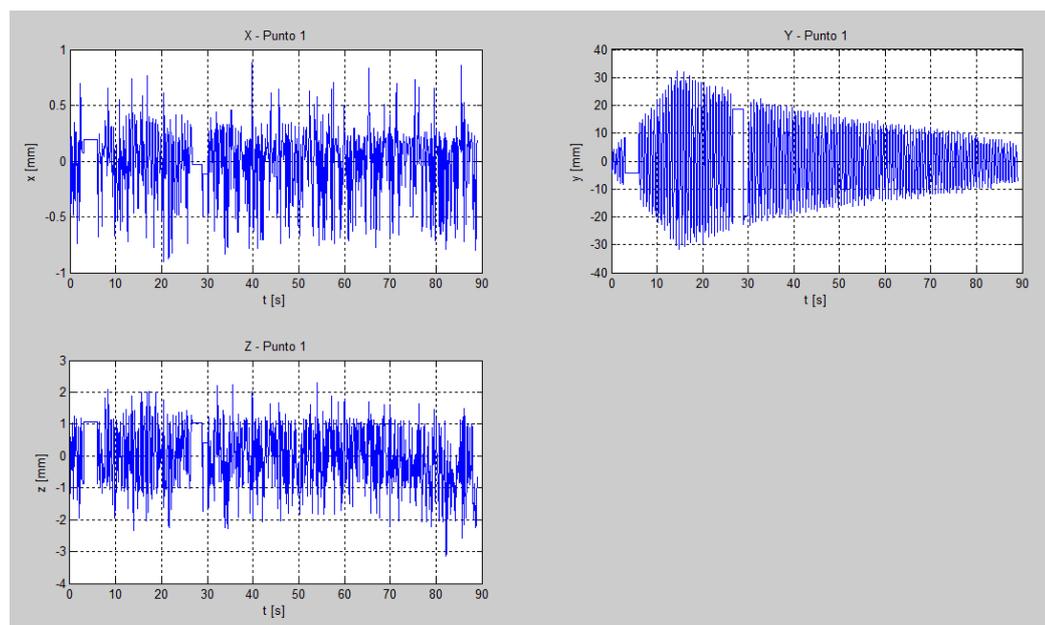
## 5.2.4 Verifica con calibratore verticale per il calcolo dell'incertezza su z

Nella prova descritta nel capitolo 5.2.2 è stato sottolineato come l'incertezza maggiore viene riscontrata sull'asse z, vale a dire l'asse perpendicolare al piano del sensore. In questa prova si vuole capire a quanto ammonta l'accuratezza sull'asse più critico e per fare ciò viene spostato il sistema di riferimento, che non si trova più sulla telecamera centrale ma nel centro del calibratore di figura 5.25, che è stato montato su una struttura che garantisce il fatto che esso sia in posizione perpendicolare al terreno. Così facendo la coordinata y del sistema di riferimento risulta perpendicolare al terreno, mentre z e x sono parallele a esso.

Facendo oscillare manualmente il modello si verifica in questa prova che vi siano componenti di spostamento solamente lungo  $y$  (spostamento verticale) e non lungo  $x$  e  $z$ . Si riporta dunque in figura 5.26 l'andamento temporale dello spostamento lungo i tre assi del punto 1. Essendo il moto del modello rigido, lo spostamento degli altri punti risulta molto simile e non viene quindi mostrato.



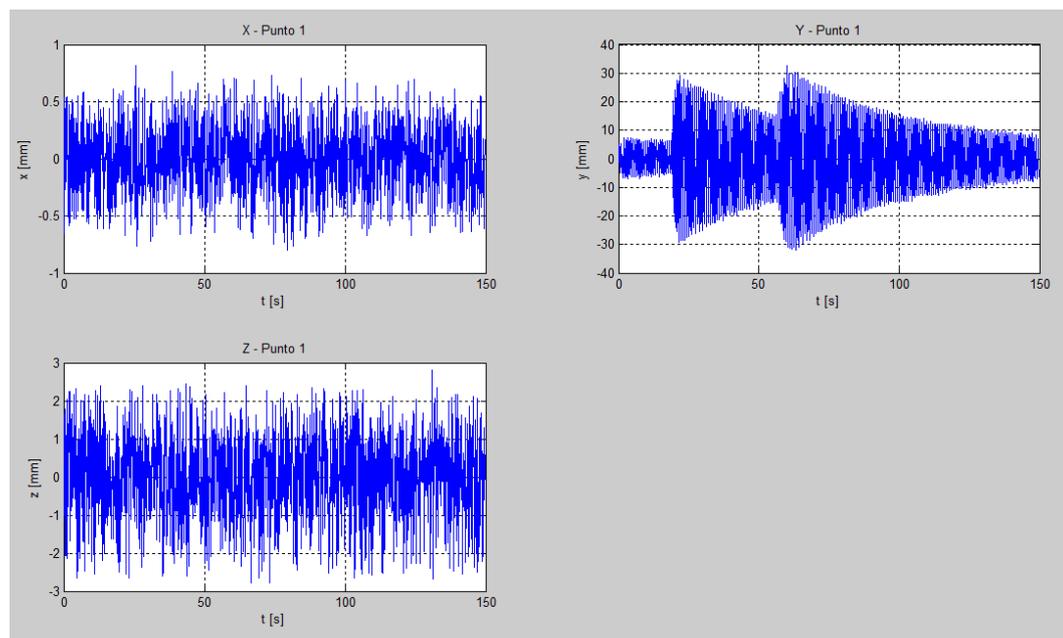
**Figura 5.25:** Calibratore sul quale è stato spostato il sistema di riferimento per questa prova. Esso è stato montato su una struttura che garantisce il fatto che esso sia in posizione perpendicolare al terreno.



