

# POLITECNICO DI MILANO

Facoltà di Ingegneria Civile, Ambientale e Territoriale



Corso di Laurea Specialistica in

Ingegneria per l' Ambiente e il Territorio

## **Posizionamento cinematico in ambiente urbano tramite il sistema integrato Kinect-GPS.**

Relatore: Prof. *Mirko Reguzzoni*

Correlatore: Ing. *Diana Pagliari*

Tesi di Laurea di:

*Marco Vavassori*

matricola 824373

*Anno Accademico 2014/2015*

Dedica

Da scrivere

Ringraziamenti

Da scrivere

# INDICE

Indice delle figure.....	7
Indice delle tabelle.....	12
1 INTRODUZIONE .....	15
2 STATO DELL'ARTE PER LA NAVIGAZIONE OUTDOOR .....	19
2.1 GNSS .....	22
2.2 HSGPS .....	27
2.3 MULTI-GNSS .....	28
2.4 INS (Inertial Navigation System).....	29
2.5 MMS (Mobile Matting System).....	31
2.6 SLAM (Simultaneous Localization and Mapping) .....	33
2.7 STAZIONE TOTALE .....	35
2.8 NAVIGAZIONE CON STRUMENTI FOTOGRAMMETRICI .....	37
2.9 NAVIGAZIONE CON MICROSOFT KINECT 2.....	38
3 SENSORE MICROSOFT KINECT .....	40
3.1 KINECT PER XBOX360 .....	41
3.2 KINECT PER XBOXONE .....	45
3.3 LIMITI E OBIETTIVI DI KINECT2 .....	48
3.4 SENSORE ToF .....	51

3.5	PIXEL DIFFERENZIALE.....	52
3.6	CALIBRAZIONE DELLE CAMERE RGB e IR.....	55
3.7	CONFRONTO TRA SENSORI KINECT1 e KINECT2 .....	57
4	POSIZIONAMENTO CINEMATICO CON TECNICHE FOTOGRAMMETRICHE.....	62
4.1	MODELLO PROIETTIVO DIRETTO E INVERSO IN FOTOGRAMMETRIA (Equazioni di Collinearità).....	63
4.2	MODELLO DI CALIBRAZIONE DELLE CAMERE .....	66
4.3	TRIANGOLAZIONE AEREA (Bundle block adjustment) .....	70
4.4	ROTOTRASLAZIONI .....	73
5	METODO FOTOGRAMMETRICO PROPOSTO .....	77
5.1	OPERAZIONI PRELIMINARI .....	80
5.1.1	Calibrazione delle camere .....	80
5.1.2	Sincronizzazione degli strumenti.....	81
5.1.3	Calibrazione geometrica del sistema di sensori .....	82
5.2	RILIEVO .....	85
5.2.1	Acquisizione delle immagini .....	89
5.2.2	Traiettoria da dati GPS e da stazione totale ad inseguimento.....	90
5.3	SOLUZIONE FOTOGRAMMETRICA.....	91
5.3.1	Sottocampionamento delle immagini.....	92
5.3.2	Correzione dalla distorsione delle camere.....	93

5.3.3	Ricampionamento delle immagini Depth.....	94
5.3.4	Costruzione della nuvola di punti.....	96
5.3.5	Costruzione dell'immagine RGBD .....	98
5.3.6	Allineamento di coppie di immagini.....	99
5.3.7	Tracciamento della traiettoria dei centri di presa.....	106
5.4	SOLUZIONE DEL SENSORE U-BLOX .....	107
5.5	FILTRO DI KALMAN .....	108
6	TEST .....	113
6.1	SCELTA DELLA STRUMENTAZIONE E DEL LUOGO DEL RILIEVO .....	113
6.2	ACQUISIZIONE DEI DATI .....	122
6.3	PROCESSAMENTO DEI DATI .....	126
6.3.1	Selezione delle coppie di immagini .....	127
6.3.2	Processamento del dato grezzo .....	129
6.3.3	Processamento delle rototraslazioni.....	132
6.4	TEST1 .....	145
6.5	TEST2 .....	151
6.6	TEST3 .....	153
7	CONCLUSIONI .....	157
	Bibliografia.....	166

## • *Indice delle figure*

<i>Figura 1 Esempio di Canyon Urbano in cui il segnale diretto dei satelliti non è sempre ricevibile (da <a href="http://www.canyon_urbani.com">www.canyon_urbani.com</a>) .</i>	<i>20</i>
<i>Figura 2. Esempio di interruzione del segnale (di ricezione durante un posizionamento cinematico (da <a href="http://www.GPS-galleria.com">www.GPS-galleria.com</a>)).</i>	<i>21</i>
<i>Figura 3 Principio di triangolazione per la determinazione della posizione con sistema GNSS (da <a href="http://www.GNSS.com">www.GNSS.com</a>).</i>	<i>24</i>
<i>Figura 4 Esempio di pessima distribuzione dei satelliti. PdoP alto (da <a href="http://www.PDoP_GNSS.com">www.PDoP_GNSS.com</a>).</i>	<i>25</i>
<i>Figura 5 Esempio di ottima distribuzione dei satelliti. PdoP basso (da <a href="http://www.PDoP_GNSS.com">www.PDoP_GNSS.com</a>).</i>	<i>25</i>
<i>Figura 6 Schema di funzionamento di un IMU (da <a href="http://www.INS">www.INS</a>).</i>	<i>30</i>
<i>Figura 7 Sistema di MMS applicato ad un'automobile (da <a href="http://www.MobileMappingSystem.com">www.MobileMappingSystem.com</a>).</i>	<i>32</i>
<i>Figura 8 a)Traiettoria 2D stimata con Microsoft Kinect e vera. B) Mmodello 3D costruito con immagini RGB-D (da Felix Endres e al. 2012)</i>	<i>34</i>
<i>Figura 9 Stazione Totale e Prisma (da <a href="http://www.Total_Station.com">www.Total_Station.com</a>).</i>	<i>36</i>
<i>Figura 10 Sensore (Microsoft) Kinect ( da <a href="http://www.Microsoft-Kinect.com">www.Microsoft-Kinect.com</a> ).</i>	<i>42</i>
<i>Figura 11 Distribuzione dei sensori che compongono il Microsoft Kinect ( da <a href="http://www.Microsoft-Kinect.com">www.Microsoft-Kinect.com</a>).</i>	<i>42</i>
<i>Figura 12.a) e b) a) Inquadratura con camera digitale del Pattern proiettato in una stanza buia dall'emettitore IR. .b) esempio di Pattern IR proiettato su una scena casuale ( da <a href="http://www.Microsoft-Kinect.com">www.Microsoft-Kinect.com</a>).</i>	<i>44</i>
<i>Figura 13 Microsoft Kinect per XboxOne (<a href="http://www.Microsoft_Kinect_One.com">www.Microsoft_Kinect_One.com</a>).</i>	<i>45</i>
<i>Figura 14 Sensori di Kinect2 che permettono la generazione di immagini RGBD (<a href="http://www.Microsoft_Kinect_One.com">www.Microsoft_Kinect_One.com</a>).</i>	<i>46</i>

<i>Figura 15 Immagine Depth catturata a 2.5 m dal soggetto (da Sell and O'Connor , 2014).</i>	49
<i>Figura 16 Esempio di cattura della figura reso possibile dal range dinamico del nuovo sensore Kinect2 (da Sell and O'Conno , 2014).</i>	50
<i>Figura 17 Sistema del sensore di immagini 3D (da Sell and O'Connor , 2014).</i>	52
<i>Figura 18 Sensore ToF e forme d'onda del segnale (da Sell and O'Connor , 2014).</i>	53
<i>Figura 19 Pannello di calibrazione tramite Calibration App di Matlab (www.calibration-app-matlab.com).</i>	57
<i>Figura 20 Deviazione standar misurata tra 100 frame RGB dei sensori Kinect 1(sx) e 2(dx) riferita all'intensità nei canali R,G e B (da Pagliari e Pinto, 2015).</i>	58
<i>Figura 21 Deviazione standard (mm) calcolata per ogni pixel dell'immagine raw Depth acquisita dal sensore Kinect1 (da Pagliari e Pinto, 2015).</i>	59
<i>Figura 22 Deviazione standard (mm) calcolata per ogni pixel dell'immagine raw Depth acquisita dal sensore Kinect2 (da Pagliari e Pinto, 2015).</i>	60
<i>Figura 23 Stima dell'errore commesso da Kinect1 e 2 come funzione della distanza tra camera ed oggetto (da Pagliari, 2015).</i>	61
<i>Figura 24 Relazioni tra immagine e sistema oggetto.</i>	65
<i>Figura 25 Modello reale dell'obiettivo e relativa violazione al Principio di Collinearità (www.PrincipiodiCollinearità.com).</i>	68
<i>Figura 26 Risustato di un processo di Triangolazione aerea per un blocco di 4 immagini(www.bundleblockadjustment.com)</i>	73
<i>Figura 27 Schema di flusso del metodo integrato proposto.</i>	80
<i>Figura 28 Immagine dei pixel con contenuto informativo dell'immagine Depth originaria.</i>	95
<i>Figura 29 Zoom di un particolare dell'immagine 28.</i>	95
<i>Figura 30 Immagine dei pixel con contenuto informativo dell'immagine Depth processata.</i>	96
<i>Figura 31 Schematizzazione delle relazioni che intercorrono tra lo spazio immagini e lo spazio oggetto.</i>	97



<i>Figura 32</i> Convoluzione con Gauss e creazione di un DoG (Lowe, 2004).....	101
<i>Figura 33</i> I DoG assoluti vengono selezionati confrontando il descrittore con quello dei pixel circostanti nella sua e nelle scale inferiori ed inferiori (Lowe,2004) .....	102
<i>Figura 34</i> a) Gradienti radiometrici predominanti in una regione circolare attorno al punto candidato. B) descrittore SIFT.....	103
<i>Figura 35</i> Piantina di Piazza Leonardo Da Vinci, Milano ( <a href="http://blog.urbanfile.org/">www.http://blog.urbanfile.org/</a> ). .....	115
<i>Figura 36</i> a) Fontana di Piazza Leonardo, Milano. b) Scultura al lato sud della fontana di Piazza Leonardo. ( <a href="http://blog.urbanfile.org/">www.http://blog.urbanfile.org/</a> ).....	117
<i>Figura 37</i> Particolari del materiale petroso di cui sono costituiti la pavimentazione e la fontana (foto da rilievo).....	117
<i>Figura 38</i> a) Il ricevitore a doppia frequenza GPS LEICA GS14 con apposito treppiede b) Multistation LEICA MS60 c) Sensore GPS miniaturizzato U-blox. d) prisma LEICA. ....	119
<i>Figura 39</i> Carrellino di 4 ruote su cui poggiano f) laptop e batteria e). Ed a cui è stata fissata la palina tramite legacci. Sulla pallina troviamo a) il sensore Kinect b) il sensore U-blox c) il prisma ed infine l'antenna GPS a doppia frequenza. ....	121
<i>Figura 40</i> Target posizionati sul lato ovest della scultura. ....	123
<i>Figura 41</i> Dislocazione del quadrato di calibrazione e dei suoi spigoli rispetto alla fontana ( <a href="http://www.googleheart.com">www.googleheart.com</a> ). ....	125
<i>Figura 42</i> Traiettoria dei centri di presa determinata dalle soluzioni dell'elaborazione delle immagini da Kinect. Nei riquadri blu con le opportune frazioni sono indicati i tratti in cui è stato necessario diminuire il sottocampionamento delle immagini.....	132
<i>Figura 43</i> Traiettoria di Giro 1 e quadrato di calibrazione ottenuto dalle traiettorie delle immagini Kinect nel sistema di riferimento Kinect. ....	134

<i>Figura 44 Traiettoria dei centri di fase dell'antenna GPS a doppia frequenza e posizionamento statico per gli spigoli del quadrato di calibrazione nel sistema di riferimento della traiettoria della prima immagine del blocco di immagini RGBD.....</i>	<i>136</i>
<i>Figura 45 Spigoli del quadrato di calibrazione ottenuti dai metodi fotogrammetrici e GPS espressi in a) nel proprio sistema di riferimento ed in b) dopo aver applicato agli spigoli ottenuti da GPS in sdr Locale GPS la rototraslazione per portarli nel sdr Kinect. ....</i>	<i>137</i>
<i>Figura 46 Traiettoria ricavata dal sistema GPS con processamento rtk(in blu) e traiettoria ottenuta dalla stazione totale(in verde).....</i>	<i>139</i>
<i>Figura 47 Particolare del secondo rettilineo delle traiettoria ottenute dal sistema GPS_rtk (in blu) e dalla stazione totale(in verde).....</i>	<i>139</i>
<i>Figura 48 Traiettoria Giro1 da metodo fotogrammetrico e metodo GPS nel sistema di riferimento della prima immagine del blocco Kinect. ....</i>	<i>141</i>
<i>Figura 49 Risultato del primo test sul filtro di Kalman in cui sono utilizzati tutti i dati disponibili delle traiettorie da Kinect e da GPS_rtk (sistema di riferimento della traiettoria da Kinect).....</i>	<i>145</i>
<i>Figura 50 Deviazione dalla traiettoria reale causata dall'ipotesi di moto rettilineo uniforme adottata nel Filtro di Kalman. ....</i>	<i>147</i>
<i>Figura 51 Risultato del primo test sul filtro di Kalman in cui sono utilizzati tutti i dati disponibili delle traiettorie da Kinect e da GPS_rtk (sistema di riferimento della traiettoria da Kinect). In a) è individuata un'imprecisione della traiettoria dovuta ad un valore di spostamento della componente dinamica anomalo. In b) il risultato della chiusura della traiettoria ottenuta dal filtro.....</i>	<i>148</i>
<i>Figura 52 Anomalia nel tracciamento della traiettoria del filtr di Kalman.....</i>	<i>149</i>
<i>Figura 53 Particolare della chiusura della traiettoria verso il punto di partenza del rilievo (Spigolo) da parte del filtro di Kalman.....</i>	<i>150</i>
<i>Figura 54 Risultato del filtro di Kalman con traiettoria GPS della seconda curva assente. ....</i>	<i>151</i>

<i>Figura 55 Particolare della deviazione nella curva in cui manca a componente di posizionamento nel filtro di Kalman.....</i>	<i>153</i>
<i>Figura 56 Traiettoria calcolata da filtro di Kalman per il test3.....</i>	<i>154</i>
<i>Figura 57 Traiettoria calcolata dal filtro di Kalman per il test3 a confronto con la traiettoria di riferimento della stazione totale. ....</i>	<i>155</i>

## • *Indice delle tabelle*

<i>Tabella 1 Specifiche tecniche dei sensori ottici del Microsoft Kinect per Xbox360 (da Pagliari 2015). .....</i>	<i>43</i>
<i>Tabella 2 Confronto tra caratteristiche della prima e della seconda versione del Kinect.....</i>	<i>48</i>
<i>Tabella 3 Parametri di OI per il sensore Kinect2 stimati dalla calibration app del software Matlab. ....</i>	<i>56</i>
<i>Tabella 4 Parametri di OE relativo della camera RGB rispetto alla camera IR del sensore Kinect2. ....</i>	<i>56</i>
<i>Tabella 5 Deviazione tra le coordinate degli spigoli ottenuti dai metodi fotogrammetrico e GPS espressi nello stesso sistema di riferimento Kinect. ....</i>	<i>138</i>
<i>Tabella 6 Valori di deviazioni nelle componenti x, y e z del sistema di riferimento della traiettoria del Kinect tra traiettoria ottenuta dall'antenna GPS e la traiettoria della stazione totale. ....</i>	<i>140</i>
<i>Tabella 7 Media delle deviazioni nelle componenti x, y e z (rispetto al sistema di riferimento della traiettoria da Kinect) della traiettoria ottenuta dal filtro di Kalman rispetto alla traiettoria di riferimento della stazione totale nel test1.....</i>	<i>150</i>
<i>Tabella 8 Medie delle deviazioni tra traiettoria del filtro di Kalman del test2 e la traiettoria di riferimento della stazione totale nel caso dell'intero Giro e nel caso della sola curva in cui è assente la traiettoria del GPS. ....</i>	<i>152</i>
<i>Tabella 9 Medie della deviazione nelle componenti x, y e z della traiettoria determinata dal filtro di Kalman nel test3 rispetto alla traiettoria di riferimento della.....</i>	<i>156</i>

DEDICA

## RINGRAZIAMENTI

# 1 INTRODUZIONE

Il posizionamento nello spazio di cose o di persone in movimento è una necessità antica quanto è antica la pratica dello scambio commerciale da parte dell'uomo. La *navigazione* (anticamente da *navicazione*, dal latino *navigatio -onis*) da enciclopedia Enciclopedia Treccani online (2016), vista come spostamento da un luogo iniziale con l'obiettivo di raggiungerne uno finale prefissato, rappresenta la tecnica, il modo di navigare (*scienza o arte della navigazione*). Essa comprende tutto il complesso delle conoscenze teoriche e pratiche necessarie per poter predisporre e seguire le rotte verso un punto prefisso, e che riguardano soprattutto i metodi di calcolo e gli strumenti di bordo che aiutano a determinare, controllare e mantenere la rotta sia sul mare (*scienza della navigazione marittima*) sia nell'aria (*scienza della navigazione aerea*), tenendo nel dovuto conto i fattori fisici che influenzano la navigazione, principalmente quelli meteorologici (forza e direzione del vento, stato del mare, ecc.), idrografici (correnti marine, ecc.) e orografici (altezze dei monti, ecc.). La navigazione, quindi, rappresenta una pratica la quale, per necessità intrinseca all'idea di rotta da perseguire, richiede come prerequisito fondamentale la conoscenza della posizione corrente.

Si ritiene che il problema della navigazione aerea e superficiale (in campi aperti) sia stata da tempo risolta dall'introduzione di sistemi GNSS (Global Navigation Satellite System) i quali rappresentano ormai una soluzione standard al problema (Hofmann-Wellenhof et al., 2008). Esso rappresenta un sistema per la radionavigazione tridimensionale (navigazione) che integra le funzioni dei satelliti di posizionamento con

quelle di alcuni satelliti di telecomunicazioni e di stazioni di controllo terrestri per la definizione con elevata precisione (qualche centimetro) della traiettoria di ogni mezzo mobile connesso al sistema.

Tuttavia si deve tenere in considerazione che molte delle applicazioni che necessitano di un posizionamento spesso richiedono soluzioni di elevate precisioni in ambienti considerati sfavorevoli per l'utilizzo di sistemi tradizionali come il GPS. In ambienti in cui non si presenta una situazione di tipo Line-of-Sight(LOS) con i satelliti della costellazione, come ad esempio canyon urbani (cioè strade parzialmente strette chiuse da palazzi alti, gallerie o semplicemente ambienti indoor (capannoni ecc), utilizzare un posizionamento con sistema GNSS diventa impossibile.

Nelle ultime due decadi è stato osservato un notevole aumento della richiesta di sviluppo tecnologico la cui applicazione fosse in grado di risolvere il problema della navigazione con ordine di precisione tali da poter essere utilizzati nelle problematiche proposte da parte sia del mercato che delle pubbliche amministrazioni. Secondo quanto espresso dall'economista Robert Solow (Solow, 1956) nel suo modello neoclassico (*esogeno*) la crescita della produzione aggregata di beni e servizi è il risultato dell'interazione tra progresso tecnologico, crescita dello stock di capitale e la crescita della forza lavoro (*La teoria della crescita e dello sviluppo economico di Solow, 1956*). L'aumento improvviso della domanda avvenuta nelle ultime due decadi quindi va alla stessa velocità con cui si muove un effettivo sviluppo di differenti tipi di tecnologie lowcost. Queste hanno aperto il mercato delle possibilità di soluzione a problemi di navigazione a strati di mercato molto più ampie rispetto agli anni precedenti. A titolo di



esempio si osservi la diffusione dell'utilizzo della tecnologia GPS negli smartphone di ultima generazione, i quali anche se permettano di raggiungere precisioni nell'ordine di solo una decina di metri (nell'ipotesi di satelliti non oscurati) hanno garantito ad una fascia di mercato decisamente più ampia rispetto alla precedente l'utilizzo di questo sistema di navigazione (tipicamente per scopi di navigazione stradale).

Per far fronte alla crescente necessità di risolvere il problema di georeferenziazione automatica anche in ambito urbano ha portato allo sviluppo di tecnologie alternative, parzialmente o totalmente, ai sistemi GNSS.

L'intento di questa tesi consiste nel proporre un sistema di navigazione con soluzione ottenuta da risoluzione di post processamento del problema fotogrammetrico. Inoltre si chiede che i dati siano ottenuti da una piattaforma di misura lowcost in un ambiente outdoor tipicamente urbano, quindi con LOS potenzialmente ridotto o totalmente privo della disponibilità dei satelliti. In particolare si descrive l'utilizzo del sensore di immagini e suoni Kinect per Xbox One, sensore di marchio Microsoft nata come interfaccia giocatore-console capace di gestire con performance elevate informazioni quali voce e movimenti dell'utente al fine di applicazioni di intrattenimento o altro.

Questo elaborato di tesi si sviluppa nel seguente modo:

- 1) Introduzione
- 2) Stato dell'arte della navigazione outdoor

*Introduzione generale a diverse tecniche che negli ultimi anni per le loro caratteristiche innovative sia dal punto di vista delle potenziali applicazioni che da quello del risparmio (tecniche lowcost) sono state sviluppate, approfondite o*

*adattate per fungere da supporto al posizionamento/navigazione in ambienti in cui l'utilizzo della soluzione tradizionale da GNSS non è sufficiente.*

- 3) Presentazione del sensore Kinect One con breve cenno alla sua prima versione
- 4) Posizionamento cinematico con tecniche fotogrammetriche
  - a) Problema fotogrammetrico (equazioni di collinearità)
  - b) Modello di calibrazione delle camere
  - c) Triangolazione aerea (Bundle block adjustment)
  - d) Rototraslazioni
- 5) Metodo proposto
  - a) Operazioni preliminari
  - b) Rilievo
  - c) Soluzione fotogrammetrica
  - d) Soluzione del sensore U-blox
  - e) Filtro di Kalman
- 6) Test
- 7) Conclusioni

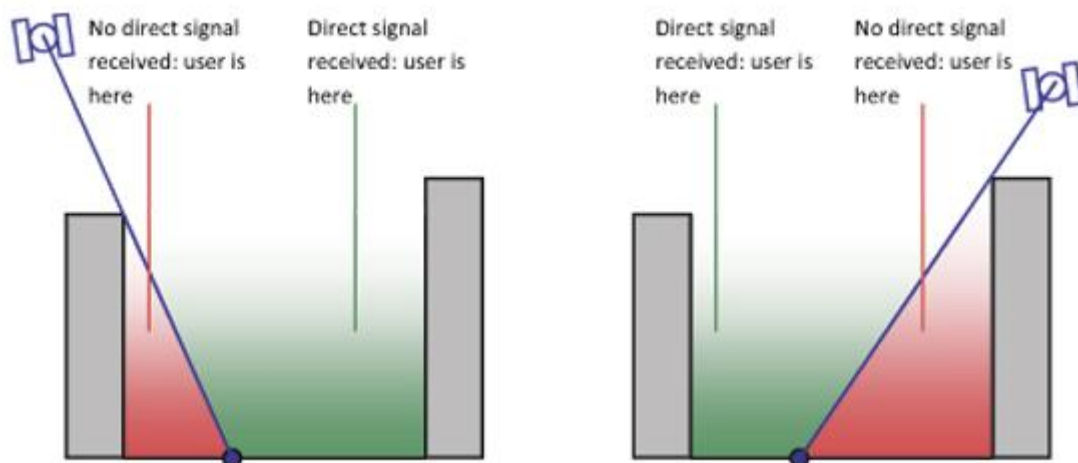
## 2 STATO DELL'ARTE PER LA NAVIGAZIONE OUTDOOR

Il sistema GNSS è la tecnica di posizionamento cinematico largamente più utilizzata per la navigazione aerea e superficiale (marittima e autostradale) perchè sebbene esistano debolezze legate a circostanze logistiche che ne limitano l'utilizzo, permette di posizionare globalmente il ricevitore indipendentemente da alcune variabili, quali ad esempio l'illuminazione, a patto che l'antenna ricevi bene il segnale dei satelliti. La misura dipende sostanzialmente dal tipo di terminale mobile e dalla quantità di segnali disponibili, il cui numero minimo necessario per determinare la posizione del ricevitore è uguale alle incognite del problema: le tre coordinate del punto di acquisizione e l'offset temporale tra l'orologio del ricevitore e gli orologi dei satelliti in vista. Nel paragrafo 2.1 viene riportata una breve introduzione al posizionamento con sistema GNSS. Per approfondimenti si rimanda comunque a Hofmann e Wellenhop in GNSS Global Navigation System (2008).

Il GNSS è un sistema che si basa sulla misura del tempo di arrivo (satellite-ricevitore) del segnale trasmesso dal satellite. Il segnale trasmesso è di alta frequenza: nel caso del GPS (abbreviazione di NAVSTAR GPS, sistema GNSS gestito dal governo degli Stati Uniti d'America) essa è di 1.575 GHz. Un segnale di tale frequenza non oltrepassa le barriere solide come pareti o quelle naturali come alberi con molte foglie. In questo caso si dice che il satellite non è disponibile al ricevitore, o oscurato. Uno dei problemi che determinano l'impossibilità di utilizzare il sistema GNSS è proprio quello di effettuare

il posizionamento del ricevitore quando quest'ultimo si trovi in ambienti dove i satelliti risultano oscurati, come ad esempio nel caso di presenza di alberi molto alti, palazzi, etc.. A causa dell'oscuramento, totale o alternato, del segnale inviato dai satelliti il sistema o non è in grado di fornire una soluzione oppure la precisione risulta essere molto scarsa, insufficiente per scopi di navigazione geodetica.

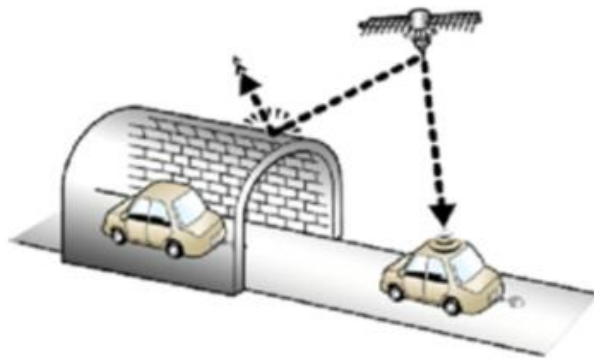
Zone ad alta densità di urbanizzazione, nei quali è alta la possibilità che vi siano presenti palazzi e strade strette (canyon urbani: in *Figura 1* è presentato un esempio), pur essendo ambienti in cui è costantemente in crescita la densità dei servizi che richiederebbero l'utilizzo del sistema di posizionamento, risultano spazi in cui la tecnica di navigazione migliore, cioè il GNSS, non è sempre utilizzabile.



*Figura 1 Esempio di Canyon Urbano in cui il segnale diretto dei satelliti non è sempre ricevibile (da [www.canyon\\_urbani.com](http://www.canyon_urbani.com)).*

Anche la domanda di tecniche per la navigazione in ambienti indoor è cresciuta molto negli ultimi anni. Si pensi ad esempio alla necessità di indagare i giacimenti minerari in

gallerie sotterranee o la necessità di localizzare i soccorritori in fase di recupero. In ambienti di lavoro indoor o in gallerie (parzialmente o completamente coperte) l'oscuramento totale di ogni satellite impedisce di disporre delle condizioni minime indispensabili per il posizionamento cinematico. In *Figura 2* un esempio di assenza di segnale durante il passaggio attraverso una galleria.



*Figura 2. Esempio di interruzione del segnale (di ricezione durante un posizionamento cinematico (da [www.GPS-galleria.com](http://www.GPS-galleria.com)))*

Considerate le difficoltà dell'utilizzo in ambiente urbano del solo sistema GNSS, vengono citati brevemente alcuni metodi alternativi che vengono aggiunti al metodo GNSS con il fine di raggiungere soluzioni di precisione accettabile anche in ambienti ostili come quello urbano. Nel capitolo 2.1 vengono introdotti i concetti fondamentali della misura della posizione con sistema GNSS. Nel paragrafi 0 e 2.3 sono presentati una raccolta di alcuni sistemi di utilizzo alternativo della classica misura son sistema GNSS. Nei paragrafi a seguire invece si parla di sistemi integrati a quello GNSS per il posizionamento cinematico, distinguendo tra quelli lowcost e non.

## 2.1 GNSS

Il sistema GNSS (*Global Navigation Satellite System*) è un servizio per il posizionamento geo-spaziale a copertura globale che permette il calcolo delle coordinate geografiche (longitudine, latitudine ed altitudine) grazie all'utilizzo di appositi ricevitori elettronici (antenne). Questo metodo di posizionamento fa affidamento ad un sistema complesso di satelliti artificiali in orbita intorno alla terra che viene chiamata Costellazione.

Il principio con cui viene effettuata una misura di posizione tramite sistemi GNSS consiste nella calcolo della distanza tra il dispositivo ricevitore ed i satelliti, orbitanti attorno alla Terra, di cui il dispositivo (un'antenna) riceve, in un preciso intervallo di tempo, il segnale trasmesso. La distanza è espressa come:

$$D_R^i = \sqrt{(x^i - x_R)^2 + (y^i - y_R)^2 + (z^i - z_R)^2} \quad (1)$$

dove  $i$  indica il satellite  $i$ -esimo,  $R$  il ricevitore  $R$ -esimo,  $D_R^i$  la distanza tra l' $i$ -esimo satellite ed il ricevitore  $R$ -esimo.

Nell'equazione 1 della distanza le incognite sono 3 e corrispondono alle coordinate dell'oggetto ricevitore. In linea di principio quindi servirebbe ricevere il segnale di almeno 3 satelliti contemporaneamente. In realtà il principio di misurazione ne richiede almeno 4.

La misura di distanza è calcolata moltiplicando il tempo misurato perchè il segnale compia il percorso satellite-ricevitore. Questa misura è possibile solo se si dispone

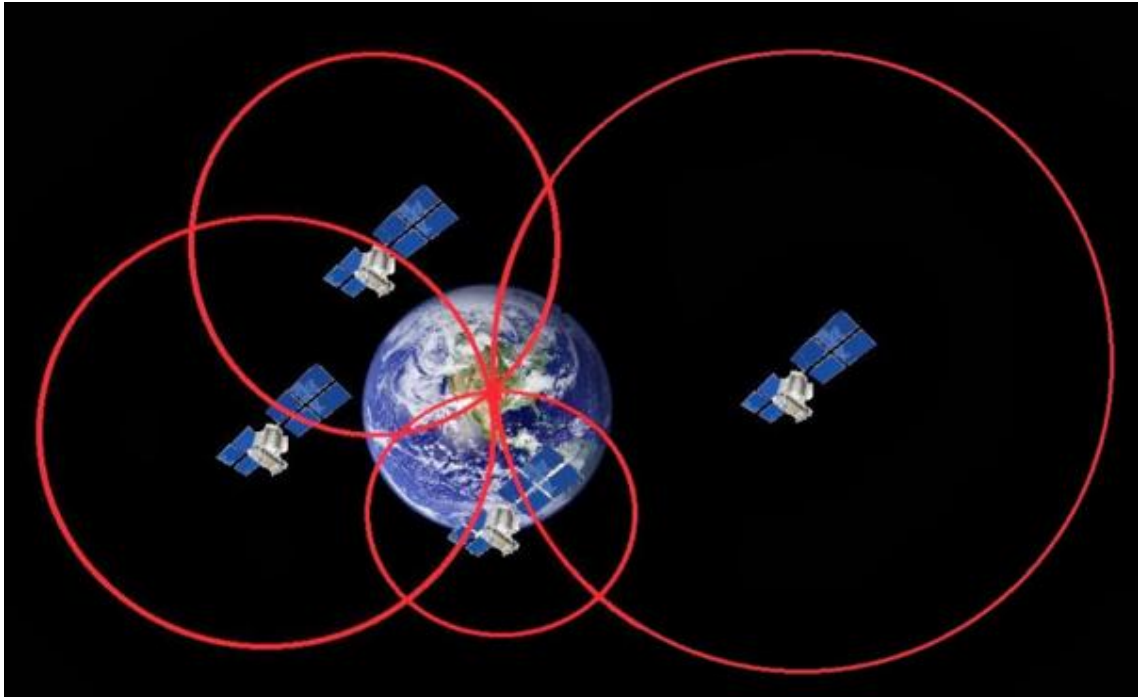
dell'orario dell'istante di invio del segnale e l'orario dell'istante in cui esso raggiunge il ricevitore. Quindi sostanzialmente serve che su ogni satellite e sul ricevitore sia presente un orologio. Tuttavia l'orologio del ricevitore non ha la stessa precisione degli oscillatori atomici montati sui satelliti. Questo offset determina un'alterazione nel calcolo della distanza che necessita di essere considerata come un'incognita addizionale al problema, da qui la necessità del quarto satellite. Quindi l'equazione della distanza diventa:

$$P_R^i = \sqrt{(x^i - x_R)^2 + (y^i - y_R)^2 + (z^i - z_R)^2} + \Delta t \cdot c = D_R^i + \Delta t \cdot c \quad (2)$$

dove  $P_R^i$  è la misura di pseudo-range,  $c$  la velocità del segnale e  $\Delta t$  l'offset tra l'orologio del satellite e quello del ricevitore.

Il segnale, crittografato, viene scomposto nelle sue componenti in frequenza da un processamento eseguito dal dispositivo ricevitore. Dalla scomposizione del segnale si misura il tempo impiegato dal segnale a percorrere la distanza satellite/ricevitore.

Dal segnale trasmesso inoltre si risale all'ora in cui il segnale è stato trasmesso dal satellite. Consultando i tabulati forniti dal produttore del sistema si individua la posizione del satellite di cui si ha ricevuto il segnale alla data ora in cui esso è stato inviato. A questo punto, conoscendo posizione dei satelliti e distanza dei satelliti dal dispositivo ricevitore, tramite triangolazione e processamento ai minimi quadrati, viene calcolata la posizione del dispositivo nella stessa ora in cui viene svolta l'acquisizione del segnale. In *Figura 3* viene raffigurato il principio di triangolazione per determinare la posizione dell'antenna ricevente partendo dal valore di distanza determinato dal segnale inviato da ogni satellite.



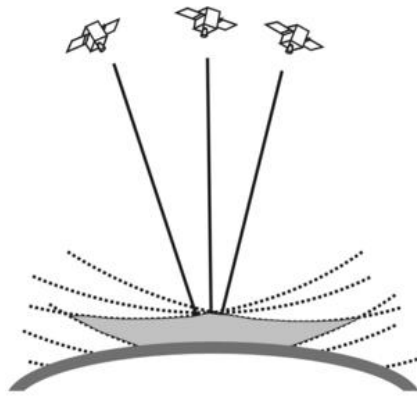
*Figura 3 Principio di triangolazione per la determinazione della posizione con sistema GNSS (da [www.GNSS.com](http://www.GNSS.com)).*

La determinazione della posizione del ricevitore con soluzione da metodo GNSS può essere svolta in diversi modi; tutti dipendono dal tipo di rilievo, dalla quantità di satelliti visibili e dalla qualità del segnale, ed infine dalla tecnica di di processamento adottata. In generale il posizionamento si divide in posizionamento statico e posizionamento cinematico. Nel primo caso si determina la posizione di un ricevitore fermo nell'intervallo di tempo di acquisizione del segnale, nel secondo il ricevitore è in movimento.

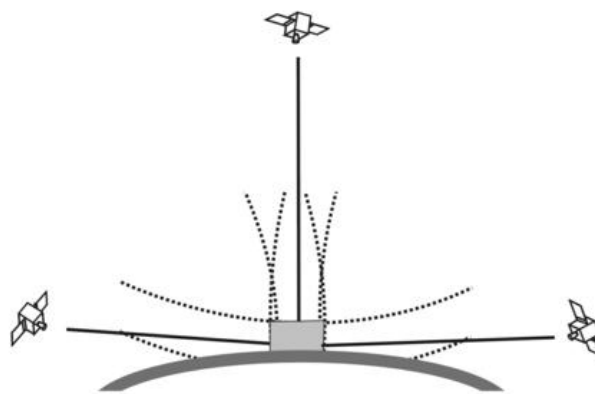
Come è stato spiegato una condizione necessaria affinché sia possibile posizionare un ricevitore tramite sistema GNSS è la disponibilità un un numero sufficiente di satelliti: tuttavia essa non è sufficiente. Infatti è necessario che la distribuzione nello spazio dei



satelliti sia idonea. Questo concetto è espresso da un coefficiente che è proporzionale al volume del solido che ha vertici nei satelliti visibili dal ricevitore. Esso è il PdoP (Position Dilution of Precision) e quando il suo valore è alto allora la geometria dei satelliti è inadatta alla triangolazione, il che determina una precisione bassa del posizionamento. In *Figura 4* e *Figura 5* sono rappresentati rispettivamente due esempi di cattiva e buona distribuzione dei satelliti.



**Figura 4** Esempio di pessima distribuzione dei satelliti. PdoP alto (da [www.PDoP\\_GNSS.com](http://www.PDoP_GNSS.com)).



**Figura 5** Esempio di ottima distribuzione dei satelliti. PdoP basso (da [www.PDoP\\_GNSS.com](http://www.PDoP_GNSS.com)).

Il posizionamento fornito da un sistema GNSS è soggetto a vari tipi di errore. Essi possono essere errori di tipo strumentale, possono essere dipendenti dal tipo di modello utilizzato o dall'osservazione. Gli errori di tipo strumentale possono essere ridotti dalla qualità del tipo di ricevitore, che ovviamente aumenta il prezzo del prodotto. Tra quelli dipendenti dal modello si include quello di offset tra orologio del satellite e orologio del ricevitore (chiamato Biases). negli errori di osservazione si distinguiamo quello di Multipath e di Cycle Slips. Il primo si verifica quando il segnale non raggiunge direttamente l'antenna, ma è soggetto a fenomeni di riflessione multipla da superfici. Il segnale risulta peggiorato nel suo rapporto segnale/rumore e la sua distanza percorsa risulta diversa da quella che collega il satellite-ricevitore. L'errore di Cycle Slips invece si verifica quando non è possibile acquisire uno o più cicli di segnale perchè i satelliti sono temporaneamente oscurati.

