

POLITECNICO DI MILANO

Scuola di Ingegneria dell'Informazione

Corso di Laurea Magistrale in
Ingegneria dell'Automazione



**QUALIFICAZIONE METROLOGICA DI TELECAMERA A TEMPO DI VOLO
MICROSOFT KINECT V2
PER APPLICAZIONI IN AMBITO NEURORIABILITATIVO**

Relatore: Ing. Remo SALA

Correlatore: Ing. Giacomo MAINETTI

Tesi di Laurea di:

Andrea CORTI Matr. 802145

Anno Accademico 2014 - 2015

*per aspera
ad astra*

Indice

Sommario	x
Abstract	xii
Introduzione	1
1 Ambito applicativo	3
2 Metodiche di scansione 3D: stato dell'arte	8
2.1 Sistemi di acquisizione con contatto	9
2.2 Sistemi di acquisizione senza contatto	10
2.2.1 Sistemi trasmissivi	10
2.2.2 Sistemi riflessivi	10
2.2.2.1 Sistemi non ottici	11
2.2.2.2 Sistemi ottici	11
3 Modello matematico di una telecamera	22
3.1 Definizione	22
3.2 Camera oscura	23
3.3 Modello pinhole	24
4 Modello matematico di un dispositivo a tempo di volo	31
4.1 Tempo di volo continuous wave (CW)	31
4.2 Telecamera a tempo di volo	35
4.2.1 Microsoft Kinect V2	36
4.2.1.1 Software di acquisizione	40
5 Qualificazione metrologica	45
5.1 Stabilità del sistema: influenza della temperatura	46
5.2 Caratterizzazione del singolo pixel	49

5.2.1	Effetto casuale relativo alla misura di un singolo pixel nello spazio	49
5.2.2	Effetto sistematico relativo alla misura di un singolo pixel	52
5.2.3	Influenza dell'angolo di riflessione	54
5.2.4	Influenza del materiale e del colore	56
5.3	Caratterizzazione dell'intero sensore	58
5.3.1	Performance nel ricostruire geometrie note	59
5.3.1.1	Piano	59
5.3.1.2	Cilindro	61
5.3.1.3	Sfera	62
5.3.2	Mixed pixels	64
5.3.3	Riflessioni multiple	65
5.4	Conclusioni	68
6	Analisi metrologica del tracciamento di una figura umana	70
6.1	Effetto casuale relativo alla misura della posizione di un giunto	74
6.2	Confronto tra Kinect V2 e sistema BTS (movimenti semplici)	75
6.3	Confronto tra Kinect V2 e sistema BTS (protocollo di trattamento neuroriabilitativo)	80
6.4	Conclusioni	86
7	Applicazione neuroriabilitativa	87
8	Conclusioni	91
8.1	Sviluppi futuri	93
A	Registrazione rigida	97
A.1	Singular Value Decomposition	97
A.2	Iterative Closest Point	101
B	Procedura di allineamento tra telecamera a tempo di volo e bersaglio planare	103
C	Articolo	107
	Acronimi	108
	Bibliografia	110
	Riferimenti citati nel testo	110
	Pubblicazioni e Manuali	110
	Materiale Online	113

Elenco delle figure

1.1	Esempi di dispositivi meccatronici per la riabilitazione. . .	4
2.1	Sistemi di ricostruzione 3D.	9
2.2	Particolare di un tastatore per misura con contatto.	10
2.3	Schema rappresentante il principio di funzionamento di un sonar.	11
2.4	Ricostruzione 3D mediante shape from shading (a) e comportamento delle superfici lambertiane (b).	12
2.5	Esempio di texture deterministica (a) e statistica (b).	13
2.6	Schema base di un sistema di visione stereoscopica in cui è evidenziato l'accoppiamento di punti omologhi (a) ed esempio di dispositivo stereoscopico sperimentale (b).	14
2.7	Principio di funzionamento dello shape from silhouettes.	16
2.8	Principio di funzionamento della triangolazione attiva.	18
2.9	Interferometria: proiezione della griglia sul misurando.	19
2.10	Principio di funzionamento di un dispositivo tof impulsato.	20
3.1	Geometria del modello di una camera oscura (a) e rappresentazione dell'effetto di ribaltamento dell'immagine (b).	24
3.2	Geometria del modello lineare di camera pinhole frontale.	25
3.3	Immagine soggetta a distorsioni radiali a botte (a) e rappresentazione delle diverse tipologie di distorsioni radiali (b).	28
4.1	Schematizzazione e principio di funzionamento di un sistema di misura lineare ottico tempo di volo CW.	32
4.2	Principio di estensione della distanza di misura utilizzando una tecnica a frequenza multipla.	34
4.3	Esempio di telecamera a tempo di volo commerciale.	35
4.4	Microsoft Kinect V2. Componenti (a) e sistema di riferimento (b).	37

4.5	Nome e disposizione dei giunti individuati da Kinect V2. . .	38
4.6	Interfaccia grafica del software di acquisizione sviluppato per la qualificazione metrologica del dispositivo.	40
4.7	Esempio di header relativo ad un file contenente le coordinate di una nuvola di punti salvata in formato pcd.	42
4.8	Interfaccia grafica del software di acquisizione appositamente sviluppato per la qualificazione metrologica del dispositivo.	43
5.1	Andamento dei dati e del valore medio relativi ad acquisizioni effettuate a ventola spenta.	46
5.2	Kinect V2 prima e dopo l'installazione del sistema di raffreddamento esterno.	47
5.3	Andamento dei dati e del valore medio relativi ad acquisizioni effettuate a ventola accesa.	48
5.4	Campioni di distanza acquisiti per 24 ore (1 Hz) senza ventola esterna (a) e con ventola esterna (b).	49
5.5	Setup utilizzato per la caratterizzazione del singolo pixel. Sulla flangia del robot è fissato un calibratore utilizzato nella procedura di allineamento.	50
5.6	Andamento dell'effetto casuale relativo alla misura restituita dal pixel centrale al variare della distanza tra camera e piano acquisito, mostrato in valore assoluto (a) e percentuale (b).	51
5.7	Andamento dell'effetto casuale relativo alla misura di distanza valutato all'interno del sensore (a) e del campo di visione (b).	52
5.8	Andamento dell'effetto sistematico lungo l'asse Z, normale al piano (a) e curva di best fitting (b).	53
5.9	Andamento dell'effetto sistematico all'interno del frame. . .	54
5.10	Andamento del valore di deviazione standard in funzione dell'angolo di incidenza tra luce e bersaglio.	54
5.11	Relazione tra effetto casuale relativo alla misura di distanza e angolo di incidenza tra raggio di luce e bersaglio.	55
5.12	Risultati delle prove eseguite con campioni di colori diversi: cartoncino (a), stoffa (b), pellicola adesiva (c).	56
5.13	Risultati delle prove eseguite con campioni di materiali differenti.	57
5.14	Setup costituito da un piano sul quale è stato applicato un marker circolare catarifrangente (a) e nuvola di punti ottenuta dalla scansione (b).	58
5.15	Nuvola di punti acquisita (a) e piano di best fitting (b). . .	59

5.16	Distanza in millimetri calcolata tra il modello ed ogni singolo punto della nuvola (a) e distribuzione dell'errore (b). . . .	60
5.17	Distanza in millimetri calcolata tra il piano e ogni singolo punto della nuvola, considerandone una porzione centrale pari a circa il 70% della superficie del sensore.	60
5.18	Nuvola di punti acquisita (a) e sovrapposizione al modello di riferimento (b).	61
5.19	Distanza in millimetri calcolata tra il modello ed ogni singolo punto della nuvola (a) e distribuzione dell'errore (b). . . .	62
5.20	Ritaglio di una porzione della nuvola di punti relativa alla superficie sferica (a) e sovrapposizione tra sfera di riferimento e punti acquisiti (b).	63
5.21	Distanza radiale in millimetri calcolata tra il modello ed ogni singolo punto della nuvola (a) e distribuzione dell'errore (b).	63
5.22	Nuvola di punti (a) e setup (b) utilizzati per l'identificazione dell'errore di mixed pixel.	64
5.23	Sezioni passanti per il centro dell'acquisizione in cui sono evidenziati i "mixed pixels".	65
5.24	Schema rappresentante il percorso compiuto dai raggi luminosi in presenza di una scena in cui si verificano riflessioni multiple (a) e acquisizione distorta relativa ad una geometria concava (b).	66
5.25	Setup utilizzato per l'analisi delle riflessioni multiple; si notino i marker utilizzati per ricavare la rototraslazione dei due sistemi di riferimento.	67
5.26	Allineamento tra nuvola di punti e piani di riferimento (a) e distanza punti-piano (b).	67
6.1	Alcuni marker posti nel volume di lavoro (a) e dettaglio di un singolo marker (b).	72
6.2	Paziente durante l'esecuzione di un esercizio (si notino il Kinect V2 posizionato frontalmente rispetto al paziente e due delle otto telecamere BTS poste sulla parete di fondo) (a) e dettaglio raffigurante una telecamera BTS SMART-DX (b).	73
6.3	Principali assi e piani del corpo umano.	74
6.4	Scomposizione e sovrapposizione delle traiettorie relative alle rotazioni compiute dal polso destro, acquisite con sistema BTS e con Kinect V2 (paziente in posizione frontale). . . .	76

6.5	Scomposizione e sovrapposizione delle traiettorie relative alle rotazioni compiute dal polso destro, acquisite con sistema BTS e con Kinect V2 (paziente in posizione laterale). . . .	78
6.6	Scomposizione e sovrapposizione delle traiettorie relative alle rotazioni compiute dal polso destro, acquisite con sistema BTS e con Kinect V2 (paziente con asse sagittale a 45° rispetto all'asse della tof camera).	79
6.7	Scomposizione e sovrapposizione delle traiettorie relative al polso destro durante l'esercizio di elevazione nel piano frontale, acquisite con sistema BTS e con Kinect V2. . . .	82
6.8	Scomposizione e sovrapposizione delle traiettorie relative al polso destro durante l'esercizio di reaching frontale, acquisite con sistema BTS e con Kinect V2.	82
6.9	Scomposizione e sovrapposizione delle traiettorie relative al polso destro durante l'esercizio di avvicinamento della mano alla bocca, acquisite con sistema BTS e con Kinect V2. . .	83
6.10	Report relativo allo svolgimento dell'esercizio di avvicinamento della mano alla bocca.	85
7.1	Esempio di esercizio presente nell'applicazione sviluppata da MediaHospital S.r.l., svolto con l'utilizzo del visore Oculus Rift.	88
7.2	Esempio di realtà virtuale, presente nell'applicazione sviluppata da MediaHospital S.r.l., in cui sono ricostruiti alcuni oggetti della vita quotidiana.	89
7.3	Sequenza fotografica relativa allo svolgimento di un esercizio presente nell'applicazione sviluppata da MediaHospital S.r.l.	90
8.1	Interfaccia dell'applicazione di analisi del cammino in cui è mostrato il paziente in posizione eretta (a) e rilevamento di una caduta (b).	95
8.2	Singola nuvola di punti acquisita (a) ed estensione del campo visivo grazie all'unione di due nuvole di punti (b).	96
B.1	Condizione di allineamento ricercata tra le terne XYZ e $X'Y'Z'$ rispettivamente associate alla camera e al bersaglio.	104
B.2	Alcune pose assunte dal calibratore durante la fase di allineamento.	105

Elenco delle tabelle

4.1	Kinect V2 for Windows - principali caratteristiche [41]. . .	37
4.2	Kinect V2 - requisiti minimi di sistema [41].	38
4.3	Esempio di parametri intrinseci di calibrazione relativi alla telecamera IR.	39
4.4	Elenco riportante i gradi di libertà relativi ai giunti tracciati da Kinect V2 ai quali sono associati un nome ed un indice numerico.	44
5.1	Caratteristiche tecniche della ventola aggiunta al fine di stabilizzare la temperatura interna al dispositivo.	47
6.1	BTS SMART-DX 7000 - principali caratteristiche [39]. . .	71
6.2	Valori di deviazione standard ottenuti dall'analisi dei dati acquisiti nelle tre prove.	75
6.3	Scostamenti medi tra le traiettorie relative alle rotazioni compiute dal polso destro acquisite con sistema BTS e con Kinect V2 (paziente in posizione frontale).	78
6.4	Scostamenti medi tra le traiettorie relative alle rotazioni compiute dal polso destro (paziente in posizione laterale). .	79
6.5	Scostamenti medi tra le traiettorie relative alle rotazioni compiute dal polso destro (paziente con asse sagittale a 45° rispetto all'asse della tof camera).	80
6.6	Dati risultanti dall'analisi dell'esercizio di elevazione nel piano frontale.	84
6.7	Dati risultanti dall'analisi dell'esercizio di reching frontale.	84
6.8	Dati risultanti dall'analisi dell'esercizio di avvicinamento della mano alla bocca.	84

Sommario

L'obiettivo del presente lavoro di tesi è la qualificazione metrologica della telecamera a tempo di volo Kinect V2. Tale dispositivo è stato presentato da Microsoft nel 2014 come accessorio per applicazioni videoludiche. Esso infatti integra algoritmi in grado di tracciare la posizione di alcuni punti di una figura umana, permettendole, grazie al movimento del corpo, il controllo di un videogioco. Inoltre, poiché prodotto su larga scala, viene venduto ad un costo decisamente inferiore rispetto a dispositivi industriali integranti la medesima tecnologia.

Si è deciso, per queste ragioni, di proporre tale strumento come supporto ad un'applicazione in fase di sviluppo da parte di MediaHospital S.r.l., in collaborazione con il Presidio dell'Ospedale Valduce "Villa Beretta" di Costa Masnaga (LC).

L'applicazione, basandosi sul concetto di serious game, si pone come obiettivo quello di permettere al paziente lo svolgimento dell'attività riabilitativa in ambiente domestico, fornendogli indicazioni riguardanti la correttezza dei movimenti compiuti durante gli esercizi prescritti dal medico curante che, a sua volta, a valle delle registrazioni effettuate nel tempo, può delineare un completo quadro evolutivo dello stato del paziente.

Lo studio del dispositivo in esame è stato quindi necessario al fine di determinare le potenzialità ed i limiti dell'apparecchio così da stabilirne l'adeguatezza, in termini di accuratezza di misura, all'applicazione considerata.

Quindi, nella presente tesi vengono inizialmente descritte alcune analisi preliminari, utili a risolvere il problema legato alla deriva della misura e causato dall'instabilità termica del dispositivo.

In seguito, sono riportati lo svolgimento ed i risultati di alcune prove finalizzate a valutare l'effetto casuale e sistematico relativi alla misura associata sia al singolo pixel sia all'intero sensore, in relazione a forma, colore, posa e superficie dell'oggetto inquadrato.

Infine, è stata valutata l'accuratezza nel tracciamento della posizione di alcuni punti dell'arto superiore sfruttando gli algoritmi di Kinect V2 e

comparando le traiettorie restituite da quest'ultimo con quelle acquisite con il sistema stereoscopico BTS già presente nella clinica.

I risultati ottenuti al termine dell'analisi rivelano che l'apparecchio risulta idoneo alla suddetta applicazione e, più in generale, a tutte le applicazioni in cui è richiesto un grado di accuratezza, relativo alle posizioni individuate, non superiore a venti millimetri.

Parole chiave: visione artificiale; 3D; telecamera a tempo di volo; Kinect V2; nuvola di punti; qualificazione metrologica; neuroriabilitazione; analisi del movimento umano

Abstract

The purpose of this thesis is the metrological qualification of the Kinect V2 time of flight camera. It was presented by Microsoft in 2014 as peripheral device for video gaming applications, and, in fact, it uses algorithms for human body tracking, allowing a player to control videogames through his body movements. Moreover, because of its large-scale production, it is cheaper than other industrial devices integrating the same technology.

For these reasons, this device has been proposed in support of a medical application developed by MediaHospital S.r.l., in collaboration with the district of the Valduce Hospital of “Villa Beretta” in Costa Masnaga (LC).

This application, based on the serious game idea, allows a patient to perform the rehabilitation activities at home and guides him in the correct execution of the movements. The trajectories described during the exercise, by his arm joints, are recorded and sent to the medical staff that can analyse the data to evaluate the progresses achieved by the patient.

The metrological characterization of the considered device has been conducted to outline its strengths and limits, and then, to establish its suitability, in terms of accuracy, for the application mentioned above.

Thus, in this thesis some preliminary analysis useful to evaluate the measure drift problem generated by the thermal instability of the device are described.

Then, the development and the results of some tests performed to quantify the random and the systematic effect are reported. In particular, these tests regard the measure performed by a single pixel or by the full sensor during the acquisitions of differently shaped, oriented and coloured objects.

Finally, the accuracy in tracking upper limb joint positions has been evaluated comparing data provided by Kinect V2 and data returned by BTS stereoscopic system already installed in the ward.

The results obtained from the analysis reveal that the device is suitable for the above-mentioned application and, more generally, for all the

applications in which the required accuracy related to the joint position does not exceed twenty millimetres.

Keywords: computer vision; 3D; time of flight camera; Kinect V2; point cloud; metrological qualification; neurorehabilitation; human motion analysis

Introduzione

Il presente lavoro di tesi sperimentale è stato svolto all'interno del Vision Bricks Laboratory (VBLab), laboratorio di ricerca che opera nel Dipartimento di Meccanica del Politecnico di Milano e che si occupa principalmente di visione artificiale bi e tridimensionale e di misure senza contatto.

La trattazione in seguito riportata riguarda l'analisi di un dispositivo commerciale a basso costo (Kinect V2) il quale è presente all'interno del laboratorio da alcuni mesi come technical preview ed è ora disponibile sul mercato.

Esso, sviluppato da Microsoft come interfaccia tra giocatore e videogioco, è in grado, servendosi della tecnologia 3D a tempo di volo (tof), di tracciare il movimento di una o più figure umane individuate all'interno della ricostruzione virtuale della scena inquadrata.

Lo strumento è stato scelto come componente hardware per un'applicazione per personal computer in grado di guidare il paziente permettendogli, a seguito di un trauma, lo svolgimento dell'esercizio riabilitativo in autonomia ed all'interno della propria abitazione.

Al fine di garantirne l'adeguatezza in termini di accuratezza nella rilevazione del movimento, è stata svolta un'approfondita qualificazione metrologica dello strumento, presentata all'interno di questo lavoro.

Outline

Il testo della presente tesi è così strutturato e suddiviso:

il primo capitolo introduce l'ambito applicativo del seguente lavoro presentando un confronto tra le tradizionali tecniche riabilitative e quelle innovative basate su serious games;

il secondo capitolo illustra lo stato dell'arte relativo alle metodiche di scansione 3D proponendo le varie tecniche presenti sul mercato e valutando la migliore soluzione al problema in esame;

nel terzo capitolo e nel quarto capitolo sono presentati rispettivamente il modello matematico di telecamera pinhole e di singolo distanziometro a tempo di volo (tof) i quali costituiscono la base di funzionamento di una telecamera a tempo di volo;

il quinto capitolo descrive una serie di prove sperimentali relative al processo di qualificazione metrologica mediante il quale è stato caratterizzato l'aspetto distanziometrico della tof camera (Kinect V2) utilizzata nel presente lavoro, così da valutarne l'adeguatezza in relazione ad un impiego scientifico;

il sesto capitolo descrive lo svolgimento di alcune prove mediante le quali è stata analizzata l'accuratezza nel tracciamento di alcuni punti del corpo umano, durante l'esecuzione di specifici esercizi neuromotori, al fine di garantire un possibile impiego del dispositivo in ambito clinico e riabilitativo;

nell'appendice A sono brevemente illustrate le basi matematiche su cui si fondano le tecniche di registrazione rigida SVD e ICP, spesso utilizzate nel presente lavoro;

nell'appendice B è illustrata e descritta la tecnica mediante la quale è stato possibile allineare un bersaglio piano al sensore della tof camera, così da poter effettuare misure relative a distanze note;

l'appendice C riporta l'abstract di un articolo, attualmente in fase di revisione da parte dell'editore Elsevier, nel quale sono presentate le prove ed i risultati relativi alla qualificazione metrologica riguardante l'aspetto distanziometrico del Kinect V2.

Capitolo 1

Ambito applicativo

Nel seguente capitolo viene inizialmente descritto lo stato dell'arte relativo alle tecniche riabilitative tradizionali, prestando particolare attenzione a quelle utilizzate nel caso di patologie neuropsicologiche causanti disordine del movimento dell'arto superiore ed analizzandone vantaggi e svantaggi in relazione alle nuove tecniche basate sul concetto di serious games. Successivamente sono delineate le caratteristiche di un'applicazione attualmente in fase di sviluppo da parte dell'azienda MediaHospital S.r.l., la quale implementa un serious game ed utilizza un dispositivo di ricostruzione 3D e tracciamento del movimento del corpo umano la cui scelta e qualificazione metrologica sono trattate nel presente lavoro.

A seguito di un trauma e/o di un'operazione chirurgica, al fine di garantire al paziente il totale o parziale recupero delle capacità motorie e sensoriali, è prassi che da parte di quest'ultimo venga intrapresa una serie di sedute nelle quali sono svolti esercizi a fine riabilitativo. Nello specifico si ha una prima fase, che avviene durante il periodo di ricovero, in cui il paziente, seguito dal personale medico, svolge tali esercizi all'interno della struttura clinica. Successivamente, in seguito alla dimissione, egli continua la riabilitazione recandosi periodicamente in ospedale e, contemporaneamente, intervallando le sedute cliniche con esercizi autonomi svolti presso la propria abitazione con l'ausilio, se necessario, di familiari e/o personale qualificato.

Nel caso specifico di patologia di natura neurologica causante disordine

del movimento, il recupero lento e talvolta solo parziale delle funzioni, determina un processo riabilitativo che si protrae per parecchi mesi.

Focalizzando l'analisi sulla riabilitazione dell'arto superiore, alcuni esempi di esercizi classici della terapia occupazionale, tutt'ora utilizzati al fine di recuperare le competenze motorie della vita quotidiana e lavorativa, consistono nell'esecuzione di movimenti semplici, quali lo spostamento ed il posizionamento di piccoli oggetti (sfere o cubi) o il raggiungimento di punti noti dello spazio o del proprio corpo. Ciò è svolto grazie all'aiuto del terapeuta il quale, soprattutto inizialmente, guida il paziente nell'esecuzione dell'azione e lo ferma nell'eventualità in cui quest'ultimo svolga movimenti scorretti e potenzialmente dannosi.

Inoltre, da alcuni anni, si è sviluppato un nuovo filone di tecniche riabilitative basate sull'utilizzo di macchine robotiche e meccatroniche (figura 1.1) le quali, sotto la supervisione del terapeuta, possono imporre movimenti controllati e ripetibili all'arto del paziente oppure esercitare una resistenza (viscosa o elastica), la quale deve essere vinta dall'azione muscolare del soggetto. Oltre ad essere un ausilio durante l'esercizio, queste macchine sono in grado di acquisire dati utili a monitorare lo stato di avanzamento della terapia.

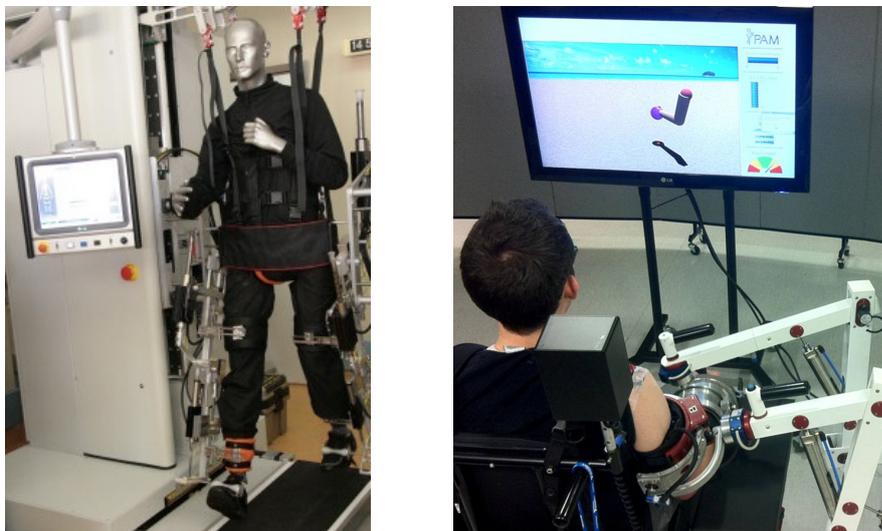


Figura 1.1. Esempi di dispositivi meccatronici per la riabilitazione.

La costanza e la dedizione nell'eseguire queste attività sono condizioni fondamentali grazie alle quali è possibile garantire ottimi risultati ma la scarsa diffusione sul territorio nazionale di strutture dotate di tecnologie all'avanguardia e di personale qualificato, comporta spesso lunghe trasferte

da parte del paziente il quale, quasi sempre accompagnato dai familiari poiché non più autosufficiente, mal tollera questi continui spostamenti. Inoltre, l'ambiente clinico non familiare e poco stimolante di certo non è d'aiuto durante il percorso riabilitativo.

Per queste e per altre simili ragioni si sta sviluppando la tendenza a fornire metodi e strumenti finalizzati a garantire lo svolgimento di una corretta attività riabilitativa svolta (molto spesso) in autonomia e presso la propria abitazione.

Da alcuni anni infatti, in relazione alla sempre maggiore diffusione di dispositivi tecnologici e videoludici a basso costo come pc, laptop, tablet, smartphone, ecc. . . vengono sviluppate applicazioni, basate sul concetto di serious games (SGs) [9], in grado di svolgere la funzione di ausilio alla riabilitazione.

Il termine serious game venne probabilmente usato la prima volta da Clark Abt, nel 1987. Nel suo libro *Serious Games* [1], il ricercatore statunitense fornisce una definizione di giochi educativi, ai quali, a prescindere dal supporto (digitale o real-world) attribuisce un primario ed esplicito scopo educativo, il quale però non esclude una componente, seppur secondaria, legata al divertimento.

Successivamente, nell'anno 2002, con la fondazione della *Serious Games Initiative* ad opera di Ben Sawyer è convenzionalmente fissata la nascita del settore videoludico dei SGs.

L'espressione serious game, a prima vista ossimorica, definisce un'attività interattiva, spesso di natura digitale, che attraverso la forma del gioco, promuove l'apprendimento di contenuti e competenze differenti in maniera autonoma. I serious game, infatti, trovano oggi applicazioni in numerosi contesti che spaziano dalla formazione aziendale alle campagne di educazione e di sensibilizzazione, dalle simulazioni di vendita telefonica, colloqui e riunioni fino all'insegnamento di materie scolastiche come la matematica o la storia. Nei SGs si ha un elevato grado di fedeltà nella ricostruzione di una realtà virtuale talvolta molto coinvolgente ed immersiva.

Negli ultimi anni si sta sviluppando un numero sempre crescente di applicazioni riabilitative basate su questi concetti [21, 31] ed utilizzabili in ambiente domestico sia in maniera autonoma, sia con l'aiuto di un familiare, a valle di una prima fase di apprendimento svolta in ambiente ospedaliero e sotto la supervisione di personale clinico e tecnico qualificato. Ciò ha praticamente eliminato la necessità da parte del paziente di doversi recare personalmente presso poli ospedalieri specializzati, permettendo quindi lo svolgimento dell'attività riabilitativa in qualsiasi momento della giornata ed all'interno di un ambiente conosciuto e familiare. In aggiunta a ciò si ha una riduzione dei costi sostenuti dal servizio sanitario nazionale.

Inoltre la grafica accattivante, la somiglianza ad un tradizionale videogame e la variabilità del contesto e della tipologia di esercizio rendono il tutto molto versatile ed attraente, soprattutto per i pazienti giovani (appartenenti alla generazione dei “nativi digitali”) spesso tediati da esercizi monotoni e noiosi [32]. È infatti molto importante, al fine di garantire una maggiore efficacia della terapia, che l’azione non sia passiva e che il paziente sia attento e partecipe durante lo svolgimento dell’esercizio.

Infatti è dimostrato che chi basa il proprio processo di riabilitazione sull’utilizzo di questi dispositivi ed applicazioni svolge gli esercizi assegnati in maniera molto più sistematica e continuativa rispetto a ciò che avviene con le tecniche tradizionali [24].

Queste tecnologie trovano applicazione anche in relazione al trattamento di disturbi cognitivi negli anziani (es. malattia di Parkinson o malattia di Alzheimer) grazie a realtà virtuali che mostrano ambienti di tipo familiare e coinvolgono il paziente in situazioni caratterizzate da attività quotidiane. In questo modo è possibile rallentare il processo di invecchiamento cerebrale ed attuare un training focalizzato sullo stimolo della memoria, dell’attenzione e delle capacità logico-esecutive, così da contrastare disturbi come il decadimento cognitivo, la demenza neurodegenerativa e la depressione.

Tutto ciò può essere monitorato da un sistema remoto di raccolta dati in grado di compilare dei report ed informare il clinico sullo svolgimento dell’attività da parte del paziente.

Si ha quindi lo svolgimento contemporaneo sia dell’attività riabilitativa sia della raccolta di dati relativi alla qualità dell’esecuzione dell’esercizio. Alcuni di questi parametri, impossibili da stimare “ad occhio” dal terapeuta, sono il jerk (indice della fluidità del movimento), l’angolo di flessione ed estensione di un arto, le distanze relative agli spostamenti compiuti, ecc. . .

Inoltre, acquisendo ed analizzando anche altri parametri come il tempo di svolgimento dell’esercizio, i risultati ottenuti, il livello di attenzione ed il ritmo cardiaco è possibile rilevare lo stato di miglioramento del paziente così da garantire una costante progressione del livello di difficoltà dell’esercizio praticamente in tempo reale e senza la necessità di effettuare numerosi colloqui personali e visite di controllo.

Il caso pratico in esame tratta una particolare applicazione per personal computer basata sul concetto di serious game la quale, per mezzo di un dispositivo in grado di ricostruire la forma tridimensionale del corpo umano e rilevarne i movimenti, permette di controllare i gesti di un avatar mostrato a schermo il cui compito è quello di eseguire semplici azioni quotidiane come la manipolazione e lo spostamento di oggetti virtuali da un contenitore ad un altro.

L'obiettivo del presente lavoro riguarda inizialmente l'individuazione e successivamente la qualificazione metrologica di un dispositivo hardware presente sul mercato in grado di ricostruire il modello 3D della figura umana e tracciarne i movimenti del corpo (in particolare dell'arto superiore) così da costituire l'interfaccia tra il paziente e l'applicazione.

Vengono ora introdotte e descritte le specifiche dell'hardware necessario all'applicazione, in relazione all'utilizzo finale ed al campo di impiego. Come prima cosa, dal momento che il prodotto finito deve poter essere installato in ambiente domestico ed utilizzato anche da personale non qualificato, si richiede che esso disponga di un hardware poco ingombrante e che sia di semplice connessione, ad esempio tramite lo standard USB presente su ogni personal computer.

Inoltre avendo la necessità di poter operare in ambienti caratterizzati da condizioni molto variabili (es. presenza e disposizione di oggetti nella stanza, illuminazione naturale o artificiale, ecc. . .) si richiede che il processo di riconoscimento attuato dal dispositivo si basi su algoritmi molto robusti aventi parametri automaticamente adattabili.

L'applicazione infine richiede una buona accuratezza (con incertezza sulla posizione inferiore a 40 mm) nell'individuare e nel tracciare la posizione nello spazio di alcuni segmenti anatomici del corpo umano ed in particolare dell'arto superiore sia che il paziente esegua l'esercizio in posizione seduta sia che lo esegua in posizione eretta. Il riconoscimento deve poi poter avvenire con qualsiasi tipo di abbigliamento (materiale e colore) e senza la necessità di indossare particolari sensori o dispositivi, così da rendere l'utilizzo del dispositivo il più semplice ed universale possibile. Come ultima richiesta, dal punto di vista economico, è bene che tutto ciò abbia un costo limitato e sostenibile da un'utenza privata, così da poter essere diffuso e commercializzato su larga scala.

Sulla base delle specifiche richieste, sopra delineate, viene condotta una ricerca volta a stilare un elenco, descritto nel successivo capitolo, delle tecnologie attualmente disponibili, dal quale estrarre la migliore soluzione.

Capitolo 2

Metodiche di scansione 3D: stato dell'arte

Il seguente capitolo propone un quadro generale dello stato dell'arte relativo ai sistemi di ricostruzione tridimensionale, indicando inizialmente una possibile tassonomia ed analizzando ogni singola tecnologia al fine di mostrarne sia il principio di funzionamento su cui si basa sia le principali caratteristiche e limitazioni, così da poter giustificare le considerazioni che hanno portato alle scelte successivamente effettuate. La presente trattazione infatti ha come finalità quella di individuare il dispositivo in grado di soddisfare al meglio le richieste precedentemente delineate, sia in termini di caratteristiche tecniche, sia di costo. Nella redazione è stata consultata la seguente bibliografia [12–14, 41, 22].

Le misure tridimensionali (3D), legate all'acquisizione automatica della forma di un oggetto, si avvalgono di un'ampia serie di tecnologie ottimizzate in relazione al diverso campo di impiego. Una possibile tassonomia [8] è riportata nella figura 2.1.

La seguente trattazione ed analisi descrive brevemente le caratteristiche principali di ciascuna di esse, presentando un quadro completo dello stato dell'arte attuale dal quale partire al fine di derivare la scelta della tecnica appropriata al caso concreto in esame.

Ad un primo livello di distinzione ed analisi, le tecniche di acquisizione di forme 3D si suddividono in sistemi con contatto e senza contatto.

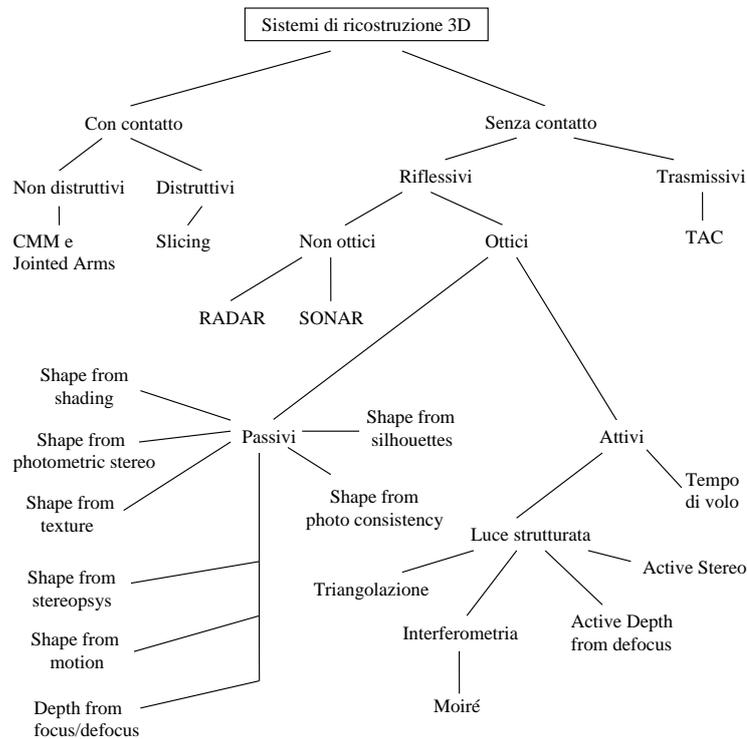


Figura 2.1. Sistemi di ricostruzione 3D.

2.1 Sistemi di acquisizione con contatto

I metodi di acquisizione con contatto prevedono un'interazione fisica tra il sistema di misura ed il misurando e possono essere di tipo non distruttivo (macchine di misura cartesiane -CMM- e jointed arms) o distruttivo (slicing).

Macchine di misura cartesiane o a giunti articolati (CMM o jointed arms): le misure vengono effettuate mediante l'utilizzo di un tastatore, il quale viene posto manualmente, o in modo automatizzato, a contatto con la superficie da analizzare. Per ogni punto vengono registrate posizione ed orientamento dell'end effector, rispetto ad un sistema di riferimento solidale all'oggetto in esame.

