

UNIVERSITÀ DEGLI STUDI DI PARMA

Dottorato di Ricerca in Tecnologie dell'Informazione

XXV Ciclo

**Sviluppo di sistemi di visione stereoscopici
per la ricostruzione tridimensionale
in scenari industriali ed automotive**

Coordinatore:

Chiar.mo Prof. Marco Locatelli

Tutor:

Chiar.mo Prof. Alberto Broggi

Dottorando: *Michele Buzzoni*

Gennaio 2013

In ricordo di mia mamma

Sommario

Introduzione	1
1 Sistema di visione per la sicurezza sul lavoro	7
1.1 Introduzione	8
1.2 Stato dell'arte	9
1.3 Sistema stereoscopico	11
1.3.1 Configurazione Hardware	13
1.3.2 Algoritmo	13
1.4 Risultati	16
1.4.1 Tempi di elaborazione	19
1.4.2 Prestazioni	19
1.4.3 Conclusioni	20
2 Sistema di visione per l'apertura automatica di porte	23
2.1 Introduzione	23
2.2 Disparità e DSI su immagini stereo IPM	25
2.2.1 Inverse Perspective Mapping	26
2.2.2 Calcolo della disparità e ricostruzione 3D su immagini IPM	31
2.2.3 L'immagine di disparità DSI su immagini IPM	35
2.3 L'algoritmo	39
2.4 Risultati e conclusioni	43

3	Sistema di individuazione di ostacoli in cantieri stradali	47
3.1	Introduzione	47
3.2	IPM Cilindrica	48
3.2.1	IPM sferica	50
3.2.2	CIPM (Cylindric Inverse Perspective Mapping)	53
3.2.3	La generazione di immagini CIPM	58
3.2.4	Disparità su immagini CIPM	60
3.2.5	Ricostruzione 3D dalla disparità cilindrica	62
3.2.6	Immagine di disparità DSI su immagini CIPM	64
3.3	L'algoritmo	69
3.3.1	Immagine accumulatore	73
3.3.2	Individuazione ostacoli	75
3.4	Risultati e conclusioni	77
3.4.1	Tempi di elaborazione	78
3.4.2	Prestazioni	79
3.4.3	Conclusioni	82
4	Sistema di identificazione ostacoli per veicoli autonomi	85
4.1	Introduzione	85
4.1.1	Configurazione Hardware	86
4.1.2	La spedizione	86
4.2	Algoritmi	88
4.2.1	Generazione mappa di disparità	89
4.2.2	Filtri applicati all'immagine DSI	91
4.2.3	Individuazione ostacoli	92
4.3	Risultati	96
4.3.1	Prestazioni	97
4.3.2	Tempi di elaborazione	98
4.3.3	Conclusioni	98
	Conclusioni	101

A	Visione stereoscopica: Disparità	105
A.1	Il modello <i>pinhole camera</i>	106
A.2	La trasformazione prospettica	107
A.3	Ricostruzione delle coordinate mondo	112
A.4	La disparità	113
A.5	La ricerca dei punti omologhi	118
A.6	L'immagine di disparità	120
	Bibliografia	129
	Ringraziamenti	131

Elenco delle figure

1.1	Macchina CNC per la lavorazione del legno.	9
1.2	Zone di sicurezza di una macchina CNC e posizione dei sistemi stereo.	10
1.3	Sistema stereoscopico.	12
1.4	Schema a blocchi del sistema.	14
1.5	Output del sistema in caso di ostacolo nella zona centrale	17
1.6	Output del sistema in caso di ostacolo nella zona sinistra	18
1.7	Grafico di confronto fra sistemi di sicurezza	21
2.1	Immagine e aree controllate da un sistema con porte automatiche	24
2.2	Schema di generazione dell'IPM.	27
2.3	Immagine prospettica e immagine IPM di una strada.	30
2.4	Immagini acquisite dalla telecamera sinistra e destra.	31
2.5	Immagini IPM: sinistra e destra.	31
2.6	Disparità su immagini IPM.	33
2.7	Immagine IPM destra e immagine di disparità	33
2.8	Calcolo della disparità massima.	38
2.9	Diagramma dell'algoritmo proposto.	40
2.10	Immagini di background, input, differenza, e maschera per la telecamera sinistra e destra	41
2.11	Output del sistema: (a) Cluster su DSI, (b) Traiettoria del pedone, (c) Output, (d) Ricostruzione 3D.	42
2.12	Output dell'algoritmo in due diverse situazioni	43

3.1	Angolo α formato dalle due rette di un ostacolo sull'IPM	50
3.2	Coordinate polari.	51
3.3	Coordinate polari dell'immagine IPM.	52
3.4	IPM: sinistra e destra (a), IPM sferica: sinistra e destra (b).	54
3.5	IPM di un ostacolo verticale: sinistra e destra.	55
3.6	IPM sferica di un ostacolo verticale: sinistra (a) e destra (b).	56
3.7	Immagini IPM classica (a) e cilindrica (b).	57
3.8	Analisi degli angoli ϕ_{min} , ϕ_{max} delle due immagini CIPM.	58
3.9	Disparità cilindrica di ostacoli verticali	63
3.10	Immagini CIPM destra e DSI.	65
3.11	Range di disparità ed errore.	65
3.12	Schema a blocchi dell'algoritmo.	70
3.13	Mappatura manuale della griglia: immagine distorta (a), IPM (b).	72
3.14	Immagini IPM, DSI e accumulatore	74
3.15	Immagini accumulatore con mattonelle di dimensione 20×10	75
3.16	Posizione di un ostacolo sull'immagine accumulatore.	76
3.17	Ostacolo localizzato sull'immagine sorgente.	76
3.18	Esempi di output	81
3.19	Ulteriori esempi di output	83
4.1	Componenti hardware del sistema stereo.	87
4.2	Il percorso di VIAC	87
4.3	Architettura del sistema.	88
4.4	Filtri sull'immagine di disparità	93
4.5	Criterio di compatibilità usato per la generazione dei <i>cluster</i>	93
4.6	Partizione dell'immagine DSI	95
4.7	Etichettatura degli ostacoli in una strada in salita	95
4.8	Alcuni esempi di output in differenti scenari	99
A.1	Il modello <i>pinhole camera</i>	106
A.2	Il sistema di riferimento mondo e immagine.	107
A.3	La trasformazione prospettica con il modello <i>pinhole camera</i>	108

A.4	Il sistema di riferimento immagine (x,y,z) e mondo (X,Y,Z)	110
A.5	L'orientamento della telecamera	111
A.6	La triangolazione stereo.	112
A.7	Proprietà della disparità al variare delle distanze degli ostacoli . . .	114
A.8	Sistema di telecamere stereo.	115
A.9	<i>Baseline</i>	116
A.10	La proiezione di un punto del mondo reale sul piano delle immagini.	117
A.11	Il piano e la retta epipolare.	119
A.12	Immagine destra ed immagine di disparità.	120
A.13	L'immagine di disparità e gli ostacoli individuati su di essa.	120

Elenco delle tabelle

1.1	Fusione dei segnali di Output	15
1.2	Tempi di elaborazione del sistema	20
2.1	Prestazioni del sistema.	44
3.1	Caratteristiche tecniche dei calcolatori utilizzati.	78
3.2	Tempi di elaborazione	78
4.1	Prestazioni dell' algoritmo SGM	91
4.2	Prestazioni dei sistemi	97
4.3	Tempi di elaborazione	98

Introduzione

La visione stereoscopica occupa un vasto campo di ricerca della visione artificiale. Questa tecnica infatti permette la ricostruzione tridimensionale della scena utilizzando due immagini riprese da due punti di vista leggermente diversi, estraendo ulteriori informazioni spaziali sulla profondità della scena.

Questa tesi si pone l'obiettivo di sviluppare nuovi sistemi di visione stereoscopici per la ricostruzione tridimensionale dell'ambiente circostante tramite l'uso di algoritmi di disparità. Tali sistemi saranno collocati all'interno di progetti con finalità diverse e riguardanti due ambiti principali: quello industriale e quello *automotive*. La principale differenza fra questi due scenari è data dalla posizione del sistema di visione, in quanto nel campo industriale si utilizzeranno telecamere fisse mentre per quello *automotive* telecamere mobili, ovviamente solidali al veicolo. I due metodi, con telecamere fisse e mobili, permettono di avere due distinte visioni del mondo circostante.

Nel caso di un approccio con telecamera fissa, l'osservatore si trova a poter inquadrare una porzione del mondo che in molti casi potrebbe avere un cambiamento a volte anche repentino, ma nella maggior parte dei casi presenta un'elevata staticità dal punto di vista della scena. Un tipico sistema di visione artificiale con telecamera fissa prevede un'installazione statica, normalmente in una posizione conosciuta, con ampia conoscenza delle problematiche relative alla scena inquadrata. Nel caso di installazione esterna sono conosciute le possibilità di cambiamenti atmosferici, di illuminazione e di temperatura; mentre nel caso di installazioni interne sono noti i cambiamenti e la qualità dell'illuminazione della scena. In entrambi i casi si possono

prevedere con un certo livello di affidabilità probabili occlusioni dell'inquadratura della telecamera o dello specifico oggetto inquadrato.

Invece, nel caso dell'approccio con telecamera mobile si ha una visione più imprevedibile del mondo circostante. Facendo riferimento al settore *automotive*, la scena inquadrata è soggetta a repentini cambiamenti per quanto riguarda la maggior parte dei parametri: illuminazione, cambiamenti atmosferici, tipologia di ambiente circostante. Il mondo circostante è in costante cambiamento per effetto del movimento del veicolo, e date le grandezze in gioco dal punto di vista delle distanze e del tempo, il cambiamento può essere anche notevole.

I due tipi di scenari, nonostante questa notevole differenza, presentano d'altra parte forti punti in comune. Infatti, i diversi algoritmi di visione applicabili a questi contesti necessitano di un'esecuzione in tempo reale dell'elaborazione delle immagini acquisite. Le tecniche per ottenere la mappa di disparità della scena dovranno pertanto fornire elevate prestazioni senza tuttavia degradare i risultati.

All'interno di questa tesi saranno quindi proposti diversi tipi di algoritmi, sia locali che globali, che si adatteranno alle necessità del progetto a cui si vogliono applicare. Per ogni sistema, una volta calcolata la mappa di disparità, sarà presentato il relativo algoritmo di individuazione ostacoli, ottenuto principalmente tramite la ricostruzione tridimensionale della scena. Le varie tecniche che saranno applicate si possono sommariamente dividere in due filoni dipendenti dalla costruzione algoritmica dell'immagine di disparità e dal metodo di calibrazione del sistema. In particolare saranno sviluppati due differenti approcci per ottenere la mappa di disparità e una ricostruzione tridimensionale della scena.

Il primo si può considerare un metodo classico e ben presente in letteratura. Si basa sulla perfetta calibrazione del sistema stereo e sul calcolo esatto di tutti i parametri estrinseci ed intrinseci delle telecamere e delle ottiche. Questo tipo di operazioni può risultare critico e può a volte introdurre errori che influiscono sulla corretta ricostruzione 3D. Infatti i parametri precedentemente calcolati sono utilizzati per la rettificazione e la rimozione della distorsione delle immagini. La coppia di immagini acquisita dallo stesso sistema stereo, dopo aver subito questa trasformazione, può quindi essere confrontata per individuare i punti omologhi e l'estrazione delle infor-

mazioni che permettono la generazione dell'immagine di disparità. Infine le coordinate mondo di ciascun punto di questa immagine potranno essere ottenute applicando equazioni strettamente dipendenti dai parametri di calibrazione.

Il secondo metodo invece verrà introdotto sperimentalmente in questa tesi e rappresenta un'innovazione nell'ambito della comunità scientifica. La principale differenza rispetto al metodo sopra descritto risiede nell'installazione e nella calibrazione del sistema. L'installazione infatti prevede un posizionamento arbitrario fra le due telecamere avendo come vincolo solo il fatto che le telecamere inquadrino la stessa scena e siano alla stessa altezza dal suolo. Una semplice calibrazione del sistema mediante una griglia permette l'eliminazione della prospettiva e della distorsione generando due immagini IPM¹, raffiguranti la stessa scena come se la fosse vista dall'alto ma nella quale solo i punti appartenenti al piano del terreno risultano correttamente trasformati. Gli stessi algoritmi per la generazione dell'immagine di disparità vengono quindi applicati su queste immagini, ma a differenza di prima la ricostruzione tridimensionale della scena non dipenderà più dai parametri intrinseci e dall'orientazione delle telecamere ma solo dalla loro posizione. Un'ulteriore trasformazione prospettica sarà introdotta per migliorare la precisione della ricostruzione tridimensionale.

Entrambi i metodi saranno implementati ed adattati a sistemi sia nel campo industriale, quindi con telecamere fisse, sia in quello *automotive* con telecamere in movimento.

I sistemi che saranno presentati in questa tesi di dottorato affronteranno inoltre le maggiori problematiche nei campi industriale ed *automotive* che hanno una ricaduta umana molto importante dal punto di vista sociale, come ad esempio la sicurezza, la videosorveglianza e la guida autonoma/assistita.

La sicurezza sul lavoro all'interno del campo industriale rappresenta un grave problema, evidenziato dalla quantità di infortuni che accadono ogni anno. Nel 2011, secondo dati INAIL [1], si sono verificati oltre mezzo milione di incidenti, di cui lo 0.1% risultati mortali. Il pericolo a cui i lavoratori sono esposti aumenta notevolmente quando si lavora a diretto contatto con macchinari in movimento. Le cause che

¹Inverse Perspective Mapping

provocano incidenti possono essere legate a diversi fattori riconducibili alle tre entità uomo-macchina-ambiente. La visione artificiale applicata a questa problematica può fornire soluzioni utili per ridurre notevolmente il numero di incidenti sul lavoro segnalando un pericolo ad un operatore o addirittura fermando una macchina con un segnale di arresto. La presenza di telecamere viene anche vista dagli operatori come deterrente per svolgere operazioni fuori norma e pericolose, riducendo anche in questo modo i rischi di incidenti.

La videosorveglianza come settore di ricerca ha trovato negli ultimi anni una forte spinta dal punto di vista dell'innovazione. I sistemi di videosorveglianza stanno passando da un controllo dell'ambiente passivo ad un tipo di controllo attivo che prevede una segnalazione da parte di sistemi software di situazioni fortemente indicative di comportamenti anomali nella scena inquadrata. Esistono studi sull'utilità dell'utilizzo di telecamere di videosorveglianza per quanto riguarda la lotta o la prevenzione al crimine. In [52] viene infatti mostrato come l'utilizzo di telecamere risulti efficace come deterrente per prevenire i reati.

All'interno del campo *automotive* la visione artificiale si occupa sia dello studio di veicoli completamente autonomi sia nella creazione di dispositivi di supporto alla guida, in grado di sopperire alle disattenzioni del conducente o di avvisarlo di un'eventuale situazione di pericolo. Grazie all'evoluzione delle tecnologie a disposizione, nell'ultimo decennio diversi settori di ricerca hanno focalizzato la loro attenzione sullo sviluppo di sistemi ADAS², e numerose risorse sono state investite per fornire sistemi sempre più efficienti, caratterizzati da elevate performance e costi ridotti. La ricerca in questo settore è sostenuta sia dalle istituzioni, per diminuire con ogni mezzo la quantità di incidenti stradali, sia dalla case automobilistiche stesse che cercano di realizzare sistemi che possano rendere più piacevole l'esperienza di guida.

²Advanced Driver Assistance Systems.

Struttura della tesi

Questa tesi di dottorato è strutturata nel seguente modo: nei primi capitoli saranno presentati due sistemi stereoscopici in scenari industriali, mentre nei successivi capitoli verranno introdotti ulteriori due sistemi stereoscopici di identificazione ostacoli nel campo *automotive*. In particolare:

- Nel Capitolo 1 verrà presentato un sistema di visione per la sicurezza sul lavoro basato su due sistemi stereoscopici e una doppia ricostruzione tridimensionale della scena.
- Il Capitolo 2 si occuperà di un sistema di visione stereoscopico per l'apertura automatica di porte. In questo capitolo sarà introdotta un'analisi della disparità alternativa e sfruttante la trasformazione geometrica IPM.
- Nel Capitolo 3 sarà illustrato un sistema installato su mezzi pesanti per l'individuazione ostacoli in cantieri stradali, basato su di un'evoluzione dell'algoritmo descritto nel capitolo precedente, ottenuto applicando un'ulteriore trasformazione prospettica definita CIMP³.
- Il Capitolo 4 presenterà il sistema di identificazione ostacoli per veicoli autonomi utilizzato durante la spedizione intercontinentale VIAC⁴ condotta dal laboratorio Laboratorio di Visione Artificiale (VisLab). Inoltre verrà dato particolare rilievo all'implementazione algoritmica delle tecniche di disparità utilizzate anche dai sistemi precedentemente descritti.
- Nell'ultimo Capitolo saranno riassunte le conclusioni di questo lavoro di tesi di dottorato.

I sistemi presentati faranno ampio uso di tecniche di visione stereoscopica, la cui teoria, per completezza, sarà esposta nell'appendice A di questo lavoro.

³Cylindric Inverse Perspective Mapping.

⁴VisLab Intercontinental Autonomous Challenge.

Capitolo 1

Sistema di visione per la sicurezza sul lavoro

In questo capitolo viene presentata un'innovativa tecnologia di sicurezza basata su visione artificiale per il controllo tridimensionale degli spazi. In particolare il sistema è stato studiato e realizzato per essere applicato ad impianti industriali per la produzione di mobili che prevedono l'utilizzo di centri di lavoro CNC ¹ per la lavorazione del legno e altri materiali.