Un altro tipo di errore è determinato dalle variabili che condizionano la propagazione di segnale ondulatorio (come quello trasmesso nel sistema GNSS) nel mezzo, quindi pressione, umidità e temperatura. La condizione meteorologia, gli effetti relativistici e gli effetti di radiopropagazione nella ionosfera e nella troposfera sono variabili che condizionano la soluzione fornita dal GNSS.

La precisione teorica per il sistema GPS (abbreviazione di NAVSTAR GPS, acronimo di NAVigation Satellite Timing And Ranging Global Positioning System) senza considerare alcun errore del modello, è tra l'1% ed il 2% della lunghezza d'onda del segnale. Tuttavia per misure di fase statiche il GPS è possibile arrivare a precisioni

molto alte, tra i 2 mm e i 5mm. Nel caso di applicazioni cinematiche con condizione di cielo scoperto le tre soluzioni del posizionamento possiedono errori dell'ordine della decina di centimetri.

Per uno studio dettagliato della risoluzione del problema del posizionamento con sistemi GNSS si rimanda a Hofmann e Wellenhop in GNSS Global Navigation System (2008).

## 2.2 HSGPS

In ambienti in cui la qualità del segnale è bassa l'attenuazione può raggiungere (ed anche superare) 20dB rispetto alla condizione normale (MacGougan 2003). Una soluzione per far fronte al problema dell'attenuazione, che non permette il posizionamento in ambienti indoor tramite tecnologie GNSS, è illustrata nel High Sensivity Global Positioning System (HSGPS). Con l'abilità di tracciare segnali deboli, il HSGPS offre maggior versatilità e più ampia applicabilità a confronto del GPS convenzionale. Il segnale è soggetto ad un processamento di guadagno in modo tale da compensare l'eccessiva attenuazione dovuta all'ambiente sfavorevole, permettendo un utilizzo migliore del segnale GPS acquisito (Shanmugan, 2008).

Sebbene HSGPS rappresenti un'interessante possibilità offerta al mercato di massa, l'accuratezza della soluzione è significativamente peggiore di quella ottenibile in una condizione di posizionamento outdoor fornita da un classico sistema GPS, ed in alcuni casi non è nemmeno possibile realizzarla. Le principali ragioni sono dovute alla mancanza di sufficienti satelliti disponibili, errori di multipath, errori associati

all'attenuazione del segnale misurato, problemi di echo e di cross-correlazione nella fase di rilevamento.

Il metodo HSGPS aumenta la possibilità di ottenere una soluzione in condizioni indoor. Tuttavia la forte attenuazione unita a forti fenomeni di interferenza del segnale (multipath ed echo) determinano errori sulla misurazione, che si ripercuotono su errori dello pseudorange (nell'esempio fornito anche di 30 metri, Shanmugan, 2008). Sono quindi necessari sviluppi sull'affidabilità e sull'accuratezza del metodo per far sì che esso risulti utilizzabile. Sebbene il processamento del segnale sia da beneficio per l'accuratezza del posizionamento, è esclusa la possibilità di usufruire del metodo HSGPS per il posizionamento in ambienti in cui i satelliti sono completamente oscurati.

## **2.3 MULTI-GNSS**

La modernizzazione delle già esistenti GPS e GLONASS consiste in parte nella trasmissione da parte dei satelliti di un numero maggiore di frequenze rispetto alla sola doppia frequenza proposta dal GPS e permette di aumentarne la precisione, l'affidabilità e la disponibilità. Inoltre con l'introduzione recente di due nuove costellazioni quali Galileo e BeiDou, è possibile pensare ad una fusione di più sistemi GNSS, aumentando notevolmente il numero di satelliti osservati ottimizzando la distribuzione spaziale dell'osservazione nel sito e quindi incrementando la capacità di convergenza, accuratezza, continuità e di affidabilità del posizionamento. L'unione di questi fattori fa

si che un Multi-GNSS risulti maggiormente applicabile in ambienti limitati come i canyon urbani, fosse aperte o in presenza di bassi valori di multipath.

In Xingxing Li e al. (2015) è proposta la determinazione della posizione di un singolo ricevitore in tempo reale con le informazioni dell'orologio e delle orbite satellitari di tutti i satelliti disponibili di tutte le costellazioni. La procedura di stima dei parametri della soluzione comune viene eseguita rigorosamente con la massima consistenza nel processamento unito di osservazioni provenienti da diversi GNSS.

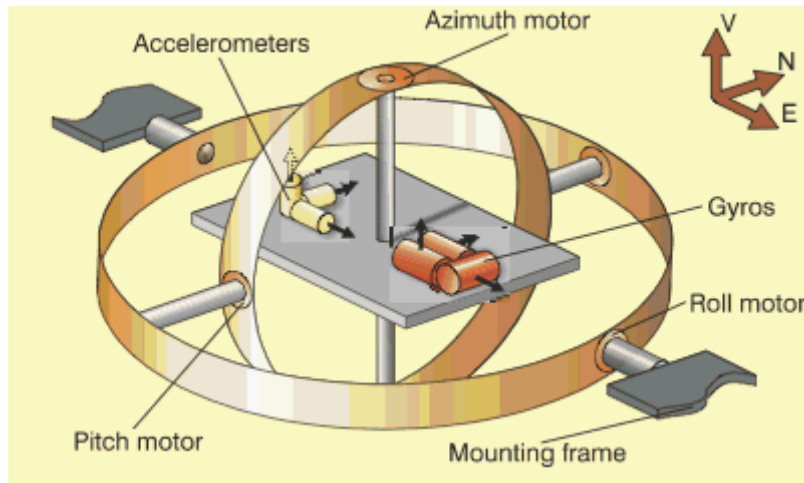
Rispetto all'utilizzo del singolo GPS, i test svolti con il Multi-GNSS hanno mostrato una riduzione del tempo di convergenza del 70% e un aumento del 25% dell'accuratezza. Inoltre l'aumento del numero di satelliti complessivo (da 32 del solo GPS ai 74 delle 4 costellazioni con la prospettiva una volta completato lo sviluppo di Galileo e BeiDou di arrivare a 100) permette anche di individuare ed eliminare gli outlier.

## **2.4 INS (Inertial Navigation System)**

Un sistema INS rappresenta un modello per il calcolo della soluzione del posizionamento cinematico a partire dalla misura del IMU (Inertial Measurement Unit).

L'IMU è uno dispositivo elettronico che permette di misurare l'accelerazione subita dall'oggetto sul quale esso è rigidamente fissato partendo dal concetto delle leggi di Newton. Lo strumento è costituito da una tripletta di accelerometri (perpendicolari tra loro) e giroscopi (per le rotazioni attorno ai tre assi). Lo strumento rappresenta un

sistema inerziale indipendente sensibile alla traslazione e alla rotazione. In *Figura 6* è rappresentato lo schema a tre accelerometri e tre giroscopi che costituiscono un IMU.



*Figura 6 Schema di funzionamento di un IMU (da [www.INS](http://www.INS.it)).*

Lo spostamento e la velocità sono calcolate dall'integrazione rispetto al tempo dell'accelerazione e necessita della condizione iniziale. Per dettagli si osservi De Agostino (2010).

Il sistema INS prevede che la determinazione della soluzione ad un passo temporale sia calcolata utilizzando come condizioni iniziali quelle ottenute dal passo temporale precedente. Un sistema di questo tipo risulta essere fortemente soggetto ad errore di deriva. Come per tutti i sistemi soggetti ad errore di deriva l'errore della soluzione tende ad aumentare all'aumentare dell'intervallo temporale rispetto cui viene calcolata la soluzione del passo successivo. Il sistema GPS non è soggetto ad errori di deriva legati all'ampiezza degli intervalli temporali, ma può fornire una soluzione solo se dispone dei segnali dei satelliti. Le caratteristiche delle soluzioni ottenute dallo strumento INS sono

complementari a quelle che si ottengono da un sistema GPS. Per questa ragione spesso INS e GPS sono integrati per ottenere una soluzione complessiva più robusta. Si veda Hassan e al. (2006).

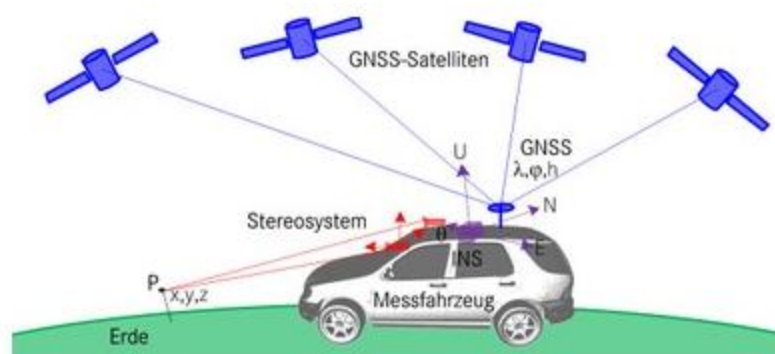
Una recente ed interessante introduzione nel mercato lowcost dei sensori IMU è il MEMS (Micro Electro Measurement System). Questo sensore consiste in una tecnologia costituita da componenti dalle dimensioni che vanno dai 1 e 100 micrometri. Le dimensioni ridottissime rendono il MEMS molto utile per applicazioni in smartphone ect. Le sue performance sono basse e in modalità stand-alone la soluzione è affetta da deriva molto alta. Per ridurre l'effetto della deriva della soluzione è necessario ridurre l'intervallo temporale rispetto cui calcolare la soluzione al passo successivo. Per approfondimenti si veda Troni e Whitcomb (2012).

## **2.5 MMS (Mobile Matting System)**

Il sistemi che prendono il nome di MMS (Mobile Mapping System) consistono nell'unione di un sistema costituito da un ricevitore GNSS con un IMU (ad esempio il sistema INS/GPS). La soluzione del sistema GNSS dipende dalla disponibilità dei satelliti (vedi paragrafo 2.1) mentre quella di un sistema INS dipende dalla durata dell'intervallo di tempo con cui viene calcolata la soluzione. L'integrazione dei dati forniti da un metodo GNSS e uno INS è uno dei metodi comuni per il Mobile Mapping perchè le due tecniche sono complementari e risolvono i problemi reciproci (Pagliari 2015).

Il MMS rappresenta un sistema integrato per il posizionamento ed per l'orientamento esterno della mappatura. Le sue componenti principali sono il sistema di navigazione (INS/GNSS) e i sensori di acquisizione delle immagini, tra i quali possono esserci quelli passivi (camere o video) o quelli attivi come Laser Scanner o LIDAR (Light Detection and Ranging). I sensori per la navigazione permettono di non dover utilizzare Ground Control Point e la bundle-block adjustment per la determinazione dell'orientamento esterno (OE), che appunto viene fornito dal sistema di navigazione INS/GNSS. Il modello base per la georeferenziazione diretta delle immagini è espressa dall'equazione che rappresentana la rototraslazione della soluzione fotogrammetrica rispetto al sistemi di riferimento della soluzione di navigazione fornita da INS/GPS. Per ulteriori dettagli circa l'implementazione matematica del problema si consulti Hassan e al. (2006).

In *Figura 7* è rappresentano lo schema di un MMS applicato a bordo di un'automobile. Si distingue l'antenna GNSS, il sistema INS, il sistema fotogrammetrico per la mappatura ed il punto P (x,y,z) mappato.



**Figura 7** Sistema di MMS applicato ad un'automobile (da [www.MobileMappingSystem.com](http://www.MobileMappingSystem.com)).



Come accennato nel paragrafo 2.4 il sistema INS permette di ottenere soluzioni affidabili in stand-alone solo per brevi intervalli di tempo. Quando è disponibile la soluzione GNSS, cioè quando i satelliti sono disponibili, la soluzione dall'INS viene aggiornata con quella del sistema GNSS. L'integrazione delle soluzioni del sistema integrato INS/GNSS è svolto dal filtro di Kalaman, spiegato nel capitolo (5.5).

Tra questi sistemi le soluzioni più economiche raggiungono i 200.000 dollari (Al-Hamad e El-Sheimu, 2004). I livelli di precisione raggiungibile con questo sistema di acquisizione è sufficiente per la risoluzione di un problema di navigazione urbana, ma i costi della strumentazione sono alti e spesso inaccessibili.

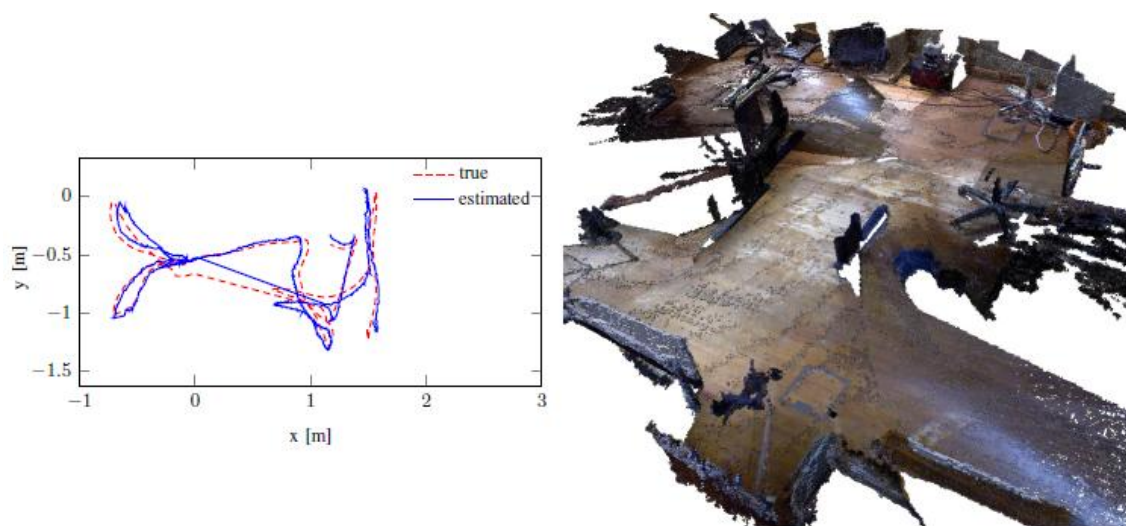
## **2.6 SLAM (Simultaneous Localization and Mapping)**

Lo SLAM rappresenta una sistema standard per la navigazione autonoma di un robot con localizzazione simultanea e la mappatura dell'ambiente circostante. Sono moltissime le applicazioni nella robotica e nella computer vision che necessitano un'acquisizione veloce della realtà 3D circostante al robot e la stima della sua posizione. Questo è un tipico dilemma “dell'uovo e della gallina” perché per localizzare il robot è necessario il modello 3D dell'ambiente, e a sua volta per la creazione del modello 3D dell'ambiente è necessario conoscere la posizione del robot in esso.

Solitamente la soluzione di un problema di SLAM è raggiungibile con l'utilizzo di Laser Scanner e odometri. Si veda in Zhao et al. (2008) il posizionamento per oggetti in movimento con questi strumenti. Tuttavia negli ultimi anni sono stati presentati

interessanti soluzioni con strumenti fotogrammetrici classici. In Bayoud e al. (2004) ad esempio è proposta la mappatura di un ambiente e la localizzazione del sistema tramite due camere CCD ed un sistema INS. Quando la soluzione fotogrammetrica non è sufficiente (per esempio a causa di bassa visibilità) viene utilizzata l'informazione del sistema INS. In questo caso vengono combinati gli output delle due tipologie di metodi tramite il filtro di Kalaman (paragrafo 5.5) per determinare la posizione al passo temporale successivo.

In Felix Endres e al. (2012) viene proposta una soluzione allo SLAM tramite una camera RGB-D, quindi con una camera che unisce informazioni RGB a quelle spaziali Depth, come quella del Microsoft Kinect. In *Figura 8 a)* si osservi la traiettoria calcolata con le immagini RGB-D realizzate da Kinect messa a confronto con la traiettoria di riferimento. *Figura 8 b)* è mostrato il modello 3D realizzato con gli stessi fotogrammi RGB-D.



**Figura 8 a)Traiettoria 2D stimata con Microsoft Kinect e vera. B) Mmodello 3D costruito con immagini RGB-D (da Felix Endres e al. 2012)**

La soluzione dello SLAM deve tener conto della necessità di ottenere una soluzione in tempo reale, quindi la velocità con cui i calcoli vengono risolti deve essere alta. Ovviamente all'aumentare delle dimensioni degli spazi di indagine aumenta anche il carico computazionale. Inoltre non sempre l'ambiente permette l'assunzione di spostamenti piani, il che significa aumentare a 6 i parametri di OE (6D SLAM).

## 2.7 STAZIONE TOTALE

La stazione totale è uno strumento costituito da un teodolite computerizzato per la misurazione degli angoli zenitali e azimutali, ed un elettrodistanziometro per la misurazione delle distanze. L'elettrodistanziometro invia un segnale ad onde e ne misura il tempo di ritorno: da qui viene calcolata la distanza della superficie riflettente (a patto che la distanza sia interna al range dello strumento, solitamente pari a 300m circa).

La stazione totale motorizzata ha come equipaggiamento aggiuntivo una componente meccanica motorizzata e un proiettore di raggio laser. La componente motorizzata fa sì che il raggio laser colpisca il bersaglio riflettore, un particolare prisma, facendo ruotare l'intera stazione totale. Il sistema fondamentale consiste nell'inseguimento a vista del bersaglio mobile. Dalle misure di angoli azimutali, zenitali e di distanza si traccia la traiettoria del prisma nel tempo. Il tracciamento automatico della traiettoria tramite stazione totale richiede che il prisma sia sempre in vista rispetto alla stazione totale.

Dalla posizione della stazione totale si può georeferenziare anche la traiettoria del prisma

Sebbene i risultati ottenuti dalla stazione totale determinino livelli di precisione alta, il suo costo non lo rende adatto al mercato di massa e quindi difficilmente usufruibili a molte applicazioni. Per questo motivo spesso si fa riferimento a questo strumento per la validazione di altre tecnologie lowcost.

Alcuni utilizzi della stazione totale sono riportati in Chiabrando e al. (2013) and Böniger e al. (2010). In *Figura 9* una stazione totale con prisma riflettore.



*Figura 9* Stazione Totale e Prisma (da [www.Total\\_Station.com](http://www.Total_Station.com)) .

## 2.8 NAVIGAZIONE CON STRUMENTI FOTOGRAMMETRICI

Un'altra interessante possibilità per la risoluzione del problema della navigazione in ambienti urbani con strumenti lowcost è rappresentato dalla fotogrammetria. Essa permette di raggiungere livelli di precisione sub-millimetrica per applicazioni di controllo qualità oppure di una decina di metri. Il livello di accuratezza dipende dal tipo di precisione richiesto dall'applicazione ed è gestibile dalla scelta del tipo di camera e obiettivo utilizzato. Da qui si intuisce la grande versatilità della soluzione con metodo fotogrammetrico.

Il recente successo dei sistemi ottici è legato principalmente agli avanzamenti tecnologici e alla miniaturizzazione dei componenti dei sensori Charged Coupled Devices (CCDs) che ne hanno anche abbassato il prezzo. In aggiunta si consideri anche le nuove capacità computazionali e lo sviluppo di algoritmi automatici per il processamento delle immagini. Un altro aspetto importante che ne aumenta la possibilità di impiego è la possibilità di utilizzarlo sia in ambienti indoor che in ambienti outdoor.

In letteratura sono diversi gli esempi in cui si utilizza il metodo fotogrammetrico per la navigazione. In Da Silva e al. (2003) il sistema fotogrammetrico è inserito all'interno del sistema di navigazione del MMS. La traiettoria ottenuta dalla soluzione fotogrammetrica viene utilizzata negli intervalli di tempo in cui è assente il segnale GPS. In Roncella (2005) viene proposta un'interessante soluzione fotogrammetrica. Dal blocco fotogrammetrico vengono estratti punti di legame automatici per eseguire il

bundle-block adjustment da cui ricavare i parametri di OE delle camere. Da essi viene ricostruita la traiettoria del centro di presa della camera.

In Pagliari (2015) è discussa l'effettiva fattibilità dell'utilizzo del metodo fotogrammetrico con strumenti lowcost per la navigazione di un mezzo quando il segnale GPS è assente. Il metodo consiste nella risoluzione del problema fotogrammetrico con un bundle-block adjustment utilizzando come punti di appoggio punti estratti dalla cartografia a grande scala (in scala 1:1000) e come punti di legame punti omologhi estratti automaticamente in dalle immagini acquisite. Test preliminari per la fattibilità del metodo proposto possono essere trovati in Barzagli e al. (2009) mentre test molto simili a situazioni reali si trovano in Cazzaniga e al. (2012) e Cazzaniga et al. (2013).

## **2.9 NAVIGAZIONE CON MICROSOFT KINECT 2**

La presenza di oggetti con tessitura uniforme impedisce l'utilizzo di una soluzione fotogrammetrica pura a sensori passivi. L'informazione dei particolari della tessitura della scena proveniente dalla camera passiva RGB viene completata dall'informazione sulla profondità degli oggetti presenti sulla scena dalla camera attiva, come il laser scanner. In letteratura sono presenti un ampio numero di applicazioni basati sull'uso di immagini RGB e Depth. Si veda ad esempio Oliver et al. (2012) o Endres et al. (2012).

Uno strumento lowcost che utilizza sensori attivi e passivi è il Microsoft Kinect. Questo sensore nasce come periferica per l'utilizzo di Xbox ed il suo costo si aggira intorno alle

poche centinaia di euro. Esso è composto da sensori attivi e passivi che permettono di integrare immagini RGB ad immagini Depth. Questo sensore rientra negli strumenti per il rilevamento degli spostamenti tramite tecnologia a raggi Infrarossi (IR) quindi potrebbe essere usato solo in ambienti indoor. Tuttavia nella sua nuova versione lanciata nel 2013 la tecnologia Time-of\_Flight permette l'utilizzo dello strumento anche in quelli outdoor.

In letteratura sono diversi le soluzioni alla navigazione con Microsoft Kinect (prima e seconda versione). Si veda Pagliari (2015). In Peter Fankhauser e al. (2015) è svolta un'analisi sulla potenzialità dello strumento per la navigazione di un robot. Nel capitolo 3 è approfondita la descrizione delle componenti del Kinect e del dato che esso permette di acquisire.

### 3 SENSORE MICROSOFT KINECT

Il Kinect viene per la prima volta annunciato al pubblico con il nome di Project Natal durante una conferenza stampa della Microsoft nel E3 (Electronic Entertainment Expo) del 2009. Viene chiamato con il suo nome odierno durante la presentazione ufficiale del prodotto svoltasi nel 2010. Il progetto originario del suo hardware parte dal lavoro della compagnia Israeliana PrimeSense, inizialmente finanziata e poi successivamente acquistata da Microsoft nel 2009. Successivamente la licenza della tecnologia viene rilasciata interamente a Microsoft. Il software del Kinect invece è stato sviluppato interamente dal MicrosoftGameStudios, e più precisamente dai programmatori della Rare.

Negli ultimi anni sono diversi i prodotti del mercato dei giochi per console che propongono al pubblico una modalità di gioco che sia controllabile dai movimenti del corpo umano. Tutto questo è possibile grazie ai passi avanti nel mondo della Robotica, dell'Ingegneria e della Computer Vision. Questi studi, combinati con la miniaturizzazione e l'abbassamento di costi dei sensori (quali ad esempio i MEMS – Micro Electro-Mechanical Systems) hanno contribuito ad importanti sviluppi nei campi della modellazione di scene 3D e nel rilevamento dei movimenti come quelli del corpo umano. Il Kinect, a differenza del Wiimote della Nintendo e al PlayStation Move della Sony, è nato dall'idea di fornire al giocatore virtuale un accessorio che fornisca il controllo del sistema senza la necessità di impugnare alcun joystick, ma solo con i movimenti del corpo e con l'utilizzo della voce. Si tratta quindi di un dispositivo



che si sviluppa a cavallo di diversi campi (Computer Graphics, Image Processing Computer Vision and Interaction Uman Machine Interaction) il che lo rende interessante per applicazioni in diversi settori, dall'Industria (giochi, robotica etc.) alla ricerca.

Tra le componenti che caratterizzano il sensore Kinect vi sono una camera RGB, un emettitore InfraRosso (IR) e una camera IR dalle quali si ottengono le immagini RGB e Depth, ovvero immagini di profondità, nelle quali ad ogni pixel è associata una misura di distanza sensore-oggetto. Questo accessorio rappresenta un'assoluta novità perché unisce informazioni di natura complementare come quelle di carattere grafico e quelle di carattere geometrico (Leandro Cruz e al. 2012). Il secondo aspetto che rende, e che ha reso, il Kinect uno strumento eccezionalmente competitivo sul mercato è il basso prezzo di vendita (poche centinaia di euro), caratteristica peculiare per quanto riguarda strumenti fotogrammetrici per la misurazione diretta della profondità.

La capacità del Kinect di acquisire dati di natura diversi permette di costruire una nuova struttura dell'informazione, concepita come l'unione dell'informazione RGB e quella ricavabile dalla Depth. Nel capitolo 5.3.3 verrà analizzata la procedura di ricampionamento adottata per la realizzazione di un'immagine RGB-Depth (RGBD).

### **3.1 KINECT PER XBOX360**

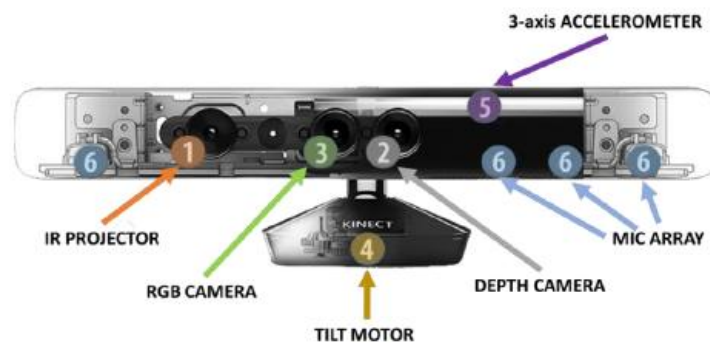
Il Kinect per Xbox360 è stato sviluppato inizialmente dalla compagnia israeliana PrimeSense, poi il progetto viene acquistato dalla Microsoft da cui viene infine concluso e lanciato al pubblico nel 2009. In *Figura 10* è raffigurato il sensore Microsoft

Kinect per Xbox360. Nel 2011 viene rilasciata anche la versione con un SDK (Software Development Kit) per Windows7.



**Figura 10 Sensore (Microsoft) Kinect ( da [www.Microsoft-Kinect.com](http://www.Microsoft-Kinect.com) ).**

Esso è composto da una camera RGB, da una camera IR, un proiettore di pattern IR, un array di microfoni, un motore per l'inclinazione ed un accelerometro a tre assi(dalla figura 2 si osservi la loro distribuzione). Il proiettore IR, la camera IR e la camera RGB sono approssimativamente sulla stesso asse come si nota in *Figura 11*. In *Tabella 1* sono riportate le specifiche tecniche dei sensori ottici del Microsoft Kinect per Xbox360.



**Figura 11 Distribuzione dei sensori che compongono il Microsoft Kinect ( da [www.Microsoft-Kinect.com](http://www.Microsoft-Kinect.com) ).**

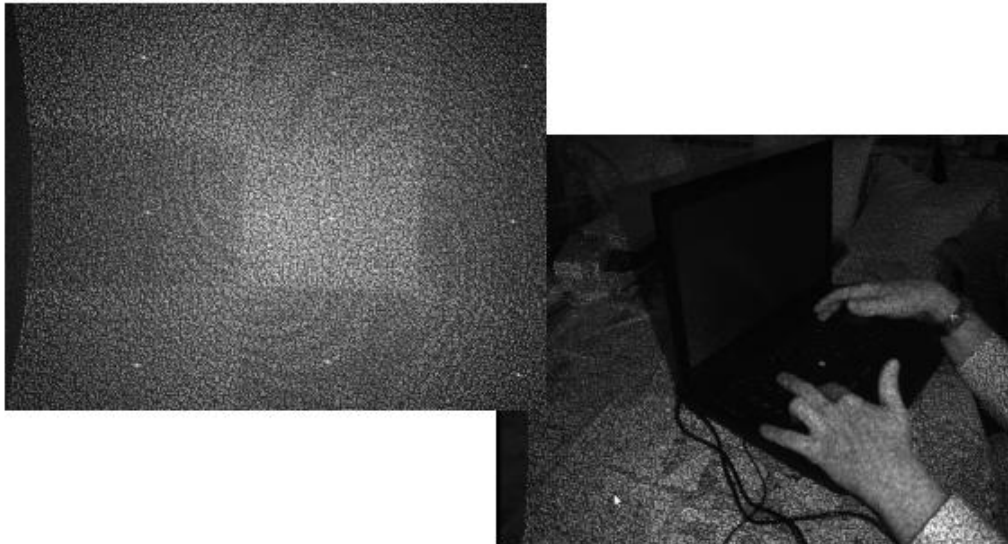
	Camera RGB	camera IR
Tipo di sensore	Aptina MT9M112 CMOS	Aptina MT9M001 CMOS
Dimensione sensore [mm]	3.58x2.87	6.66x5.32
Dimensione Pixel [ $\mu$ m]	2.8	5.2
Formato immagine grezza [pixel]	1280x1024	1200x960
Formato immagine in Output [pixel]	1280x1024 or 640x480	640x480
Lunghezza focale [mm]	2.9	6.0
FOV H [degrees]	63	57
FOV V [degrees]	50	45

**Tabella 1** Specifiche tecniche dei sensori ottici del Microsoft Kinect per Xbox360 (da Pagliari 2015).

Come già anticipato la novità introdotta dal Kinect è l'introduzione della misura di profondità (Depth) tramite strumentazione lowcost. Tale misura è realizzata secondo un principio di triangolazione da un proiettore IR e di una camera IR. Il primo serve a proiettare uno schema di pixel (Pattern) sulla scena, mentre la camera serve per l'acquisizione della scena illuminata dal Pattern proiettato su di essa. La stessa Microsoft dichiara che la versione del Microsoft Kinect per Xbox360 possa acquisire misure di profondità con un range dagli 0.80m ai 4m.

Il Pattern proiettato consiste in un insieme di spot luminosi nell'IR la cui distribuzione a prima vista sembra casuale. Esso è generato facendo passare il fascio di luce nell'infrarosso attraverso un reticolo di diffrazione. In *Figura 12.a)* si osserva l'inquadratura effettuata con camera digitale del Pattern proiettato dall'emettitore IR in

una stanza buia. Per approfondimenti sul pattern proiettato riferirsi a **Leandro Cruz, (2012)**. In *Figura 12.b*) si osserva il Pattern IR proiettato su una scena casuale.



*Figura 12.a) e b) a) Inquadratura con camera digitale del Pattern proiettato in una stanza buia dall'emettitore IR.  
.b) esempio di Pattern IR proiettato su una scena casuale ( da [www.Microsoft-Kinect.com](http://www.Microsoft-Kinect.com)).*

Nella memoria del dispositivo sono salvate una serie di immagini del Pattern di riferimento, acquisite in corrispondenza di una superficie piana parallela al piano dell'immagine e a distanze note. La distanza oggetto sensore viene calcolata confrontando le immagini acquisite con quelle di riferimento. Per ulteriori dettagli circa il funzionamento della prima generazione di sensore Microsoft Kinect si veda Khoshelham (2011).

## 3.2 KINECT PER XBOXONE

La nuova versione del Kinect (che chiameremo Kinect2) è disponibile alla distribuzione in commercio da luglio 2013. All'interno del presente lavoro di tesi è stata utilizzata la versione per Windows di questa nuova generazione di sensori. In *Figura 13* è mostrata la nuova versione del sensore Microsoft Kinect.

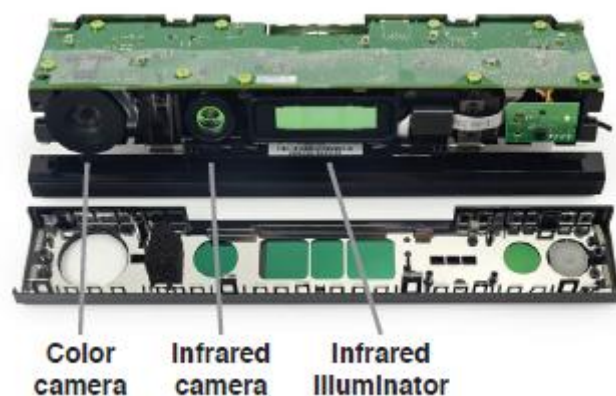


*Figura 13 Microsoft Kinect per XboxOne (www.Microsoft\_Kinect\_One.com).*

Il sensore Kinect2 basa la capacità di acquisizione dell'informazione di profondità sul principio di Time-of-Flight (ToF). Esso è caratterizzato da una maggiore risoluzione e permette di acquisire dati di profondità anche in presenza di luce solare, il che ne amplia il campo di applicazione (estendendolo anche ad ambienti outdoor). Dal confronto tra le performance della versione nuova e vecchia si osserva che per quanto riguarda le immagini Depth realizzate da Kinect2 rispetto a quelle ottenute dalla versione precedente esse sono caratterizzate da maggiore stabilità (Pagliari e Pinto, 2015). Inoltre è stato mostrato che Kinect2 è in grado di acquisire dati anche quando in ombra o colpito da fascio diretto di luce. Si veda Zennaro (2014), Butkiewicz (2014) e Lachat (2015). Tra gli altri aspetti possono essere all'origine di altre fonti d'errore si ricordano

la variabilità dell'informazione Depth in base al colore o alla rugosità del materiale riflettente o all'attraversamento di acqua (etc). Per approfondimenti si rimanda a Lachat e al. (2015). Un interessante studio di Fankhauser (2014) affronta la problematica di come l'errore nella misura delle Depth sia dipendente dall'angolo di incidenza dei raggi di luce.

La nuova versione di Microsoft Kinect è composta dalle stesse tipologie di sensori ottici presenti nella versione precedente a meno del proiettore del Pattern, che è sostituito da un un altro tipo di emettitore IR. La camera RGB è di qualità più alta rispetto alla versione precedente e l'immagine RGB e quella IR non risultano perfettamente sovrapponibili, perchè la prima risulta avere una copertura orizzontale più ampia rispetto a quella IR, la quale a sua volta ha una copertura maggiore in verticale. In *Figura 14* si osservino le componenti del sensore Kinect2 per la generazione di immagini RGBD.



*Figura 14* Sensori di Kinect2 che permettono la generazione di immagini RGBD ([www.Microsoft.Kinect.One.com](http://www.Microsoft.Kinect.One.com)).

L'immagine Depth è realizzata da un sensore immagine CMOS 3D e da un sistema basato sulla tecnologia Time-of-Flight (Sell and O'Connor – 2014). Un fascio di luce (originato dall'emettitore) illumina la scena, quindi gli ostacoli illuminati riflettono il segnale, che viene recepito dalla camera IR. In ogni pixel dell'immagine IR viene immagazzinata l'informazione del tempo di ritorno del segnale (ToF). In una fase di processamento interna allo strumento (paragrafi 3.4 e 3.5) tramite modulazione e phase-detection dell'onda si determina la distanza degli ostacoli.

Un'altra importante caratteristica introdotta nel Kinect2 è la capacità di auto adattamento al tempo di esposizione, che si applica tramite l'acquisizione automatica una di seguito all'altra di due immagini con esposizione diverse e prefissate (paragrafo 3.7). Pixel per pixel sono scelti tra le due immagini quelli a brillantezza maggiore, creando quindi un'immagine più brillante, ma limitando la frequenza di acquisizione dei frame a 15fps. In *Tabella 2* sono riportate le caratteristiche specifiche dei sensori Kinect1 e Kinect2.

	<b>Kinect 1.0</b>	<b>Kinect 2.0</b>
Camera RGB [pixel]	1280x1024 or 640x480	1920x1080
Camera Depth [pixel]	640x480	512x424
Max Depth [m]	4.0	4.5
Min Depth [m]	0.8	0.5
FOV H[degrees]	57	70
FOV V [degrees]	43	60
Motore inclinometro	Yes	No
Definizione giunto scheletro	20	26
Acquisizione completa dello scheletroFull skeleton tracking	2	6

USB	2.0	3.0
prezzo [euro]	80	199

*Tabella 2 Confronto tra caratteristiche della prima e della seconda versione del Kinect.*

### 3.3 LIMITI E OBIETTIVI DI KINECT2

Il sensore Microsoft Kinect sebbene possieda caratteristiche che lo rendono disponibili ad applicazioni in diversi ambiti, nasce come periferica della console Xbox necessaria per il riconoscimento degli utenti/giocatori e dei loro movimenti. Date alcune criticità rivelate nell'utilizzo della prima versione, la nuova versione di Microsoft Kinect è stata progettata con l'obiettivo di migliorare diversi aspetti che permettano di rendere l'esperienza di gioco più realistica. I cambiamenti sono stati resi possibili grazie alla creazione ad opera di Microsoft di una sensore immagine 3D CMOS, grazie all'utilizzo della tecnologia ToF e del SoC (System-on a Chip) che amministra il sensore immagine e la comunicazione con la console XboxOne. In Sell and O'Connor (2014) sono riportate alcune di queste migliorie.

Uno dei limiti della prima versione di Kinect è di non permettere ai bambini di utilizzare il sensore per Xbox a causa dell'impossibilità da parte della tecnologia per l'acquisizione dell'immagine Depth di rilevare gli arti troppo sottili. In questo caso la criticità è rappresentata dalla risoluzione. Il nuovo sensore è stato progettato in modo che la risoluzione della misura di Depth sia almeno di 2.5 cm, misura che rappresenta indicativamente il diametro del polso di un bambino. Questo valore di risoluzione di



Depth è raggiungibile se la risoluzione interna è minore dell'1% (rispetto alla distanza massima). In *Figura 15* si può osservare come l'accuratezza della misura di profondità del nuovo sensore Depth di Kinect2 permette di riconoscere anche particolari del viso o le pieghe dei vestiti.