Slicing: il misurando viene suddiviso in sezioni parallele e le misure sulle parti ottenute sono effettuate mediante telecamere.

È evidente come nel caso in esame si debba scartare qualsiasi sistema eccessivamente invasivo o addirittura distruttivo.

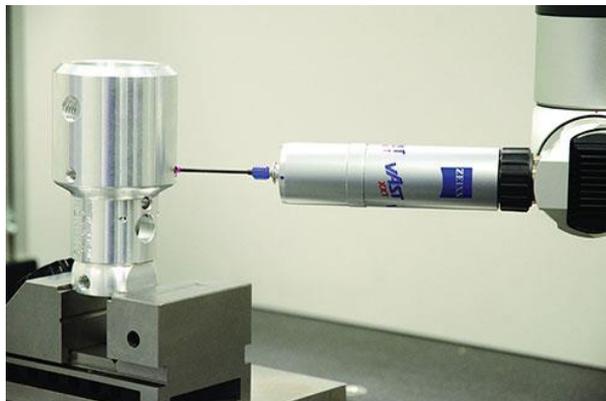


Figura 2.2. Particolare di un tastatore per misura con contatto.

Alla luce delle presenti considerazioni consegue la necessità di utilizzare sistemi di acquisizione senza contatto.

2.2 Sistemi di acquisizione senza contatto

I metodi di acquisizione senza contatto si basano sulla misura dell'interazione di onde elettromagnetiche o sonore con il misurando e possono essere suddivisi in due categorie: trasmissivi e riflessivi.

2.2.1 Sistemi trasmissivi

I metodi trasmissivi sfruttano la proprietà dei raggi X di attraversare un corpo ed essere parzialmente o totalmente assorbiti in funzione delle caratteristiche del materiale che incontrano lungo il percorso tra emettitore e ricevitore. La procedura viene ripetuta punto per punto, a breve distanza, così da ottenere una mappa completa che, a seguito di un'opportuna analisi, è convertita in immagine 3D. Un esempio di questa tecnica, applicata in ambito medico, è la tomografia assiale computerizzata (TAC).

È ben evidente come questo approccio richieda molto tempo e renda impossibile l'acquisizione in real-time.

2.2.2 Sistemi riflessivi

I metodi riflessivi, invece, si fondano sulla capacità delle superfici di riflettere le onde elettromagnetiche e/o le onde sonore. Essi si suddividono in non ottici ed ottici.

2.2.2.1 Sistemi non ottici

I due principali sistemi riflessivi non ottici sono il RADAR (RADio Detection And Ranging) e il SONAR (SOund Navigation And Ranging). Entrambi misurano la distanza tra emettitore e bersaglio registrando il ritardo temporale tra l'invio e la ricezione di un impulso elettromagnetico nel primo caso o sonoro nel secondo.

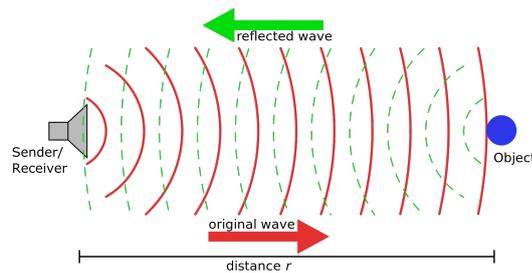


Figura 2.3. Schema rappresentante il principio di funzionamento di un sonar.

Questi metodi sono utilizzabili in settori (aeronautico e navale) in cui le grandi dimensioni dei campi di misura non richiedono elevate risoluzioni assolute.

È chiaro come nel caso in esame, tali metodi siano inutilizzabili.

2.2.2.2 Sistemi ottici

I metodi riflessivi ottici, spesso basati sull'impiego di telecamere, consentono un'analisi veloce ed esente dal contatto tra strumento di misura e misurando, applicabile in settori differenti in quanto a natura e vastità. Essi presentano il vantaggio di essere in generale più veloci ed economici dei sistemi non ottici anche se il campo di misura (relativo alla sola parte di superficie visibile), le proprietà dei materiali (brillantezza, colore, trasparenza, rugosità superficiale, ecc.) e le condizioni ambientali (prima fra tutte l'illuminazione) possono introdurre alcune limitazioni.

Ad un primo livello, queste tecniche si suddividono in passive ed attive. Nel primo caso l'illuminazione della scena, ripresa mediante telecamere, è unicamente affidata all'ambiente. Si hanno quindi sistemi economici e privi di alcun vincolo di applicabilità, affetti però da ridotta accuratezza. Al contrario le tecniche attive, sfruttando fonti di illuminazione speciali, in generale conducono a risultati aventi migliori accuratezze ma spesso comportano maggiori costi e presentano limiti di applicabilità.

Sistemi ottici passivi

Ricostruzione dal chiaroscuro (Shape from shading): la superficie 3D viene generata analizzando la variazione di luminosità di un'immagine bidimensionale (2D) dell'oggetto e ricavando la normale associata ad ogni punto. Questo metodo ha notevoli svantaggi allorché si operi in condizioni non ideali (superfici lambertiane, albedo uniforme ed illuminazione parallela).

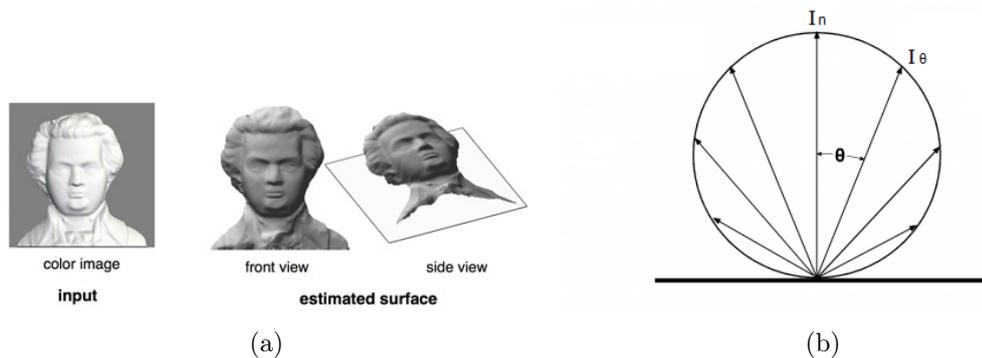


Figura 2.4. Ricostruzione 3D mediante shape from shading (a) e comportamento delle superfici lambertiane (b).

Le superfici lambertiane sono caratterizzate dalla loro completa opacità tale da diffondere la luce incidente senza alcuna direzione preferenziale.

Esse seguono la legge del coseno di Lambert:

$$I_\vartheta = I_n \cdot \cos(\vartheta) \quad (2.1)$$

dove ϑ è l'angolo presente tra il raggio (di intensità I_n) normale alla superficie ed un qualsiasi raggio (di intensità I_ϑ) diffuso da essa.

L'albedo di una superficie è invece definito come il rapporto tra il flusso luminoso riflesso e diffuso in tutte le direzioni dalla superficie e quello da essa ricevuto mediante un fascio di raggi paralleli.

Dovendo realizzare un'applicazione in grado di operare in condizioni ambientali reali ed ampiamente variabili, la tecnica dello shape from shading viene quindi ritenuta non applicabile.

Ricostruzione dallo stereo fotometrico (Shape from photometric stereo): questo metodo si basa sul medesimo principio concettuale dello shape from shading; esso utilizza tre differenti immagini acquisite variando, tra uno scatto ed il successivo, le condizioni di illuminazione, al fine di risolvere

alcune problematiche relative alla definizione univoca della normale propria di ogni punto dell'immagine.

Il metodo presenta però limitazioni simili a quelle della tecnica precedentemente illustrata e risulta quindi inutilizzabile per la ricostruzione tridimensionale necessaria all'applicazione in esame.

Ricostruzione dalla tessitura (Shape from texture): questo metodo è in grado di ricostruire la forma acquisita da una superficie su cui sono presenti pattern (texture) ripetuti in modo deterministico (es. un muro costituito da mattoni o una tappezzeria) oppure statistico (es. disposizione di sassi nella spiaggia o dei fili di erba in un prato). L'operazione è effettuata analizzando l'entità delle distorsioni e la velocità delle variazioni di opportune caratteristiche degli elementi della tessitura (chiamati texel).



Figura 2.5. Esempio di texture deterministica (a) e statistica (b).

Poichè nell'applicazione in esame gli oggetti da ricostruire non presentano necessariamente superfici texturizzate, questo metodo non è applicabile al presente caso di studio.

Visione stereoscopica o stereoscopia (Shape from stereopsis): questo metodo simula il sistema di visione umano ricavando le informazioni di profondità dall'analisi di immagini provenienti da due o più telecamere inquadranti la stessa scena da posizioni diverse.

La ricostruzione 3D viene suddivisa in due fasi consecutive: la prima, basata sul principio di somiglianza, è finalizzata al reperimento, nelle due (o più) immagini, di punti omologhi (cioè che si riferiscono al medesimo punto fisico), la seconda, basata sui parametri di calibrazione di ogni singola camera (intrinseci), sulle loro posizioni reciproche (estrinseci) e su

opportuni algoritmi, permette di effettuare la ricostruzione vera e propria determinando, per ogni n-upla di punti omologhi, il passaggio da coordinate 2D in pixel a coordinate 3D in millimetri.

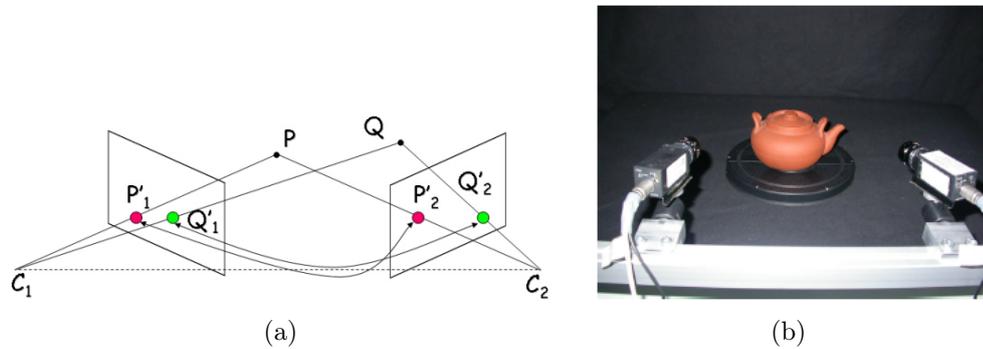


Figura 2.6. Schema base di un sistema di visione stereoscopica in cui è evidenziato l'accoppiamento di punti omologhi (a) ed esempio di dispositivo stereoscopico sperimentale (b).

È chiaro come la ricerca di punti corrispondenti sia ostacolata dal verificarsi di problemi di accoppiamento legati ad occlusioni (punti visibili in alcune immagini ma nascosti in altre a causa di oggetti presenti nella scena), distorsioni fotometriche (superfici non perfettamente lambertiane) e distorsioni proiettive (fenomeno della prospettiva associato a due differenti punti di vista relativi allo stesso oggetto).

In aggiunta alle problematiche sopra citate, la principale controindicazione del presente metodo, in riferimento all'applicazione in esame, è dovuta alla necessità di acquisire forme caratterizzate da superfici uniformi e di poter operare in condizioni altamente variabili.

Ciò determina una crescente complessità computazionale nell'estrazione di un numero elevato di punti omologhi e nella ricostruzione 3D in real-time, la quale rende questa tecnica inapplicabile al caso in esame.

Ricostruzione dal movimento (Shape from motion): questo metodo si basa sullo studio di immagini catturate con una singola telecamera (di cui sono noti i parametri intrinseci) in moto relativo rispetto all'oggetto da ricostruire.

Si ottiene quindi una serie di acquisizioni, riprese da punti di vista diversi, le quali sono paragonabili ad immagini scattate simultaneamente con un sistema di visione stereoscopico; i due metodi infatti sono concettualmente simili. È chiaro come l'ottenimento di buoni risultati in termini di ricostruzione 3D sia subordinato al fatto che non vi debba

essere alcuno spostamento relativo tra le componenti dell'oggetto, così da garantire la medesima posa nell'intero set di immagini. Esistono due diversi approcci al problema: i metodi discreti ed i metodi differenziali. I primi fondano il loro funzionamento sul reperimento dei punti omologhi (analogamente alla stereoscopia), mentre i secondi analizzano le velocità dei punti dell'immagine.

Nel caso in esame questo metodo è inutilizzabile poiché il soggetto da acquisire non si presenta come un unico corpo rigido.

Fochettatura (Depth from focus): la posizione di un punto nel sistema di riferimento della camera viene determinata dall'analisi di diverse immagini, scattate in sequenza, in cui l'oggetto in esame è acquisito con un livello di messa a fuoco via via migliore. Questa tecnica è sia costosa, richiede ottiche dotate di lenti di alta qualità, sia lenta, per ogni punto dell'immagine deve essere ricercata la posizione di fuoco ottimale (condizione non sempre banale da stabilire).

Le suddette considerazioni determinano l'incompatibilità del presente metodo con l'applicazione analizzata.

Sfocamento (Depth from defocus): la distanza di ciascun punto viene determinata basando l'analisi di varie immagini, aventi differenti livelli di messa a fuoco, sulla relazione che lega profondità, sfocamento e caratteristiche dell'ottica (focale, ecc...).

La necessità di dover acquisire numerose immagini, di calibrare la relazione sfocamento-profondità e di misurare in modo accurato lo sfocamento, in aggiunta all'utilizzo di ottiche professionali (al fine di ottenere risultati accurati), caratterizzate da elevate prestazioni, rende questo metodo sia lento sia costoso e quindi non adatto all'applicazione in esame.

Ricostruzione dalle sagome (Shape from silhouettes): la ricostruzione 3D generata da questo metodo si fonda sull'acquisizione di immagini binarizzate (silhouette), ciascuna delle quali è composta da pixel o bianchi o neri corrispondenti a raggi ottici rispettivamente intersecanti o non intersecanti la superficie in esame. A ciascuna immagine può essere associato un volume, a forma di cono generalizzato, all'interno del quale è contenuto l'oggetto dell'analisi. Diverse viste definiscono altrettanti coni dalla cui intersezione deriva la ricostruzione 3D (immagine 2.7).

Dalla precedente trattazione è chiaro come sia necessario che il soggetto acquisito non cambi forma durante l'intero processo di acquisizione.

Analizzando lo spazio tridimensionale, si comprende come questo metodo definisca un volume noto (tipicamente un cubo) al cui interno giace

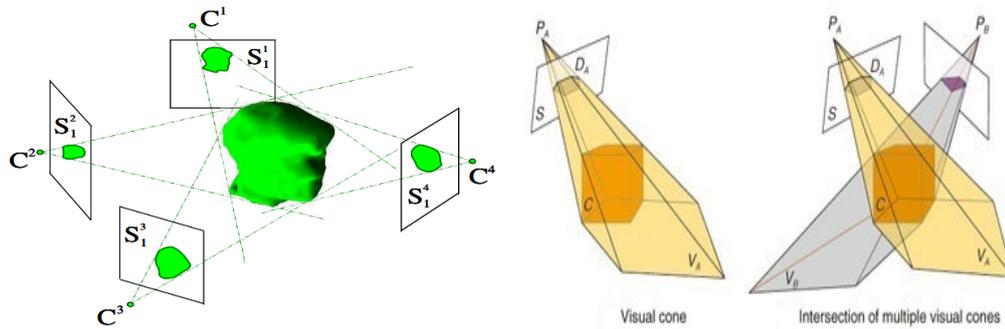


Figura 2.7. Principio di funzionamento dello shape from silhouettes.

l'elemento tridimensionale di interesse. Lo spazio considerato è suddiviso in mattoncini elementari (voxel) a ciascuno dei quali viene associato un valore booleano a seconda che sia o meno occupato dal volume in esame. Operativamente l'acquisizione può essere effettuata ruotando l'oggetto mentre quest'ultimo viene inquadrato da una sola telecamera oppure utilizzando un sistema di visione costituito da più telecamere.

I principali vantaggi di questo metodo consistono nel non dovere effettuare la complessa operazione di ricerca di punti omologhi, e nella possibilità di ottenere direttamente un modello tridimensionale della forma analizzata. Tuttavia lo shape from silhouettes è un metodo concettualmente non banale e, nonostante l'esistenza di diversi algoritmi ottimizzanti, computazionalmente oneroso. In aggiunta a ciò si evidenzia come esso non sia adatto all'acquisizione di parti tra di loro in movimento relativo.

Alla luce di questi fattori, ai quali si aggiunge l'impossibilità di ruotare l'oggetto e quindi la necessità di impiegare un numero elevato di telecamere, si sconsiglia l'utilizzo di tale metodo nella presente applicazione.

Ricostruzione da foto coerenza (Shape from photo consistency): è un secondo metodo di ricostruzione volumetrica il quale concettualmente si presenta come un'estensione del precedente. Esso infatti si basa sull'utilizzo di immagini a colori (o in scala di grigi) le quali aggiungono vincoli al problema migliorando il processo di ricostruzione.

Si ha quindi una tecnica da un lato più performante della precedente ma dall'altro computazionalmente più onerosa e quindi non ottimale ai fini della presente applicazione.

Sistemi ottici attivi

I sistemi ottici attivi, caratterizzati dall'utilizzo di una fonte di illuminazione di tipo artificiale, si suddividono in tecniche a tempo di volo (time of flight) e metodi basati su luce strutturata (sfocamento attivo, stereoscopia attiva, triangolazione laser, interferometria). Questa seconda categoria accomuna dispositivi in cui la fonte di illuminazione diventa parte integrante della logica di funzionamento del sistema stesso; ciascuno di essi, infatti, richiede una specifica tipologia di sorgente luminosa. Possono essere identificate due sottoclassi: la prima comprende metodi che migliorano analoghi sistemi passivi (active stereo, active depth from defocus) mentre la seconda è relativa a tecniche non aventi controparte passiva (triangolazione, interferometria).

Sfocamento attivo (Active depth from defocus): questa tecnica presenta il medesimo principio di funzionamento del corrispondente metodo passivo, precedentemente descritto, al quale viene aggiunta una fonte di illuminazione di tipo strutturato. Analogamente al caso passivo, al fine di ottenere un'accurata ricostruzione, sono richieste un numero elevato di immagini, ottiche professionali e una procedura di calibrazione finalizzata ad ottenere la relazione tra distanza e sfocamento.

Si conclude quindi affermando che questo metodo è inadatto alla presente applicazione poiché sia lento sia costoso.

Stereoscopia attiva (Active stereo): questa tecnica, come quella passiva precedentemente descritta, consente di ricostruire la terza dimensione analizzando punti corrispondenti ricavati nelle immagini acquisite mediante l'uso di due o più telecamere opportunamente calibrate. In aggiunta, un sistema di illuminazione a luce strutturata spesso infrarossa, facilita l'accoppiamento di punti omologhi, semplificando l'identificazione della posizione di opportuni marker catarifrangenti, situati in corrispondenza di punti notevoli appartenenti agli oggetti da misurare. Si possono così effettuare misure dinamiche delle loro coordinate 3D con risultati molto accurati ma limitate ad un numero ristretto di punti.

La presente metodica è spesso utilizzata per la ricostruzione del movimento del corpo umano, applicando blob sferici catarifrangenti in punti rilevanti, corrispondenti ai giunti delle articolazioni. Questa soluzione, considerata lo stato dell'arte attuale, richiede un'attenta preparazione del soggetto, il quale deve indossare un apposito abbigliamento dotato di marker.

La ricostruzione può essere estesa ad un numero decisamente superiore di punti proiettando sulla scena un pattern di luce strutturata costituito

da un insieme di pixel a cui è associato un colore, o una gradazione di scala di grigi, con distribuzione casuale (rumore bianco). In questo modo, basandosi sulle corrispondenze del pattern acquisito nelle due immagini, viene facilitato il reperimento di punti omologhi.

Queste tecniche possono presentare problemi in situazioni in cui la luce ambiente, altamente variabile, rende i marker catarifrangenti o il pattern difficilmente visibili dalle camere. Inoltre è richiesto un hardware costoso composto da due o più telecamere e da altrettanti illuminatori o proiettori.

Triangolazione laser: il sistema di acquisizione è composto da un proiettore in grado di generare un lama laser ed una telecamera, solidali alla medesima struttura ed opportunamente orientati così che i due componenti rispettivamente illuminino ed inquadrino la scena da due differenti posizioni. La misura è effettuata intersecando il raggio ottico della camera con la lama laser proiettata sull'oggetto ed effettuando calcoli trigonometrici di triangolazione. Il dispositivo necessita di una preliminare calibrazione al fine di rilevarne il modello geometrico ma presenta il vantaggio di non dover individuare punti omologhi e di ottenere la ricostruzione molto accurata di un intero oggetto.

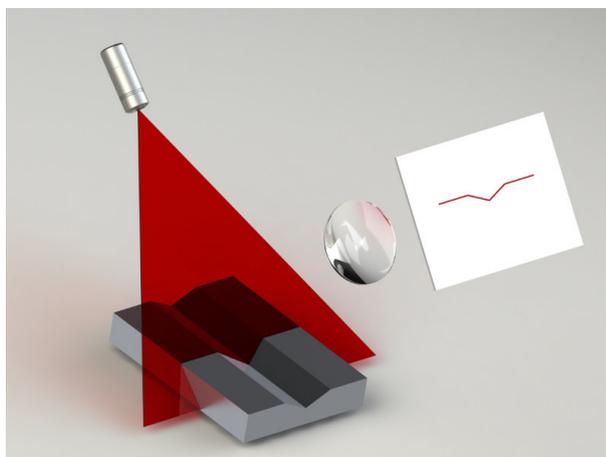


Figura 2.8. Principio di funzionamento della triangolazione attiva.

La necessità di acquisire un elevato numero di immagini da diversi punti di vista rende la ricostruzione molto lenta e determina l'impossibilità di effettuare misure dinamiche.

Al fine di ridurre i tempi di ricostruzione, così da poter effettuare acquisizioni in real-time di scene in movimento, viene sostituito l'emettitore di lama laser con un dispositivo in grado di proiettare sulla scena un pattern casuale di spot luminosi ottenuto dalla diffrazione, mediante opportuno

filtro, di un singolo raggio infrarosso (IR). Il reticolo si deforma assumendo curvature e dimensioni in relazione alla posizione degli oggetti presenti nella scena. L'acquisizione del pattern deformato, effettuata mediante camera sensibile solamente alla radiazione infrarossa, viene confrontata con un modello indeformato di riferimento acquisito in fase di calibrazione ad una distanza nota e memorizzato all'interno del sensore. A seguito di questa operazione, le coordinate 3D di ciascun punto vengono estratte mediante una procedura basata sulla triangolazione geometrica.

Questa tecnologia è implementata in un dispositivo commerciale di fascia consumer chiamato Kinect e presentato da Microsoft nel 2009. Il presente hardware, in grado di ricostruire e tracciare il movimento di un soggetto umano inquadrato, potrebbe essere una soluzione impiegabile nell'applicazione in esame.

Interferometria (Metodo delle frange di interferenza di Moirè): l'idea di fondo, basata su una teoria comune dalla quale sono state derivate diverse tecniche di scansione, è quella di proiettare sulla scena un pattern luminoso (barre di luce parallele ed equispaziate o anelli concentrici) e filtrare l'acquisizione attraverso il medesimo pattern leggermente ruotato, così da generare frange di interferenza dalla cui analisi è possibile effettuare la ricostruzione 3D. Il metodo è estremamente accurato ma computazio-

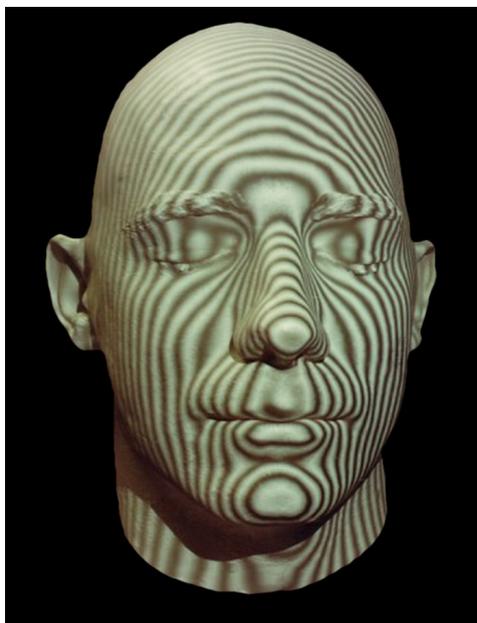


Figura 2.9. Interferometria: proiezione della griglia sul misurando.

nalmente oneroso e, in presenza di brusche discontinuità presenti nella superficie in esame, richiede numerose iterazioni successive con accuratezza crescente al fine di raffinare sempre più il modello ricostruito.

Anche questa tecnologia, come la precedente, limitata da tempi di elaborazione elevati non permette di effettuare una ricostruzione in real-time.

Tempo di volo (Time-of-flight): questa tecnologia permette di effettuare misure puntuali di distanza e si divide in due sottocategorie: tempo di volo impulsato e a onda continua (CW). La prima tipologia, di semplice concezione teorica, si basa sulla misura del tempo (t) trascorso tra l'invio di un impulso luminoso e la sua ricezione a seguito di una riflessione sulla scena. La distanza (d) tra dispositivo e bersaglio, pari alla metà di quella realmente percorsa dalla luce, è calcolabile attraverso la seguente relazione:

$$d = \frac{c \cdot t}{2} \quad (2.2)$$

in cui c è pari alla velocità della luce (nell'aria).

Per misure di distanza dell'ordine del metro (o inferiori) sono necessari contatori molto accurati e costosi.

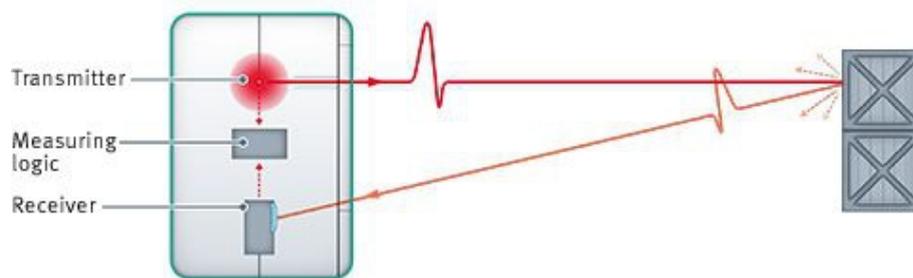


Figura 2.10. Principio di funzionamento di un dispositivo tof impulsato.

Nei sistemi CW, invece, il segnale emesso viene modulato sinusoidalmente (o a onda quadra) alla frequenza f e lo spazio tra emettitore e bersaglio è calcolato misurando lo sfasamento tra segnale emesso e ricevuto (φ) e convertendolo in distanza mediante la seguente formula:

$$d = \frac{c}{4\pi f} \varphi \quad (2.3)$$

La scansione di un'intera scena o di un oggetto 3D può essere effettuata per mezzo di misure ripetute, a distanze angolari note, così da acquisire un numero elevato di punti. Ciò è reso possibile, in strumenti di misura

chiamati laser scanner, conoscendo il modello geometrico del dispositivo ed utilizzando un sistema di movimentazione in grado di orientare opportunamente il raggio laser.

In particolare, all'interno dell'apparecchio è presente uno specchio posto in rotazione per mezzo di un motore elettrico, attorno ad un asse verticale, così da deflettere il raggio laser all'interno di un piano ed acquisire punti aventi il medesimo angolo di elevazione. Contemporaneamente, la rotazione dell'intero dispositivo attorno ad un asse orizzontale, permette di variarne l'elevazione ed ottenere così, grazie ai due movimenti combinati, un insieme di punti 3D, detto nuvola di punti. È chiaro come ciò si traduca in un processo molto lento attraverso il quale è impossibile effettuare misure dinamiche.

Questo problema può essere superato utilizzando illuminatori a luce IR diffusa e sensori a matrice (spesso telecamere IR) grazie ai quali si ha la possibilità di acquisire l'intera scena a frequenze intorno a 30 Hz.

Negli ultimi anni, la disponibilità sul mercato di dispositivi elettronici a basso costo ha permesso di realizzare telecamere low cost basate sulla tecnologia tof ad onda continua.

Uno di essi è rappresentato dalla Creative Senz3D [40]. Questo dispositivo, presentato al pubblico nel mese di giugno dell'anno 2013, viene posizionato, come una comune webcam, sulla parte superiore dello schermo ed è in grado di riconoscere e tracciare, all'interno di un volume di lavoro ampio un metro lungo l'asse ottico della camera, i movimenti delle dita delle mani compiuti dall'utilizzatore, il quale può in questo modo controllare il pc senza l'utilizzo del mouse.

Il range di misura limitato e l'impossibilità di tracciare l'intera figura umana non permettono l'utilizzo del presente dispositivo nell'applicazione in esame.

Un secondo esempio di telecamera a tempo di volo è costituito dal Kinect V2, dispositivo commercializzato da Microsoft a partire dall'anno 2014 il quale, presentando caratteristiche tecniche dichiarate dal costruttore, sia relativamente alla ricostruzione 3D sia al tracciamento del corpo umano, nettamente superiori rispetto alla precedente versione [7], sembra essere un'ottima soluzione hardware per l'applicazione in esame.

La scelta cade quindi su questo dispositivo, disponibile in anteprima presso il VBLab del Politecnico di Milano come technical preview.

Nei seguenti capitoli (3 e 4) saranno descritti i modelli matematici di pinhole camera e di distanziometro tof i quali, costituendone i componenti base, sono utili alla comprensione del funzionamento di una telecamera a tempo di volo.

Capitolo 3

Modello matematico di una telecamera

Il seguente capitolo ha come finalità quelle di descrivere matematicamente il modello di telecamera utilizzato in questo lavoro e di illustrare come avviene la mappatura in uno spazio bidimensionale di punti aventi coordinate tridimensionali. La trattazione propone inizialmente una modellizzazione semplice alla quale vengono aggiunte, passo dopo passo, le più comuni non idealità. I concetti delineati in seguito, uniti al modello di tof monodimensionale presentato successivamente nel capitolo 4, permetteranno di derivare un modello di tof camera in grado di convertire un'immagine di distanza in nuvola di punti tridimensionale.

3.1 Definizione

Con il termine telecamera viene definito uno strumento in grado di acquisire la realtà e di convertirla in immagine bidimensionale. I componenti principali di una telecamera digitale moderna sono la parte ottica (obiettivo) composta da una o più lenti mobili oppure fisse, il sensore (in alcuni casi CCD: Charge-Coupled Device, in altri CMOS: Complementary Metal-Oxide Semiconductor) sensibile ad una precisa porzione dello spettro luminoso, ed un circuito elettronico di controllo. I fotoni, provenienti da

un'opportuna fonte di illuminazione e riflessi dagli oggetti presenti nella scena, vengono catturati dall'obiettivo e focalizzati sul sensore, il quale, composto da una matrice di pixel, li converte in elettroni così da ottenere, grazie al circuito di controllo, un dato memorizzabile ed elaborabile digitalmente.

Un'immagine digitale, quindi, non è altro che una matrice ordinata contenente in ogni cella un valore numerico proporzionale alla quantità di fotoni acquisita dal singolo pixel nell'intervallo di tempo. Nella scala di grigi a 8 bit si rappresenta con il valore 0 un pixel nero, cioè che ha catturato pochi (idealmente zero) fotoni, mentre con il valore 255 (pari a $2^8 - 1$) un pixel bianco, il quale è stato colpito da un numero molto elevato di fotoni.

Ad un primo livello, le telecamere presenti sul mercato posso essere distinte in proiettive (con centro di proiezione situato in posizione finita) e telecamere affini (con centro di proiezione posto all'infinito) [14]. Quest'ultima tipologia, poiché utilizza ottiche aventi elevata lunghezza focale al fine di ottenere elevati ingrandimenti, non è di interesse per il presente lavoro il quale, al contrario, richiede ottiche corte e ampi campi di visione.

Tra i modelli geometrico-matematici presenti in letteratura, relativi a telecamere proiettive, il più comunemente utilizzato è chiamato pinhole camera model [11, 14].

Esso descrive la relazione di mappatura di punti 3D in immagini 2D e nella sua più semplice formulazione accomuna il funzionamento di una telecamera a quello di una camera oscura.

3.2 Camera oscura

Questo oggetto è costituito da un volume chiuso (spesso a forma di parallelepipedo) avente, al centro di una faccia, un piccolo foro, detto foro stenopeico (o pinhole), il quale svolge la funzione di obiettivo selezionando i raggi luminosi, riflessi in maniera diffusa dagli oggetti presenti in scena, e proiettando al proprio interno una copia bidimensionale specchiata della realtà (figura 3.1(b)). Al fine di ottenere un risultato il più nitido e focalizzato possibile è necessario ridurre le dimensioni del foro così da aumentare l'accuratezza nella selezione dei raggi luminosi; ciò per contro riduce la quantità di fotoni catturati e quindi la luminosità delle immagini ottenute. È buona norma avere un rapporto di 1:100 tra il diametro del foro (diaframma) e la distanza tra piano di proiezione e foro (detta focale).

Storicamente, il primo a descrivere il fenomeno che sta alla base di questo dispositivo fu Aristotele nel IV secolo a.C. Col passare dei secoli la

tecnologia fu migliorata sempre più da scienziati, astronomi e matematici come Leonardo da Vinci e Gerolamo Cardano. Nel XVIII secolo, pittori vedutisti come Bellotto e Canaletto utilizzarono camere oscure dotate di lenti per inquadrare paesaggi reali ed impostare quadri altamente prospettici [30].

Viene ora introdotto e descritto il modello matematico di telecamera pinhole.

3.3 Modello pinhole

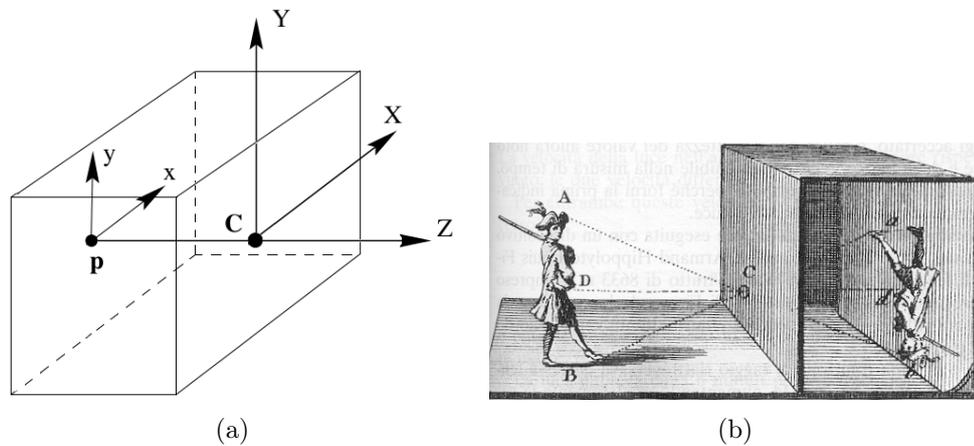


Figura 3.1. Geometria del modello di una camera oscura (a) e rappresentazione dell'effetto di ribaltamento dell'immagine (b).

In riferimento all'immagine 3.1(a) il punto corrispondente al foro stenopeico viene chiamato centro ottico ed indicato con la lettera C . In questo punto viene posizionato un sistema di riferimento XYZ definito da una terna destra cartesiana avente asse Z uscente dalla camera. Su quest'ultimo, ad una distanza f (detta focale) dal punto C , viene posto un secondo sistema di riferimento cartesiano (xy) , centrato nel punto p ed allineato con il precedente. Gli assi xy individuano un piano ortogonale a Z detto piano dell'immagine.

Nelle moderne telecamere a CCD o CMOS questo piano corrisponde alla superficie del sensore.