Il sistema sarà in grado di definire liberamente degli spazi protetti e di rilevare e comunicare l'accesso di oggetti in spazi predefiniti sottoposti ad allarme. In particolare dovrà fermare l'esecuzione della macchina nel caso fosse presente un ostacolo nel campo di azione della macchina stessa, considerando che normalmente la lavorazione non viene fermata neanche durante le fasi di carico/scarico dei pezzi di lavorazione.

Il grado di sicurezza apportato da questo tipo di sistema potrà permettere la rimozione di eventuali altri sistemi di sicurezza basati su di una sensoristica diversa. Un sistema di visione fornisce infatti vantaggi che questi tipi di sensori non possono garantire, ad esempio il funzionamento anche in caso di ambienti polverosi, il rilevamento di oggetti che non toccano il terreno, ed un basso impatto strutturale nella scena.

¹Controllo numerico computerizzato

La realizzazione fisica del sistema comporta l'installazione di due sistemi stereoscopici ai lati opposti dell'area da sorvegliare per generare una ridondanza dei risultati. La presenza di due sistemi visivi garantisce un grado di affidabilità assoluta gestendo i possibili guasti di un singolo sistema. Una particolare attenzione a tutti i possibili guasti del sistema e la scrittura del codice sotto determinate regole deve essere garantita per poter ottenere una certificazione ufficiale di sicurezza.

1.1 Introduzione

L'applicazione, che il sistema proposto intende fornire, è l'arresto immediato di una macchina utensile nel caso in cui un operatore entri in un perimetro di sicurezza definito a priori. La tipologia di macchine CNC presa in esame è formata da un banco di lavoro statico con lunghezza variabile dai 4 ai 9 metri, sul quale scorre una testa utensile mobile, come mostrato in figura 1.1. La dimensione dell'utensile mobile risulta essere molto più ingombrante del banco ed ha un volume di ingombro frontale molto accentuato in quanto sporge di decine di centimetri rispetto alla zona statica della macchina. Questo tipo di macchina, per la lavorazione legno, prevede inoltre la presenza di due zone di carico materiale intervallate da una zona di inibizione. Molto spesso infatti, la lavorazione del materiale è continua; quando un operatore carica il materiale in una delle due postazioni la lavorazione viene spostata nella zona opposta traslando la testa mobile nella zona opportuna. Quindi risulta evidente la necessità di un sistema di sicurezza che impedisca al macchinario di spostarsi nella zona interdotta se è occupata da un operatore. Secondo le norme di sicurezza specifiche, la zona di sicurezza necessaria ad una macchina di questo tipo è individuata dalla zona antistante alla stessa, figura 1.2. La dimensione di questa zona viene calcolata in base alla minima distanza consentita fra operatore ed elemento mobile, considerando la massima velocità di movimento della testa mobile ($90m/s$). Tenendo presente queste caratteristiche si viene ad individuare un rettangolo con lunghezza uguale a quella della macchina ed una larghezza di circa $1.5m$ dalla base del banco di lavoro. Nelle attuali norme non è previsto la necessità di controllare un volume ma solo un'area quindi non viene specificata nessuna altezza, anche se in futuro è plausibile che an-

che questo vincolo sia richiesto dalla legge. Un'ulteriore richiesta per poter lavorare simultaneamente su due zone di lavoro, come descritto precedentemente, è quella di dividere l'area di sicurezza in tre distinte zone. L'accesso o il divieto di ingresso nelle singole aree sarà stabilito real-time in base alla posizione di lavoro della testa mobile: mentre la zona centrale risulterà sempre inibita, per garantire comunque una distanza di tranquillità, le due laterali saranno inibite alternativamente consentendo agli operatori di caricare il materiale in una postazione quando contemporaneamente la testa si trova in lavoro sull'altra.



Figura 1.1: Macchina CNC per la lavorazione del legno.

1.2 Stato dell'arte

Esistono diversi studi e sistemi attualmente in commercio che cercano di affrontare questo tipo di problema e presentano soluzioni utilizzando diversi sensori. Di seguito



Figura 1.2: Zone di sicurezza di una macchina CNC e posizione dei sistemi stereo.

ne saranno presentati alcuni, descrivendone brevemente le caratteristiche:

- Tappeti a pressione - Posti al suolo lungo tutta l'area da controllare, segnalano la presenza di un ostacolo in base alla pressione che questo esercita sul tappeto. Permettono la divisione modulare delle aree ed una installazione molto semplice, non apportano ingombri materiali ai lavoratori, e sono insensibili alla polvere e ai piccoli scarti prodotti dalla lavorazione. Presentano svantaggi per quanto riguardano la durata nel tempo, in quanto sono soggetti a rotture in caso di cadute di attrezzi o schiacciamenti dovuti ai mezzi utilizzati per caricare i materiali sulla macchina. Analizzano solo il contatto a terra quindi non sono in grado di individuare oggetti fluttuanti.
- Paraurti sensibili - *Bumber* fisici applicati alla testa mobile che permettono il blocco della macchina in caso di pressione al contatto con un ostacolo. Consentiti solo a basse velocità della macchina, quindi prevedono necessariamente l'utilizzo di altri sistemi di sicurezza per rallentare la macchina.
- Laser scanner - Effettuano una scansione a raggiera su uno o più piani e forniscono con precisione la posizione degli ostacoli presenti nelle zone analizzate.

Essendo di facile installazione e di poco ingombro, esistono diverse possibilità di posizionamento, possono essere montati solidali alla testa mobile oppure in posizione fissa sotto la macchina. Presentano un grave problema in caso di polvere o trucioli prodotti durante la lavorazione, segnalando spesso blocchi non voluti alla macchina.

- Barriere ad infrarosso - Due barriere virtuale poste ai lati della zona da monitorare, che sfruttano l'interruzione di un raggio infrarosso in presenza di un ostacolo. Presentano una difficile installazione e una regolazione poco durevole nel tempo, dovuta a possibili disallineamenti fra ricevitore e trasmettitore.
- Sistemi di visione - Esiste solo un sensore di visione attualmente in commercio [59]: un sistema trinoculare che prevede un'installazione sopra la zona da sorvegliare ed effettua una duplice ricostruzione tridimensionale della scena sfruttando alternativamente le tre telecamere. Sebbene il sistema presenta una parziale ridondanza dei risultati, la similarità dei punti di vista delle tre telecamere e la sua posizione aerea non permette l'individuazione sicura di tutta l'area. L'ingombro della testa mobile non permette infatti al sensore di vedere la zona sottostante a questa, generando zone occluse in posizioni molto pericolose.

1.3 Sistema stereoscopico

Il sistema di visione stereoscopico proposto introduce un *Safety Integrity Level (SIL) 2* secondo la norma IEC 61508 [2] e un *Performance Level (PL) C* secondo la norma EN ISO 13849-1 [4]. Potrà quindi permettere la rimozione di eventuali altri sistemi di sicurezza basati su di una sensoristica diversa. Un sistema di visione fornisce infatti vantaggi che questi tipi di sensori non possono garantire, come ad esempio: funzionamento anche in caso di ambienti polverosi, rilevamento di oggetti che non toccano il terreno, basso impatto strutturale.

Le caratteristiche principali ed innovative che il sistema presenta sono:

- Riconoscimento di persone/ostacoli sospesi all'interno della zona di sicurezza;

- Eliminazione del numero di arresti errati della macchina durante la lavorazione a causa di caduta materiali di scarto della lavorazione o utensileria di dimensioni contenute;
- Modificazione real-time e ad-hoc mediante software della dimensione e forma delle zone da considerarsi di pericolo;
- Modificazione real-time e ad-hoc delle possibili azioni associate alla presenza di una persona/oggetto nella zona di pericolo.

Il sistema è composto da due coppie di telecamere 1.3 e da un'unità di elaborazione collegata ad una centralina per l'arresto fisico della macchina. Attraverso un algoritmo di individuazione ostacoli basato su mappa di disparità, il sistema è in grado di effettuare una ricostruzione tridimensionale della scena osservata e quindi individuare eventuali persone all'interno del perimetro di sicurezza. La zona da monitorare è divisa in tre aree di carico per permettere la selezione automatica delle zone da monitorare a seconda della posizione di lavoro della testa mobile. Il sistema è in grado di testare in modo continuo il proprio funzionamento ed in caso di guasto fermare la macchina attraverso la centralina. Le telecamere sono posizionate ai lati della zona da monitorare inquadrando anche la zona sottostante la testa mobile, questo elimina la possibilità di eventuali occlusioni da parte della testa stessa.



Figura 1.3: Sistema stereoscopico.

L'utilizzo di due blocchi di percezione per monitorare la stessa scena da due punti di vista diversi, assicura anche l'eliminazione di possibili zone occluse dovute alla presenza di operatori in una zona di carico consentita, in quanto almeno un sistema stereo avrà sempre la visuale libera. Utilizzando questo accorgimento tecnico, viene aumentata la ridondanza del sistema e quindi la sua affidabilità permettendo l'autorilevazione di possibili guasti al dispositivo di sicurezza.

Il software garantisce un massimo tempo di risposta pari a $100ms$ in grado di garantire l'arresto istantaneo della macchina al rilevamento di un ostacolo.

1.3.1 Configurazione Hardware

Le due teste stereoscopiche utilizzano ciascuna due telecamere USB2 distanti fra loro $20cm$ e montano lenti con la focale da $3.6mm$. L'utilizzo di una *baseline* così corta e di lenti di questo tipo portano ad un posizionamento laterale rispetto alla zona da vigilare di circa 1 metro, permettendo ad entrambe le camere di avere la visuale di tutta l'area, sia in larghezza che in lunghezza. Le camere si trovano ad un'altezza di $50cm$ da terra ed inquadrano un volume con altezza di un metro. Le immagini usate per la ricostruzione stereo sono acquisite alla risoluzione di 752×480 pixel ma ridotte a 376×240 per permettere un'elaborazione real-time. Tutte le camere sono sincronizzate da un segnale *trigger* hardware proveniente dal pc. L'unità di elaborazione è realizzata da un pc industriale con un processore Intel® Core™ i5-M520 2.4GHz e 4 GB di RAM al quale sono collegate tutte le camere ed dotato di una scheda IO digitale per l'interfacciamento con la centralina del quadro di comando della macchina CNC. Tutto l'hardware descritto è selezionato all'interno di una gamma di prodotti certificati per permettere successivamente al sistema stesso di ottenere una certificazione di affidabilità.

1.3.2 Algoritmo

In figura 1.4 sono illustrati i passi principali dell'algoritmo:

- Acquisizione delle immagini dalle telecamere;

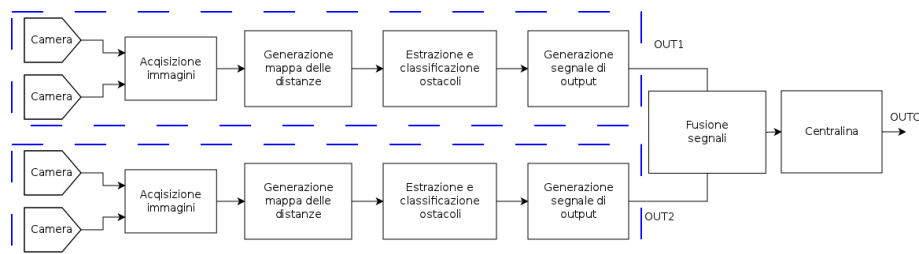


Figura 1.4: Schema a blocchi del sistema.

- Generazione della mappa di disparità per tutti i pixel dell'immagine;
- Estrazione degli ostacoli e verifica della loro potenziale pericolosità;
- Generazione del segnale di output.

Le teste stereo sono calibrate relativamente in fase di costruzione e quindi le immagini acquisite possono semplicemente essere rettificare usando parametri già calcolati. Successivamente viene generata la mappa di disparità con algoritmo globale SGM [15], descritto nel Cap 4.2.1. Per poter effettuare la ricostruzione tridimensionale, oltre al calcolo dell'immagine DSI (Disparity Space Map), serve una calibrazione del sistema per estrarre i parametri di imbardata, beccheggio e rollio di una sola camera per sistema stereo. La calibrazione si ottiene tramite una procedura semi-automatica che rileva dei marker con coordinate mondo note posti sui bordi della zona da sorvegliare. La presenza fissa dei marker, una volta finita la funzione di calibrazione, permette di effettuare dei controlli sulla possibile perdita di calibrazione del sistema. A seconda della gravità dell'errore sarà possibile effettuare una ricalibrazione automatica del sistema oppure segnalare la necessità di un intervento manuale da parte di un operatore. L'estrazione degli ostacoli presenti nella scena viene effettuata calcolando per ciascuna area tridimensionale il volume complessivo occupato dai punti dell'immagine di disparità. Se questo volume supera una soglia, determinata dall'ingombro volumetrico di un determinato oggetto campione, si segnala la presenza di un ostacolo all'interno di quella zona.

I possibili segnali di output, OUT-Destro e OUT-Sinistro, potranno assumere per zona gli stati di Libero, Occupato od Occluso. Infatti ogni sistema che rileva un ostacolo segnala come occluse tutte le zone poste più lontane di esse anche se queste sono state segnalate come occupate. Questo è dovuto al fatto che un sistema non può essere sicuro di quello che effettivamente avviene in una zona occlusa.

Inoltre il sistema è anche in grado di fornire un segnale indicante un possibile malfunzionamento di esso tramite un ulteriore segnale in aggiunta a quello delle aree di sicurezza. Causa di possibili errori potrebbero essere un degrado dell'immagine DSI tale da non garantire il funzionamento del sistema, oppure l'individuazione di un ostacolo posto fra la zona di sicurezza ed il sistema stereo, situazione di possibile pericolo da segnalare immediatamente.

I segnali dei due output relativi alle tre zone di riconoscimento presenza ostacolo/operatore vengono fuse alla fine della fase di elaborazione attraverso un'opportuna logica che genera un unico segnale di OUT per ogni zona e lo invia alla centralina. La tabella 1.1 mostra lo schema di gestione dei segnali da applicare ad ogni zona, indicando con il valore digitale 1 l'assenza di ostacoli e col valore 0 la presenza di ostacoli nella zona di sicurezza. La scelta di queste uscite è data dal fatto che in caso di spegnimento dell'unità di elaborazione l'uscita digitale andrebbe a valore basso, comportando l'arresto della macchina.

Sistema Sinistro	Sistema Destro	OUT
OCCLUSO	LIBERO	1
LIBERO	LIBERO	1
LIBERO	OCCLUSO	1
LIBERO	OCCUPATO	0
OCCUPATO	LIBERO	0
OCCUPATO	OCCUPATO	0
OCCLUSO	OCCLUSO	0

Tabella 1.1: Fusione dei segnali di Output

La centralina è collegata al controllo della macchina ed è in grado di bloccarla alla presenza di un qualsiasi output posto a zero o di un eventuale errore.

Si può notare che i due sistemi stereo effettuano due elaborazioni uguali ma distinte all'interno della stessa applicazione. Solo al livello di fusione degli output i due rami dell'algoritmo si unificano. All'interno dell'algoritmo quindi le due operazioni sono state divise su due *thread* separati indipendenti l'uno dall'altro, migliorando in questo modo il tempo di elaborazione totale.

1.4 Risultati

Alcuni esempi di funzionamento del sistema, durante un'installazione su una macchina campione, sono visibili nelle figure 1.5 e 1.6. Le immagini di ciascuna figura si riferiscono ad un caso specifico e presentano le immagini DSI con le relative ricostruzioni tridimensionali della stessa scena inquadrata dai due sistemi stereoscopici. L'immagine inferiore invece è un esempio di possibile interfaccia uomo-macchina (HMI) del sistema nella quale sono visibili in alto 3 riquadri che rappresentano gli output finali del sistema. In particolare nell'immagine 1.5 la scena presenta un pedone all'interno dell'area centrale di sicurezza. Nelle immagini DSI ottenute dalle due teste stereo, l'ostacolo risulta evidentemente al centro dell'immagine con colore verde; la differenza di colore fra le due immagini è data dalla diversa distanza dell'ostacolo dal sistema. La relativa ricostruzione tridimensionale mostra come il pedone sia l'unico ostacolo presente all'interno delle aree di sicurezza, ed in entrambe le visuali viene individuato nella stessa posizione centrale, illuminando di rosso la zona. Nell'immagine 1.5.e si vede come solo l'area centrale viene segnalata occupata accendendo un quadrato di colore rosso nella posizione opportuna. Le altre zone sono invece rappresentate dal colore verde in quanto sono libere.