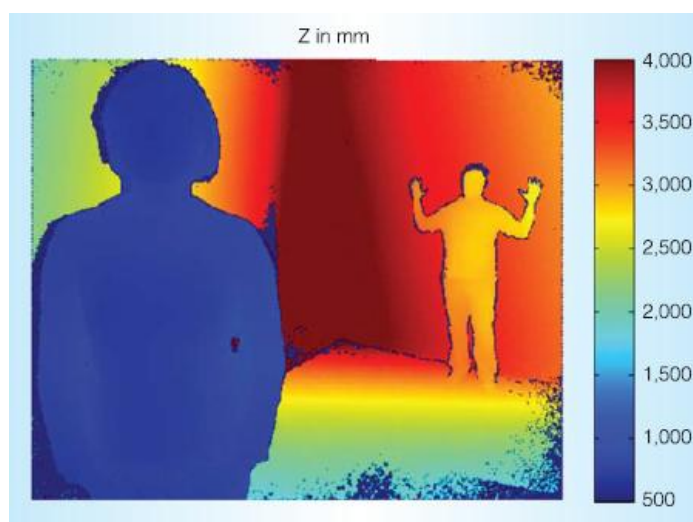
*Figura 15 Immagine Depth catturata a 2.5 m dal soggetto (da Sell and O'Connor , 2014).*

Per quanto riguarda l'operatività alla misura di profondità, il Kinect2 permette il riconoscimento di oggetti all'interno di un range che va dai 0.5m ai 4.50m, più ampio rispetto al range del Kinect1 che va tra gli 0.8m e 4m. Essendo diminuita la distanza minima di acquisizione della misura di profondità, per assicurare una visione sufficientemente ampia della scena nel caso di iterazione con gli utenti a distanze ravvicinate o in spazi ampi, è stato necessario aumentare la Field of View (FoV) delle camere. In

	<b>Kinect 1.0</b>	<b>Kinect 2.0</b>
Camera RGB [pixel]	1280x1024 or 640x480	1920x1080

Camera Depth [pixel]	640x480	512x424
Max Depth [m]	4.0	4.5
Min Depth [m]	0.8	0.5
FOV H[degrees]	57	70
FOV V [degrees]	43	60
Motore inclinometro	Yes	No
Definizione giunto scheletro	20	26
Acquisizione completa dello scheletroFull skeleton tracking	2	6
USB	2.0	3.0
prezzo [euro]	80	199

Tabella 2 sono riportati i rispettivi valori di FoV per il primo ed il secondo sensore.



**Figura 16 Esempio di cattura della figura reso possibile dal range dinamico del nuovo sensore Kinect2 (da Sell and O'Conno , 2014).**

Per far sì che l'utente sia libero di poter indossare indumenti di colore diverso e la sua posizione nella scena (all'interno dei 4.50m) non influisca sul riconoscimento dei suoi movimenti, è stato necessario imporre al nuovo sensore un'esposizione massima di 14ms che permette di ottenere un range di dinamica sufficiente agli scopi sopra esposti.

In *Figura 16* è possibile osservare come le figure siano chiaramente distinguibili a prescindere della loro distanza dal sensore.

In aggiunta grazie alla nuova tecnologia ToF (Time of Flight), spiegata nel paragrafo 3.4, il sensore è indipendente dalla luce presente nella stanza (quindi adatto anche ad applicazione Outdoor).

Altri aspetti curati dal nuovo prodotto sono la latenza più bassa rispetto al vecchio sensore (latenza minore di 20 ms: dalla USB3.0 al software principale del sistema) che permette un'iterazione uomo/giocatore-virtuale più realistica.

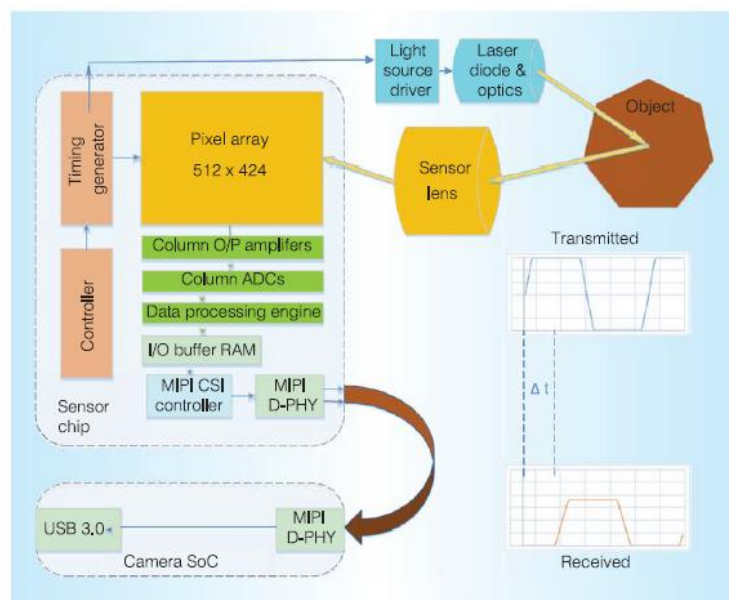
In generale il nuovo sensore possiede maggior stabilità e ripetibilità della misura di Depth e nell'acquisizione delle immagini RGB.

### **3.4 SENSORE ToF**

Il sistema ToF (descritto in *Figura 17*) utilizza un'onda quadra generata da un orologio. Tramite questa onda vengono modulate sia la sorgente luminosa IR (trasmettitore) sia la ricezione da parte dei pixel della camera IR. Il laser diodo (il trasmettitore di luce IR) è usato per illuminare la scena di oggetti. Essi riflettendo la luce e illuminano i pixel dell'immagine IR con una certa intensità luminosa attenuata rispetto a quella generata dall'emettitore. L'informazione salvata in ogni pixel corrisponde all'intensità di luce attenuata da essi ricevuta. Combinando la misura dell'attenuazione (si conosce l'intensità della luce emessa) a quella del ritardo (scift di phase), calcolabile grazie

all'onda quadra che funge da riferimento, si determina  $\Delta t$ . Esso è il tempo impiegato dalla luce per percorrere il tratto trasmettitore/oggetto più quello di riflessione/camera IR. Moltiplicando  $\Delta t$  per la velocità del segnale, nota e pari a 1cm in 33 picosecondi, si determina la misura di profondità.

Il sistema del sensore di acquisizione di immagini 3D rappresentato in *Figura 17* comprende un chip, la camera Soc, il sensore per l'illuminazione ed i sensori ottici.

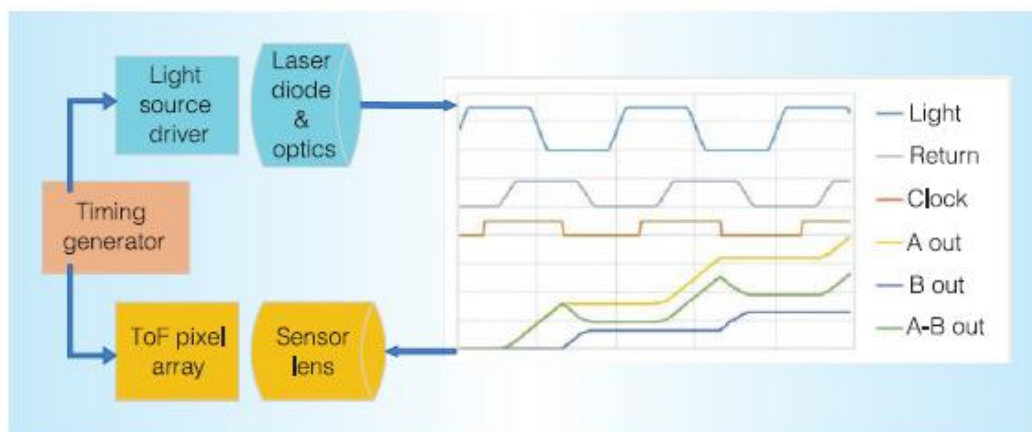


*Figura 17 Sistema del sensore di immagini 3D (da Sell and O'Connor , 2014).*

### 3.5 PIXEL DIFFERENZIALE

L'intensità di luce attenuata rappresenta l'informazione immagazzinata all'interno dei pixel. Tuttavia più che semplici pixel sarebbe più corretto parlare di pixel differenziali.

Infatti ognuno di essi è separato in due accumulatori differenti di informazione. L'orologio controlla quale parte di pixel deve essere utilizzata per immagazzinare l'informazione. Quando l'onda quadrata è positiva allora il pixel immagazzina nella porzione di pixel A, viceversa nella porzione B. In *Figura 18* si osservi le forme di segnale necessarie per la misurazione della profondità tramite tecnologia ToF. "Light" e "return" denotano la composizione modulata della luce trasmessa e ricevuta. "Clock" è il gate per il pixel selezionato. "Aout" e "Bout" sono i voltage output della forma d'onda del pixel selezionati.



*Figura 18 Sensore ToF e forme d'onda del segnale (da Sell and O'Connor, 2014).*

Un sensore a ToF che fa utilizzo dei pixel differenziali si distingue molto da un normale sensore per acquisizione di immagini; il loro utilizzo permette di ricavare differenti immagini che descrivono la scena ripresa:

- (A+B) normale immagine a scala di grigio che dipende dall'illuminazione dell'ambiente (immagine dell'ambiente)
- (A-B) informazione di fase dopo un calcolo di arctang (immagine Depth)

- $\sqrt{(\sum(A - B)^2)}$  immagine in scala di grigio indipendente dall'illuminazione dell'ambiente (immagine attiva)

L'immagine risultante dipende molto da il chip ottico e dai parametri utilizzati, mentre dipende poco da fattori meccanici.

I vantaggi del sensore ToF sono numerosi rispetto al precedente sistema a triangolazione, tra cui i principali sono:

- La risoluzione della Depth è funzione del rapporto segnale rumore e della modulazione di frequenza dove il segnale è la potenza della luce ricevuta e il rumore è la combinazione di quello emesso dal generatore della luce e quello dal sensore elettrico.
- Una frequenza di acquisizione dei frame più alta considerando che il rapporto fase e distanza dipende direttamente dalla modulazione della frequenza.
- Il sistema meccanico e circuitale è semplificato.

La profondità è calcolata a partire dalla misura di shift di fase del segnale modulato tramite la formula

$$2d = \frac{fase}{2\pi} \cdot \frac{c}{f_{mod}}$$

Dove d è la profondità, c è la velocità della luce e  $f_{mod}$  è la frequenza di modulazione e fase rappresenta la fase dell'onda generata.

### 3.6 CALIBRAZIONE DELLE CAMERE RGB e IR

Con l'obiettivo di calcolare i parametri di Orientamento Interno (OI) delle singole camere e di Orientamento Esterno relativo (OEr) tra le camere stesse del sensore Kinect2 è stato necessario eseguire una procedura di calibrazione del sensore. In questo caso si è utilizzato l'applicazione Calibration App presente all'interno del software Matlab.

I risultati ottenuti dalla calibrazione dei parametri di OI e delle relative deviazioni sono riassunti in *Tabella 3*.

Camera name				
	Kinect2 RGB Camera		Kinect2 IR Camera	
Parametri di OI				
	<i>Valore</i>	<i>Dev St.</i>	<i>Valore</i>	<i>Dev St.</i>
Lunghezza focale [Pixel]	1.062e+03	1.572	3.653e+02	0.544
	1.061e+03	1.588	3.655e+02	0.531
Ampiezza immagine [pixels]	1920		512	
Altezza immagine [pixels]	1080		424	
Punto principale coo.x [Pixel]	9.552e+02	0.926	2.546e+02	0.424
Punto principale coo.y [Pixel]	5.351e+02	0.950	2.053e+02	0.414
Parametri aggiuntivi				
K1	3.529e-02	1.972e-03	0.102	2.894e-03
K2	-2.596 e-02	5.613 e-03	-0.234	7.801e-03
P1	-8.049e-04	2.487e-04	-6.429e-04	3.377e-04

P2	1.221 e-03	3.071e-04	0.00218	4.351e-04
----	------------	-----------	---------	-----------

**Tabella 3 Parametri di OI per il sensore Kinect2 stimati dalla calibration app del software Matlab.**

Mentre in *Tabella 4* sono riportati i risultati della calibrazione dei parametri di OE relativo tra le camere IR e RGB.

	X[mm]	Y[mm]	Z[mm]	$\omega$ [gon]	$\varphi$ [gon]	$\kappa$ [gon]
Media	50.438	-1.956	2.771	-6.357e-04	-3.182e-03	-6.799e-03
Dev. St	1.556	0.764	1.250	2.739e-03	3.017e-03	3.247e-03

**Tabella 4 Parametri di OE relativo della camera RGB rispetto alla camera IR del sensore Kinect2.**

La procedura consiste nell'acquisire una serie di immagini muovendo un pannello di calibrazione costituito da una scacchiera bianca e nera (riportato in *Figura 19*). Tale movimento è analogo a lasciare il pannello fermo e ruotare e traslare la camera in fronte allo stesso. Una volta acquisite le immagini i sensori ottici sono stati calibrati utilizzando l'applicazione di Matlab Calibration App: essa riconosce automaticamente i punti gli spigoli della scacchiera sul pannello e calcola i parametri di calibrazione. Per approfondimenti si consulti The Camera Calibration App (Matlab) nei manuali del software Matlab.



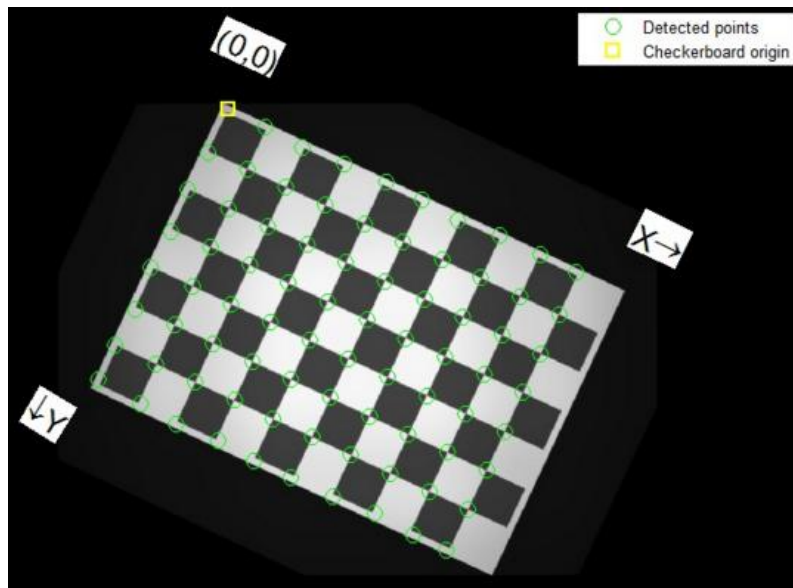


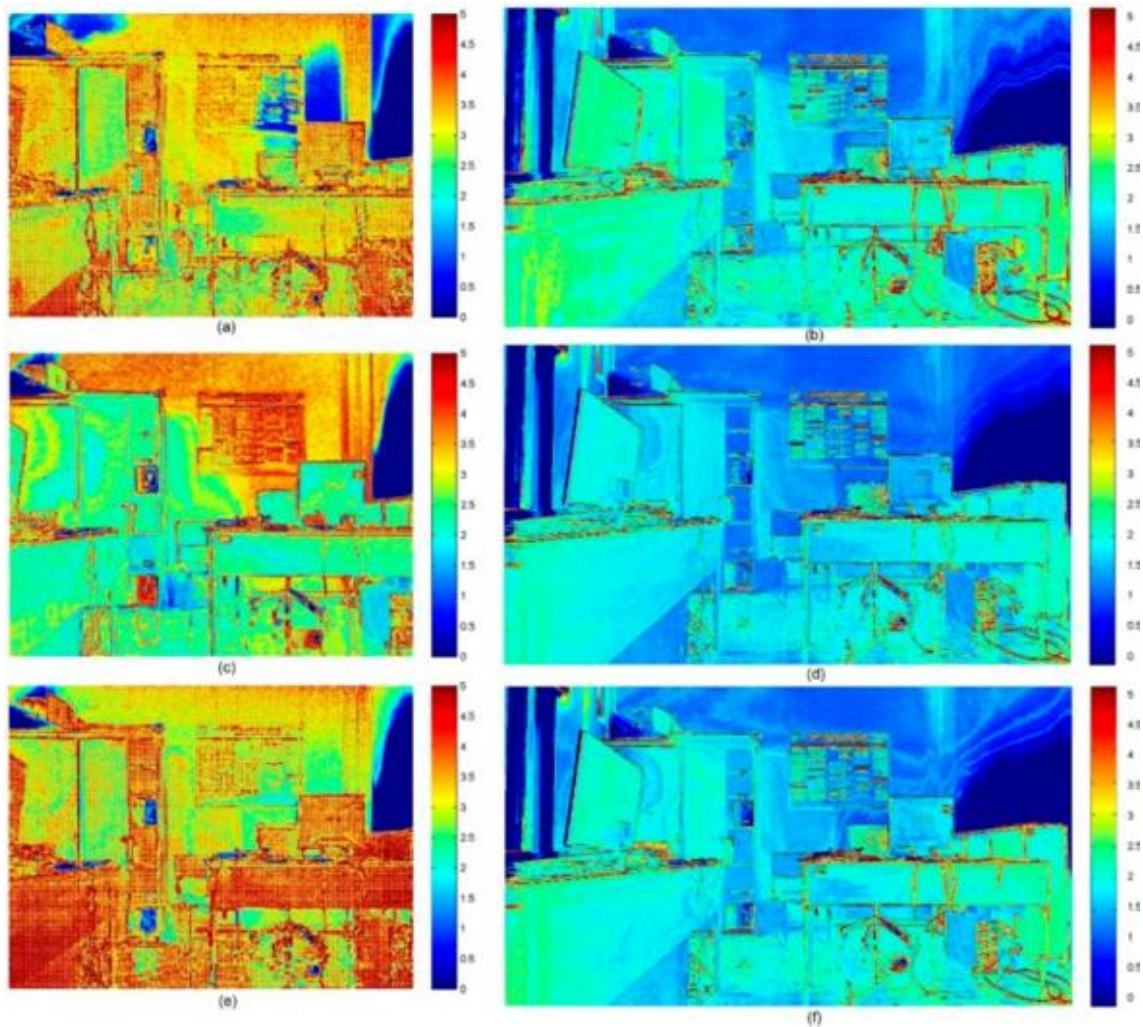
Figura 19 Pannello di calibrazione tramite Calibration App di Matlab ([www.calibration-app-matlab.com](http://www.calibration-app-matlab.com)).

### 3.7 CONFRONTO TRA SENSORI KINECT1 e KINECT2

Lo scopo della seguente tesi è quello di sviluppare un'analisi sulle potenzialità di Kinect2 per la navigazione (in post-processamento) in ambiente urbano (Outdoor). Come già spiegato la tecnologia ToF permette l'utilizzo del sensore Kinect2 anche in ambienti Outdoor, quindi la nostra scelta tra sensore Kinect1 e Kinect2 sarebbe a prescindere obbligata a favore del secondo. Tuttavia è interessante osservare le differenze tra le due tipologie di sensore.

Con lo scopo di valutare le differenze di precisione per quanto riguarda la stabilità dei dati acquisiti dalle due versioni di Kinect si faccia riferimento Pagliari (2015) e Pagliari e Pinto (2015), in cui sono state svolte varie analisi sui sensori immagine.

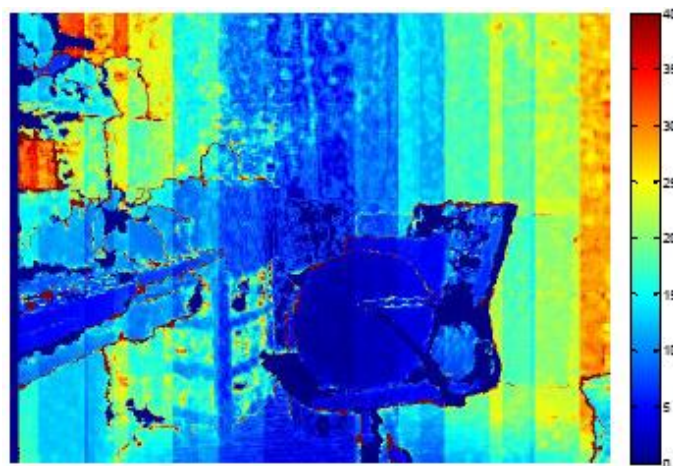
Per prima cosa è stata valutata la stabilità della camera RGB confrontando la deviazione standard nei livelli di intensità registrati in una successione di 100 frame di immagini RGB acquisite rispettivamente con le due versioni del sensore Microsoft Kinect. L'intensità è rappresentata in una scala di colori a 8-bit (256 tonalità di colore).



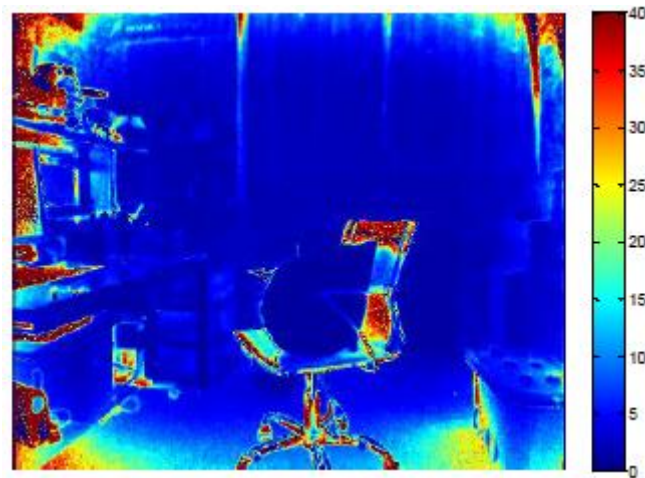
**Figura 20** Deviazione standar misurata tra 100 frame RGB dei sensori Kinect 1(sx) e 2(dx) riferita all'intensità nei canali R,G e B (da Pagliari e Pinto, 2015).

Osservando i risultati riassunti in *Figura 20* risulta evidente che tra i due sensori vi siano variazioni dei valori di intensità, soprattutto nei bordi degli oggetti. Inoltre è interessante notare come in entrambi i sensori il canale del Verde sia soggetto a minor variazioni di intensità rispetto a quello del Rosso e del Blu, che invece si assomigliano. Questo è dovuto al fatto che nel filtro di Bayer gli elementi sensibili alla luce verde sono doppi rispetto a quelli sensibili alla luce blu o rossa.

La stessa analisi è stata ripetuta anche per le immagini Depth, calcolando e confrontando la deviazione standard ottenuta per 100 frame con il sensore Kinect1 e Kinect2.



*Figura 21 Deviazione standard (mm) calcolata per ogni pixel dell'immagine raw Depth acquisita dal sensore Kinect1 (da Pagliari e Pinto, 2015).*

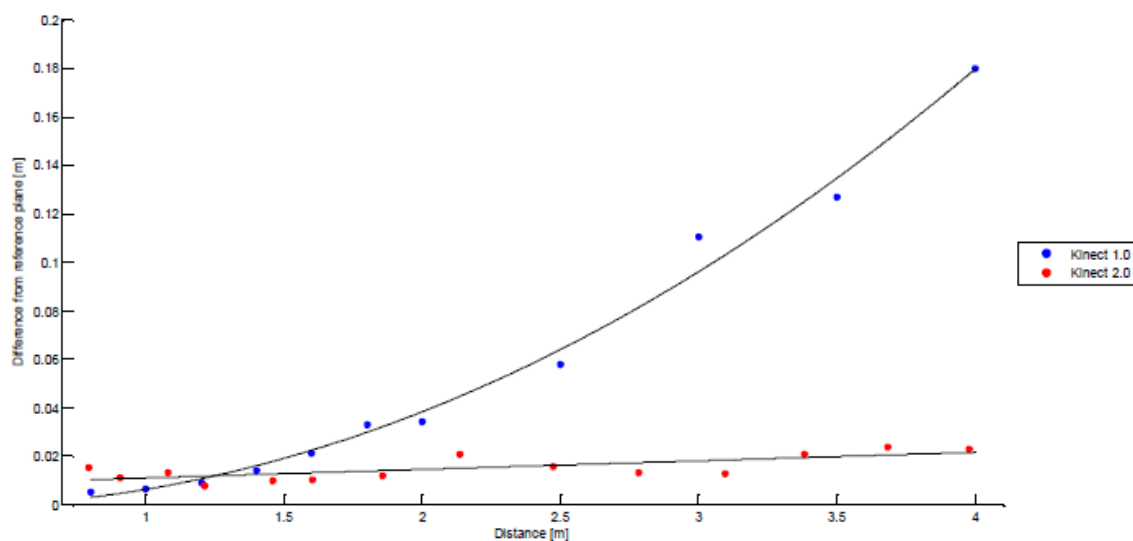


**Figura 22** Deviazione standard (mm) calcolata per ogni pixel dell'immagine raw Depth acquisita dal sensore Kinect2 (da Pagliari e Pinto, 2015).

In *Figura 21* sono riportati i risultati ottenuti dall'analisi svolta per le immagini Depth del sensore Kinect1, mentre in *Figura 22* quelli derivanti dal sensore Kinect2. Oltre a constatare una stabilità molto maggiore per quanto riguarda la misura Depth ottenuta dal Kinect2 rispetto a quella ottenuta dal Kinect1, si nota per l'ultima versione un'interessante fenomeno radiale.

Dopo aver analizzato la stabilità della misura di profondità è opportuno soffermarsi anche sull'aspetto riguardante l'errore associato alla sua misura. Con lo scopo di valutare l'errore commesso dalle due versioni del Kinect quando impiegate per la misura di distanza, sono stati svolti una serie di test. In particolare è stata svolta una analisi statistica selezionando una regione centrata dell'immagine, corrispondente ad una porzione di muro (ovvero la superficie piana scelta come riferimento), in modo tale da limitare gli effetti ai bordi immagine.

Data una successione di 100 immagini per ogni pixel sono calcolate valore medio e deviazione standard. Inoltre è stata calcolata la deviazione standard rispetto al piano interpolato determinato per minimi quadrati. Considerando la distanza rispetto al piano interpolato, e ipotizzando effetti di rotazione nulli, è stata valutata la deviazione rispetto alle misure ottenute dalle due versioni di Kinect.



**Figura 23** Stima dell'errore commesso da Kinect1 e 2 come funzione della distanza tra camera ed oggetto (da Pagliari, 2015).

In *Figura 23* sono rappresentate le stime delle deviazioni rispetto alle misure ottenute dalle due versioni di Kinect. Per il Kinect1 l'espressione analitica che descrive l'errore commesso nel misurare le distanze nell'ipotesi pseudo-nadirale dipende dal quadrato della distanza (Krauss,1997), mentre nel caso del sensore Kinect2 (che sfrutta la tecnologia ToF) è di tipo lineare.

Il risultato è che all'aumentare della distanza dell'oggetto l'errore sull'accuratezza della seconda versione di Kinect è molto minore rispetto a quello della prima versione.

## **4 POSIZIONAMENTO CINEMATICO CON TECNICHE FOTOGRAMMETRICHE**

La fotogrammetria è una disciplina, presente da molti anni, il cui utilizzo mira all'investigazione della realtà 3D utilizzando delle immagini, cioè informazioni sviluppate in un piano 2D. Essa è una tecnica che permette di ricostruire metricamente oggetti, senza richiedere alcun tipo di contatto, a condizione che l'oggetto indagato sia ripreso da almeno due punti di vista differenti, garantendo un certo grado di ricoprimento. Infatti, solo per i punti oggetto visibili su due o più immagini è possibile risalire alla loro posizione tridimensionale. Le relazioni che intercorrono tra realtà 3D e immagini 2D acquisite dai sensori ottici sono rappresentate dalle ben note equazioni di Collinearità (paragrafo 4.1) che si basano sul principio di proiezione centrale (paragrafo 4.1).

Uno dei motivi per cui negli ultimi anni la fotogrammetria è stata rivalutata come una delle discipline chiave per la soluzione del posizionamento degli oggetti è la disponibilità di processori sufficientemente potenti e di algoritmi innovativi che permettono di risolvere un problema di Triangolazione Aerea (paragrafo 4.3) anche per blocchi estesi di immagini. Un esempio delle nuove capacità espresse dagli innovativi algoritmi di calcolo sono rappresentate dalla possibilità di sostituire la ricerca manuale di punti di legame tra le immagini con una ricerca automatica (matching automatico) che ovviamente diminuisce moltissimo i costi ed i tempi di lavoro, aumentando però

allo stesso tempo il numero di punti misurabili, tanto che l'oggetto può essere ricostruito da una nube densa di punti che ne descrive molto bene la superficie.

L'obiettivo della fotogrammetria consiste nel fornire informazioni metriche della realtà indagata. La misura è caratterizzata, come tutte le misure di grandezze reali, da un'incertezza la quale in prima analisi dipende dalla qualità dei sensori ottici utilizzati (tipicamente dalla camera). In particolare le lenti della camera determinano una distorsione delle immagini e di conseguenza anche delle misure sui modelli 3D da esse realizzati. La distorsione è descritta da modelli caratterizzati da parametri caratteristici (paragrafo 4.2) stimabili da un processo di calibrazione. Dalle equazioni del modello si risale alla correzione da applicare per eliminare parzialmente l'errore della distorsione delle lenti.

Solitamente la fotogrammetria viene utilizzata per risolvere il problema diretto, ovvero ricostruire geometricamente un oggetto in modo tridimensionale. Essa però può anche essere utilizzata in modo inverso quando lo scopo è calcolare una soluzione di navigazione, ovvero determinare la traiettoria descritta da una camera in moto nel tempo.

## **4.1 MODELLO PROIETTIVO DIRETTO E INVERSO IN FOTOGRAMMETRIA (Equazioni di Collinearità)**

Alla base del modello proiettivo diretto ed inverso della fotogrammetria sta il Principio di Collinearità secondo cui un punto oggetto  $(X,Y,Z)$ , il centro di presa  $O$



dell'immagine e il punto  $(\xi, \eta)$ , proiezione del punto oggetto sul piano immagine  $\Omega$ , giacciono sulla stessa retta. Esprimendo la trasformazione che porta dal punto immagine  $(\xi, \eta)$  nel sistema di riferimento immagine al punto oggetto  $(X, Y, Z)$  espresso rispetto al sistema di riferimento oggetto otteniamo l'Equazione 1

**Equazione 1**

$$\begin{bmatrix} \xi \\ \eta \\ \zeta \end{bmatrix} = \lambda R \begin{bmatrix} X - X_0 \\ Y - Y_0 \\ Z - Z_0 \end{bmatrix}$$

Dove  $\lambda$  è il fattore di scala,  $\underline{X}_0 = [X_0 \ Y_0 \ Z_0]^t$  sono le coordinate del centro di presa O della camera.  $R$  è la matrice di rotazione descritta dai tre angoli di cardano  $\omega$ ,  $\varphi$  e  $\kappa$  ed è determinata dal prodotto delle tre matrici di rotazione  $R_\omega$ ,  $R_\varphi$  e  $R_\kappa$  definite come

$$R_\omega = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos \omega & \sin \omega \\ 0 & -\sin \omega & \cos \omega \end{bmatrix}, R_\varphi = \begin{bmatrix} \cos \varphi & 0 & -\sin \varphi \\ 0 & 1 & 0 \\ \sin \varphi & 0 & \cos \varphi \end{bmatrix}, R_\kappa = \begin{bmatrix} \cos \kappa & -\sin \kappa & 0 \\ \sin \kappa & \cos \kappa & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Da cui definiamo infine la matrice di rotazione  $R$  come

$$R = \begin{bmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{bmatrix}$$

La proiezione del centro di presa O sul piano immagine non coincide con il centro fiduciale dell'immagine e rappresenta il Punto Principale (PP) di coordinate  $(\xi_0, \eta_0)$ . Per come è definito il sistema di riferimento immagine, la coordinata  $\zeta$  del centro di presa O coincide con la focale della camera  $c$  (vedi Figura 24). Le grandezze  $\xi_0, \eta_0$  e  $c$  sono chiamate parametri di Orientamento Interno (OI) e sono grandezze caratteristiche



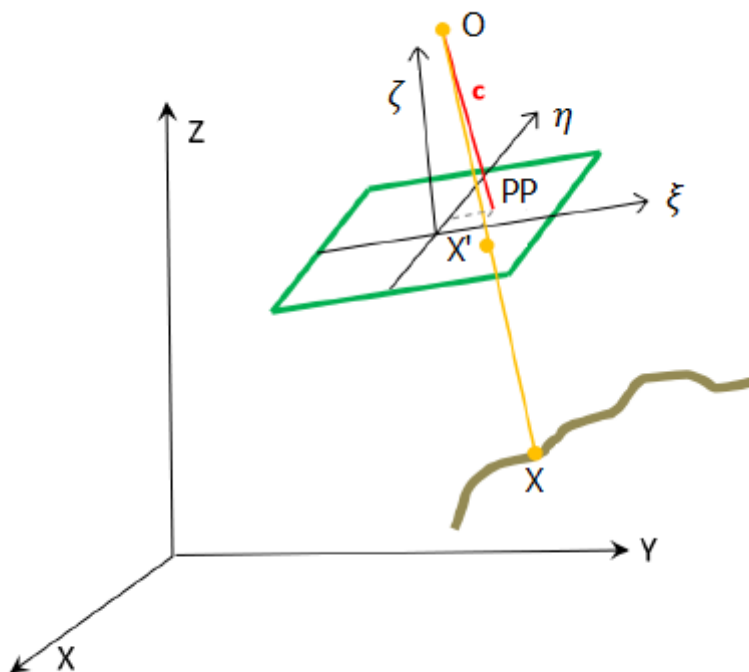
della camera, determinabili da processo di calibrazione (paragrafo 0) (Krauss - 2008, Fraser - 2005).

Tenendo conto di questo offset l'Equazione 1 diventa

**Equazione 2**

$$\begin{bmatrix} \xi - \xi_0 \\ \eta - \eta_0 \\ -c \end{bmatrix} = \lambda R \begin{bmatrix} X - X_0 \\ Y - Y_0 \\ Z - Z_0 \end{bmatrix}$$

In *Figura 24* viene rappresentato lo schema del problema fotogrammetrico rispetto al sistema di assi definito.



**Figura 24** Relazioni tra immagine e sistema oggetto.

L'Equazione 2 può essere riscritta come

*Equazione 3*

$$\begin{cases} \xi - \xi_0 = \lambda[r_{11}(X - X_0) + r_{12}(Y - Y_0) + r_{13}(Z - Z_0)] \\ \eta - \eta_0 = \lambda[r_{21}(X - X_0) + r_{22}(Y - Y_0) + r_{23}(Z - Z_0)] \\ -c = \lambda[r_{31}(X - X_0) + r_{32}(Y - Y_0) + r_{33}(Z - Z_0)] \end{cases}$$

Da cui dividendo per la terza equazione del sistema dell'*Equazione 3* è possibile eliminare il parametro del fattore di scala. L'*Equazione 3* diventa quindi

*Equazione 4*

$$\begin{cases} \xi = \xi_0 - c \frac{r_{11}(X - X_0) + r_{12}(Y - Y_0) + r_{13}(Z - Z_0)}{r_{31}(X - X_0) + r_{32}(Y - Y_0) + r_{33}(Z - Z_0)} \\ \eta = \eta_0 - c \frac{r_{21}(X - X_0) + r_{22}(Y - Y_0) + r_{23}(Z - Z_0)}{r_{31}(X - X_0) + r_{32}(Y - Y_0) + r_{33}(Z - Z_0)} \end{cases}$$

Le equazioni espresse in *Equazione 4* prendono il nome di equazioni di Collinearità Diretta e portano dalle coordinate immagine a quelle oggetto. Esprimendo l'*Equazione 4* rispetto X e Y si ottengono le equazioni di Collinearità inversa. Le incognite sono i parametri di OI (Orientamento Interno) e quelli di OE (Orientamento Esterno).

Le equazioni di Collinearità possono venire espresse anche rispetto alla matrice K di calibrazione in un'espressione che più facilmente viene impiegata nel settore informatico.

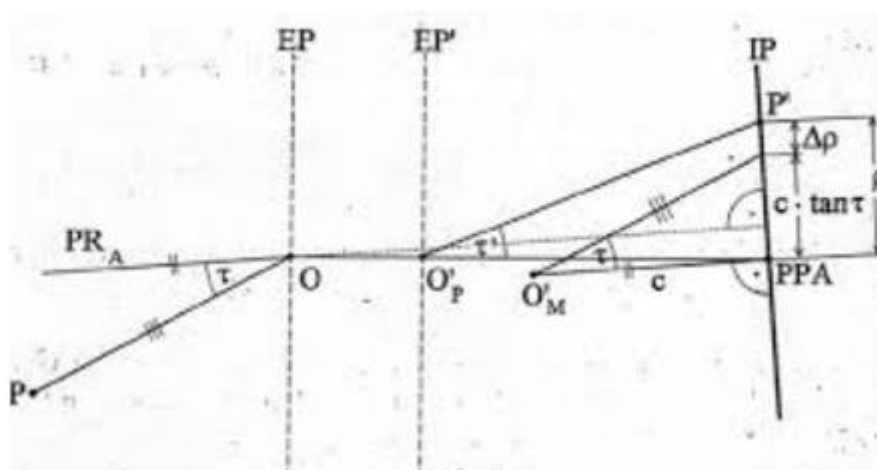
*Equazione 5*

$$\begin{bmatrix} \xi \\ \eta \\ \zeta \end{bmatrix} = \begin{bmatrix} c & 0 & p_x \\ 0 & c & p_y \\ 0 & 0 & 1 \end{bmatrix} [R, t] = K[R, t]$$

## 4.2 MODELLO DI CALIBRAZIONE DELLE CAMERE

Come già accennato la fotogrammetria rappresenta un utile strumento per la misurazione di grandezze metriche su modelli della realtà 3D generati da immagini 2D.