La retta passante per C e p è detta asse principale della camera.

La schematizzazione geometrica più utilizzata prevede che il punto p si trovi in corrispondenza di una coordinata positiva dell'asse Z ; in questo modo si ottiene il modello pinhole frontale rappresentato nell'immagine 3.2.

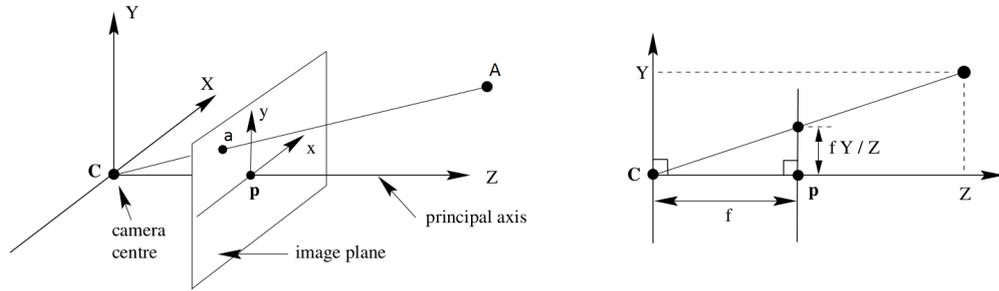


Figura 3.2. Geometria del modello lineare di camera pinhole frontale.

Esprimendo le coordinate (X, Y, Z) dei punti mondo mediante vettori colonna, è possibile, attraverso l'analisi del modello geometrico riportato in figura 3.2, derivare una mappatura da uno spazio euclideo in R^3 in uno spazio euclideo in R^2 .

$$(X, Y, Z)^T \rightarrow (x, y)^T \quad (3.1)$$

In coordinate omogenee, riferendosi ai sistemi di assi cartesiani di figura 3.2, la relazione tra realtà 3D (nel sistema di riferimento camera) ed immagine 2D (nel sistema di riferimento sensore centrato in p) può essere espressa con la seguente formulazione lineare

$$(\lambda x, \lambda y, \lambda)^T = (Xf, Yf, Z)^T \quad (3.2)$$

esprimibile in forma matriciale esplicita come:

$$\lambda \begin{pmatrix} x \\ y \\ 1 \end{pmatrix} = \begin{bmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix} \quad (3.3)$$

La matrice 3x4 che si ottiene è chiamata matrice di proiezione ed è indicata con la lettera \mathbf{P} . Il termine λ rappresenta un fattore di scala il cui valore è pari a Z .

Utilizzando una notazione compatta la 3.3 può essere riscritta come

$$\lambda \mathbf{x} = \mathbf{P} \mathbf{X}_{\text{cam}} \quad (3.4)$$

dove il vettore \mathbf{x} rappresenta le coordinate omogenee bidimensionali di un punto a (figura 3.2) giacente nel piano immagine mentre il vettore \mathbf{X}_{cam} rappresenta le coordinate omogenee 3D del corrispondente punto A situato nello spazio.

In molti dispositivi di visione, il sistema di riferimento associato al sensore ha l'origine posta nell'angolo in alto a sinistra di quest'ultimo e coordinate, orizzontale e verticale, rispettivamente chiamate u e v .

Il modello viene quindi leggermente modificato ottenendo:

$$\lambda \begin{pmatrix} u \\ v \\ 1 \end{pmatrix} = \begin{bmatrix} f & 0 & O_x & 0 \\ 0 & f & O_y & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix} \quad (3.5)$$

dove O_x e O_y sono le coordinate del punto p nel nuovo sistema di riferimento centrato nell'angolo del sensore.

Il modello fin qui descritto è molto semplificato e non abbastanza accurato per la maggior parte delle applicazioni in cui è richiesto tenere in considerazione le non idealità della telecamera.

In primo luogo si nota come spesso il baricentro del sensore non coincida con il punto principale (intersezione tra asse principale della camera e piano dell'immagine). Vengono quindi aggiunti i termini p_x e p_y rappresentanti le coordinate del punto principale nel vecchio sistema di riferimento posizionato nel centro del sensore.

L'espressione del modello diventa:

$$\lambda \begin{pmatrix} u \\ v \\ 1 \end{pmatrix} = \begin{bmatrix} f & 0 & O_x + p_x & 0 \\ 0 & f & O_y + p_y & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix} \quad (3.6)$$

Introducendo la matrice \mathbf{K} :

$$\mathbf{K} = \begin{bmatrix} f & 0 & O_x + p_x \\ 0 & f & O_y + p_y \\ 0 & 0 & 1 \end{bmatrix} \quad (3.7)$$

detta matrice di calibrazione della telecamera, la 3.6 può essere scritta in forma compatta come:

$$\lambda \bar{\mathbf{x}} = \mathbf{K}[\mathbf{I}|\mathbf{0}]\mathbf{X}_{\text{cam}} \quad (3.8)$$

dove il vettore $\bar{\mathbf{x}}$ contiene le coordinate u e v del punto a .

Inoltre, nel caso in cui non si vogliano esprimere le coordinate dei punti appartenenti alla realtà nel sistema di riferimento camera ma bensì in coordinate mondo, è necessario introdurre una rototraslazione presente tra i due sistemi di riferimento, ottenendo:

$$\lambda \bar{\mathbf{x}} = \mathbf{KR}[\mathbf{I} - \tilde{\mathbf{C}}]\mathbf{X} \quad (3.9)$$

dove \mathbf{R} è la matrice di rotazione tra i due sistemi di riferimento e $\tilde{\mathbf{C}}$ è il vettore che indica la posizione del centro della telecamera in coordinate mondo. Il vettore \mathbf{X} contiene le coordinate del punto A espresse nel sistema di riferimento mondo.

La matrice \mathbf{K} è composta da parametri che dipendono unicamente dalla telecamera, chiamati quindi intrinseci, mentre i valori di \mathbf{R} e $\tilde{\mathbf{C}}$ dipendono dalla posizione e dall'orientamento assunti dalla telecamera rispetto al mondo esterno; per tale ragione i parametri che li compongono sono chiamati estrinseci.

Poiché la superficie del sensore è fisicamente composta da una matrice di pixel, è prassi utilizzare quest'ultimo come unità di misura base della dimensione di un'immagine. Si avranno quindi punti mondo aventi coordinate 3D espresse in metri (nel sistema di assi XYZ) e punti immagine aventi coordinate 2D espresse in pixel (nel sistema di riferimento uv).

Al fine di mantenere la coerenza dimensionale all'interno del modello presentato, è bene sottolineare la necessità di esprimere in pixel le componenti della matrice \mathbf{K} . In particolare vengono introdotti i termini $f_x = f \cdot m_x$ ed $f_y = f \cdot m_y$ in cui m_x ed m_y rappresentano il numero di pixel per unità di distanza.

La necessità di diversificare i termini di focale (lungo l'asse x e lungo l'asse y) deriva dal fatto che, sebbene in prima approssimazione i pixel possano essere considerati di forma quadrata, in realtà essi hanno lati di dimensioni differenti; inoltre, spesso, questi ultimi non sono tra di loro perpendicolari.

Viene quindi introdotto un ulteriore termine s detto skew e definito dalla seguente relazione:

$$s = \cos(\alpha) \cdot f_x \quad (3.10)$$

dove α è l'angolo formato dai lati del pixel.

In definitiva la matrice \mathbf{K} diventa:

$$\mathbf{K} = \begin{bmatrix} f_x & s & O_x + p_x \\ 0 & f_y & O_y + p_y \\ 0 & 0 & 1 \end{bmatrix} \quad (3.11)$$

Il modello pinhole di telecamera descritto fino ad ora presenta 11 gradi di libertà (GDL) e contiene parametri legati unicamente da relazioni di tipo lineare, che lo rendono estremamente vantaggioso, soprattutto dal punto di

vista computazionale. In molti casi, però, in cui sono presenti distorsioni ottiche considerevoli, il presente modello non è più sufficientemente accurato ed è quindi necessario introdurre nuovi termini.

Tali non idealità delle lenti sono individuabili come non linearità del sistema di visione e necessitano, per essere tenute in considerazione, di un modello ulteriormente complicato in cui un generico punto A nel mondo e la rispettiva proiezione a sull'immagine non sono collegabili attraverso una linea retta passante per il centro ottico.

Tale effetto è ben visibile in figura 3.3(a) in cui le linee verticali ed orizzontali, che dovrebbero essere rette e parallele tra loro, sono invece curvate in maniera tanto più accentuata quanto più si trovano in prossimità dei bordi dell'immagine.

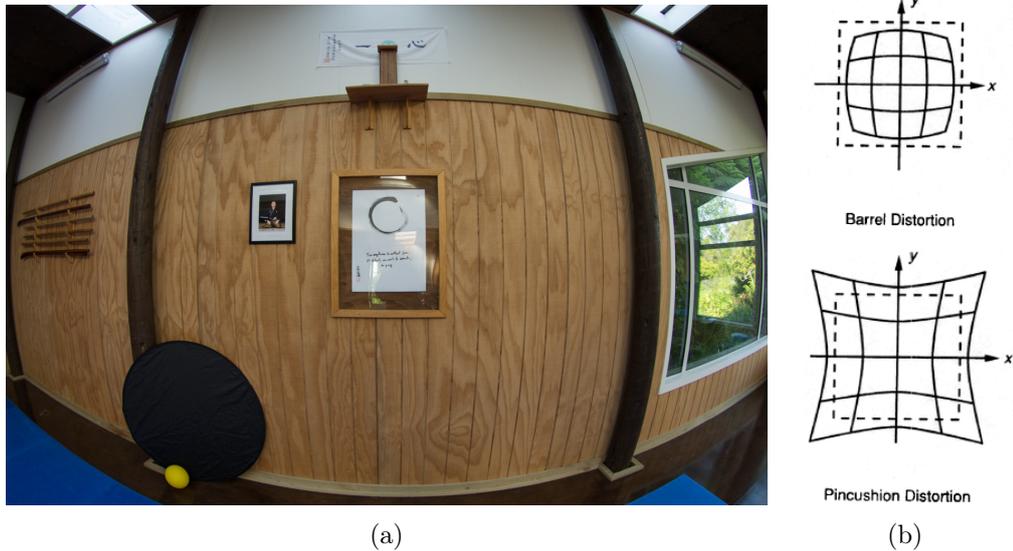


Figura 3.3. Immagine soggetta a distorsioni radiali a botte (a) e rappresentazione delle diverse tipologie di distorsioni radiali (b).

Tale distorsione viene chiamata radiale e può verificarsi, a seconda del segno (negativo o positivo), in due modalità: a botte oppure a puntaspilli (figura 3.3(b)). Questa tipologia di distorsione, solitamente, diventa più accentuata con il diminuire della distanza focale e può essere eliminata applicando la seguente correzione:

$$\begin{aligned} X_d &= L(r)X_n \\ &e \\ Y_d &= L(r)Y_n \end{aligned} \quad (3.12)$$

in cui $X_n = \frac{X}{Z}$ e $Y_n = \frac{Y}{Z}$ sono le coordinate normalizzate del punto A nel sistema di riferimento camera e X_d e Y_d sono le nuove coordinate normalizzate, le quali considerano l'effetto di distorsione introdotto dalle aberrazioni radiali.

Inoltre la formula 3.13 mostra come il termine $L(r)$ sia un polinomio sviluppato secondo Taylor mentre la 3.14 evidenzia come il raggio (r) sia pari alla distanza euclidea normalizzata misurata tra il punto considerato (avente coordinate X_n e Y_n) ed il centro del sistema di riferimento camera.

$$L(r) = 1 + k_1 r + k_2 r^2 + k_3 r^3 + \dots \quad (3.13)$$

$$r^2 = X_n^2 + Y_n^2 \quad (3.14)$$

Infine i coefficienti k_1, k_2, k_3 , ecc. . . sono ricavabili tramite opportune procedure di calibrazione.

Nei modelli di distorsione radiale più comuni, lo sviluppo in serie è arrestato al sesto ordine e vengono considerati solo i termini di ordine pari ($k_2 r^2, k_4 r^4, k_6 r^6$).

Un modello più accurato, in grado di applicare una migliore correzione delle aberrazioni ottiche, fu introdotto, col nome "Plumb Bob" (filo a piombo), da D. C. Brown nel 1966 [4].

Indicando con \mathbf{X}_n il vettore contenente le coordinate normalizzate nel sistema di riferimento camera, la correzione delle distorsioni diventa:

$$\mathbf{X}_d = (1 + k_2 r^2 + k_4 r^4 + k_6 r^6) \mathbf{X}_n + \mathbf{dx} \quad (3.15)$$

con

$$\mathbf{X}_d = \begin{pmatrix} X_d \\ Y_d \end{pmatrix} \quad e \quad \mathbf{X}_n = \begin{pmatrix} X_n \\ Y_n \end{pmatrix} \quad (3.16)$$

Lo spostamento \mathbf{dx} del centro delle aberrazioni ottiche radiali viene modellizzato attraverso una distorsione tangenziale dovuta ad un problema di allineamento della lente o ad un difetto costruttivo.

$$\mathbf{dx} = \begin{bmatrix} 2p_1 X_n Y_n + p_2 (r^2 + 2X_n^2) \\ p_1 (r^2 + 2Y_n^2) + 2p_2 X_n Y_n \end{bmatrix} \quad (3.17)$$

Inserendo le coordinate \mathbf{X}_d così ottenute al posto delle \mathbf{X} nella 3.9 si ottiene l'espressione completa esprimibile come:

$$\lambda \bar{\mathbf{x}} = \mathbf{KR}[\mathbf{I} - \tilde{\mathbf{C}}]\mathbf{X}_d \quad (3.18)$$

Si è infine giunti al modello completo per descrivere la telecamera, caratterizzato da 16 GDL, separabili in intrinseci, estrinseci e di distorsione [42]. Tali parametri, in seguito riassunti ed elencati, possono essere determinati attraverso un insieme di operazioni e procedure di calibrazione.

5 parametri intrinseci:

- f_x e f_y chiamati distanze focali ed espressi lungo x e lungo y ;
- x_c e y_c relativi alle coordinate del centro ottico;
- s chiamato skew e proporzionale all'angolo α presente tra i lati del pixel.

6 parametri estrinseci:

- r_1, r_2, r_3 utilizzati per esprimere la rotazione tra SDR camera e SDR mondo;
- t_1, t_2, t_3 coordinate del vettore traslazione del centro camera nel SDR mondo.

5 parametri di distorsione:

- k_2, k_4, k_6 relativi alle distorsioni radiali;
- p_1, p_2 relativi alle distorsioni tangenziali.

Non sempre però è necessario ricorrere all'utilizzo del modello completo; in alcuni casi, infatti, esso introduce una complessità tale da rivelarsi eccessiva per l'applicazione in esame. Spesso le distorsioni sono trascurabili ed in particolare è abbastanza frequente che non si considerino i coefficienti p_1 e p_2 relativi alle distorsioni tangenziali ed il coefficiente k_6 di distorsione radiale [11].

Capitolo 4

Modello matematico di un dispositivo a tempo di volo

Questa sezione espande la descrizione riportata nel capitolo 2 analizzando nel dettaglio il principio di funzionamento ed il modello matematico di un dispositivo time-of-flight ad onda continua (tof CW) con tecnologia basata sulla misura della variazione di fase tra segnale inviato e ricevuto. Viene successivamente presentata la possibilità di utilizzare come sensore una telecamera sensibile alla radiazione infrarossa, così da ottenere una mappatura tridimensionale della realtà. Sono inoltre descritti il principio di funzionamento e le caratteristiche tecniche di un dispositivo commerciale (Kinect V2), basato sulla presente tecnologia e potenzialmente in grado di soddisfare le richieste dell'applicazione presentata in questo lavoro. Infine sono descritti due software, appositamente realizzati per controllare il Kinect V2 e permettere l'acquisizione ed il salvataggio di dati necessari alla successiva fase di qualificazione metrologica.

4.1 Tempo di volo continuous wave (CW)

Una schematizzazione base di un sistema di misura lineare ottico a tempo di volo continuous wave (CW) è riportata in figura 4.1. Esso si compone di un emettitore di luce (spesso diodi led o laser) con lunghezza

d'onda contenuta nello spettro dell'infrarosso (IR), un sensore (fotodiodo) opportunamente filtrato e sensibile alla luce IR ed un circuito elettronico in grado sia di controllare il segnale inviato dall'emettitore sia di acquisire ed elaborare i dati provenienti dal ricevitore [22].

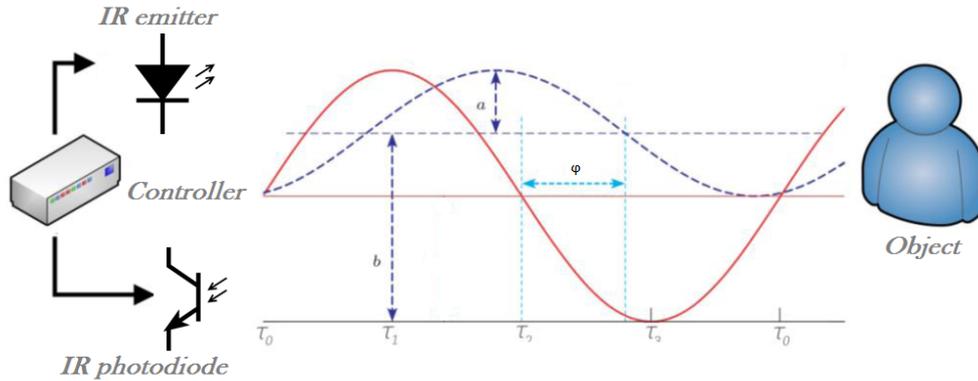


Figura 4.1. Schematizzazione e principio di funzionamento di un sistema di misura lineare ottico tempo di volo CW.

Verrà di seguito descritto il modello matematico che sta alla base del tof CW riferendosi alla situazione rappresentata in figura 4.1, in cui un raggio luminoso, modulato sinusoidalmente alla frequenza f_{mod} , viene catturato da un fotorivelatore a seguito di una riflessione da parte di un bersaglio.

Analizzando nell'immagine la sovrapposizione tra segnale emesso (rosso) e ricevuto (blu), si nota come quest'ultimo, dal confronto tra i due, presenti una minore ampiezza a dovuta principalmente alla diffusività della superficie riflettente, la quale disperde parte delle luce ricevuta, ed un offset costante b causato dal contributo delle componenti di luce ambiente non bloccate dal filtro IR montato sull'ottica della camera. Infine è chiaramente visibile uno sfasamento φ proporzionale alla distanza presente tra emettitore e target.

Al fine di misurare il ritardo in fase associato al segnale riflesso, una tecnica comunemente utilizzata prevede di acquisire 4 valori di ampiezza ($Q1, Q2, Q3, Q4$) relativi al segnale ricevuto, corrispondenti ad istanti temporali ($\tau_0, \tau_1, \tau_2, \tau_3$) sfasati di 90° l'uno dall'altro e sincronizzati col periodo del segnale inviato.

Il valore di fase φ viene ricavato applicando la seguente formula:

$$\varphi = \text{atan}\left(\frac{Q2 - Q4}{Q1 - Q3}\right) \quad (4.1)$$

Analizzando la relazione 4.1 si nota come un eventuale offset, sovrapposto al segnale ricevuto, venga eliminato dalle differenze $Q2 - Q4$ e $Q1 - Q3$ inoltre il rapporto $\frac{Q2-Q4}{Q1-Q3}$ ha lo scopo di normalizzare i valori e di eliminare effetti dipendenti da variazioni di intensità luminosa, dovute a differenti riflettività dei bersagli. Queste caratteristiche rendono il modello CW a variazione di fase molto robusto.

Inoltre, combinando opportunamente i valori $Q1, Q2, Q3, Q4$ è possibile ottenere le quantità:

$$a = \frac{\sqrt{(Q1 - Q3)^2 + (Q2 - Q4)^2}}{2}$$

e

$$b = \frac{Q1 + Q2 + Q3 + Q4}{4}$$
(4.2)

rispettivamente pari all'ampiezza della luce riflessa ed all'offset sovrapposto al segnale ricevuto precedentemente trattati.

Poiché la frequenza f_{mod} è dell'ordine di 10^6 Hz [10] mentre la frequenza f_{acq} , con cui vengono restituiti i valori di distanza, è dell'ordine delle decine di Hertz (tipicamente 30-100 Hz), il sistema è in grado di effettuare svariate misure ripetute del valore di φ e restituirne il valore medio così da migliorare notevolmente l'accuratezza del sistema.

Infine la distanza d presente tra sensore e bersaglio viene ricavata convertendo il valore di fase φ attraverso la seguente formulazione:

$$d = \frac{c}{4\pi f_{mod}} \varphi$$
(4.3)

in cui c è la velocità della luce nell'aria.

Un'attenta analisi della relazione 4.3 rivela l'esistenza di una massima distanza misurabile, detta distanza ambigua (d_{amb}) [22] il cui valore è assunto dalla variabile d in corrispondenza di uno sfasamento pari a 2π radianti.

$$d_{amb} = \frac{c}{2f_{mod}}$$
(4.4)

Questa limitazione della tecnologia tof CW a singola frequenza è chiaramente estendibile solo a patto di ridurre il valore di f_{mod} e di conseguenza di peggiorare l'accuratezza delle misure.

Sistemi di nuova generazione permettono di scavalcare il precedente compromesso basandosi su tecniche a frequenza multipla le quali fondano la propria teoria di funzionamento sull'utilizzo di due o più frequenze

con le quali viene modulato il segnale luminoso emesso. Ciascuna misura, effettuata mediante l'utilizzo di una diversa frequenza modulante, restituisce un set di distanze tra di loro multiple della quota associata ad uno sfasamento pari a 2π , alle quali il sistema attribuisce la stima della posizione dell'oggetto misurato.

Nell'ipotesi semplice di utilizzo di doppia modulazione (figura 4.2), si otterranno valori di distanza d_1 e d_2 rispettivamente attribuibili a misure effettuate con frequenza f_1 ed f_2 e matematicamente esprimibili come:

$$d_1 = \frac{c}{4\pi f_1}(\varphi_1 + h2\pi)$$

e

$$d_2 = \frac{c}{4\pi f_2}(\varphi_2 + k2\pi)$$
(4.5)

con h e k interi e positivi.

Il sistema, tramite un opportuno algoritmo, determina la reale distanza d tale per cui un ben preciso valore di h ed uno di k fanno sì che d_1 sia uguale a d_2 .

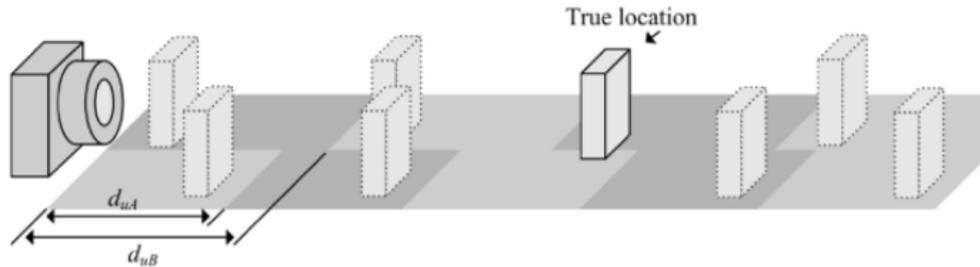


Figura 4.2. Principio di estensione della distanza di misura utilizzando una tecnica a frequenza multipla.

La tecnologia tof fin qui descritta si basa su una sola coppia emettitore-ricevitore ed è quindi limitata a singole misure puntiformi di distanza. La ricostruzione di nuvole di punti 3D è ottenibile, come descritto nella sezione riguardante la tecnologia a tempo di volo, presente nel capitolo 2, deflettendo il raggio luminoso mediante specchi rotanti e conoscendo il modello geometrico del dispositivo. Ciò richiede però tempi di acquisizione dell'ordine del secondo e quindi non permette di effettuare acquisizioni in real-time.

La presente problematica può essere superata sostituendo la singola coppia formata da laser IR e fotorivelatore con un emettitore a luce diffusa

ed un sensore matriciale (telecamera). I dispositivi così composti, descritti nel seguente capitolo, vengono chiamati telecamere a tempo di volo.

4.2 Telecamera a tempo di volo

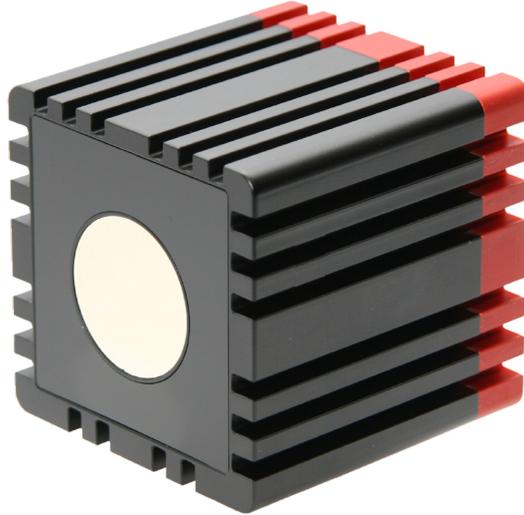


Figura 4.3. Esempio di telecamera a tempo di volo commerciale.

Le telecamere a tempo di volo integrano in un unico oggetto le caratteristiche tecniche proprie sia di un distanziometro a tempo di volo, sia di una telecamera tradizionale. Questi dispositivi, dal momento che presentano al proprio interno una telecamera sensibile all'infrarosso, possiedono un sensore composto da una matrice rettangolare di fotorivelatori, a ciascuno dei quali può essere applicata la trattazione matematica riportata nel paragrafo 4.1.

È quindi possibile ottenere, con una singola acquisizione, una mappa di distanze completa, relativa all'intero sensore. Un ulteriore vantaggio, legato all'utilizzo di una camera IR, è quello di poter raggiungere frame rate caratteristici della visione artificiale con telecamera (tipicamente aventi valori compresi tra 30 e 100 Hz) e quindi di effettuare acquisizioni in real-time. Conoscendo per ogni punto acquisito la distanza lungo Z , nel sistema di riferimento (SDR) camera, e le relative coordinate (u e v) in pixel nella matrice del sensore, è possibile, a valle di un opportuno processo di calibrazione, atto ad estrarre i parametri intrinseci, applicare un modello di telecamera (ad esempio il modello pinhole presentato nel paragrafo 3.3) e ricavare le relative coordinate X ed Y .

A valle di questo procedimento sono note le posizioni 3D dei punti acquisiti; è quindi possibile ricostruire la nuvola di punti dell'intera scena osservata.

Fino a pochi anni fa le uniche telecamere a tempo di volo disponibili sul mercato, a causa del loro elevato costo, erano indirizzate ad un bacino di utenza ristretto ed il loro utilizzo era limitato all'ambito industriale e della ricerca. Inoltre esse disponevano di ridotti angoli di visione e di scarsa risoluzione.

Ad esempio la Mesa Imaging SwissRanger 4000 (SR4000), riportata nell'immagine 4.3, è probabilmente la più conosciuta telecamera a tempo di volo. Essa presenta un range di misura ampio 5 m, un campo di vista di $69 \times 55^\circ$ ed una risoluzione di 176×144 pixel. È in grado di acquisire fino a 30 frame al secondo (fps), e viene venduta a circa 9000 \$ [43].

Un altro modello di telecamera a tempo di volo industriale è la PMD Technologies CamCube 2.0. Essa presenta un range di lavoro pari a 7 m ed una risoluzione di 204×204 pixel con un campo di vista di $40 \times 40^\circ$. È in grado di acquisire alla frequenza di 25 fps ed è venduta a circa 12000 \$ [47].

A partire dagli ultimi mesi dell'anno 2013, Microsoft ha realizzato e commercializzato un dispositivo chiamato Kinect V2 basato su un sensore tof CW a basso costo (meno di 100 \$) sviluppato da 3DV Systems (e Canesta). Le caratteristiche di questo sistema di misura ed acquisizione verranno descritte ed analizzate nelle prossime sezioni, al fine di derivare una caratterizzazione metrologica utile a stabilire se l'hardware in esame è in grado di soddisfare, sia in termini di costo che di accuratezza, le richieste della presente applicazione (capitolo 1).

4.2.1 Microsoft Kinect V2

Una telecamera a tempo di volo commerciale prodotta da Microsoft è chiamata Kinect V2 ed è presente sul mercato a partire dall'anno 2014 [41]. In figura 4.4 è riportata un'immagine pubblicitaria del dispositivo. Questo oggetto, venduto a circa 150 \$, nasce come accessorio della console Xbox One al fine di permettere il controllo di videogame attraverso i gesti ed il movimento del corpo, senza l'ausilio di alcun controller o altro dispositivo impugnato o indossato dal giocatore.

Nel dettaglio esso si basa su una telecamera tof CW a variazione di fase composta da un emettitore IR realizzato mediante un array di 3 laser dotati di apposito filtro diffusore e da una telecamera monocromatica sensibile alla radiazione infrarossa con risoluzione di 512×424 pixel e campo di vista

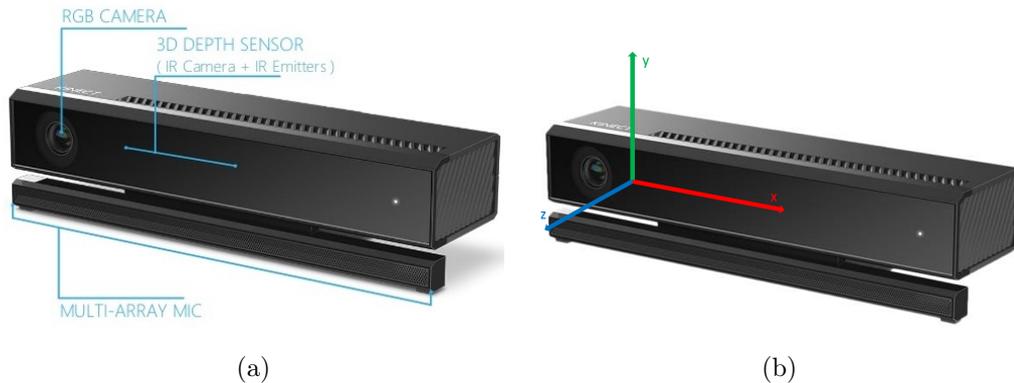


Figura 4.4. Microsoft Kinect V2. Componenti (a) e sistema di riferimento (b).

orizzontale ampio 70° e verticale ampio 60° . Il range di misura si estende, lungo l'asse Z uscente dalla camera, da 500 mm a 4500 mm (tabella 4.1).

Inoltre all'interno dell'oggetto sono presenti una telecamera RGB con risoluzione pari a 1080×1920 pixel ed un array di 4 microfoni in grado di interpretare comandi vocali (figura 4.4).

Tabella 4.1. Kinect V2 for Windows - principali caratteristiche [41].

Risoluzione camera IR [pix]	512 x 424
Risoluzione camera RGB [pix]	1080 x 1920
Frame rate massimo [Hz]	30
Campo visivo (HxV)	$70^\circ \times 60^\circ$
Range di misura [mm]	500-4500
Dimensioni [mm]	250 x 66 x 67
Peso [g]	966
Lunghezza cavo [m]	3
Collegamento	USB 3.0
Alimentazione	Adattatore per rete elettrica
Altro	4 microfoni

Poiché il dispositivo presenta un cavo di collegamento terminato con un connettore proprietario, al fine di fornire alimentazione elettrica e stabilire la connessione ad un pc, così da poterne acquisire i dati, viene venduto dal costruttore, a circa 50 \$, un apposito kit composto da un convertitore AC-DC e da un adattatore USB 3.0 il quale permette la trasformazione da Kinect V2 for Xbox a Kinect V2 for Windows.

Al fine di ottenere il corretto funzionamento della periferica sono richiesti i requisiti di sistema mostrati nella tabella 4.2.

Tabella 4.2. Kinect V2 - requisiti minimi di sistema [41].

Operating System	Windows 8, 8.1
CPU	Intel Core i7 3.1 GHz (or higher)
RAM	4 GB (or more)
GPU	DirectX 11 supported
USB	USB 3.0
Compiler	Visual Studio 2012, 2013
Language	C++, C#, VB.Net, HTML

In aggiunta è fornito un Software Developmet Kit (SDK) contenente numerose funzioni utili alla scrittura di software, basato su piattaforma Microsoft Windows, in grado di visualizzare in tempo reale ed acquisire immagini a colori, immagini IR, immagini di profondità, stream audio e posizione ed orientamento di 25 giunti del corpo umano. Questi ultimi sono estratti attraverso un'analisi della nuvola di punti, relativa all'intera scena, svolta mediante opportuni algoritmi proprietari in grado di riconoscere e tracciare fino a 6 corpi presenti nel volume inquadrato. Quest'ultima funzione, abbastanza insolita per una telecamera tof, è chiaramente l'obiettivo ultimo della casa costruttrice la quale, come precedentemente accennato, commercializza il presente dispositivo al fine di essere utilizzato come interfaccia di controllo tra giocatore e videogioco.

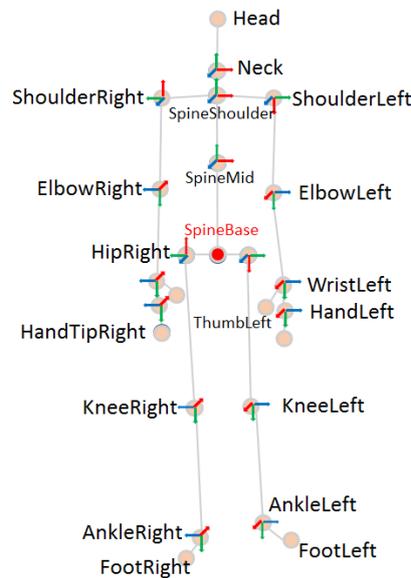


Figura 4.5. Nome e disposizione dei giunti individuati da Kinect V2.

Per questa ragione, nel caso in cui venga individuata nella nuvola di punti relativa alla scena inquadrata una porzione sufficientemente estesa ed approssimabile mediante un piano orizzontale (corrispondente, nella realtà, al pavimento della stanza), il dispositivo restituisce i coefficienti a, b, c, d relativi al corrispettivo piano di best fitting definito dall'equazione $y = ax + by + cz + d$ ed utilizzato come riferimento su cui basare l'identificazione della posizione relativa tra giocatore e pavimento della stanza.

Inoltre l'SDK, attraverso le funzioni chiamate “coordinate mapper”, permette di convertire l'immagine di profondità in nuvola di punti e di mappare su di essa il colore acquisito dalla telecamera RGB, ottenendo così una nuvola di punti (chiamata anche point cloud) colorata. La prima delle due conversioni (da immagine di depth a point cloud) è effettuata tramite l'applicazione di una look-up table (LUT) basata su un modello di camera pinhole con sole distorsioni radiali e contenente i coefficienti che, moltiplicati per il valore di distanza lungo l'asse Z corrispondente a ciascun pixel, permettono di ricavare, per ogni punto dell'immagine, le rispettive coordinate X e Y nel sistema di riferimento della camera IR mostrato in figura 4.4(b).

La presente LUT ed i parametri intrinseci da cui è stata derivata sono estraibili dal dispositivo. La tabella 4.3 riporta i valori di calibrazione relativi al dispositivo Kinect V2 utilizzato nel presente lavoro.

Tabella 4.3. Esempio di parametri intrinseci di calibrazione relativi alla telecamera IR.

<i>Parameter</i>	<i>Value</i>	<i>Symbol</i>
Focal Length X	366.8153 [pix]	f_x
Focal Length Y	366.8153 [pix]	f_y
Principal Point X	257.0098 [pix]	cc_1
Principal Point Y	203.6805 [pix]	cc_2
Radial Distortion Second Order	0.09059434	k_1
Radial Distortion Fourth Order	-0.2688895	k_2
Radial Distortion Sixth Order	0.09351667	k_3

Infine il dispositivo è in grado di ricostruire la forma dei volti inquadrati generando una mesh composta da oltre 2000 punti utile per creare avatar realistici aventi fisionomia molto simile a quella del giocatore. A partire da questi dati il sistema riconosce una serie di espressioni facciali tra cui l'apertura o la chiusura degli occhi o della bocca, lo sguardo rivolto alla camera, ecc. . .