Nell'altra scena, descritta dalle immagini in fig.1.6, l'operatore si trova nella zona di sicurezza sinistra ed è individuato correttamente anche in questo caso nella ricostruzione 3D da entrambi i sistemi stereoscopici. A differenza del caso precedente l'area occupata è evidenziata dal colore giallo nel singolo sistema in quanto le zone esterne potrebbero essere permesse nel caso in cui la macchina si trovi a lavo-

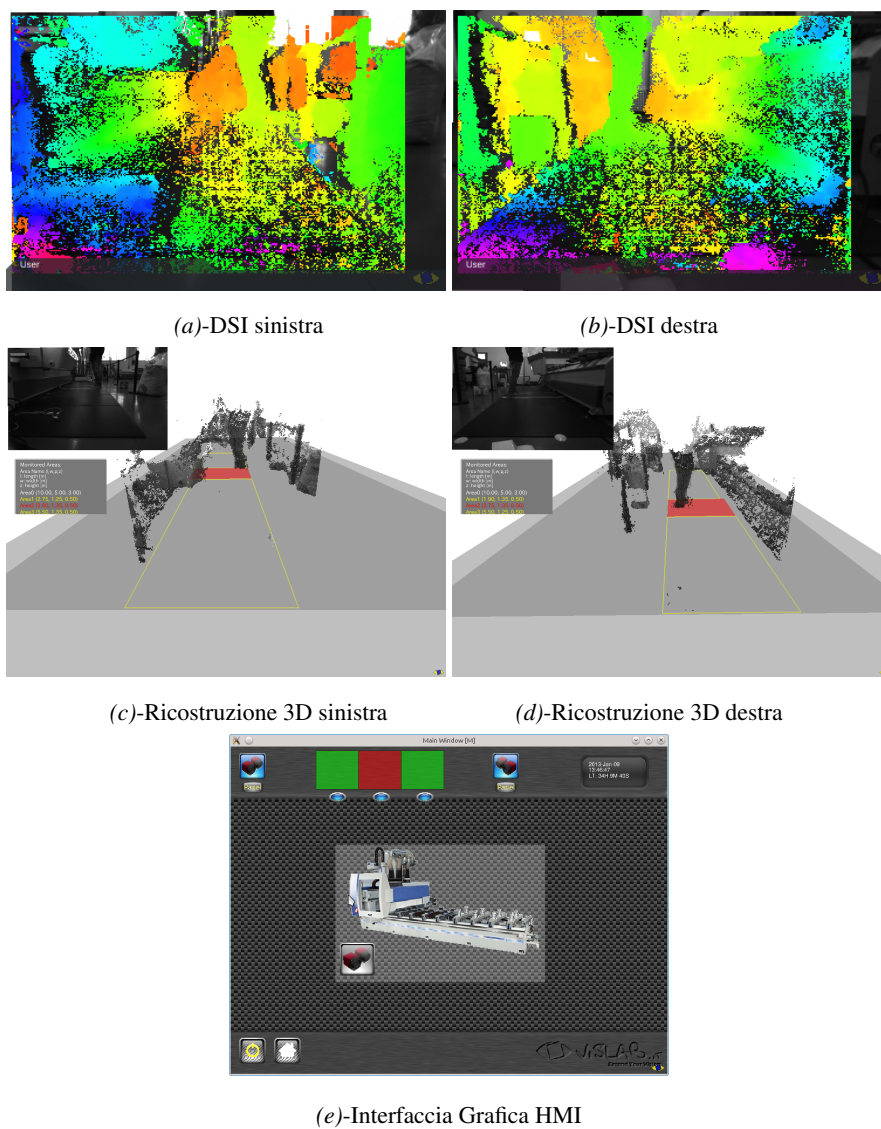


Figura 1.5: Output del sistema in caso di ostacolo nella zona centrale: (a,c)-DSI del sistema sinistro con relativa ricostruzione 3D della scena; (b,d)-DSI del sistema destro con relativa ricostruzione 3D della scena; (e)-Interfaccia Grafica HMI con output zone occupate.

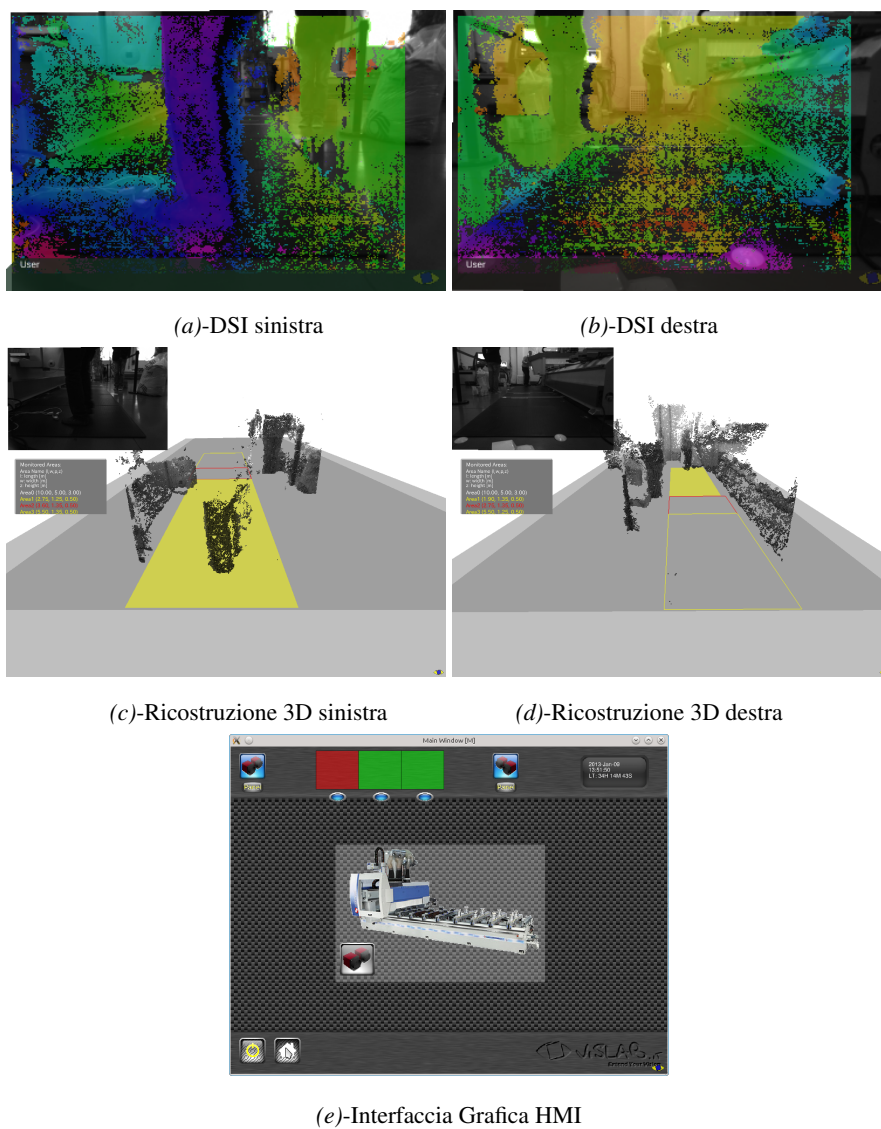


Figura 1.6: Output del sistema in caso di ostacolo nella zona sinistra: (a,c)-DSI del sistema sinistro con relativa ricostruzione 3D della scena; (b,d)-DSI del sistema destro con relativa ricostruzione 3D della scena; (e)-Interfaccia Grafica HMI con output zone occupate.

rare nell'area opposta. Nell'HMI 1.5.e invece questa zona viene evidenziata sempre col colore rosso per segnalare in ogni caso all'operatore l'intrusione di un ostacolo all'interno di qualsiasi area.

All'interno dell'interfaccia HMI sono presenti diversi *tool* che permettono: la disattivazione di determinate aree o di uno dei due sistemi, la calibrazione e la modifica di alcuni parametri del sistema, ed infine la segnalazione di possibili guasti tramite segnali luminosi

1.4.1 Tempi di elaborazione

L'algoritmo è stato testato su due diversi hardware: il primo è un desktop PC con un processore Intel® Core™ i7-2600K 3.4GHz e 8 GB RAM, mentre il secondo è il PC industriale installato sul campo, descritto nel Cap. 1.3.1. La sequenza di immagini su cui sono stati testati gli algoritmi è stata registrata utilizzando il secondo PC ed è stato possibile riprodurre il funzionamento in laboratorio. La tabella 1.2 contiene i tempi di elaborazione suddivisi per ognuna delle operazioni svolte nello schema dell'algoritmo. Le prestazioni dei rami stereo sono paragonabili fra loro e il tempo totale di elaborazione del singolo frame, ottenuto dalla somma dei due algoritmi più le operazioni comuni, rimane in tutti i casi sotto i 100ms, garantendo l'esecuzione real-time del sistema.

Visto che i due sistemi stereo effettuano due elaborazioni uguali ma distinte all'interno della stessa applicazione sarebbe possibile una divisione fisica dell'algoritmo su due unità di elaborazione diverse, aumentandone così le prestazioni e la frequenza di funzionamento.

1.4.2 Prestazioni

Per testare scientificamente le prestazioni del sistema sono stati effettuati vari tipi di test su una macchina CNC in funzione. Il sistema proposto è stato installato su una macchina che prevedeva già un sistema di sicurezza ottenuto dalla disposizione di tappeti sensibili nelle tre aree di sorveglianza. Nel grafico 1.7 viene mostrato il confronto dei segnali di output delle tre aree fra i due sistemi. Il test effettuato si

Fasi dell'algo- ritmo	Tempo di Elaborazione [ms]			
	Intel® Core™ i7-2600K 3.4GHz		Intel® Core™ i5-M520 2.4GHz	
	Sinistro	Destro	Sinistro	Destro
Preproc.	2.5	2.7	4.6	5.3
DSI	8.9	9.1	19.0	19.2
Obstacle	3.6	3.0	6.3	6.9
Tot.Stereo	15.0	14.8	29.9	31.4
Fusione	0.7		1.3	
Output	1.4		4.1	
Totale	31.9		66.7	

Tabella 1.2: Tempi di elaborazione del sistema

riferisce ad un operatore che entra ed esce varie volte dalle aree di arresto spostandosi fra di esse.

I segnali risultano fundamentalmente uguali, mostrando solo qualche piccola differenza. Ad esempio il sistema di visione è più sensibile perché analizza un volume e quindi arresta la macchina anche se viene individuato un ostacolo che non tocca il terreno. Alcuni ritardi di segnalazione sono anche da attribuire alla leggera differenza di disposizione delle aree da sorvegliare, in particolar modo nella parte frontale l'area impostata dal sistema di visione è risultata essere leggermente più stretta di quella ottenuta tramite la posa dei tappeti sensibili.

1.4.3 Conclusioni

Il sistema di visione stereoscopica per la sicurezza sul lavoro presentato e realizzato in questa tesi ha dimostrato di poter essere utilizzato come unico sistema di sicurezza per il controllo di aree di lavorazione adiacenti a macchine CNC per la lavorazione del legno. Il sistema è stato installato e testato in diverse situazioni e si è dimostrato

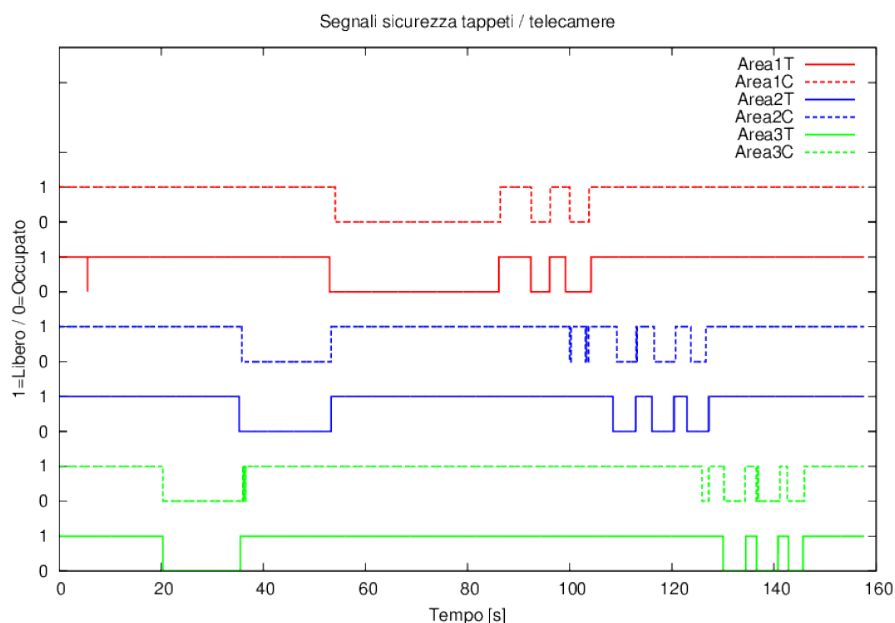


Figura 1.7: Grafico di confronto fra sistemi di sicurezza: i colori rappresentano i segnali di una stessa area; le linee continue sono riferite alle uscite dei tappeti, mentre quelle tratteggiate al segnale proveniente dal sistema di visione; un segnale alto rappresenta una zona libera, un segnale basso una zona occupata.

robusto ai test eseguiti. In particolare è stata testata la stabilità del sistema alla variazione di luminosità media dello scenario ed alla ripetibilità delle rilevazioni effettuate nelle varie aree di pericolo. I test hanno quindi dimostrato che si potranno rimuovere eventuali altri sistemi di sicurezza come griglie di protezione, laser, barriere ad infrarossi o tappeti sensibili alla pressione. Il sistema di visione fornisce infatti vantaggi che questi tipi di sensori non possono garantire: funzionamento anche in caso di ambienti polverosi, rilevamento di oggetti che non toccano il terreno, basso impatto strutturale nella scena. Rispetto ad altri sistemi di visione, inoltre i due sistemi stereo fissati ai lati opposti dell'area da vigilare garantiscono la ridondanza dei risultati e

l'esclusione di possibili zone occluse, garantendo quindi l'affidabilità del sistema. La presenza di procedure di autodiagnosi che segnalano l'insorgere di possibili problemi dovuti a guasti hardware o errori software dovuti alla perdita della calibrazione, l'utilizzo di hardware industriali certificati e una scrittura del codice seguendo le norme di sicurezza imposte dalla legge, hanno reso possibile una certificazione ufficiale di sicurezza. Il sistema è stato brevettato [12] e se ne prevede una produzione industriale nei prossimi anni.

Capitolo 2

Sistema di visione per l'apertura automatica di porte

In questo capitolo viene presentato un sistema stereo per individuare ed estrarre le traiettorie dei pedoni in prossimità di accessi con porte scorrevoli per permettere l'apertura automatica delle porte. Il sistema comanda l'apertura delle porte scorrevoli solo se il pedone cammina in direzione delle stesse, mentre nel caso in cui la traiettoria del pedone sia parallela alla porta, il sistema evita l'apertura interpretando l'intenzione della persona di volere solo passare e non attraversare il varco. Questo permette di evitare aperture non volute ed inutili, consentendo risparmi per la manutenzione fisica della porta e l'aumento dell'efficienza delle strutture per quanto riguarda la temperatura (ad es. riscaldamento ed aria condizionata).

L'innovazione e la robustezza di questo sistema è data dall'implementazione di un nuovo metodo per l'estrazione dell'immagine di disparità. L'algoritmo è stato testato in condizioni reali per misurarne l'efficienza e stimare le sue prestazioni.

2.1 Introduzione

Gli attuali sistemi per il controllo degli accessi per porte automatiche richiedono un sensore in grado di individuare la presenza di un oggetto o di un pedone che si trova

nella zona antistante una porta. Questo approccio non analizza tuttavia la traiettoria dei pedoni e non può neppure stimarne i movimenti. Ad esempio se un pedone attraversasse l'area di fronte alla porta ma senza volerla attraversarla, un sensore ad infrarosso individuerebbe la sua presenza e comanderebbe in ogni caso l'apertura. In questo caso il sistema non sarebbe efficiente producendo una piccola perdita di energia in termini di elettricità, aria condizionata, riscaldamento ed andando a diminuire il ciclo di vita meccanico del sistema stesso effettuando un'apertura e una chiusura non necessarie. Se si considera le migliaia di passaggi a cui sono soggette alcuni accessi risulta evidente come nel tempo questo tipo di soluzione porterebbe ad evidenti sprechi.

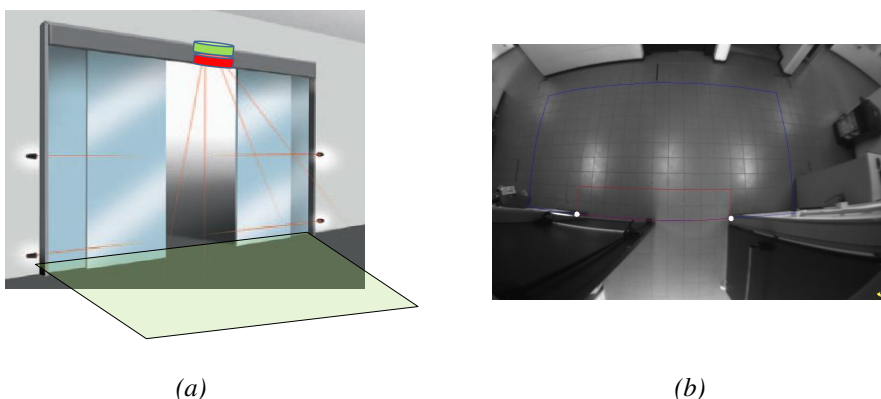


Figura 2.1: (a) Il sistema stereo (in verde) collocato sulla porta, sostituisce quello classico ad infrarossi (in rosso). (b) Un'immagine acquisita da una singola camera installata su un accesso con porte automatiche. L'area monitorata dal nuovo sistema di visione (piano verde) è più grande di quella inquadrata dai sensori classici (linea rossa) (b) e quindi permette l'individuazione della traiettoria dei pedoni.

La scena inquadrata dalle telecamere, visibile in figura 2.1.b, mostra come l'area monitorata dal sistema è ampia ed in grado di individuare pedoni anche a distanza dalla porta. Le immagini acquisite sono poi elaborate da un computer che tramite uno specifico algoritmo controllerà una scheda di controllo per la gestione dell'apertura della porta

In letteratura sono presenti altri sistemi di questo tipo [54, 47, 23, 37, 55, 57], ma la caratteristica innovativa di questo sistema risiede proprio nel metodo di ricostruzione 3D, in quanto la mappa di disparità non viene calcolata sulle immagini acquisite ma su di una loro trasformata, Inverse Perspective Mapping (IPM [44]). L'utilizzo della trasformata IPM porta vantaggi dal punto di vista della calibrazione e dell'installazione del sistema come verrà descritto nei capitoli successivi.

In dettaglio il nuovo metodo per la costruzione dell'immagine di disparità sarà trattato nel Cap. 2.2; l'algoritmo per l'identificazione della traiettoria nel Cap. 2.3; ed infine i risultati, considerazioni finali ed idee di sviluppi futuri verranno presentati nel Cap. 2.4.