Quindi il processo con cui viene ricreato il modello 3D della realtà (processo di restituzione fotogrammetrica) acquisisce un'importanza fondamentale in termini di precisioni della misurazione finale. Per rendere ottimale l'accuratezza del processo di restituzione è necessario disporre dei parametri di calibrazione della camera utilizzata per l'acquisizione delle immagini del rilievo. Come per ogni elemento di fabbricazione umana anche le camere sono soggette ad imperfezioni dell'ottica. Oltre a questi errori casuali si aggiungono errori sistematici. Dal modello reale del obiettivo di una camera si sa infatti che il Principio di Collinearità non è rispettato con esattezza. Esiste quindi una violazione delle equazioni di Collinearità che determinano errori residui sistematici. In *Figura 25* si veda lo schema reale di un obiettivo e la relativa violazione al Principio di Collinearità. Nelle immagini questo errore determina il noto effetto “bombato” ai bordi dell'immagine. L'errore sistematico determina la distorsione delle immagini e può essere modellizzato. Per dettagli sull'aspetto geometrico e matematico della violazione del Principio di Collinearità si rimanda a Brown (1966) e Brown (1987).



*Figura 25 Modello reale dell'obiettivo e relativa violazione al Principio di Collinearità  
(www.PrincipiodiCollinearità.com).*

I parametri di interesse sono quelli di OI a cui si aggiungono quelli che modellano le distorsioni delle immagini, cioè i tre coefficienti dell'equazione di distorsione radiale ( $K_1, K_2$  e  $K_3$ ) e i due coefficienti dell'equazione della distorsione tangenziale ( $p_1, p_2$ ) per un totale di 8 parametri (in alcuni casi si utilizzano 10 parametri aggiungendo al modello anche i parametri  $b_1$  e  $b_2$  che modellizzano le distorsioni di taglio). (Brown (1966) e Brown (1987)).

Il modello della distorsione radiale può essere modellizzato utilizzando un'equazione polinomiale del tipo

$$\delta\rho = K_1\rho^3 + K_2\rho^5 + e K_3\rho^7$$

Dove  $\delta\rho$  è la distanza di un punto immagine generico  $(\xi, \eta)$  dal PP  $(\xi_0, \eta_0)$

$$\rho = \sqrt{(\xi - \xi_0)^2 + (\eta - \eta_0)^2}$$

Quindi si esprime la distorsione radiale nelle direzioni del sistema di riferimento immagine  $\xi$  e  $\eta$

$$\delta\xi = (\xi - \xi_0) \frac{\delta\rho}{\rho} \quad \delta\eta = (\eta - \eta_0) \frac{\delta\rho}{\rho}$$

La distorsione tangenziale è definita dal disallineamento delle lenti dall'asse ottico  $\Delta\xi$  e  $\Delta\eta$  calcolato come

$$\Delta\xi = p_1[\rho^2 + 2(\xi - \xi_0)^2] + 2p_2(\xi - \xi_0)(\eta - \eta_0)$$

$$\Delta\eta = p_2[\rho^2 + 2(\eta - \eta_0)^2] + 2p_1(\xi - \xi_0)(\eta - \eta_0)$$

Le equazioni di Collinearità espresse nella forme dell'*Equazione 2*, introducendo anche le componenti di correzione delle distorsioni delle immagini diventa

*Equazione 6*

$$\begin{bmatrix} \xi - \xi_0 + \delta\xi + \Delta\xi \\ \eta - \eta_0 + \delta\eta + \Delta\eta \\ -c \end{bmatrix} = \lambda R \begin{bmatrix} X - X_0 \\ Y - Y_0 \\ Z - Z_0 \end{bmatrix}$$

Dall'Le equazioni di Collinearità espresse nella forme dell'*Equazione 2*, introducendo anche le componenti di correzione delle distorsioni delle immagini diventa

*Equazione 6* si osserva che la correzione delle equazioni che modellano la deformazioni dell'immagine è applicata nella forma di una traslazione della posizione della coordinate immagine misurate. Si può esprimere la correzione della coordinata rispetto ad un fattore di distorsione  $L(\rho)$  dipendente solo dalla distanza radiale

$$\xi_{dif} = \xi_c + L(\rho)(\xi - \xi_c) \quad \eta_{dif} = \eta_c + L(\rho)(\eta - \eta_c)$$

dove  $(\xi_{dif}, \eta_{dif})$  sono le coordinate corrette,  $(\xi_c, \eta_c)$  il centro di distorsione radiale e  $(\xi, \eta)$  le coordinate misurate.

I parametri di OI e i coefficienti delle equazioni che descrivono il modello di distorsione delle lenti sono calcolate tramite un processamento per Minimi Quadrati. Il modello presentato nell'Le equazioni di Collinearità espresse nella forme dell'*Equazione 2*, introducendo anche le componenti di correzione delle distorsioni delle immagini diventa *Equazione 6* non è lineare quindi per risolvere i Minimi Quadrati serve un'inizializzazione, che solitamente è rappresentata da una soluzione approssimata.

Il processo ai Minimi Quadrati è calcolato dalle soluzioni ottenute da una serie di immagini convergenti acquisite nell'intorno di un opportuno pannello di calibrazione

(fornito dal produttore dell'algoritmo di calibrazione). Per minimizzare gli effetti di correlazione è necessario ruotare la camera attorno al suo centro immagine di angoli di  $\mp 90^\circ$ .

### **4.3 TRIANGOLAZIONE AEREA (Bundle block adjustment)**

Per risolvere il problema della restituzione o della presa da un blocco di immagini 2D è necessario conoscere la posizione del centro di presa della camera e l'orientazione delle immagini nell'istante di presa. Per ogni immagine si hanno 6 incognite di cui tre sono le componenti traslazionali che definiscono la posizione del centro di presa della camera rispetto al sistema di riferimento, e tre sono gli angoli che definiscono la matrice di rotazione  $R$ , cioè l'orientamento dell'immagine rispetto al sistema di riferimento. I sei parametri appena descritti prendono il nome di parametri di Orientamento Esterno (OE). Il processo fotogrammetrico noto come Triangolazione Aerea (TA) (bundle block adjustment) consiste nella stima simultanea dei parametri di OE e delle coordinate dei punti di legame nel sistema di riferimento oggetto di un gruppo di immagini che riprende la scena di indagine. Per il calcolo dell'allineamento delle immagini si richiede che tra esse vi sia una buona percentuale di ricoprimento necessaria per il riconoscimento di un numero sufficiente di punti di legame.

Per georeferenziare il blocco fotogrammetrico di immagini è necessario disporre anche di un numero minimo di GCP (Ground Control Point) che siano visibili nelle immagini. I GCP sono punti di cui si conoscono le coordinate oggetto. Solitamente sono punti acquisiti con sistemi quali antenne GPS a doppia frequenza o stazione totale. Per

georeferenziare un blocco fotogrammetrico è necessario determinare i 7 parametri della rototraslazione con fattore di scala, quindi servono almeno 3 GCP. Ovviamente un numero maggiore migliora la ridondanza della soluzione dei parametri di OE delle immagini.

Il processo di risoluzione di un problema di TA permette di orientare il blocco di immagini partendo dal fotogramma per arrivare direttamente all'oggetto, basandosi sulle equazioni di Collinearità. Il loro utilizzo diretto, senza il passaggio creazione di un modello, rende la TA un metodo rigoroso di ricostruzione della realtà fisica dell'oggetto. La TA ricerca la migliore combinazione dei parametri di OE delle immagini che minimizza l'errore di proiezione dei punti dal piano immagine al sistema oggetto. L'errore è la somma degli scarti degli errori di posa di ogni punto ed è espresso da una funzione non lineare. La minimizzazione dell'errore, che rappresenta la soluzione del problema di TA, è svolta tramite implementazione di algoritmo non lineare di Minimi Quadrati. La non linearità del problema da affrontare richiede la necessità di una soluzione approssimata per inizializzare il problema. La soluzione approssimata del problema può essere calcolata ad esempio utilizzando la DLT (Direct Linear Transformation) (Abdel Aziz e Karara, 1971).

Considerato un blocco di  $n$  immagini per la soluzione del problema fotogrammetrico di orientamento delle immagini è necessario un numero sufficiente di punti di legame per fissare tra loro le immagini e GCP (Ground Control Point) a sufficienza per georeferenziare il blocco fotogrammetrico. Punti di legame e GCP vanno forniti con la loro precisione. Aggiungere informazioni anche da strumenti di posizionamento diversi (come sistema GNSS o sistemi integrati GNSS/INS etc) migliora il bilanciamento tra

incognite ed equazioni, aumentando la ridondanza e quindi la precisione della soluzione.

Considerate  $n$  immagini da orientare e  $m$  punti oggetto 3D l'obiettivo del processo di Triangolazione Aerea consiste nel minimizzare la distanza tra proiezione sul piano immagine dei punti oggetto 3D e i relativi punti immagine. Nella forma di equazione si ha

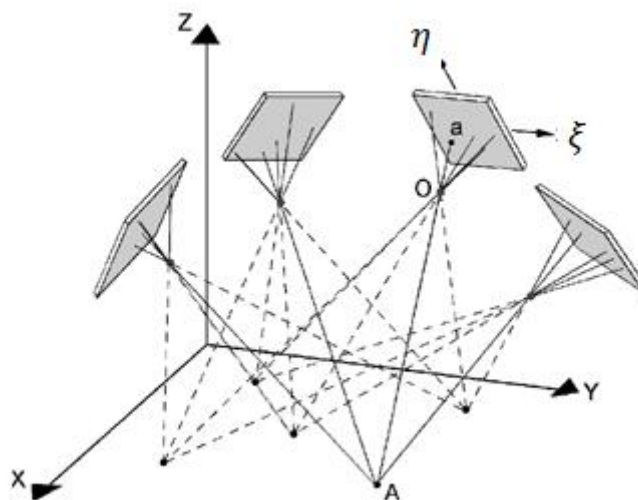
*Equazione 7*

$$\min \sum_{i=1}^n \sum_{j=1}^m w_{ij} \|q_{ij} - P_i(C_i, X_i)\|^2$$

dove  $q_{ij}$  è il punto immagine osservato,  $P_i$  la matrice di rotazione,  $C_i$  i parametri della camera dell'immagine e  $X_i$  il generico punto oggetto 3D.

In Figura 26 è presentato un semplice esempio di Triangolazione Aerea di un blocco fotogrammetrico di 4 immagini. Si osservi che le 4 immagini sono distribuite attorno ad una stessa scena, a rimarcare la necessità che tra le immagini ci sia buona sovrapposizione per la ricerca dei punti di legame. Sempre nella *Figura 26* a terra si distinguono i punti di legami fissati (ad esempio il punto A).





*Figura 26 Risustato di un processo di Triangolazione aerea per un blocco di 4 immagini(www.bundleblockadjustment.com)*

Per dettagli sulla Triangolazione Aerea si consulti Krauss (2008).

## 4.4 ROTOTRASLAZIONI

Per la fotogrammetria che fa utilizzo di camere passive la creazione di scene 3D, cioè la profondità, è realizzabile disponendo necessariamente di almeno 2 immagini che riprendano la stessa scena. Questo limite coincide con il principio su cui si basa la Stereoscopia. Il Kinect invece, disponendo di un sensore passivo e di uno attivo, permette di acquisire della stessa scena sia un'immagine RGB che un'immagine Depth. Dalla calibrazione delle camere del sensore Kinect (eseguibile nella fase preliminare di un rilievo, vedere paragrafo 5.1.1) vengono determinati i parametri di OI delle camere RGB e IR. In particolare quelli della camera IR permettono di ricostruire la nuvola di

punti della scena 3D rispetto al sistema di riferimento della camera IR che ha per origine il suo centro di presa, asse x definito come la larghezza del sensore IR, asse z uscente dal sensore e asse y a chiudere la terna destrorsa. Quindi da una singola immagine Depth si dispone già della ricostruzione 3D della scena ripresa. Un'immagine RGBD (paragrafo generazione immagine RGBD 5.3.5) consiste in una immagine RGB alla quale è associata l'informazione della nuvole di punti generata dall'immagine Depth (paragrafo proiezione e generazione PC 5.3.4).

Il metodo di navigazione proposto in questa tesi consiste nel tracciare la traiettoria dei centri di presa da cui è acquisito il blocco di immagini RGB e Depth concatenando tra loro coppie di immagini RGBD consecutive (paragrafo metodo di posizionamento fotogrammetrico 4). Con il termine "concatenare" coppie di immagini RGBD si intende il loro allineamento nello spazio, cioè la ricostruzione della loro posizione e del loro assetto nell'istante di presa. In altre parole i loro parametri di OE.

Per ogni coppia di immagini RGBD il sistema di riferimento secondo cui definire il loro orientamento è fissato nel centro di presa della prima immagine (secondo l'ordine del tempo della loro acquisizione), con asse x diretto come la larghezza del sensore, asse y definita come l'altezza del sensore e asse z uscente dal piano immagine, a chiudere la terna destrorsa. In questo modo mancano da definire solo i parametri di OE della seconda immagine. Questi possono essere ricavati dalla matrice di rototraslazione che permette di concatenare la nuvola di punti 3D della prima immagine RGBD con quella della seconda immagine RGBD.

Una rototraslazione  $Rt$  è definita come una trasformazione rigida che permette di sovrapporre le due nuvole di punti delle due immagini RGBD. Essa è espressa come

$$Rt = \begin{bmatrix} R & t \\ 0 & 1 \end{bmatrix}$$

in cui  $R$  è la matrice 3x3 che definisce la rotazione tra le due immagini

$$R = \begin{bmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{bmatrix}$$

e  $t$  è il vettore 3x1 che definisce la traslazione tra i centri di presa delle due immagini

$$t = [t_x \quad t_y \quad t_z]^T$$

Nel paragrafo 5.3.6 vi è la spiegazione nel dettaglio del processo a cui è soggetta ogni immagine per poterne calcolare la rototraslazione rispetto alla precedente. Le rototraslazioni sono determinate dall'algoritmo Align2RGBD del Princeton Vision (paragrafo 5.3.6).

Infine, considerando che ogni immagine RGBD è collegata alla precedente secondo una rototraslazione, per determinare la traiettoria dell'intero blocco fotogrammetrico acquisito da Kinect è necessario applicare la trasformazione

$$X'_i = \prod_i^{i=1} Rt_i^{i+1} X'_{i+1}$$

dove  $X'_i$  è la coordinata del centro di presa dell'immagine  $i$ -esima del blocco,  $X'_{i+1}$  quella  $i$ -esima+1 dello stesso blocco, e  $Rt_i^{i+1}$  la rototraslazione dalla nuvola di punti dell'immagine  $i$ -esima alla nuvola di punti dell'immagine  $i$ -esima+1. Si tenga presente

che la trasformazione va applicata al contrario, cioè partendo dalla rototraslazione tra la coppia di immagini  $(i,i+1)$  fino alla rototraslazione tra la coppia  $(1,2)$ .

## 5 METODO FOTOGRAMMETRICO PROPOSTO

Come è stato già discusso del capitolo (4) la fotogrammetria rappresenta un'interessante soluzione al problema del posizionamento e della navigazione a basso costo sia in ambienti indoor che in ambienti outdoor. Una caratteristica peculiare delle immagini acquisite con camere RGB è proprio quella di permettere di distinguere una grande densità di feature sulla base delle variazioni cromatiche. Tuttavia quando la superficie rappresentata dalle immagini risulta a tessitura uniforme la ricerca di punti di legame utilizzando processi di matching automatico, e in alcuni casi anche quello manuale, è difficile. In questo contesto sarebbe impossibile risolvere la Triangolazione Aerea. Inoltre la soluzione puramente fotogrammetrica necessita di qualche informazione metrica per scalare la soluzione stessa, quali ad esempio la base di presa nel caso di una bicamera o le coordinate di alcuni punti d'appoggio. L'utilizzo di punti noti permette poi anche di correggere eventuali effetti di deriva.

Molto spesso, però, le sole camere passive non sono sufficienti a risolvere un problema di posizionamento e risulta evidente la necessità di integrare un sensore attivo in grado di acquisire informazioni metriche spaziali, come le immagini Depth. In questa tesi un approccio analogo è stato adottato con il sensore Microsoft Kinect. La presenza in esso di camere passive e camere attive rende questo strumento più versatile grazie ad un pacchetto informativo più robusto rispetto ad un sensore passivo normale, perchè combina tra loro informazioni di tipo complementare.

L'intento di questa tesi è di utilizzare le immagini ottenute dal sensore Kinect per supportare la navigazione con un sistema GPS in tratti in cui il suo segnale non risulti oscurato dagli ostacoli presenti in ambiente urbano.

Negli ultimi anni sono stati proposti diversi algoritmi automatici per la ricerca di feature tra immagini le quali condividono una sufficiente percentuale della stessa scena. Tra questi troviamo SIFT (Lowe, 2004) (paragrafo 5.3.6.1), l'operatore sicuramente più utilizzato grazie alle ottime prestazioni in numerose applicazioni. Nell'ottica dello scopo della tesi si è preferito post-processare i dati, preferendo una soluzione che dia più importanza all'accuratezza finale piuttosto che una soluzione calcolata in modo speditivo per la navigazione in tempo reale.

Per il calcolo delle rototraslazioni tra coppie di immagini RGBD il metodo proposto presuppone dell'utilizzo dell'algoritmo distribuito in download libero con licenza opensource dalla Princeton Vision `align2RGBD`, il quale implementa al suo interno anche SIFT (paragrafo 5.3.6).

Prima di lanciare la funzione `align2RGBD` è stato necessario eseguire diverse fasi quali: accoppiamento delle immagini RGB e Depth, il miglioramento della loro qualità tramite eliminazione delle distorsioni ottiche e processo di ricampionamento delle immagini Depth. Quindi si passa alle fasi di selezione delle coppie con il sottocampionamento dell'intero blocco ed infine la generazione delle immagini RGBD. Si veda nel dettaglio nel paragrafo 5.3.

Lo scopo del progetto di tesi è verificare la possibilità di effettuare un posizionamento cinematico tramite l'utilizzo di un sistema integrato fotogrammetrico/GPS

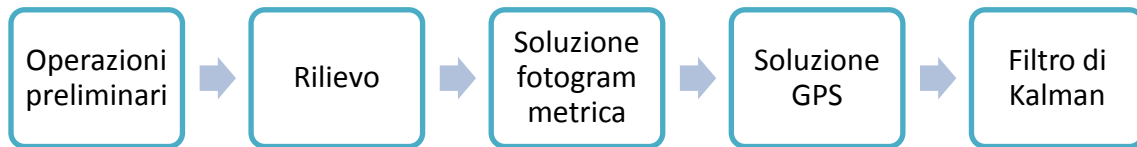
rigorosamente costituito da sensori lowcost. La componente fotogrammetrica del sistema è costituita dal sensore Microsoft Kinect la quale è collegata rigidamente alla componente GPS, l'U-blox (paragrafo 5.2.2). Il metodo proposto nel progetto di tesi consiste nell'integrazione della traiettoria calcolata dal sensore Kinect con quella determinata dall'U-blox utilizzando filtro di Kalman (paragrafo 5.5).

Nel rilievo effettuato sono stati previsti anche l'utilizzo di un'antenna GPS a doppia frequenza e di una stazione totale per l'inseguimento di un prisma. Sia antenna GPS a doppia frequenza che prisma per l'inseguimento sono montati rigidamente alla struttura del sistema integrato Kinect/GPS. Le traiettorie fornite da GPS a doppia frequenza e da stazione totale rappresentano un riferimento per stabilire la qualità del posizionamento fornito dal sistema integrato Kinect/U-blox.

L'integrazione della traiettoria calcolata dal Kinect con quella determinata da U-blox è possibile grazie alla sincronizzazione dei tempi degli strumenti Kinect e U-blox (paragrafo 5.1.2): per entrambi gli strumenti è infatti noto anche il tempo PC al quale è stato acquisito il dato.

Tra le operazioni preliminari riguardanti il rilievo è stata eseguita una calibrazione geometrica per determinare la posizione relativa tra i sensori che compongono il sistema integrato. Da questo processo di calibrazione si determinano le rototraslazioni che permettono di integrare le soluzioni del Kinect, dell'U-blox e di confrontare le soluzioni dell'inseguimento della stazione totale e dell'antenna GPS a doppia frequenza

(paragrafo 5.1.3). In *Figura 27* è riportato lo schema di flusso del metodo proposto.



*Figura 27 Schema di flusso del metodo integrato proposto.*

## 5.1 OPERAZIONI PRELIMINARI

Le operazioni preliminari consistono nella fase di calibrazione delle camere del sensore Kinect, nella fase di calibrazione geometrica del sistema di strumenti montati sul carrello nel momento del rilievo ed infine nella loro sincronizzazione temporale, necessaria come riferimento dei dati acquisiti.

### 5.1.1 Calibrazione delle camere

Con calibrazione delle camere del sensore del sensore Microsoft Kinect2 si intende nello specifico la determinazione dei parametri di OI della camera RGB e della camera IR e i parametri di OE relativo tra le camere stesse. A supporto di questa operazione si è utilizzato l'applicazione Calibration App presente all'interno del software Matlab<sup>®</sup>. Dalla stessa applicazione viene fornita un'immagine del pannello di calibrazione da stampare su carta necessario al processo (figura *Figura 19*). I dati per effettuare la calibrazione consistono in una serie di immagini del pannello da caricare all'interno del



software Matlab<sup>®</sup>. Caricando l'eseguibile dell'applicazione Calibration App viene lanciato il calcolo dei parametri il cui valore risultante è riportato nelle tabelle *Tabella 3* *Tabella 4*. Per ulteriori dettagli della fase di calibrazione delle camere del sensore Kinect consultare il paragrafo 3.

### 5.1.2 Sincronizzazione degli strumenti

Il sensore Kinect acquisisce immagini Depth e le relative immagini RGB contemporaneamente, ma il processo di salvataggio nel laptop utilizzato determina una latenza nell'ordine di qualche millisecondo [6 o 7 ms]. Considerato che dall'utilizzo del sensore Kinect ci si aspetta di ottenere un'accuratezza del posizionamento nell'ordine di una decina di centimetri, la latenza tra acquisizione di un'immagine Depth e RGB può essere trascurata. Le immagini vengono salvate con un nome uguale a quello del tempo interno dello strumento (espresso in millisecondi) nel quale vengono acquisite ed è facile associare le coppie di immagini Depth e RGB ordinandole rispetto ad esso. Per convenzione alle immagini RGBD generate dalle coppie Depth ed RGB viene associato il tempo di acquisizione dell'immagine Depth. Di conseguenza anche alle coordinate dei centri di presa determinati dall'allineamento delle immagini RGBD è associato il tempo dell'immagine Depth. Ai punti della traiettoria determinati dall'antenna GPS a doppia frequenza e dall' U-blox invece vengono associati i tempi del GPS.

Per poter applicare il filtro di Kalman alle soluzioni della traiettoria calcolata dal Kinect e dall'U-blox è necessario che alle coordinate dei punti delle due traiettorie sia associato lo stesso tempo. Tuttavia esiste un possibile legame dato dal fatto che sia U-blox che Kinect salvano il dato sul medesimo laptop. Tramite il software di acquisizione delle

immagini del Kinect viene scritto un file di testo nel quale sono riportati anche i tempi PC di acquisizione. Parallelamente un secondo software registra il tempo dei dati del Ublox il tempo del laptop in cui il dato è salvato. Durante la fase elaborazione dei dati è stato associato sia alle immagini RGBD che alle soluzioni di U-blox il corrispondente tempo del laptop, che in definitiva è quello utilizzato per l'integrazione con filtro di Kalman.

### **5.1.3 Calibrazione geometrica del sistema di sensori**

Il metodo proposto considera la traiettoria tracciata da quattro tipi diversi di sensori: il sensore Kinect, il GPS U-blox, la stazione totale e l'antenna GPS a doppia frequenza. Delle soluzioni determinate da questi sensori le due traiettorie determinate dal sensore Kinect e quella determinata dal GPS U-blox vengono integrate tramite Filtro di Kalman. Le traiettorie determinate con antenna GPS a doppia frequenza e quella con la stazione totale invece servono da riferimento per valutare la precisione della soluzione del sistema integrato Kinect/U-blox.

Per poter integrare le soluzioni del Kinect e dell'U-blox e per confrontare i risultati ottenuti con le traiettorie ricavate dalla stazione totale e dall'antenna GPS a doppia frequenza è necessario conoscere le rototraslazioni che portano il sistema di riferimento rispetto a cui è definita la traiettoria del primo sensore nel sistema di riferimento in cui è ricavata la soluzione dell'altro sensore. Questo è possibile solo se, e per noi è verificato (vedere capitolo 6.1), tutti i sensori sono rigidamente fissati ad un sistema solido per tutta la durata dell'acquisizione dei dati.

In sostanza il problema della calibrazione geometrica del sistema di sensori montati rigidamente tra di loro consiste nella determinazione delle rototraslazioni tra le diverse soluzioni. Si consideri che una rototraslazione necessita di almeno 3 punti 3D per determinare i parametri della trasformazione che porta da un sistema di riferimento all'altro, quindi con 4 punti si dispone di sufficiente ridondanza per la risoluzione del sistema di incognite.

Il metodo adottato per la determinazione della trasformazione consiste nel rilievo di un quadrato di calibrazione (4 punti). Una volta installati rigidamente tutti i sensori sul carrello (spiegazione della modalità del rilievo nel paragrafo 5.2) con esso si disegna un quadrato ipotetico muovendo il carrellino a velocità bassa e facendo in modo che esso non cambi mai direzione e verso. In questo modo le immagini acquisite dalle camere del sensore Kinect riprendono un'alta percentuale di scena identica facilitando il riconoscimento di feature e quindi di punti di legame tra immagini. Per ogni spigolo del quadrato invece si sosta qualche minuto per effettuare un posizionamento statico con il GPS U-blox e con l'antenna GPS a doppia frequenza.

La soluzione di U-blox e antenna GPS a doppia frequenza è fornita rispetto al sistema globale WGS84. Lo stesso vale per la soluzione tracciata con la stazione totale. In questo caso si risale alla traiettoria del prisma (per i dettagli di come vengono tracciate le traiettorie nel caso della stazione totale vedere paragrafo 2.7) in sistema WGS84 perchè la stazione totale è stata posta su apposito treppiede precedentemente georeferenziato nel sistema di riferimento WGS84 tramite rilievo statico con l'antenna

GPS a doppia frequenza del punto di stazione stesso e di un secondo punto di orientamento

La soluzione della traiettoria dei centri di presa fornita dai dati acquisiti dal sensore Kinect in fase di calibrazione è determinata per risoluzione del problema di Triangolazione Aerea (TA) dal blocco di immagini RGB acquisite negli istanti in cui il carrellino è in sosta nei vertici del quadrato (si veda paragrafo 5.1.3). Vengono utilizzati come GCP alcuni punti su appositi target posti in un oggetto 3D presente sulla scena delle immagini. Le coordinate in sistema di riferimento WGS84 dei GCP si conoscono grazie al posizionamento effettuato con la stazione totale su treppiede georeferenziato. Se nelle 4 immagini prese dagli spigoli non fossero ben visibili i target presenti sull'oggetto 3D si può selezionare qualche foto presa lungo i lati del quadrato e aiutare la risoluzione del problema di TA vincolando qualche punto di legame tra le immagini. L'individuazione dei GCP sui target e dei punti di legame, è praticabile utilizzando il software Agisoft) Photoscan ([www.Photoscan.com](http://www.Photoscan.com)) da cui è possibile lanciare la soluzione della TA. Infine anche i centri di presa delle immagini prese dagli spigoli del quadrato sono in sistema di riferimento WGS84.

I parametri della trasformazione tra le traiettorie dei centri dei sensori viene calcolata utilizzando la funzione di Matlab *Helmert3D.m* (Wasmeier, 2016). Successivamente si utilizza la funzione di Matlab *d3trafo.m* (Wasmeier, 2016) per applicare la trasformazione calcolata sopra alle traiettorie dei centri di presa delle immagini RGB del Kinect. Arrivati a questo punto è possibile confrontare le coordinate dei vertici del

quadrato e calcolare la deviazione standard dei parametri di traslazione della trasformazione determinata.

## 5.2 RILIEVO

Il metodo proposto presuppone l'utilizzo di strumenti fotogrammetrici in grado di acquisire dati anche in ambienti outdoor (paragrafo 3.4). Lo strumento utilizzato, il sensore Microsoft Kinect, acquisisce immagini Depth e RGB, quindi si predilige la presenza elementi geometrici 3D di dimensione maggiore della risoluzione minima della camera Depth. Durante la fase di progettazione del percorso del rilievo si faccia presente che il sensore Depth ha un range di acquisizione tra i 0,50m e i 4,50m.

La soluzione fotogrammetrica va integrata, almeno a tratti, con la soluzione da strumenti GPS, quindi si presuppone che l'ambiente sia sufficientemente aperto per assicurare la disponibilità di un numero sufficiente di satelliti ad effettuare un posizionamento cinematico. L'utilizzo della stazione totale per l'inseguimento del prisma impone che per gran parte della traiettoria vi sia una linea diretta di visione tra prisma e stazione totale.

Come spiegato nel paragrafo 5.1.3 affinché vi sia una corrispondenza biunivoca tra le traiettorie determinate dai sistemi Kinect, dall'U-blox, dall'antenna GPS a doppia frequenza e dalla stazione totale (riferendosi sempre al prisma) è necessario che tutti e quattro gli strumenti siano rigidamente fissati e quindi soggetti a vincolo per impedirne movimento relativo (cioè rispetto agli altri sensori) indipendente. Solo sotto questa

ipotesi è possibile utilizzare le trasformazioni determinate dal processo di calibrazione geometrica del sistema di sensori per ogni punto del rilievo. Nel caso in cui si integrino i dati del Kinect con quelli di una antenna GPS posta a sbraccio rispetto al device Microsoft, sarebbe necessario stimare il vettore 3D che collega i due sensori. Esso è costante nel sistema di riferimento camera (il Kinect) e deve essere rototraslato ad ogni istante considerando l'orientamento della camera stessa. In prima approssimazione questo problema non è stato considerato all'interno del presente lavoro, poiché l'antenna u-blox (i cui dati saranno integrati nel filtro di Kalman) è collocata sopra la camera RGB e l'eventuale sbraccio tra i due sensori è inferiore alle precisioni dei sensori stessi.

Utilizzando il supporto rigido di una palina con opportune staffe per il fissaggio è possibile vincolare gli strumenti affinché essi non subiscano movimenti relativi indipendenti. Il prisma, elemento a cui fa riferimento la stazione totale per il tracciamento ad inseguimento della traiettoria, va fissato tramite l'utilizzo di staffe nel punto più alto disponibile della palina. La posizione elevata fa sì che sia limitata la possibilità che vi sia un'interruzione del campo visivo tra stazione totale e prisma. L'antenna GPS a doppia frequenza (antenna Leica gs14) va fissata nell'apposita vite di fissaggio presente al di sopra del prisma (Accessori Geosystems Leica) e risulta essere lo strumento più alto dell'intero sistema di strumenti. In questo modo si assicura la miglior visibilità possibile dei satelliti disponibili, evitando il possibile oscuramento di persone o cose nella vicinanza. Si richiede inoltre che l'antenna sia disposta più orizzontalmente

possibile, condizione necessaria per non oscurare porzioni di cielo e quindi di possibili satelliti in vista.

Riferendoci agli assi del sistema di riferimento interno del sensore Microsoft Kinect, che ha centro nel centro di presa della camera RGB, il sensore Kinect appunto va posizionato con asse x orizzontale, in modo che camera Depth e RGB siano approssimativamente alla stessa altezza. Inoltre si inclina lo strumento in avanti in modo che l'asse z non sia orizzontale, ma inclinato verso il basso, cioè entrante alla superficie del terreno con angolo di circa  $30^\circ$  o  $40^\circ$ . Quest'ultimo accorgimento fa sì che nelle immagini acquisite sia presente oltre agli oggetti 3D che sorgono verticalmente, anche parte del terreno, che nel caso di superficie a tessitura non uniforme (nel nostro test è presente del ciottolato capitolo 6.1) risulta ideale sia per la ricerca di feature automatico per differenze cromatiche (algoritmo SIFT). Inoltre, nel caso nella scena ripresa non ci sia presenza di alcun elemento 3D, si assicura la disponibilità assicurata della metà inferiore delle immagine per la ricerca di punti di legame tramite algoritmo SIFT, condizione non assicurata invece nel caso di asse z orizzontale. In questa ottica si fa in modo che il Kinect abbia una distanza dal terreno inferiore ai 4 metri. facendo attenzione a scegliere una distanza dal terreno sufficiente per avere un campo di visuale ampio. Si consiglia di scegliere zone in cui l'asfalto non sia a tessitura uniforme.

Il sensore U-blox è una antenna planare, motivo per cui esso deve essere posto il più orizzontale possibile. Esso è stato fissato sulla superficie del case al disopra della camera RGB del Kinect, cioè in prossimità del centro di presa scelto come origine del sistema di riferimento interno della soluzione fotogrammetrica. Del nastro adesivo è

sufficiente allo scopo dato il peso ridotto del sensore U-blox. Essendo anch'esso un sensore per la ricezione del segnale dei satelliti GPS è necessario seguire le stesse indicazioni che valgono per l'antenna a doppia frequenza. Considerata l'inclinazione dell'asse z e quindi del Kinect stesso, si dispone un elemento al disotto del U-blox per renderlo orizzontale. Per quanto riguarda l'altezza la condizione di fissaggio accoppiata con il Kinect è soddisfacente. Per evitare effetti di oscuramento dall'antenna a doppia frequenza (posta nel punto più alto della palina) si consiglia l'utilizzo di una seconda palina distanziata dalla prima. In questo caso però bisogna assicurarsi che entrambe le paline siano fissate rigidamente al mezzo di trasporto ed impedito anche alla rotazione. Nel caso fosse impossibile disporre dell'attrezzatura per fissare gli strumenti in modo rigido è accettato accontentarsi di una sola palina. L'U-blox è alimentato direttamente dal laptop, mentre il Kinect richiede di essere alimentato tramite batteria; essi devono essere collegati allo stesso laptop, sia per acquisire i dati sia per sincronizzarne i tempi.

Il rilievo verrà effettuato per il tracciamento di una soluzione di posizionamento cinematico a velocità costante. Quindi è necessario disporre di un mezzo con ruote che sia possibile trascinare comodamente lungo il percorso almeno in direzione rettilinea, preferibilmente anche in direzione perpendicolare. Esso deve presentare anche elementi verticali su cui fissare rigidamente la palina e spazio sufficiente per disporre batteria e laptop.

Per il rilievo sono necessari dei target stampati su foglietti di carta da bloccare con nastro adesivo sugli elementi verticali 3D in modo tale da non coprire la trama



stampata. La posizione deve essere tale perchè siano visibili dalla presa delle immagini Kinect e visibili dalla posizione su cui verrà posta la stazione totale. Da questi target verranno estratti i GCP posizionati tramite stazione totale necessari per risolvere il problema della TA del blocco di immagini del quadrato di calibrazione (paragrafo 5.1.3). La fase di progettazione del rilievo del quadrato per la calibrazione geometrica del sistema di strumenti deve tener conto della necessità che i target siano visibili dal maggior parte di posizioni del quadrato da cui vengono acquisite le immagini. Impostare la traiettoria del rilievo in modo tale che il primo spigolo del quadrato sia il punto iniziale e finale dell'intero giro ed che il primo lato del quadrato corrisponda al primo tratto del rettilineo del giro.

Infine si consideri qualche accorgimento necessario per effettuare un rilievo le cui soluzione siano confacenti alle ipotesi assunte nel nostro filtro di Kalman. Come verrà spiegato nel paragrafo 5.5 il filtro di Kalman che utilizzeremo pone l'ipotesi di moto a velocità costante. Durante la fase del rilievo quindi si consiglia di mantenere una velocità del carrellino costante. Per quanto riguarda i punti in cui è necessario effettuare delle curve si preferisce che queste, nel limite del possibile, siano a raggio ampio e sempre a velocità costante.

### **5.2.1 Acquisizione delle immagini**

Le immagini acquisite dal sensore Kinect sono immagazzinate nella memoria del laptop collegato allo strumento durante la fase di rilievo. Le immagini sono salvate con nome

identificativo del tempo interno di acquisizione dello strumento (paragrafo 5.1.2 per dettagli su latenza e scelte adottate)..

Per evitare la perdita di immagini necessarie al calcolo della traiettoria causa eventi imprevedibili, quali attraversamento improvviso di soggetti, si consiglia l'acquisizione di un numero elevato di frame per secondo. In questo modo nel caso servisse è possibile eliminare la striscia di immagini inadatte senza compromettere la riuscita del progetto. Di fatto questo accorgimento è già insito nelle modalità di acquisizioni del Kinect: esso permette di acquisire fino a 30 frame per secondo, anche se di fatto è stato verificato come con un laptop di fascia medio-alta alimentato a batteria sia in grado di acquisire 5 fps, dovendo gestire sia il salvataggio dei dati GPS che di quelli acquisiti dal Kinect.

### **5.2.2 Traiettoria da dati GPS e da stazione totale ad inseguimento**

Il posizionamento con tecniche GPS e tecniche fotogrammetriche hanno caratteristiche molto diverse tra loro che è necessario tener conto per la progettazione di un rilievo che possa ottenere risultati soddisfacenti.

Il rilievo con sensore Kinect necessita della presenza di elementi geometrici 3D dove siano evidenti delle variazioni di profondità, in modo che sia possibile applicare una ricerca della trasformazione dall'informazione Depth delle due immagini RGBD (paragrafo 5.3.6). Per quanto riguarda la linea di inseguimento del prisma dalla stazione totale, si consideri che oltre alla sua traiettoria è tracciata anche quella dell'antenna GPS a doppia frequenza, quindi, sebbene il rilievo presenti alcuni tratti di linea interruzione

dell'inseguimento del prisma (comunque preferibilmente da evitare) si dispone comunque della soluzione della traiettoria riferimento.

Per effettuare un posizionamento statico e per fissare la posizione iniziale di un posizionamento cinematico tramite sistemi GPS è necessario sostare fermi per qualche minuto. In questo arco di tempo, il sensore Kinect acquisisce un numero altissimo di immagini tra loro identiche ed inutili al fine della determinazione della navigazione. Questo va a scapito dell'uso efficace della memoria del laptop, aspetto questo aggravato dal fatto che la camera del nuovo sensore Kinect acquisisce immagini RGB ad alta definizione, quindi piuttosto ingombranti (circa 6/7 MB per immagine).