La capacità di riconoscimento dei giunti del corpo umano in aggiunta sia alle caratteristiche tecniche di campo di vista e di risoluzione sia al basso costo fanno del Kinect V2 un'ottima ipotesi di soluzione hardware per l'applicazione in esame. Inoltre le ridotte dimensioni del dispositivo e la connessione plug-and-play ad un comune personal computer forniscono all'utente finale la possibilità di utilizzo dell'applicazione in ambiente domestico.

Verrà mostrata in seguito una serie di prove sperimentali, descritta nei capitoli 5 e 6, utile alla qualificazione metrologica del dispositivo. L'obiettivo della seguente trattazione è quello di stabilire se l'hardware in esame presenta caratteristiche adeguate all'applicazione ed in linea con le richieste di accuratezza, costo, ecc... delineate nel capitolo 1; effettuando misure in differenti condizioni di setup sarà possibile analizzare, a livelli di complessità via via crescenti, se le prestazioni di misura del dispositivo sono soddisfacenti.

4.2.1.1 Software di acquisizione

Al fine di acquisire i dati necessari alla qualificazione metrologica del dispositivo in esame sono stati appositamente realizzati due software, scritti in linguaggio C# e basati sull'SDK fornito da Microsoft. Il primo, la cui

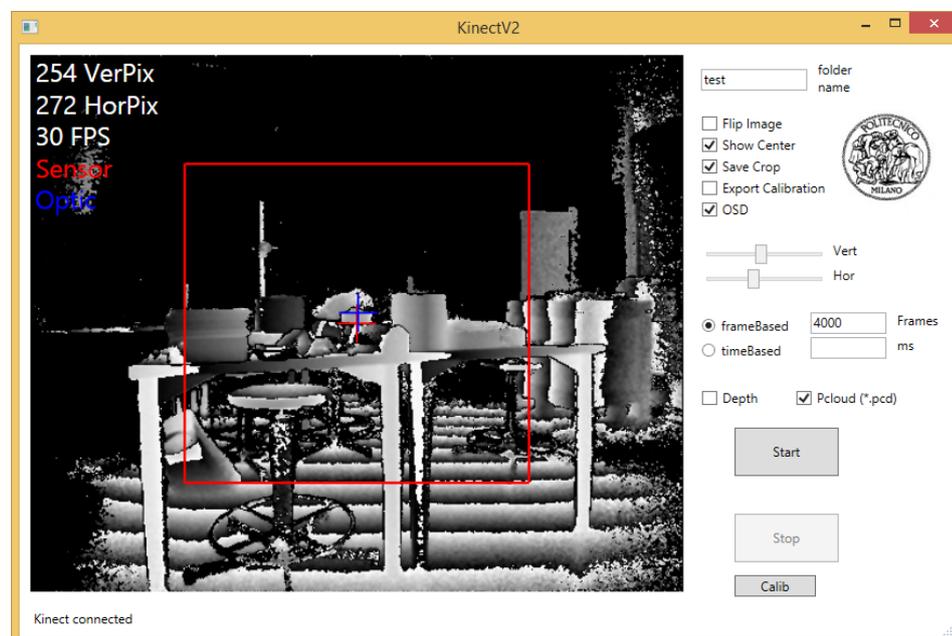


Figura 4.6. Interfaccia grafica del software di acquisizione sviluppato per la qualificazione metrologica del dispositivo.

interfaccia è mostrata in figura 4.6, è in grado di visualizzare in tempo reale la mappa di profondità acquisita, codificata mediante scala di grigi, e di salvare in modalità sequenziale un numero prestabilito di acquisizioni relative all'intero sensore o ad una sua porzione (Save Crop) rettangolare simmetrica rispetto al centro; possono essere scelti sia il formato depth (solo valore associato alla coordinata Z) sia il formato pointcloud. Quest'ultimo genera file aventi estensione .pcd normati secondo gli standard dettati dalle point cloud library (pcl) [46], ossia librerie open source composte da algoritmi per la gestione di nuvole di punti comunemente utilizzate nell'ambito della visione artificiale tridimensionale. Esse permettono di effettuare ricostruzione di superfici, registrazione di due o più nuvole, model fitting, segmentazione, ecc. . .

Inoltre il software è in grado di visualizzare a schermo la velocità di acquisizione (in fps), ribaltare l'immagine attorno ad un asse verticale (Flip Image), estrarre e salvare i parametri intrinseci relativi alla camera IR (Export Calibration) e mostrare a schermo una croce in corrispondenza del centro ottico ed una in corrispondenza del centro del sensore (Show Center).

Entrambi i formati scelti per il salvataggio (depth e point cloud) sono di tipo binario e contengono rispettivamente una sequenza di valori di distanza misurata lungo Z (in millimetri) o di triplette di coordinate XYZ (in metri) associati ai singoli pixel, ordinati per righe, a partire da quello corrispondente al vertice posto in alto a sinistra nell'immagine di preview.

Nel caso di file di depth, ogni valore binario è codificato utilizzando 16 bit e quindi l'intero set di distanze relativo ad un'acquisizione (217088 valori) richiede 434176 byte di memoria. Per le nuvole di punti, invece, si hanno 96 bit per ogni tripletta di coordinate e quindi un totale di 2605056 byte per acquisizione.

Inoltre all'inizio di ciascuna tipologia di file è presente un'intestazione (header) in formato ASCII contenente varie informazioni relative alle caratteristiche e alle dimensioni dell'acquisizione. In particolare, per quanto riguarda i file di tipo pcd, viene mostrato un esempio di header in figura 4.7 in cui, tra le varie informazioni, vengono riportati la larghezza (WIDTH) e l'altezza (HEIGHT) in pixel, il numero totale di punti (POINTS) e il tipo di codifica (DATA) utilizzata, la quale può essere ASCII o binaria. Quest'ultima è stata scelta per il salvataggio delle acquisizioni a causa del minor quantitativo di memoria richiesto a parità di dato memorizzato.

Il secondo software (figura 4.8), utilizzato nell'analisi della ricostruzione del corpo umano, permette la visualizzazione, in tempo reale, di un modello di corpo umano (body) composto da link (segmenti) e joints (punti)

```
1 # .PCD v.7 - Point Cloud Data file format
2 VERSION .7
3 FIELDS x y z
4 SIZE 4 4 4
5 TYPE F F F
6 COUNT 1 1 1
7 WIDTH 512
8 HEIGHT 424
9 POINTS 217088
10 DATA binary
```

Figura 4.7. Esempio di header relativo ad un file contenente le coordinate di una nuvola di punti salvata in formato pcd.

ricostruito mediante le coordinate spaziali dei punti (in seguito chiamati giunti) restituite dal Kinect V2 ed aggiornato alla frequenza di circa 30 Hz.

Inoltre, con il tasto Start, si ha la possibilità di salvare l'acquisizione in corso all'interno di un file di testo (*.txt) contenente una sequenza di blocchi ciascuno dei quali composto da 25 righe (una per ciascun giunto) riportanti, nell'ordine, un indice progressivo indicante il tipo di giunto, le coordinate di posizione spaziale (XYZ espresse in metri) nel sistema di riferimento camera, la rotazione nel SDR camera (quaternioni $XYZW$) ed un flag che indica se la posizione è tracciata (Tracked: giunto visibile dalla camera), stimata (Inferred: giunto occluso e quindi non visibile dalla camera) o non tracciata (Not Tracked: giunto occluso e posizione impossibile da stimare). Inoltre, all'inizio di ogni blocco, viene memorizzato il timestamp relativo all'istante in cui è avvenuto il salvataggio. Quest'ultimo riporta il tempo trascorso da quando il sensore è stato inizializzato ed è codificato come nel seguente esempio: #t 27.22:48:09.4912426 in cui i vari campi contengono, nell'ordine, i giorni (27), le ore (22), i minuti primi (48), i secondi (09) e i decimi di microsecondo (4912426).

Nella tabella 4.4 è riportato l'elenco dei giunti rilevati dal Kinect V2 ai quali è associato un nome e un indice; si noti come la progressione numerica si basi su una rappresentazione "ad albero" del corpo umano avente come radice il giunto SpineBase e come foglie i giunti relativi alla testa, alle mani ed ai piedi. Inoltre, per ciascuna articolazione, è indicato se essa viene modellizzata solo tramite una traslazione o anche tramite una rotazione.

In seguito alla fase di acquisizione, l'analisi dei dati restituiti da entrambi i software è stata svolta mediante opportuni script Matlab in grado di

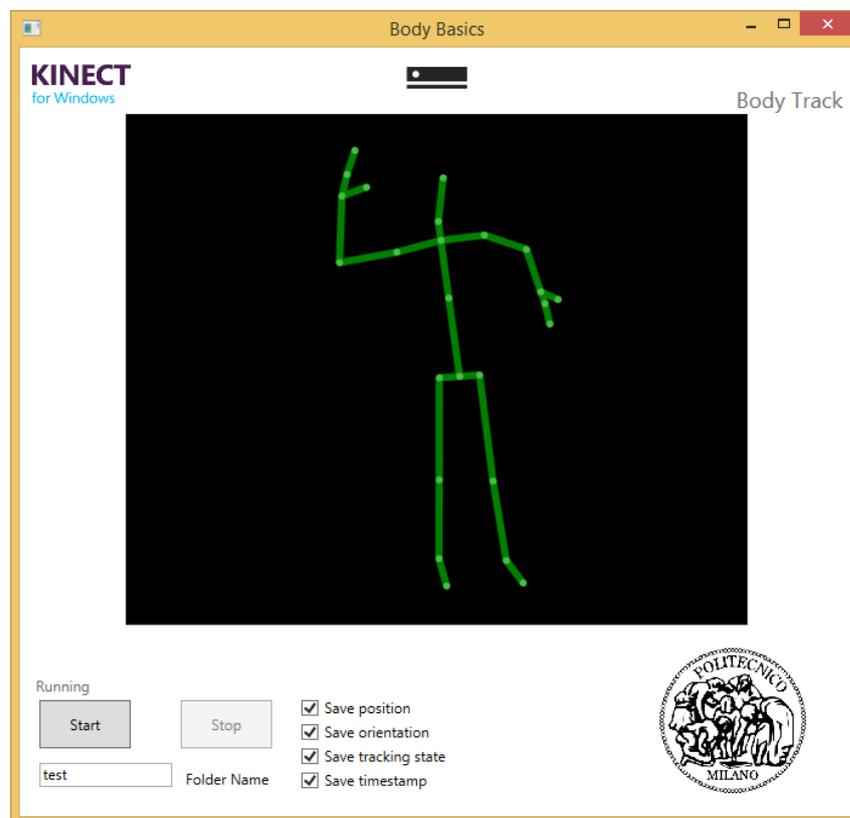


Figura 4.8. Interfaccia grafica del software di acquisizione appositamente sviluppato per la qualificazione metrologica del dispositivo.

aprire i file precedentemente salvati ed estrarne il contenuto.

In particolare, nel caso di dati provenienti dal primo software, si ha la possibilità di mostrare la mappa di distanze o la nuvola di punti in un visualizzatore 3D grazie al quale è possibile ruotarla, traslarla, ingrandirla, rimpicciolirla e, se necessario, selezionarne una porzione da ritagliare. Inoltre tali script eseguono calcoli finalizzati all'estrazione di parametri fondamentali per la qualificazione metrologica riportando i risultati ottenuti in opportuni grafici (capitolo 5).

Per quanto riguarda i file contenenti le traiettorie restituite dal secondo software, essi vengono aperti ed analizzati mediante apposite routine di elaborazione le quali, come ampiamente descritto in seguito nel capitolo 6, estraggono una cifra di merito in grado di definire il livello di accuratezza con cui viene tracciata la posizione dei segmenti del corpo umano da parte di Kinect V2, confrontando le traiettorie acquisite da quest'ultimo dispositivo con un riferimento generato da un sistema stereoscopico.

Tabella 4.4. Elenco riportante i gradi di libertà relativi ai giunti tracciati da Kinect V2 ai quali sono associati un nome ed un indice numerico.

<i>Joint Index</i>	<i>Joint Name</i>	<i>DOF</i>
01	SpineBase	Position-Orientation
02	SpineMid	Position-Orientation
03	Neck	Position-Orientation
04	Head	Position
05	ShoulderLeft	Position-Orientation
06	ElbowLeft	Position-Orientation
07	WristLeft	Position-Orientation
08	HandLeft	Position-Orientation
09	ShoulderRight	Position-Orientation
10	ElbowRight	Position-Orientation
11	WristRight	Position-Orientation
12	HandRight	Position-Orientation
13	HipLeft	Position-Orientation
14	KneeLeft	Position-Orientation
15	AnkleLeft	Position-Orientation
16	FootLeft	Position
17	HipRight	Position-Orientation
18	KneeRight	Position-Orientation
19	AnkleRight	Position-Orientation
20	FootRight	Position
21	SpineShoulder	Position-Orientation
22	HandTipLeft	Position
23	ThumbLeft	Position
24	HandTipRight	Position
25	ThumbRight	Position

Capitolo 5

Qualificazione metrologica

Il seguente capitolo descrive il processo di qualificazione metrologica ed analisi dello strumento in esame (Kinect V2) condotto al fine di identificare le performance ed i limiti del sistema di misura, così da poter valutare la possibilità di un impiego in ambito scientifico e nella fattispecie nell'applicazione precedentemente descritta. Basandosi sulla "Guide to the expression of Uncertainty in Measurement" [34] e riferendosi ad alcuni articoli presenti in letteratura [6, 18, 20, 26] vengono condotti diversi test in grado di coprire ed analizzare i differenti aspetti del dispositivo. La prima prova svolta riguarda la verifica della stabilità della misura di distanza effettuata a scena fissa, così da identificare un possibile transitorio derivante dall'influenza di fenomeni termici interni al dispositivo. Successivamente, nei diversi test, vengono valutati l'effetto casuale (deviazione standard) e l'effetto sistematico (scostamento del valor medio) sulla misura di distanza al variare della posizione del pixel nello spazio di visione, dell'angolo di incidenza della luce infrarossa e delle caratteristiche (materiale e colore) della superficie in esame. Infine viene testata la capacità del sensore nel ricostruire oggetti di forma nota (piano, cilindro, sfera) ed analizzati e caratterizzati gli effetti di mixed pixel e di riflessioni multiple.

5.1 Stabilità del sistema: influenza della temperatura

Una principale problematica relativa agli strumenti di misura elettronici è l'influenza della temperatura, la quale, surriscaldando i sensori ed i circuiti di condizionamento e campionamento del segnale, può causare, in situazioni statiche, drift dei valori forniti in uscita.

Dal momento che l'apparecchio in esame raggiunge, dopo alcune decine di minuti di funzionamento, elevate temperature, si decide di analizzare l'andamento di una serie di misure acquisite in un intervallo di tempo prolungato, al fine di identificare un possibile drift ed un conseguente periodo di preriscaldamento. Inoltre, poiché il dispositivo presenta al proprio interno una ventola attuata da un termostato, si vogliono mettere in luce le possibili differenze tra acquisizioni effettuate mentre quest'ultima è in rotazione oppure ferma.

La prova è condotta acquisendo 20000 campioni di distanza, alla frequenza di 30 Hz, ponendo il sensore a circa un metro da una parete planare di colore bianco. Si calcola la media mobile (su 500 campioni) al fine di filtrare il rumore di misura ed evidenziare l'andamento medio del valore restituito dal pixel centrale del sensore.

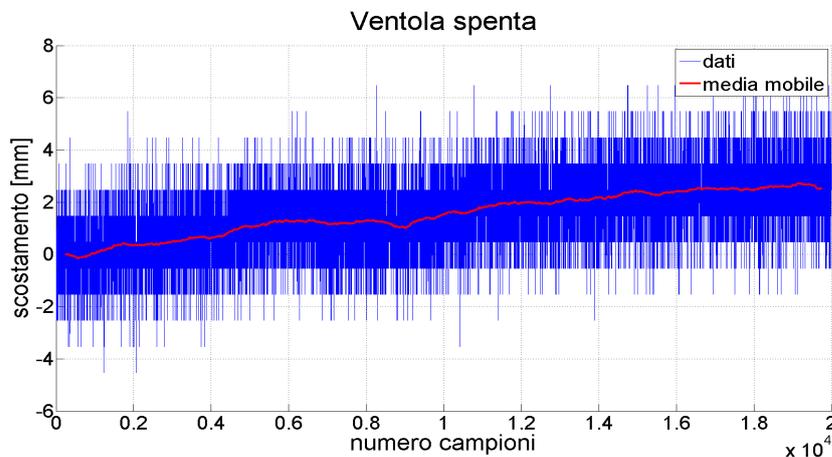


Figura 5.1. Andamento dei dati e del valore medio relativi ad acquisizioni effettuate a ventola spenta.

Durante l'intero test (circa 10 minuti) si è verificato che la ventola, presente all'interno della camera, non si è posta in rotazione. Ciò è dovuto al fatto che, nei primi minuti di acquisizione, la temperatura del sensore non raggiunge un valore tale da far scattare la soglia di attivazione del

termostato. Dall'analisi dei dati (figura 5.1) si nota una differenza di circa 2.5 mm tra il valore risultante dalla media dei primi valori acquisiti e quello relativo ai campioni finali.

Dopo aver verificato che il set-up non presentasse parti mobili, e che la camera si mantenesse alla medesima distanza dalla parete per tutta la durata dell'acquisizione, si è stabilito che questo drift dipende esclusivamente da un innalzamento della temperatura interna al sensore. Si è quindi cercato di mantenere quest'ultima il più possibile costante, applicando una ventola esternamente a quella inserita dal costruttore così da poterne replicare il comportamento e attuarne la rotazione indipendentemente dalla temperatura interna alla camera. L'immagine 5.2 mostra il posizionamento della ventola aggiuntiva la quale presenta le caratteristiche elencate nella tabella 5.1.

Tabella 5.1. Caratteristiche tecniche della ventola aggiunta al fine di stabilizzare la temperatura interna al dispositivo.

Tensione di alimentazione [V]	12
Corrente assorbita [A]	0.04
Velocità di rotazione [rpm]	5000
Volume di aria spostato [m^3/s]	0.0026



Figura 5.2. Kinect V2 prima e dopo l'installazione del sistema di raffreddamento esterno.

Nelle stesse condizioni della precedente prova vengono nuovamente acquisiti 20000 campioni, mantenendo la ventola esterna in continua rotazione. Dall'analisi dei dati (figura 5.3) si nota come, con questa configurazione, si possano garantire migliori prestazioni, limitando notevolmente la deriva

della misura al passare del tempo. Infatti, dopo un breve transitorio della durata di alcune decine di secondi, il valor medio si mantiene all'interno di una banda ampia 0.6 mm.

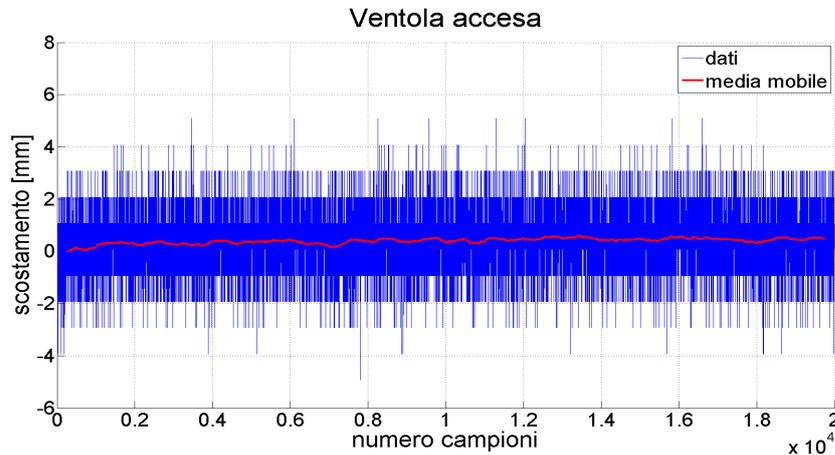


Figura 5.3. Andamento dei dati e del valore medio relativi ad acquisizioni effettuate a ventola accesa.

Con questo test preliminare è stata messa in luce l'influenza della temperatura sulla misura di distanza fornita dal dispositivo Kinect V2 notando come la ventola presente al proprio interno, ed il relativo controllo ad isteresi, non siano in grado di stabilizzare la temperatura del sensore. Il sistema di raffreddamento aggiunto, invece, generando un flusso d'aria continuo, garantisce un regime termico circa costante (mantenendo la temperatura al di sotto della soglia di attivazione della ventola interna) così da produrre valori di distanza caratterizzati da un andamento stabile.

Al fine di ottenere maggiori informazioni, e validare i risultati derivanti dalla precedente prova, si sono effettuate acquisizioni della durata di 24 ore ciascuna, in cui i campioni di distanza, relativi al pixel centrale, sono acquisiti alla frequenza di circa 1 Hz. La prima prova è stata eseguita senza l'utilizzo della ventola esterna. La figura 5.4(a) illustra i risultati delle misure. In blu sono riportati i dati grezzi, mentre in rosso è rappresentato l'andamento della media mobile effettuata su 500 campioni. Si nota come l'acquisizione sia caratterizzata da un transitorio iniziale, della durata di circa quattro ore, in cui la misura presenta un andamento oscillante, dovuto ad un funzionamento alternato della ventola interna al Kinect V2. Trascorso ciò, la temperatura del dispositivo raggiunge un valore tale da non consentire più lo spegnimento del sistema di raffreddamento nativo, il quale si mantiene in costante rotazione fino al termine della prova. Ciò permette lo stabilizzarsi delle misure ed avvalorare la tesi per cui l'utilizzo

di una ventola mantenuta in costante rotazione possa garantire una rapida stabilizzazione della misura. Durante l'acquisizione del secondo set di dati, la ventola esterna viene quindi mantenuta in funzione per tutta la durata del test. Ancora una volta, in figura 5.4(b), sono rappresentati in blu i dati acquisiti e in rosso la media mobile calcolata su 500 campioni.

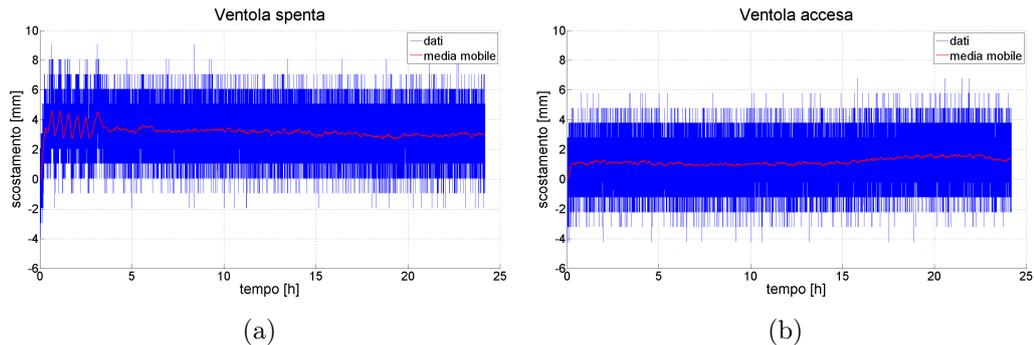


Figura 5.4. Campioni di distanza acquisiti per 24 ore (1 Hz) senza ventola esterna (a) e con ventola esterna (b).

La prova mette in risalto un andamento costante e stabile delle misure a seguito di un transitorio iniziale, di tipo esponenziale, della durata di alcuni minuti. A valle di queste considerazioni, in tutte le prove di seguito mostrate, verrà utilizzata una ventola fissata sulla parte posteriore del Kinect V2 e mantenuta in costante rotazione, così da raggiungere in pochi minuti il regime termico e stabilizzare la misura per l'intera durata dell'acquisizione.

5.2 Caratterizzazione del singolo pixel

Ad un primo livello di analisi, si considera l'aspetto distanziometrico del Kinect V2 effettuando misure di distanza lungo l'asse Z al fine di valutare e caratterizzare l'andamento della misura restituita dal singolo pixel e le principali fonti che ne influenzano l'incertezza.

5.2.1 Effetto casuale relativo alla misura di un singolo pixel nello spazio

In questa prova viene valutato l'andamento dell'effetto casuale all'aumentare della distanza misurata. Il setup utilizzato è composto da un

braccio robotico (ABB IRB 1600), in grado di compiere spostamenti caratterizzati da una ripetibilità pari a 0.02 mm [38], sulla flangia del quale è stato avvitato un piano in alluminio di colore bianco su cui è riportato un pattern regolare utile alla successiva fase di calibrazione. Quest'ultimo è realizzato e fissato in modo tale da far corrispondere, a meno di una traslazione lungo l'asse Z uscente dal tool, il proprio baricentro con il centro della terna relativa alla flangia del robot.

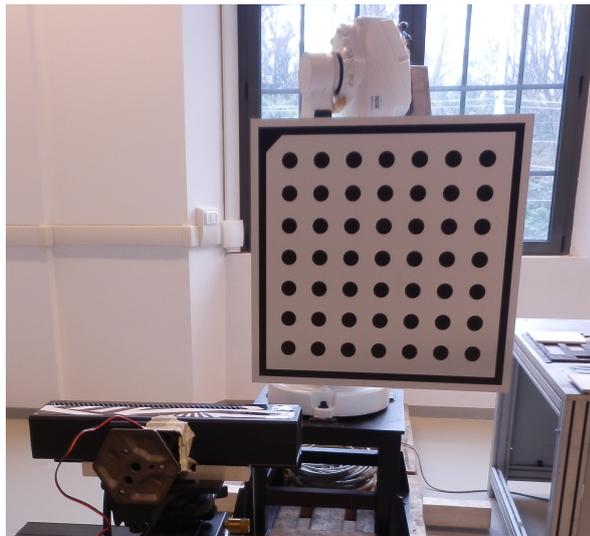


Figura 5.5. Setup utilizzato per la caratterizzazione del singolo pixel. Sulla flangia del robot è fissato un calibratore utilizzato nella procedura di allineamento.

Preliminarmente, mediante procedura di calibrazione (descritta nell'appendice B) basata su appositi algoritmi legati all'analisi delle immagini provenienti dalla camera IR [14], il bersaglio è stato allineato al sistema di riferimento del Kinect V2 al fine di poter effettuare traslazioni lungo l'asse Z uscente dalla camera. A seguito di quest'operazione vengono effettuati 51 posizionamenti, partendo da una distanza tra piano e camera di 800 mm ed incrementando questo valore, tra un posizionamento e il successivo, di una quota pari a 20 mm. Per ogni posa vengono acquisiti 4000 campioni di distanza relativi al pixel centrale.

Poiché il braccio robotico permette un'escursione massima pari a circa 1000 mm, al fine di poter coprire l'intero range di misura, la procedura di allineamento e acquisizione viene ripetuta per altre tre volte, partendo da distanze iniziali pari a 1600 mm, 2400 mm e 3200 mm così da avere zone di sovrapposizione ampie 200 mm. I risultati ottenuti sono mostrati nella

figura 5.6(a) in cui vengono riportati, per ogni set di acquisizioni, sull'asse X il valor medio della distanza e sull'asse Y la deviazione standard.

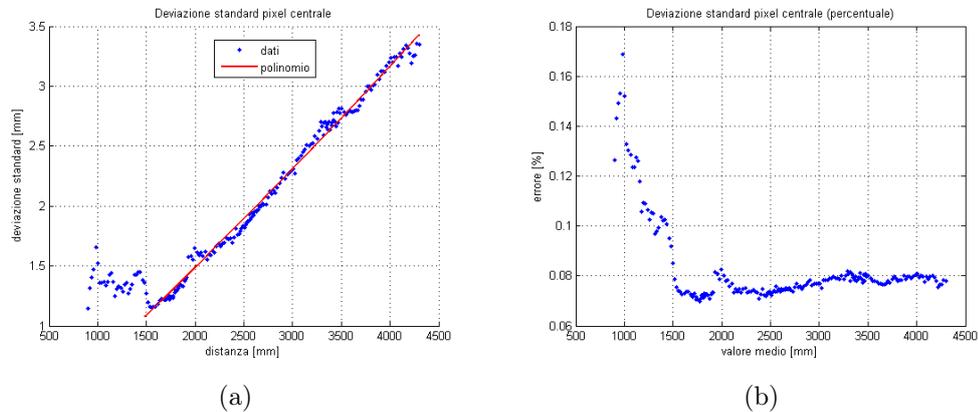


Figura 5.6. Andamento dell'effetto casuale relativo alla misura restituita dal pixel centrale al variare della distanza tra camera e piano acquisito, mostrato in valore assoluto (a) e percentuale (b).

Queste prove mostrano un andamento dell'effetto casuale relativo alla misura il quale cresce all'aumentare della distanza tra sensore e scena. Nei primi 1500 mm di acquisizione il valore di deviazione standard si mantiene al di sotto di 1.5 mm; a metà del range di misura (2500 mm) esso ha un valore inferiore a 2 mm, il quale cresce fino a raggiungere, in corrispondenza dell'ultima misura (4200 mm), una quota pari a 3.3 mm.

Si denota quindi un andamento lineare dovuto al fatto che, all'aumentare dello spazio tra sensore e bersaglio, si ha una riduzione della quantità di luce IR che ritorna al sensore, e quindi un deterioramento del rapporto segnale-rumore.

In seguito viene considerato il set di acquisizioni relativo alle misure effettuate ponendo il sensore a circa 1250 mm dalla parete. Si calcola la deviazione standard relativa ad ogni pixel acquisito e la si riporta in un'immagine avente lo stesso rapporto (in pixel) del sensore presente all'interno della telecamera IR del Kinect V2. Dalla figura 5.7(a) si nota come la deviazione standard cresca con il crescere della coordinata radiale. Anche in questo caso, come nel precedente, una minore quantità di luce IR proveniente dalla periferia del cono di illuminazione, comporta una degradazione del rapporto segnale-rumore, il che si traduce, in corrispondenza dei pixel relativi alle zone adiacenti ai vertici del sensore, ad un aumento della rumorosità delle misure.

Infine in figura 5.7(b) sono riportati i risultati ottenuti analizzando sette immagini di profondità relative ad una superficie planare acquisita traslando lungo l'asse Z della camera in un campo di visione esteso da 750 mm a 3750 mm. I punti appartenenti agli angoli del sensore, ai quali è associato un valore di deviazione standard maggiore di 15 mm, sono molto probabilmente relativi a zone della scena non illuminate dal cono di luce IR e quindi i loro valori di distanza, restituiti dal Kinect V2, non sono affidabili.

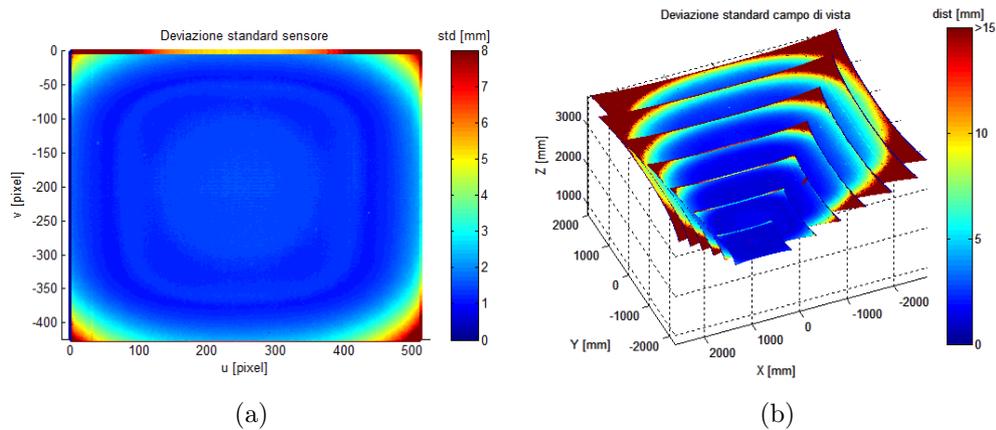


Figura 5.7. Andamento dell'effetto casuale relativo alla misura di distanza valutato all'interno del sensore (a) e del campo di visione (b).

5.2.2 Effetto sistematico relativo alla misura di un singolo pixel

Al fine di valutare l'effetto sistematico relativo al singolo pixel, sono state effettuate misure di distanza relative ad un piano al quale vengono imposti spostamenti noti. Il setup utilizzato in questa prova è lo stesso della precedente in cui un bersaglio planare è fissato alla flangia del medesimo braccio robotico ed allineato al sistema di riferimento del Kinect V2, mediante algoritmi di calibrazione applicati alle immagini restituite dalla camera IR (appendice B). Successivamente viene calcolato l'andamento della differenza tra lo spostamento realmente effettuato e il valor medio dei campioni acquisiti nelle quattro prove (paragrafo 5.2.1). I set di risultati sono allineati in un unico grafico (figura 5.8(a)), sovrapponendo i punti appartenenti ai tratti comuni tramite la minimizzazione delle distanze.

I risultati ottenuti denotano un errore caratterizzato da un andamento armonico, presente in letteratura col nome di "wiggling error" [16]. Ciò

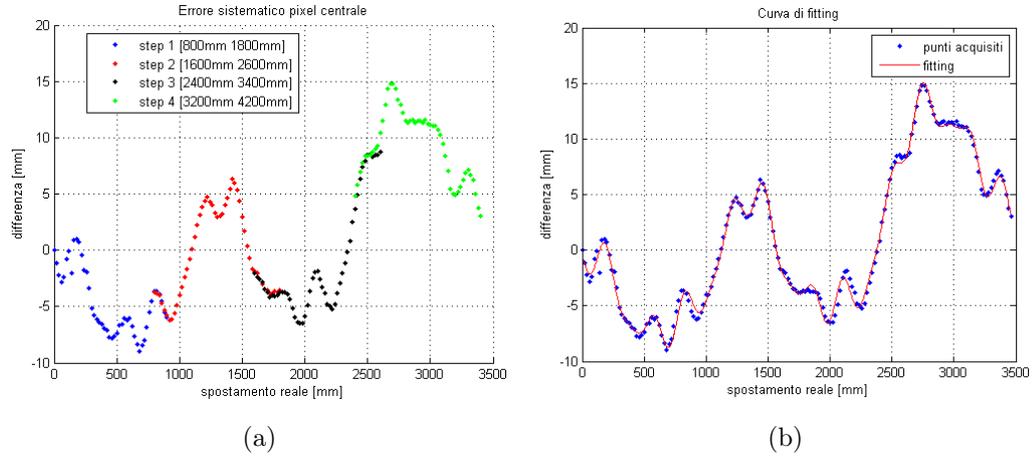


Figura 5.8. Andamento dell'effetto sistematico lungo l'asse Z, normale al piano (a) e curva di best fitting (b).

deriva dalla non perfetta modulazione sinusoidale del segnale IR il quale, essendo generato da un circuito elettronico digitale, presenta componenti armoniche superiori di ordine dispari [28]. Il modello matematico utilizzato dal Kinect V2 per convertire il valore di sfasamento in valore di distanza invece, prevede che il segnale riflesso, acquisito dalla camera IR, sia perfettamente sinusoidale e composto da una sola armonica. Da un'analisi dell'andamento dei punti ottenuti è stata calcolata una curva di fitting (figura 5.8(b)), definita come somma di funzioni sinusoidali, avente un coefficiente R^2 pari a 0.9915.