2.2 Disparità e DSI su immagini stereo IPM

In un sistema di visione stereoscopico, per effettuare una ricostruzione 3D della scena osservata, sfruttando il concetto di disparità, devono essere noti i parametri estrinseci ed intrinseci del sistema stereo. Per fare sì che il sistema sia efficiente, le telecamere devono essere posizionate in modo opportuno con gli stessi angoli di imbardata, beccheggio e rollio. In alternativa è possibile effettuare una rettificazione dell'immagine tramite la calibrazione del sistema, operazione lunga e complicata, che necessita della conoscenza esatta dei parametri della telecamera. Il sistema proposto su cui si vuole applicare l'algoritmo di disparità non presenta invece vincoli se non quello della conoscenza della posizione delle telecamere. Il sistema non necessita di una calibrazione complessa, ma con una semplice mappatura di una griglia viene effettuata la rimozione della distorsione e della prospettiva con la generazione *offline* di una LUT per ogni telecamera. L'applicazione della LUT all'immagine acquisita dalla telecamera grandangolare genera l'*Inverse Perspective Mapping* dell'immagine (IPM).

L'IPM è una trasformazione geometrica che permette di rimuovere l'effetto prospettico delle immagini acquisite [10, 9]. Essa ricampiona l'immagine sorgente mappando ogni pixel in una posizione differente creando una nuova immagine IPM che rappresenta una vista dall'alto della scena osservata. Questa tecnica consente di omo-

geneizzare il contenuto informativo di ogni singolo pixel dell'immagine, sottocampionando i pixel dell'immagine sorgente vicini alle telecamere e sovracampionando i punti lontani. Purtroppo non risulta possibile rimappare in modo corretto tutti i pixel dell'immagine sorgente ma solo quelli appartenenti ad un piano noto a priori (in genere il terreno, $Z = 0$). I punti non appartenenti al suddetto piano vengono rimappati in una posizione non corretta che dipende dalla loro posizione nel mondo e dalla posizione delle telecamere.

La LUT che definisce l'*inverse perspective mapping* può essere definita a partire dai parametri intrinseci ed estrinseci del sistema di visione oppure tramite una mappatura manuale dei punti di una griglia di calibrazione e una successiva interpolazione. In quest'ultimo caso la LUT ottenuta permette di rimuovere sia la prospettiva sia la distorsione dell'immagine sorgente [24, 26, 53].

2.2.1 Inverse Perspective Mapping

Si vuole ora calcolare come un punto non appartenente al piano $Z = 0$ viene rimappato sull'immagine IPM, analizzando analiticamente il processo di generazione dell'immagine.

L'immagine IPM è il risultato di due trasformazioni prospettiche: una diretta e l'altra inversa, come rappresentato in figura 2.2. I punti del mondo sono mappati in un'immagine prospettica¹ tramite una trasformazione prospettica e successivamente in un'immagine IPM imponendo che tutti i punti si trovino all'altezza del suolo.

Nei passaggi seguenti i punti del mondo (X, Y, Z) e quelli dell'IPM (X', Y') saranno espressi nella stessa unità di misura, cioè in metri. Nella realtà le coordinate dell'IPM, trattandosi di un'immagine, dovranno poi essere convertite da metri a pixel tramite un opportuno fattore di scala e di *offset*.

Si prenda l'equazione base A.10 di una semplice trasformazione prospettica PM, descritta nell'appendice A.2:

¹Questa trasformazione avviene all'interno della telecamera nell'atto di acquisizione di un'immagine.

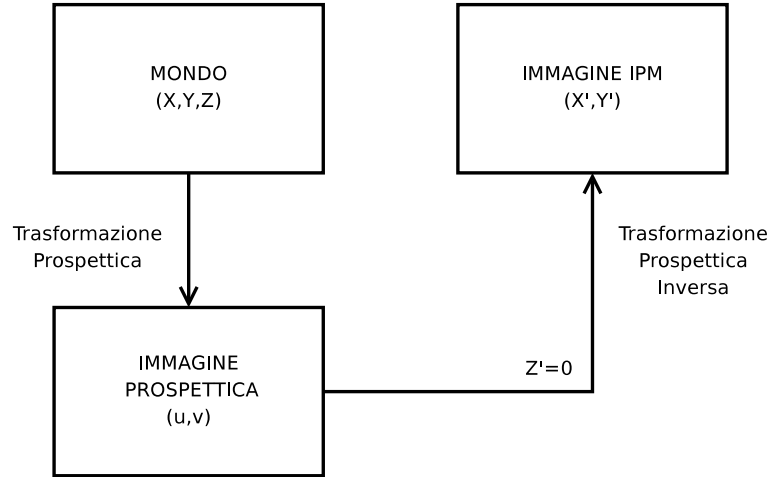


Figura 2.2: Schema di generazione dell'IPM.

$$\mathbf{c} = \mathbf{A}\mathbf{R}_{\psi\phi\theta}\mathbf{T}\mathbf{w}$$

Questa equazione è espressa in coordinate omogenee su quattro dimensioni. È possibile riscrivere l'equazione portandosi in coordinate omogenee tridimensionali:

$$\mathbf{c} = \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \mathbf{A}\mathbf{R}_{\psi\phi\theta} \left(\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} - \begin{bmatrix} x_0 \\ y_0 \\ z_0 \end{bmatrix} \right) \quad (2.1)$$

dove (X, Y, Z) sono le coordinate dei punti nel mondo, (x_0, y_0, z_0) sono le coordinate della posizione nel mondo della telecamera (u, v) le coordinate del punto nell'immagine prospettica. \mathbf{A} ed \mathbf{R} sono rispettivamente le matrici (3x3) dei parametri intrinseci e di rotazione.

Definiamo ora la trasformazione prospettica inversa (IPM), che riproietta i punti di un'immagine prospettica nell'immagine IPM. L'equazione di questa trasformazione si ricava invertendo la trasformazione PM appena ricavata, eq. 2.1, moltiplicando entrambi i membri prima per \mathbf{A}^{-1} e poi per $\mathbf{R}_{\psi\phi\theta}^{-1}$:

$$\left(\begin{bmatrix} X' \\ Y' \\ Z' \end{bmatrix} - \begin{bmatrix} x_0 \\ y_0 \\ z_0 \end{bmatrix} \right) = \mathbf{R}_{\psi\phi\theta}^{-1} \mathbf{A}^{-1} \begin{bmatrix} u \\ v \\ 1 \end{bmatrix}$$

Dove Z' è l'altezza dal suolo del punto da proiettare. Ma sappiamo che l'IPM ignora la reale altezza dei punti considerandoli tutti uguali all'altezza del suolo, cioè $Z' = 0$. Essendo (X', Y', Z') le coordinate dell'immagine IPM poniamo quindi $Z' = 0$ e sostituiamo alle coordinate $(u, v, 1)$ l'equazione 2.1:

$$\left(\begin{bmatrix} X' \\ Y' \\ 0 \end{bmatrix} - \begin{bmatrix} x_0 \\ y_0 \\ z_0 \end{bmatrix} \right) = \mathbf{R}_{\psi\phi\theta}^{-1} \mathbf{A}^{-1} \mathbf{A} \mathbf{R}_{\psi\phi\theta} \left(\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} - \begin{bmatrix} x_0 \\ y_0 \\ z_0 \end{bmatrix} \right)$$

e semplificando:

$$\begin{bmatrix} X' - x_0 \\ Y' - y_0 \\ -z_0 \end{bmatrix} = \begin{bmatrix} X - x_0 \\ Y - y_0 \\ Z - z_0 \end{bmatrix}$$

Sapendo che le coordinate omogenee sono definite a meno di un coefficiente di proporzionalità, otteniamo:

$$\begin{bmatrix} X' - x_0 \\ Y' - y_0 \\ -z_0 \end{bmatrix} = \begin{bmatrix} X - x_0 \\ Y - y_0 \\ Z - z_0 \end{bmatrix} \lambda$$

Risolviamo ora il sistema ricavando λ dalla terza equazione:

$$\lambda = \frac{-z_0}{Z - z_0}$$

$$X' = -z_0 \frac{X - x_0}{(Z - z_0)} + x_0 \quad (2.2)$$

$$Y' = -z_0 \frac{Y - y_0}{(Z - z_0)} + y_0 \quad (2.3)$$

Si è, così, ottenuto un notevole risultato, infatti le equazioni 2.2 e 2.3 non dipendono dai parametri intrinseci della lente e della telecamera, esclusa la loro posizione nel mondo. L'IPM quindi non dipende dal metodo con cui è stata calcolata (calibrazione o LUT).

Da questi risultati si nota che lo stesso punto nel mondo (X, Y, Z) viene rimappato in coordinate IPM diverse (X', Y') al variare della posizione della telecamera nel mondo. Se il punto mondo ha $Z = 0$, cioè appartiene al suolo, le equazioni si semplificano e risulta che $X' = X$ e $Y' = Y$ indipendentemente dalla posizione della telecamera.

Nella figura 2.3 è possibile notare come l'immagine di una strada viene trasformata in una vista dall'alto nell'immagine IPM, eliminando la prospettiva all'immagine acquisita. Le linee della strada risultano parallele perché poste sul terreno e quindi rimappate correttamente. Gli ostacoli con altezza dal suolo maggiore di zero sono invece rimappati in forme diverse in base alla loro posizione e per questo risultano deformati.

Le equazioni 2.2 e 2.3 possono essere combinate insieme, eliminando la variabile Z , ottenendo:

$$\frac{X' - x_0}{X - x_0} = \frac{Y' - y_0}{Y - y_0} \quad (2.4)$$

Fissando ora X e Y , il che equivale a prendere un ostacolo con base puntiforme che si erge dal suolo, l'equazione 2.4 diventa una retta con coefficiente angolare generale m' :

$$m' = \frac{X - x_0}{Y - y_0}$$

Tutti i punti mondo che si trovano su una stessa retta verticale, perpendicolare al terreno, con coordinate (X, Y) , saranno rimappati nell'IPM sulla retta che collega il *pin-hole* alla posizione della base dell'ostacolo. La retta dipende dalle coordinate (X, Y) del punto del mondo che si vuole proiettare e dalla posizione della

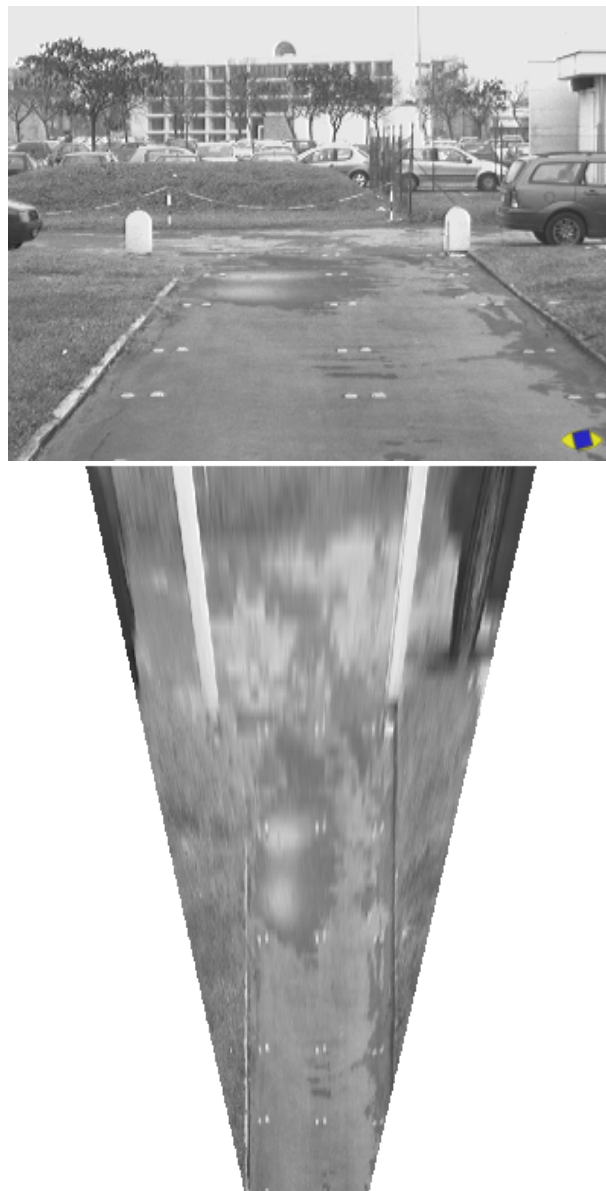


Figura 2.3: Immagine prospettica e immagine IPM di una strada.

telecamera (x_0, y_0, z_0) . La coordinata z_0 non influenza però la retta su cui gli ostacoli verticali saranno proiettati, ma solo la loro posizione su di essa².

Nelle figure 2.4 si vede il corpo di un pedone, che, tolta la distorsione delle immagini, può considerarsi un oggetto verticale. Risulta evidente come il pedone venga rimappato nelle immagini 2.5 su due rette con angolazione diversa nelle due IPM.



Figura 2.4: Immagini acquisite dalla telecamera sinistra e destra.

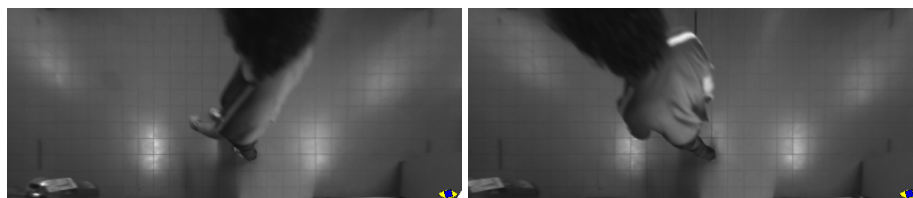


Figura 2.5: Immagini IPM: sinistra e destra.

2.2.2 Calcolo della disparità e ricostruzione 3D su immagini IPM

Prendiamo in esame un sistema composto da due telecamere stereo allineate su una retta parallela all'asse Y , poste alla stessa altezza h e distanziate fra di loro di una *baseline* b . La telecamera destra si troverà in $(x_0, -b_d, h)$; quella sinistra in (x_0, b_s, h) .

²vedere eq. 2.2, eq. 2.3.

Calcoliamo ora, tramite le equazioni precedentemente trovate (eq. 2.2, eq. 2.3), le coordinate IPM di un generico punto del mondo $P(X, Y, Z)$, rimappato nelle due IPM generate dalle immagini acquisite dalle due telecamere:

$$X'_d = X'_s = -h \frac{X - x_0}{Z - h} + x_0$$

$$Y'_d = -h \frac{Y + b_d}{Z - h} - b_d$$

$$Y'_s = -h \frac{Y - b_s}{Z - h} + b_s$$

Esprimiamo la coordinata Y'_d in funzione della coordinata Y'_s :

$$Y'_d = Y'_s - b \left(\frac{h}{Z - h} + 1 \right) = Y'_s + b \frac{Z}{h - Z}$$

Considerando che la coordinata X' è uguale per entrambe le immagini, $X'_d = X'_s$, possiamo definire il concetto di disparità:

La disparità di un punto mondo di coordinate (X, Y, Z) proiettato su due immagini IPM è pari a:

$$\Delta = Y'_d - Y'_s = b \frac{Z}{h - Z} \quad (2.5)$$

Dalla definizione di disparità su IPM data dalla formula 2.5, derivano molte proprietà che possono essere estratte analizzando la formula 2.5 e la figura 2.6:

- Dato che nell'immagine IPM sono rimappati solo punti del mondo con un'altezza dal suolo minore rispetto a quella della telecamera, cioè con $Z < h$, la disparità risulterà sempre positiva.
- La disparità di un punto non dipende dalle coordinate (X, Y) del punto nel mondo, ma solo dalla sua coordinata Z . Questo implica che tutti i punti posti ad una stessa altezza Z nel mondo, possiedono la stessa disparità Δ nell'immagine IPM. Il suolo, avendo un'altezza nulla ($Z = 0$), avrà una disparità nulla

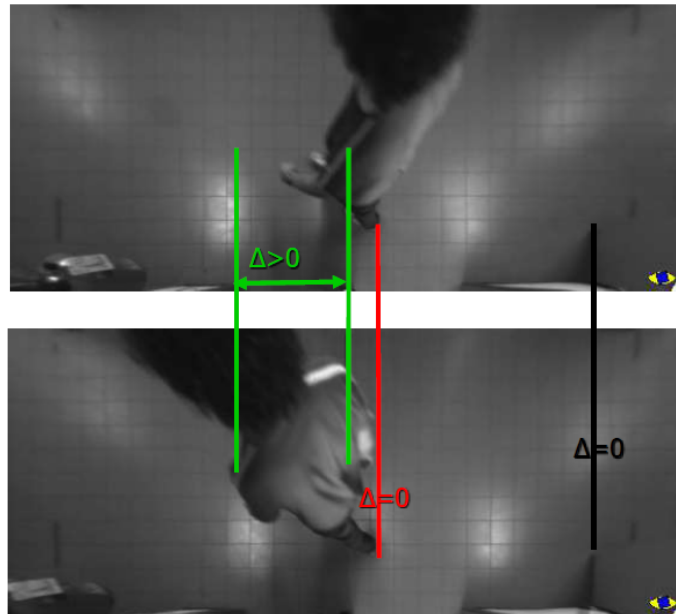


Figura 2.6: Disparità su immagini IPM.

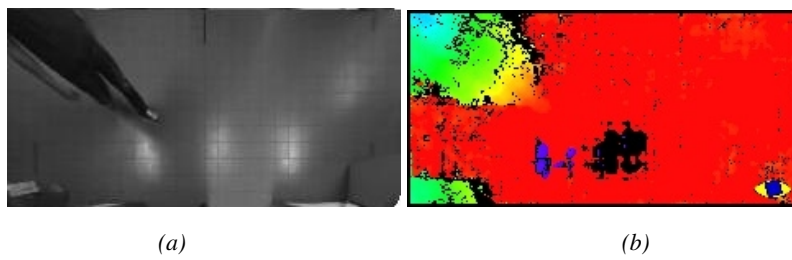


Figura 2.7: (a) Immagine IPM destra, (b) Immagine di disparità.