## **5.3 SOLUZIONE FOTOGRAMMETRICA**

L'obiettivo del progetto di tesi consiste nel posizionamento cinematico con un sistema fotogrammetrico a supporto di un sistema GPS che a causa del contesto urbano può essere soggetto a errore di cycle slip o addirittura di mancanza totale del segnale dei satelliti (paragrafo 2.1). Il metodo fotogrammetrico proposto quindi serve per risolvere un problema inverso rispetto a quello classico di restituzione di una realtà 3D. L'obiettivo quindi consiste nella determinazione delle coordinate dei centri di presa delle immagini della strisciata di acquisizione, generando la traiettoria del sistema. Nel presente progetto di tesi si è scelto di risolvere il problema dell'allineamento delle immagini utilizzando una ricerca automatica dei punti di legami tramite SIFT (paragrafo 5.3.6.1). La funzione align2RGBD della Princeton Vision permette la determinazione

della rototraslazione tra le coppie di immagini (paragrafo 5.3.6). Applicando tutto il blocco di rototraslazioni in maniera sequenziale si determina la traiettoria dei centri di presa del rilievo (paragrafo 5.3.7). Prima di lanciare l'algoritmo dell'allineamento è necessario processare le immagini per migliorarne sia la qualità che la precisione stessa (paragrafi 5.3.2 e 5.3.3). Inoltre l'algoritmo della Princeton Vision lavora con immagini RGBD le quali vanno create a partire dalle coppie di immagini Depth-RGB (paragrafi 5.3.4 e 5.3.5).

Come spiegato nel paragrafo Sincronizzazione degli strumenti (paragrafo 5.1.2) le immagini sono salvate direttamente su laptop e nominate secondo un nome che coincide con il tempo interno del Kinect in cui sono state scattate. Il sensore acquisisce le immagini Depth e RGB in sequenza partendo dalle Depth. Al fine del tracciamento della traiettoria della camera non sono necessarie le serie di immagini acquisite mentre il sistema è in sosta per il posizionamento con sistemi GPS. Queste immagini, riconoscibili facendo scorrere il file delle immagini RGB, vanno eliminate insieme alle rispettive immagini Depth. Si eliminano anche le serie di immagini RGB e Depth in cui vi sia la presenza di elementi in movimento nella scena dell'immagine.

### **5.3.1 Sottocampionamento delle immagini**

Dalla traiettoria calcolata da dati del sensore Kinect ci si aspetta un errore dell'ordine di poche decine di centimetri. Si ritiene sia poco significativo lanciare l'algoritmo di calcolo dello spostamento per coppie di immagini per cui lo spostamento tra esse

calcolato sia atteso essere molto inferiore rispetto alla precisione dello strumento stesso. Si sceglie quindi di sottocampionare il blocco di immagini. Da un altro punto di vista il sottocampionamento riduce anche il numero di rototraslazioni da calcolare. Considerando che l'algoritmo della Pinceton Vision richiede tempi di calcolo ampi (circa 20 minuti per ogni coppia di immagini nel caso di laptop ASUS con processore AMD E-450 APU con radeon(tm) HD Graphics 1.65 GHz di 3.60 GB di RAM utilizzabile, con sistema operativo a 64bit), il sottocampionamento riduce notevolmente il tempo di calcolo.

Si tenga comunque presente che un sottocampionamento eccessivo non permette una sufficiente sovrapposizione delle immagini, che determina l'impossibilità di calcolarne l'allineamento tra le coppie. Ai fini della stima del sottocampionamento si tenga in considerazione la necessità di una sovrapposizione tra coppie di immagini circa dell'80%.

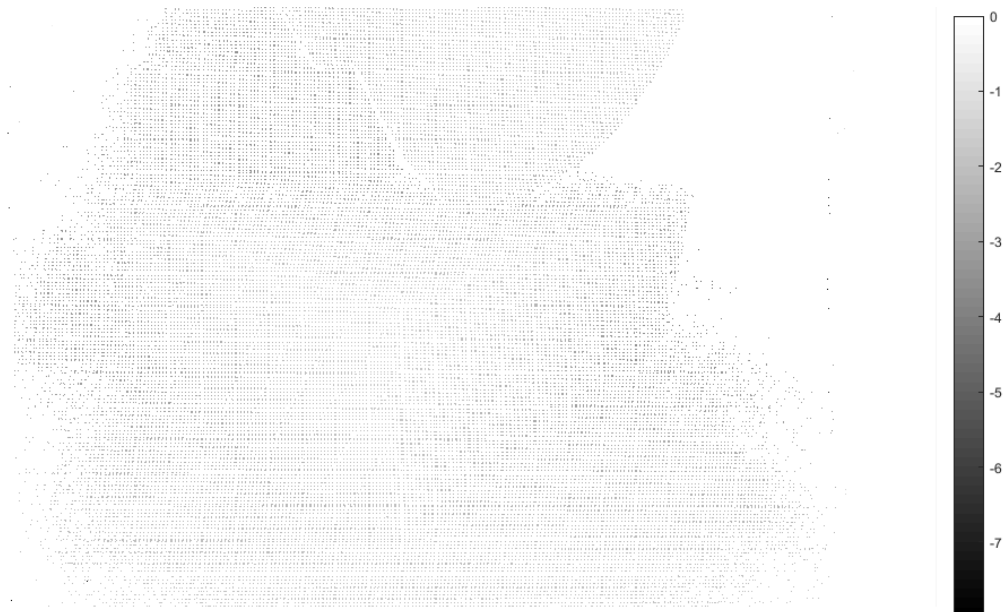
### **5.3.2 Correzione dalla distorsione delle camere**

Dalla fase descritta nel paragrafo Calibrazione delle camere 5.1.1 si ottengono i parametri di OI delle camere RGB e IR e i rispettivi coefficienti che descrivono le equazioni delle distorsioni ottiche presenti nel paragrafo di introduzione teorica 4.2. Le immagini (sia depth che RGB) sono quindi ricampionate in modo da ottenere immagini prive di distorsione. Tale operazione viene svolta per tutte le immagini Depth e RGB selezionate nella fase di sottocampionamento. Questo processo dovrebbe assicurare una

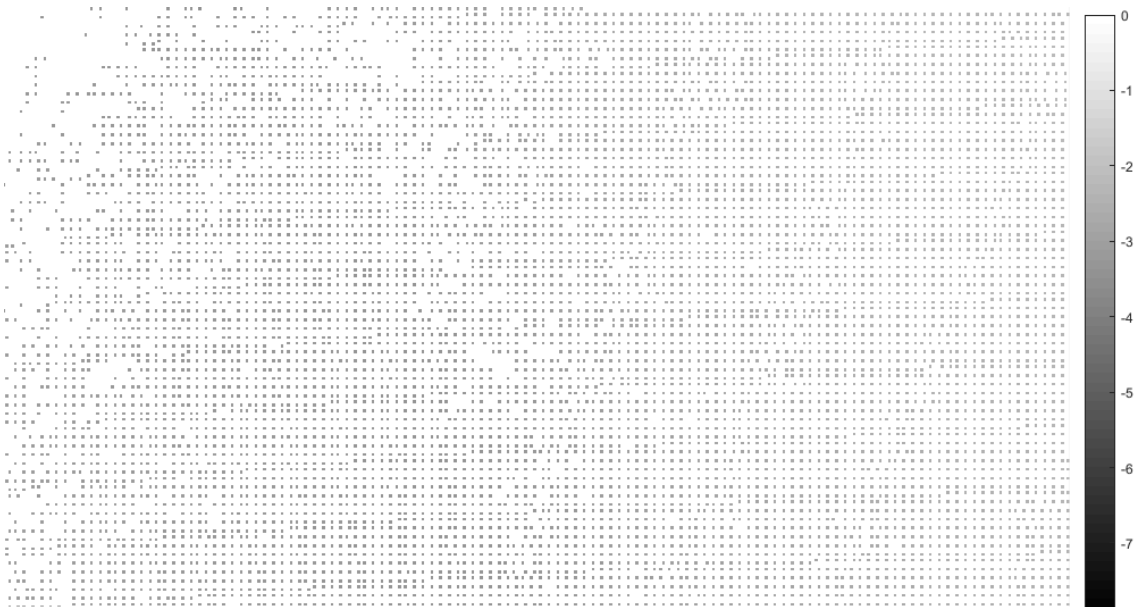
migliore qualità soprattutto ai bordi immagine, dove gli effetti della distorsione sono maggiori.

### 5.3.3 Ricampionamento delle immagini Depth

Da alcuni test precedenti alla fase di rilievo si è constatata la necessità di ricampionare le immagini Depth, aumentandone in modo fittizio la risoluzione. Sostanzialmente si divide ogni pixel in 4 pixel più piccoli, ai quali viene associata la stessa informazione del pixel precedente. Questo processamento risulta necessario al fine di poter utilizzare delle immagini RGBD nelle quali l'informazione legata alla depth non sia troppo scarsa. Ciò è legato alla risoluzione molto diversa tra i due sensori ottici del Kinect v2. Aumentando il numero di pixel distribuendo in essi la stessa informazione del pixel originario si rende più fitta l'immagine Depth. Di conseguenza i punti che nell'immagine Depth originale sono isolati (spigoli di oggetti o quant'altro), nella nuova immagine Depth risultano essere un gruppo di punti che di conseguenza sono maggiormente riconoscibili per la fase del calcolo delle trasformazioni tra nuvole di punti di immagini Depth. In *Figura 28* ed in *Figura 30* si osservino le differenze tra l'informazione contenuta in un'immagine Depth prima e dopo il ricampionamento.

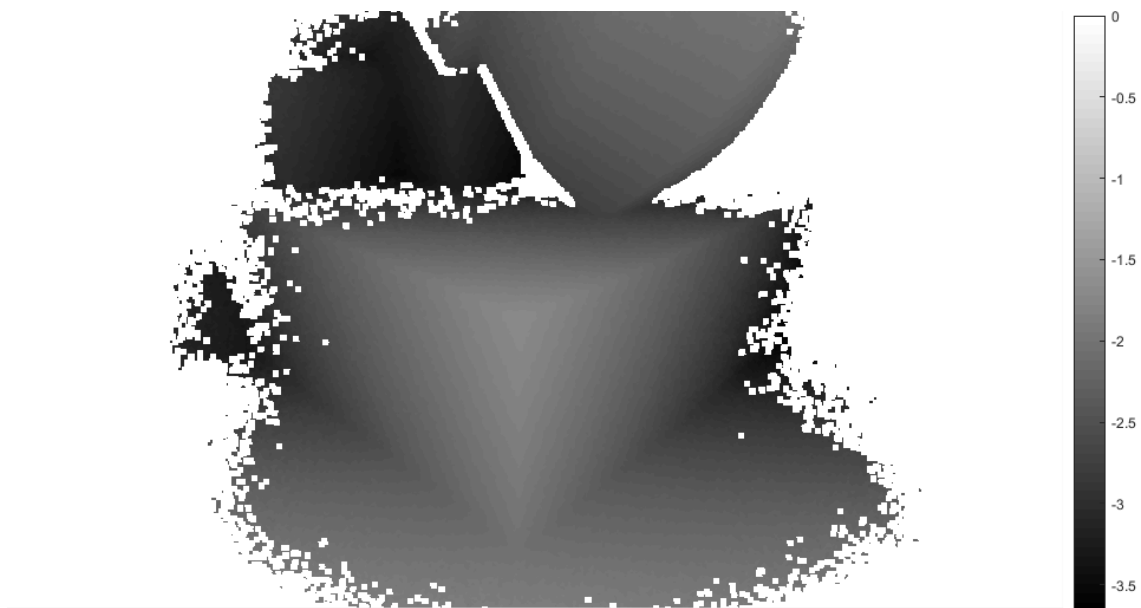


**Figura 28** Immagine dei pixel con contenuto informativo dell'immagine Depth originaria.



**Figura 29** Zoom di un particolare dell'immagine 28.

In *Figura 29* si osservi lo zoom di un particolare della nuvola di punti rappresentata in *Figura 29*. L'effetto striato presente in *Figura 29* nello zoom non è presente, quindi esso è determinato solo dalla rappresentazione grafica.



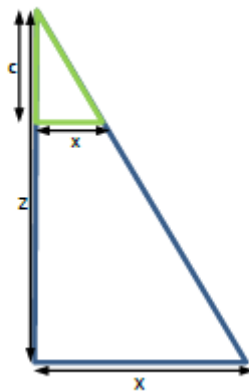
*Figura 30 Immagine dei pixel con contenuto informativo dell'immagine Depth processata.*

### **5.3.4 Costruzione della nuvola di punti**

Le immagini acquisite dal Kinect sono immagini speculari della scena ripresa, quindi vanno capovolte da destra verso sinistra. In particolare, il contenuto informativo dei pixel di una immagine Depth è l'informazione metrica della profondità (paragrafo 3). Considerando un sistema di riferimento centrato nel centro di proiezione della camera IR, in cui  $x$  è l'asse è parallelo alla dimensione maggiore del sensore della alla camera



IR, l'asse y è parallelo alla altezza del sensore della camera IR e l'asse z entrante nel sensore per completare l'orientamento destrorso, risulta facile calcolare la nuvola di punti originata dall'immagine Depth. Infatti, dai triangoli simili presentati in *Figura 31* sono riassunte le relazioni tra spazio immagini e spazio oggetto necessarie per la determinazione dei punti della nuvola di punti (PC) con sistema di riferimento sopra descritto.



**Figura 31 Schematizzazione delle relazioni che intercorrono tra lo spazio immagini e lo spazio oggetto.**

Le formule per la proiezione dei punti immagine in punti oggetto sono

$$X = \frac{\xi}{c} \cdot Z \quad Y = \frac{\eta}{c} \cdot Z$$

.dove X e Y sono le coordinate del punto oggetto proiezione del punto immagine di coordinate  $(\xi, \eta)$ . Queste sono calcolate per ogni pixel dell'immagine Depth utilizzando i parametri di OI

$$\xi = i \frac{W}{w} \cdot c - \xi_0 \quad \eta = \eta_0 - j \frac{H}{h} \cdot c$$

. dove  $i$  e  $j$  sono rispettivamente la riga e la colonna del pixel  $(i,j)$ ,  $w$  e  $h$  sono la larghezza e l'altezza dell'immagine Depth,  $W$  e  $H$  sono la larghezza e l'altezza del sensore IR.  $\xi_0$  e  $\eta_0$  sono le coordinate immagine del punto principale.

### 5.3.5 Costruzione dell'immagine RGBD

Con immagine RGBD si intende un'immagine generata dall'unione di un'immagine RGB con l'informazione contenuta in un'immagine Depth. Questo formato di immagine è richiesto per il calcolo della rototraslazione dell'algoritmo align2RGBD della Princeton Vision.

Un'immagine RGBD è costituita da 6 canali di egual numero di pixel (1920x1080): i primi tre sono rispettivamente destinati alle immagini in toni del rosso (R), verde (G) e del blu (B), normalizzati. I canali 4, 5 e 6 sono destinati alle coordinate rispettivamente Z, X e Y della PC proiettata tramite equazioni di Collinearità di presa sull'immagine RGB. Il problema della generazione delle immagini RGBD quindi si risolve come un problema fotogrammetrico di presa. Le equazioni di Collinearità in fase di presa sono espresse nel paragrafo 4.1 nell'equazione *Equazione 4*. In questo caso i parametri di OE coincidono con i parametri di OE relativo delle camere stimati nella fase di calibrazione delle camere (si veda paragrafo 5.1.1). I termini della matrice di rotazione R sono stimati partendo dagli angoli di Cardano  $\omega$ ,  $\varphi$  e  $\kappa$ , soluzione dei parametri di OE relativo. Si stima la matrice R dagli angoli di cardano tramite l'*Equazione 8*

$$R = \begin{bmatrix} \cos \varphi \cos \kappa & \cos \omega \sin \kappa + \sin \omega \sin \varphi \cos \kappa & \sin \omega \sin \kappa - \cos \omega \sin \varphi \cos \kappa \\ -\cos \varphi \sin \kappa & \cos \omega \cos \kappa - \sin \omega \sin \varphi \sin \kappa & \sin \omega \sin \kappa + \cos \omega \sin \varphi \cos \kappa \\ \sin \varphi & -\sin \omega \cos \varphi & \cos \omega \cos \varphi \end{bmatrix}$$

Calle coordinate immagine  $(\xi, \eta)$  tramite le equazioni

$$i = (\xi_0 + \xi) \frac{w}{W} \quad j = (\eta_0 - \eta) \frac{h}{H}$$

(ed un arrotondamento all'intero più vicino +1) si stima la coordinata pixel in cui allocare le coordinate Z, X e Y nei canali 4, 5 e 6 dell'immagine RGBD.

### 5.3.6 Allineamento di coppie di immagini

Per allineare le immagini RGBD si utilizza l'algoritmo align2RGBD della Princeton Vision, Esso allinea coppie di immagini per volta, quindi successivamente è necessario combinare le trasformazioni di ogni coppia per generare la traiettoria completa (paragrafo Tracciamento della traiettoria dei centri di presa 5.3.7).

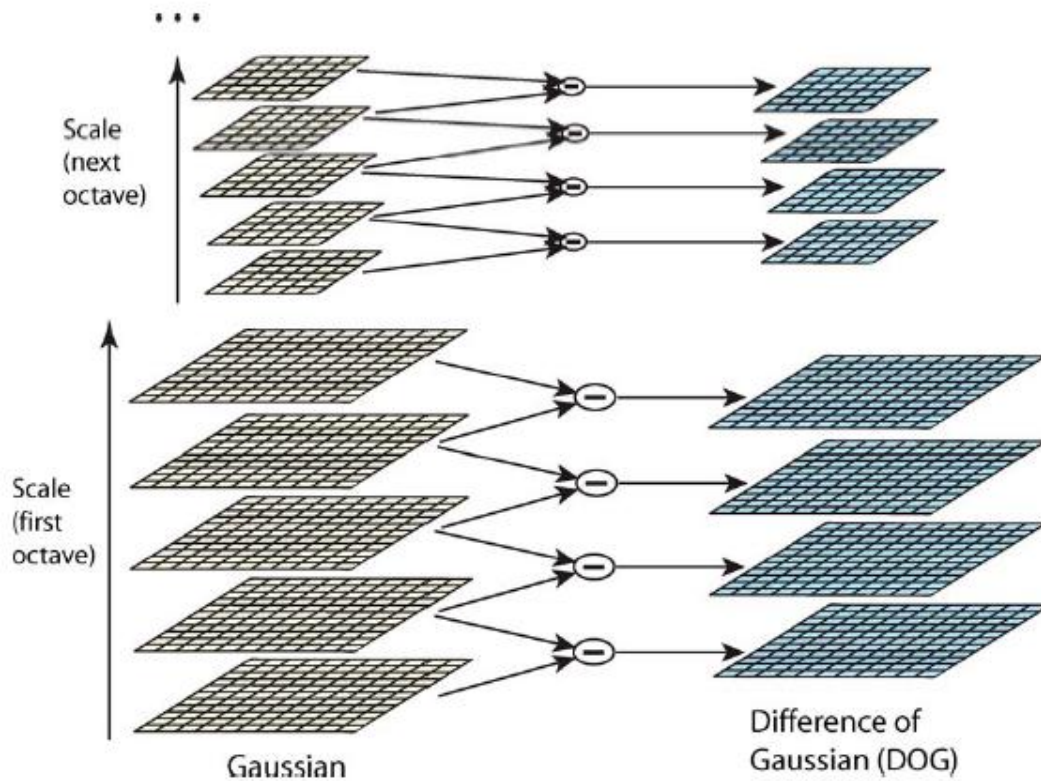
L'algoritmo inizia con la ricerca di punti di legame tramite un matching tra le due immagini RGB eseguita da SIFT (Lowe, 2004). Quindi vengono eliminati i matching poco significativi tramite test sul rapporto di quelli selezionati (paragrafo *SIFT (Scale Invariant Feature Transform)* 5.3.6.1). All'interno del set di punti di legame selezionati da SIFT vengono scelti quelli con informazione Depth (nei canali 4, 5 e 6) valida allo scopo. Quindi viene utilizzato un algoritmo per la ricerca della trasformazione che porti

la nuvola di punti selezionata dai punti di legame dalla prima immagine RGBD a quella della seconda immagine RGBD (paragrafo ICP 5.3.6.2). Quest'ultimo algoritmo lavora all'interno di un ciclo di RANSAC (Nist'er, 2003) il quale esegue un processo di reiezione degli outlier (Fischer e Bolles, 1981). Per ulteriori informazioni circa il funzionamento dell'algoritmo si rimanda a Jianxiong Xiao e al. (2013).

#### **5.3.6.1 SIFT (Scale Invariant Feature Trasform)**

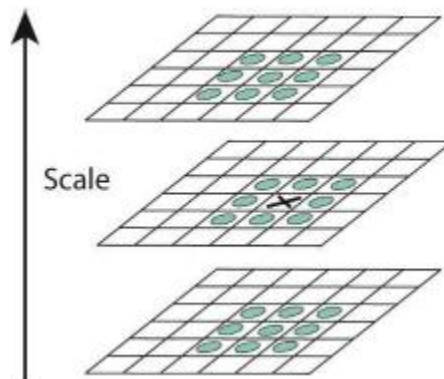
SIFT è l'algoritmo più utilizzato per la ricerca di punti di legame. L'algoritmo ricerca i punti in una regione dell'immagine limitata, onde evitare errori dovute ad occlusioni o deformazioni. Viene quindi classificato come operatore di tipo regionale.

La procedura dell'algoritmo SIFT inizia con la ricerca di tutti i punti interessanti in tutti a diverse scale di ricampionamento dell'immagine (svolta per ottave di pixel). Alle immagini quindi viene applicato un filtro di convoluzione di Gauss. Si applica per ogni scala una differenza con la funzione di Gauss adiacente (Differenze di Gauss, DoG) e quindi vengono selezionati i punti con differenza massima e minima. In *Figura 32* si veda come sono realizzate le immagini di diverse scale e come vengono create le DoG.



**Figura 32** Convoluzione con Gauss e creazione di un DoG (Lowe, 2004).

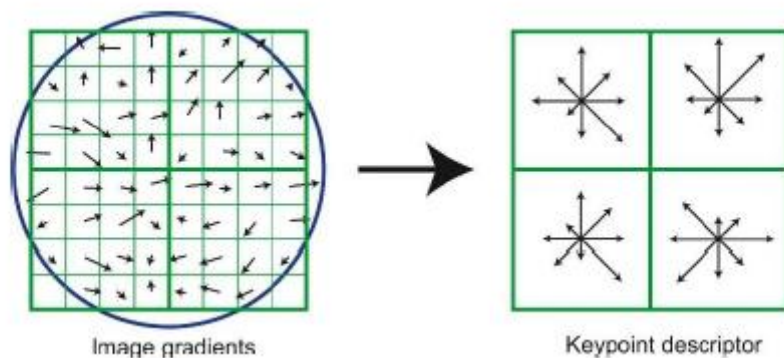
Per ogni punto di ogni scala selezionato dalle DoG viene definito un descrittore in funzione della suo gradiente radiometrico. Questo valore viene confrontato quello degli 8 pixel a lui circostanti e con quelli dei 9 circostanti nelle scale superiori ed inferiori. Questo confronto serve a selezionare attraverso le diverse scale DoG di valori massimi e minimi assoluti. Vengono selezionati quei pixel la cui DoG assoluta supera una certa soglia prefissata. In *Figura 33* l'esempio di un DoG e della fase di confronto con i punti adiacenti nelle scale superiori ed inferiori.



**Figura 33** I DoG assoluti vengono selezionati confrontando il descrittore con quello dei pixel circostanti nella sua e nelle scale inferiori ed inferiori (Lowe,2004)

Partendo dal valore di direzione ottenuto dall'analisi del gradiente dell'immagine, ad ogni candidato punto omologo è assegnato uno o più valori di orientazione. Quindi ad ogni punto viene applicata una rotazione pari al valore di orientazione. Questa procedura assicura invarianza per tutti i livelli di scala.

Un buon descrittore di un punto omologo deve essere il meno suscettibile a fattori variabili quali illuminazione e prospettiva. Va quindi analizzata tramite una Gaussiana ogni variazione all'interno di una finestra che ha per centro il candidato punto omologo. Questa analisi viene realizzata dividendo i 16 pixel che circondano il candidato punto omologo in una regione di dimensioni 4x4. Per ognuna di queste regioni sono considerate 8 orientazioni, quindi ogni descrittore è composto da 128 vettori. Quindi il vettore è normalizzato con l'obiettivo di ridurre gli effetti dei cambi di illuminazione.



**Figura 34 a) Gradienti radiometrici predominanti in una regione circolare attorno al punto candidato. B) descrittore SIFT**

L'ultimo step di SIFT consiste nella fase di Matching, cioè l'associazione dei punti omologhi tra le due immagini. Per valutare il miglior matching, viene valutata la distanza tra descrittori di candidati punti omologhi delle due immagini. I punti caratterizzati da distanza Euclidea minore sono classificati come omologhi. Per eliminare i falsi punti omologhi Lowe (2004) propone di confrontare la distanza dei descrittori di un punto e del suo omologo con quella calcolata considerando il secondo punto più vicino. Sono accettate le coppie di punti omologhi la cui distanza è superiore di un livello di soglia fissato. Per informazioni dettagliate consultare Lowe (2004).

### 5.3.6.2 ICP

L'algoritmo SIFT permette di ricavare le coppie di punti omologhi per le due immagini in analisi. Da questi punti omologhi è possibile già svolgere un allineamento delle due

immagini e quindi il calcolo della rototraslazione che le allinea. Tuttavia la rototraslazione ottenuta da SIFT è solo una soluzione che non tiene conto delle informazioni ricavate direttamente dalle immagini depth (spazio oggetto). L'algoritmo di ICP (Iterative Closest Point) (Chen and Medioni (1981) e Basl and McKay (1992)). prende in input due nuvole di punti e calcola la rototraslazione (matrice di rotazione R e vettore traslazione t) che meglio allinea le due nuvole di punti. Questo algoritmo necessita di essere inizializzato e la soluzione ottenuta da SIFT è una buona approssimazione da cui partire.

Considerando le due nuvole di punti  $P = \{p_1, p_2, \dots, p_n\}$  e  $Q = \{q_1, q_2, \dots, q_n\}$  il problema dell'algoritmo ICP è di minimizzare l'errore tra i due set di dati

$$E = \sum_{i=1}^n \|\bar{R}p_i + \bar{t} - q_i\|^2$$

dove E rappresenta lo scarto in metri tra le due nuvole di punti,  $\bar{R}$  è la matrice di rotazione  $\bar{t}$  la traslazione della matrice di rototraslazione che si sta cercando.

Si consideri che la ricerca della soluzione è resa difficile dal fatto che non si conosce la corrispondenza tra i punti delle due nuvole di punti. Quindi ICP è un algoritmo che può raggiungere la soluzione solo per modifiche iterative della rototraslazione fino a che non si raggiunge una soluzione che minimizza E. La convergenza della soluzione dipende nel maggior dei casi dalla qualità della soluzione approssimata di inizializzazione.



### 5.3.6.3 RANSAC

RANSAC sta per “RANdom Sample Consensus” e consiste in un metodo iterativo per la stima dei parametri di un modello matematico a partire da un insieme di dati contenenti uno o più outlier. RANSAC è un algoritmo non deterministico, cioè produce un risultato corretto solo con un livello di soglia di probabilità, detto confidenza. All’aumentare della confidenza,  $p$ , viene aumentato anche il numero di iterazioni,  $k$ , e quindi del tempo di calcolo (Fischler e Bolles, 1981). L’assunzione di base per il funzionamento dell’algoritmo RANSAC è che gli  $n$  dati siano costituiti da inlier ed outlier. I dati inlier sono rappresentabili da un modello definito da parametri. Alcuni di essi sono in grado di descrivere in modo ottimale il modello. I dati outlier invece non possono essere rappresentati da tale modello. Essi derivano da misurazione erronea dell’osservazione o da ipotesi non corrette.

L’algoritmo di RANSAC viene applicato sia per valutare la presenza di outlier dalla soluzione già ottenuta da SIFT e da ICP. L’algoritmo di RANSAC esegue una procedura in loop che consiste in

1. Selezione di  $n$  coppie omologhe ipotetiche (per SIFT coppie di punti immagini omologhi, mentre per ICP coppie di punti oggetto omologhi).
2. Stima di tutti i parametri liberi del modello (cioè della rototraslazione) utilizzando le  $n$  coppie ipotetiche di punti omologhi. Quindi applicazione del modello stimato a tutti i dati del problema. I dati che verificano il modello vengono considerati nuovi inlier ipotetici.

3. Il modello viene considerato accettabile se nella fase 2 un numero sufficienti di dati è stato classificato come inlier ipotetico.
4. Stima di nuovi parametri del modello utilizzando tutti i dati classificati come inlier nella fase 2.
5. Stima dell'errore degli inlier rispetto al modello calcolato nella fase 4.

La procedura descritta da questo loop di operazioni viene eseguita un numero di volte prefissato  $k$ . In ogni ripetizione della procedura è calcolato un modello il quale che nella fase 3 viene accettato o escluso. Dai fallimenti è possibile risalire agli outlier.

Per ulteriori informazione su RANSAC si rimanda a (Fischler e Bolles, 1981).

### **5.3.7 Tracciamento della traiettoria dei centri di presa**

Si faccia coincidere il l'origine di un sistema di riferimento locale con il centro di presa della camera RGB della prima immagine del blocco di immagini. Quindi si considera come sistema di riferimento locale quello precedentemente definito in fase di creazione delle immagini RGBD (ovvero origine coincidente con il centro di presa della camera RGB, asse  $x$  diretto come la larghezza del sensore, asse  $y$  diretto come l'altezza del sensore e asse  $z$  a chiudere la terna destrorsa)

A seguito del processo di calcolo dell'algoritmo di Princeton Vision si ottengono le rototraslazioni tra tutte le coppie di immagini ordinate secondo il tempo di acquisizione. Ogni rototraslazione è calcolata tra due immagine consecutive (si da per scontato il sottocampionamento) quindi la coordinata della seconda immagine della coppia rispetto

al sistema di riferimento interno nella prima immagine coincide con le componenti traslazionali della traslazione tra le due immagini.

L'ordine con cui applicare le rototraslazioni per calcolare la traiettoria finale è fondamentale. Ogni rototraslazione determinata tra coppie di immagini RGBD definisce la trasformazione da applicare ai punti oggetto nel sistema di riferimento della seconda immagine RGBD della coppia affinché siano definiti rispetto al sistema di riferimento della prima immagine RGBD della coppia. Questo impone che per determinare la traiettoria di tutti i centri di presa del blocco fotogrammetrico nel sistema di riferimento della prima immagine RGBD del blocco stesso sia necessario applicare per ogni immagine del blocco le rototraslazioni di tutte le coppie di immagini che precedono l'immagine stessa (con precedono si intende acquisite temporalmente prima). Il calcolo va effettuato in ordine retrogrado, cioè dalla trasformazione della coppia precedente all'immagine stessa fino alla trasformazione della prima coppia dell'intero blocco.

## **5.4 SOLUZIONE DEL SENSORE U-BLOX**

La soluzione al posizionamento cinematico della stazione tramite sistema GPS misurata dal sensore U-blox viene determinata per mezzo del software goGPS) Realini e Reguzzoni, 2013), software realizzato in ambiente Matlab<sup>®</sup>. Per mezzo della sincronizzazione degli strumenti Kinect e U-blox (paragrafo 5.1.2) si può risalire al tempo GPS degli istanti iniziali e finali in cui il sistema era in sosta nei punti prefissati o era in rilievo del giro andando ad osservare le immagini RGB del blocco fotogrammetrico. Gli intervalli di sosta coincidono a quelli in cui il sistema è nei vertici

del quadrato di calibrazione e a quelli esattamente precedenti all'inizio del rilievo del giro (posizione spigolo1 del quadrato, vedi paragrafo 5.2) e a quelli esattamente dopo la fine del rilievo del giro (ancora posizione 1 del quadrato). Per questi punti si calcola una soluzione di posizionamento statico. Per quanto riguarda gli intervalli di rilievo del giro si calcola un posizionamento cinematico.

## 5.5 FILTRO DI KALMAN

Il filtro di Kalman è l'algoritmo più conosciuto per l'integrazione misure di posizionamento assoluto di diversi strumenti. Utilizzato moltissimo per integrare misure navigazione di sistemi INS/GPS in questa tesi viene utilizzato per l'integrazione della traiettoria calcolata con un sistema fotogrammetrico e con un sistema GPS. Tramite questo strumento ci poniamo l'obiettivo di stimare la posizione del sistema all'istante  $i+1$  utilizzando le osservazioni disponibili fino all'istante precedente  $i$ . Il filtro di Kalman agisce sulla stima della soluzione al passo successivo minimizzando ai Minimi Quadrati la funzione-target (*Equazione 9*) calcolata come somma degli errori.

*Equazione 9*

$$\phi = e_i^+ (C_i^e)^{-1} e_i + \varepsilon_{i+1}^+ (C_{i+1}^\varepsilon)^{-1} \varepsilon_{i+1} + v_{i+1}^+ (C_{i+1}^v)^{-1} v_{i+1}$$

in cui  $\varepsilon_{i+1}$  è l'errore del modello,  $v_{i+1}$  il rumore dell'osservazione e  $e_i$  è l'errore della stima al passo precedente. I tre errori sono tra loro stocasticamente indipendenti

Per ulteriori approfondimenti sul filtro di Kalman si rimanda a Kalman (1960).

Il filtro di Kalman implementato nel modello di questa tesi prevede tre componenti spaziali. Si lavora in ipotesi di sistema in moto a velocità costante. La discretizzazione

del tempo è scandita dai tempi dalle osservazioni immagini RGBD e posizionamento con Ublox. I tempi dei due strumenti fanno riferimento al tempo PC (paragrafo Sincronizzazione degli strumenti 5.1.2). La condizione di inizializzazione del modello è quindi determinata dal vettore colonna con e tre posizioni iniziali e le componenti della velocità nel primo tratto.

$$X_1 = [0 \quad 0 \quad 0 \quad v_{x_{12}} \quad v_{y_{12}} \quad v_{z_{12}}]^T$$

in cui  $v_{s_{12}}$  è la componente della velocità nella direzione  $i$ , calcolata come  $s_{s_{12}}/t_{12}$  dove  $t_{12}$  è l'intervallo temporale tra l'osservazione dell'istante iniziale e dell'istante successivo, e dove  $s_{s_{12}}$  è la componente lungo la direzione  $s$  dello spostamento osservato l'intervallo  $t_{12}$ .

La matrice di covarianza degli errori della stima che minimizza la funzione target, per il primo passo, coincide con la matrice di covarianza dell'osservazione. Quindi rispettivamente  $C_1^e = C_1^\varepsilon$

$$C_1^e = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \sigma_{Eps_{12}}^2 & 0 & 0 \\ 0 & 0 & 0 & 0 & \sigma_{Eps_{12}}^2 & 0 \\ 0 & 0 & 0 & 0 & 0 & \sigma_{Eps_{12}}^2 \end{bmatrix}$$

.in cui  $\sigma_{Eps_{12}}^2$  equivale alla precisione con cui viene determinata la rototraslazione dall'algorithm align2RGBD tra la prima coppia di immagini RGBD1 e RGBD2.

$$C_\varepsilon = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \sigma_{\bar{v}_x}^2 & 0 & 0 \\ 0 & 0 & 0 & 0 & \sigma_{\bar{v}_y}^2 & 0 \\ 0 & 0 & 0 & 0 & 0 & \sigma_{\bar{v}_z}^2 \end{bmatrix}$$

Dove

$$\sigma_{\bar{v}_s}^2 = (v_{s_t} - v_{\bar{s}})^2 \quad v_{s_i} = \frac{s_{i+1,i}}{(t_{i+1}-t_i)} \quad v_{\bar{s}} = \frac{\sum_1^n v_{s_i}}{n}$$

In cui  $C_\varepsilon$  rappresenta la matrice di covarianza dell'errore del modello e  $\sigma_{\bar{v}_s}^2$  la varianza del vettore velocità delle osservazioni della traiettoria del Kinect nella componente di direzione s.  $v_{s_i}$  è la velocità nel tratto dall'osservazione i-esima all'osservazione i+1-esima nella componente i.  $v_{\bar{s}}$  è la media della velocità nella componente s delle osservazione da Kinect. La matrice  $C_\varepsilon$  è mantenuta per ipotesi stazionaria per l'intera durata del rilievo.

La matrice T per calcolare la lunghezza dell'intervallo di tempo da un'osservazione all'altra è

$$T = \begin{bmatrix} 1 & 0 & 0 & \Delta t_{i+1,i} & 0 & 0 \\ 0 & 1 & 0 & 0 & \Delta t_{i+1,i} & 0 \\ 0 & 0 & 1 & 0 & 0 & \Delta t_{i+1,i} \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

in cui gli elementi  $\Delta t_{i+1,i}$  rappresentano l'intervallo di tempo calcolato come il tempo dell'osservazione i+1 esima meno il tempo dell'osservazione i-esima. Si tenga presente che i tempi delle osservazioni comprendono sia quelle del sensore Kinect sia quelle del sensore UGPS. Quindi si può costruire un vettore colonna t dei tempi degli strumenti da

cui ricavare il vettore  $\Delta t$  degli intervalli di tempi delle osservazioni. Esso possiede  $n+m$  righe, quanto la somma delle osservazioni da Kinect,  $n$ , e quelle da GPS,  $m$ . Nel vettore dei tempi delle osservazioni si conserva la provenienza dello strumento in apposita colonna. Ad esempio

$$t = \begin{bmatrix} t_1 & Kin \\ t_2 & Kin \\ t_3 & Kin \\ t_4 & GPS \\ t_5 & Kin \\ \vdots & \vdots \\ t_{n+m} & Kin \end{bmatrix}$$

La matrice disegno dell' modello dell'osservazione Kinect e dell'osservazione GPS sono

$$H_{Kin} = \begin{bmatrix} 0 & 0 & 0 & \Delta t_{k+1,k} & 0 & 0 \\ 0 & 0 & 0 & 0 & \Delta t_{k+1,k} & 0 \\ 0 & 0 & 0 & 0 & 0 & \Delta t_{k+1,k} \end{bmatrix}$$

$$H_{GPS} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \end{bmatrix}$$

E quindi le matrici di covarianza degli errori dell'osservazioni causati dal rumore

$$C_{Kin}^{vv} = \begin{bmatrix} Eps_{k+1}^2 & 0 & 0 \\ 0 & Eps_{k+1}^2 & 0 \\ 0 & 0 & Eps_{k+1}^2 \end{bmatrix}$$

$$C_{GPS}^{vv} = \begin{bmatrix} \sigma_{x_{GPS}}^2 & \sigma_{yx} & \sigma_{zx} \\ \sigma_{xy} & \sigma_{y_{GPS}}^2 & \sigma_{zy} \\ \sigma_{xz} & \sigma_{yz} & \sigma_{z_{GPS}}^2 \end{bmatrix}$$

in cui  $C_{GPS}^{vv}$  è espressa nel sistema di riferimento GPS. Dalla trasformazione

$$C_{GPSKin}^{vv} = AC_{GPS}^{vv}A^T$$

dove  $A$  è la rototraslazione dal sistema di riferimento della traiettoria tracciata del Kinect (quindi con origine nella prima immagine del blocco, vedi paragrafo 5.3.7).