Inoltre, utilizzando un set di dati composto da 4000 frame di distanza relativi ad un piano acquisito a 1250 mm, viene calcolato il valor medio relativo ad ogni pixel del sensore. Attraverso la look up table relativa ai parametri intrinseci di calibrazione, estratta dal dispositivo grazie all'apposita funzione presente nell'SDK, viene generata la nuvola di punti corrispondente, a partire dalla quale sono calcolati i coefficienti del piano di best fitting passante per il baricentro. In figura 5.9 è riportata la distanza, in millimetri, tra il valor medio dei punti acquisiti e la superficie di riferimento. L'immagine riporta un andamento confinato in una banda inferiore a 20 mm in cui i punti relativi alla zona centrale presentano uno scostamento positivo mentre quelli appartenenti agli angoli sono caratterizzati da valori di distanza negativi. Inoltre si nota, in prossimità del centro, un punto avente scostamento elevato, probabilmente dovuto ad un difetto nella matrice del sensore.

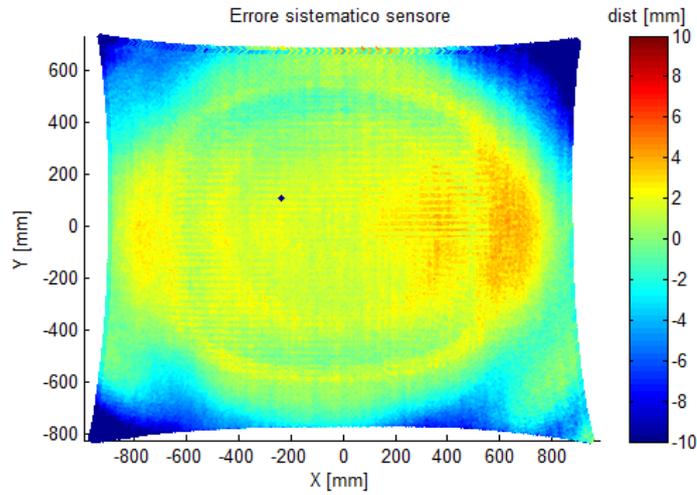


Figura 5.9. Andamento dell'effetto sistematico all'interno del frame.

5.2.3 Influenza dell'angolo di riflessione

In questa sezione viene valutata l'influenza, sulla misura di distanza, dell'angolo presente tra la luce incidente e il bersaglio.

La prova viene eseguita servendosi dello stesso setup utilizzato precedentemente, costituito da un target planare in alluminio fissato alla flangia di un braccio robotico ed allineato al piano del sensore.

Ciò permette di imporre al bersaglio rotazioni attorno ad assi paralleli agli assi X ed Y del sensore ed intersecanti la normale uscente dalla camera.

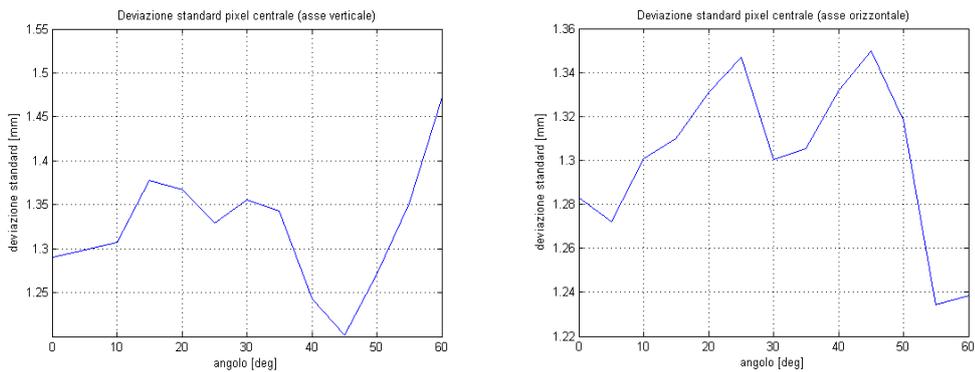


Figura 5.10. Andamento del valore di deviazione standard in funzione dell'angolo di incidenza tra luce e bersaglio.

Vengono eseguiti, partendo dalla posizione di piano parallelo, 13 posizionamenti ruotando il piano attorno all'asse verticale e 13 ruotandolo attorno all'asse orizzontale, compiendo step di 5° . Per ogni posa si acquisiscono 4000 frame di distanza calcolando la deviazione standard relativa alle misure restituite dal pixel centrale.

Dai grafici (figura 5.10) non emerge alcun andamento notevole che permette di legare l'angolo di riflessione all'errore di misure relativo al pixel centrale.

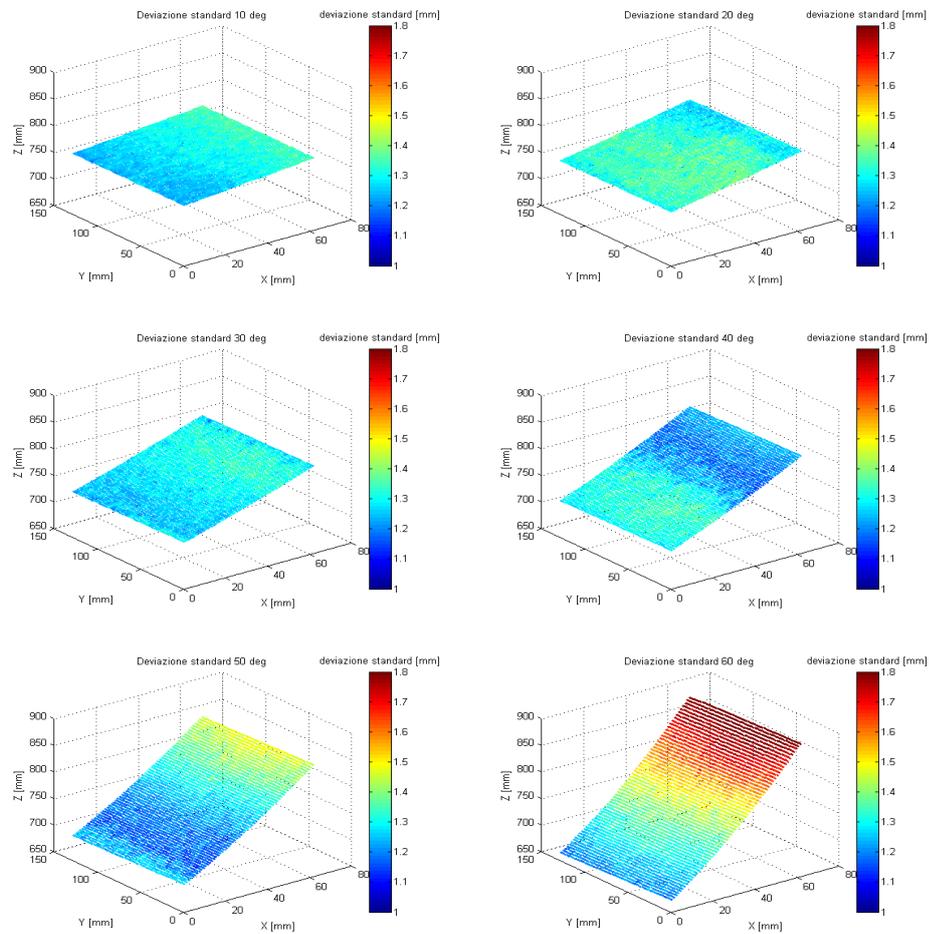


Figura 5.11. Relazione tra effetto casuale relativo alla misura di distanza e angolo di incidenza tra raggio di luce e bersaglio.

La prova viene quindi ripetuta estendendo l'analisi ad un ritaglio dell'immagine relativo ad una porzione centrale del sensore (62 x 130 pixel) interamente coperta dal bersaglio ed acquisendo per ciascuna posa 4000

frame di distanza. Per ogni pixel viene calcolata la deviazione standard il cui valore è rappresentato nelle immagini di figura 5.11, le quali, relative ad alcune pose significative, evidenziano una distribuzione a bande parallele all'asse di rotazione.

In particolare si nota come le acquisizioni relative alle ultime due pose (55 e 60°) presentino misure molto rumorose con valori di deviazione standard fino a 1.8 mm.

La medesima prova, eseguita analizzando le misure derivanti da rotazioni effettuate attorno all'asse orizzontale ha restituito simili risultati.

5.2.4 Influenza del materiale e del colore

Alcuni articoli presenti in letteratura, riguardanti telecamere tof, riportano la possibilità di riscontrare errori di offset riguardanti la misura di oggetti aventi superfici diverse, posti alla medesima distanza dal sensore [19]. La seguente prova è eseguita acquisendo provini di colori e materiali differenti, fissati, uno per volta, su di un sostegno piano alla distanza di circa 775 mm dal sensore. Per ognuno di essi vengono salvati 4000 campioni di distanza, restituiti dal pixel centrale. Al fine di limitare le variabili e disaccoppiare i problemi, si analizzano inizialmente campioni dello stesso materiale aventi colori diversi. Nella fattispecie, per la seguente analisi, sono stati utilizzati provini di cartoncino, stoffa e pellicola adesiva opaca. Per ciascuna tipologia vengono scelti i seguenti colori: nero, bianco, giallo, rosso, blu e verde. Per ogni set di acquisizioni vengono calcolati il valor medio e la deviazione standard relativi ad ogni singolo colore ed i risultati ottenuti sono mostrati in figura 5.12.

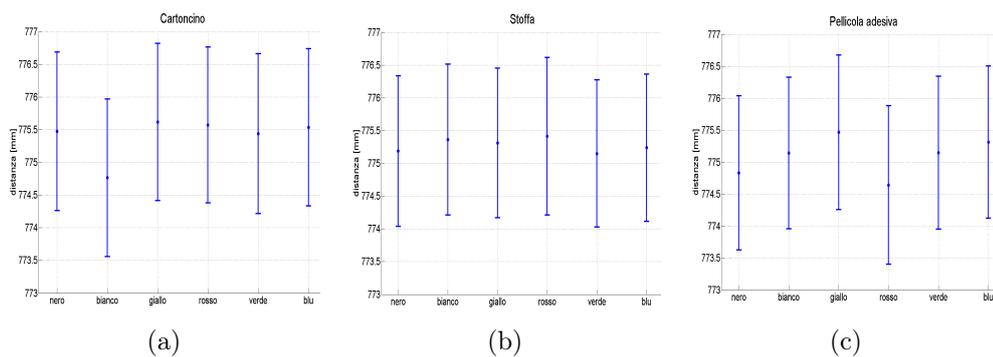


Figura 5.12. Risultati delle prove eseguite con campioni di colori diversi: cartoncino (a), stoffa (b), pellicola adesiva (c).

Analizzando i singoli grafici si nota come l'errore casuale, all'incirca dell'ordine del millimetro per tutte le acquisizioni, non cambia al variare del colore. Al contrario, in ogni prova, è presente un offset sistematico derivante da una differenza nei valori medi acquisiti. In particolare si evidenzia come, per quanto riguarda i provini di cartone, ci sia una differenza di quasi 0.8 mm tra il campione a cui è attribuita la distanza minore e quello a cui è attribuita quella maggiore. Uno scostamento simile è rilevato per i provini di pellicola adesiva. Infine, per i provini di stoffa, i valori medi delle misure sono confinati in una banda inferiore a 0.5 mm. Non avendo però riscontrato alcuna correlazione sistematica tra colore acquisito ed offset rilevato si estende la prova ad oggetti di materiali differenti. Con il medesimo setup della precedente prova, vengono effettuate misure relative ad oggetti di materiali diversi. In particolare, oltre a quelli precedentemente analizzati, vengono introdotti campioni composti da pellicola di alluminio, lastra di legno e tessuto riflettente. Un'analisi simile a quella precedentemente riportata conduce ai risultati riportati in figura 5.13.

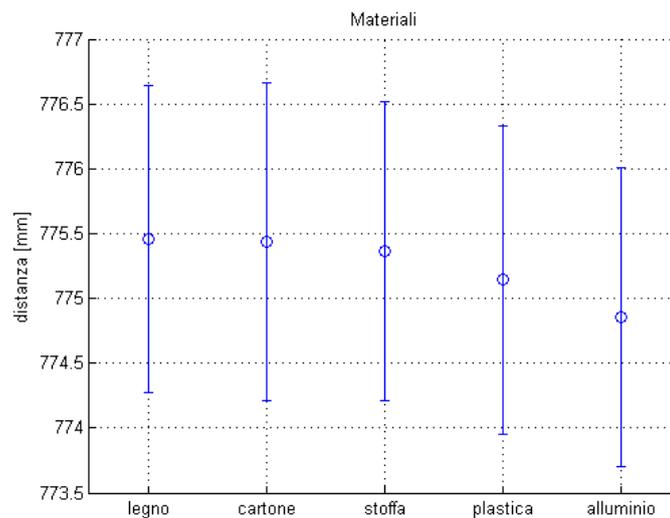


Figura 5.13. Risultati delle prove eseguite con campioni di materiali differenti.

Anche questo grafico mostra valori di deviazione standard pressoché costanti e una banda di circa 0.7 mm in cui sono confinati i valori medi delle acquisizioni. L'oggetto rilevato più vicino è il campione costituito dalla pellicola di alluminio mentre quello rilevato più lontano è la lastra di legno. La misura di distanza relativa all'oggetto costituito da tessuto riflettente restituisce un valore pari a zero (il che, convertito in nuvola di punti, si trasforma in assenza di dato). Considerando l'intero set

di prove si può concludere che la causa legata all'errore di offset è da attribuirsi alla correlazione tra risposta non lineare dei singoli pixel che compongono il sensore IR ed una differente riflettività, alla radiazione infrarossa, dei materiali. Infatti all'estremo superiore del grafico troviamo il legno, caratterizzato da una superficie ruvida mentre a quello inferiore troviamo l'alluminio, il quale presenta una superficie molto riflettente. Articoli presenti in letteratura, riguardanti l'analisi di telecamere tof, riportano la presenza di errori di offset derivanti dall'utilizzo di differenti tempi di integrazione [28]. Possiamo ipotizzare che il Kinect V2 adatti automaticamente il proprio tempo di integrazione in relazione alla quantità di luce retroriflessa dall'oggetto in analisi. La prova effettuata con il tessuto riflettente in aggiunta ad altre condotte con provini aventi superficie bianca, posti nelle vicinanze del sensore, ha rivelato che il Kinect V2 non è in grado di rilevare oggetti caratterizzati da un'elevata riflettività.

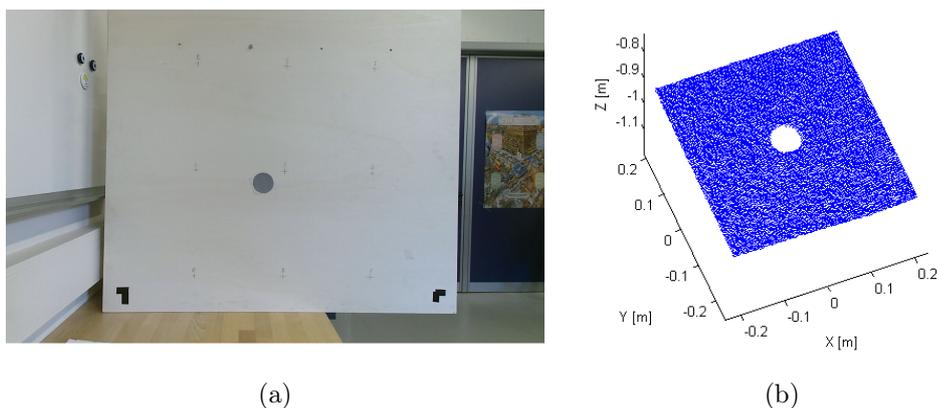


Figura 5.14. Setup costituito da un piano sul quale è stato applicato un marker circolare catarifrangente (a) e nuvola di punti ottenuta dalla scansione (b).

L'immagine e la nuvola di punti di figura 5.14 riportano il setup e i risultati della prova eseguita con materiale catarifrangente.

5.3 Caratterizzazione dell'intero sensore

Nella precedente sezione sono state eseguite prove mirate a valutare la qualità delle misure di distanza restituite dai singoli pixel. I test svolti in questa parte, invece, hanno la finalità di analizzare l'intera scena acquisita considerando la nuvola di punti ricostruita a partire dai dati restituiti dall'intero sensore. Viene quindi abbandonata la caratterizzazione

dell'aspetto distanziometrico per passare ad una visione globale in cui ad ogni punto sono associate coordinate 3D.

5.3.1 Performance nel ricostruire geometrie note

In questa sezione viene analizzata la qualità della ricostruzione di geometrie note come piani, cilindri e sfere misurando la distribuzione delle distanze tra i punti acquisiti e un modello matematico opportunamente allineato.

5.3.1.1 Piano

Inizialmente si valuta la qualità della ricostruzione di un piano acquisito parallelamente al sensore ed approssimativamente posto alla distanza di 1000 mm.

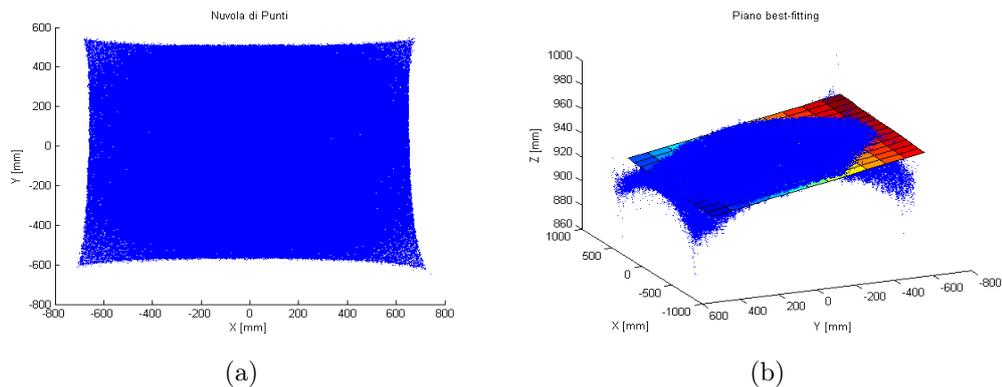


Figura 5.15. Nuvola di punti acquisita (a) e piano di best fitting (b).

La nuvola di punti ottenuta (figura 5.15(a)) ha una dimensione di circa 1400 mm nella direzione dell'asse X e 1200 mm in quella dell'asse Y . L'angolo di visione è di 70° lungo l'asse orizzontale e di 60° lungo l'asse verticale, la risoluzione a 1000 mm, lungo la direzione identificata dagli assi X e Y , è di 2.7 mm. Questi dati confermano le caratteristiche dichiarate dal costruttore.

A questa nuvola di punti viene applicata un'operazione di best fitting planare. Nel grafico riportato in figura 5.15(b) viene mostrato il piano ottenuto passante per il baricentro della nuvola di punti (rappresentata in colore blu).

Si nota un andamento concavo dell'acquisizione del piano. La maggiore distorsione, dell'ordine delle decine di millimetri, è presente agli angoli

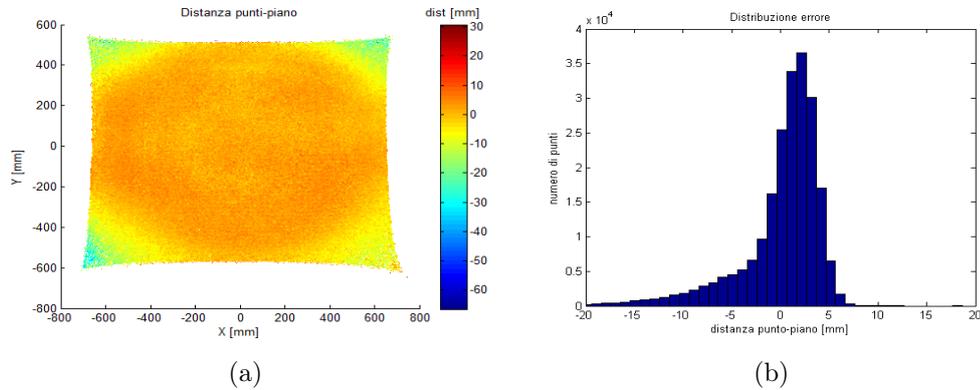


Figura 5.16. Distanza in millimetri calcolata tra il modello ed ogni singolo punto della nuvola (a) e distribuzione dell'errore (b).

dell'immagine. La figura 5.16(a) illustra l'andamento dell'errore di distanza, il quale, nella porzione centrale dell'immagine, è confinato in una banda di circa ± 5 mm, mentre in corrispondenza degli angoli, presenta valori di distanza pari ad alcune decine di millimetri.

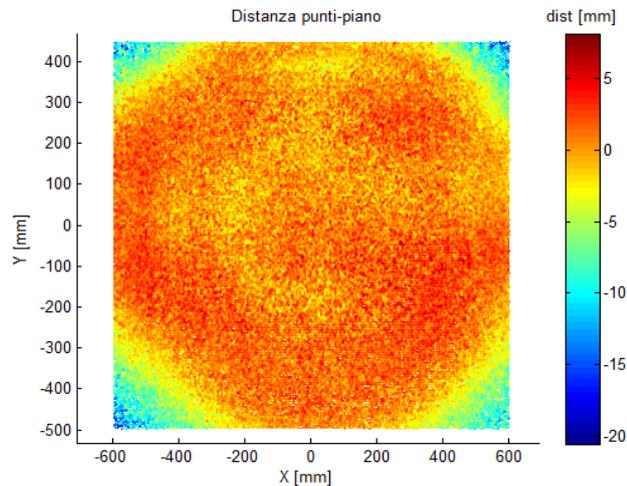


Figura 5.17. Distanza in millimetri calcolata tra il piano e ogni singolo punto della nuvola, considerandone una porzione centrale pari a circa il 70% della superficie del sensore.

Infine, analizzandone la distribuzione (figura 5.16(b)), si nota come la maggior parte dei punti presenta distanze confinate in un intorno di zero millimetri ampio poco più di 10 mm. La coda a sinistra dello zero rappresenta i punti appartenenti agli angoli del sensore.

La figura 5.17 mostra un ritaglio del sensore che ne prende in considerazione una porzione centrale avente una superficie pari al 70% di quella totale. In questo caso la distanza tra punti acquisiti e piano di best fitting (calcolato sui soli punti appartenenti alla nuvola ritagliata) è confinata in una banda che si estende da -15 a +5 mm.

Dalle precedenti analisi si evince un netto scostamento dal piano di riferimento dei punti appartenenti agli angoli del sensore. Ciò è in parte dovuto alle distorsioni ottiche introdotte dalle lenti della telecamera IR, le quali non sono completamente compensate dal modello che permette di convertire le immagini di profondità in nuvola di punti.

5.3.1.2 Cilindro

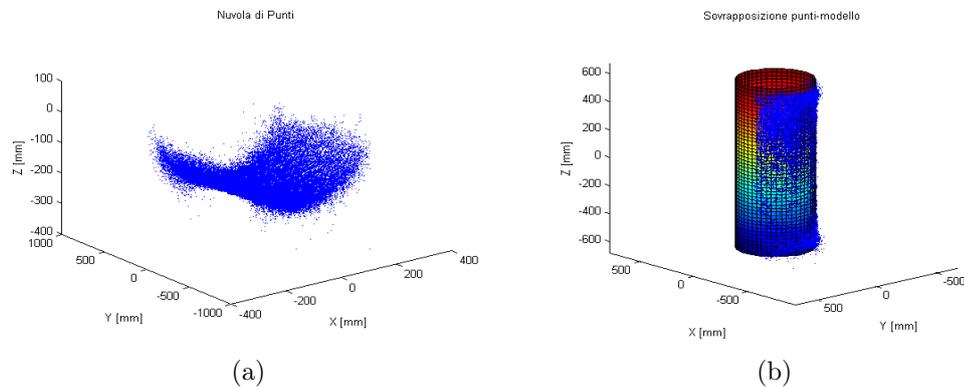


Figura 5.18. Nuvola di punti acquisita (a) e sovrapposizione al modello di riferimento (b).

Un altro oggetto semplice, di cui viene effettuata la ricostruzione 3D al fine di valutarne la qualità, è costituito dal cilindro.

Si acquisisce, ponendolo ad una distanza di circa 1000 mm dal sensore, un oggetto di forma cilindrica avente diametro nominale, dichiarato dal costruttore, pari a 500 mm e tolleranze di valore trascurabile rispetto all'accuratezza del sistema di misura. L'immagine 5.18(a) riporta la nuvola di punti relativa alla porzione di superficie acquisita.

Mediante script Matlab viene inizialmente generato un modello matematico di cilindro avente diametro esattamente pari a quello dell'oggetto in esame. Successivamente, applicando la matrice di rototraslazione ricavata mediante un algoritmo di registrazione rigida chiamato ICP (appendice A.2), viene sovrapposta ed allineata al modello matematico la nuvola di

punti acquisita (figura 5.18(b)). Questa operazione garantisce la minimizzazione delle distanze tra i punti restituiti da Kinect V2 e quelli del cilindro virtuale.

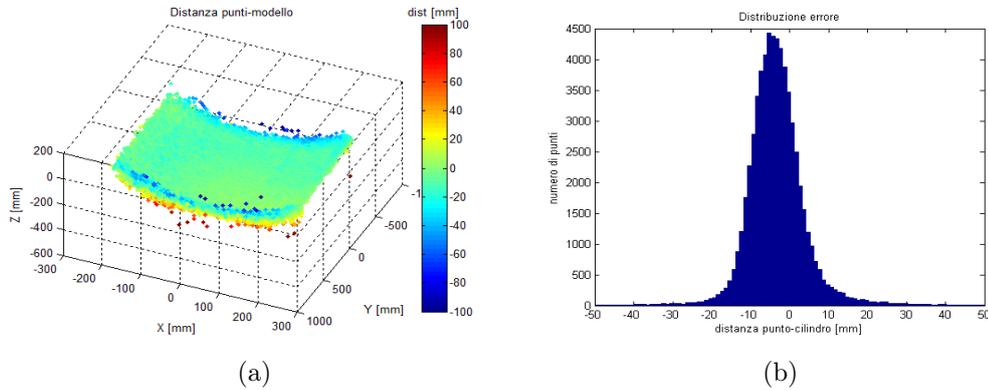


Figura 5.19. Distanza in millimetri calcolata tra il modello ed ogni singolo punto della nuvola (a) e distribuzione dell'errore (b).

Queste ultime vengono riportate nell'immagine 5.19(a) la quale mostra come la maggior parte dei punti acquisiti presenti uno scostamento inferiore a 25 mm. Inoltre, come precedentemente sottolineato, si nota che le zone connotate da un maggior distacco dal modello (dell'ordine delle decine di millimetri) siano quelle corrispondenti al bordo superiore ed inferiore del sensore. Viene riportata in figura 5.19(b) la distribuzione dell'errore di distanza. In questo grafico è ben visibile come la maggior parte dei valori sia confinata in un intorno dello zero ampio meno di 40 mm.

5.3.1.3 Sfera

L'ultimo oggetto, del quale si vuole valutare l'accuratezza della ricostruzione 3D, è rappresentato dalla sfera.

Si procede quindi acquisendo la porzione di superficie visibile di un solido avente forma sferica e raggio pari a 105 mm, con tolleranze trascurabili rispetto all'accuratezza del sistema di misura, posto ad una distanza di circa 1000 mm dal sensore (figura 5.20(a)).

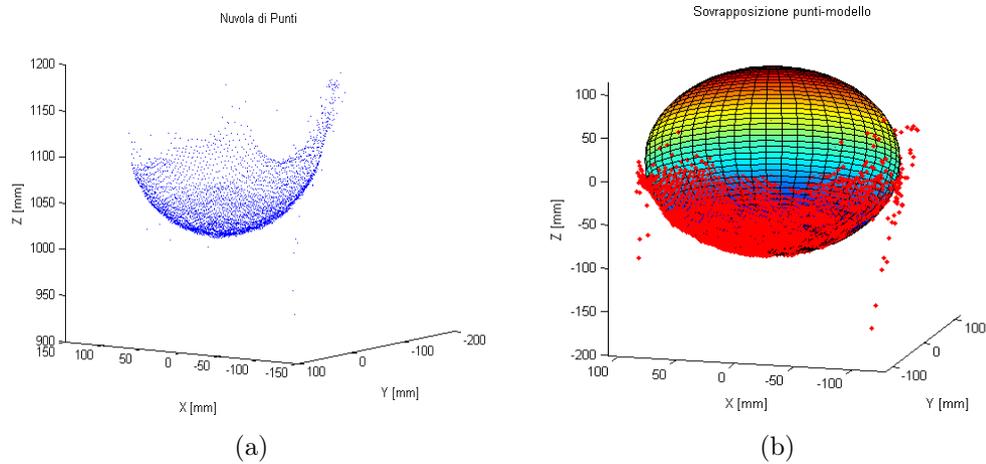


Figura 5.20. Ritaglio di una porzione della nuvola di punti relativa alla superficie sferica (a) e sovrapposizione tra sfera di riferimento e punti acquisiti (b).

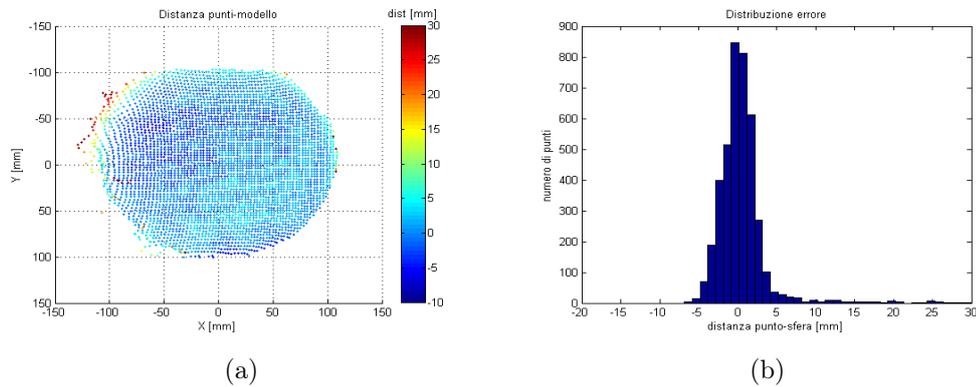
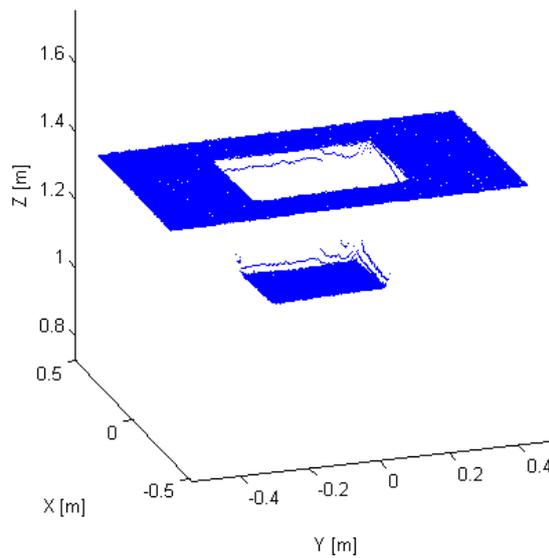


Figura 5.21. Distanza radiale in millimetri calcolata tra il modello ed ogni singolo punto della nuvola (a) e distribuzione dell'errore (b).

Anche in questo caso l'acquisizione viene allineata ad un modello matematico di sfera, tramite registrazione rigida ICP (figura 5.20(b)). Dalle immagini riportate in figura 5.21 si può notare come la maggior parte dei punti acquisiti disti meno di 10 mm dalla superficie di riferimento. Alcuni punti appartenenti al contorno della nuvola presentano scostamenti decisamente superiori dovuti con elevata probabilità all'errore di mixed pixel analizzato nel seguente paragrafo.

5.3.2 Mixed pixels

La seguente prova analizza la fedeltà delle acquisizioni riguardanti elementi posti ad alcune decine di centimetri da uno sfondo ed in particolare la ricostruzione della zona di transizione tra bordo dell'oggetto in esame e scena retrostante. Il setup utilizzato per la seguente prova (figura 5.22(b)) è composto da una superficie piana (di dimensioni pari a circa 250 x 300 mm), costituita da materiale rigido e sottile (3 mm), posta alla distanza di circa 350 mm rispetto ad un fondale anch'esso piano. Le acquisizioni sono eseguite ponendo il Kinect V2 a circa 1050 mm dall'oggetto in esame. La nuvola acquisita (figura 5.22(a)) presenta dei punti spuri nella zona di stacco tra oggetto e sfondo, i quali compongono una sorta di raccordo non presente nella realtà.



(a)



(b)

Figura 5.22. Nuvola di punti (a) e setup (b) utilizzati per l'identificazione dell'errore di mixed pixel.

L'effetto di mixed pixel (o jump edge [29]) è ben visibile analizzando due sezioni, della nuvola di punti acquisita, passanti per il centro del sensore e rispettivamente parallele ai piani YZ e XZ individuati dagli assi cartesiani relativi al sensore della tof camera (figura 5.23).

Questo fenomeno deriva dall'operazione di filtraggio applicato alla mappa di profondità al fine di ridurne la rumorosità [16]. Ciò è eseguito

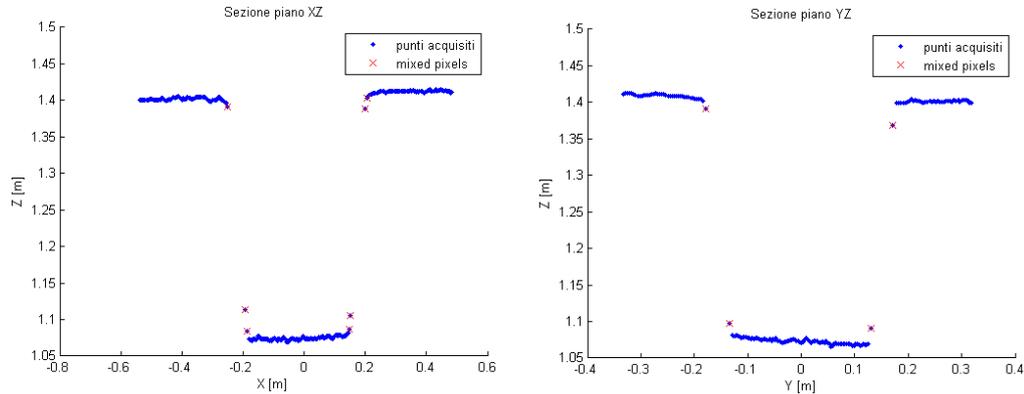


Figura 5.23. Sezioni passanti per il centro dell'acquisizione in cui sono evidenziati i “mixed pixels”.

mediando iterativamente, con opportuni pesi, il valore di gruppi di pixel adiacenti, appartenenti a sottomatrici del sensore. È ben evidente come, in presenza di bordi, e quindi di pixel consecutivi aventi valori di profondità molto diversi, l'operazione in esame attribuisca ad alcuni punti quote intermedie. Inoltre è stato verificato che nel caso in cui lo sfondo sia ad una distanza tale da non essere acquisito, questo fenomeno non si verifica.

5.3.3 Riflessioni multiple

Infine, un noto problema di misura relativo alle telecamere tof è quello generato dalle riflessioni multiple (multipath). Questo fenomeno ha luogo in presenza di geometrie concave schematizzabili attraverso un diedro costituito da due piani secanti.

L'errore di misura si verifica poiché parte della luce IR, incidente su ciascuna delle due superfici, ritorna alla telecamera solo dopo aver compiuto una seconda riflessione sul piano adiacente, percorrendo così, come mostrato in figura 5.24(a), un tragitto più lungo di quello ideale e generando distorsioni (figura 5.24(b)) nella nuvola di punti acquisita [3, 10].

Per analizzare e misurare l'entità di tale differenza è stato realizzato un apposito set-up (figura 5.25) costituito da due superfici piane (parete bianca e tavola di legno) opportunamente posizionate ed allineate.