$\Delta = 0$. All'aumentare dell'altezza aumenterà anche la disparità, fino ad arrivare all'altezza $Z = h$ dove l'equazione presenta una singolarità.

- La disparità dipende proporzionalmente dalla distanza fra le due telecamere (*baseline*).
- La disparità non dipende dalla posizione x_0 delle telecamere.

La conoscenza della disparità dei punti di un immagine IPM permette la ricostruzione tridimensionale della scena. Le formule per ricavare le coordinate mondo di un punto IPM derivano dalla stessa definizione di disparità data precedentemente. In particolare la coordinata Z si ottiene direttamente dall'eq. 2.5:

$$Z = h \left(1 - \frac{b}{\Delta + b} \right) = h \frac{\Delta}{\Delta + b} \quad (2.6)$$

dove h e b sono rispettivamente l'altezza dal suolo e la *baseline* delle telecamere. Sostituiamo adesso all'equazioni 2.2 e 2.3 il valore di Z appena trovato ed otteniamo:

$$X'_d = -h \frac{X - x_0}{\left(h \frac{\Delta}{\Delta + b} - h \right)} + x_0$$

$$Y'_d = -h \frac{Y + b_d}{\left(h \frac{\Delta}{\Delta + b} - h \right)} - b_d$$

da cui è possibile ricavare i valori delle coordinate mondo (X, Y) in funzione della disparità:

$$X = X'_d \left(\frac{b}{\Delta + b} \right) + x_0 \left(\frac{\Delta}{\Delta + b} \right) \quad (2.7)$$

$$Y = Y'_d \left(\frac{b}{\Delta + b} \right) - b_d \left(\frac{\Delta}{\Delta + b} \right) \quad (2.8)$$

2.2.3 L'immagine di disparità DSI su immagini IPM

L'immagine di disparità DSI, ovvero *Disparity Space Image*, è tipicamente utilizzata dai sistemi di visione stereo classici, vedere appendice A.6, ma può essere usata efficientemente anche su immagini IPM. Infatti gli algoritmi per l'implementazione della DSI, come l'algoritmo SAD³, richiedono di cercare per ogni pixel dell'immagine destra il corrispondente pixel nell'immagine sinistra, detti punti omologhi. Anche nel caso di due immagini IPM è possibile effettuare una ricerca di punti omologhi per ricavare la disparità di tali punti. In figura 2.7.b è rappresentata un esempio di DSI, in cui ad ogni pixel è assegnato un valore pari alla propria disparità. Conoscendo il valore di disparità di un punto è poi possibile ricavare le sue coordinate mondo, come illustrato nel Cap. 2.2.2, e quindi, effettuare una ricostruzione 3D completa della scena osservata.

La costruzione dell'immagine di disparità richiede che, per ogni punto dell'immagine destra, si ricerchi il pixel omologo sull'immagine sinistra. Il pixel omologo, per ragioni di efficienza, viene ricercato sulla stessa riga del pixel dell'immagine destra. Infatti dall'equazione 2.5 sappiamo che la coordinata X' di un punto mondo dipende solo dalla sua altezza e quindi sarà uguale nelle due immagini. La ricerca dei pixel omologhi viene effettuata tramite *match* fra finestre.

Prima di effettuare la ricerca dei punti omologhi viene applicato alle immagini IPM un filtro di Sobel, che permette di eliminare le zone con basso contenuto informativo che produrrebbero falsi *match*.

Facciamo ora alcune considerazioni sull'immagine di disparità ottenuta dalle immagini IPM:

- La disparità di un pixel è sempre positiva, quindi vengono provati solo *match* corrispondenti a disparità positive.
- Il suolo ha disparità uguale a zero.

³Sum of Absolute Difference [27]

- Se la correlazione fra due finestre non è abbastanza alta, al pixel viene assegnato un valore riservato di disparità, per indicare che il pixel non ha riscontri nell'altra immagine.
- La disparità di un oggetto, che nel mondo è perpendicolare al suolo, aumenta all'aumentare della sua altezza. Gli oggetti verticali presentano una crescita di disparità lungo la retta definita dall'eq.2.4.
- La trasformazione IPM deforma in modo diverso gli oggetti che si ergono dal suolo, quindi gli oggetti che in prospettiva sono simili, riproiettati possono risultare differenti. Questo comporta una maggior difficoltà nell'associazione fra punti omologhi.
- La dimensione delle immagini IPM è un parametro che influenza notevolmente i tempi di calcolo della DSI. I confronti per la ricerca dei punti omologhi e quindi anche il tempo dell'elaborazione, aumentano con un ordine cubico rispetto alle dimensioni dell'immagine.
- Per diminuire il tempo di elaborazione è possibile far effettuare la ricerca del punto omologo solo entro un determinato *range* di ricerca. In quanto, un punto in un'immagine IPM può avere un valore di disparità limitato che dipende dalla sua posizione.

Calcolo del *range* di ricerca

La ricerca dei punti omologhi viene effettuata sull'immagine IPM sinistra sulla stessa riga del punto dell'immagine destra. Il punto omologo però non potrà trovarsi su tutti i punti della riga, ma solo in una determinata zona.

Sappiamo dalla definizione data precedentemente che la disparità è sempre positiva e il valore minimo che può assumere è uguale a zero se si riferisce ad un punto appartenente al suolo. Quindi il punto omologo si potrà trovare solo in una posizione con coordinata IPM maggiore o uguale rispetto a quella del punto dell'immagine destra.

Allo stesso modo ogni punto possiede anche una possibile disparità massima, dipendente dalla posizione nell'immagine. Cerchiamo di calcolare questo valore per poter così definire per ogni punto dell'immagine destra il *range* di ricerca della disparità.

Dato un punto $P'_d = (X'_d, Y'_d)$ nell'immagine IPM destra si vuole calcolare il valore massimo di disparità che questo punto può assumere. Per fare questo definiamo un valore X_{min} che rappresenta la distanza minima dove si può trovare la base di un ostacolo verticale. Sappiamo che un ostacolo verticale al terreno con la base nel punto di coordinate $(X, Y, 0)$ viene rimappato nell'immagine IPM destra, equazione 2.4, su una retta passante per il punto (X, Y) e la proiezione del suo *pin-hole*:

$$\frac{X - x_0}{X'_d - x_0} = \frac{Y + b_d}{Y'_d + b_d} \quad (2.9)$$

Nel caso peggiore la base dell'ostacolo verticale rimappato su questa retta si può trovare nel punto di coordinata $X = X_{min}$ (questo è il caso che massimizza l'altezza di un ostacolo e quindi la sua disparità). Sostituiamo quindi questo valore alla retta 2.9:

$$\frac{X_{min} - x_0}{X'_d - x_0} = \frac{Y + b_d}{Y'_d + b_d} \quad (2.10)$$

Definiamo ora la retta sull'immagine IPM sinistra passante per il punto di coordinate (X, Y) e la proiezione sul terreno del suo *pin-hole*, sostituendo anche in questo caso il valore di X con X_{min} :

$$\frac{X_{min} - x_0}{X'_d - x_0} = \frac{Y - b_s}{Y'_s - b_s} \quad (2.11)$$

Nell'equazione 2.11 è stata sfruttata l'uguaglianza che lega le coordinate X' di uno stesso punto mondo, sostituendo ad X'_s la coordinata X'_d .

Ricaviamo ora dall'equazioni 2.10 e 2.11 il valore di Y'_d e Y'_s :

$$Y'_d = \frac{(Y + b_d)(X'_d - x_0)}{X_{min} - x_0} - b_d$$

$$Y'_s = \frac{(Y - b_s)(X'_d - x_0)}{X_{min} - x_0} + b_s$$

Sostituendo queste equazioni alla definizione di disparità 2.5 si ricava il valore della disparità massima:

$$\Delta_{max} = Y'_d - Y'_s = \frac{(X'_d - x_0)}{X_{min} - x_0} (Y + b_d - Y + b_s) - b_d - b_s$$

semplificando si ottiene il risultato:

$$\Delta_{max} = b \frac{X'_d - x_0}{X_{min} - x_0} - b = b \frac{X'_d - X_{min}}{X_{min} - x_0} \quad (2.12)$$

Questa equazione può anche essere ricavata geometricamente. Nella figura 2.8 si può individuare una similitudine fra il triangolo con vertici nei punti P'_d, P'_s, I e quello

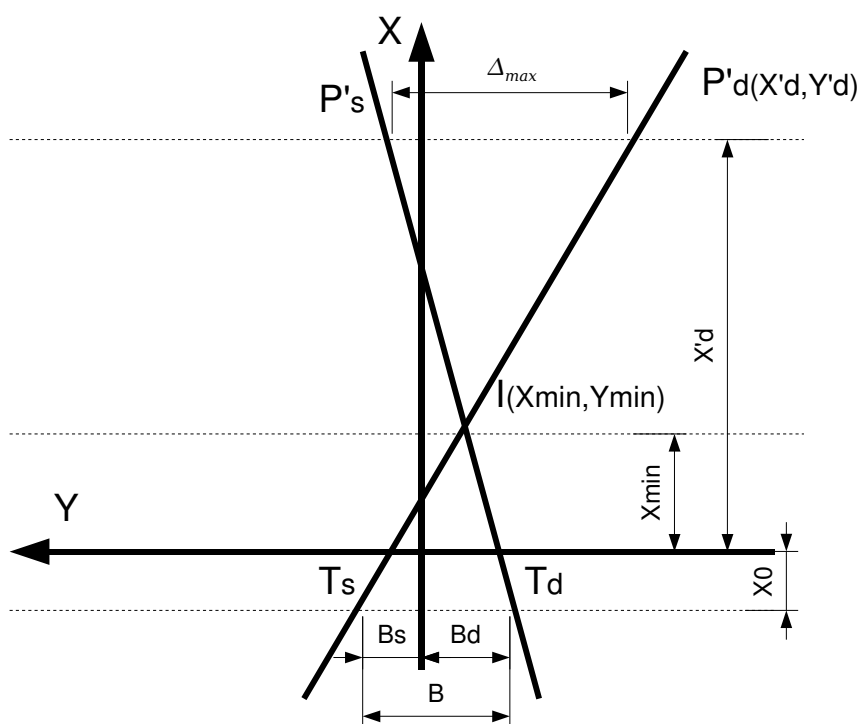


Figura 2.8: Calcolo della disparità massima.

con vertici T_d, T_s, I . Dalla similitudine fra le basi e le altezze di questi triangoli, si ricava l'equazione 2.12 trovata precedentemente.

$$\Delta_{max} : b = X'_d - X_{min} : X_{min} - x_0$$

La disparità massima che un punto dell'IPM può assumere non dipende dalla sua coordinata Y' , ma solo dalla coordinata X' . Quindi ogni riga orizzontale dell'immagine IPM ha la stessa disparità massima. La disparità massima dipende proporzionalmente dalla *baseline*, cioè solo dalla distanza fra le due telecamere ma non dalla loro posizione y . Con questa formula possiamo così calcolare per ogni punto dell'immagine IPM il *range* di ricerca dell'immagine di disparità. Per fare ciò è necessario convertire la Δ_{max} da metri a pixel, tramite un fattore di scala, tenendo conto delle dimensioni dell'immagine IPM.

2.3 L'algoritmo

In questo capitolo viene presentata l'effettiva implementazione dell'algoritmo usato per realizzare il sistema. Lo schema 2.9 mostra il diagramma a blocchi dei passi principali dell'algoritmo.

Date due immagini stereo, viene identificata e analizzata la direzione degli oggetti che si muovono nella scena per controllare l'apertura delle porte. Come visto nel capitolo precedente 2.2, l'algoritmo non richiede una calibrazione complessa: in quanto le immagini IPM della scena sono generate tramite una Look Up Table (LUT) calcolata *off-line*. Le immagini così ottenute sono utilizzate per la generazione di un'immagine di background per ogni telecamera. La generazione delle due immagini di riferimento avviene utilizzando un algoritmo di generazione standard che calcola una media pesata delle ultime N immagini salvate all'interno di un vettore circolare [5, 50, 43]. Successivamente le immagini di background sono confrontate tramite differenza con l'immagine di input, generando immagini di maschera a toni di grigio. Queste immagini sono poi binarizzate attraverso una soglia ed elaborate tramite un'espansione per renderle meno sensibili a possibili rumori, fig. 2.10.

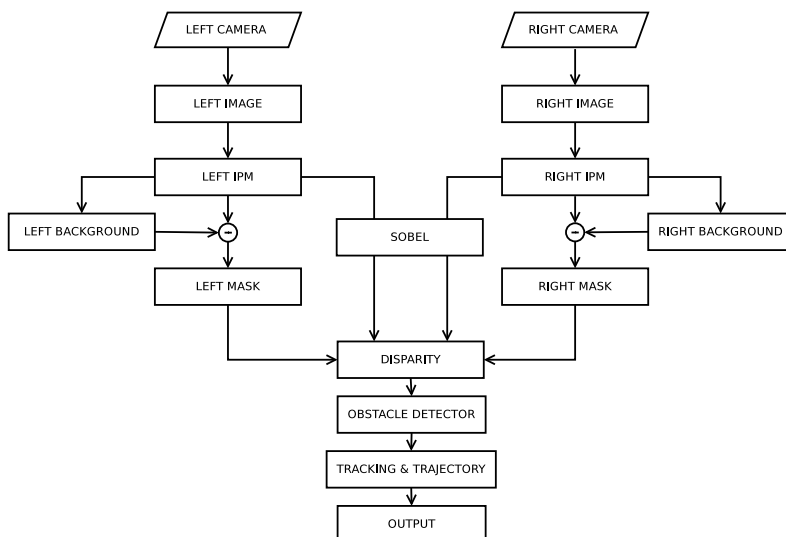


Figura 2.9: Diagramma dell'algoritmo proposto.

Successivamente un filtro di Sobel è applicato alle immagini per ridurre le problematiche legate al cambio di condizioni di luce ambientale. Il risultato di questa operazione sono due immagini che enfatizzano i bordi sulle quali è possibile applicare l'algoritmo di disparità [27] per la generazione della Disparity Space Image (DSI). Grazie all'uso preliminare della trasformata IPM, l'immagine DSI risultante può essere interpretata usando l'approccio specifico sviluppato nel capitolo precedente, vedere Cap. 2.2.

L'immagine DSI è quindi filtrata applicando le due maschere calcolate precedentemente: vengono presi in esame solo i punti della DSI appartenenti alla maschera destra, di questi sono validati solo quelli che sommandovi la corrispondente disparità risultano appartenere alla maschera sinistra. Questo approccio riduce i falsi positivi e gli errori nel calcolo della DSI, dovuti ad esempio a pattern ripetitivi e alle riflessioni.

L'immagine così ottenuta è usata dall'algoritmo di *obstacle detection* per individuare gli ostacoli all'interno della scena. L'immagine DSI è quindi clusterizzata usando sia la contiguità dei punti sull'immagine sia vincoli spaziali nella loro posizione del mondo, ottenuta tramite le operazioni descritte nel capitolo 2.2. Il risultato

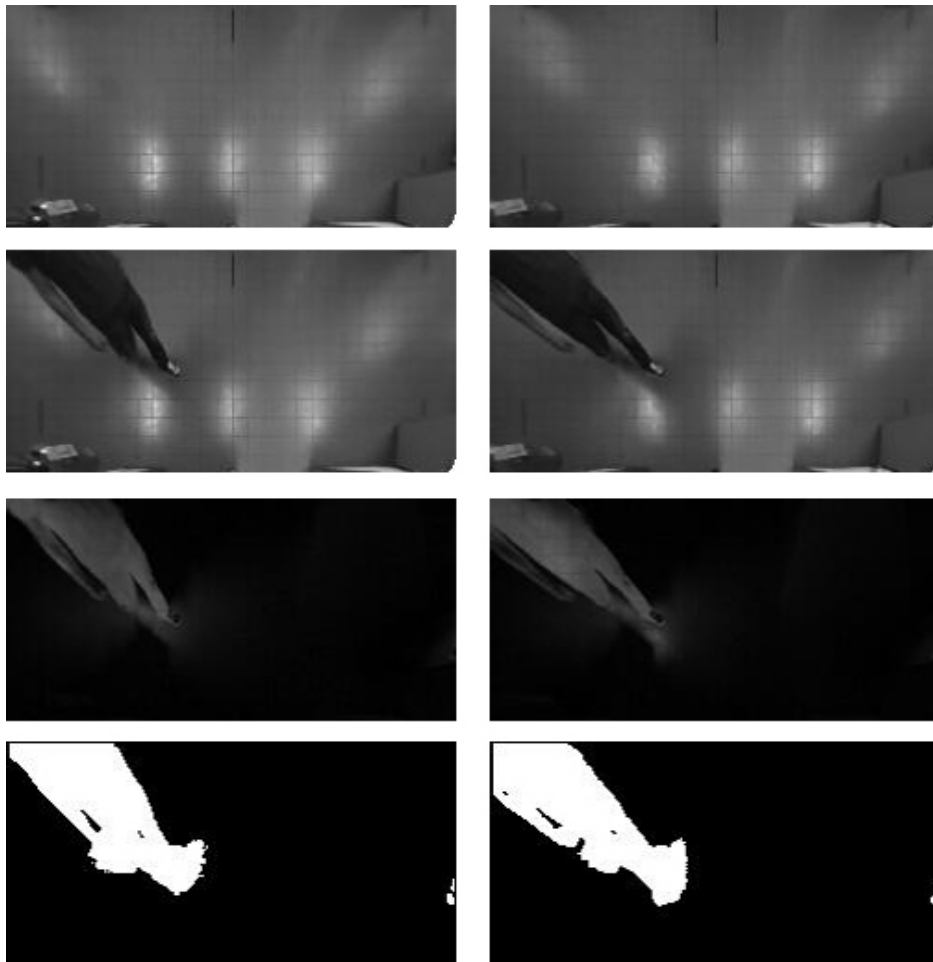


Figura 2.10: Dall'alto: immagini di background, input, differenza e maschera per la telecamera sinistra e destra.

di questa operazione è una lista di *blob* che rappresentano gli ostacoli nella scena corrente (fig. 2.11.a). Attraverso il calcolo della trasformazione tridimensionale, per ogni *blob* si ottiene la dimensione fisica e la posizione del mondo dell'ostacolo corrispondente. Gli ostacoli con una dimensione ridotta e con un'altezza minore di 20cm sono scartati poiché non dovrebbero in teoria rappresentare dei pedoni. Successivamente viene implementata un'analisi temporale in cui sono seguiti gli ostacoli che si muovono nella scena utilizzando come vincolo di associazione la posizione e la dimensione nel mondo dei relativi *blob*. Sono utilizzati vincoli sulla velocità degli ostacoli e sul mutamento delle dimensioni dovuti a possibili cambiamenti di inquadratura. La traiettoria di ogni pedone è quindi calcolata attraverso una regressione lineare [48] delle posizioni passate (fig. 2.11.b). Il sistema segnala l'apertura delle porte solo se la traiettoria così calcolata è diretta verso la porta per due frame successivi o se il pedone si trova nell'area di prossimità della porta. Per ragioni di sicurezza questa piccola area di prossimità localizzata vicino alla porta garantisce l'apertura della stessa se qualsiasi ostacolo viene individuato al suo interno, indipendentemente dalla sua direzione di movimento.