$C_{GPS_{Kin}}^{vv}$  è la matrice di  $C_{GPS}^{vv}$  nel sistema di riferimento Kinect. Nel filtro viene utilizzata  $C_{GPS_{Kin}}^{vv}$ .

Si applichi quindi in modo iterativo la soluzione del filtro di Kalman, costituita dalle equazioni

$$\begin{aligned}
 K_{i+1} &= C_\varepsilon + T_{i+1} C_i^e T_{i+1}^+ \\
 G_{i+1} &= K_{i+1} H_{i+1}^+ [C_{i+1}^v + H_{i+1} K_{i+1} H_{i+1}^+]^{-1} \\
 X_{i+1|i+1} &= G_{i+1} Y_{i+1} + (I - G_{i+1} H_{i+1}) T_{i+1} X_{i|i} \\
 C_{i+1}^e &= (I - G_{i+1} H_{i+1}) K_{i+1}
 \end{aligned}$$

In cui  $X_{i+1|i+1}$  e  $C_{i+1}^e$  sono rispettivamente la soluzione stimata che minimizza la funzione-target (Equazione 9) e la matrice di covarianza della sua stima.

Per i valori di inizzializzazione di  $X$  si consideri il vettore già indicato  $X_1$ , mentre per quello di  $C^e$  si consideri  $C_1^e$ . In riferimento al tipo di strumento da cui deriva l'osservazione (consultare la seconda colonna del vettore  $t$ ) si associ ad  $H_{i+1}^+$  ed a  $C_{i+1}^v$  opportunamente tra  $H_{Kin}$  o  $H_{GPS}$  e tra  $C_{Kin}^{vv}$  o  $C_{GPS_{Kin}}^{vv}$ .



### 6.1 SCELTA DELLA STRUMENTAZIONE E DEL LUOGO DEL RILIEVO

Come già spiegato nel capitolo 2 la tecnica di riferimento per la navigazione è quella GNSS ma a causa del problema dell'oscuramento essa risulta inaffidabile in ambiente urbano. Con l'obiettivo di fornire una soluzione di supporto alle difficoltà del metodo GNSS in ambito urbano con questa tesi si propone una soluzione frutto dell'integrazione tra quella fornita da tecnica GPS e quella fornita da tecnica fotogrammetrica. Il sensore fotogrammetrico scelto è il Microsoft Kinect nella sua ultima versione (capitolo 3.2). Per realizzare una soluzione al posizionamento cinematico in ambiente urbano integrando al metodo GPS quello fotogrammetrico è stato necessario scegliere un ambiente favorevole allo scopo (paragrafo Rilievo 5.2).

Considerando che il metodo proposto necessita dell'utilizzo di un'antenna GPS è stato necessario scegliere un luogo che presentasse sufficiente visione di cielo da permettere il posizionamento cinematico. Nell'emisfero nord la maggior parte dei satelliti sono visibili nella porzione sud di cielo quindi è stato necessario scegliere un luogo aperto in cui non vi fossero elementi verticali molto alti nel lato sud. Il rilievo è stato svolto in Piazza Leonardo Da Vinci a Milano, sfruttando l'elemento geometrico della scultura che compone la fontana presente in prossimità della Scuola Primaria Statale Leonardo Da Vinci. Piazza Leonardo Da Vinci, a differenza del lato nord in cui è presente il

Rettorato del Politecnico, nel suo lato sud risulta priva di palazzi molto alti che coprono la visuale di cielo. A questo si aggiunga anche la comodità della sua vicinanza con il Dipartimento di Ingegneria Civile e Ambientale (DICA) – Sezione Geodesia e Geomatica, sede del magazzino degli strumenti utilizzati per la rilevazione. In *Figura 35* è presentata una rappresentazione schematica dall'alto di Piazza Leonardo da Vinci a Milano. Nella stessa figura in basso si osservi la presenza della fontana in cui è stato scelto di effettuare il rilievo.

La fontana di Piazza Leonardo si trova circa nel centro della porzione sud della piazza a circa 50-60 metri di distanza dai palazzi ai lati della piazza, quindi vi è una porzione di cielo visibile sufficiente per l'acquisizione del segnale dei satelliti GPS.



*Figura 35 Piantina di Piazza Leonardo Da Vinci, Milano ([www.http://blog.urbanfile.org/](http://blog.urbanfile.org/)).*

Come spiegato nel capitolo 3, il sensore Kinect permette di acquisire sia immagini RGB che Depth. Le seconde, affinché risultino utili al tracciamento della traiettoria, necessitano che in esse sia rappresentata una scena in un range tra gli 0.50 e i 4.50 metri in cui sia presente un elemento geometrico modellato nelle tre dimensioni. L'allineamento delle immagini Depth infatti è possibile solo se dalla nuvole di punti di

immagini consecutive siano presenti le stesse feature interessanti, cioè appunto forme geometriche nelle tre dimensioni, e non semplici superfici piatte. La fontana costituisce questo elemento geometrico di spicco. Questa presenta sul suo lato minore a sud una scultura di altezza di circa 4 metri connessa alla vasca e la sua struttura è costituita da un cornicione alto poco meno di un metro che percorre l'intero perimetro rettangolare della vasca. La presenza di questo elemento peraltro ha ispirato l'idea di svolgere una traiettoria circolare attorno al perimetro della fontana. L'altezza ridotta della fontana ha permesso di avvicinare il sistema rigido degli strumenti all'interno del range di acquisizione della camera Depth (paragrafo 3.2) senza che l'altezza della scultura determinasse un intralcio importante all'acquisizione del segnale dei satelliti GPS. La sua forma geometrica originale invece ha permesso di acquisire un blocco fotogrammetrico di immagini Depth sufficienti allo scopo del tracciamento della traiettoria circolare. In *Figura 36 a)* si osservi la fontana con il cornicione lungo il perimetro rettangolare. In *Figura 36 b)* il particolare della scultura dalla modellazione della forma articolata.

Si sottolinei che l'idea originale era quella di svolgere un rilievo di più di un giro attorno alla struttura della fontana. L'aspettativa era di poter individuare per ogni giro attorno alla fontana un valore di deriva della traiettoria ottenuta dal Kinect rispetto alla traiettoria di riferimento. Tuttavia un imprevisto ha interrotto l'acquisizione durante il secondo giro, compromettendone i dati e obbligandoci ad accontentarci del solo primo giro.



**Figura 36 a) Fontana di Piazza Leonardo, Milano. b) Scultura al lato sud della fontana di Piazza Leonardo. ([www.http://blog.urbanfile.org/](http://blog.urbanfile.org/)).**



**Figura 37 Particolari del materiale petroso di cui sono costituiti la pavimentazione e la fontana (foto da rilievo).**

La fontana di Piazza Leonardo è stata un'interessante location anche per il materiale scabroso con il quale è stata realizzata la scultura, il cornicione e per il tipo di pavimentazione che circonda tutto il perimetro. Il materiale di cui è costituita la scultura

ed il cornicione è granito grigio, mentre la pavimentazione è costituita da sampietrini o parti di pietra in generale. Entrambi i materiali sono opachi alla luce del sole e sono caratterizzati da molte variazioni cromatiche, quindi ideali alla ricerca di feature da parte dell'algoritmo SIFT (paragrafo 5.3.6.1). Nella *Figura 37* si osservino in a) i particolari del materiale di cui è costituita la pavimentazione e in b) della fontana stessa.

Come già anticipato per acquisire le immagini Depth e le immagini RGB è stato utilizzata l'ultima versione del sensore Microsoft Kinect. Nel capitolo 3.2 vi è una presentazione dettagliata delle sue componenti e del suo funzionamento. Per quanto riguarda la rilevazione del segnale GPS invece si è utilizzato il ricevitore U-blox AEK6T, sensore GPS di basso costo (*Figura 38 c*)), e il ricevitore a doppia frequenza GPS LEICA GS14 (*Figura 38 a*)). La stazione totale utilizzata è la Multistation LEICA MS60 (*Figura 38 b*)) ed il prisma sempre della LEICA (*Figura 38 d*)).



**Figura 38 a) Il ricevitore a doppia frequenza GPS LEICA GS14 con apposito treppiede b) Multistation LEICA MS60 c) Sensore GPS miniaturizzato U-blox. d) prisma LEICA.**

Nel capitolo 5.2 è stata più volte sottolineata la necessità di fissare gli strumenti Kinect, Ublox, antenna GPS a doppia frequenza e il prisma ad una struttura rigida indeformabile in modo tale da formare un sistema in cui le soluzioni alla traiettoria fossero solidali tra loro e quindi confrontabili. Per lo scopo si è utilizzata una palina a cui gli strumenti sono stati solidamente fissati tramite l'utilizzo di apposite staffe. Come anticipato nel paragrafo (rilievo 5.2) è stato di fondamentale importanza che gli strumenti per la ricezione del segnale GPS siano stati disposti orizzontalmente e che siano stati posti ad altezza tale da non poter essere oscurati da persone o da altri strumenti durante il rilievo. Quindi l'antenna GPS a doppia frequenza è stata fissata nell'apposita vite sopra il prisma il quale a sua volta è stato fissato sul punto più alto

della palina. Non disponendo di un'altra palina si è optato per fissare il Kinect sotto il prisma. il Kinect è stato disposto in modo tale che le due camere IR e RGB fossero alla stessa altezza e rivolte leggermente verso il basso, a riprendere cioè oltre che parte di scena verticale anche parte del suolo. Questo accorgimento ha permesso di avere per ogni immagine parte di oggetto (appunto la pavimentazione) assicurata per la fase di matching tra foto consecutive. Il sensore U-blox è stato fissato con del semplice nastro adesivo in corrispondenza della camera RGB, sulla parte superiore del Kinect in modo minimizzare la distanza tra il centro di presa del sensore Kinect, rispetto a cui si voleva determinare la traiettoria, e il centro di fase dell'antenna U-blox. Questo accorgimento fa sì che la distanza dei due centri degli strumenti sia di circa 2-3cm, cioè effettivamente minore in confronto alla precisione della soluzione attesa dai due strumenti, per entrambi nell'ordine dei 10-20cm. Con questo accorgimento è possibile confrontare le traiettorie direttamente (dopo la loro georeferenziazione nel medesimo sistema di riferimento) senza considerare alcun vettore eccentricità tra i due sensori. Infine la palina stessa, con tutti gli strumenti collegati ad essa, è stata bloccata con fascette in plastica alla struttura di un carrellino di 4 ruote. Il carrellino è stato utilizzato come mezzo per il trasporto del sistema rigido ed è risultato sufficiente capiente anche per il posizionamento del laptop che gestisce il Kinect e l'antenna U-blox e della batteria con cui è alimentato il device Microsoft. In *Figura 39* si osservi la disposizione degli strumenti sul carrellino.





***Figura 39 Carrellino di 4 ruote su cui poggiano f) laptop e batteria e). Ed a cui è stata fissata la palina tramite legacci. Sulla pallina troviamo a) il sensore Kinect b) il sensore U-blox c) il prisma ed infine l'antenna GPS a doppia frequenza.***

L'integrazione delle soluzioni ottenute da Kinect e dal sistema GPS necessita la conoscenza dell'ora dell'acquisizione dei due strumenti. Kinect salva le immagini su laptop nominandole con il loro orario di acquisizione interno al Kinect stesso. I dati GPS acquisiti invece si riferiscono all'orario dell'orologio GPS. Ai fini della soluzione di posizionamento cinematico è stato utilizzato il tempo PC, cioè il tempo in cui un dato proveniente da uno strumento veniva immagazzinato sul laptop. Il tempo PC è quindi stato utilizzato da ponte per sincronizzare temporalmente i dati acquisiti dal Kinect e quelli dal GPS. Il tempo PC corrispondente alle acquisizioni del Kinect è stato salvato su un file di testo. In particolare i tempi di acquisizione delle immagini Depth e RGB

sono stati salvati nella scala dei tempi di acquisizione del Kinect, insieme al corrispondente tempo PC. Per minimizzare i tempi di latenza a qualche millisecondo i tempi sono stati salvati nella memoria del PC e, solo al termine dell'acquisizione stessa, sono stati scritti in un file di testo. I tempi di acquisizione del sensore U-blox sono stati invece acquisiti con un apposito logger che restituisce in output un file testuale contenente il tempo GPS e il corrispondente tempo PC di ognuna delle acquisizioni fatte. Nel paragrafo 5.1.2 è discussa la trascurabilità della latenza tra tempo di acquisizione effettiva del segnale e tempo PC che corrisponde al tempo di salvataggio su laptop.

## 6.2 ACQUISIZIONE DEI DATI

Il rilievo è stato svolto in data 24 Marzo 2016 in una giornata di cielo sereno. La prima fase del rilievo è stata occupata dal posizionamento statico dei due punti in cui sono stati posti i treppiedi. Come già anticipato nel paragrafo 5.2, per disporre di una traiettoria di riferimento aggiuntiva rispetto a quella dell'antenna GPS è stato scelto di tracciare il percorso del prisma con stazione totale. Per inseguire il prisma durante il rilievo del giro la stazione totale è stata posizionata sul treppiede disposto a circa 10 metri dal lato Nord della vasca della fontana. Il secondo treppiede è stato posizionato nell'angolo sud-ovest della fontana ad una distanza di una decina di metri da essa. Questa posizione del treppiede è servita come punto di orientamento del rilievo con stazione totale e, allo stesso tempo, da tale posizione sono state misurate le coordinate

di alcuni target opportunamente fissati sul lato più alto della scultura della fontana, in prossimità del punto di partenza del rilievo cinematico (si vedano i target in *Figura 40*). Per ognuno dei due punti è stato eseguito un rilievo GNSS statico, stando su ognuno dei punti per 15 minuti circa: tale operazione si è resa necessaria con lo scopo di inquadrare i dati misurati dalla stazione totale in un sistema di riferimento globale. Con lo scopo di garantire la possibilità di ricostruire la traiettoria anche per via puramente fotogrammetrica sono state misurate le posizioni di alcuni target lungo i lati della fontana: le coordinate dei target sono state misurate con la stazione totale. In *Figura 40* si veda una foto dei target disposti sul lato ovest della scultura.



*Figura 40 Target posizionati sul lato ovest della scultura.*

Come spiegato nel paragrafo 5.1.3 per poter confrontare le traiettorie determinate da Kinect e da sistema GPS è necessario eseguire la calibrazione geometrica del sistema integrato. Essa è stata ottenuta acquisendo con tutti gli strumenti disponibili la traiettoria del quadrato di calibrazione. Dalle traiettorie solidali ma in sistemi di riferimento diverso è stata determinata la rototraslazione da applicare ad una delle due traiettorie in modo che esse fossero nello stesso sistema di riferimento. Il quadrato di calibrazione è stato disposto sul fianco ovest della scultura, in modo tale che dalle immagini acquisite da Kinect fosse possibile osservare il maggior numero di target presenti sulla stessa facciata della scultura. Quindi utilizzando del nastro adesivo si è tracciato il quadrato a terra in modo da avere una base di riferimento dello spostamento che avrebbe dovuto avere il carrellino. Il lato del quadrato disegnato era di lato di circa 3 metri, un valore sufficientemente più grande rispetto alla sensibilità del posizionamento che si aveva ipotizzato di ottenere con gli strumenti GPS e con il Kinect. Si è fatto in modo di disporre il quadrato in modo che il suo primo lato coincidesse con la parte iniziale del rettilineo del giro.

Gli spigoli del quadrato sono stati posizionati nel sistema di riferimento del sistema GPS (quindi Locale e WGS84) tramite posizionamento statico con gli strumenti GPS di circa 2 minuti per lo *spigolo1* ed n minuto per gli *spigoli2, 3 e 4*. Tra uno spigolo e l'altro si è eseguito un posizionamento cinematico dei lati. Durante l'acquisizione dei segnali GPS sono state acquisite anche le immagini Depth ed RGB con frequenza di circa 5 immagini per secondo, oltre alla traiettoria tracciata dalla stazione totale ad

inseguimento del prisma. In *Figura 41* si osservi la dislocazione del *quadrato di calibrazione* e degli spigoli rispetto alla fontana.



**Figura 41** *Dislocazione del quadrato di calibrazione e dei suoi spigoli rispetto alla fontana* ([www.googleheart.com](http://www.googleheart.com)).

Dopo aver acquisito i dati per il quadrato di calibrazione si è passati al posizionamento statico di circa due minuti del punto di partenza del giro, che è stato scelto essere lo *spigolo1* del quadrato. Prima del posizionamento statico dello *spigolo1* è stata fatta iniziare anche l'acquisizione delle immagini da Kinect. Quindi si è passati alla fase di rilievo del giro attorno alla fontana. Come già spiegato nel paragrafo 5.2 è stato

fondamentale muovere il carrellino a velocità costante e a far attenzione che nulla passasse sia nel campo visivo tra stazione totale e prisma sia nella scena ripresa dalle camere RGB ed IR (si veda paragrafo 5.2). Si è fatto concludere il primo giro esattamente nello *spigolo* del quadrato di calibrazione in cui infine è stato svolto un ulteriore posizionamento statico di circa due minuti.

Come già anticipato l'intenzione era quella di svolgere più di un giro tuttavia un evento imprevedibile ha smosso il treppiede sul lato nord in cui era posizionata la stazione totale per l'inseguimento del prisma. Essendo stato smosso dalla sua posizione originale, poiché il punto di stazione non era stato segnalizzato a terra con apposito chiodo topografico, sarebbe stato impossibile confrontare le nuove misure acquisite con quelle precedenti, a meno di non ripetere nuovamente il rilievo statico GNSS.

## **6.3 PROCESSAMENTO DEI DATI**

Nel processamento dei dati viene discusso principalmente la modalità di trattamento del dato fotogrammetrico. Il cuore del presente lavoro di tesi consiste appunto nel trattare le immagini RGBD per ricavarne la traiettoria dei centri di presa lungo il perimetro della fontana. Il trattamento del dato GPS invece verrà per lo più citato considerata l'importanza per l'implementazione del filtro di Kalman.

Il processamento dei dati fotogrammetrici si suddivide sostanzialmente nelle fasi di selezione delle immagini, di processamento del dato grezzo, cioè delle immagini così

come sono state acquisite, e quella del processamento delle rototraslazioni con cui è stata calcolata la traiettoria finale.

### **6.3.1 Selezione delle coppie di immagini**

Durante l'intero rilievo sono state acquisite immagini con frequenza di 5 fps, il che significa che durante le soste per il posizionamento statico degli spigoli del quadrato e dell'inizio e della fine del giro sono state acquisite un numero pari ad almeno 300 immagini tra loro praticamente identiche. Nel caso in cui la sosta risultasse più lunga il numero ovviamente è maggiore. Ai fini della determinazione della traiettoria finale non è assolutamente utile possedere un pacchetto di immagini uguali tra loro. Se per il posizionamento con sistemi GPS all'aumentare del tempo di stazionamento durante la fase di acquisizione del segnale aumenta anche la ridondanza della sua soluzione, per il metodo fotogrammetrico proposto semplicemente non ha senso sapendo che già si era fermi e visto il tempo richiesto per il calcolo della soluzione. Questa constatazione rappresenta la sostanziale differenza tra metodo satellitare e metodo fotogrammetrico. In quello satellitare la soluzione è la posizione, mentre in quello fotogrammetrico proposto la soluzione è lo spostamento della posizione da cui le due coppie di immagini RGB e depth sono state acquisite (o per meglio la velocità nel tratto tra l'acquisizione delle due immagini). Quindi se per il primo metodo non muoversi significa aumentare la ridondanza della soluzione per il secondo metodo è conveniente, in termini di tempo risparmiato di calcolo, che tra due immagini acquisite in tempi diversi ci sia un effettivo spostamento del centro di presa.

Come primo passo del processamento del dato grezzo quindi si considera la suddivisione delle immagini nei pacchetti di interesse, cioè quello del primo giro e quello del quadrato di calibrazione. Quindi si passa all'esclusione di tutte quelle immagini in cui non si è effettuato alcuno movimento del carrellino. Il processo di deselezionazione è possibile dall'osservazione delle immagini RGB. Si risale alla relativa Depth accoppiata alla RGB deselezionata osservandone il tempo di acquisizione, che è precedente di alcuni millesimi di secondo rispetto alla RGB accoppiata.

Una volta selezionate le sole coppie di immagini in cui si è verificato effettivamente un movimento si è poi passati alla ricerca di quelle immagini rese inutilizzabili a causa dell'attraversamento di persone/bambini. La frequenza effettiva di acquisizione, pari a 5 fps permette di evitare che si verificano buchi derivanti dall'esclusione di immagini nelle quali erano presenti soggetti in movimento all'interno che avrebbero potuto portare l'impossibilità di determinare la rototraslazione tra coppie estreme dell'esclusione.

Si è valutato che allineamento di immagini RGBD ottenute da Kinect potesse raggiungere precisione dell'ordine di 5-10cm. Da questo punto di vista sarebbe stato inutile determinare la rototraslazioni per coppie di immagini acquisite consecutivamente con la frequenza di 5 fps, anche perchè la velocità con cui è stato mosso il carrellino era piuttosto bassa. Per completare un intero giro corrispondente a circa 40 metri di distanza percorsa sono stati impiegati 5 minuti circa. Facendo un rapido calcolo significa che tra coppie di immagini consecutive a frequenza di 5 fps si stima uno spostamento medio di 2 o 3 centimetri, valore minore rispetto alla precisione attesa dello strumento



(paragrafo descrizione Kinect 3.2). Per questo motivo si è deciso di sottocampionare le immagini. In questa elaborazione dei dati si è scelto di sottocampionarle per una coppia Depth RGB ogni 20 coppie. Come verrà spiegato più avanti nel paragrafo per alcuni tratti critici è stato necessario diminuire il sottocampionamento per l'impossibilità da parte di SIFT e ICP di determinare sufficienti punti di legame.

### 6.3.2 Processamento del dato grezzo

Dopo che sono state selezionate le coppie di immagini considerate accettabili è stato eseguito il trattamento dell'immagine grezza per renderla adatta al processamento da cui ne verrà estratta l'informazione della rototraslazione. Il processamento dell'immagine grezza è stato svolto interamente nel software Matlab<sup>®</sup>.

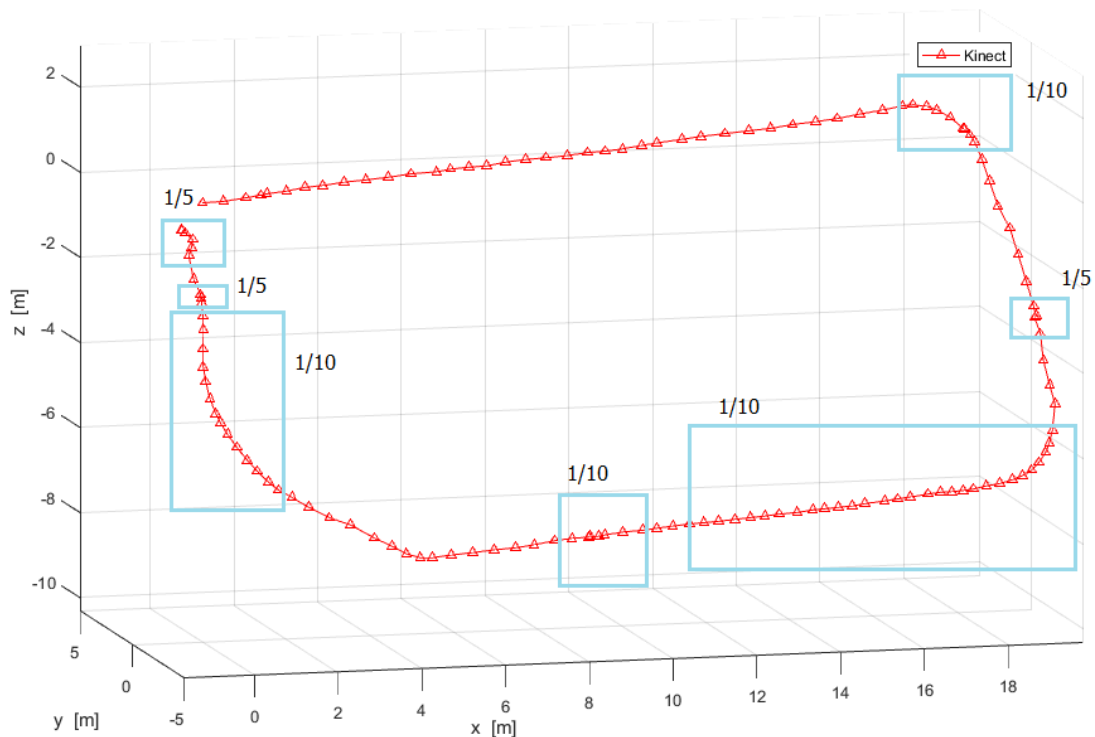
Per ogni immagine RGB e Depth è stata eseguita la correzione alle distorsioni ottiche già descritte nel paragrafo 5.1.1. La funzione utilizzata si chiama *undistorImage* ed appartiene al pacchetto applicativo per la processione delle immagini. Per quanto riguarda le immagini Depth è stata eseguita una funzione Matlab<sup>®</sup> per il ricampionamento *imresize*, che ha permesso di sovracampionare l'immagine con un rapporto 4:1, permettendo di densificarne la nuvola di punti (paragrafo 5.3.3). Prima della creazione della nuvola di punti si è proceduto specchiando le immagini Depth e RGB: questa operazione è necessaria poiché esse sono acquisite in modo speculare rispetto alla realtà. A questo punto si è passati alla creazione delle nuvole di punti partendo dalle immagini Depth. Nel paragrafo 5.3.4 sono presentate le equazioni

necessarie allo scopo. Per ogni nuvola di punti si è effettuato una proiezione della stessa sull'immagine RGB a cui l'immagine Depth è accoppiata. Per la proiezione sono state utilizzate le equazione di collinearità di presa (paragrafo 4.1) in cui si tiene conto anche dell'OE relativo tra le camere IR e RGB. Quindi è stata creata un nuovo tipo di immagine nominata RGBD (RGB più Depth) già descritta nel paragrafo 5.3.5.

Dalle immagini RGBD si è passati alla determinazione delle rototraslazioni tra le coppie di esse. Risulta non di poco conto il fatto che il blocco fotogrammetrico di immagini RGBD sia piuttosto ingombrante. Per lanciare l'algoritmo align2RGBD per la ricerca della rototraslazione tra immagini consecutive, a causa di limiti del carico computazionale del laptop in uso (laptop ASUS con processore AMD E-450 APU con radeon(tm) HD Graphics 1.65 GHz di 3.60 GB di RAM utilizzabile, con sistema operativo a 64bit)) è stato necessario suddividere il blocco fotogrammetrico in ulteriori sottoblocchi non superiori a 30 immagini RGBD. La suddivisione in sottoblocchi ha imposto il problema del loro collegamento affinché non vi fosse la perdita della rototraslazione tra coppie adiacenti di blocchi diversi. Quindi le immagini RGBD estreme di ogni blocco coincidono con quelle estreme dei blocchi adiacenti.

Come già anticipato nel paragrafo 0 si sono presentati alcuni tratti critici in cui il sottocampionamento di una coppia di immagini ogni venti è stato eccessivo. Nel paragrafo 5.3.6.1 riservato alla descrizione di SIFT si accenna al fatto che tra le immagini è necessario ci sia una corrispondenza di almeno 80% dei punti. Tale percentuale è stata assicurata in tratti in cui il carrello veniva mosso a velocità costante lungo tratti rettilinei. Nei tratti curvilinei dove la scena cambiava rapidamente o

dove è stato necessario eliminare alcune sequenze di immagini causa attraversamento di persone invece questa condizione non era rispettata. Per questi tratti specifici è stato necessario diminuire il sottocampionamento da 1 coppia ogni 20 a una ogni 10 (1/10), ed in alcuni casi fino a una ogni 5 (1/5). Per questi tratti è valso lo stesso approccio adottato dell'allacciamento tra sottoblocchi adiacenti descritto sopra. In *Figura 42* è raffigurata la traiettoria dei centri di presa ottenuta dall'elaborazione delle immagini da Kinect. Il sistema di riferimento rispetto a cui è rappresentata la traiettoria in *Figura 42* è quello del sistema Kinect, che ha l'origine nel centro di presa della camera RGB della prima immagine RGBD del blocco fotogrammetrico, quindi asse x rivolto verso la larghezza dell'immagine e l'asse z come l'altezza dell'immagine e l'asse y per terna destrorsa. In blu sono indicati con i rispettivi valori di sottocampionamento i tratti sottoposti a sottocampionamento minore rispetto ad una coppia ogni 20. Si noti come effettivamente in questi tratti lo spostamento dei centri di presa risulti minore rispetto ai restanti tratti.



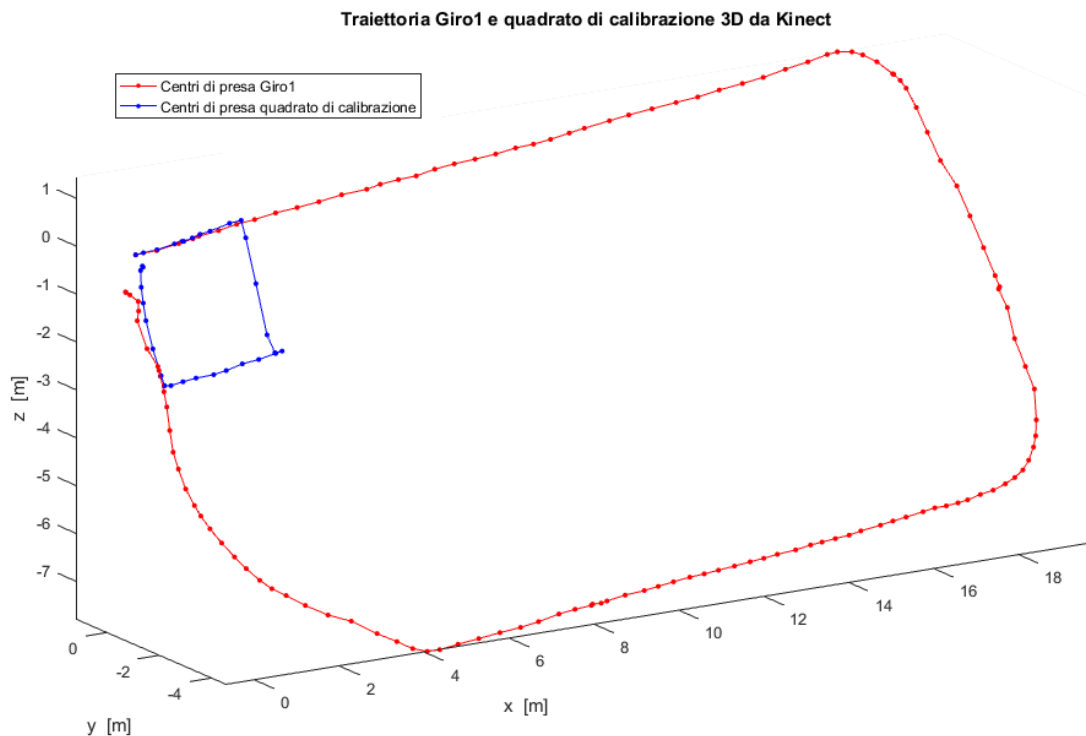
**Figura 42** Traiettoria dei centri di presa determinata dalle soluzioni dell'elaborazione delle immagini da Kinect. Nei riquadri blu con le opportune frazioni sono indicati i tratti in cui è stato necessario diminuire il sottocampionamento delle immagini.

### 6.3.3 Processamento delle rototraslazioni

Il risultato ottenuto dall'algoritmo *align2RGBD* per ogni sottoblocco o tratto di sottocampionamento minore si suddivide nelle matrici di rototraslazione 3x4 (paragrafo rototraslazioni 4.4) allocate alla variabile *Rt*, e nella precisione con cui queste sono state determinate (nella stessa unità di misura della componente traslazionale), allocate nella variabile *Eps*. Considerato che variabili *Rt* ed *Eps* si riferiscono ai sottoblocchi o ai tratti di diverso sottocampionamento prima di tutto è stato necessario collegare tra loro tutte le soluzioni rispettando l'ordine di acquisizione. In questa fase si è rivelato di grande

aiuto riferirsi all'informazione del tempo PC delle immagini RGBD allocato ad una variabile chiamata *time\_RGBD* della stessa lunghezza delle immagini RGBD che è stata creata nella fase di selezione e sottocampionamento delle immagini. Questa informazione inoltre è indispensabile per l'implementazione del filtro di Kalman.

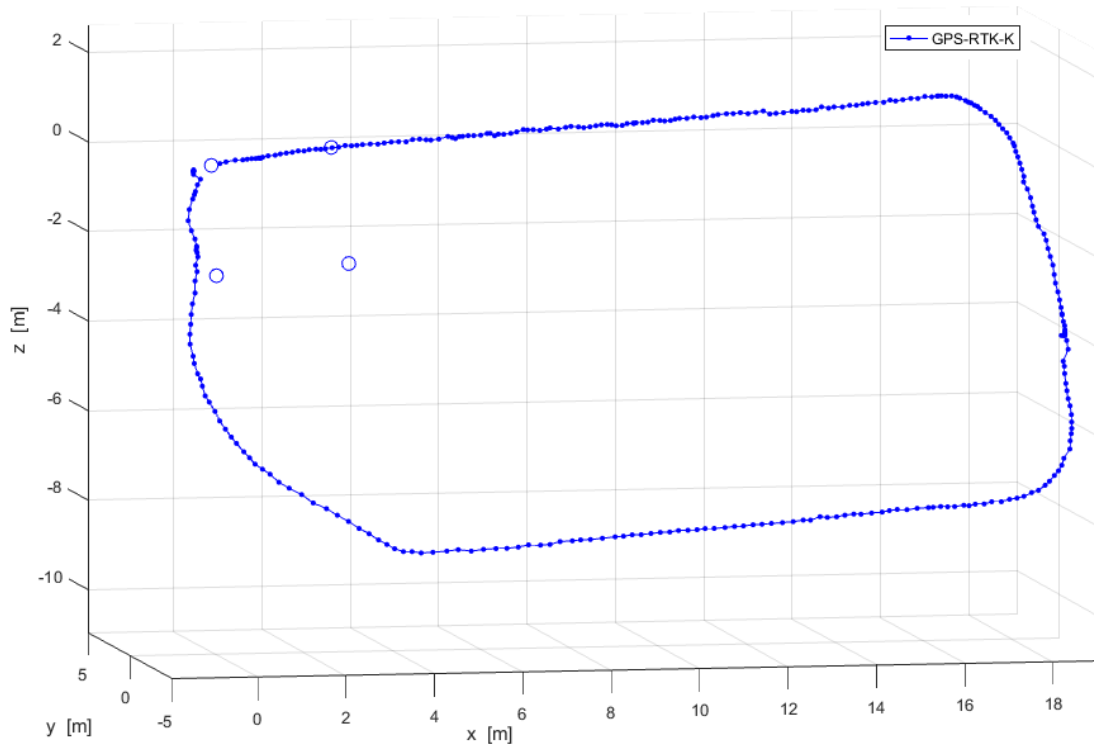
Quindi dalla lista completa di rototraslazioni  $Rt$  e di relative accuratèzze  $Eps$  è stato possibile determinare dalla formula presentata nel paragrafo 4.4 la traiettoria di tutti i centri di presa del blocco fotogrammetrico *Giro1* e del blocco *quadrato di calibrazione*. Si ricordi ancora una volta che questo risultato è nel sistema di riferimento della prima immagine RGBD acquisita nel punto *spigolo1*. In *Figura 43* si osservi la traiettoria dei centri di presa in *Giro1* e *quadrato di calibrazione*. Si noti come avendo inclinato il sensore Kinect verso il basso per acquisire parte di asfalto si ottenga una traiettoria obliqua della stessa inclinazione. In *Figura 43* risulta evidente come il *lato1* del *quadrato di calibrazione* coincida con il tratto iniziale del rettilinea di *Giro1*. Il sistema di riferimento è lo stesso adottato per quello di *Figura 42*.



**Figura 43** Traiettoria di Giro 1 e quadrato di calibrazione ottenuto dalle traiettorie delle immagini Kinect nel sistema di riferimento Kinect.

Riferendoci ai tempi PC delle immagini RGBD presenti nella variabile *time\_RGBD* è stato impostato il calcolo delle soluzioni al posizionamento statico per gli spigoli del *quadrato di calibrazione*, sia per l'inizio che per la fine del *giro1*; quindi per quello del posizionamento cinematico del *Giro1*. Nel selezionare gli intervalli di tempo PC dei dati del segnale GPS si è fatta attenzione nel considerare l'intervallo di tempo in cui effettivamente il carrellino era fermo per il posizionamento statico e in cui il carrellino era effettivamente in movimento per il posizionamento cinematico. I dati GPS dell'antenna GPS a doppia frequenza e di U-blox sono stati analizzati dal software goGPS (paragrafo 5.2.2).