Acquisendo mediante una stazione totale (Leica modello TS06) 9 punti su ciascun piano, è stato possibile ricavare un riferimento accurato delle due superfici e del loro posizionamento relativo, con il quale confrontare le nuvole di punti generate da Kinect V2. Poiché i due dispositivi di

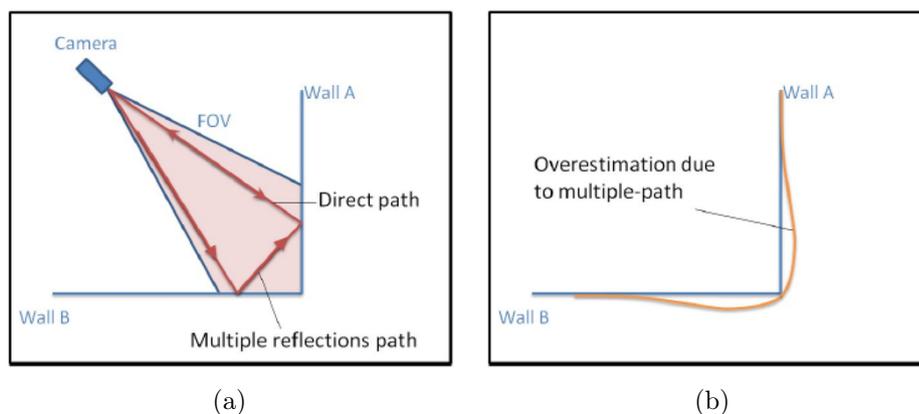


Figura 5.24. Schema rappresentante il percorso compiuto dai raggi luminosi in presenza di una scena in cui si verificano riflessioni multiple (a) e acquisizione distorta relativa ad una geometria concava (b).

acquisizione presentano differenti pose nella stanza e quindi differenti sistemi di riferimento, è necessario effettuare una procedura di registrazione che permetta di ricavare la matrice di rototraslazione presente tra di essi, utile al fine di allineare e sovrapporre la nuvola di punti acquisita ai piani di riferimento ottenuti tramite stazione totale.

Operativamente, ciò è stato ottenuto ponendo su ciascuna parete tre marker circolari di colore nero, ben visibili in figura 5.25, e ricavando le coordinate 3D associate al centro di ciascuno di essi in entrambi i sistemi di riferimento.

In particolare, per quanto riguarda quello relativo alla stazione totale, si è proceduto misurando, con quest'ultima, la posizione dei centri, opportunamente evidenziati, di ciascun disco, mentre in quello associato al Kinect V2, le coordinate ricercate sono state ricavate tramite l'analisi dell'immagine IR e della nuvola di punti. Nel dettaglio è stata utilizzata l'immagine IR per ricavare le aree corrispondenti ai marker (indistinguibili nella point cloud) delle quali sono stati ricavati i baricentri le cui coordinate 2D (in pixel) hanno permesso di individuare, nella nuvola di punti, le rispettive coordinate 3D.

Successivamente, attraverso un'analisi svolta mediante un apposito script Matlab, in cui è implementato un algoritmo di minimizzazione di distanze, basato su decomposizione ai valori singolari (appendice A.1), è stata ricavata, a partire dalle coordinate, relative ai centri dei marker, ricavate in precedenza, la matrice di rototraslazione presente tra il sistema di assi cartesiani utilizzato dal Kinect V2 e quello relativo alla stazione

totale.



Figura 5.25. Setup utilizzato per l'analisi delle riflessioni multiple; si notino i marker utilizzati per ricavare la rototraslazione dei due sistemi di riferimento.

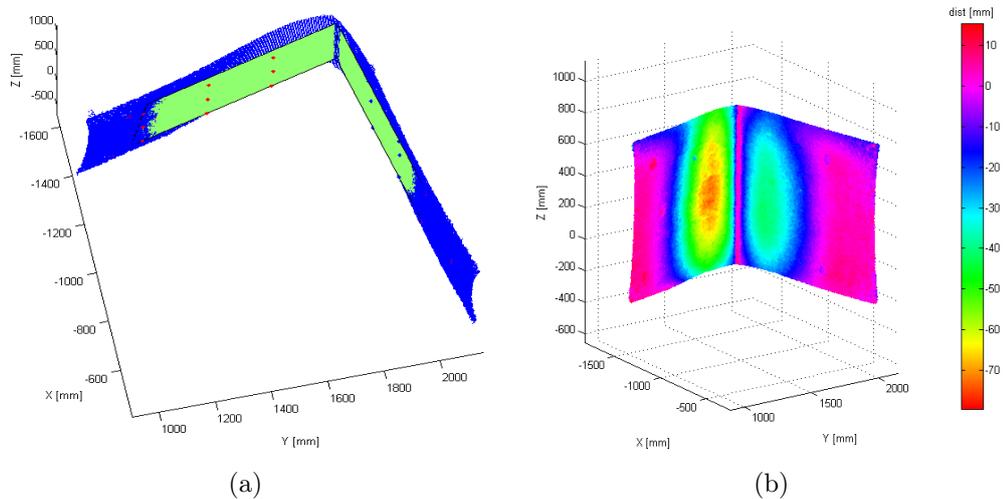


Figura 5.26. Allineamento tra nuvola di punti e piani di riferimento (a) e distanza punti-piano (b).

Nell'immagine 5.26(a) viene riportato l'allineamento tra nuvola di punti acquisita (blu) e piani di riferimento (verde). Dall'analisi della distanza tra nuvola di punti e relativo piano di riferimento (figura 5.26(b)) si ricava come le maggiori distorsioni, dell'ordine delle decine di millimetri, siano relative a

punti situati ad una distanza dallo spigolo tra i piani compresa tra 150 e 250 millimetri. Inoltre si noti come la superficie (in muratura), rappresentata nella parte destra dell'immagine, presenti una minore distorsione, con valori massimi pari a 50 mm, rispetto a quella di sinistra (in legno) in cui alcuni punti sono caratterizzati da un valore di distorsione pari ad 80 mm.

5.4 Conclusioni

In questa sezione è stata presentata una serie di prove relative alla qualificazione metrologica di una telecamera tof a basso costo prodotta da Microsoft (Kinect V2) atte a verificarne l'accuratezza nella misura e la qualità nella ricostruzione tridimensionale.

Da uno studio preliminare è emerso come il valore relativo alla misura di distanza effettuata in condizioni statiche sia fortemente influenzato dalla variazione della temperatura interna al dispositivo. L'aggiunta di una ventola esterna ha permesso di stabilizzare l'andamento della misura e migliorare le prestazioni.

L'analisi del singolo pixel (centrale) ha mostrato come l'incertezza di misura dovuta all'effetto casuale, cresca linearmente con la distanza tra bersaglio e camera fino a raggiungere il valore massimo pari a 3.3 mm in corrispondenza delle misure effettuate in prossimità dell'estremo superiore del range di misura (4200 mm).

Inoltre, le mappe di distanza acquisite dall'intero sensore rivelano un legame tra coordinata radiale e incertezza di misura; quest'ultima infatti aumenta all'allontanarsi dal centro dell'immagine. In particolare, in acquisizioni relative a piani posti a più di 1500 mm dalla camera, sono presenti misure di distanza molto rumorose e non affidabili, associate ai pixel relativi alle zone adiacenti agli angoli del sensore. Ciò è dovuto alla scarsa illuminazione incidente sulla porzione periferica della scena.

L'effetto sistematico associato alla misura di distanza restituita dal pixel centrale presenta un andamento armonico che assume valori confinati in un intervallo compreso tra -10 e 15 mm.

I test eseguiti inclinando il bersaglio, rispetto alla condizione di allineamento tra quest'ultimo e la camera, hanno evidenziato una correlazione tra il valore di deviazione standard calcolata e l'angolo di incidenza tra raggio di luce e piano.

Le misure effettuate utilizzando bersagli di diverso materiale e colore, posti alla medesima distanza dal sensore, hanno presentato un offset legato alla diversa riflettività (nello spettro IR) della superficie. Questo scostamento, mantenendosi inferiore al millimetro, può essere trascurato

nella maggior parte delle applicazioni. Solo in casi particolari, in cui il target è molto riflettente, il Kinect V2 non è in grado di restituire valori di distanza affidabili.

Per quanto riguarda l'analisi basata sull'acquisizione e ricostruzione di geometrie note (piano, cilindro, sfera) è stato riscontrato uno scostamento massimo, misurato tra modello matematico e nuvola di punti, avente un valore assoluto pari a 20 mm.

Si è inoltre rilevato che la problematica denominata mixed-pixel si verifica solo nel caso in cui l'oggetto acquisito si trovi di fronte ad uno sfondo situato nel campo di misura del Kinect V2. Infine in alcune acquisizioni, relative a particolari geometrie, è stato riscontrato l'effetto dovuto alle riflessioni multiple, nota limitazione della tecnologia a tempo di volo; ciò ha generato distorsioni, le quali, nelle aree di maggiore scostamento tra nuvola di punti e modello, assumono valori pari a 80 mm.

Capitolo 6

Analisi metrologica del tracciamento di una figura umana

Questa sezione illustra il setup, lo svolgimento ed i risultati relativi ad alcuni test svolti presso il presidio dell'Ospedale Valduce "Villa Beretta" di Costa Masnaga (LC), al fine di qualificare sia a livello quantitativo che qualitativo il tracciamento e la ricostruzione della posizione dei giunti del corpo umano, effettuati tramite Kinect V2. Vengono preliminarmente eseguite ed analizzate una prova di tipo statico, finalizzata a determinare il grado di rumorosità della misura, ed una di tipo dinamico, utile a ricavare la miglior posa relativa tra paziente e camera e a stabilire il grado di correttezza della traiettoria acquisita. Infine, seguendo un protocollo riabilitativo relativo all'arto superiore [5], stilato all'interno della clinica, sono acquisiti i movimenti compiuti da un paziente durante lo svolgimento di alcuni esercizi riabilitativi classici, al fine di valutare un possibile impiego del presente dispositivo come ausilio alla riabilitazione.

Le prove riportate in questo capitolo hanno come scopo l'analisi dell'accuratezza relativa alla ricostruzione, effettuata mediante Kinect V2, della posizione di alcuni particolari punti (in seguito chiamati giunti) di una figura umana. Ciò permette di qualificare la misura a più alto livello restituita dal dispositivo, così da determinare il grado di affidabilità dello

strumento durante un ipotetico utilizzo in ambito clinico e riabilitativo.

Come precedentemente descritto (paragrafo 4.2.1), il Kinect V2 è in grado di identificare, dalla nuvola di punti acquisita, le coordinate spaziali associate a 25 giunti del corpo umano.

La prova viene svolta confrontando le coordinate 3D, e quindi le traiettorie, relative ai giunti di un arto umano in movimento, restituite da Kinect V2 con quelle acquisite, contemporaneamente, servendosi di un sistema stereoscopico BTS composto da otto telecamere sensibili all'infrarosso e considerato come riferimento.

Quest'ultimo, le cui caratteristiche tecniche sono riportate nella tabella 6.1, è ritenuto, in termini di accuratezza, lo stato dell'arte attuale per quanto riguarda l'analisi del movimento umano, (accuratezza <0.1 mm all'interno di un volume di $6 \times 6 \times 3$ m [39]) ed è in grado di ricavare le coordinate 3D del centro di specifici marker catarifrangenti, sferici o semisferici, applicati sul corpo del paziente in corrispondenza dei punti anatomici di cui si vuole misurare la posizione rispetto ad un sistema di riferimento opportunamente posizionato all'interno della stanza.

Tabella 6.1. BTS SMART-DX 7000 - principali caratteristiche [39].

Risoluzione [pix]	2048 x 2048
F_{acq} alla massima risoluzione [fps]	500
F_{acq} massima [fps]	2000
Accuratezza [mm]	<0.1 (volume di $6 \times 6 \times 3$ m)
Lunghezza d'onda illuminatore [nm]	850
Massimo numero di marker rilevati	illimitato
Tipologia di marker	passivi e retroriflettenti
Diametro marker [mm]	3-20
Tipologia di connessione	Gigabit Ethernet
Alimentazione	fornita dalla datastation

La presente tecnologia, a causa del complesso hardware e del relativo costo, è quindi utilizzabile solo all'interno di strutture specializzate e da parte di personale qualificato. Il test è stato infatti condotto presso il presidio dell'Ospedale Valduce "Villa Beretta" di Costa Masnaga (LC).

Poiché i due dispositivi (Kinect V2 e sistema stereoscopico BTS) presentano sistemi di riferimento differenti, si è reso necessario condurre una preliminare procedura di registrazione, al fine di ricavare la matrice di rototraslazione utile a definire la posizione relativa tra gli assi cartesiani associati al sensore della tof camera Kinect V2 e il sistema di riferimento utilizzato dal dispositivo stereoscopico BTS, così da poter riportare entram-

bi i set di dati in un comune SDR, che nel caso specifico si è deciso essere quello utilizzato dal sistema BTS. Viene di seguito brevemente descritta la strategia adottata.

Inizialmente il dispositivo Kinect V2 è installato su un treppiede fotografico opportunamente orientato così da poter inquadrare l'intero volume in cui si muoverà il paziente. Da questo momento in poi il sistema di acquisizione non sarà più spostato e manterrà la medesima posa ed orientamento durante tutte le prove svolte.

In seguito sono distribuiti, in punti casuali della scena, otto dischi (marker), aventi un diametro pari a 8 cm, realizzati in materiale sottile e di colore nero, così da poter contrastare con le superfici su cui sono posizionati (figura 6.1(a)).

Acquisendo un'immagine in scala di grigi proveniente dalla camera IR del Kinect V2 ed analizzandola mediante l'applicazione di una soglia sull'intensità, è possibile individuare le aree della scena corrispondenti ai marker ed estrarne, con accuratezza inferiore al pixel, le coordinate relative ai baricentri. Grazie a queste ultime è possibile quindi ricavare dalla nuvola di punti (anch'essa ricostruita tramite dati provenienti dalla camera IR) relativa alla scena statica, le coordinate 3D corrispondenti alla posizione del centro di ciascun disco.

Successivamente viene posata su ciascun marker una semisfera catari-frangente (figura 6.1(b)), opportunamente centrata, avente diametro pari a 15 mm, utile ad individuare le coordinate 3D dei medesimi punti nel SDR relativo al dispositivo stereoscopico.

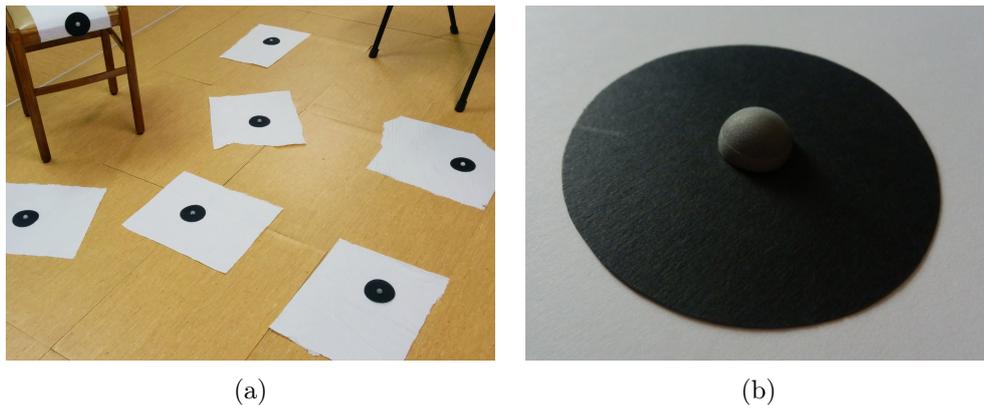


Figura 6.1. Alcuni marker posti nel volume di lavoro (a) e dettaglio di un singolo marker (b).

Sono così ottenute sedici triplette di coordinate (XYZ) mutuamente

accoppiate in base all'associazione tra punti omologhi.

Esse vengono processate mediante script Matlab in cui sono implementati opportuni algoritmi di minimizzazione di distanze basati su Singular Value Decomposition (SVD) (appendice A.1) al fine di ricavare la matrice \mathbf{R} di rotazione e il vettore \mathbf{T} di traslazione, utilizzati successivamente per effettuare la trasformazione di coordinate.

In seguito alla procedura di calibrazione è stata condotta l'acquisizione dei dati.

Servendosi del software descritto nel paragrafo 4.2.1.1 vengono salvati all'interno di un file di testo, ad una frequenza di 30 Hz, sia le coordinate XYZ di ciascun giunto del corpo umano sia il timestamp relativo all'istante di acquisizione.

Similmente, il sistema stereoscopico fornisce come output un file contenente le coordinate 3D, acquisite a 250 Hz, relative alla posizione spaziale assunta dai blob catarifrangenti, posti in corrispondenza dei punti del corpo umano di cui si vuole tracciare il movimento.

Durante tutte le prove, il paziente si posiziona su di una sedia situata a circa 2.5 m (metà dal campo di operatività) dalla telecamera a tempo di volo e si mantiene centrato nell'inquadratura, così da garantire che, nell'eseguire l'esercizio, tutti i giunti del proprio corpo vengano correttamente visti ed identificati da entrambi i sistemi (figura 6.2(a)).

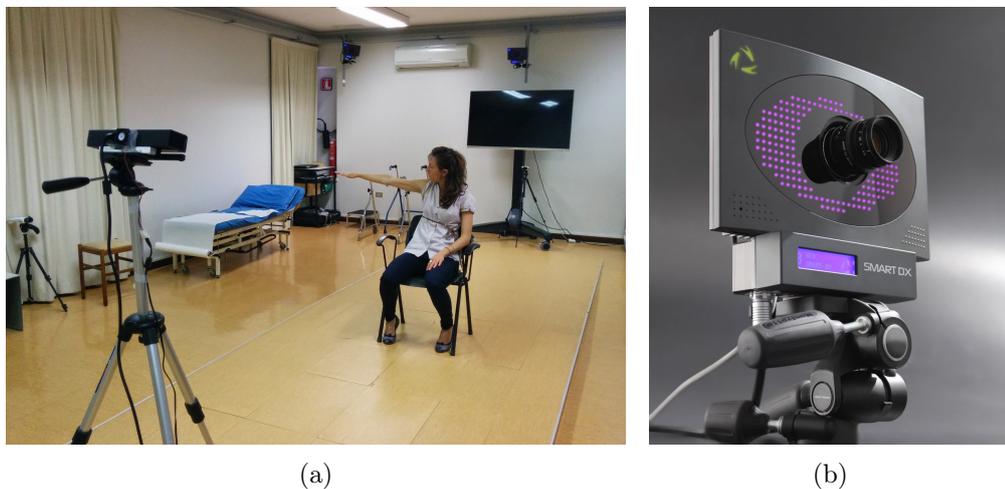


Figura 6.2. Paziente durante l'esecuzione di un esercizio (si notino il Kinect V2 posizionato frontalmente rispetto al paziente e due delle otto telecamere BTS poste sulla parete di fondo) (a) e dettaglio raffigurante una telecamera BTS SMART-DX (b).

6.1 Effetto casuale relativo alla misura della posizione di un giunto

La prima prova, di tipo statico, ha come obiettivo quello di valutare la rumorosità della misura relativa alla tracciatura della posizione di un giunto umano, acquisito in condizioni di assenza di movimento. In particolar modo, riferendosi alle necessità e alle caratteristiche dell'applicazione in esame, viene scelto il giunto corrispondente al polso; quest'ultimo infatti, nell'applicazione, viene tracciato al fine di rilevarne il movimento da applicare ad un avatar, in grado di afferrare e spostare gli oggetti virtuali presenti nella scena.

Vengono quindi svolte tre campagne di acquisizione, in ciascuna delle quali il paziente, posto in posizione seduta di fronte alla camera, mantiene per 60 secondi il braccio fermo e teso parallelamente all'asse frontale del proprio corpo (6.3). Al fine di garantire l'immobilità dell'arto è stata prevista una struttura metallica di sostegno opportunamente realizzata e posizionata per non interferire con l'acquisizione.

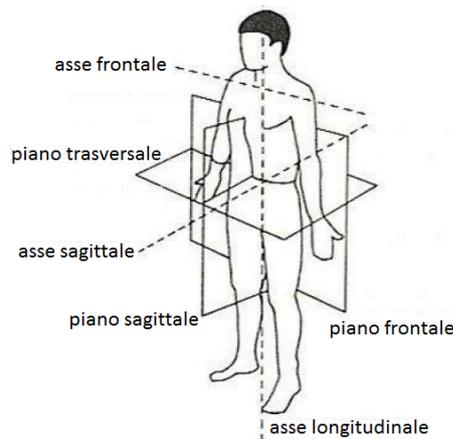


Figura 6.3. Principali assi e piani del corpo umano.

Calcolando, per ciascuno dei tre set di dati, la deviazione standard associata ai valori di coordinata X , Y e Z , viene ricavata la rumorosità della misura relativa alla posizione del giunto acquisita in condizioni statiche.

Ai risultati dell'analisi, riportati in tabella 6.2, viene applicata la formula 6.1, ottenendo così lo scarto tipo complessivo $\bar{\sigma}_i$ pari, rispettivamente nei tre casi, a $\bar{\sigma}_1 = 1.10 \text{ mm}$, $\bar{\sigma}_2 = 1.10 \text{ mm}$, $\bar{\sigma}_3 = 1.09 \text{ mm}$.

$$\bar{\sigma}_i = \sqrt{\sigma_{iX}^2 + \sigma_{iY}^2 + \sigma_{iZ}^2} \quad (6.1)$$

Tabella 6.2. Valori di deviazione standard ottenuti dall'analisi dei dati acquisiti nelle tre prove.

<i>Dev. standard</i>	<i>Valore</i>
σ_{1X} [mm]	0.82
σ_{1Y} [mm]	0.57
σ_{1Z} [mm]	0.52
σ_{2X} [mm]	0.51
σ_{2Y} [mm]	0.78
σ_{2Z} [mm]	0.57
σ_{3X} [mm]	0.75
σ_{3Y} [mm]	0.30
σ_{3Z} [mm]	0.73

I valori risultanti dalla presente prova, ed in particolare quelli relativi alle quote misurate lungo l'asse Z , mostrano un grado di rumorosità inferiore a quello presentato nella precedente analisi e qualificazione metrologica relativa al singolo pixel (paragrafo 5.2).

Ciò è indice del fatto che il sistema ricava i valori di X , Y e Z , associati alla posizione assunta da ciascun giunto, mediando opportunamente le coordinate 3D relative ai punti appartenenti ai segmenti della nuvola in cui viene identificata la presenza di una figura umana. In questo modo la misura ottenuta è caratterizzata da un'inferiore rumorosità.

6.2 Confronto tra Kinect V2 e sistema BTS (movimenti semplici)

La seconda prova, di tipo dinamico, prevede che il paziente, seduto, ruoti il braccio destro, teso inizialmente lungo il fianco, attorno al proprio asse frontale, così da far definire al polso una traiettoria approssimativamente circolare, avente il centro posto in corrispondenza della spalla.

Durante l'esercizio vengono acquisite e salvate, contemporaneamente con Kinect V2 e con il sistema stereoscopico BTS, le coordinate X , Y e Z relative alla posizione spaziale assunta dal polso in ogni istante di tempo.

Al fine di individuare una possibile condizione preferenziale, in cui l'acquisizione del movimento risulti essere più fedele alla realtà, il test viene suddiviso in tre fasi: nella prima il paziente esegue l'esercizio ponendosi frontalmente alla tof camera (asse sagittale parallelo all'asse Z del sensore), nella seconda si pone lateralmente (asse sagittale ortogonale all'asse Z),

nella terza assume una posa intermedia rispetto alle due precedenti (asse sagittale a circa 45° rispetto all'asse Z).

Viene inizialmente analizzato il primo set di dati, riportando nei grafici di figura 6.4 la sovrapposizione delle componenti X , Y e Z , relative alla traiettoria percorsa, rispettivamente acquisite con il sistema BTS (in blu) e con Kinect V2 (in rosso).

Queste ultime, al fine di poter essere confrontate con la relativa controparte fornita dal sistema stereoscopico, vengono dapprima interpolate e ricampionate alla frequenza di acquisizione del sistema BTS (250 Hz), successivamente sono trasformate applicando l'opportuno cambio di sistema di coordinate descritto in precedenza ed infine sincronizzate alle rispettive traiettorie di riferimento. Quest'ultima operazione è svolta calcolando il valore di cross-correlazione presente tra ogni coppia di serie di dati, così da ricavarne lo sfasamento temporale.

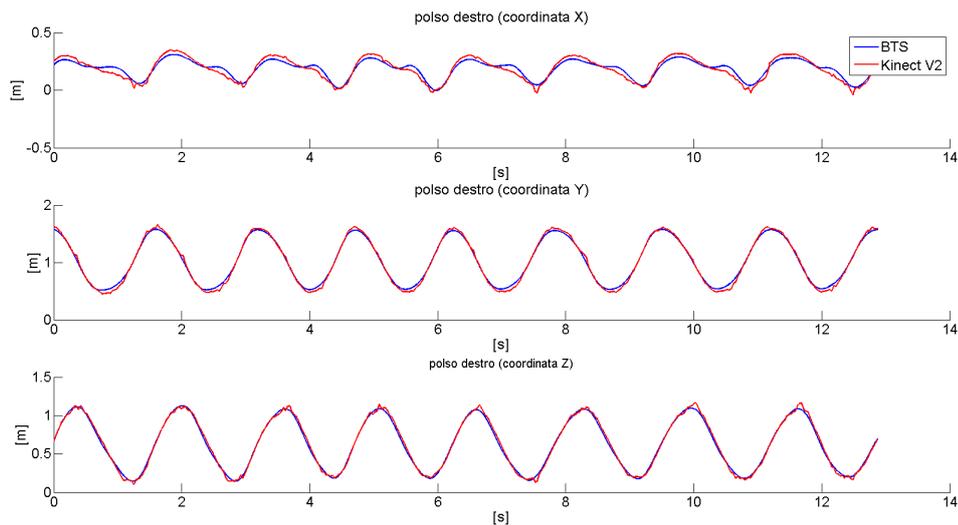


Figura 6.4. Scomposizione e sovrapposizione delle traiettorie relative alle rotazioni compiute dal polso destro, acquisite con sistema BTS e con Kinect V2 (paziente in posizione frontale).

Analizzando qualitativamente il grafico di figura 6.4, si nota come le traiettorie appartenenti a ciascuna coppia, presentando un andamento molto simile, siano tra di loro ben allineate.

Nello specifico, riferendosi al tracciato rilevato da Kinect V2, si identifica che i punti a massimo scostamento, pari a 20 mm, spesso caratterizzati da un andamento rumoroso, sono localizzati nell'intorno degli istanti in cui le curve presentano il massimo (o minimo) valore di ampiezza, indice di

un'inversione della direzione del movimento. Per contro si nota come nei tratti in cui le funzioni assumono valori di derivata sufficientemente elevati (in valore assoluto), ci sia un'ottima corrispondenza tra le due traiettorie acquisite.

Si può quindi ipotizzare che l'algoritmo basi la stima della posizione attuale del giunto su quelle assunte in precedenza e quindi restituisca risultati incerti in presenza di improvvisi cambi di direzione.

Inoltre alcune differenze, sistematicamente riscontrabili tra le coppie di traiettorie, possono essere attribuite all'offset presente tra il centro del marker catarifrangente, posto sullo stiloide ulnare, e il baricentro del polso umano identificato dal Kinect V2. Ciò infatti, nel caso di rotazione dell'arto superiore, determina la descrizione di due circonferenze aventi raggi differenti, il cui scostamento presenta un ordine di grandezza confrontabile con lo scarto tra le curve precedentemente analizzato.

Infine si nota, soprattutto nel grafico relativo alla componente Z in cui le funzioni assumono valori di ampiezza ridotti, come la rumorosità di misura non permetta di evidenziare piccole variazioni di movimento.

Al fine di derivare una cifra di merito, indicante l'entità della differenza presente tra la traiettoria acquisita da Kinect V2 e il relativo riferimento assoluto, e quindi la bontà dell'acquisizione effettuata con il sistema a tempo di volo, viene calcolato per ciascuna coordinata, lo scostamento medio (\bar{d}_\bullet) tra le due traiettorie. Esso, valutato su un insieme di campioni relativi ad un numero intero di rotazioni, è definito, ad esempio nel caso specifico di coordinata X , come

$$\bar{d}_X = \frac{\sum_{i=1}^{N_{samp}} |d_i|}{N_{samp}} \quad (6.2)$$

in cui N_{samp} corrisponde al numero di campioni considerati e d_i è pari a $X_{iKin} - X_{iBTS}$, dove X_{iKin} e X_{iBTS} rappresentano la coordinata X acquisita all' i -esimo istante rispettivamente da Kinect V2 e dal sistema BTS. Il medesimo calcolo è ripetuto per i campioni relativi alle coordinate Y e Z ed i valori ottenuti sono riportati nella tabella 6.3.

La stessa procedura di allineamento e sincronizzazione, utilizzata nel confronto delle traiettorie relative alla rotazione dell'arto superiore effettuata dal paziente in posizione frontale, viene applicata al secondo set di dati (paziente in posizione laterale) ottenendo il grafico di figura 6.5.

In questo caso i valori di distanza media calcolati sono riportati nella tabella 6.4.

Dal confronto di questi risultati si ricava come la migliore ricostruzione è ottenuta nel caso in cui il paziente si pone con asse sagittale ortogonale

Tabella 6.3. Scostamenti medi tra le traiettorie relative alle rotazioni compiute dal polso destro acquisite con sistema BTS e con Kinect V2 (paziente in posizione frontale).

<i>Scostamento medio</i>	<i>Valore</i>
$\bar{d}_X [mm]$	28
$\bar{d}_Y [mm]$	31
$\bar{d}_Z [mm]$	16

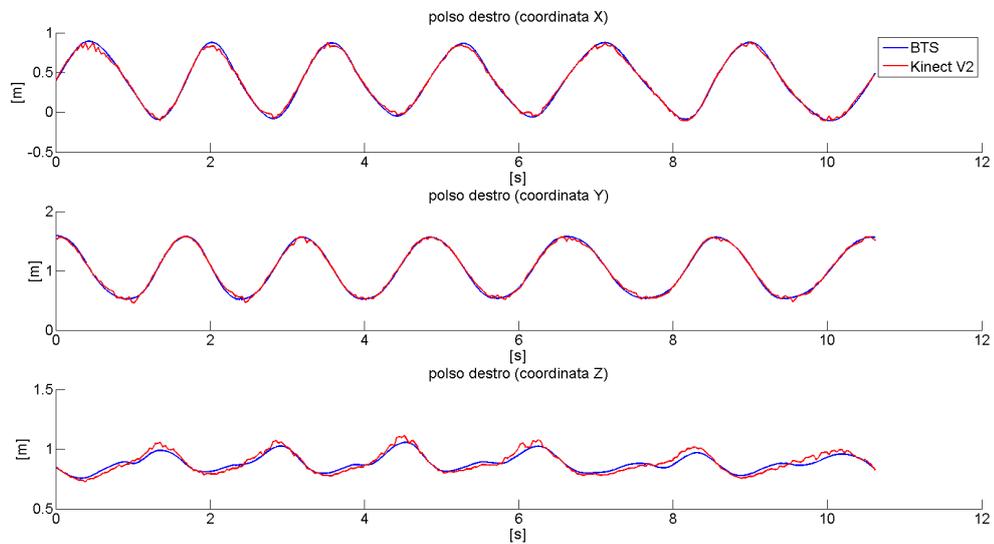


Figura 6.5. Scomposizione e sovrapposizione delle traiettorie relative alle rotazioni compiute dal polso destro, acquisite con sistema BTS e con Kinect V2 (paziente in posizione laterale).

rispetto all'asse Z della camera. Ciò è giustificato dal fatto che, durante il movimento eseguito in quest'ultima posizione, il giunto relativo al polso è sempre visibile da parte della camera, mentre nel caso di posa frontale si possono presentare situazioni in cui l'articolazione tracciata risulti essere occlusa, poiché parzialmente coperta dalla spalla o dal busto del paziente. Si può quindi affermare come la presente prova mostri la necessità di variare e scegliere opportunamente la posa relativa tra paziente e camera, in accordo con il differente esercizio eseguito, così da minimizzare il rischio di avere occlusioni.

Viene ora analizzato il terzo set di coordinate, in cui il paziente assume un orientamento intermedio rispetto a quelli relativi alle due precedenti acquisizioni.

Tabella 6.4. Scostamenti medi tra le traiettorie relative alle rotazioni compiute dal polso destro (paziente in posizione laterale).

<i>Scostamento medio</i>	<i>Valore</i>
$\bar{d}_X [mm]$	17
$\bar{d}_Y [mm]$	18
$\bar{d}_Z [mm]$	23

Anche in questo caso viene applicata ai dati la procedura di allineamento e sincronizzazione utilizzata in precedenza, ottenendo i grafici di figura 6.6 e gli scostamenti medi riportati nella tabella 6.5.

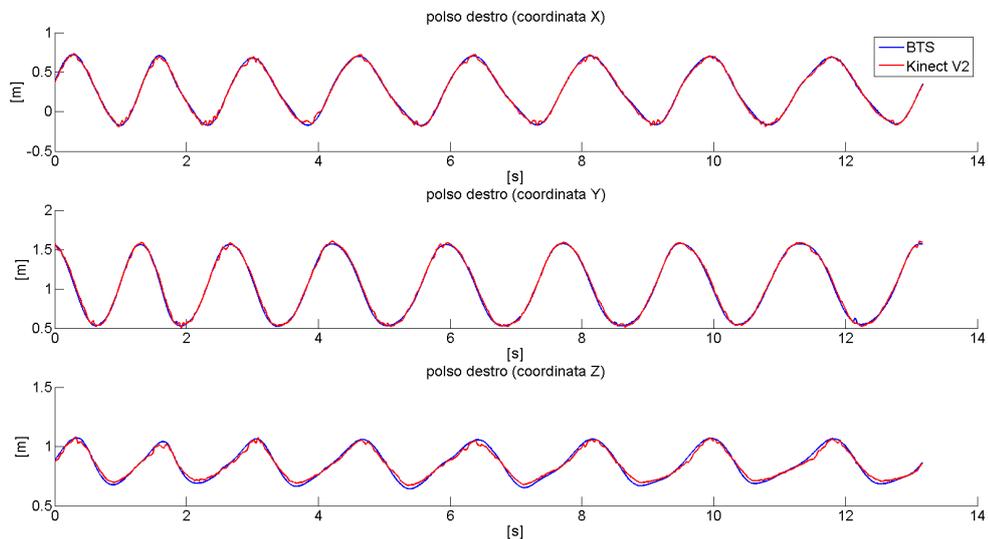


Figura 6.6. Scomposizione e sovrapposizione delle traiettorie relative alle rotazioni compiute dal polso destro, acquisite con sistema BTS e con Kinect V2 (paziente con asse sagittale a 45° rispetto all'asse della tof camera).

Confrontando i risultati di quest'ultima prova, relativi sia alla sovrapposizione delle traiettorie sia agli scostamenti medi calcolati, con quelli ottenuti in precedenza, si nota come la presente posa risulti essere la migliore in termini di tracciamento e ricostruzione della posizione del polso, relativamente al movimento svolto. Inoltre i valori di scostamento rilevati sono ampiamente in linea con le richieste dettate dall'applicazione, brevemente descritta nel capitolo 1.

Tabella 6.5. Scostamenti medi tra le traiettorie relative alle rotazioni compiute dal polso destro (paziente con asse sagittale a 45° rispetto all'asse della tof camera).

<i>Scostamento medio</i>	<i>Valore</i>
$\bar{d}_X [mm]$	15
$\bar{d}_Y [mm]$	16
$\bar{d}_Z [mm]$	17

6.3 Confronto tra Kinect V2 e sistema BTS (protocollo di trattamento neuroriabilitativo)

In ultimo vengono svolti, da parte del paziente, alcuni esercizi basati su un protocollo riabilitativo [5] relativo all'arto superiore, definito ed utilizzato all'interno del Centro di Riabilitazione "Villa Beretta" di Costa Masnaga (LC), il quale prevede, per l'analisi del movimento del singolo arto, l'utilizzo di sole cinque sfere catarifrangenti. Questa semplice soluzione, sebbene non permetta di individuare l'esatta posizione del centro di rotazione delle articolazioni, garantisce un buon grado di somiglianza tra il reale movimento eseguito e i dati acquisiti, avendo come obiettivo primario quello di evitare al paziente di dover sopportare lunghe attese causate dal tempo necessario al posizionamento di numerosi marker.