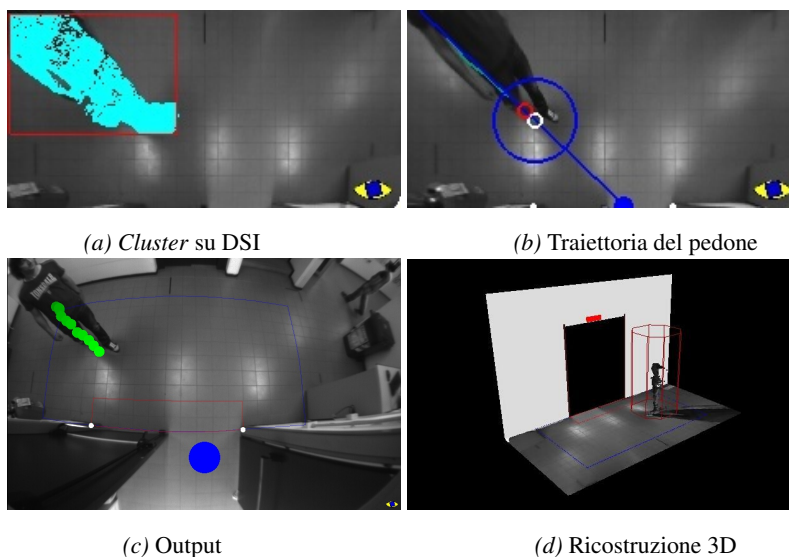


Figura 2.11: Output del sistema.

2.4 Risultati e conclusioni

Il sistema è stato testato in condizioni reali per misurare le sue capacità e valutarne le prestazioni su di computer industriale con un processore Intel Core 2 Duo 2.5GHz e 2Gb DDR di RAM

Il sistema prevede l'installazione di due telecamere solidali fra di loro con una risoluzione video di 752×480 pixel. Per permettere un'esecuzione real time a 10 HZ dell'algoritmo, le immagini sono state ridotte, per mezzo della trasformazione in immagini IMP, ad una risoluzione di 240×120 pixel. Le immagini sono stati analizzate in laboratorio ed è stato possibile testare l'algoritmo in molte situazioni.

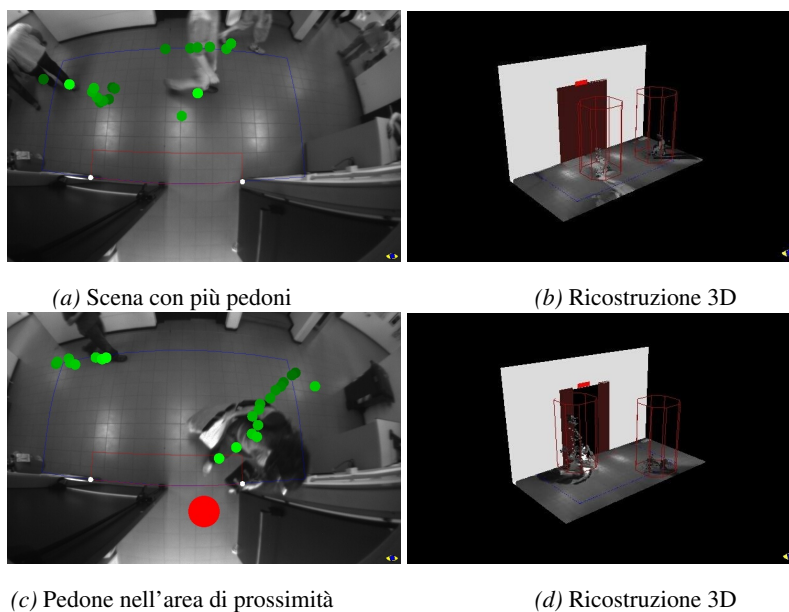


Figura 2.12: Output dell'algoritmo in due diverse situazioni: (a) Esempio di funzionamento del sistema con la presenza di più pedoni e relativa ricostruzione 3D (b); (b) Esempio di pedone nell'area di prossimità e relativa ricostruzione 3D (d).

Il sistema ha dimostrato di lavorare correttamente e fornire ottimi risultati in diverse situazioni (fig. 2.12). La situazione più critica che influenza le prestazioni del

sistema è data dalla presenza di un eccessivo numero di pedoni nella scena. In questo caso, la traiettoria degli ostacoli non è sempre calcolata correttamente, per la difficoltà nell'associazione temporale degli ostacoli, e genera qualche falso positivo che porta ad aperture della porte non volute. Tuttavia, il sistema assicura sempre l'apertura quando una persona vuole realmente attraversare la soglia della porta grazie all'implementazione dell'area di prossimità (fig. 2.12.c).

Le prestazioni del sistema, visibili nella tabella 2.1, sono state ottenute mettendo a confronto gli output del sistema di visione implementato e la simulazione di due sistemi classici con sensori ad infrarosso. Per questo tipo di sensore si è pensato di segnalare un'effettiva apertura della porta nel caso in cui un pedone, riconosciuto tramite l'algoritmo di visione, si trovi all'interno di due aree, rispettivamente con una larghezza dalla soglia della porta di 0.5 metri e 1.5 metri

Algoritmo	Errore Aperture [%]	Errore Passaggi [%]	Tempo di reazione medio [s]
Sistema proposto	15.34 %	7.02 %	0.53 s
Simulazione sensore IR 0.5 m	13.48 %	6.17 %	1.24 s
Simulazione sensore IR 1.5 m	65.11 %	29.78 %	0.56 s

Tabella 2.1: Prestazioni del sistema.

I dati sono stati ottenuti analizzando una sequenza di un intero giorno registrata presso un'uscita del dipartimento di Ingegneria dell'Informazione dell'Università di Parma. Nella tabella 2.1 sono evidenziati gli errori del sistema nelle aperture delle porte segnalate, rispetto al *ground truth* degli effettivi passaggi dei pedoni all'interno della porta; ottenuta tramite la formula:

$$ErroreAperture = 100(AperturaSegnalate - ApertureReali)/ApertureReali.$$

Per quanto riguarda gli errori di apertura segnalata rispetto al *ground truth* dei pedoni di solo passaggio, cioè quelli che passano vicino alla porta ma non l'attraver-

sano, si è considerata la seguente formula:

$$ErrorePassaggi = 100(AperturaSegnalate - ApertureReali) / Passaggi.$$

Analizzando i dati si osserva che il sistema ha un funzionamento simile ad un sensore ad infrarosso con una zona di apertura molto piccola, solo $0.5m$. Il sistema di visione proposto apporta però un vantaggio nel tempo di apertura delle porte stesse, in quanto l'apertura viene segnalata prima dell'effettiva uscita della persona riducendo il fastidio di dover attendere qualche istante davanti alla porta prima che questa si apra. Aumentando infatti la larghezza della zona controllata dal sensore infrarosso ad $1.5m$, si riduce ad un livello simile il tempo di attesa, ma con un degrado notevole delle prestazioni sulle aperture non volute, più del 50% di errori rispetto al caso precedente. Anche se i dati dei sistemi infrarossi sono stati simulati si capiscono le potenzialità del sistema che è stato sviluppato.

Il sistema presentato ha inoltre introdotto un grande vantaggio rispetto ad un classico sistema di visione stereoscopica, infatti l'applicazione della disparità su immagini IPM richiede un'installazione e una calibrazione semplice. La LUT per la rimozione dell'effetto di prospettiva e della distorsione della lente viene calcolata con una semplice griglia posizionata nell'area d'interesse durante l'installazione fisica del sistema. Questa area è facilmente regolabile attraverso parametri software e permette l'utilizzo e l'adattabilità del sistema a diverse porte e ambienti.

Il sistema è stato brevettato [13] per tutelare quest'approccio e poter successivamente trasformarlo in un prodotto industriale.

Capitolo 3

Sistema di individuazione di ostacoli in cantieri stradali

La sicurezza sul lavoro nei cantieri è un problema di notevole rilevanza e ultimamente ampiamente discusso. Il pericolo a cui i lavoratori sono esposti aumenta notevolmente quando si lavora a diretto contatto con automezzi di grandi dimensioni in movimento. Le cause che provocano incidenti possono essere legate a diversi fattori riconducibili alle tre entità uomo-macchina-ambiente. Alcune delle cause di questi tipi di incidenti possono essere riconducibili alla distrazione o disattenzione degli operatori, e alla ridotta visibilità che comporta la guida di questi automezzi. La visione artificiale applicata a questa problematica può fornire soluzioni utili per ridurre notevolmente il numero di incidenti sul lavoro. Infatti l'applicazione di un sistema di telecamere ad automezzi può sia ampliare il campo visivo dell'operatore sia intervenire in caso di pericolo. In questo progetto si è implementato un nuovo algoritmo di *obstacle detection* per rilevare e localizzare ostacoli presenti in cantieri stradali.

3.1 Introduzione

Il sistema è stato progettato per essere installato su un automezzo operante in cantieri stradali a basse velocità, ed è in grado di individuare ostacoli a breve distanza

e segnalare la loro presenza al guidatore tramite segnale visivo o acustico. Per fare questo sono utilizzate due telecamere stereo con ottiche grandangolari allineate in modo arbitrario: una semplice calibrazione del sistema mediante una griglia permette l'eliminazione della prospettiva e della distorsione generando nuove immagini IPM (Inverse Perspective Mapping), raffiguranti una vista dall'alto della scena.

L'analisi dell'algoritmo di generazione dell'immagine di disparità su immagini IPM trattato nel Capitolo 2.2 ha fornito l'idea per migliorare le sue prestazioni mediante la generazione di nuove immagini chiamate CIPM (Cylindric Inverse Perspective Mapping) ottenute con un'opportuna trasformazione di coordinate. Nell'immagine CIPM gli ostacoli verticali sono rimappati in modo da subire la stessa deformazione indipendentemente dalla posizione della telecamera da cui le immagini sono state generate. L'algoritmo di disparità, ha mostrato un funzionamento migliore grazie ad una ricerca dei punti omologhi più performante.

L'algoritmo di individuazione ostacoli implementato sfrutta i dati forniti dalle immagini di disparità derivanti dalle immagini CIPM per localizzare gli ostacoli con una dimensione rilevante, attraverso un'aggregazione dei punti mondo. Gli ostacoli sono poi evidenziati sull'immagine acquisita con la loro relativa distanza.

3.2 IPM Cilindrica

In questo capitolo sarà descritta l'implementazione di una nuova trasformazione alternativa all'IPM chiamata Cylindric Inverse Perspective Mapping (CIPM). Sulle immagini risultanti da questa trasformazione sarà poi definita la disparità cilindrica con un'analisi approfondita del suo utilizzo.

Nel Capitolo 2.2 si sono analizzate in dettaglio le immagini IPM e la disparità ad esse applicata e si è osservato come gli oggetti che emergono dal terreno subiscono una trasformazione differente nelle due IPM. Questo può provocare errori nell'individuazione dei punti omologhi e quindi nel calcolo della disparità. L'immagine di disparità infatti viene costruita effettuando correlazioni tra finestre appartenenti all'immagine destra e sinistra. Poiché le finestre omologhe hanno un diverso con-

tenuto informativo, non sempre è possibile individuare il *match* corretto e, quindi, l'immagine di disparità risulta rumorosa.

Prendiamo in esame il classico sistema stereo utilizzato in questa tesi, cioè due telecamere allineate lungo l'asse y , poste alla stessa altezza h , e distanti fra di loro b . Sotto queste ipotesi un oggetto nel mondo, con X e Z costanti e Y variabile, viene rimappato su di un'immagine IPM in un oggetto con X' costante e Y' che varia lungo una retta parallela all'asse y . La forma dell'oggetto non viene alterata, cambia solo la sua dimensione secondo un fattore di scala uguale per ogni oggetto. La forma dell'oggetto cambia invece se l'oggetto ha Z variabile.

Come già detto un ostacolo perpendicolare al suolo viene rimappato sull'IPM sulla retta passante per la proiezione del *pin-hole* e la base dell'ostacolo, secondo l'equazione 2.4.

Si prendano le due rette formate dallo stesso oggetto, una sull'immagine IPM destra e l'altra sull'immagine IPM sinistra. Queste due rette si incontreranno nel punto $P(X, Y)$, cioè il punto di contatto tra l'ostacolo e il suolo. Le coordinate di P saranno le stesse sia nell'IPM destra sia in quella sinistra. All'aumentare dell'altezza Z dell'oggetto, la distanza fra queste rette aumenterà.

In figura 3.1, è rappresentata la situazione analizzata: l'angolo α generato dalle suddette rette è uguale alla somma degli angoli β e γ .

Essendo:

$$\beta = \arctan\left(\frac{\frac{b}{2} - e + Y}{X - x_0}\right)$$

$$\gamma = \arctan\left(\frac{\frac{b}{2} + e - Y}{X - x_0}\right)$$

si ricava:

$$\tan \alpha = b \frac{X - x_0}{(X - x_0)^2 - \left(\frac{b}{2}\right)^2 + (Y - e)^2} \quad (3.1)$$

L'angolo α dipende pertanto dalla posizione dell'ostacolo, dalla *baseline* e anche dalla posizione delle due telecamere.

Lo scopo di questa sezione sarà quello di introdurre alcune trasformazioni per rimappare le immagini IPM destra e sinistra in nuove immagini in cui gli ostacoli

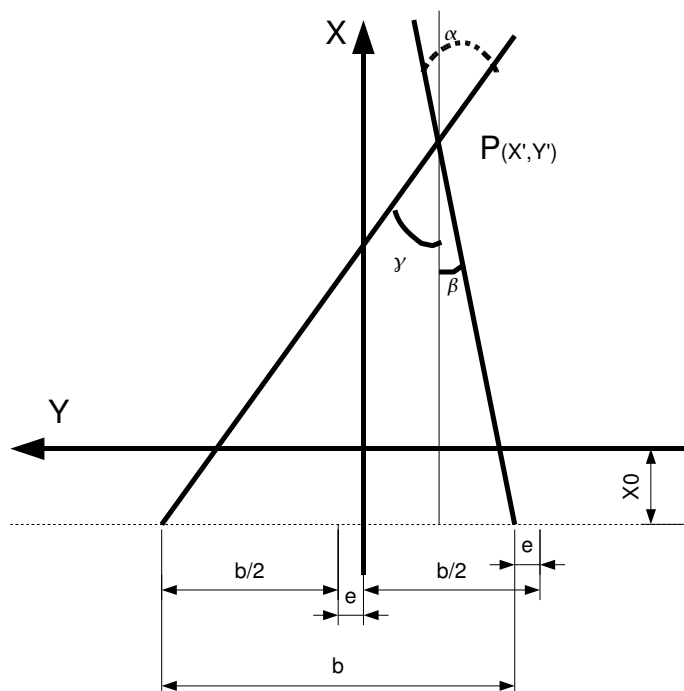


Figura 3.1: Un ostacolo puntiforme viene mappato nelle due immagini IPM su due rette con un'inclinazione diversa: α è l'angolo compreso fra di esse.

perpendicolari al terreno non abbiano più un orientamento diverso fra di loro. Questo faciliterà il calcolo dell'immagine di disparità in quanto le finestre omologhe risulteranno simili per punti appartenenti ad ostacoli verticali.

3.2.1 IPM sferica

Le immagini IPM fino ad ora analizzate adottano un sistema di riferimento cartesiano, con coordinate (X', Y') . Per la generazione delle immagini IPM sferiche si effettua una trasformazione del sistema di riferimento passando alle coordinate polari (ρ, ϑ, \cdot) .

Un generico sistema di riferimento polare, descritto nella figura 3.2, è formato da due coordinate indicate con ϑ e ρ :

- ρ indica la distanza del punto considerato dall'origine del sistema, cioè il modulo del vettore $\vec{\rho}$ che congiunge l'origine degli assi con il punto P .
- ϑ indica l'angolo che si forma tra l'asse positivo delle ascisse e il vettore $\vec{\rho}$ considerato prima.