**Figura 5.26:** Andamento temporale dello spostamento del punto 1.

Si nota come lo spostamento lungo  $y$  rappresenti lo spostamento verticale del modello (oscillazione smorzata), mentre lungo  $x$  si hanno spostamenti di massimo 1 mm e su  $z$  di massimo 3 mm. Questi ultimi non sono in realtà spostamenti dato che il modello si muove solo verticalmente, ma rappresentano rumore nella misura. In figura 5.26, per  $t \cong 5s$  e  $t \cong 28s$ , si nota che lo spostamento risulta costante su tutti gli assi per alcuni istanti. Questo è dovuto a una mancata acquisizione della posizione dei blob causata dal fatto che qualcuno è passato davanti alle telecamere oppure che la blob detection non ha trovato tutti i marker.

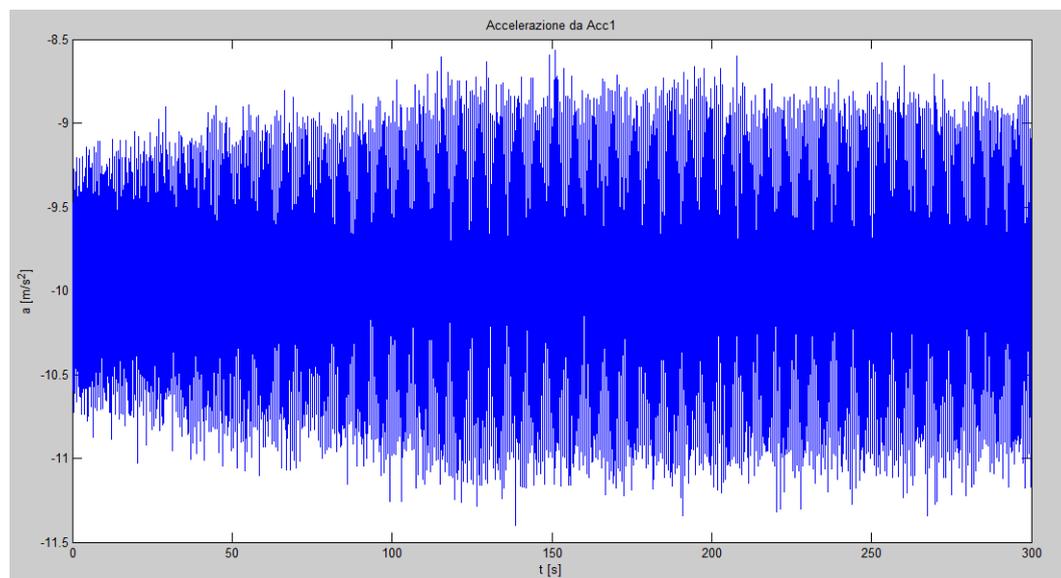
Per ulteriore conferma è stata eseguita una seconda prova uguale alla precedente. Gli andamenti temporali della posizione del blob 1 lungo i tre assi sono riportati in figura 5.27. Si nota come l'accuratezza lungo  $x$  sia ancora di circa 1 mm mentre lungo  $z$  non superi i 3 mm.