Inoltre avendo cura di porre le sfere in corrispondenza di parti dell'osso facilmente individuabili (es. sporgenze ben visibili), è possibile ripetere con buona accuratezza tale operazione nelle sedute successive, ottenendo così dati relativi ai medesimi punti, utili al fine di confrontare tra di loro le varie acquisizioni e quantificare i progressi ottenuti.

Il seguente test si divide in tre fasi distinte, durante ciascuna delle quali il paziente esegue, ripetutamente per dieci volte, un diverso movimento costituito da semplici azioni, svolte in posizione seduta; queste sono utili all'analisi delle capacità di compiere gesti caratteristici della vita quotidiana, grazie ai quali è possibile esplorare lo spazio circostante.

Nel dettaglio, la prima fase prevede l'elevazione del braccio teso all'interno del piano frontale, partendo da una posizione in cui l'arto è rilassato e steso lungo il fianco.

Al paziente viene richiesto di vincere la forza di gravità, senza utilizzare alcun dispositivo di compensazione. A seguito delle acquisizioni viene valutata dal clinico la massima ampiezza raggiunta dall'angolo presente

tra l'arto e l'asse longitudinale del corpo, verificando che il paziente non sfrutti la flessione della colonna vertebrale per compensare l'incapacità nell'eseguire il sollevamento oltre una certa quota.

Nella seconda fase viene richiesto di compiere un movimento, chiamato *reaching frontale*, concettualmente molto simile al precedente, con l'unica differenza che in questo caso l'elevazione è svolta nel piano sagittale. In questa situazione vengono valutate non solo la capacità da parte del paziente di vincere la forza di gravità ma anche la fluidità della traiettoria seguita nell'avvicinare la mano ad un bersaglio posto all'altezza della spalla, ad una distanza dal corpo leggermente superiore alla lunghezza totale dell'arto.

Nella terza ed ultima fase, viene chiesto al paziente di indirizzare la mano, inizialmente posta sulla coscia, verso la propria bocca così da raggiungerla e toccarla. È chiaro come quest'ultima prova sia utile a valutare la capacità, da parte della persona, di potersi alimentare autonomamente.

Durante il test, sono acquisiti, sia attraverso Kinect V2 sia attraverso il sistema stereoscopico BTS, i dati relativi alle posizioni delle articolazioni di polso, gomito e spalla destri e delle vertebre cervicale C7 e toracica T5; queste ultime due (corrispondenti nel sistema Kinect V2 ai giunti rispettivamente chiamati *SpineShoulder* e *SpineMid*) sono utili alla definizione della direzione assunta dall'asse longitudinale del corpo.

Le traiettorie restituite da entrambi i sistemi vengono inizialmente allineate e sovrapposte, come fatto in precedenza, ed i risultati ottenuti sono riportati nei grafici di figura 6.7, 6.8 e 6.9 nei quali i punti acquisiti mediante Kinect V2 sono processati tramite l'applicazione di un filtro passa basso del secondo ordine di tipo Butterworth con frequenza di taglio a 5 Hz, così da eliminare parte del rumore di misura.

Un'analisi qualitativa dei risultati ottenuti rivela come, nei grafici relativi ai primi due movimenti (elevazione nel piano frontale e *reaching frontale*), siano presenti coppie di traiettorie, associate alla medesima coordinata, mutuamente ben sovrapponibili ed aventi un andamento confrontabile.

Valutando i grafici derivanti dall'ultimo esercizio, al contrario, si evidenziano alcuni tratti in cui è presente un sistematico scostamento tra le coppie di traiettorie associate alla medesima coordinata, ed in particolar modo tra quelle relative alla componente del movimento eseguita lungo gli assi *X* e *Z*.

La motivazione del presente problema, oltre a quelle precedentemente enunciate relativamente allo scostamento presente tra il baricentro del marker e la reale posizione dell'articolazione, può essere ricercata in una difficoltà, dimostrata da parte del dispositivo Kinect V2, nell'individuare le esatte coordinate relative ad un giunto che si trova a contatto con una parte del corpo umano o nell'immediata prossimità di esso.

6.3. Confronto tra Kinect V2 e sistema BTS (protocollo di trattamento neuroriabilitativo)

Inoltre la posizione relativa tra paziente e camera utilizzata in questa prova, in cui il sensore è posto frontalmente al soggetto, genera acquisizioni nelle quali parte del busto, durante il movimento, risulta non essere presente poiché nascosta dal braccio, rendendo più difficoltoso il processo di individuazione dei giunti svolto dall'algorithm.

È evidente come queste sfavorevoli condizioni, non presenti nelle precedenti acquisizioni, si verifichino spesso in quest'ultimo esercizio.

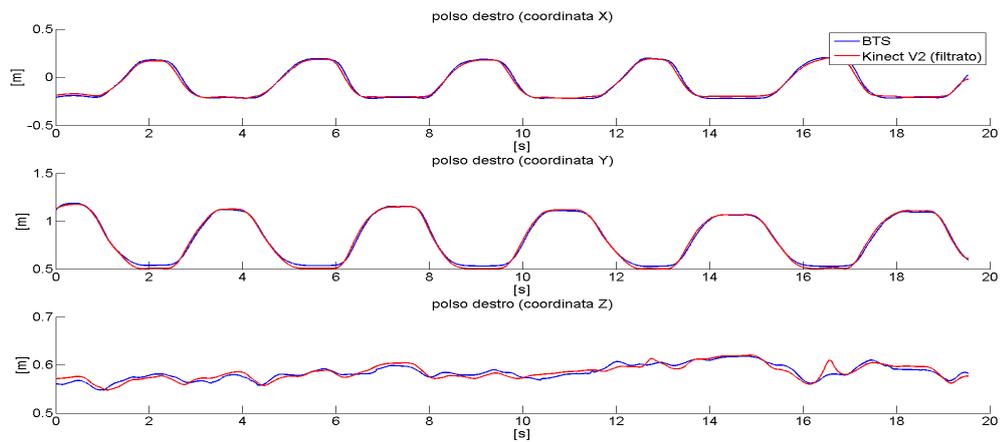


Figura 6.7. Scomposizione e sovrapposizione delle traiettorie relative al polso destro durante l'esercizio di elevazione nel piano frontale, acquisite con sistema BTS e con Kinect V2.

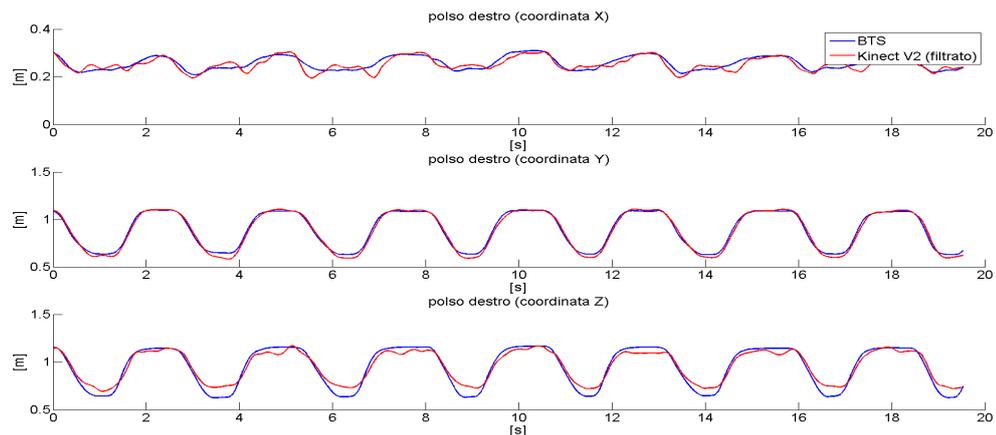


Figura 6.8. Scomposizione e sovrapposizione delle traiettorie relative al polso destro durante l'esercizio di reaching frontale, acquisite con sistema BTS e con Kinect V2.

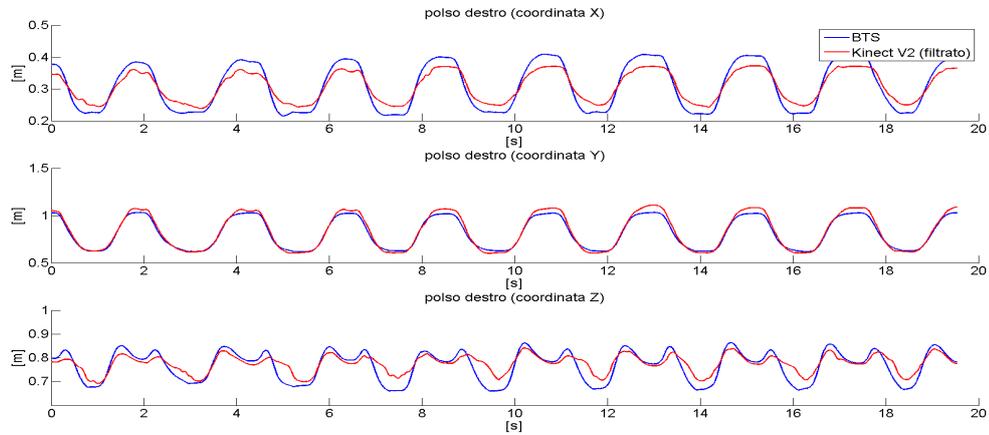


Figura 6.9. Scomposizione e sovrapposizione delle traiettorie relative al polso destro durante l'esercizio di avvicinamento della mano alla bocca, acquisite con sistema BTS e con Kinect V2.

Successivamente viene svolta, attraverso algoritmi sviluppati dal laboratorio di bioingegneria presente all'interno della clinica, l'analisi dell'intero set di dati acquisiti, al fine di estrarre parametri utilizzati dal medico per effettuare la diagnosi. In particolare, suddividendo ciascun movimento (ciclo) in due fasi, "andata" e "ritorno", sono calcolati, relativamente ad ogni fase, alcuni parametri, dei quali sono in seguito riportati la descrizione e i valori mediati sul numero totale di cicli eseguiti.

Nel dettaglio i valori analizzati, relativi sia a ciascuna delle due fasi sia all'intero ciclo, sono il tempo impiegato (espresso in percentuale e in secondi), l'angolo massimo e minimo (espressi in gradi) relativi a diversi segmenti anatomici, scelti in base al differente esercizio svolto, e la velocità di esecuzione del movimento (espressa in gradi al secondo).

I risultati dell'analisi, riassunti nelle tabelle 6.6, 6.7 e 6.8, mostrano una buona corrispondenza tra i valori derivanti dalle acquisizioni relative al dispositivo Kinect V2 e quelli ottenuti con il sistema stereoscopico BTS.

6.3. Confronto tra Kinect V2 e sistema BTS (protocollo di trattamento neuroriabilitativo) 84

Tabella 6.6. Dati risultanti dall'analisi dell'esercizio di elevazione nel piano frontale.

	KinectV2			BTS		
	Ciclo	Andata	Ritorno	Ciclo	Andata	Ritorno
Tempo [%]		49	51		47	53
Tempo [s]	3.53	1.72	1.81	3.50	1.65	1.85
Angolo spalla [°]		100	9		82	13
Velocità [°/s]	51	53	50	40	39	42

Tabella 6.7. Dati risultanti dall'analisi dell'esercizio di reching frontale.

	KinectV2			BTS		
	Ciclo	Andata	Ritorno	Ciclo	Andata	Ritorno
Tempo [%]		51	49		52	48
Tempo [s]	2.57	1.32	1.25	2.56	1.30	1.26
Angolo spalla [°]		103	17		93	8
Velocità [°/s]	67	65	69	66	65	67

Tabella 6.8. Dati risultanti dall'analisi dell'esercizio di avvicinamento della mano alla bocca.

	KinectV2			BTS		
	Ciclo	Andata	Ritorno	Ciclo	Andata	Ritorno
Tempo [%]		54	46		52	48
Tempo [s]	4.85	2.60	2.25	5.44	2.82	2.63
Angolo gomito [°]		133	62		137	77
Velocità [°/s]	30	71	84	22	54	58

Infine viene mostrato, nell'immagine 6.10, il report completo emesso dalla clinica al termine dell'esercizio di avvicinamento della mano destra alla bocca. Si noti, nella tabella, la presenza di alcuni parametri aggiuntivi, rispetto a quelli precedentemente riportati, quali l'indice di fluidità del movimento (NAJ), decrescente al diminuire dell'ampiezza degli "scatti" o degli "strappi" rilevati nell'esecuzione dell'azione, l'indice di ripetibilità del movimento (CPA), che descrive il grado di somiglianza tra i differenti cicli di movimento eseguiti nell'avvicinare la mano alla bocca, e l'indice

6.3. Confronto tra Kinect V2 e sistema BTS (protocollo di trattamento neuroriabilitativo) 85

di rilevazione di strategie compensatorie di spalla (Comp) utilizzate dal paziente per aggirare le proprie limitazioni motorie.

Essi, in aggiunta a quelli precedentemente elencati e relativi ai tempi di esecuzione, alle velocità e agli angoli raggiunti, sono di particolare interesse biomedico ed utili alla stima e valutazione del progresso mostrato dal paziente. Analizzando la seguente tabella è possibile verificare come, anche tra questi ultimi termini, sia presente un sufficiente grado di somiglianza tra i valori associati al medesimo parametro.

Congregazione Suore Infermiere dell'Addolorata
Ospedale Generale di Zona "VALDUCE"
Centro di Riabilitazione "VILLA BERETTA"
 Costamasnaga (LC)
 Tel: 031/8544211 - Fax: 031/8544347
 Centro di Bioingegneria
 Laboratorio di Analisi del Movimento

Costamasnaga, 11/05/2015

Cognome e Nome: prova kinect bts
 Data di Nascita:
 Patologia:
 Note:

MANO ALLA BOCCA (Arto destro)

	Kinect V2						BTS					
	Ciclo		Andata		Ritorno		Ciclo		Andata		Ritorno	
	Mean	Std	Mean	Std	Mean	Std	Mean	Std	Mean	Std	Mean	Std
Time [%]			54	6	46	6			52	5	48	5
Time [s]	4.85	0.69 (0.97-2.51)	2.60	0.41 (0.50-1.14)	2.25	0.51 (0.47-1.37)	5.44	0.34 (0.97-2.51)	2.82	0.35 (0.50-1.14)	2.63	0.34 (0.47-1.37)
Elbow [°]			133	2	62	3			137	1	77	2
Comp [-]	-0.09	0.05 (0.00-0.22)					0.02	0.01 (0.00-0.22)				
Elbow Vel [°/s]	30	3 (41-103)	71	11 (38-99)	84	11 (42-107)	22	1 (41-103)	54	6 (38-99)	58	6 (42-107)
CPA [-]	0.57	(0.76-0.98)	0.56	(0.83-0.99)	0.86	(0.73-0.99)	0.70	(0.76-0.98)	0.75	(0.83-0.99)	0.94	(0.73-0.99)
NAJ [-]	27	0 (9-38)	28	13 (8-36)	26	13 (9-48)	36	0 (9-38)	34	21 (8-36)	37	8 (9-48)

NOTE:
 Comp: indice di rilevazione di strategie compensatorie di spalla
 CPA: indice compreso tra 0 e 1 crescente con il livello di ripetibilità de
 NAJ: indice di fluidità, decrescente al crescere della fluidità

Figura 6.10. Report relativo allo svolgimento dell'esercizio di avvicinamento della mano alla bocca.

6.4 Conclusioni

In questa sezione è stata presentata una serie di test mediante i quali si sono analizzati i dati, restituiti dal Kinect V2, relativi al tracciamento dei giunti del corpo umano, e in particolare a quello associato al polso destro.

Dalla prima prova è emerso come la deviazione standard, relativa a misure statiche, assuma valori di poco superiori al millimetro.

La seconda prova ha mostrato come i movimenti eseguiti dal paziente vengano ben tracciati e siano ben sovrapponibili alle traiettorie di riferimento misurate dal sistema BTS, presentando uno scostamento medio inferiore a 18 mm, nel caso di paziente con asse sagittale posto a 45° rispetto alla normale uscente dalla camera. Ciò dimostra come sia di fondamentale importanza posizionare il dispositivo Kinect V2 in maniera opportuna e tale da evitare che, durante il movimento, i giunti di interesse risultino occlusi anche solo per un breve lasso di tempo.

Nella terza prova è stata svolta un'analisi dei dati acquisiti mediante appositi algoritmi sviluppati all'interno del Centro di Riabilitazione "Villa Beretta" di Costa Masnaga (LC). Da questo test è emerso come, a valle di un filtraggio a 5 Hz, vi sia una buona somiglianza, con scostamenti inferiori alle massime tolleranze ammissibili, tra i parametri biomedici estratti da ciascuna coppia di storie temporali relativa al medesimo esercizio.

Ciò fornisce un'ulteriore conferma di come il sistema a tempo di volo, analizzato nel presente lavoro, sia adeguato per l'applicazione in esame. Le leggere differenze, in termini di angoli raggiunti e di velocità angolari medie associate all'esecuzione del movimento, possono essere attribuite all'impossibilità di avere un'esatta corrispondenza tra i punti tracciati da ambedue i dispositivi, problematica principalmente dovuta alla non conoscenza della reale posizione dei giunti stimati da Kinect V2. Infatti è bene precisare come ciò obblighi a svolgere solamente prove di tipo qualitativo e non permetta di ottenere dati tali da poter essere oggettivamente confrontati, ma solamente sovrapposti ed allineati al fine di ricavare un ordine di grandezza indicante la verosimiglianza tra il movimento eseguito e quello tracciato.

Capitolo 7

Applicazione neuroriabilitativa

In questa sezione viene presentato il caso applicativo per il quale sono state svolte la ricerca del dispositivo di ricostruzione 3D più adatto e l'analisi metrologica riportate nel presente lavoro. In particolare il prodotto in seguito descritto, sviluppato dall'azienda MediaHospital S.r.l., avvalendosi della capacità del Kinect V2 di identificare una figura umana all'interno di una scena e di tracciarne la posizione dei giunti del corpo, si rivela essere un ottimo dispositivo, in quanto a rapporto costo-efficienza, in grado di fornire ausilio alla neuroriabilitazione motoria e cognitiva svolta sia in ambiente clinico sia in ambiente domestico.

Il sistema in seguito descritto, sviluppato da MediaHospital S.r.l. in collaborazione con il Presidio dell'Ospedale Valduce "Villa Beretta" di Costa Masnaga (LC), ha come principale caratteristica quella di essere un dispositivo di semplice utilizzo, in grado di fornire ausilio, in ambiente sia clinico sia domestico, nel processo di riabilitazione svolto in presenza di patologie di natura neurologica causanti disordine del movimento.

L'hardware minimo richiesto è infatti composto da un Kinect V2 for Windows e da un personal computer al quale può essere collegato uno schermo di maggiori dimensioni (es. televisore LCD) o un videoproiettore. Attraverso questi ultimi viene mostrata una realtà virtuale, in cui sono presenti semplici oggetti tridimensionali, i quali possono essere afferrati, spostati e rilasciati da un avatar movimentato dal paziente tramite i gesti della propria mano, tracciati dal sistema Kinect V2. L'esercizio può essere

svolto sia in posizione seduta sia in posizione eretta.

Un grande punto di forza della presente applicazione, che la differenzia dalla maggior parte dei videogames in cui sono proposti esercizi motori e di fitness, è quello di poter fornire in output una serie di dati clinici, in parte acquisiti attraverso sensori esterni, utili a calibrare il livello di difficoltà, accordandolo in relazione alle capacità del singolo paziente, così da inseguirne i progressivi miglioramenti. Viene quindi generato un database in cui sono salvati i parametri monitorati ed i risultati ottenuti, così da poter stilare un quadro, consultabile dal medico, ed utile al fine di emettere una diagnosi relativa all'evoluzione della condizione clinica del paziente.

Inoltre questo tipo di esercizio permette di eseguire contemporaneamente attività riabilitativa sia di tipo motorio sia di tipo cognitivo. Esistono infatti casi in cui il paziente, seppur in grado di svolgere semplici movimenti, interrompe l'azione nel caso in cui gli venga chiesto di parlare o di rispondere ad una domanda. Un tempo i due tipi di riabilitazione venivano svolti in sedi differenti e da personale clinico specializzato nell'uno o nell'altro settore; recenti studi [35] hanno invece dimostrato che è possibile, grazie a specifici esercizi, avere un'evoluzione parallela dei miglioramenti relativi ad entrambe le patologie.

È quindi buona norma mantenere alta la concentrazione del paziente durante l'esecuzione dell'attività ed assicurarsi che quest'ultimo svolga l'esercizio fisico in maniera attiva, prestando attenzione ai movimenti compiuti. Per queste ragioni, un ulteriore sviluppo della presente tecnologia, volto ad aumentare il realismo delle azioni compiute, consiste nella possibilità di utilizzare, in alternativa ad un comune schermo televisivo, il visore Oculus Rift [45] il quale, indossato dal paziente, lo proietta all'interno di

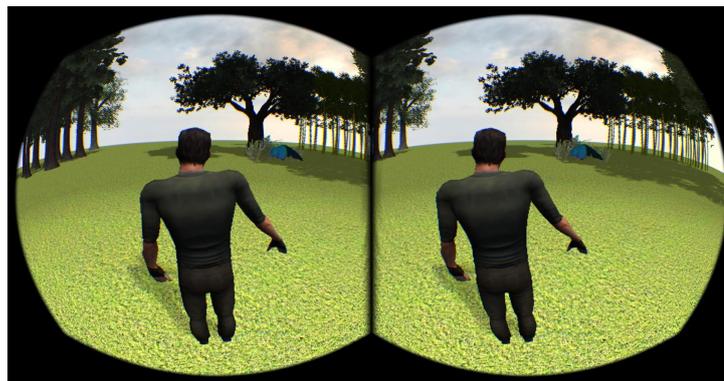


Figura 7.1. Esempio di esercizio presente nell'applicazione sviluppata da MediaHospital S.r.l., svolto con l'utilizzo del visore Oculus Rift.

una realtà tridimensionale altamente coinvolgente.

Nella figura 7.1 è riportato uno screenshot dell'immagine stereoscopica mostrata all'interno del visore durante un esercizio in cui il paziente, immerso nella realtà virtuale, deve avvicinare la propria mano ad alcune farfalle che compaiono in punti casuali della scena.

È chiaro come questa applicazione possa disporre di una variabilità potenzialmente infinita di esercizi e livelli di difficoltà, così da potersi adattare al meglio non solo alle diverse patologie e problematiche del paziente ma anche alle differenti attitudini del singolo individuo. Ad esempio è possibile seguire il movimento di una sola mano o di entrambe, variare l'ambientazione virtuale ed il numero di oggetti presenti, sceglierne la tipologia, la dimensione e la disposizione all'interno della scena.

Grazie alla propria flessibilità, l'applicazione è di grande aiuto ad esempio in presenza di patologie, caratterizzate da negligenza spaziale unilaterale (NSU), nelle quali, in seguito ad un trauma localizzato in un emisfero cerebrale, è resa difficoltosa l'esplorazione del semispazio opposto all'area di cervello danneggiata. In casi come questo è possibile chiedere al software di concentrare il numero di oggetti generati in una specifica porzione di schermo, così da spingere l'attenzione del paziente nella parte di spazio visivo in cui risulta più difficile la ricerca e stimolare il recupero funzionale.

Inoltre, grazie all'elevata variabilità di oggetti e scenari, si ha la possibilità di mostrare realtà virtuali legate alla vita trascorsa dal paziente e alle sue passioni (figura 7.2), così da coinvolgerlo maggiormente attirandone l'attenzione ed ottenendo, come spiegato in precedenza, maggiori benefici.



Figura 7.2. Esempio di realtà virtuale, presente nell'applicazione sviluppata da MediaHospital S.r.l., in cui sono ricostruiti alcuni oggetti della vita quotidiana.

Questi sono solo alcuni degli esempi attraverso i quali è possibile mettere in luce come la natura digitale del dispositivo permetta di variare e personalizzare la tipologia di esercizio riabilitativo in modo molto più semplice ed immediato rispetto a ciò che si potrebbe fare con oggetti fisici e metodiche tradizionali.

Nella figura 7.3 viene mostrata una sequenza fotografica in cui sono presenti tre fasi relative allo svolgimento di uno degli esercizi proposti dall'applicazione. Nel dettaglio si vede come sia ricostruita una classica scena di vita quotidiana in cui alcuni oggetti, presenti in un armadio, devono essere riposti, uno ad uno, in una valigia.

Il paziente, movimentando la mano virtuale presente sullo schermo, la avvicina all'oggetto richiesto dal medico (immagine 7.3(a)) così da afferrarlo e trascinarlo verso la valigia (immagine 7.3(b)) nella quale lo rilascia al termine dell'esercizio (immagine 7.3(c)). L'applicazione conferma con un simbolo di spunta l'esatta esecuzione del compito.

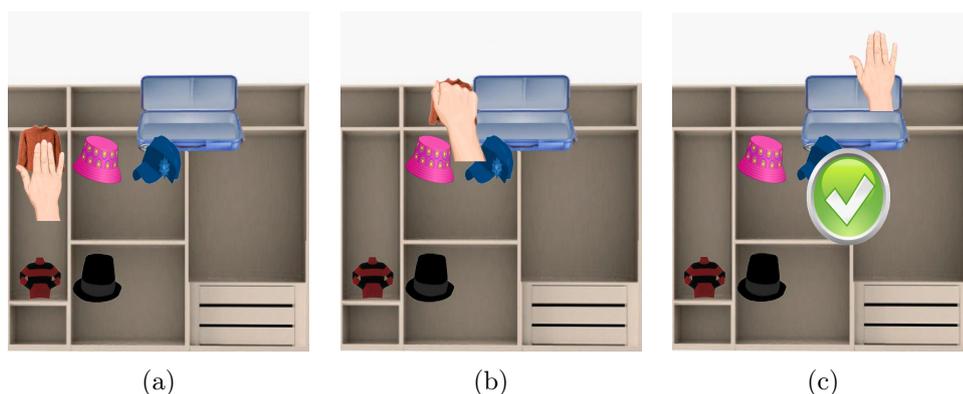


Figura 7.3. Sequenza fotografica relativa allo svolgimento di un esercizio presente nell'applicazione sviluppata da MediaHospital S.r.l.

Capitolo 8

Conclusioni

In questa tesi è stata effettuata la qualificazione metrologica di un dispositivo in grado di tracciare, in real-time, la posizione nello spazio dei principali segmenti anatomici e articolazioni del corpo umano, così da valutarne un possibile impiego in un'applicazione biomedica utilizzata nell'ambito della neuroriabilitazione motoria e cognitiva, condotta sia in ambiente ospedaliero sia domestico.

A seguito di un'approfondita ricerca iniziale, durante la quale sono stati vagliati differenti dispositivi e diverse tecnologie attualmente disponibili sul mercato ed in grado di ricostruire scene e corpi tridimensionali, in accordo con le specifiche richieste dall'applicazione e delineate nel capitolo 1, la scelta è ricaduta su un dispositivo di nuova generazione, prodotto da Microsoft e chiamato Kinect V2. Esso integra una telecamera a tempo di volo, in grado di ricostruire la nuvola di punti 3D relativa alla porzione di scena inquadrata, alla frequenza di 30 Hz, con una risoluzione di 512 x 424 pixel. Inoltre, mediante opportuni algoritmi, il dispositivo ha la possibilità di rilevare la presenza di una o più figure umane all'interno del proprio campo di vista e di tracciare il movimento di alcuni punti notevoli del corpo senza la necessità di indossare una particolare tipologia di abbigliamento o di impugnare specifici controller o altri dispositivi.

Inoltre l'oggetto, posizionandosi all'interno della fascia consumer (prezzo di vendita inferiore a 200 \$ [41]) si mantiene in linea con le specifiche dettate dall'applicazione, richiedendo, per il funzionamento, una connessione USB 3.0 ad un comune personal computer e presentando ridotte dimensioni che ne semplificano il trasporto e l'installazione.

A valle della scelta riguardante l'hardware, è stata svolta un'analisi metrologica al fine di determinare la qualità dei dati restituiti e stabilire i limiti di accuratezza e robustezza delle misure entro i quali è possibile ipotizzare un impiego del dispositivo in ambito scientifico e misuristico.

Al termine della prima parte del lavoro, focalizzata principalmente sul livello distanziometrico dell'analisi metrologica, è emerso come sia possibile eseguire misure senza contatto di distanza, dimensione e volume nonché ricostruzioni tridimensionali di interi oggetti all'interno di una scena ampia $70 \times 60^\circ$ avente un campo di vista che si estende da 0.5 a 4.5 m. I risultati ottenuti, riportati al termine del capitolo 5 e qui brevemente riassunti, riguardano una serie di prove interamente svolte con l'ausilio di una ventola fissata esternamente ed indispensabile al mantenimento di una temperatura pressoché stabile all'interno al dispositivo. È stato infatti verificato come, affidandosi semplicemente al sistema di raffreddamento integrato nel dispositivo, si ottengono sequenze di valori di profondità, acquisiti a scena fissa, affette da un'evidente deriva.

Dall'analisi del pixel centrale è emerso che l'incertezza di misura, dovuta all'effetto casuale, cresce con l'aumentare della distanza tra sensore e target fino a raggiungere il valore di 3.3 mm in prossimità dell'estremo superiore del campo di operatività (valore medio delle misure pari a 4200 mm). Inoltre, le acquisizioni relative all'intero sensore rivelano un'incertezza di misura del singolo pixel proporzionale alla coordinata radiale dell'immagine e dovuta ad una scarsa illuminazione dei bordi della scena; quest'ultima problematica, in presenza di misure effettuate a distanze superiori a 1500 mm, si traduce nella non affidabilità delle coordinate 3D relative ai punti associati ai pixel prossimi ai vertici del sensore. Dall'analisi di misure effettuate a distanze note, emerge la presenza di un effetto sistematico caratterizzato da un andamento armonico, confinato in una banda che si estende da -10 a 15 mm.

I test eseguiti con bersaglio non parallelo al sensore mostrano una correlazione tra angolo di riflessione del cono di illuminazione e valore di deviazione standard relativo a ciascun pixel.

L'utilizzo di bersagli posti alla medesima distanza e realizzati con materiali e colori differenti ha evidenziato come il diverso grado di riflessività delle superfici, nella banda relativa alla luce IR emessa dal dispositivo, causi un offset il quale però, mantenendosi sempre inferiore al millimetro, può, nella maggior parte delle applicazioni, essere trascurato; è stata inoltre riscontrata l'impossibilità di effettuare misure di distanza relative ad oggetti aventi superfici estremamente riflettenti (es. materiale catarifrangente).

Le nuvole ottenute dall'acquisizione di geometrie note risultano essere composte da punti aventi uno scostamento, misurato rispetto ad un modello di riferimento, molto spesso inferiore a venti millimetri.

È stato inoltre verificato che la problematica di mixed-pixel è presente solo in acquisizioni in cui lo sfondo della scena, appartenente al campo di misura, viene acquisito insieme all'oggetto di interesse.

Infine, in acquisizioni relative a particolari geometrie, è presente la nota limitazione relativa alla tecnologia a tempo di volo, derivante dalle multiple riflessioni del cono di luce IR, la quale, nelle prove, ha generato scostamenti, tra acquisizione e modello, aventi grandezze dell'ordine di alcune decine di millimetri.

La seconda fase di test, caratterizzata da prove di biomediche, ha permesso di valutare la rumorosità della misura e l'accuratezza del dispositivo nell'identificare la posizione assunta dall'arto superiore. Le analisi sono state svolte acquisendo contemporaneamente, con Kinect V2 e sistema BTS, le coordinate 3D di alcuni particolari punti relativi all'arto destro. Ciò ha permesso di confrontare i dati restituiti dal dispositivo a tempo di volo con le traiettorie generate dal sistema stereoscopico.

Dai risultati, riportati nel capitolo 6, è emerso come la deviazione standard, relativa alla misura statica della posizione di un giunto, assuma valori di poco superiori al millimetro.

Inoltre, nel caso di dispositivo Kinect V2 posizionato in maniera opportuna, e tale da evitare l'occlusione dei giunti di interesse, è stata ottenuta una corretta tracciatura dei movimenti eseguiti dal paziente, la quale presenta scostamenti medi, misurati rispetto alle traiettorie di riferimento, aventi valori inferiori a 18 mm.

Infine sono stati analizzati, mediante appositi algoritmi, alcuni particolari esercizi relativi ad un protocollo neuroriabilitativo. Per ciascun movimento, i parametri restituiti dall'elaborazione delle acquisizioni effettuate con il dispositivo Kinect V2 mostrano valori confrontabili a quelli ottenuti in parallelo dal sistema BTS ed aventi scostamenti inferiori alle massime tolleranze ammissibili.

Alla luce di ciò si può definire il Kinect V2 idoneo, come dispositivo markerless di tracciamento del movimento, in ambiti in cui non sono richiesti valori di incertezza inferiori a venti millimetri. Per queste ragioni il presente sensore si rivela adatto all'applicazione descritta nel capitolo 7 poiché in linea con le specifiche, sia di accuratezza sia di costo e di semplicità di impiego, richieste.

8.1 Sviluppi futuri

Per quanto riguarda gli sviluppi futuri, sono attualmente in corso, all'interno del VBLab, alcune prove relative alla possibilità di utilizzare contemporaneamente due o più dispositivi, opportunamente calibrati e registrati, così da poter acquisire la scena da due o più punti di vista diversi. In questo modo sarà possibile scansionare volumi di lavoro ed oggetti di

dimensioni superiori a quelle contenute nel cono di visione del singolo dispositivo e, nel caso di utilizzo in ambito biomedicale e di analisi del movimento, si avrà la garanzia di una notevole riduzione delle occlusioni, a vantaggio di una migliore e più robusta ricostruzione della posizione del corpo nello spazio.

Inoltre grazie alla calibrazione della camera RGB, della quale è ricavata la rototraslazione rispetto al sensore IR, viene garantita la possibilità di sovrapporre alla nuvola di punti 3D le informazioni di colore utili sia al riconoscimento di particolari oggetti nella scena sia come appoggio per l'allineamento di due o più scansioni.

Un'applicazione, il cui sviluppo futuro è legato alla possibilità di estendere il volume di lavoro grazie alla connessione di due o più Kinect V2 al medesimo PC, ha come obiettivo quello di monitorare, in ambiente chiuso, il cammino di un individuo avente difficoltà nella deambulazione, così da poterne rilevare un'eventuale caduta e inviare una richiesta di soccorso.

Un setup base può prevedere l'installazione di almeno quattro Kinect V2 posizionati ai vertici alti di una stanza (ad esempio camera di ospedale).

Alcuni articoli presenti in letteratura [17, 27] descrivono una possibile soluzione basata sull'utilizzo della precedente versione di Kinect.

Queste applicazioni, analizzando il movimento del corpo umano ne rilevano la caduta attraverso l'utilizzo di due algoritmi i quali, conoscendo la posizione spaziale di 5 specifici giunti (Head, ShoulderCenter, HipCenter, AnkleRight, AnkleLeft) ne calcolano, istante per istante, sia la distanza rispetto alla superficie su cui si sta muovendo il paziente (individuata da Kinect all'interno della nuvola di punti e modellizzata mediante un piano di best fitting) sia la velocità lungo la direzione determinata dal vettore normale al piano. In particolare, definendo con a, b, c, d i coefficienti del piano rappresentato dall'equazione $y = ax + by + cz + d$ e con x_i, y_i, z_i le coordinate dell' i -esimo giunto, la distanza d_i tra giunto e piano è pari a

$$d_i = \frac{ax_i + by_i + cz_i + d}{\sqrt{a^2 + b^2 + c^2}} \quad (8.1)$$

Per quanto riguarda la velocità $v_{i,k}$ in m/s (relativa al giunto i e calcolata all'istante k) essa viene ricavata applicando la seguente relazione:

$$v_{i,k} = \frac{d_{i,k} - d_{i,k-1}}{t_k - t_{k-1}} \quad (8.2)$$

in cui $d_{i,k}$ e $d_{i,k-1}$ sono le distanze in metri dell' i -esimo giunto calcolate rispettivamente all'istante k e $k - 1$ mentre t_k e t_{k-1} sono i timestamp, convertiti in secondi, relativi agli istanti k e $k - 1$.