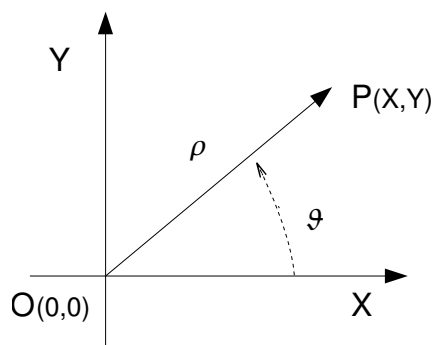


Figura 3.2: Coordinate polari.

Il passaggio dalle coordinate polari a quelle cartesiane si ottiene applicando le seguenti formule:

$$X = \rho \cos \vartheta$$

$$Y = \rho \sin \vartheta$$

È necessario adattare questa trasformazione al sistema di riferimento dell'immagine IPM effettuando un cambiamento degli assi, definendo ϑ come l'angolo fra l'asse positivo X e il vettore $\vec{\rho}$, come in figura 3.3.

$$X' = \rho \cos \vartheta$$

$$Y' = -\rho \sin \vartheta$$

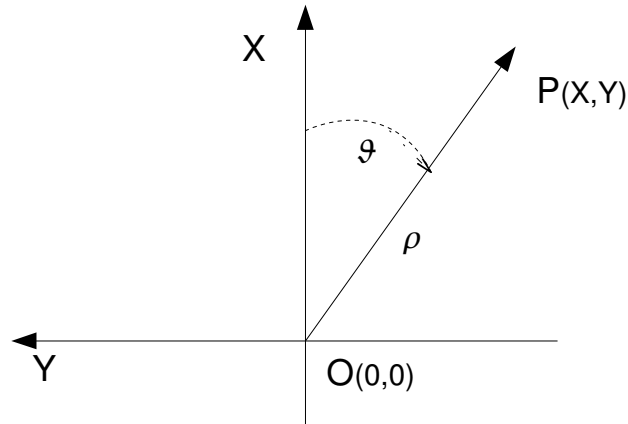


Figura 3.3: Coordinate polari dell'immagine IPM.

Per fare in modo che le rette, sulle quali l'IPM rimappa gli ostacoli verticali, risultino verticali, è necessario che l'origine del nuovo sistema di riferimento sia traslato sulla rispettiva proiezione del *pin-hole* della telecamera (x_0, y_0) :

$$\begin{aligned} X' &= \rho \cos \vartheta + x_0 \\ Y' &= -\rho \sin \vartheta + y_0 \end{aligned}$$

con $y_0 = -b_d$ per la telecamera destra e $y_0 = b_s$ per la telecamera sinistra.

Invertendo le equazioni si avrà rispettivamente sulle ascisse e sulle ordinate della nuova immagine:

$$\begin{aligned} \vartheta &= -\arctan\left(\frac{Y' - y_0}{X' - x_0}\right) \\ \rho &= \sqrt{(X' - x_0)^2 + (Y' - y_0)^2} \end{aligned}$$

con ϑ compreso nell'intervallo $[-\frac{\pi}{2}, +\frac{\pi}{2}]$ e $\rho \geq 0$.

Sostituendo i valori di X' e Y' ricavati nelle equazioni 2.2 e 2.3, si ottiene:

$$\vartheta = -\arctan\left(\frac{Y - y_0}{X - x_0}\right)$$

$$\rho = \left|\frac{z_0}{(Z - z_0)}\right| \sqrt{(X - x_0)^2 + (Y - z_0)^2}$$

Queste equazioni, essendo riferite all'IPM, sono valide solo per valori di Z inferiori a z_0 .

In figura 3.4 sono illustrate le immagini IPM destra e sinistra e le rispettive IPM sferiche. Come si può osservare nelle immagini 3.4. *b* le barre verticali dell'oggetto, ripreso dalle telecamere, risultano verticali e parallele, proprio come ci si aspettava di ottenere. Questa trasformazione, però, rimappa uno stesso punto mondo in due punti immagine con ordinata differente, in quanto ρ dipende da Y' , che è diverso nelle due immagini IPM per punti con $Z \neq 0$, e da y_0 che è la posizione della telecamera.

L'algoritmo utilizzato per generare l'immagine DSI ipotizza che i punti omologhi delle due immagini si trovino sulla stessa riga in entrambe le immagini. Questo vincolo è necessario per poter garantire all'algoritmo un tempo di elaborazione compatibile con la modalità *real-time*. Risulta quindi evidente, che applicando una simile trasformazione, questo algoritmo non può più essere utilizzato. Si noti però, che il *match* tra due finestre delimitanti ostacoli verticali risulta migliore in quanto questi ultimi vengono deformati in egual maniera nelle due immagini.

Come si può osservare nelle immagini in figura 3.6, generate dalle immagini IPM (fig. 3.5), la disparità \mathbf{d} di un ostacolo verticale risulta costante su tutto l'oggetto rimappato. Lo stesso ostacolo presenterà nelle due immagini una diversa posizione e una diversa altezza, ($\overline{AB} \neq \overline{CD}$), che dipendono dalla posizione dell'ostacolo e delle telecamere rispetto all'origine del sistema di riferimento del mondo.

3.2.2 CIPM (Cylindric Inverse Perspective Mapping)

La trasformazione in coordinate polari proposta in precedenza crea notevoli problemi nell'utilizzo dell'algoritmo di disparità. Si è quindi cercato di trovare una trasformazione di coordinate che, come le coordinate polari, rimappi un oggetto verticale su rette verticali, ma che mantenga uguale in entrambe le immagini la coordinata X .

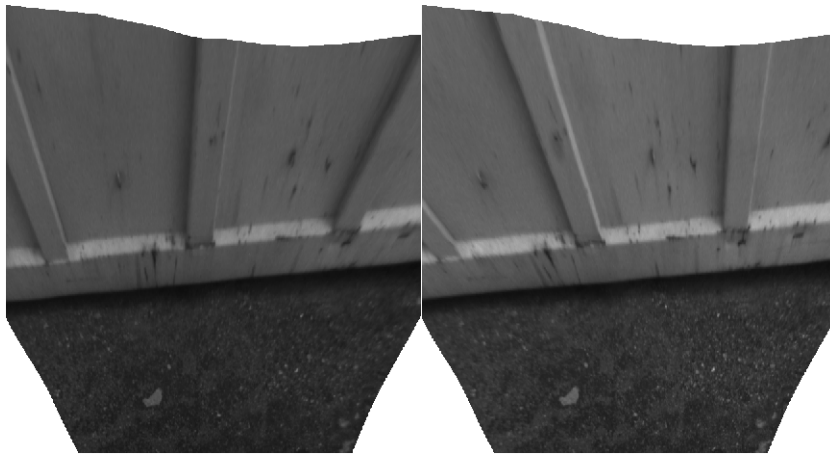
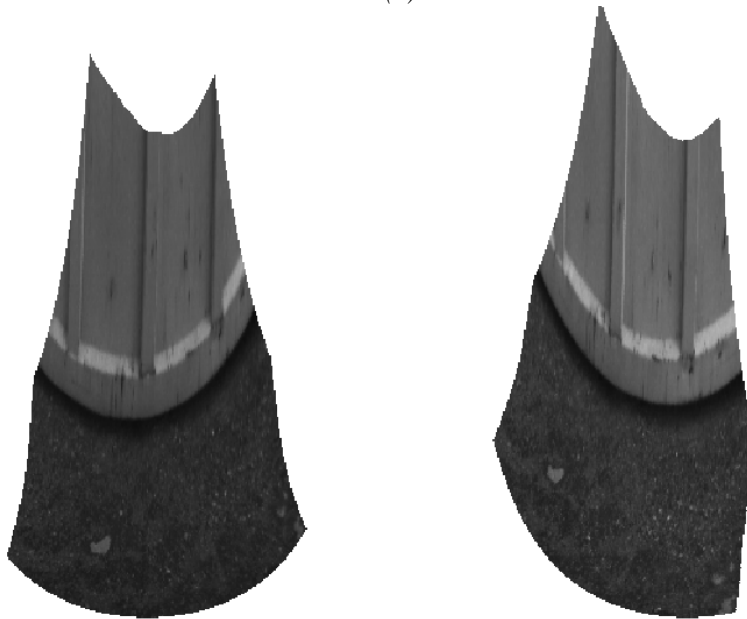
*(a)**(b)*

Figura 3.4: IPM: sinistra e destra (a), IPM sferica: sinistra e destra (b).

Questo permetterebbe l'utilizzo corretto dell'algoritmo di disparità per la creazione dell'immagine DSI.

Per ottenere questo risultato, si è pensato di utilizzare una trasformazione opportuna, che per le sue proprietà può essere chiamata quasi cilindrica. La trasformazione è così definita:

$$\vartheta = -\arctan\left(\frac{Y' - y_0}{X' - x_0}\right) \quad (3.2)$$
$$X'' = X'$$

dove ϑ rappresenta l'angolo formato rispetto l'asse X , mentre X'' la distanza dalla telecamera lungo l'asse X .

La nuova immagine cilindrica dell'IPM, chiamata Cylindric Inverse Perspective Mapping (CIPM), avrà sulle ascisse la variabile ϑ , che potrà assumere valori compresi nell'intervallo $[-\frac{\pi}{2}, +\frac{\pi}{2}]$ e sulle ordinate la variabile X'' .

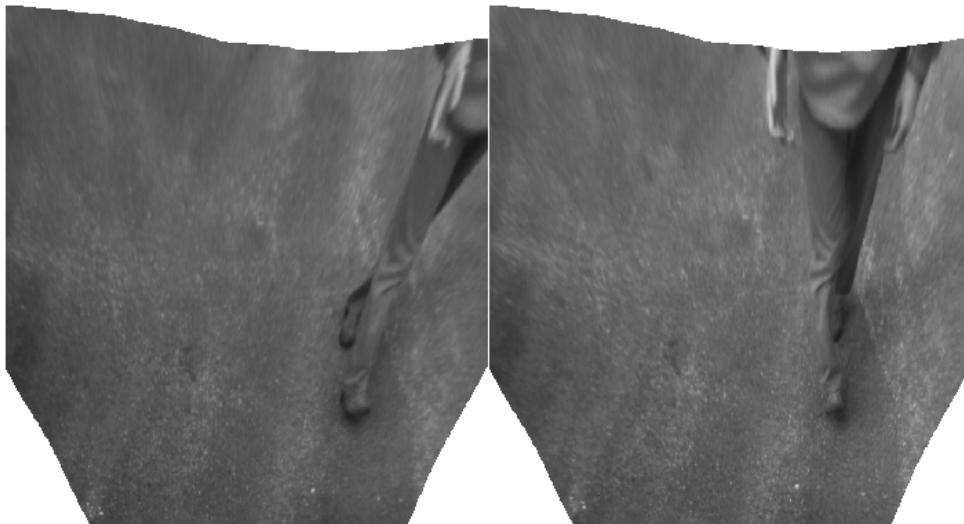


Figura 3.5: IPM di un ostacolo verticale: sinistra e destra.

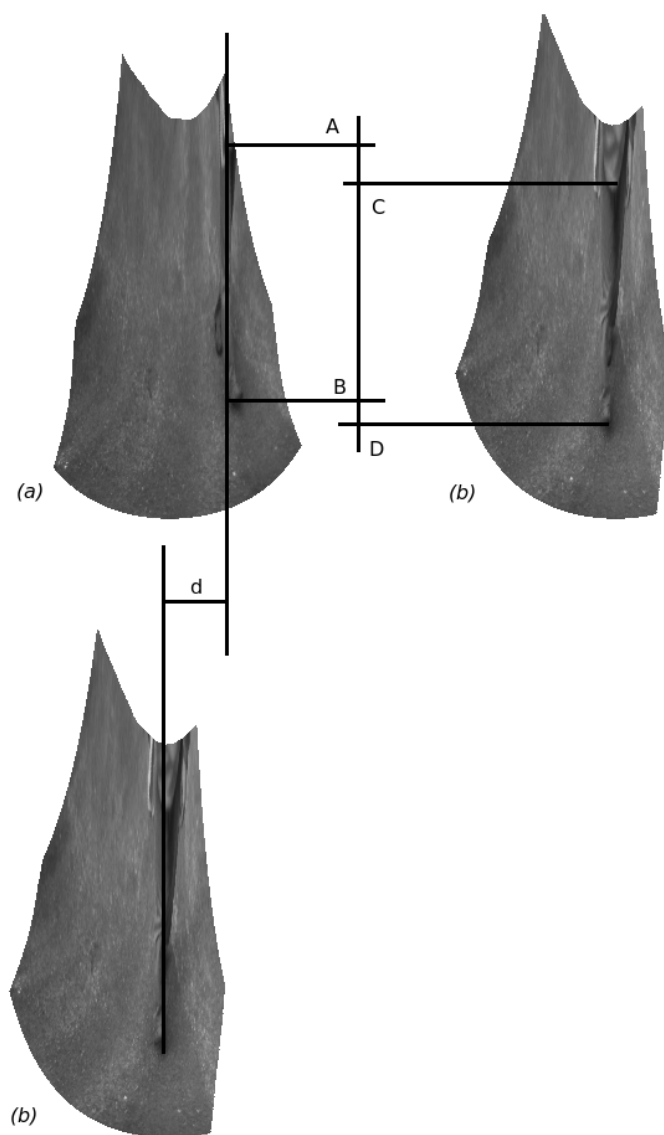


Figura 3.6: IPM sferica di un ostacolo verticale: sinistra (a) e destra (b).

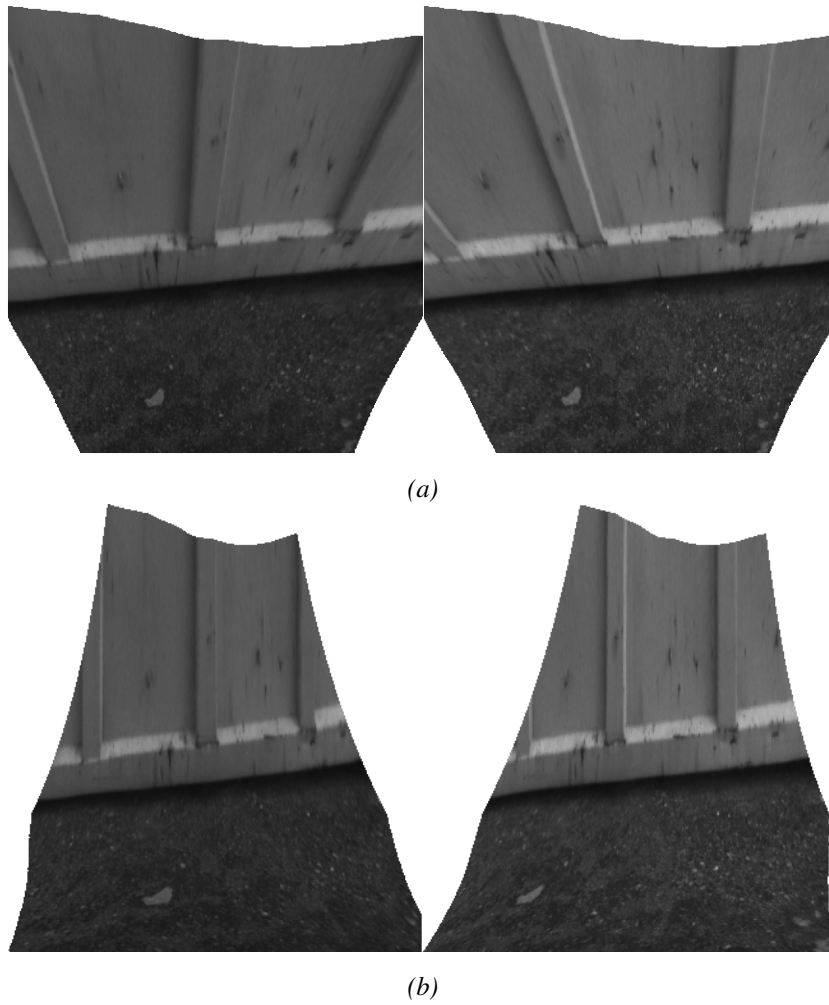


Figura 3.7: Immagini IPM: sinistra e destra (a); Immagini IPM cilindrica: sinistra e destra (b).

Le immagini 3.7, mostrano come un ostacolo verticale viene *raddrizzato* allo stesso modo in entrambe le immagini CIPM, fig 3.7.b. I punti omologhi in queste immagini si trovano come supposto sulla stessa riga dell'immagine come accade nelle immagini IPM.

3.2.3 La generazione di immagini CIPM

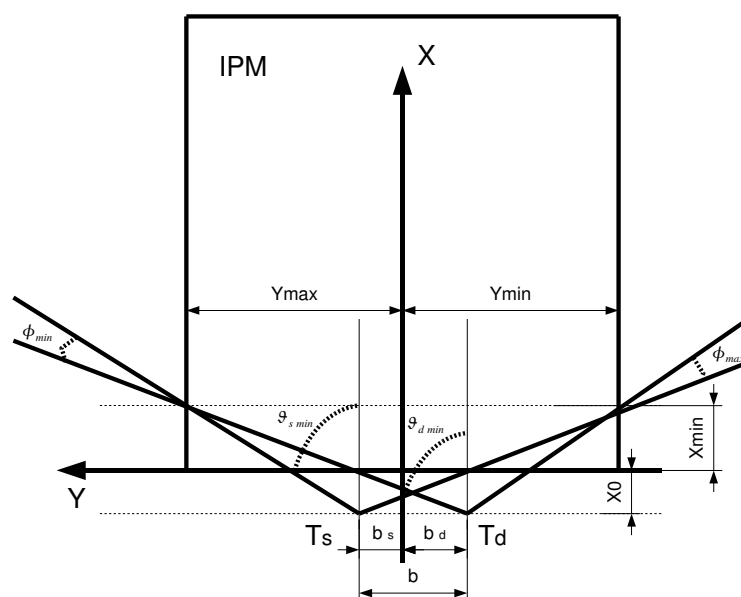


Figura 3.8: Analisi degli angoli ϕ_{min} , ϕ_{max} delle due immagini CIPM.