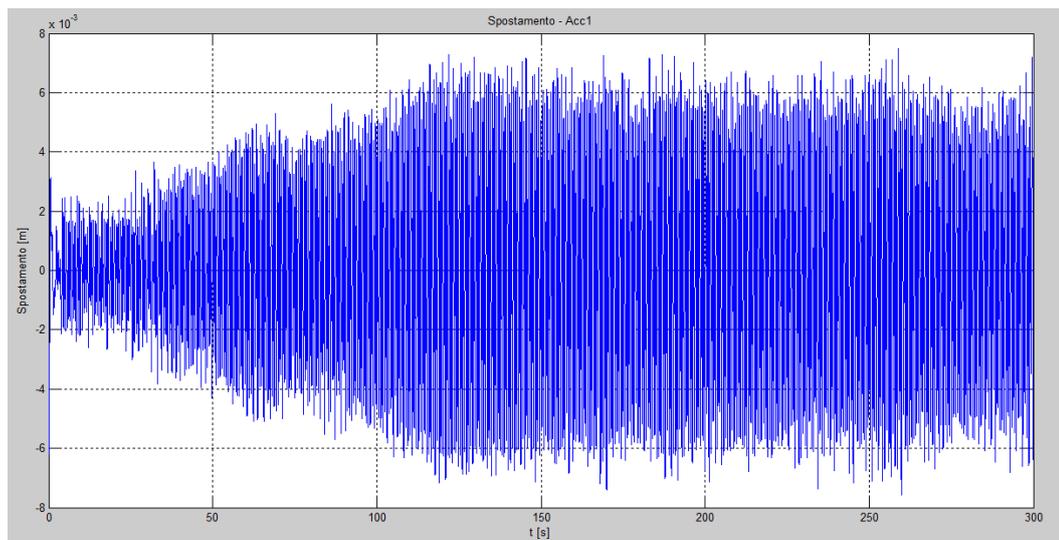
**Figura 5.27:** Andamento temporale dello spostamento del punto 1.

## 5.2.5 Confronto tra dati di Galleria e misure del sistema stereoscopico

Sul modello sono stati montati tre accelerometri che consentono di misurare l'accelerazione verticale e orizzontale dell'oggetto e verificare che non vi siano movimenti torsionali. Si sono dunque confrontate le misure effettuate con il sistema stereoscopico con quelle ricavate dalle acquisizioni degli accelerometri. In particolare si è scelto di confrontare la prova 130, in cui non vi è spostamento orizzontale e non vi sono movimenti torsionali. Il modello è dunque soggetto solamente a spostamenti verticali e sono di interesse solamente i dati acquisiti dall'accelerometro 1 che misura l'accelerazione verticale, mostrata in figura 5.28. I dati restituiti dall'accelerometro sono stati integrati due volte nel dominio delle frequenze (così da poter eliminare le componenti in bassa frequenza dovute all'accelerazione di gravità), trovando l'andamento temporale dello spostamento rappresentato in figura 5.29. In particolare il segnale di accelerazione è stato trasformato nel dominio delle frequenze e ogni armonica è stata divisa per  $\omega^2$  annullando le armoniche inferiori a 0.5 Hz. Infine il segnale è stato anti trasformato.



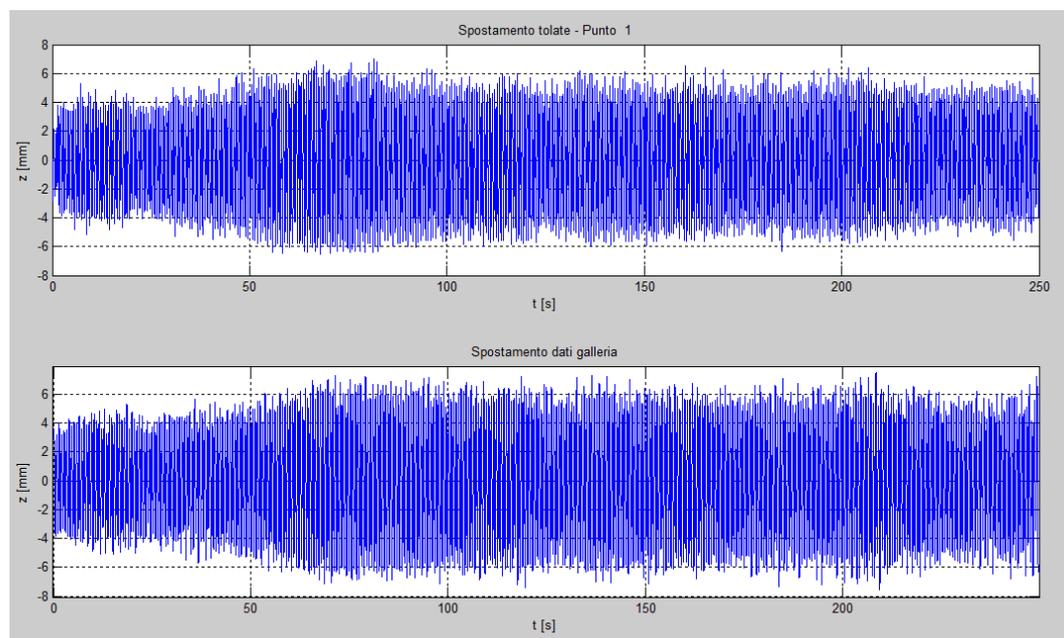
**Figura 5.28:** Andamento dell'accelerazione verticale del modello misurata dall'accelerometro 1.