Si consideri che i risultati ottenuti da U-blox sono stati talmente insoddisfacenti rispetto alla precisione attesa che si è scelto di scartarne completamente i risultati. Si è quindi proseguito il lavoro considerando solo i risultati dell'antenna GPS a doppia frequenza, analizzati secondo processo rtk. I risultati sono riportati in *Figura 44* dove si osserva la traiettoria del centro di fase dell'antenna GPS a doppia frequenza per il *Giro1* e per il posizionamento statico degli spigoli del *quadrato di calibrazione* nel sistema di riferimento della traiettoria Kinec (si veda sistema di riferimento *Figura 42*). Il risultato GPS è stato determinato impostando una soluzione ad ogni secondo da qui il motivo per cui la traiettoria del *Giro1* risulta così fitta.

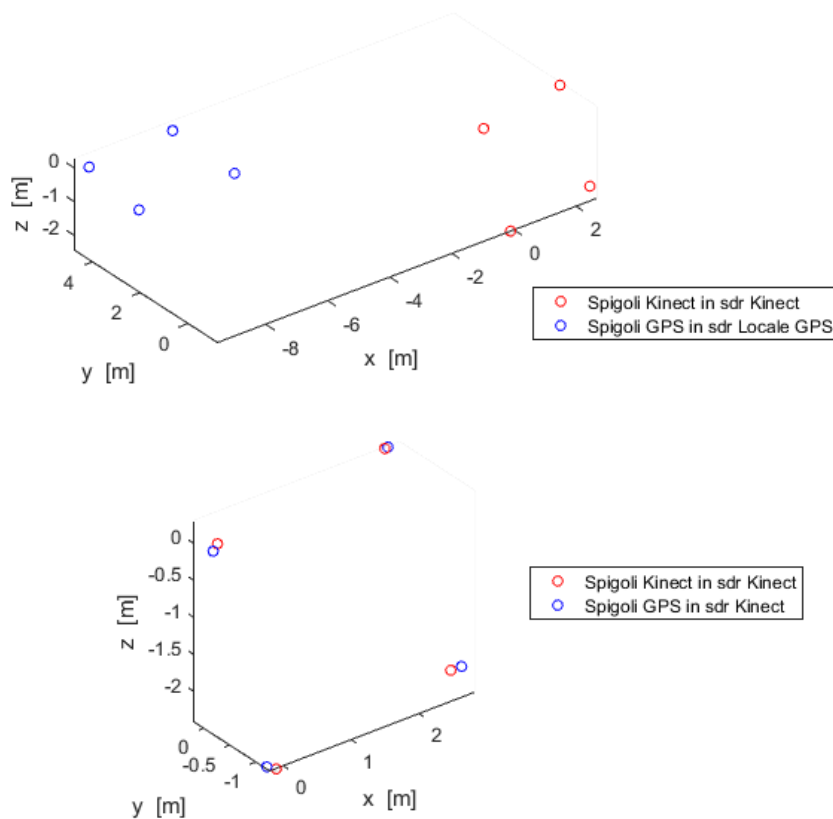


**Figura 44** Traiettorie dei centri di fase dell'antenna GPS a doppia frequenza e posizionamento statico per gli spigoli del quadrato di calibrazione nel sistema di riferimento della traiettoria della prima immagine del blocco di immagini RGBD.

A questo punto è stata determinata la rototraslazione per portare la traiettoria ottenuta dal sistema GPS dal sistema di riferimento Locale GPS al sistema di riferimento rispetto cui è definita la traiettoria ottenuta dal Kinect. Per determinare i 6 parametri della trasformazione è stata utilizzata la funzione *Helmerd3d* disponibile in codice Matlab<sup>®</sup> utilizzando come input le traiettorie dei 4 spigoli del *quadrato di calibrazione* ottenute dal metodo fotogrammetrico con Kinect e dal metodo GPS con antenna GPS a doppia frequenza (paragrafo 6.2). Ognuna delle due traiettorie degli spigoli sono espresse secondo il proprio sistema di riferimento, cioè rispettivamente quello riferito alla prima immagine del blocco fotogrammetrico del Kinect e quello locale GPS. In *Figura 45* le



posizioni dei 4 spigoli ottenuti dal metodo fotogrammetrico e dal metodo GPS. In a) i punti blu sono le posizioni degli spigoli ottenute dall'antenna GPS a doppia frequenza nel sistema GPS Locale mentre i punti rossi sono le posizioni degli spigoli ottenute da Kinect nel sistema di riferimento dell'immagine RGBD acquisita dallo *spigolo1*. In b) i punti blu corrispondono agli spigoli acquisiti con GPS dopo aver subito la trasformazione che li porta a sovrapporsi agli spigoli rossi acquisiti con Kinect nel sistema di riferimento della traiettoria del Kinect.



**Figura 45** Spigoli del quadrato di calibrazione ottenuti dai metodi fotogrammetrici e GPS espressi in a) nel proprio sistema di riferimento ed in b) dopo aver applicato agli spigoli ottenuti da GPS in sdr Locale GPS la rototraslazione per portarli nel sdr Kinect.

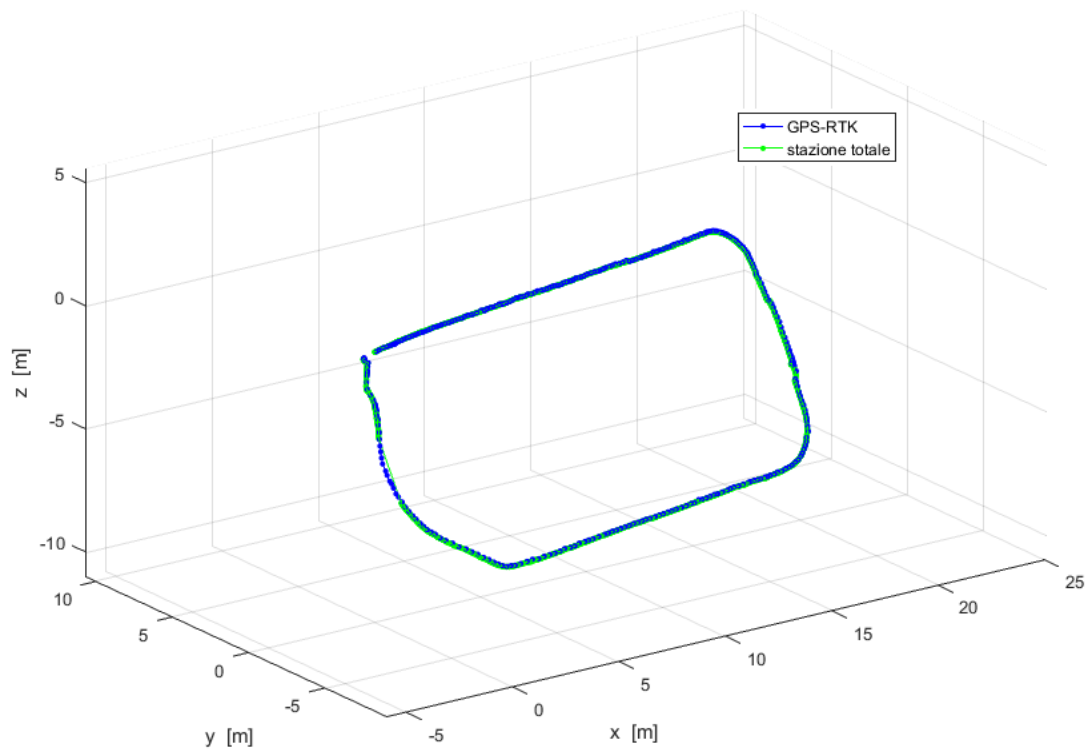
In *Tabella 5* sono presenti le medie delle deviazioni delle coordinate degli spigoli determinati con il metodo fotogrammetrico da sensore Kinect rispetto agli spigoli determinati dal metodo GPS con antenna a doppia frequenza. In *Figura 45 b)* si osservino gli spigoli per il calcolo della *Tabella 5*.

	<b><i>Spigolo1</i></b>	<b><i>Spigolo2</i></b>	<b><i>Spigolo3</i></b>	<b><i>Spigolo4</i></b>
<b>Medie delle deviazioni [m]</b>	0,127	0,075	0,145	0,121

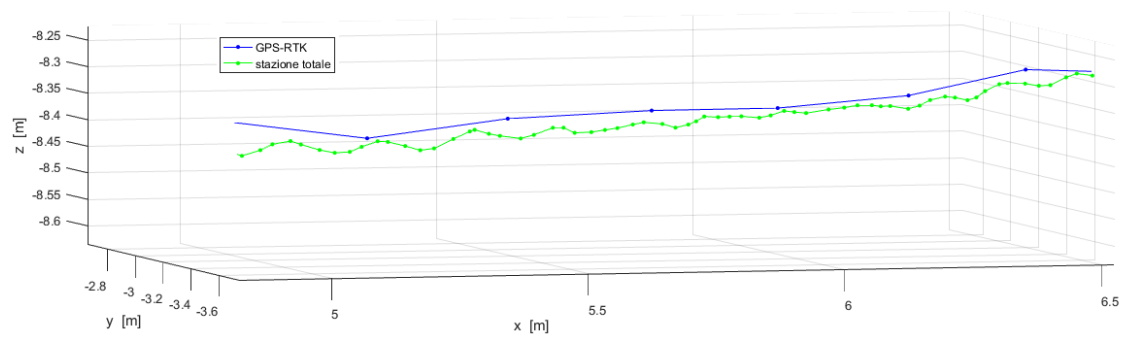
***Tabella 5 Deviazione tra le coordinate degli spigoli ottenuti dai metodi fotogrammetrico e GPS espressi nello stesso sistema di riferimento Kinect.***

Le medie delle deviazioni delle coordinate degli spigoli riportate in *Tabella 5* hanno valori che rientrano nelle precisioni attese dal posizionamento cinematico tramite l'uso di sensori lowcost come Kinect.

Si osservino in *Figura 46* la traiettoria ricavata dal sistema GPS con processamento del dato rtk (in blu) ed (in verde) la traiettoria ricavata dall stazione totale. Entrambe le traiettorie sono raffigurate nel sistema di riferimento della traiettoria Kinect.



**Figura 46** Traiettoria ricavata dal sistema GPS con processamento rtk (in blu) e traiettoria ottenuta dalla stazione totale (in verde).



**Figura 47** Particolare del secondo rettilineo delle traiettoria ottenute dal sistema GPS\_rtk (in blu) e dalla stazione totale (in verde).

Come messo in risalto in *Figura 47* (particolare del secondo rettilineo estratto dalla *Figura 46*) le due traiettorie sono molto simili tra loro. In particolare la traiettoria del prisma è di qualche centimetro direzione z più in basso rispetto a quella ottenuta dal GPS\_rtk a causa della composizione degli strumenti sulla palina (capitolo rilievo 6.1). La traiettoria ottenuta dalla stazione totale presenta un punto per ogni centimetro di spostamento e risulta più fitta rispetto a quella dell'antenna GPS, che presenta un punto per ogni secondo di acquisizione (vedi *Figura 47*). La traiettoria della stazione totale è sufficientemente fitta per rappresentare il moto oscillatorio del sistema rigido palina/carellino, effetto non visibile dalla traiettoria GPS\_rtk, che risulta più lisciata. Dall'interpolazione delle traiettorie rispetto al tempo di acquisizione del dato, tempo espresso nel tempo PC, è stato possibile determinare la media della deviazione delle due traiettorie nelle componenti x, y e z (nel sistema di riferimento della traiettoria del Kinect) ed i valori di picco massimo (vedi *Tabella 6*).

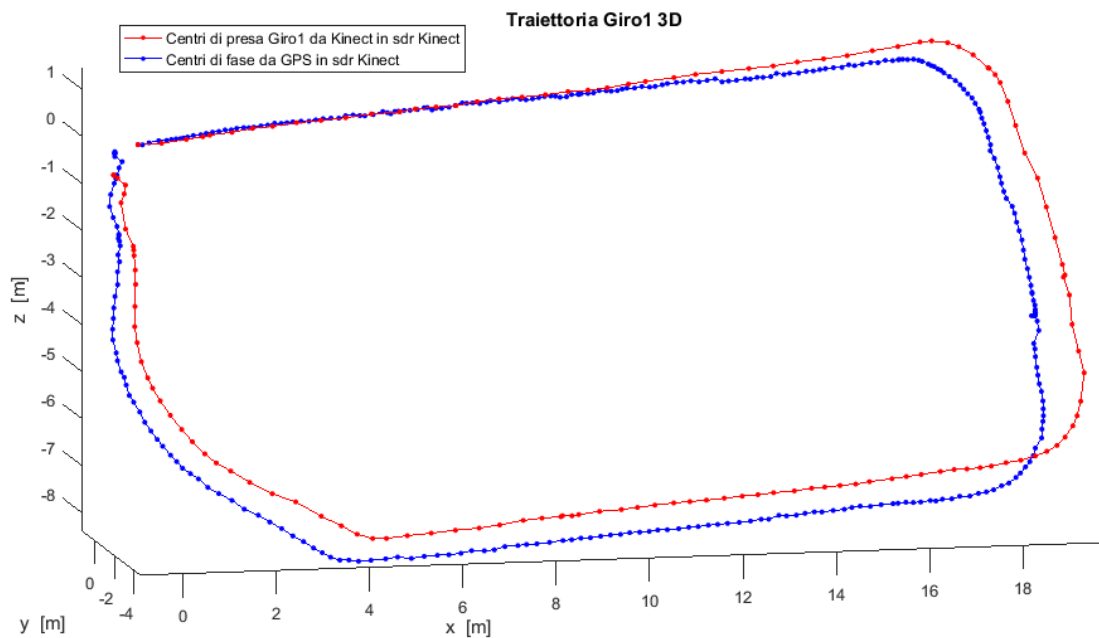
	<i>componente x</i>	<i>componente y</i>	<i>componente z</i>
<b>Media delle deviazioni [m]</b>	0.045	0.020	0.017
<b>Massimi di deviazione [m]</b>	0.190	0.053	0.136

*Tabella 6 Valori di deviazioni nelle componenti x, y e z del sistema di riferimento della traiettoria del Kinect tra traiettoria ottenuta dall'antenna GPS e la traiettoria della stazione totale.*

A parte qualche picco di 20 cm circa (dovute forse più alla traiettoria GPS che da quella della stazione totale), la media della deviazione nelle tre componenti è abbastanza al di sotto del livello di precisione atteso dalla traiettoria del Kinect. Da qui in avanti quindi si utilizza la traiettoria della stazione totale come traiettoria di riferimento rispetto a cui

verranno definite le medie delle deviazioni delle traiettorie nelle componenti x, y e z ottenute dal Kinect e dal sistema integrato.

Quindi è stata applicata alla traiettoria del *Giro1*, ricavata dal metodo GPS, la rototraslazione determinata dalla calibrazione geometrica del sistema integrato (paragrafo 5.1.3) e una traslazione per far sì che il suo primo punto coincida con il primo punto della traiettoria determinata con Kinect. In *Figura 48* si osservino la traiettoria ricavata dai dati acquisiti dall'antenna GPS a doppia frequenza (in blu) insieme alla traiettoria ricavata dalle immagini acquisite dal Kinect (in rosso) rappresentate nel sistema di riferimento della prima immagine del blocco fotogrammetrico del sensore Kinect.



**Figura 48** Traiettoria Giro1 da metodo fotogrammetrico e metodo GPS nel sistema di riferimento della prima immagine del blocco Kinect.

Le due traiettorie rappresentate in *Figura 48* risultano praticamente coincidenti per quasi tutto il rettilineo. La deviazione nella prima curva della traiettoria ottenuta dal Kinect rispetto a quella ottenuta dall'antenna GPS mette in risalto il problema di deriva che ci si aspettava di ottenere adottando il metodo fotogrammetrico, il quale determina la posizione del passo successivo dal calcolo dallo spostamento tra due immagini consecutive (paragrafo calcolo della traiettoria 5.3.7). Parte della deriva è composta anche un'importante differenza delle quote delle due traiettorie (differenza lungo  $z$ , quindi in *Figura 48* lungo l'asse  $z$ ). Questa deviazione potrebbe essere derivante anche dall'imprecisione con cui è stata determinata la trasformazione che ha portato la traiettoria GPS nel sistema di riferimento in cui è espressa la traiettoria Kinect, imprecisione già messa in risalto in *Tabella 5*, ma potrebbe essere anche frutto della scelta di aver inclinato il sensore Kinect verso il basso (paragrafo rilievo 5.2.1).

Le traiettorie ricavate dai due sistemi dovrebbero chiudersi nel punto di inizio, infatti il punto di partenza e di fine coincidono con lo *Spigolo1* (paragrafo rilievo 5.2). La traiettoria determinata da Kinect si chiude ad una distanza di circa 1 metro rispetto alla posizione originale, mentre quella dell'antenna GPS di mezzo metro (vedi *Figura 48*).

Infine è stato applicato il filtro di Kalman ai risultati delle traiettorie ottenute da Kinect e dall'antenna GPS a doppia frequenza. I tempi utilizzati per l'applicazione del filtro di Kalman sono quelli del PC associati alle immagini RGBD, cioè ai centri di presa della traiettoria ottenuta da Kinect (punti rossi in *Figura 48*) e quelli relativi alle soluzioni di posizionamento con metodo GPS (punti blu in *Figura 48*). Questi tempi PC sono stati

ordinati in ordine crescente in una variabile chiamata *time\_PC* costituita da due colonne. Nella prima colonna vengono allocati i valori dei tempi PC mentre nella seconda colonna viene salvato un elemento che renda riconoscibile a che metodo è associato il tempo PC allocato nella stessa riga.

Come già precedentemente osservato lo strumento Kinect osserva gli spostamenti (o velocità), mentre un'antenna GPS osserva le posizioni. Il filtro di Kalman serve per integrare i due tipi di osservazioni. Esso quindi sarà caratterizzato da un elemento di posizionamento statico che considera l'osservazione GPS, dipendente dalla precisione con cui è stata determinata la posizione, ed un elemento dinamico che considera lo spostamento (o la velocità) tra i punti da cui sono acquisite le immagini. In questo secondo caso l'osservazione è dipendente dalla bontà con cui sono state determinate le matrici di rototraslazione tra coppie di immagini RGBD. Quindi ci si aspetta un risultato in cui l'osservazione GPS fissa la posizione nello spazio, mentre l'osservazione Kinect ne dirige la direzione dello spostamento. Infine il risultato ottenibile dal filtro di Kalman dipende dall'ipotesi di moto scelta e da come essa sia effettivamente rispettata dai dati osservati. In questo elaborato di tesi è stata scelta l'ipotesi di moto rettilineo uniforme. La direzione rettilinea consona all'ipotesi non è stata rispettata nelle curve mentre quella di moto a velocità costante è stata oggetto di attenzione durante la fase di rilievo (paragrafo rilievo 5.2).

Il calcolo del filtro di Kalman è descritto nelle equazioni nel paragrafo 5.5. Grazie all'elemento presente nella seconda colonna della variabile *time\_PC* descritta sopra è stato possibile distinguere ad ogni iterazione quali delle rispettive matrici disegno

( $H_{Kin}$  e  $H_{GPS}$ ) e delle matrici di covarianza del rumore dell'osservazione ( $C_{Kin}^{vv}$  e  $C_{GPS_{Kin}}^{vv}$ ) utilizzare. A seconda del metodo corrispondente al tempo PC dell'iterazione i-esima è necessario cambiare anche il vettore delle osservazioni da implementare nel calcolo. Quindi si distingue tra il vettore delle osservazioni da metodo fotogrammetrico  $Y_{Kin_k}$  e quello ottenuto dalle osservazioni da metodo GPS  $Y_{GPS_g}$

$$Y_{Kin_k} = \begin{bmatrix} t_{x_k} \\ t_{y_k} \\ t_{z_k} \end{bmatrix} \quad Y_{GPS_g} = \begin{bmatrix} X_g \\ Y_g \\ Z_g \end{bmatrix}$$

in cui  $t_{x_k}$ ,  $t_{y_k}$  e  $t_{z_k}$  sono le componenti della traslazione k-esima determinata dalle rototraslazioni nella coppia k-esima di immagini RGBD provenienti da Kinect mentre  $X_g$ ,  $Y_g$  e  $Z_g$  sono le coordinate del posizionamento effettuato con l'antenna GPS a doppia frequenza.

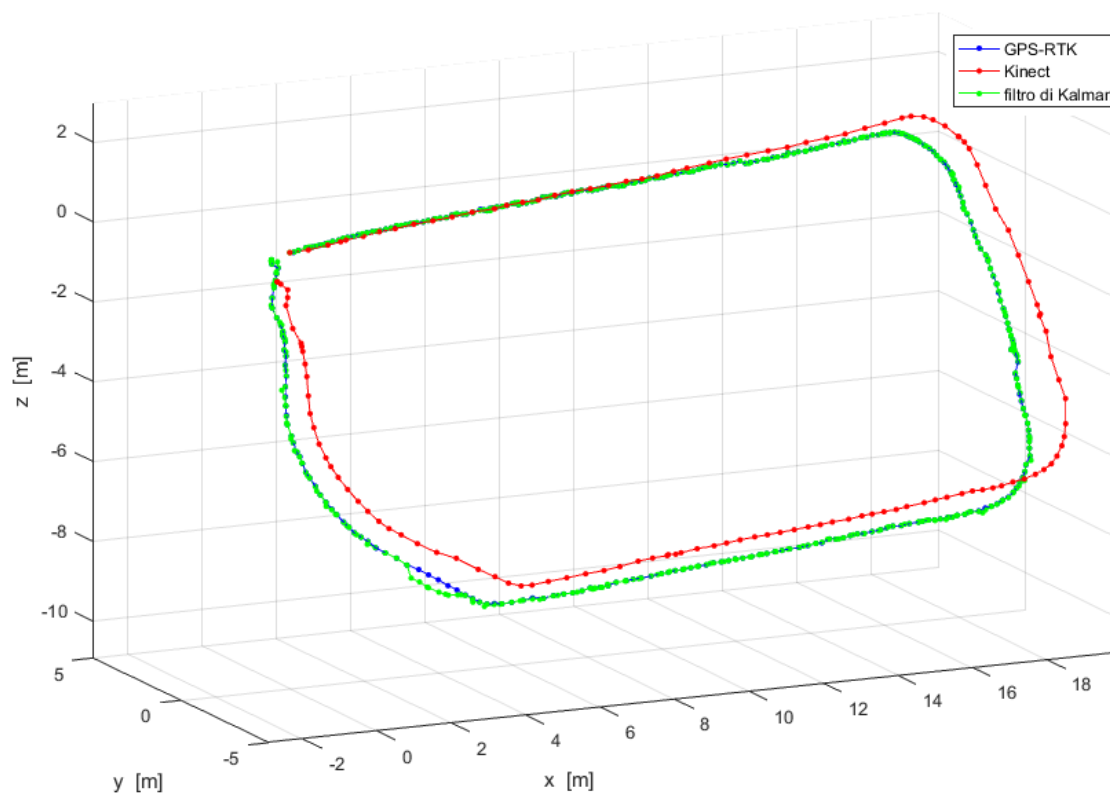
La matrice  $C_i^e$ , matrice di covarianza della stima della traiettoria al passo i-esimo è compilata utilizzando i valori della variabile  $Eps$  della rotazione k-esima, cioè i valori della bontà con cui è stata stimata la rototraslazione (ma di cui non conosciamo l'unità di misura) tra la coppia k-esima di immagini RGBD del blocco. La matrice  $C_{GPS_{Kin}}^{vv}$  è determinata applicando la matrice di rotazione della rototraslazione, determinata dai quadrati di calibrazione, alla matrice di covarianza delle osservazioni del segnale GPS.

Le matrici di covarianza nel filtro di Kalman svolgono il compito di regolare quanta fiducia si associa all'osservazione proveniente dallo strumento. Più la varianza è alta più il sistema è meno vincolato.



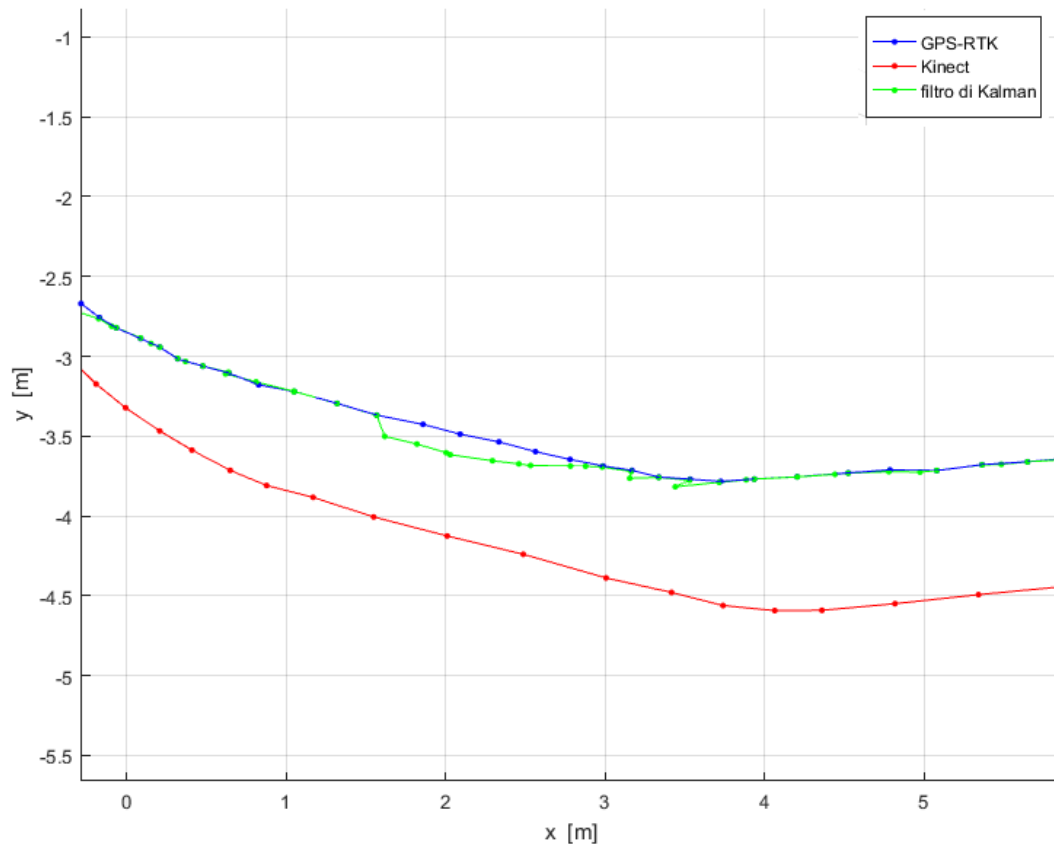
## 6.4 TEST1

Il primo test effettuato con il filtro di Kalman è stato realizzato usando tutti i punti delle traiettorie ottenute dal Kinect e dall'antenna GPS a doppia frequenza. Questo primo test non è di particolare interesse dal punto di vista dell'applicazione del sistema integrato per scopi di posizionamento cinematico in ambiente urbano in cui il segnale GPS è interrotto, ma è interessante per capire il comportamento. In *Figura 49* si osservi il risultato nel sistema di riferimento della traiettoria ottenuta dal Kinect.



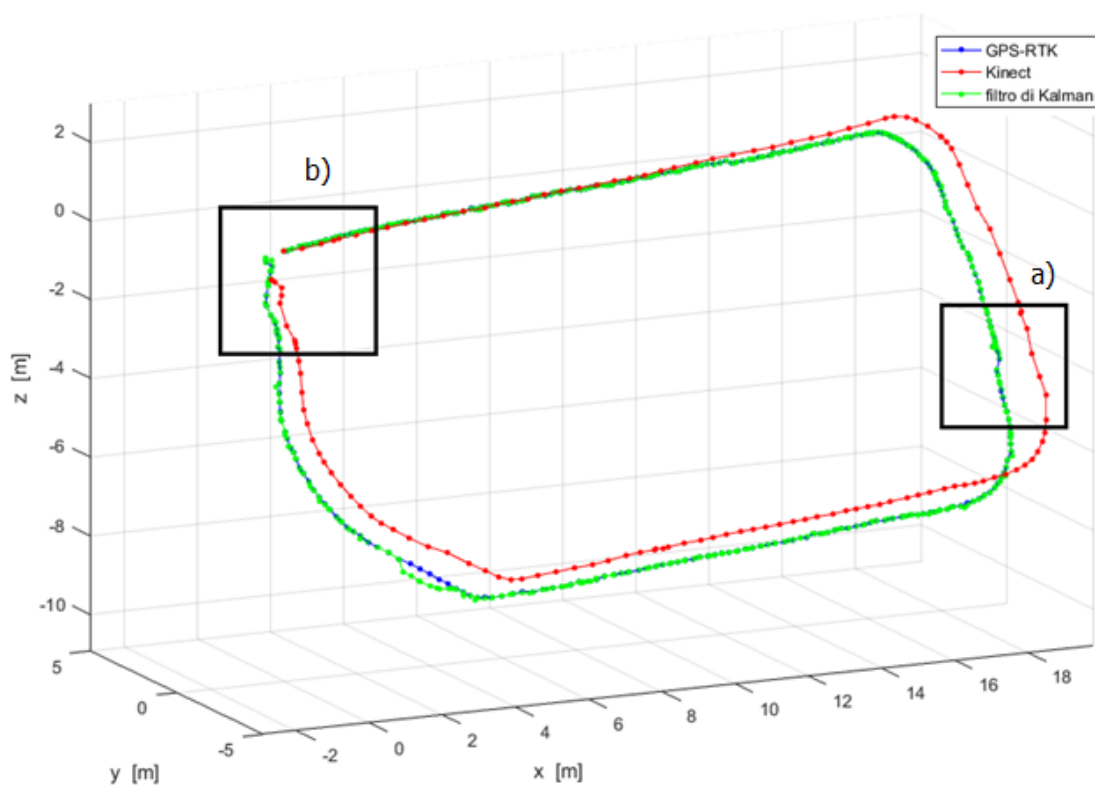
**Figura 49** Risultato del primo test sul filtro di Kalman in cui sono utilizzati tutti i dati disponibili delle traiettorie da Kinect e da GPS\_rtk (sistema di riferimento della traiettoria da Kinect).

Come ci si aspettava la traiettoria ottenuta dal filtro di Kalman è per lo più coincidente con quella ottenuta dal GPS. Questo è per lo più garantito dal fatto che le osservazioni di posizione da GPS sono molte di più rispetto a quelle di spostamento del Kinect. Si osserva tuttavia che in due occasioni si verifica una deviazione della traiettoria ottenuta dal filtro rispetto a quella ottenuta dal GPS. In particolare si osservi l'uscita dalla terza ed ultima curva, in cui la traiettoria ottenuta dal Kinect tende ad "allungare" la curva del filtro. Questo effetto è dovuto alla tendenza del filtro a "raddrizzare" la traiettoria finale. Il filtro lavora nell'ipotesi di moto uniforme rettilineo quindi interpreta la variazione della direzione della traiettoria come errore. Il risultato è un effetto di ritardo sulla variazione della direzione della traiettoria che determina il breve tratto deviato (vedi *Figura 50*). Per quanto riguarda questo aspetto si rimanda al Test2 nel paragrafo 6.5 dove questa tendenza risulta molto più evidente.

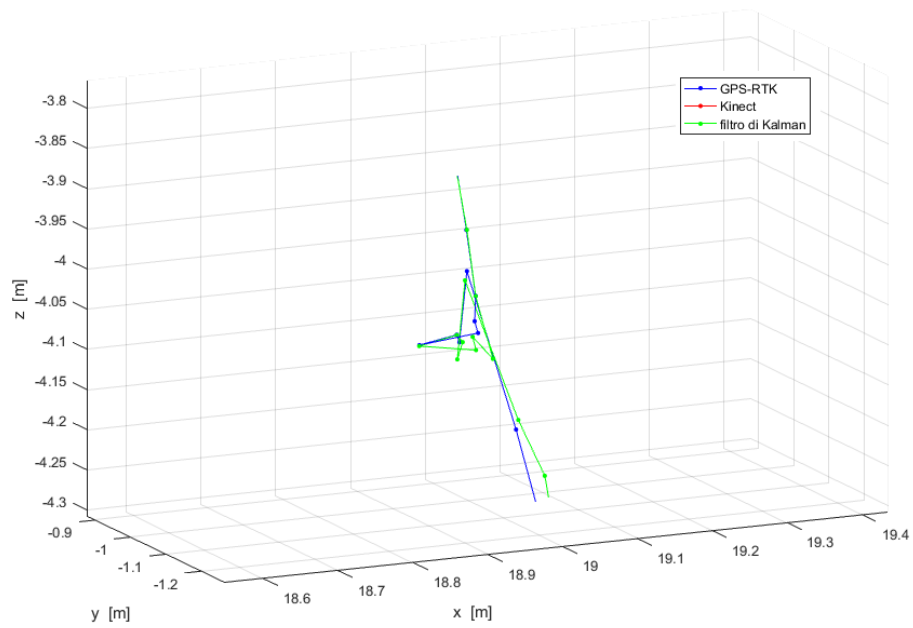


**Figura 50** Deviazione dalla traiettoria reale causata dall'ipotesi di moto rettilineo uniforme adottata nel Filtro di Kalman.

Nel riquadro *a)* di *Figura 51*, raffigurato anche in *Figura 52*, si osserva un'anomalia nel tracciamento della traiettoria. Proprio in quel punto del rilievo si è verificato l'attraversamento di una persona, che è presente in posizioni diverse nella breve sequenza di immagini dal sensore Kinect. Nello stesso momento il carrellino è stato arrestato per pochi secondi. Evidentemente la persona in movimento nelle immagini è stata interpretata dall'algoritmo align2RGBD come un falso movimento del centro di presa dell'immagine. Questa informazione si scontra con quella della traiettoria GPS e determina l'anomalia in *Figura 52*.

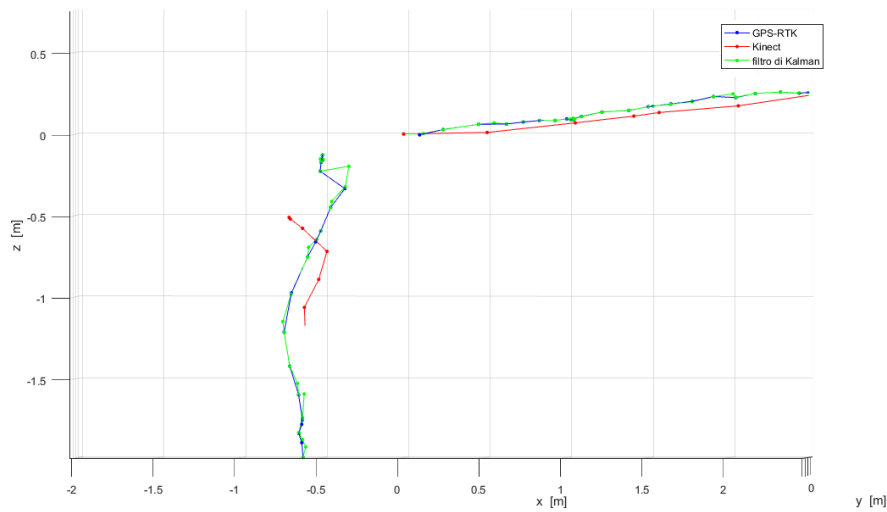


**Figura 51 Risultato del primo test sul filtro di Kalman in cui sono utilizzati tutti i dati disponibili delle traiettorie da Kinect e da GPS\_rtk (sistema di riferimento della traiettoria da Kinect). In a) è individuata un'imprecisione della traiettoria dovuta ad un valore di spostamento della componente dinamica anomalo. In b) il risultato della chiusura della traiettoria ottenuta dal filtro.**



**Figura 52 Anomalia nel tracciamento della traiettoria del filtr di Kalman.**

In *Figura 53* invece si osserva come il filtro di Kalman utilizza l'informazione della posizione del GPS per cercare di chiudere la traiettoria, avvicinandola al punto di partenza, cioè allo *Spigolo1*.



**Figura 53 Particolare della chiusura della traiettoria verso il punto di partenza del rilievo (Spigolo) da parte del filtro di Kalman.**

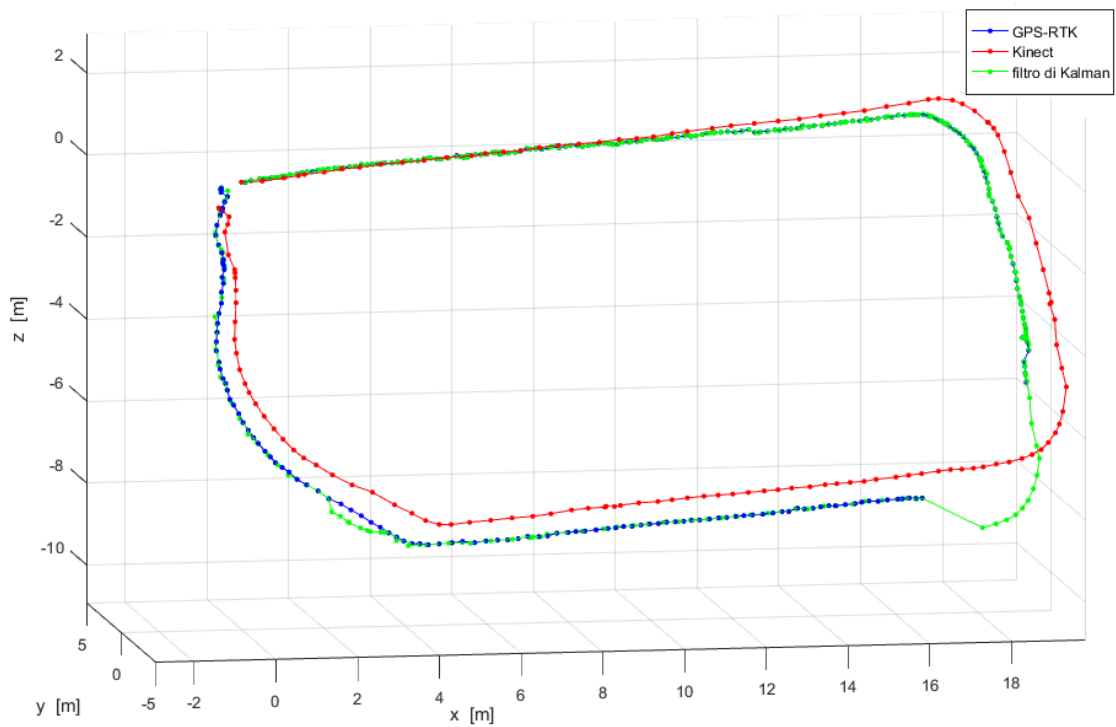
In *Tabella 5* sono riportate le medie delle deviazioni nelle componenti  $x$ ,  $y$  e  $z$  (riferite al sistema di riferimento della traiettoria del Kinect) della traiettoria ottenuta con il filtro di Kalman del *test1* rispetto alla traiettoria di riferimento della stazione totale. Si osservi come i valori siano sostanzialmente uguali a quelli del confronto tra traiettoria del GPS e traiettoria della stazione totale, risultato in linea con il fatto che nel *test1* la traiettoria ottenuta dal filtro di Kalman coincide sostanzialmente con quella del GPS.