Nel caso in cui i valori di distanza (di tutti i giunti) e di velocità calcolati scendano al di sotto di specifiche soglie preimpostate, il software ipotizza una probabile caduta ed invia una richiesta di soccorso.

Applicando i medesimi algoritmi ai dati acquisiti in real-time da Kinect V2 sono state effettuate alcune preliminari prove utilizzando il software mostrato in figura 8.1 che nella versione attuale gestisce un singolo dispositivo il quale garantisce prestazioni decisamente superiori rispetto alla precedente versione [7], come mostrato nel presente lavoro di tesi.

Nell'immagine 8.1(a) si nota come l'individuo, in posizione eretta, viene definito dal software "Safe" mentre nell'immagine 8.1(b) viene mostrato come il software segnala la caduta rilevata col messaggio "Fall".

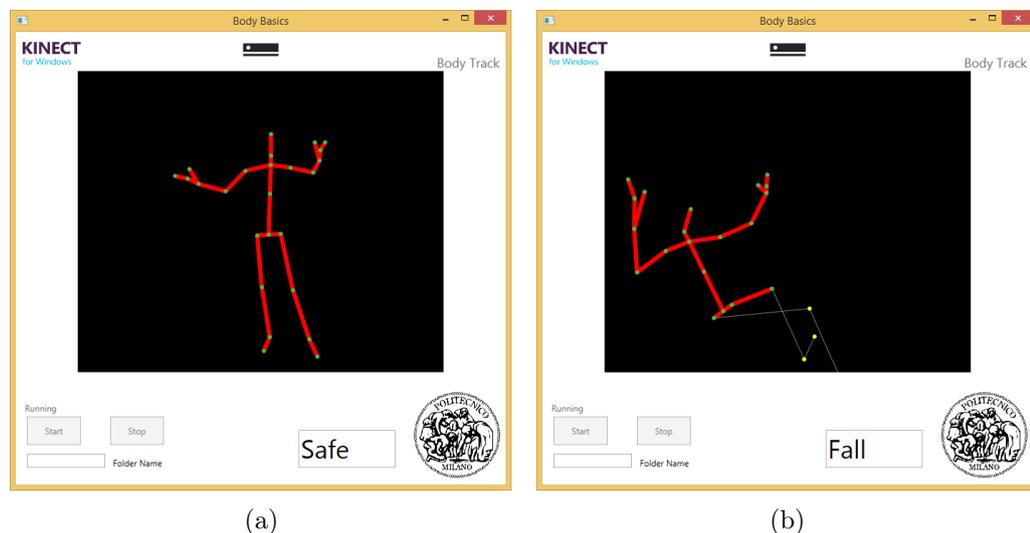


Figura 8.1. Interfaccia dell'applicazione di analisi del cammino in cui è mostrato il paziente in posizione eretta (a) e rilevamento di una caduta (b).

Inoltre è possibile salvare la storia temporale relativa alla posizione spaziale di ciascun giunto così da poter analizzare varie problematiche e disturbi della deambulazione tra cui un progressivo rallentamento della velocità di marcia, una deviazione della fluidità, della simmetria e della sincronia dei movimenti corporei.

Relativamente ad un diverso campo di impiego, non legato al settore biomedico, un ulteriore sviluppo futuro, condotto all'interno del VBLab riguarda la realizzazione di un'applicazione che, sfruttando la possibilità di unire nuvole di punti provenienti da due o più dispositivi, opportunamente registrati, acquisisce sia la forma sia il colore di un oggetto 3D, così da poter

scansionare elementi architettonici e strutture civili di notevoli dimensioni e ricostruirne il modello virtuale tridimensionale.

Nell'immagine 8.2 viene mostrato l'allineamento relativo a due nuvole di punti acquisite mediante due dispositivi Kinect V2.

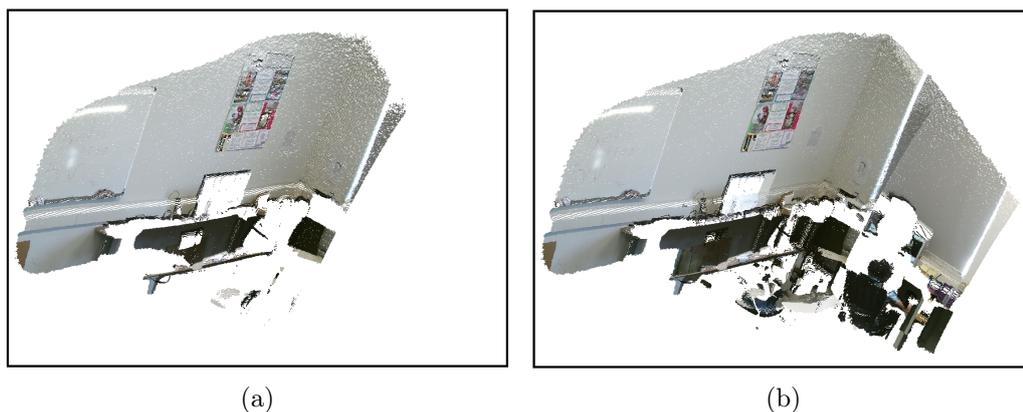


Figura 8.2. Singola nuvola di punti acquisita (a) ed estensione del campo visivo grazie all'unione di due nuvole di punti (b).

Appendice A

Registrazione rigida

La presente appendice descrive i due più comuni algoritmi in grado di determinare la registrazione rigida tra due insiemi di punti che permette di portarli a coincidere o, nell'impossibilità di farlo, di minimizzarne le distanze. Ciò è necessario, come visto più volte nel presente lavoro, al fine di riportare nel medesimo sistema di riferimento due set di punti aventi coordinate 3D espresse mediante riferimenti differenti. Nella prima parte del capitolo è riportata una trattazione matematica attraverso la quale si dimostra come l'applicazione dell'algoritmo di Singular Value Decomposition (SVD) [2] risolva il problema di minimizzazione delle distanze, a patto che ciascun set contenga lo stesso numero di punti e che il loro mutuo accoppiamento sia noto a priori. Successivamente, introducendo l'algoritmo di Iterative Closest Point (ICP) [36], viene mostrato come quest'ultimo, estendendo il precedente, permetta l'esecuzione della registrazione rigida anche nel caso di insiemi di punti non accoppiati ed aventi dimensioni arbitrarie.

A.1 Singular Value Decomposition

Si supponga di disporre di due insiemi di coordinate 3D relative ai medesimi punti acquisiti con due differenti dispositivi e quindi espressi

in due sistemi di riferimento (SDR) diversi. Nella seguente trattazione il primo insieme sarà chiamato X mentre il secondo Y .

Il problema da risolvere consiste nel trovare la trasformazione rigida 3D (composta da matrice di rotazione e vettore di traslazione) la quale, applicata ai punti dell'insieme Y permetta di farli coincidere coi corrispondenti appartenenti ad X [33].

Matematicamente, considerati N punti, si può scrivere

$$X_i = RY_i + t \quad \text{con } i = 1 \dots N \quad (\text{A.1})$$

dove R è una matrice di rotazione avente dimensione 3×3 , t è un vettore di traslazione di dimensione 3×1 e, per ogni valore di i , le coordinate X_i ed Y_i si riferiscono al medesimo punto espresso nei due differenti SDR.

In presenza di rumore, l'obiettivo della registrazione è quindi quello di trovare R e t tali per cui sia minimizzata la distanza euclidea E presente tra ogni punto e la rispettiva controparte rototraslata e definita come

$$E = \sum_{i=1}^N \| X_i - (RY_i + t) \|^2 \quad (\text{A.2})$$

Al fine di disaccoppiare le variabili del problema viene inizialmente ricavata un'espressione del vettore t la quale, inserita nella A.2 la renderà indipendente dal parametro di traslazione.

Assumendo R costante si può scrivere la seguente funzione del parametro t :

$$F(t) = \sum_{i=1}^N \| X_i - (RY_i + t) \|^2 \quad (\text{A.3})$$

Il valore ottimo del vettore di traslazione t si ottiene dalle radici della derivata di $F(t)$ calcolata rispetto a t .

$$0 = \frac{\partial F}{\partial t} = \sum_{i=1}^N 2(X_i - RY_i + t) = 2 \sum_{i=1}^N X_i - 2R \sum_{i=1}^N Y_i + 2Nt \quad (\text{A.4})$$

Risolvendo la A.4 rispetto a t si ottiene

$$t = \frac{\sum_{i=1}^N X_i}{N} - R \frac{\sum_{i=1}^N Y_i}{N} \quad (\text{A.5})$$

e definendo i termini

$$X_b = \frac{1}{N} \sum_{i=1}^N X_i$$

e

$$Y_b = \frac{1}{N} \sum_{i=1}^N Y_i$$
(A.6)

come baricentri (o centroidi) dei rispettivi set di dati, la A.5 può essere riscritta nella seguente forma:

$$t = X_b - RY_b$$
(A.7)

Sottraendo ad ogni insieme di dati la quota relativa al corrispondente baricentro si ottengono le quantità

$$\bar{X}_i = X_i - X_b$$

e

$$\bar{Y}_i = Y_i - Y_b$$
(A.8)

rappresentanti le coordinate dei punti, appartenenti agli insiemi X e Y , riferite ad un SDR posto nel corrispettivo centroide.

Sostituendo la A.7 e la A.8 nella A.2, la cifra da minimizzare diventa quindi

$$E = \sum_{i=1}^N \| \bar{X}_i - R\bar{Y}_i \|^2$$
(A.9)

A questo punto per ricavare la matrice di rotazione R si può scrivere la A.9 come

$$E = \sum_{i=1}^N (\bar{X}_i - R\bar{Y}_i)^T (\bar{X}_i - R\bar{Y}_i)$$
(A.10)

che sviluppata diventa

$$E = \sum_{i=1}^N (\bar{X}_i^T \bar{X}_i + \bar{Y}_i^T \bar{Y}_i - 2\bar{X}_i^T R\bar{Y}_i)$$
(A.11)

Dalla A.11 si nota chiaramente che minimizzare E significa massimizzare una quantità chiamata E' , definita come

$$E' = \sum_{i=1}^N \bar{X}_i^T R \bar{Y}_i = Tr(X^T R Y) \quad (\text{A.12})$$

in cui le matrici X e Y sono composte rispettivamente dall'accostamento per colonne dei vettori \bar{X}_i e \bar{Y}_i .

Ricordando che l'obiettivo è quello di calcolare una matrice R tale per cui sia massima la quantità $Tr(Y^T R X)$ applicando la proprietà per cui

$$Tr(AB) = Tr(BA) \quad (\text{A.13})$$

si ottiene

$$Tr(Y^T R X) = Tr(R X Y^T) = Tr(R H) \quad \text{con} \quad H = X Y^T \quad (\text{A.14})$$

detta matrice di covarianza e relativa ai due set di dati.

Utilizzando un procedimento basato su una scomposizione ai valori singolari (SVD) applicato alla matrice H si ottiene

$$H = U D V^T \quad (\text{A.15})$$

Quindi, inserendo la A.15 nella A.14 si ottiene

$$Tr(R H) = Tr(R U D V^T) = Tr(D V^T R U) = Tr(D M) \quad (\text{A.16})$$

con $M = V^T R U$ matrice ortogonale e cioè avente colonne composte da vettori ortonormali ($|m_{ij}| < 1$). Inoltre, come noto, la matrice D è diagonale con valori $\sigma_i \geq 0$.

Da ciò deriva che la traccia ($Tr(D M)$) è massima se $m_{ii} = 1$ e cioè se $M = I$.

Con semplici passaggi algebrici si ottiene

$$I = M = V^T R U \rightarrow V = R U \rightarrow R = V U^T \quad (\text{A.17})$$

Il processo finora descritto permette di trovare la matrice ortogonale ottima R , la quale può contenere, oltre alla rotazione, anche un contributo indesiderato di riflessione.

Per verificare ciò è sufficiente calcolarne il determinante.

Se $\det(V U^T) = -1$ si è in presenza di una matrice R contenente una riflessione mentre se $\det(V U^T) = 1$ non si ha riflessione.

Quindi la formulazione generale che comprende entrambi i casi può essere espressa nel modo seguente:

$$R = V \begin{pmatrix} 1 & & & \\ & 1 & & \\ & & \ddots & \\ & & & 1 \\ & & & & \det(VU^T) \end{pmatrix} U^T \quad (\text{A.18})$$

Ora che è stata calcolata la matrice R la si inserisce nella A.7 ottenendo il vettore t .

A.2 Iterative Closest Point

Una seconda tipologia di registrazione rigida è quella basata sugli algoritmi di Iterative Closest Point (ICP). Essi, essendo un'estensione di quelli precedentemente descritti (sezione A.1), risultano applicabili anche ad insiemi di punti non accoppiati ed aventi dimensioni differenti. È chiaro come, non potendo in questa situazione avere un'esatta sovrapposizione tra le coppie di punti, l'obiettivo che ci si pone è semplicemente quello di minimizzare le distanze.

Si supponga di avere due set di punti (chiamati anche nuvole di punti) descritti per mezzo di coordinate 3D. Il primo dei due sarà chiamato riferimento o modello mentre il secondo sorgente. L'obiettivo del presente algoritmo è quello di trovare la matrice di rototraslazione tale per cui, applicata alla nuvola sorgente, la faccia corrispondere al meglio a quella di riferimento (minimizzando le distanze tra i punti). Il procedimento si compone di 4 passi.

1. Per ogni punto della nuvola sorgente viene trovato il più vicino punto nel modello con il quale essere accoppiato. I punti non accoppiati, dovuti ad una possibile differenza di dimensioni tra i due gruppi, non verranno utilizzati nel successivo passo.
2. Le coppie così formate sono utilizzate come input di un algoritmo, basato su SVD, in grado di minimizzare una funzione di costo il quale, come visto nel paragrafo precedente A.1, restituisce una matrice R ed un vettore t .
3. Ai punti della nuvola sorgente viene applicata la rototraslazione appena calcolata.
4. Si reitera l'algoritmo più volte fino a portare le distanze, via via minori, tra i punti al di sotto di una soglia prefissata. Nel caso di raggiungimento del numero massimo di iterazioni prefissato, l'algoritmo si ferma.

Per questa ragione la procedura di ICP può non convergere. Inoltre, partendo da una posa iniziale in cui le due nuvole di punti risultano essere particolarmente disallineate, può accadere che l'algoritmo, seppur terminando correttamente, si porti in una condizione di minimizzazione delle distanze diversa da quella desiderata.

Dalla precedente spiegazione si evince come la ricerca degli accoppiamenti tra i punti costituisca la parte dell'algoritmo computazionalmente più onerosa. Un iniziale allineamento manuale, seppur grossolano, può notevolmente ridurre i tempi di calcolo. Inoltre, fornendo alla procedura di registrazione ICP la struttura a k-d tree della nuvola di punti (come proposto da Zhang [37]), grazie alla quale sono note in ciascuna point cloud le distanze relative tra i punti, si garantisce all'algoritmo la possibilità di scartare a priori gli accoppiamenti ritenuti scorretti, così da restringere il campo di ricerca. Viene in questo modo nettamente ridotta la durata della fase 1, garantendo una maggiore efficienza e velocità di esecuzione.

Appendice B

Procedura di allineamento tra telecamera a tempo di volo e bersaglio planare

Il presente capitolo di appendice descrive il procedimento adottato durante la fase di allineamento della terna cartesiana, relativa alla telecamera tof (e quindi al sensore IR presente all'interno del Kinect V2), con il SDR posto al centro del target planare fissato alla flangia del robot antropomorfo (ABB IRB 1600). Quest'ultimo, utilizzato come ausilio durante l'analisi dell'effetto sistematico relativo alla misura di distanza, permette di posizionare il bersaglio, compiendo spostamenti noti e caratterizzati da una maggiore accuratezza [38] rispetto a quella del sistema di misura in analisi. Vengono in seguito illustrati, nel dettaglio, le necessità, la strategia ed il setup utilizzati nonché gli algoritmi su cui è stata basata la procedura di allineamento.

Nella fase di caratterizzazione dell'effetto sistematico (sezione 5.2.2) si è fatto uso di una tecnica di tipo differenziale, mediante la quale è stata imposta ad un bersaglio planare, opportunamente fissato alla flangia di un robot antropomorfo, una traslazione nota, la cui ampiezza è stata poi confrontata con la differenza (in valore assoluto) calcolata tra due misure di distanza effettuate rispettivamente prima e dopo lo spostamento. Lo scostamento (ε) tra il valore reale di distanza e quello restituito da

Kinect V2 viene infatti definito come $\varepsilon = r - |d_1 - d_2|$ dove d_1 ed d_2 sono i valori relativi alle misure di distanza restituite dallo strumento in esame ed r è l'ampiezza della traslazione imposta alla flangia del robot. La medesima procedura è stata ripetuta più volte all'interno del campo di misura.

È chiaro come questa strategia sia applicabile solo nel caso in cui il bersaglio venga orientato in modo tale da traslare lungo la direzione del vettore normale (Z') ad esso associato e la traiettoria (rettilinea) percorsa dal proprio baricentro risulti essere parallela (al più coincidente) alla direzione definita dall'asse Z uscente dalla camera IR.

L'obiettivo primario è quindi quello di rilevare la posizione relativa (rotazione e traslazione) tra camera e piano così da determinare il movimento da imporre alla flangia del robot al fine di raggiungere il corretto allineamento. Quest'ultimo, definendo con XYZ la terna associata alla camera e con $X'Y'Z'$ la terna tool del robot (associata al target), può essere ottenuto imponendo contemporaneamente il parallelismo tra le rispettive coppie di assi omologhi individuanti i piani (del sensore e del target) e garantendo la giacenza degli assi Z e Z' sulla medesima retta. In questo modo si ha inoltre la possibilità di imporre al piano rotazioni attorno ad assi paralleli a quelli della camera, condizione necessaria per il test descritto nella sezione 5.2.3.

La condizione di allineamento ottenuta è mostrata in figura B.1.

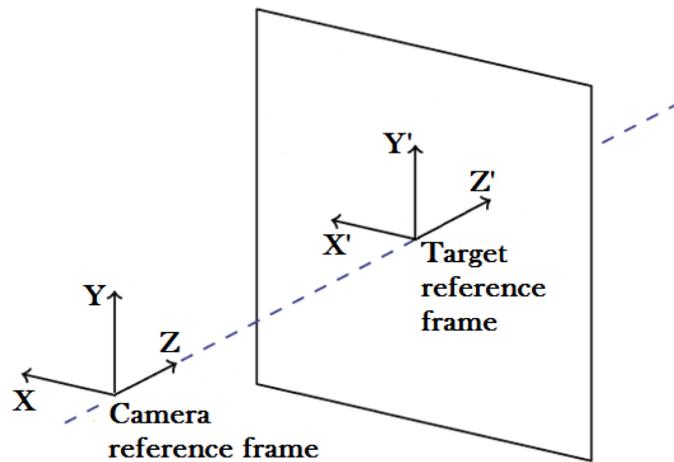


Figura B.1. Condizione di allineamento ricercata tra le terne XYZ e $X'Y'Z'$ rispettivamente associate alla camera e al bersaglio.

Per fare ciò è stata utilizzata una tecnica di calibrazione di telecamera basata sul metodo di J. Heikkilä [15], ed implementata mediante algoritmi presenti nelle librerie commerciali HALCON, sviluppate da MVTec [44], i

cui risultati sono stati verificati in un precedente lavoro svolto all'interno del laboratorio [25].

La procedura, di seguito illustrata, ha come punto di forza la semplicità ed economicità dell'hardware necessario il quale è costituito da un calibratore planare di dimensioni pari a 800 x 800 mm realizzato mediante strati alternati di alluminio e materiale polimerico, così da garantirne un'elevata planarità e una limitata deformabilità. Su di esso è riportato un pattern regolare composto da una matrice quadrata formata da 49 (7x7) cerchi neri stampati su fondo bianco e racchiusi all'interno di una cornice avente un'asimmetria, in corrispondenza di un vertice, realizzata mediante un triangolo nero; quest'ultimo accorgimento permette di definire in maniera univoca la posa della terna cartesiana relativa al calibratore.

La fase preliminare alla calibrazione consiste nell'acquisire immagini codificate in scala di grigi, provenienti dalla camera IR del Kinect V2, contenenti il target. La documentazione relativa alla libreria consiglia, al fine di ottenere buoni risultati, di disporre di almeno 20 immagini le quali devono presentare una superficie occupata per non meno del 30% dal calibratore il quale deve essere interamente presente ed assumere, nella successione di scatti, pose il più possibile casuali e differenti. Inoltre è necessario ottenere fotografie nitide e ben focalizzate servendosi di una corretta fonte di illuminazione tale da generare un netto contrasto tra i punti ed il fondo, evitando però possibili saturazioni.

Rispettando queste condizioni si otterranno ottimi risultati nella successiva fase automatizzata di ricerca del calibratore all'interno dell'immagine, basata su algoritmi di blob detection [23].

Date le dimensioni del calibratore, per semplificare le operazioni di acquisizione si è quindi deciso di fissare il target su di un sostegno e di orientare manualmente la camera, variandone la posa tra uno scatto ed il successivo. Utilizzando un'apposita applicazione basata sull'SDK fornito da Microsoft, sono state acquisite e salvate in formato *.jpg, le immagini utili alla calibrazione, alcune delle quali sono mostrate in figura B.2.

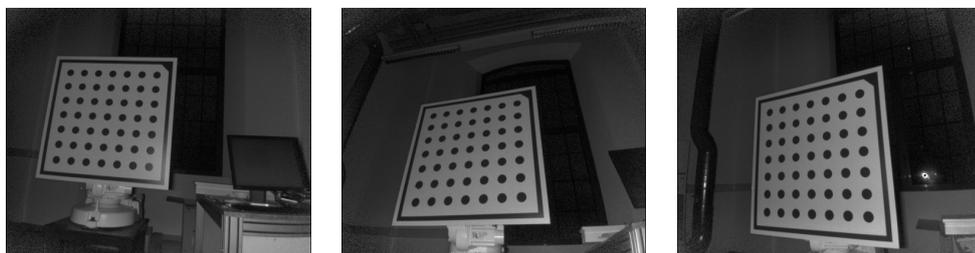


Figura B.2. Alcune pose assunte dal calibratore durante la fase di allineamento.

Successivamente viene fissata la camera ad un sostegno molto stabile e solidale al terreno (es. cavalletto fotografico) ed orientata in modo da inquadrare l'intero calibratore ora avvitato alla flangia del robot. Questa sarà la posa assunta dal dispositivo di acquisizione durante l'intero test. È chiaro come a questo punto della procedura di calibrazione, il piano di riferimento assuma una posa casuale e non sia allineato con il sensore del dispositivo a tempo di volo. Si procede quindi scattando un'ulteriore immagine, contenente il calibratore, e l'intero set di acquisizioni viene fornito come input alla procedura di calibrazione implementata in ambiente HALCON.

Dopo pochi secondi di elaborazione, vengono ricavati sia i parametri intrinseci della camera sia la posa relativa (rotazione e traslazione) presente tra la terna cartesiana associata alla camera ed il SDR del calibratore, così come è stato acquisito nell'ultima immagine, utilizzata come riferimento.

Essendo a questo punto nota la posizione assunta dalla flangia in corrispondenza dell'ultimo scatto, e quindi la differenza presente tra la posa realmente assunta e la condizione di allineamento desiderata, è possibile far compiere al robot il corretto movimento così da riportarne la flangia nella condizione di allineamento desiderata.

In conclusione è bene specificare come questa procedura si basi sulla forte ipotesi che la terna del tool, alla quale si impone il movimento, e quella del target, alla quale è riferita la rototraslazione, siano coincidenti. Questa condizione è stata ottenuta realizzando, mediante lavorazione a controllo numerico, una piastra in alluminio in grado di accoppiare con elevata accuratezza il calibratore alla flangia del robot, garantendo così la sovrapposizione delle relative terne.

Appendice C

Articolo

A Metrological Characterization of the Kinect V2 Time-of-Flight Camera

Andrea Corti, Silvio Giancola*, Giacomo Mainetti, Remo Sala

*Department of Mechanical Engineering
Politecnico di Milano
Milano, Italy, 20156*

Abstract

A metrological characterization process for time-of-flight (TOF) cameras is proposed in this paper and applied to the Microsoft Kinect V2. Based on the Guide to the Expression of Uncertainty in Measurement (GUM), the uncertainty of a three-dimensional (3D) scene reconstruction is analysed. In particular, the random and the systematic components of the uncertainty are evaluated for the single sensor pixel and for the complete depth camera. The manufacturer declares an uncertainty in the measurement of the central pixel of the sensor of about few millimetres, which is considerably better than the first version of the Microsoft Kinect. This work points out that performances are highly influenced by measuring conditions and environmental parameters of the scene; 3D point reconstruction uncertainty can vary from 1.5 to tens of millimetres.

Keywords: Time-of-Flight, Kinect V2, characterization, Point Cloud Library, GUM, metrological qualification

*Principal Corresponding author

Email address: silvio.giancola@polimi.it (Silvio Giancola)

Acronimi

2D	Bidimensionale
3D	Tridimensionale
a.C.	avanti Cristo
AC	Alternating Current
ASCII	American Standard Code for Information Interchange
CCD	Charge-Coupled Device
CMM	Coordinate Measuring Machine
CMOS	Complementary Metal-Oxide Semiconductor
CPA	Coefficiente di Periodicità sull'Accelerazione
CPU	Central Processing Unit
CW	Continuous Wave
DC	Direct Current
DOF	Degree Of Freedom
fps	fotogrammi per secondo
GDL	Grado Di Libertà
GPU	Graphics Processing Unit
HTML	HyperText Markup Language
ICP	Iterative Closest Point
IR	Infrarosso

JPEG	Joint Photographic Experts Group
LCD	Liquid Crystal Display
LUT	Look-Up Table
NAJ	Normalized Acceleration Jerk
NSU	Negligenza Spaziale Unilaterale
OSD	On Screen Display
PC	Personal Computer
pcl	point cloud library
RADAR	RAdio Detection And Ranging
RAM	Random Access Memory
RGB	Red Green Blue
rpm	revolutions per minute
SDK	Software Development Kit
SDR	Sistema Di Riferimento
SGs	Serious Games
SONAR	SOund Navigation And Ranging
SVD	Singular Value Decomposition
TAC	Tomografia Assiale Computerizzata
tof	time of flight
USB	Universal Serial Bus
VB	Visual Basic

Bibliografia

Riferimenti citati nel testo

Pubblicazioni e Manuali

- [1] Clark C Abt. *Serious games*. University Press of America, 1987 (cit. a p. 5).
- [2] K Somani Arun, Thomas S Huang e Steven D Blostein. «Least-squares fitting of two 3-D point sets». In: *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 5 (1987), pp. 698–700 (cit. a p. 97).
- [3] Ayush Bhandari et al. «Resolving multipath interference in time-of-flight imaging via modulation frequency diversity and sparse regularization». In: *Optics letters* 39.6 (2014), pp. 1705–1708 (cit. a p. 65).
- [4] Duane C Brown. «Decentering distortion of lenses». In: *Photometric Engineering* 32.3 (1966), pp. 444–462 (cit. a p. 29).
- [5] Marco Caimmi et al. «Using kinematic analysis to evaluate constraint-induced movement therapy in chronic stroke patients». In: *Neurorehabilitation and neural repair* 22.1 (2008), pp. 31–39 (cit. alle pp. 70, 80).
- [6] Filiberto Chiabrando et al. «Sensors for 3D imaging: metric evaluation and calibration of a CCD/CMOS time-of-flight camera». In: *Sensors* 9.12 (2009), pp. 10080–10096 (cit. a p. 45).
- [7] J Chow et al. «Performance analysis of a low-cost triangulation-based 3d camera: Microsoft kinect system». In: *International Society for Photogrammetry and Remote Sensing Congress (ISPRS)*. Vol. 39. 2012, pp. 175–180 (cit. alle pp. 21, 95).
- [8] Brian Curless. «Overview of active vision techniques». In: () (cit. a p. 8).

- [9] Damien Djaouti^{1&2} et al. «Origins of Serious Games». In: () (cit. a p. 5).
- [10] Adrian A Dorrington et al. «Separating true range measurements from multi-path and scattering interference in commercial range cameras». In: *IS&T/SPIE Electronic Imaging*. International Society for Optics e Photonics. 2011, pp. 786404–786404 (cit. alle pp. 33, 65).
- [11] Andrea Fusiello. «Visione computazionale». In: *Appunti delle lezioni. Pubblicato a cura dell'autore* (2008) (cit. alle pp. 23, 30).
- [12] Andrea Fusiello, Emanuele Trucco e Alessandro Verri. «A compact algorithm for rectification of stereo pairs». In: *Machine Vision and Applications* 12.1 (2000), pp. 16–22 (cit. a p. 8).
- [13] Gabriele Guidi, Michele Russo e Jean-Angelo Beraldin. *Acquisizione 3D e modellazione poligonale*. McGraw-Hill, 2010 (cit. a p. 8).
- [14] Richard Hartley e Andrew Zisserman. *Multiple view geometry in computer vision*. Cambridge university press, 2003 (cit. alle pp. 8, 23, 50).
- [15] Janne Heikkilä e Olli Silvén. «A four-step camera calibration procedure with implicit image correction». In: *Computer Vision and Pattern Recognition, 1997. Proceedings., 1997 IEEE Computer Society Conference on*. IEEE. 1997, pp. 1106–1112 (cit. a p. 104).
- [16] Texas Instruments. «Introduction to the Time-of-Flight (ToF) – System Design». In: *User's Guide* (2014) (cit. alle pp. 52, 64).
- [17] Christopher Kawatsu, Jiaying Li e CJ Chung. «Development of a fall detection system with Microsoft Kinect». In: *Robot Intelligence Technology and Applications 2012*. Springer, 2013, pp. 623–630 (cit. a p. 94).
- [18] Kouros Khoshelham e Sander Oude Elberink. «Accuracy and resolution of kinect depth data for indoor mapping applications». In: *Sensors* 12.2 (2012), pp. 1437–1454 (cit. a p. 45).
- [19] Ralph Klose, Jaime Penlington e Arno Ruckelshausen. «Usability study of 3D time-of-flight cameras for automatic plant phenotyping». In: *Bornimer Agrartechnische Berichte* 69.93-105 (2009), p. 12 (cit. a p. 56).
- [20] Benjamin Langmann, Klaus Hartmann e Otmar Loffeld. «Depth Camera Technology Comparison and Performance Evaluation.» In: *ICPRAM (2)*. 2012, pp. 438–444 (cit. a p. 45).

- [21] Kate Laver et al. «Virtual reality for stroke rehabilitation». In: *Stroke* 43.2 (2012), e20–e21 (cit. a p. 5).
- [22] Larry Li. «Time-of-Flight Camera – An Introduction». In: *Texas Instruments-Technical White Paper* (2014) (cit. alle pp. 8, 32, 33).
- [23] Tony Lindeberg. «Feature detection with automatic scale selection». In: *International journal of computer vision* 30.2 (1998), pp. 79–116 (cit. a p. 105).
- [24] Keith Lohse et al. «Video games and rehabilitation: using design principles to enhance engagement in physical therapy». In: *Journal of Neurologic Physical Therapy* 37.4 (2013), pp. 166–175 (cit. a p. 6).
- [25] GIACOMO MAINETTI. «Calibrazione di telecamere per sistemi di visione stereoscopica: Confronto tra algoritmi genetici e tecniche tradizionali». In: (2011) (cit. a p. 105).
- [26] Stefan May et al. «3D pose estimation and mapping with time-of-flight cameras». In: *International Conference on Intelligent Robots and Systems (IROS), 3D Mapping workshop, Nice, France*. 2008 (cit. a p. 45).
- [27] Zaid A Mundher e Jiaofei Zhong. «A Real-Time Fall Detection System in Elderly Care Using Mobile Robot and Kinect Sensor». In: () (cit. a p. 94).
- [28] Holger Rapp. «Experimental and theoretical investigation of correlating TOF-camera systems». In: (2007) (cit. alle pp. 53, 58).
- [29] Thikhathali Terence Ratshidaho et al. «An investigation into trajectory estimation in underground mining environments using a time-of-flight camera and an inertial measurement unit». In: *South African Journal of Industrial Engineering* 25.1 (2014), pp. 145–161 (cit. a p. 64).
- [30] Anna Carla Russo. «L'avventura del cinematografo. Storia di un'arte e di un linguaggio». In: () (cit. a p. 24).
- [31] Gustavo Saposnik, Mindy Levin et al. «Virtual reality in stroke rehabilitation a meta-analysis and implications for clinicians». In: *Stroke* 42.5 (2011), pp. 1380–1386 (cit. a p. 5).
- [32] Laurie Snider, Annette Majnemer e Vasiliki Darsaklis. «Virtual reality as a therapeutic modality for children with cerebral palsy». In: *Developmental Neurorehabilitation* 13.2 (2010), pp. 120–128 (cit. a p. 6).

- [33] Olga Sorkine. «Least-squares rigid motion using svd». In: *Technical notes* 120 (2009), p. 3 (cit. a p. 98).
- [34] SUBJECT TO CHANGE THEREFORE e A SAUDI STANDARD UNTIL APPROVED. «Guide to the Expression of Uncertainty in Measurement». In: (1995) (cit. a p. 45).
- [35] Y Tripovic et al. «Neuromotor Rehabilitation and Cognitive Outcomes in Patients with Traumatic Brain Injury through the Method BAPNE». In: *Procedia-Social and Behavioral Sciences* 152 (2014), pp. 1050–1056 (cit. a p. 88).
- [36] Ziv Yaniv. «Rigid Registration: The Iterative Closest Point Algorithm». In: *School of Engineering and Computer Science, The Hebrew University, Israel* (2001) (cit. a p. 97).
- [37] Zhengyou Zhang. «Iterative point matching for registration of free-form curves and surfaces». In: *International journal of computer vision* 13.2 (1994), pp. 119–152 (cit. a p. 102).

Materiale Online

- [38] *ABB Asea Brown Boveri*. Accessed: 2015-04-16. URL: <http://new.abb.com/it> (cit. alle pp. 50, 103).
- [39] *BTS true technology for true life*. Accessed: 2015-05-05. URL: <http://www.btsbioengineering.com/it/> (cit. a p. 71).
- [40] *Creative Senz3D*. Accessed: 2015-06-06. URL: <http://it.creative.com/p/web-cameras/creative-senz3d> (cit. a p. 21).
- [41] *Kinect for Windows features*. Accessed: 2015-01-06. URL: <http://www.microsoft.com/> (cit. alle pp. 8, 36–38, 91).
- [42] *Matlab Camera Calibration Toolbox*. Accessed: 2015-04-21. URL: http://www.vision.caltech.edu/bouguetj/calib_doc/ (cit. a p. 30).
- [43] *Mesa Imaging*. Accessed: 2015-04-16. URL: <http://www.mesa-imaging.ch/home/> (cit. a p. 36).
- [44] *MVTec HALCON*. Accessed: 2015-01-24. URL: <http://www.halcon.com/> (cit. a p. 104).
- [45] *Oculus VR*. Accessed: 2015-05-18. URL: <http://www.oculus.com/> (cit. a p. 88).
- [46] *pcl*. Accessed: 2015-01-22. URL: <http://pointclouds.org/> (cit. a p. 41).

- [47] *PMD Technologies*. Accessed: 2015-04-16. URL: <http://www.pmdtec.com/> (cit. a p. 36).