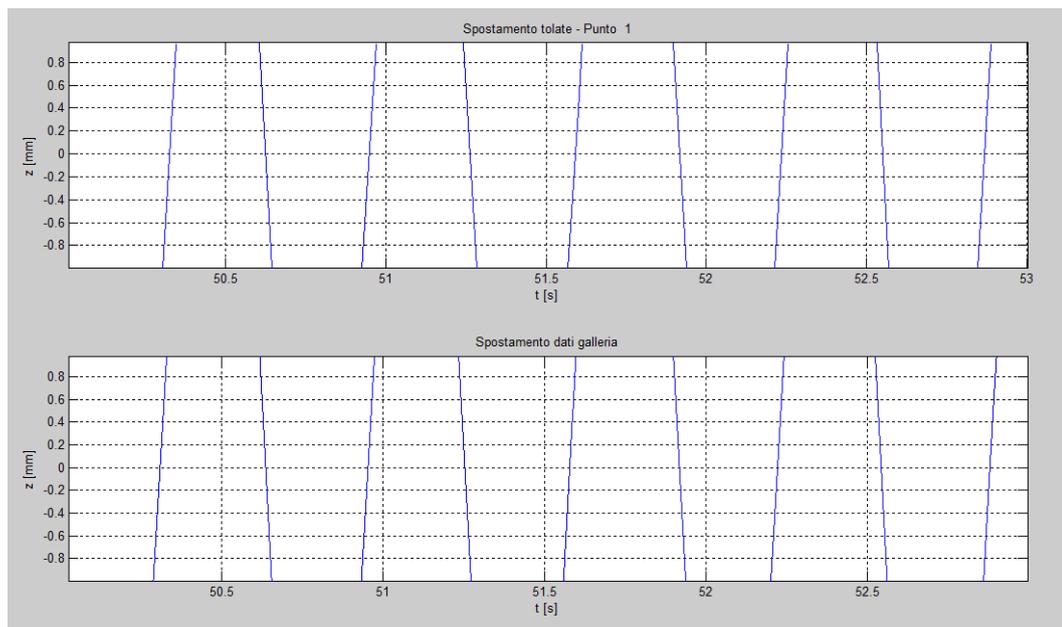
**Figura 5.29:** *Spostamento verticale ricavato integrando nel dominio delle frequenze i dati acquisiti dall'accelerometro 1.*

I dati acquisiti dal sistema stereoscopico invece sono riferiti ad un sistema di riferimento differente da quelli misurati dall'accelerometro. Infatti la tripletta è stata tarata imponendo come sistema di riferimento la posizione e la posa della telecamera centrale. Nelle misure allora si avrà una componente dominante di spostamento sull'asse y del sistema di riferimento e componenti inferiori di spostamento su x e z. I dati quindi sono stati rielaborati in modo da ricavare lo spostamento verticale reale, vale a dire lo spostamento sull'asse perpendicolare al terreno. Inoltre i segnali trovati con l'accelerometro e con il sistema stereoscopico non sono sincronizzati, poiché durante le prove non è stato possibile prendere alcun accorgimento che garantisse che l'acquisizione dei due differenti sistemi di misura cominciasse e terminasse nello stesso istante. I due segnali sono vanno dunque correlati in modo da essere paragonabili. In figura 5.30 quindi vengono mostrati i due segnali messi a confronto. In particolare si riporta nel primo grafico lo spostamento nel tempo del marker 1, dato che lo spostamento degli altri marker è quasi identico poiché il modello è rigido, ricavato dalle misure del sistema stereoscopico. Nel secondo grafico invece è raffigurato lo spostamento del modello trovato grazie all'accelerometro. Già da una prima analisi dei due segnali si nota che essi sono piuttosto simili, con oscillazioni che differiscono in ampiezza di circa 1 mm. Dall'ingrandimento di entrambi i segnali, mostrato in figura 5.31,