	<b>Componente <math>x</math></b>	<b>Componente <math>y</math></b>	<b>Componente <math>z</math></b>
<b>Media della deviazione [m]</b>	0.047	0.022	0.023

**Tabella 7 Media delle deviazioni nelle componenti  $x$ ,  $y$  e  $z$  (rispetto al sistema di riferimento della traiettoria da Kinect) della traiettoria ottenuta dal filtro di Kalman rispetto alla traiettoria di riferimento della stazione totale nel *test1*.**

## 6.5 TEST2

Il secondo test è stato realizzato fingendo l'oscuramento dei satelliti GPS nel tratto della seconda curva della traiettoria. In questo tratto è stata fatta risultare nulla l'informazione di posizionamento del GPS, quindi il filtro di Kalman per tracciare la curva utilizza solo la componente dinamica proveniente dalla sequenza di immagini RGBD del Kinect. In *Figura 54* si osservi il tratto di curva in cui è stato annullata la traiettoria dal GPS.



*Figura 54 Risultato del filtro di Kalman con traiettoria GPS della seconda curva assente.*

Come constatato per l'uscita della terza curva nel TEST1 anche in questo caso si osserva sulla traiettoria calcolata dal filtro di Kalman l'effetto della deriva della

componente dinamica del filtro, in questo caso più evidente a causa della mancanza prolungata dell'osservazione di posizionamento del GPS. Inoltre si ricordi che il filtro descritto poggia sull'ipotesi di moto uniforme rettilineo. L'ipotesi di moto rettilineo ha fatto sì che in prossimità dell'inizio della curva il risultato del filtro allargato la curva, tracciando una traiettoria più ampia del reale (vedi particolare rappresentato in *Figura 55*). In *Tabella 8* vi sono le medie delle deviazioni nelle componenti x,y e z (sempre nello stesso sistema di riferimento della traiettoria del Kinect) per tutto il Giro (*Media della deviazione Giro1*) ed in corrispondenza del tratto in cui manca la traiettoria GPS (*Media della deviazione Curva*).

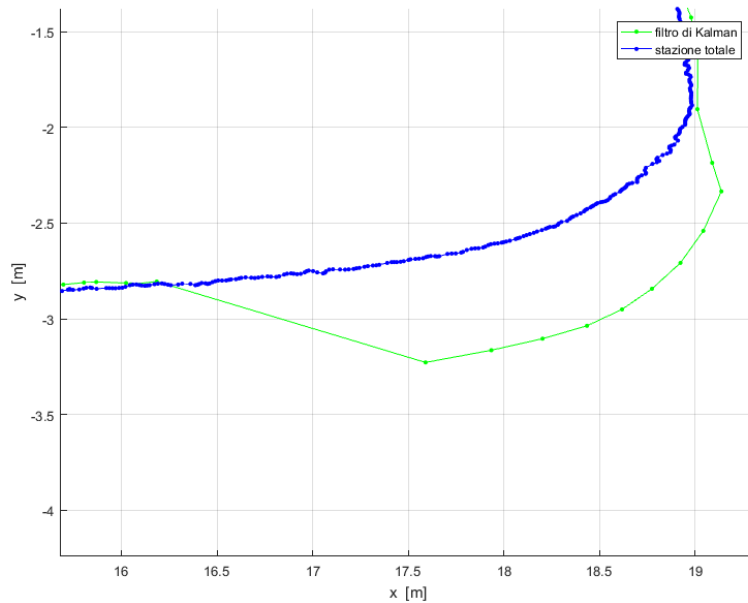
	<b>Componente x</b>	<b>Componente y</b>	<b>Componente z</b>
<b>Media della deviazione Giro1 [m]</b>	0.067	0.026	0.032
<b>Media della deviazione Curva [m]</b>	0.667	0.167	0.277

*Tabella 8* Medie delle deviazioni tra traiettoria del filtro di Kalman del test2 e la traiettoria di riferimento della stazione totale nel caso dell'intero Giro e nel caso della sola curva in cui è assente la traiettoria del GPS.

Dai dati presentati in *Tabella 8* risulta evidente l'aumento della deviazione proprio in prossimità della curva in cui manca la traiettoria GPS e soprattutto per la componente x, cioè lungo la direzioni dei rettilinei, in cui lo spostamento è maggiore. In questo tratto ha un forte peso nella determinazione di un simile valore di media della deviazione in componente x il fatto che il rettilino in entrata della curva sia in direzione perpendicolare a x. Infatti nell'entrata della curva finché la componente x di velocità è piccola il filtro tende a trascurarla, considerandola un errore dell'osservazione (si ricordi l'ipotesi di moto uniforme rettilineo), quindi prosegue nella stessa direzione y . Quando



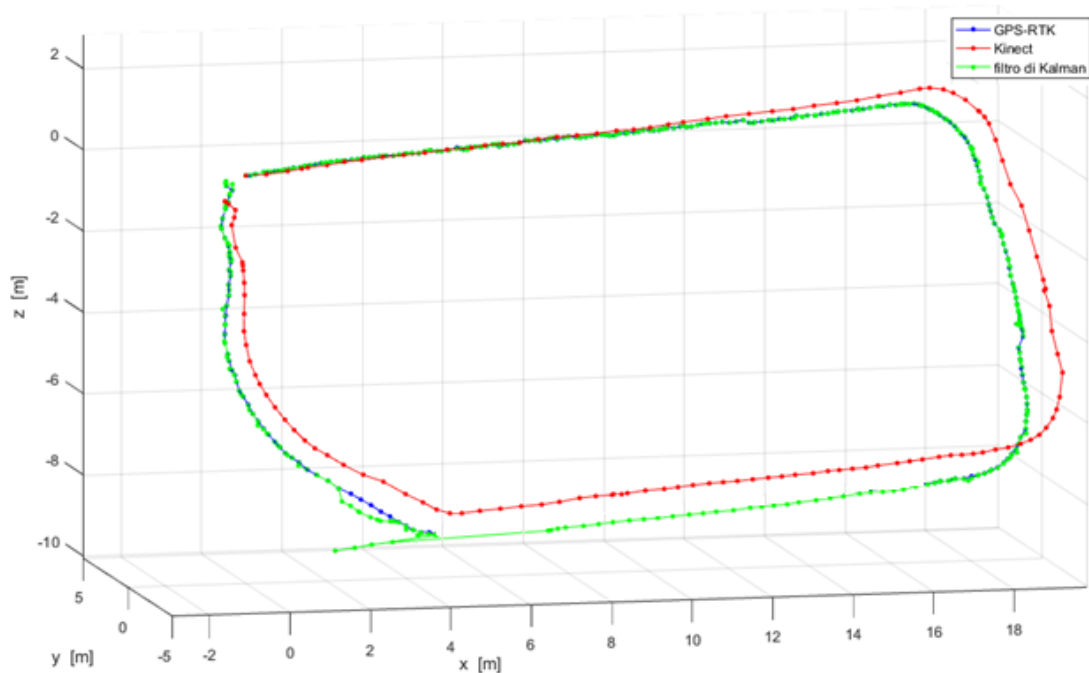
la componente x di velocità diventa evidente allora il filtro comincia ad applicare la curva, ma ormai la traiettoria è in ritardo rispetto a quella originale e quindi la deviazione risulta notevole.



**Figura 55** Particolare della deviazione nella curva in cui manca la componente di posizionamento nel filtro di Kalman.

## 6.6 TEST3

Nel terzo ed ultimo test si è simulata l'assenza di satelliti GPS per tutto il tratto del secondo rettilineo. Come per il TEST2 quindi è stato eseguito il calcolo del filtro di Kalman utilizzando per questo tratto solo la componente dinamica proveniente dalle osservazioni di velocità di Kinect. In Figura 56 si osservi il risultato ottenuto.



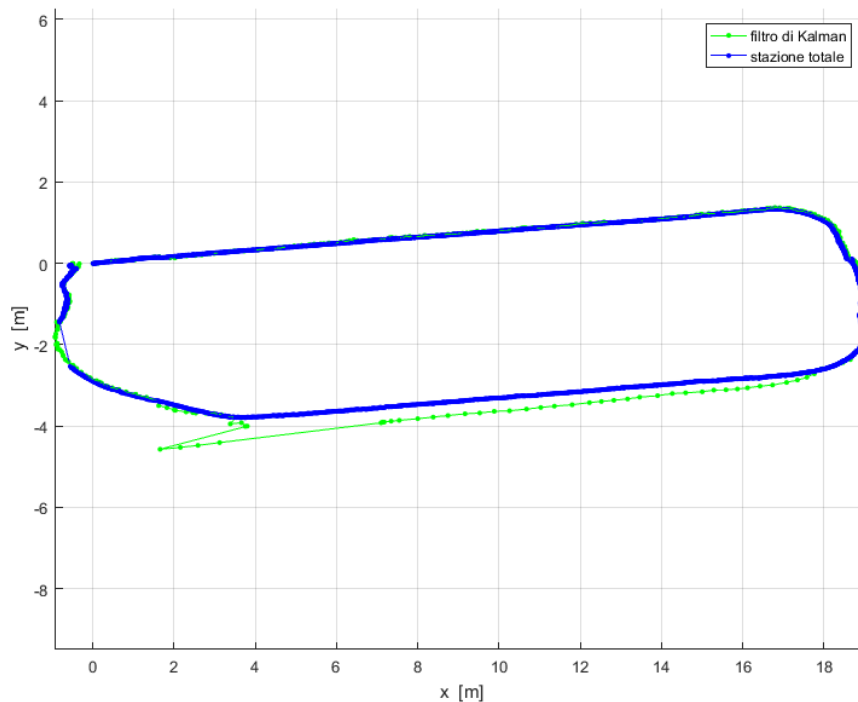
**Figura 56 Traiettoria calcolata da filtro di Kalman per il test3.**

Si osservi in *Figura 57* la traiettoria ottenuta dal filtro di Kalman messa a confronto con quella della stazione totale nel sistema di riferimento della traiettoria del Kinect. Si nota come il filtro di Kalman unito alla componente di osservazione dinamica interpreti bene gran parte del rettilineo sebbene ci sia una tendenza ad aprire la traiettoria, causata forse dalla mancata chiusura completa della seconda curva.

Intorno alla fine del rettilineo un aumento improvviso della componente dinamica del filtro determina un'improvvisa deriva della traiettoria del filtro di Kalman, il quale traccia un rettilineo più lungo del reale di quasi un metro e mezzo. L'aumento della componente dinamica potrebbe essere dovuto ad un errore nell'assegnazione dei tempi da cui viene calcolata la componente di velocità. Ne consegue una media della deviazione nella componente x per il secondo rettilineo molto maggiore rispetto a quelle

delle altre componenti (vedi *Tabella 9*). Le componenti  $y$  e  $z$  della media della deviazione del solo tratto privo di traiettoria del GPS invece sono minori rispetto a quelle ottenute nel TEST2 nel tratto di curva. La spiegazione sta ancora nel fatto che il filtro di Kalman interpreta meglio i tratti rettilinei.

Con la fine del rettilineo ed il ritorno della posizione GPS la traiettoria del filtro di Kalman torna a coincidere molto bene con la traiettoria descritta dal GPS



**Figura 57** Traiettoria calcolata dal filtro di Kalman per il test3 a confronto con la traiettoria di riferimento della stazione totale.

	<b>Componente <math>x</math></b>	<b>Componente <math>y</math></b>	<b>Componente <math>z</math></b>
<b>Media della deviazione Giro1 [m]</b>	0.087	0.033	0.031

<b>Media deviazione rettilineo [m]</b>	0.510	0.136	0.103
--	-------	-------	-------

***Tabella 9 Medie della deviazione nelle componenti x, y e z della traiettoria determinata dal filtro di Kalman nel test3 rispetto alla traiettoria di riferimento della***

## 7 CONCLUSIONI

Se si disponesse di un metodo lowcost affidabile e di precisione sufficiente in grado di risolvere il problema del posizionamento cinematico in ambiente urbano le possibili applicazioni sarebbero numerosissime. La fotogrammetria con fine navigazionale ha subito negli ultimi anni importanti sviluppi che ne hanno aumentato notevolmente i livelli di accuratezza del risultato finale. Sono importanti anche gli sviluppi nel processamento automatico delle immagini, aspetto che rende molto più utilizzabile il metodo fotogrammetrico. La fotogrammetria inoltre presenta il vantaggio di raggiungere livelli di precisione decisamente alti utilizzando camere dal costo molto più basso rispetto, ad esempio, agli accelerometri del metodo INS che raggiungono lo stesso livello di accuratezza.

Il lavoro fin qui descritto si è sviluppato nel tentativo di implementare la soluzione del problema della navigazione utilizzando un metodo GNSS integrato ad un metodo fotogrammetrico. Questi due tipi di metodo forniscono soluzioni che hanno caratteristiche molto diverse tra loro, e che in un certo senso si completano. Il metodo GNSS è indiscutibilmente il metodo principale per il posizionamento statico e dinamico. Da esso è possibile ottenere la posizione in sistema di riferimento geocentrico. Il suo punto debole consiste nel dipendere dalla disponibilità di satelliti della costellazione del sistema. In ambiente urbano questo aspetto non è trascurabile data la presenza di numerosi ostacoli che oscurano la visibilità dei satelliti (canyon urbani, palazzi, alberi ecc). Il metodo fotogrammetrico proposto permette di determinare la traiettoria dei centri di presa della camera nel suo assetto di presa nell'istante di

acquisizione delle immagini. La soluzione al posizionamento cinematico, come è spiegato nel paragrafo 4.4, viene determinata allacciando le trasformazioni rigide tra coppie di immagini consecutive. L'osservazione del metodo fotogrammetrico proposto quindi è lo spostamento (o velocità dal momento che è facile conoscere l'istante di acquisizione di ogni immagine). L'allacciamento di rototraslazioni consecutive fa sì che il posizionamento cinematico determinato dal metodo fotogrammetrico sia soggetto a problemi di deriva della traiettoria.

L'integrazione delle soluzioni ottenute dai due metodi ha lo scopo di compensare le loro reciproche debolezze. L'integrazione è stata svolta tramite il filtro di Kalman. Lo scopo del progetto quindi è stato quello di testare il comportamento della soluzione integrata tramite filtro di Kalman in diversi scenari di disponibilità della traiettoria del GPS.

Il metodo fotogrammetrico proposto è stato implementato con l'utilizzo di immagini acquisite dal sensore Microsoft Kinect per XboxOne. Questo strumento è costituito da una camera passiva per l'acquisizione di immagini RGB e di una camera attiva per le immagini Depth, acquisite sfruttando la tecnologia ToF (paragrafo 3.4). Lo strumento risulta essere piuttosto economico (poche centinaia di euro) e grazie alla tecnologia ToF è utilizzabile anche in ambienti outdoor.

Il metodo fotogrammetrico proposto per il posizionamento cinematico in ambiente urbano è stato messo in pratica per il rilievo svoltosi in Piazza Leonardo Da Vinci (paragrafo 5.2). Il blocco di immagini RGB e Depth e i dati acquisiti dalle antenne GPS sono stati elaborati rispettivamente con il software Matlab<sup>®</sup> e goGPS.

Per quanto riguarda le immagini da Kinect con una fase di selezione sono state scartate le immagini in cui il carrellino era in sosta per il posizionamento statico degli spigoli del *quadrato di calibrazione* e del punto iniziale e finale del tragitto attorno alla fontana denominato *Giro1* (cioè ancora lo *spigolo1* del *quadrato di calibrazione*). Quelle significative sono state ulteriormente sottocampionate. Le immagini RGB e Depth rimanenti sono state processate per eliminarne le distorsioni ottiche e nel caso dell'immagine Depth per migliorare la qualità dell'informazione contenuta (paragrafo ricampionamento delle immagini Depth 5.3.3). Quindi tramite le equazioni di collinearità di presa ogni nuvola di punti dell'immagine Depth è stata proiettata sulla rispettiva immagine RGB creando l'immagine chiamata RGBD (paragrafo creazione immagine RGBD 5.3.5). La rototraslazione tra il centro di presa di un'immagine RGBD e la successiva è stata determinata per tutta la sequenza di immagini RGBD. L'algoritmo di utilizzato è *align2RGBD* della Princeton Vision che comprende un algoritmo di matching automatico (SIFT, paragrafo 5.3.6.1), un algoritmo di ICP (paragrafo 5.3.6.2) e l'algoritmo RANSAC (paragrafo 5.3.6.3). La qualità della traiettoria determinata dalle immagini RGBD e dell'allineamento dei centri di presa delle immagini acquisite dagli spigoli del *quadrato di calibrazione* è fortemente legata alla precisione della soluzione determinata dall'algoritmo automatico *align2RGBD*.

In *Figura 48* è raffigurata la traiettoria dei centri di presa del Kinect e quella ottenuta dall'antenna GPS a doppia frequenza lungo il tragitto di *Giro1* mentre in *Figura 45* è raffigurato l'allineamento dei centri di presa degli spigoli del *quadrato di calibrazione* per il sensore Kinect e i punti posizionamento statico per l'antenna GPS. In (Tabella

deviazione traiettoria Giro1 Kinect e GPS) è riportata la media della deviazione nelle componenti  $x$ ,  $y$  e  $z$  tra traiettoria ottenuta dal Kinect e quella ottenuta dall'antenna GPS, mentre in (tabella prima deviazione per quadrato di calibrazione) vengono riportate le deviazioni dei punti del *quadrato di calibrazione* ottenuti da kinect e dall'antenna GPS. Si faccia presente che gli assi  $x$ ,  $y$  e  $z$  fanno riferimento al sistema di riferimento che ha origine nel centro di presa della prima immagine RGB del blocco fotogrammetrico acquisito per il rilievo di *Giro1*, cioè nello *Spigolo1*. Con asse  $x$  rivolto verso la larghezza dell'immagine e l'asse  $z$  come l'altezza dell'immagine e l'asse  $y$  per terna destrorsa.

Dalla *Tabella 5* si osserva che gli spigoli del quadrato determinato dall'allineamento delle immagini RGBD, acquisite dagli stessi, non coincidono perfettamente con quelli ottenuti dall'antenna GPS. Sebbene la deviazione per ogni spigolo risultino interne alla precisione attesa dal metodo fotogrammetrico proposto è pur sempre sintomo del fatto che l'allineamento delle immagini RGBD svolto da *align2RGBD* è soggetto ad imprecisione, la quale può determinare una notevole divergenza delle traiettorie per un rilievo di 40 metri di lunghezza totale. Questo tipo di imprecisione si ripercuote sulla determinazione della rototraslazione che porta dal sistema di riferimento GPS Locale (rispetto a cui sono definite inizialmente le traiettoria del GPS e della stazione totale) a quello della prima immagine RGBD del blocco del *Giro1*.

Se si osservano le due traiettorie raffigurate in *Figura 48* si nota che lungo il primo rettilineo le due traiettorie sono quasi coincidenti per due terzi della sua lunghezza, mentre con l'inizio della prima curva cominciano separarsi. Questo significa che prima



di tutto la rototraslazione determinata dagli spigoli del *quadrato di calibrazione* permette di allineare bene il primo tratto di rettilineo. Ne consegue che evidentemente l'allineamento di *align2RGBD* delle immagini RGBD acquisite dallo *Spigolo1* e *Spigolo2* *Figura 45* è di buona precisione come lo è anche per tutte le immagini del rettilineo. Quindi per esclusione è chiaro che la precisione dell'algoritmo *align2RGBD* viene meno soprattutto per l'allineamento di coppie di immagini in cui lo spostamento non sia lungo una direzione parallela al piano immagine, quindi appunto nelle curve del *Giro1* ma anche nei *lati2* e *lato4* del quadrato di calibrazione, in cui il carrellino è stato sollevato e spostato trasversalmente al piano immagine.

In sviluppi futuri del metodo di posizionamento cinematico proposto potrebbe essere interessante, ed utile, determinare l'effettiva imprecisione dell'algoritmo *align2RGBD* nel determinare l'allineamento per immagini che non poggino sullo stesso piano, quindi ad esempio per i tratti curvilinei delle traiettorie. Per far questo sarebbe utile disporre di una trasformazione, che porti la traiettoria di riferimento in un sistema di riferimento diverso rispetto a quella ottenuta dal Kinect nel sistema di riferimento rispetto a cui è definita la traiettoria ottenuta dal Kinect, indipendente dall'allineamento delle immagini RGBD svolto da *align2RGB*. Si propone quindi di determinare la posizione dei centri di presa delle immagini acquisite dagli spigoli del *quadrato di calibrazione* utilizzando il metodo fotogrammetrico classico, cioè risolvendo il problema di triangolazione aerea. Questo è realizzabile utilizzando le immagini RGB già acquisite con Kinect e i GCP nel sistema di riferimento GPS acquisiti dai target disposti sulla scultura della fontana. Per quanto riguarda le posizioni degli spigoli nel sistema di riferimento GPS possono essere

utilizzate quelle già ottenuta con l'antenna GPS a doppia frequenza. Applicando alla traiettoria di riferimento la rototraslazione indipendente dai risultati di *align2RGBD* si potrebbe valutare la vera natura dell'errore proveniente dall'allineamento calcolata da *align2RGBD* lungo la traiettoria di *Giro1* svolta con il Kinect.

Per testare il comportamento della soluzione ottenibile dall'integrazione della soluzione della traiettoria ottenuta da Kinect e traiettoria ottenuta dall'antenna GPS tramite il filtro di Kalman si è scelto di utilizzare in un primo caso tutti i punti della traiettoria GPS. In altri due casi invece si è simulata l'interruzione del segnale dei satelliti semplicemente eliminando parte della traiettoria. In particolare nel secondo test è stata eliminata la parte di traiettoria della seconda curva e nel terzo test quella del secondo rettilineo. I risultati, visibili nelle tre Figure *Figura 49* *Figura 54* e *Figura 56* sono stati analizzati a alla luce dell'ipotesi di moto rettilineo uniforme in cui è stato implementato il filtro di Kalman.

Nel primo test si osserva come grazie alla presenza di numerose soluzioni di posizione fornita dalla traiettoria de GPS (il triplo circa rispetto a quelle della traiettoria Kinect) la traiettoria del filtro risulti per lo più coincidente a quella del GPS. Questo è vero ad esclusione dell'entrata dalla terza curva (*Figura 50*) in cui l'elemento dinamico del filtro (cioè dell'osservazione di velocità proveniente dalle coppie di immagini RGBD) tende a ritardarne l'inizio. Questo si verifica sia perchè la traiettoria del Kinect è leggermente in deriva rispetto a quella del GPS (in questo caso in ritardo nell'affrontare la curva) sia perchè il filtro nei primi punti di entrata tende a mal interpretare la curva, mantenendo la traiettoria nel moto rettilineo.

Nel secondo test nella seconda curva succede in maniera molto più evidente quello già visto nell'entrata della terza curva del primo test. Il filtro tende a mal interpretare il cambio di direzione della componente della velocità dell'elemento dinamico e percorre per un paio di punti ancora un moto rettilineo, portando la curva fuori dalla direzione originaria. Questo comportamento è accentuato dal fatto che nella seconda curva non c'è l'elemento di posizionamento di riferimento al filtro.

Nel terzo test invece si osserva come il filtro ben interpreti l'intero rettilineo. Sebbene la traiettoria tracciata dal filtro tenda ad aprirsi leggermente verso sinistra, le deviazioni trasversali rispetto alla direzione di percorrenza del tratto sono piuttosto basse, soprattutto se considerate alla lunghezza di traiettoria tracciata. Se si mettono a confronto le medie delle deviazioni (della componente trasversale al moto) della tabella *Tabella 8* e della tabella **Tabella 9** si osserva che per il tratto di curva del secondo test, che interessa pochi metri di tragitto, la deviazione trasversale al moto di percorrenza (per quanto riguarda la seconda curva intendiamo per trasversale la direzione  $x$ , perpendicolare al tratto di entrata della curva che ha direzione  $y$ ) è maggiore rispetto a quella del terzo test, in cui invece i metri percorsi sono circa venti (nel caso del test3 la direzione trasversale al moto è quella lungo  $y$ ). Per le direzioni si faccia riferimento al sistema di riferimento delle figure.

Questo tipo di risultato deriva dal fatto che in prima analisi nel test3 il filtro di Kalman traccia la traiettoria utilizzando la componente dinamica di un tratto rettilineo. Come già detto in precedenza la componente dinamica deriva dallo spostamento tra coppie di immagini RGBD determinato dall'algoritmo *align2RGBD*. In conclusioni già tratte si è

constatato che l'algoritmo *align2RGBD* ottiene soluzioni di precisione migliore per tratti di spostamento parallelo al piano immagine. Quindi il terzo test ha ottenuto risultati migliori rispetto al secondo grazie sicuramente alla qualità migliore della sua componente dinamica. A questo si aggiunge il fatto, già osservato nel test1, che il filtro mal interpreta le entrate nelle curve per l'ipotesi di moto rettilineo uniforme.

Il filtro di Kalman è stato uno strumento molto utile per integrare all'osservazione GPS quella del Kinect nei tratti in cui il satellite risulta oscurato. Risulta evidente l'effetto di deriva della soluzione ottenuta dalla sola componente dinamica soprattutto nei tratti di curva. Come già anticipato nel paragrafo del test 6.3 la matrice di covarianza dell'osservazione della componente dinamica non ha un preciso significato metrico, ma è una bontà indicativa con cui è stata stimata da *align2RGBD* la rototraslazione tra una coppia di immagini RGBD fornita da *align2RGBD* stesso. Per sviluppi futuri è sicuramente necessario approfondire l'aspetto della precisione dell'osservazione proveniente da Kinect.

Risulta di interesse anche capire il comportamento dell'integrazione delle traiettorie ottenute dal Kinect e dal GPS con filtro di Kalman al variare delle matrici di covarianza della componente dinamica (spostamenti tra coppie di immagini RGBD) e della componente di posizionamento (fornite dal GPS).

Lo scopo della tesi era quello di ricercare una soluzione al posizionamento cinematico utilizzando dati acquisiti con soli sensori lowcost. Nell'elaborazione dei dati tuttavia è stato necessario utilizzare il dato proveniente dall'antenna GPS a doppia frequenza, perchè il sensore U-blox, causa malfunzionamento durante il rilievo, ha fornito

soluzioni dalla precisione inaccettabile. Si propone quindi di ripetere lo stesso rilievo con attrezzatura funzionante.

Infine partendo dai problemi tecnici verificatisi vengono dati alcuni consigli pratici su come svolgere il rilievo.

Prima di tutto si consiglia l'utilizzo di un carrellino che renda possibile il movimento trasversale rispetto al posizionamento delle camere Kinect. Questo accorgimento potrebbe migliorare l'acquisizione di immagini lungo i tratti *lato2* e *lato4* del *quadrato di calibrazione*, e quindi permetterebbe la determinazione della posizione degli spigoli con tecnica fotogrammetrica con una precisione migliore.

Si consiglia inoltre di organizzare il rilievo in orario in cui la zona risulti tranquilla e sgombra da soggetti dal comportamento imprevedibile. Questo accorgimento serve ad evitare l'attraversamento continuo di soggetti nella scena di acquisizione delle immagini rendendo più veloce il processo di selezione delle immagini utili ed evitando anomalie simili a quelle presentate in figura anomalia *Figura 52*.

## • Bibliografia:

Abdel Aziz, Y.I., Karara, H.M., (1971), *Direct Linear Transformation from Comparator Coordinates into Object Space Coordinates, Close-range Photogrammetry, American Society of Photogrammetry, Falls Church (Virginia), vol. 443, pp. 1-18.*

Al-Hamad, A., El-Sheimy, N., 2014. *Smartphones Based Mobile Mapping Systems. In: The International Archives of Photogrammetry and Remote Sensing and Spatial Information Sciences. Vol. XL-5, Proceedings on the ISPRS Technical Commission V Symposium, Riva del Garda, Italy.*

Barzaghi, R., Carrion, D., Cazzaniga, N.E., Forlani, G., 2009. *Vehicle Positioning in Urban Areas Using Photogrammetry and Digital Maps, Proceedings of the ENC-GNSS09, Napoli.*

Bayoud F.A., Skaloud, J., Merminod, B., (2004), *Photogrammetry-derived navigation parameters for INS Kalman Filter Updates, Proceedings of the XXth ISPRS Congress Istanbul, Turkey, ISPRS Archives, Vol. XXXV Part B5, pp.262-267.*

Böniger, U., Troncke, J. (2010), *On the Potential of Kinematic GPR Surveying Using a Self-Tracking Total Station: Evaluating System Crosstalk and Latency, IEEE Transaction on Geoscience and Remote Sensing, vol.28, n°10, pp.3792-3798. Doi:10.1109/TGRS.2010.2048332*

Brown, D., (1966), *Decentering Distortion of Lenses, Photo Eng, vol. XXXII, n°3, pp. 444-462.*

Brown, D., (1987), *Close-range Camera Calibration, Photo Eng, vol. XXXVII, n° 8, pp. 855-866.*

Cazzaniga, N.E., Pagliari, D., Pinto, P., 2012. *Photogrammetry for Mapping Underground Utility Lines with Ground Penetrating Radar in Urban Areas. In: International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences. Volume XXXIX-B1, Melbourne, Australia*

Cazzaniga, N.E., Pagliari, D., Pinto, L., 2013. *Fotogrammetria e GPS per posizionamento di veicoli dedicati al rilievo dei sottoservizi. Bollettino SIFET, n°1.*

Chiabrando, F., Lingua, A., Piras, M., (2013), *Direct Photogrammetry using UAV: test and first results*, *International Archives of Photogrammetry and Remote Sensing and Spatial Information Sciences*, Vol. XL-1/W2/2013, UAV-g 2013, Rockstock, Germny, 4-6 September.

Da Silva, J.F. de Oliviera Carmago, P., Gallis R. B. A., 2003. *Development of a low-cost mobile mapping system: a South American experience*. *The Photogrammetric Record*. Vol. 18 (101), pp. 5-26.

De Agostino M. (2010), *I sensori inerziali di basso costo per la navigazione geodetica*, *Tesi di Dottorato, Dottorato di Ricerca in Ambiente e Territorio (XXII ciclo)*, Politecnico di Torino.

Endres, F., Hess, J., Engelhard, N., Sturn, J., Cremer, D., Burgard, W., 2012. *An Evaluation of RGB-D SLAM System*. *2012 IEEE International Conference on Robotics and Automation*. River Centre, Saint Paul, Minnesota.

Fankhauser, M. Bloesch, C. Gehring, M. Hutter, and R. Siegwart, "Robot-Centric Elevation Mapping with Uncertainty Estimates," in *International Conference on Climbing and Walking Robots (CLAWAR)*, Poznan, Poland, 2014.

Fankhauser\_, Michael Bloesch\_, Diego Rodriguez\_y,Ralf Kaestner\_, Marco Hutter\_, Roland "Kinect v2 for Mobile Robot Navigation: Evaluation and Modeling" Siegwart \_Autonomous Systems Lab, ETH Zurich, Switzerland yRobotics and Embedded Systems, Technische Universit"at M"unchen, Germany 2015.

Fischer, M.A., Bolles, R. C., (1981), *Random Samples Consensus: A paradigm for model fitting with application to image analysis and application cartography*, *Communication of the ACM*, pp 381-395.

Fraser, C. (2005), *Network Orientation Model for Image Based 3D Measurement*, *International Archives of Photogrammetry, Remote Sensing and Spatial Information Science*, 36(5/W-17), 9 pages

Hassan, T., Ellum, C., El-Sheimy, N., 2006. *Bridging land-based mobile mapping using photogrammetric adjustments*. *ISPRS Commission I Symposium. From Sensors to Imagery*.

Hofmann-Wellenhof , Lichtenegger e Wasle in *GNSS Global Navigation Sysem GPS, GLONASS, Galileo and more di Springer Wien NewYork , ISBN 978-3-211-73012-6 ,2008).*

Kalman, "A New Approach to Linear Filtering and Prediction Problems," *Trans. ASME Journal of Basic Engineering*, vol. 82, no. 1, pp. 3545, Mar. 1960.

Khoshelham, K., (2011), *Accuracy analysis of Kinect depth data*, *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Science*, Vol. 38(5/W12).

Krauss, K., (2008), *Photogrammetry: Geometry from Images and Laser Scans*, Second Edition, Walter de Gruyter, 459 pages.

Lachat, H. Macher, M.-A. Mittet, T. Landes, and P. Grussenmeyer, "First Experiences With Kinect V2 Sensor for Close Range 3D Modelling," in *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences (ISPRS)*, 2015

Lachat, H. Macher, M.-A. Mittet, T. Landes, and P. Grussenmeyer, "First Experiences With Kinect V2 Sensor for Close Range 3D Modelling," in *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences (ISPRS)*, 2015.

Leandro Cruz, Djalma Lucio, Luiz Velho IMPA - VISGRAF Lab Rio de Janeiro, Brazil Kinect and GBD Images: Challenges and Applications, 2012.

Lowe. Distinctive image features from scale-invariant key points. *IJCV*, 2004. (SIFT) [fz matlab svilupata da VLFeat: vl\_sift <http://www.vlfeat.org/overview/sift.html>. esiste anche la versione Matlab di Lowe scaricabile al sito <http://www.cs.ubc.ca/~lowe/keypoints/>]

MacGougan G. (2003) *High Sensitivity GPS Performance Analysis in Degraded Signal Environments*, M.Sc. thesis, UCGE Report No. 20176,, Department of Geomatics Engineering, University of Calgary, Canada.

Medioni, G. Chen, Y., (1991), *Object Modelling by registration of Multiple Range Images*, *Proceedings of 1991 IEEE International Conference on Robotics and Automation*, pp. 2724-2729, vol. 3.

Nister. Preemptive RANSAC for live structure and motion estimation. In *Proc. of the IEEE Intl. Conf. on Computer Vision (ICCV)*, 2003.

Oliver, A., Kong, S., Wünsche, B., MacDonald, B., 2012. *Using the Kinect as a Navigation Sensor for Mobile Robotics*. *Proceedings of the 27th Conference on Image and Vision Computing*, pp. 505-514.



*Pagliari Diana outdoor and indoor navigation based on photogrammetry with low-cost system, Politecnico di Milano- Department of Civil and Environmental Engineering Doctoral Program in Environmental and Infrastructure Engineering, 2015.*

*Pinto A. M., P. Costa, A. P. Moreira, L. F. Rocha, G. Veiga, and E. Moreira, "Evaluation of Depth Sensors for Robotic Applications," in IEEE International Conference on Autonomous Robot Systems and Competitions, 2015.*

*Realini E., Reguzzoni M., 2013, goGPS: open source software for enhancing the accuracy of low-cost receivers by single-frequency relative kinematic positioning, Measurement Science and Technology 24 (11), 115010, October 2013 (DOI 10.1088/0957-0233/24/11/115010; ISSN 1361-6501).*

*Robert M. Solow (Solow, 1956) A Contribution to the Theory of Economic Growth The Quarterly Journal of Economics, Vol. 70, No. 1. (Feb., 1956), pp. 65-94*

*Roncella, R., Remondino, F., Forlani, G., 2005. Photogrammetric Bridging of GPS outages in mobile mapping. Electronic Imaging. pp.308-319.*

*Sell, J., O'Connor, P, (2014), The XboxOne System in a Chip and Kinect Sensor, Micro IEEE, vol. 34, issue n°2, pp. 44-53, March-April, 2014.*

*Treccani online enciclopedia (<http://www.treccani.it/>), 2016*

*Troni G. e Whitcomb L. L., Experimental evaluation of a MEMS inertial measurements unit for Dopplernavigation of underwater vehicles Oceans, 2012, Pages: 1-7, DOI: 10.1109/OCEANS.2012.6405003.*

*Wasmeier (2006) Geodetic Transformations Toolbox della Mathworks funzioni matlab Helmert e d3trafo (<http://www.mathworks.com/matlabcentral/fileexchange/9696-geodetic-transformations-toolbox/content/GeodeticToolbox/d3trafo.m>)*

*Xiao (Princeton University), A. Owens (MIT) and A. Torralba (MIT) sun3D\_a database of big spaces reconstructed using SfM and Object Labels [2013 IEEE International Conference on Computer Vision].*

*Xingxing Li e al. 2015 Accuracy and reliability of multi-GNSS real-time precise positioning: GPS, GLONASS, BeiDou, and Galileo.*

*Zennaro, "Evaluation of Microsoft Kinect 360 and Microsoft Kinect One for robotics and computer vision applications," Master's thesis, University of Padova, Italy, 2014.*

*Zhao, H., Masaki. C., Shibasaki, R., Xiaowei S., Cui, J.; Zha, H., SLAM in a dynamic large outdoor environment using a laser scanner, ICRA 2008. IEEE International Conference on Robotics and Automation, 2008, pp.1455-1462, 19-23 May 2008, doi: 10.1109/ROBOT.2008.4543407.*

*@inproceedings{SUN3D, author = {Jianxiong Xiao and Andrew Owens and Antonio Torralba}, title = {{SUN3D}: A Database of Big Spaces Reconstructed using SfM and Object Labels}, booktitle = {Proceedings of IEEE International Conference on Computer Vision (ICCV)}, year = {2013}, }*