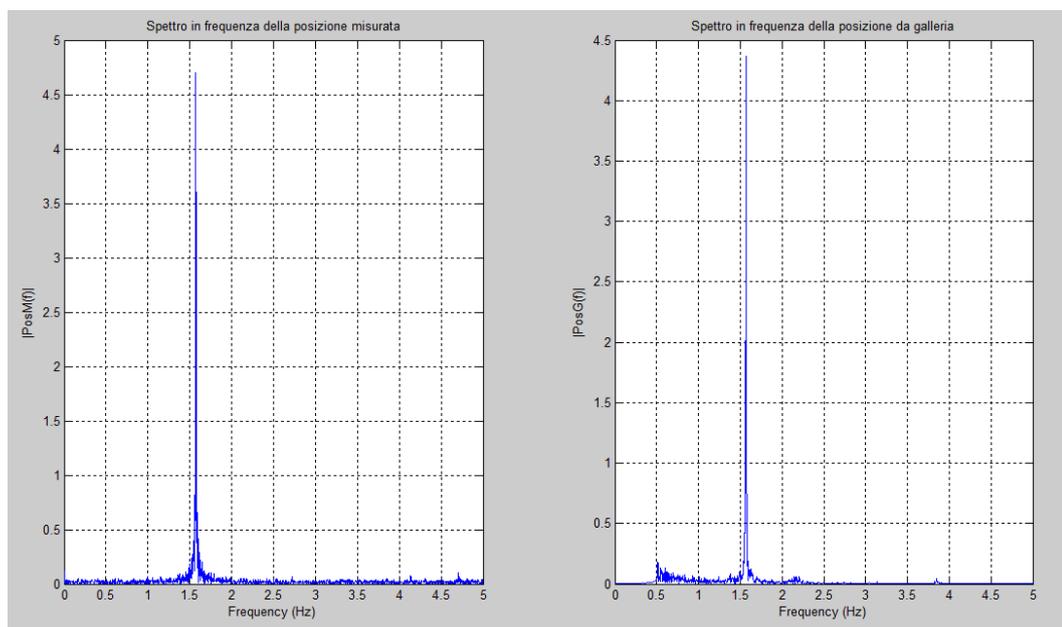
si può notare che anche la frequenza delle oscillazioni è la stessa per entrambi i segnali ed è circa 1.5 Hz. Per un'analisi più dettagliata quindi si è ricavato lo spettro in frequenza dei due segnali, mostrato in figura 5.32. Da questo è facile verificare che i due segnali hanno entrambi frequenza propria pari a 1.6 Hz. A tale frequenza corrisponde nel segnale misurato dal sistema stereoscopico un modulo di 4.7 mm, mentre nel segnale ricavato dall'accelerometro un modulo di 4.37 mm. L'incertezza in questo caso è dunque circa mezzo millimetro.



**Figura 5.30:** Confronto tra lo spostamento del modello misurato dal sistema stereoscopico e dall'accelerometro.



**Figura 5.31:** Ingrandimento degli andamenti temporali dello spostamento del modello misurato dal sistema stereoscopico e dall'accelerometro.



**Figura 5.32:** Spettro in frequenza dello spostamento del modello misurato dal sistema stereoscopico e dall'accelerometro.

## 6. Conclusioni

In questa tesi è stato realizzato uno strumento di misura ottico stereoscopico che consente di effettuare misure dinamiche delle coordinate tridimensionali di alcuni punti notevoli nella scena, con frequenze di acquisizione fino a 17 Hz. La scelta della stereoscopia rispetto ad altre tecniche ottiche di misura 3D è stata dettata da specifiche legate al volume di lavoro e all'incertezza di misura. Tali specifiche nascono dalle necessità della Galleria del Vento del Politecnico di Milano di effettuare misure dinamiche senza contatto, in modo da non disturbare la vena fluida del vento, anche se in realtà si è voluto realizzare uno sistema di misura che potesse essere utilizzato in un'ampia varietà di ambiti applicativi, con diverse condizioni e volumi di lavoro e diverse situazioni di illuminazione. Al fine di soddisfare i requisiti di accuratezza, è molto importante prestare particolare attenzione alla procedura di taratura dello strumento. Per questo motivo una parte rilevante del lavoro è stata dedicata allo studio e al confronto di diverse tecniche di calibrazione di telecamere. In particolare si è deciso di confrontare il metodo di Zhang ottimizzato con una particolare formulazione dell'algoritmo di Levenberg-Marquardt specifica per il problema della calibrazione, con una serie di metodi basati sugli algoritmi genetici (GA) e con una tecnica contenuta nella libreria commerciale per la visione artificiale HALCON, sviluppata da MVTEC. Mentre il

primo metodo è già ampiamente utilizzato e validato, per quanto riguarda gli algoritmi genetici si è inizialmente implementata una tecnica già presente in letteratura (*tecnica di Yang*) e si è dimostrato come essa impieghi un tempo eccessivamente elevato per raggiungere risultati paragonabili a quelli di Zhang-LM. Si è quindi deciso di implementare altre tecniche, nel tentativo di migliorare le prestazioni di metodi di calibrazione basati su algoritmi genetici. I risultati ottenuti mostrano che il metodo di Zhang-LM, all'attuale stato dell'arte, fornisce risultati migliori rispetto ai GA, poichè converge a una soluzione ottima in tempi di gran lunga inferiori. Tutte le tecniche che si basano sugli algoritmi genetici invece presentano lo svantaggio di essere eccessivamente lente e impiegano tempi molto elevati per arrivare a risultati paragonabili a quelli trovati con Zhang-LM, dunque per convergere alla soluzione ottima. Grazie però all'implementazione di una tecnica basata su algoritmi genetici che minimizzano l'errore di ricostruzione tridimensionale, invece che l'errore di retroproiezione come in Zhang-LM, è stato possibile dimostrare che la soluzione trovata con il metodo di Zhang-LM non solo è ottima per quanto riguarda l'errore di retroproiezione, ma è molto vicina all'ottimo globale anche per quanto concerne l'errore di ricostruzione tridimensionale.

Per quanto riguarda la tecnica di HALCON, invece, gli svantaggi consistono nel fatto che essa richiede la conoscenza della dimensione dei pixel del sensore e che raggiunge risultati accurati solamente con molte immagini di calibrazione, rispettando più possibile i vincoli imposti dalla procedura di calibrazione. Siccome per le telecamere industriali la dimensione dei pixel è ricavabile dal datasheet, il primo svantaggio non comporta in realtà alcun problema. Vantaggi della procedura di HALCON sono invece che essa permette di controllare la qualità delle immagini acquisite e che è veloce e facile da utilizzare, anche da parte di un utente meno esperto. Inoltre da alcuni esperimenti è stato provato che i risultati della tecnica di HALCON sono compatibili a quelli ricavati con il metodo di Zhang-LM, se le immagini sono acquisite rispettando i vincoli imposti. Non sempre è però possibile acquisire immagini che rispettino fedelmente tali vincoli, riducendo così l'accuratezza della procedura di calibrazione rispetto alla tecnica di

Zhang-LM. Si è comunque deciso di utilizzare la tecnica di HALCON, ma si sottolinea che se è richiesta un'accuratezza particolarmente elevata è consigliabile controllare i risultati di essa con la tecnica di Zhang-LM. Tale scelta è stata effettuata soprattutto per garantire una più agevole manutenibilità del software di gestione del sistema. Si è infatti creato un tool di calibrazione inserito nel pacchetto software, utilizzando il linguaggio di programmazione C# e la libreria HALCON. È vero che la taratura della tripletta viene fatta offline a monte delle acquisizioni, ma è stato ritenuto più adeguato l'utilizzo dello stesso linguaggio di programmazione per tutto il pacchetto software.

Il software che gestisce lo strumento è stato interamente sviluppato durante la tesi e consente di acquisire immagini, effettuare misure stereoscopiche e registrare diversi strumenti. Quest'ultima caratteristica è stata prevista poiché in applicazioni, in cui ad esempio il misurando è grande o presenta curvature elevate, può essere necessario utilizzare più strumenti di misura contemporaneamente. In un sistema costituito da più strumenti stereoscopici, è indispensabile che la procedura di registrazione non introduca incertezze sensibili nelle misure. Per questo si è voluto dedicare parte del lavoro all'implementazione di una tecnica di registrazione basata sulla decomposizione ai valori singolari (*svd*) e alla sua verifica metrologica. La procedura di registrazione è stata verificata mediante una serie di prove che hanno dimostrato che l'incertezza introdotta dipende dal numero di punti utilizzato nell'algoritmo. In particolare utilizzando meno di 4 punti si introduce un'incertezza superiore al millimetro, mentre con un numero di punti maggiore di 8 è possibile limitare le incertezze al decimo di millimetro. Inoltre è stato dimostrato che, nel caso si decida di utilizzare pochi punti, è conveniente scegliere punti sparsi nell'immagine, che ricoprono il più possibile la superficie del sensore.

L'intero sistema di misura è stato validato metrologicamente mediante una serie di prove in laboratorio su un calibratore planare. Si è così verificato che l'incertezza dello strumento è pari a 0,5 mm, come richiesto dalle specifiche. Infine lo strumento è stato applicato a una campagna sperimentale presso la Galleria del Vento del Politecnico di Milano. In particolare si è studiato il

comportamento dinamico di un modello di una struttura che servirà per aumentare la stabilità di uno dei palazzi del nuovo complesso CityLife di Milano, soggetto alla forza del vento. Da queste prove si è potuto verificare il corretto funzionamento dello strumento e l'efficacia delle sue caratteristiche di flessibilità e scalabilità.

## **6.1 Sviluppi futuri**

### **6.1.1 Calibrazione di telecamere**

In questa tesi si è voluto verificare le prestazioni di metodi di calibrazione basati sugli algoritmi genetici, confrontandoli con l'algoritmo di Zhang ottimizzato mediante Levenberg-Marquardt, già utilizzato e validato anche presso il VBLab. Inoltre si è deciso di verificare l'accuratezza di una tecnica contenuta nella libreria commerciale per la visione artificiale HALCON, che è stata utilizzata per la taratura del sistema stereoscopico sviluppato per ragioni di mantenibilità del software. In letteratura esistono numerose altre metodologie basate su tecniche anche molto diverse fra loro. Un esempio è dato da alcuni metodi di calibrazione che si basano sull'utilizzo delle reti neurali, in qualche caso anche ottimizzate con gli algoritmi genetici [18]. Nonostante il gran numero di tecniche di calibrazione non esiste uno standard che definisce quale sia la metodologia migliore da utilizzare, anche perché non è presente in letteratura nessuno studio comparativo completo sui vari metodi presenti. Sarebbe di grande interesse allargare l'analisi e il confronto proposto in questa tesi alle molte altre tecniche di calibrazione. Da uno studio accurato sarebbe interessante capire se esiste una tecnica di calibrazione più accurata delle altre o se la scelta della metodologia dipende dall'ambito applicativo e definire una linea guida per decidere quale metodo utilizzare.

## 6.1.2 Sistema di misura

Lo strumento di misura stereoscopico sviluppato in questa tesi rispetta le specifiche imposte ed è stato utilizzato in una campagna di prove presso la Galleria del Vento, con risultati soddisfacenti. Durante le prove sono stati riscontrati alcuni limiti del sistema, tra cui il principale è costituito dalle prestazioni dinamiche. Lo strumento è infatti in grado di acquisire a frequenze inferiori a 17 Hz. Per superare tale problematica è necessario utilizzare telecamere con prestazioni maggiori. Ciò non è sufficiente poiché si saturerebbe la capacità di trasmissione del sistema. Si potrebbe allora utilizzare più schede di acquisizione, aumentando però eccessivamente i costi dello strumento. Un'alternativa migliore consiste nel limitare il flusso di dati alle informazioni di interesse per le misure, e dunque implementare algoritmi intelligenti che riducano la ROI (*region of interest*) fisica acquisita dai dispositivi all'area che contiene informazioni utili alla misura. In tal caso sono indispensabili telecamere con un sensore a tecnologia CMOS. Inoltre, l'utilizzo di PC embedded per ogni dispositivo, e dunque interamente dedicati all'acquisizione dalla singola telecamera, costituirebbe un'ulteriore soluzione per l'aumento delle prestazioni dinamiche del sistema. Ciò consentirebbe anche di realizzare un sistema più compatto e maneggevole, rendendo le procedure di allestimento più semplici e veloci. A tale scopo è in progetto presso il VBLab la realizzazione di un nuovo strumento che possa presentare queste caratteristiche. Per questioni di comodità e maneggevolezza sarebbe anche utile poter comandare lo strumento via wireless da remoto e sfruttare la nuova tecnologia PoE (Power over Ethernet), che consente di alimentare le telecamere attraverso il cavo ethernet. Lo sviluppo di uno strumento compatto e comandabile da remoto risulterà particolarmente utile per le applicazioni in Galleria del Vento, dove è importante ridurre al minimo i tempi di allestimento e poter comandare il sistema dalla sala di controllo.



# Bibliografia

- [1] R. I. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Second edition. Cambridge University Press, 2003.
- [2] Zhengyou Zhang. *A Flexible New Technique for Camera Calibration*. Microsoft Research, One Microsoft Way, Redmond, USA. December 1998.
- [3] W. H. Press, S. A. Teukolsky, W. T. Vetterling, B. P. Flannery. *Numerical Recipes. Third edition*. Cambridge University Press, 2007.
- [4] Pradit Mittrapiyanuruk, *A Memo on How to Use the Levenberg-Marquardt Algorithm for Refining Camera Calibration Parameters*, Robot Vision Laboratory, Purdue University, West Lafayette, IN, USA.
- [5] Manolis I.A. Lourakis and Antonis A. Argyros. *Is Levenberg-Marquardt the Most Efficient Optimization Algorithm for Implementing Bundle Adjustment?*. Institute of Computer Science, Foundation for Research and Technology, Crete, GREECE.
- [6] Andrea Fusiello. *Visione Computazionale*. Università degli Studi di Verona, Italy, 2008.
- [7] Matteo Pedraglio. *Analisi e qualificazione metrologica di calibrazioni di telecamere matriciali per applicazioni di Bin-Picking*. Politecnico di Milano, Italy, 2009.

- [8] Roger Y. Tsai. *A Versatile Camera Calibration Techniaue for High-Accuracy 3D Machine Vision Metrology Using Off-the-shelf TV Cameras and Lenses*. IEEE, Journal of Robotics and Automation, Vol. RA-3, August 1987.
- [9] Wang Qi, Fu Li, Liu Zhenzhong. *Review on Camera Calibration*. Shenyang Institute of Aeronautical Engineering, Shenyang 110136, China.
- [10] Hou Zhenjie, Zhao Jing, Gu Ligu, Lv Guoling. *Automatic Calibration Method Based on Traditional Camera Calibration Approach*. Inner Mongolia Agriculture University, Hohhot, China.
- [11] D.C. Brown. *Decentering Distorsion of Lenses*. Photometric Engineering, pages 444-462, Vol. 32, No. 3, 1966.
- [12] R.C. Gonzales, R.E. Woods, S.L. Eddins. *Digital Image Processing Using Matlab*. Pearson Education, 2004.
- [13] K.S. Fu, R.C. Gonzalez, C.S.G. Lee. *Robotics: Control, Sensing, Vision and Intelligence*. Mc Graw-Hill, 1987.
- [14] Z.J.Yang, F.Chen, J.Zhao, H.W.Zhao. *A Novel Camera Calibration Method Based on Genetic Algorithm*. School of Mechanical Science and Engineering, Jilin University Changchun, China.
- [15] Mooz Bourhooicha, Mahamed Ben Kbelifa, William Puceh. *A Non Linear Camera Calibration with Genetic Algorithms*. University of Toulon and University of Montpellier, France.

- [16] Yongnian Zhang and Qiang Jit. *Camera Calibration With Genetic Algorithms*. IEEE, Transactions on Systems, Man and Cybernetics, Part A: Systems and Humans, Vol. 31, No. 2, March 2001.
- [17] Moez Bouchouicha. *A Genetic Algorithm Application to Stereo Calibration*. Toulon-Var Southern University, France.
- [18] Liu Wan-Yu, Xie Kai. *A Camera Calibration Method based on Neural Network optimized by Genetic Algorithm*. Heilongjiang Scientific Research Foundation and the Scientific Research Foundation for the Returned Overseas Chinese Scholars.
- [19] Melanie Mitchell. *An Introduction to Genetic Algorithms*. Cambridge, Massachusetts and London, England. Fifth printing, 1999.
- [20] Andrew Chipperfield, Peter Fleming, Hartmut Pohlheim, Carlos Fonseca. *Genetic Algorithm Toolbox*. University of Sheffield.
- [21] R.L Haupt, S.E. Haupt. *Practical Genetic Algorithms*. May. Wiley, 2004.
- [22] J.R Koza, *Genetic Programming: on the Programming of Computers by Means of Natural Selection*, MIT Press, Cambridge, MA, 1992.
- [23] Y. Ma, S. Soatto, J. Kosecka, S. S. Sastry. *An Innovation to 3-D Vision*. Springer, 2004.
- [24] E. Trucco, A. Verri. *Introductory Techniques for 3-D Computer Vision*. Prentice Hall, 1998.

- [25] C. Harris and M. Stephens. *A Combined Corner and Edge Detector*. Proceedings of the 4th Alvey Vision Conference. pp. 147–151. 1988.
- [26] P.E. Gill and W. Murray. *Algorithms for the Solution of Non-Linear Leastsquares Problem*. SIAM J Num Anal,15(5):977-992, 1978.
- [27] Arun, Huang, Blostein. *Least-Squares fitting of Two 3-D Point Sets*. IEEE Transaction on Pattern Analysis and Machine Intelligence. Vol. PAMI-9, No. 5, September 1987.
- [28] Simone Pavan. *Confronto fra un sensore “CMOS HDR” ed un sensore “CMOS NIR” per la misura di mappe termiche*. Politecnico di Milano, Italy, 2011.
- [29] [www.mvtec.com/halcon/](http://www.mvtec.com/halcon/)