Le immagini CIPM possono essere costruite semplicemente applicando le formule della trasformazione appena descritta ai punti delle immagini IPM. Ad ogni pixel $P(\vartheta, X'')$ dell'immagine CIPM, viene assegnato il valore del pixel nell'immagine

IPM corrispondente:

$$Y' = y_0 - (X' - x_0) \tan(\vartheta) \quad (3.3)$$

$$X' = X''$$

Se le coordinate del punto IPM sono fuori dall'immagine allora al pixel CIPM viene assegnato un valore nullo. Per alcuni valori iniziali e finali di ϑ , le immagini CIPM possono presentare zone senza nessuna informazione perché esterni all'immagine IPM. Questo permette di ridurre il range di variazione di ϑ in un intervallo $[\vartheta_{min}, \vartheta_{max}]$. Le immagini CIPM generate dall'immagine IPM destra e dall'immagine IPM sinistra avranno angoli massimi e minimi diversi perché diversa è la posizione del *pin-hole* delle due telecamere.

I valori di questi angoli possono essere ricavati dalla conoscenza dei punti limite sinistro Y_{max} e destro Y_{min} delle immagini IPM, fig 3.8, mediante le seguenti formule:

$$\vartheta_{d,min} = -\arctan\left(\frac{Y_{max} + b_d}{X_{min} - x_0}\right)$$

$$\vartheta_{s,min} = -\arctan\left(\frac{Y_{max} - b_s}{X_{min} - x_0}\right)$$

$$\vartheta_{d,max} = -\arctan\left(\frac{Y_{min} + b_d}{X_{min} - x_0}\right)$$

$$\vartheta_{s,max} = -\arctan\left(\frac{Y_{min} - b_s}{X_{min} - x_0}\right)$$

Dovendo generare immagini della stessa larghezza W , si definisce un passo S :

$$S = \frac{\max[(\vartheta_{d,max} - \vartheta_{d,min}), (\vartheta_{s,max} - \vartheta_{s,min})]}{W - 1}$$

Il passo S definisce come quantizzare l'immagine sorgente.

È possibile, dato il valore quantizzato ϑ' in pixel, risalire al valore reale ϑ :

$$\vartheta_d = \vartheta_{d,min} + S\vartheta'_d \quad (3.4)$$

$$\vartheta_s = \vartheta_{s,min} + S\vartheta'_s$$

La differenza fra i due intervalli di ϑ da cui deriva S dipende dalla posizione della *baseline* delle telecamere rispetto al sistema mondo. Se i due intervalli di ϑ sono uguali, ovvero quando il centro della *baseline* coincide con l'origine dell'IPM, le immagini CIPM risultano rimappate in immagini della stessa larghezza. Se questo non si verifica, avendo scelto il passo derivante dall'intervallo ϑ massimo, un'immagine avrà una piccola zona priva di informazione nel lato destro dell'immagine.

Da come si può intuire dalla costruzione delle stesse immagini CIPM, la zona nella parte sinistra eliminata nell'immagine destra risulterà più piccola rispetto a quella dell'immagine sinistra, fig 3.8. Per questo le immagini sinistre risulteranno traslate verso sinistra rispetto a quelle destra di un valore ϕ_{min} da considerare nella trattazione successiva di queste immagini.

Il valore dell'angolo ϕ_{min} si può ricavare dalla differenza fra l'angolo minimo nell'immagine sinistra e destra:

$$\phi_{min} = \vartheta_{s,min} - \vartheta_{d,min}$$

3.2.4 Disparità su immagini CIPM

Si prenda una coppia di immagini CIPM definite sull'intervallo $\vartheta = [-\frac{\pi}{2}, +\frac{\pi}{2}]$ ricavate dalle immagini IPM di una coppia di telecamere stereo. Possiamo definire la disparità cilindrica fra due punti omologhi come la distanza fra la posizione nell'immagine sinistra e quella destra.

$$\Delta_c = \vartheta_s - \vartheta_d \quad (3.5)$$

Se si effettua un'eliminazione delle parti con contenuto informativo nullo dell'immagine mediante il restringimento dell'intervallo ϑ , dalla definizione 3.4, si ottiene:

$$\begin{aligned} \Delta_c &= (\vartheta_{s,min} + S\vartheta'_s) - (\vartheta_{d,min} + S\vartheta'_d) \\ \Delta_c &= S(\vartheta'_s - \vartheta'_d) + \phi_{min} \end{aligned} \quad (3.6)$$

Quindi la disparità cilindrica a meno di un fattore di scale S e di un *offset* risulta comunque sempre dipendente dalla differenza fra due punti omologhi.

Prendiamo ora l'equazione 3.2, questa può essere riscritta sostituendo a Y' e X' i loro valori presi dalle equazioni 2.2 e 2.3 ottenendo:

$$\vartheta = -\arctan\left(\frac{Y-y_0}{X-x_0}\right) \quad (3.7)$$

Considerando che, per ogni punto dell'immagine CIPM destra, $X_d'' = X_s''$, si può affermare che:

La disparità cilindrica di un punto mondo di coordinate (X, Y, Z) proiettato su due immagini CIPM è pari a:

$$\Delta_c = \vartheta_s - \vartheta_d = \arctan\left(\frac{Y+b_d}{X-x_0}\right) - \arctan\left(\frac{Y-b_s}{X-x_0}\right) \quad (3.8)$$

Calcoliamo la tangente di Δ_c applicando la formula della tangente della differenza di due angoli:

$$\tan\Delta_c = \frac{\tan\vartheta_s - \tan\vartheta_d}{1 + \tan\vartheta_d \tan\vartheta_s} = \frac{b(X-x_0)}{(X-x_0)^2 + (Y+b_d)(Y-b_s)} \quad (3.9)$$

che risulta uguale all'equazione della tangente dell'angolo α calcolato nella 3.1, in quanto per costruzione $\alpha = \Delta_c$.

Facciamo qualche considerazione sulla disparità:

- La disparità cilindrica è sempre positiva. L'angolo della retta dell'IPM destra su cui viene rimappato un ostacolo è infatti sempre maggiore rispetto a quello della retta sull'IPM sinistra.
- La disparità cilindrica dipende solo dalle coordinate (X, Y) del punto nel mondo.
- Ostacoli perpendicolari al terreno avranno la stessa disparità cilindrica, fig.3.9. Infatti la coordinata Z del punto mondo non appare nella formula 3.9.
- La disparità cilindrica di un punto del suolo non sarà più uguale a zero, fig.3.9, come accadeva per la disparità sull'immagine IPM.

- La disparità cilindrica dipende dalla posizione (x_0, y_0) delle due telecamere ma non dalla loro altezza z_0 .

3.2.5 Ricostruzione 3D dalla disparità cilindrica

In modo analogo alla disparità sull'immagine IPM, la conoscenza della disparità dei punti di un'immagine CIPM permette la ricostruzione tridimensionale della scena. Le formule per ricavare le coordinate mondo di un punto nota Δ_c , possono essere ricavate utilizzando le formule della disparità sull'IPM. Per fare questo si deve trovare l'equazione che lega la disparità IPM a quella cilindrica appena descritta.

Riscriviamo l'equazione 3.3 per le immagini destra e sinistra:

$$Y'_d = -\tan(\vartheta_d)(X'_d - x_{0d}) - b_d \quad (3.10)$$

$$Y'_s = -\tan(\vartheta_s)(X'_s - x_{0s}) + b_s \quad (3.11)$$

Osservando che in un'immagine IPM: $X'_s = X'_d = X'$, $x_{0s} = x_{0d} = x_0$ e $b_d + b_s = b$, è possibile esprimere la disparità calcolata sull'immagine IPM, eq.2.5, con la seguente equazione:

$$\Delta = Y'_d - Y'_s = (\tan(\vartheta_s) - \tan(\vartheta_d))(X' - x_0) + b$$

Possiamo quindi sostituire a ϑ_s il valore ricavato dall'equazione della disparità cilindrica 3.8:

$$\Delta = Y'_d - Y'_s = (\tan(\Delta_c + \vartheta_d) - \tan(\vartheta_d))(X' - x_0) + b \quad (3.12)$$

La disparità Δ su immagini IPM, dipende quindi, dalla disparità cilindrica, dalla coordinata (ϑ, X'') del punto dell'immagine cilindrica e dalla *baseline* e posizione x_0 delle telecamere. È possibile quindi risalire alle coordinate mondo applicando le formule 2.6 e 2.7 sostituendo a Δ il valore della 3.12:

$$Z = h \frac{\Delta}{\Delta + b} = h \frac{(\tan(\Delta_c + \vartheta_d) - \tan(\vartheta_d))(X'' - x_0) + b}{(\tan(\Delta_c + \vartheta_d) - \tan(\vartheta_d))(X'' - x_0) + 2b} \quad (3.13)$$

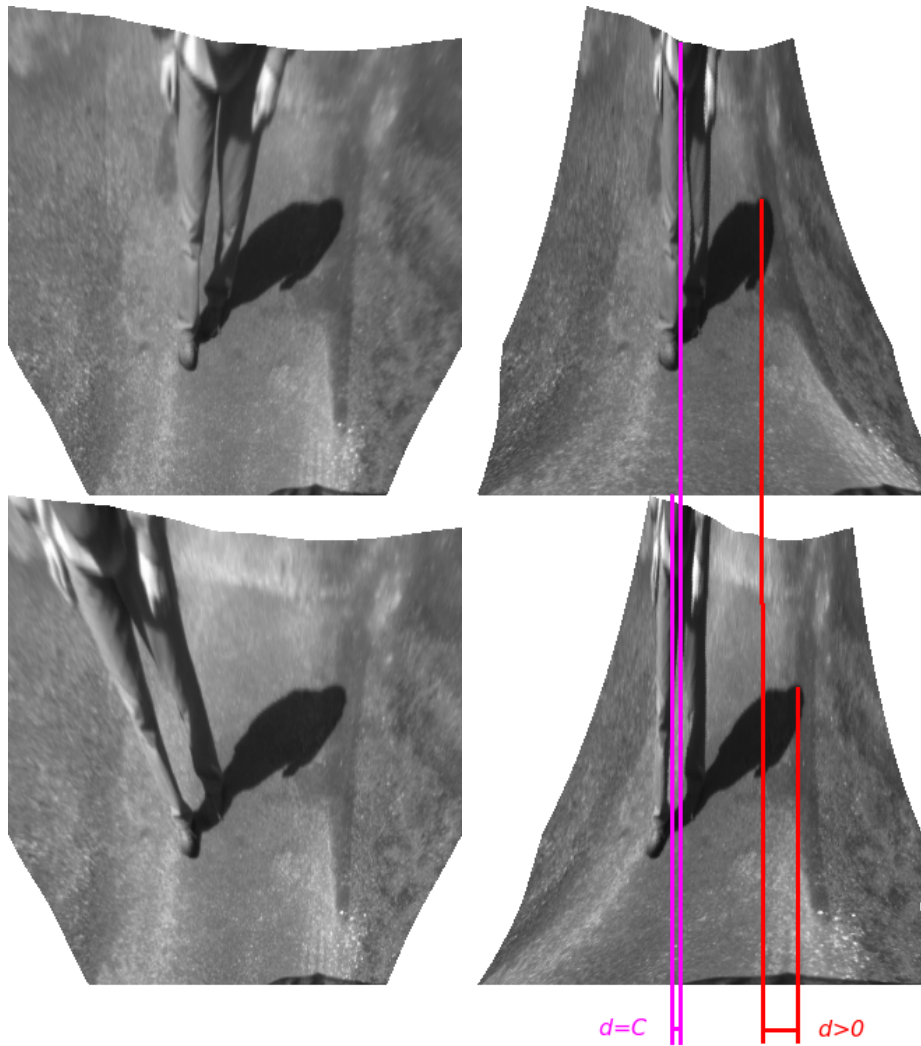


Figura 3.9: Gli ostacoli verticali (pedone) presentano una disparità cilindrica costante; i punti appartenenti al suolo (ombre) hanno una disparità variabile.

$$X = \frac{X'b + x_0\Delta}{\Delta + b} = \frac{X''b + x_0((\tan(\Delta_c + \vartheta_d) - \tan(\vartheta_d))(X'' - x_0) + b)}{(\tan(\Delta_c + \vartheta_d) - \tan(\vartheta_d))(X'' - x_0) + 2b} \quad (3.14)$$

L'equazione della Y può essere ricavata dall'equazione 3.7 andando a sostituire a X il valore appena trovato:

$$Y = -\tan(\vartheta_d) \left(\frac{X''b + x_0((\tan(\Delta_c + \vartheta_d) - \tan(\vartheta_d))(X'' - x_0) + b)}{(\tan(\Delta_c + \vartheta_d) - \tan(\vartheta_d))(X'' - x_0) + 2b} - x_0 \right) - b_d \quad (3.15)$$

3.2.6 Immagine di disparità DSI su immagini CIPM

Il calcolo dell'immagine di disparità DSI, figura 3.10, permette, come già descritto nei capitoli precedenti, di assegnare ad ogni pixel dell'immagine destra il valore della sua disparità. La costruzione dell'immagine CIPM permette un miglior utilizzo dell'algoritmo della DSI poiché permette di uniformare la deformazione degli ostacoli che emergono dal terreno. A differenza delle immagini IPM sferiche, la ricerca dei punti omologhi può essere limitata alla stessa riga del pixel destro. A differenza dell'applicazione dell'algoritmo sulle immagini IPM, con l'applicazione su quelle CIPM il confronto fra due finestre in entrambe le immagini è più preciso per quanto riguarda gli ostacoli verticali, mentre si ha un peggioramento nel confronto dei pixel appartenenti al suolo, ad esempio un'ombra.

Facciamo ora alcune considerazioni sull'immagine di disparità CIPM:

- La disparità cilindrica di un pixel è sempre positiva, ma avendo effettuato una traslazione ϕ delle immagini CIPM, l'immagine DSI può assumere anche valori negativi.
- Il suolo ha disparità diversa per ogni suo punto.
- Se la funzione di correlazione fra finestre non supera determinate soglie, al pixel viene assegnato un valore standard di disparità, per indicare che il pixel non ha riscontri nell'altra immagine.

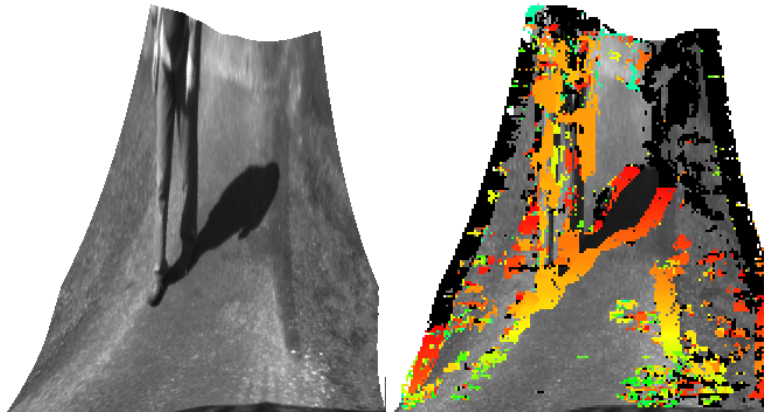


Figura 3.10: Immagine CIPM destra e DSI.

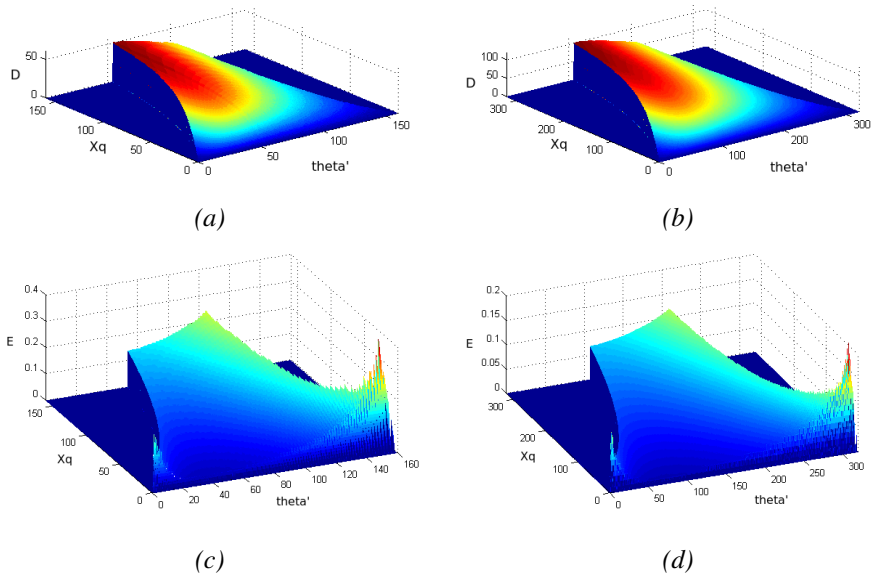


Figura 3.11: Range di disparità ed errore.

- La disparità di un oggetto, che nel mondo è perpendicolare al suolo, rimane uguale per tutto l'oggetto.
- Anche in questo caso le dimensioni delle immagini CIPM sono un parametro che influenza notevolmente i tempi di calcolo della DSI. I confronti per la ricerca dei punti omologhi e quindi anche il tempo dell'elaborazione, aumentano con un ordine cubico rispetto alle dimensioni dell'immagine.
- Per diminuire il tempo di elaborazione è possibile far effettuare la ricerca del punto omologo solo entro un determinato *range* di ricerca. Ogni punto dell'immagine CIPM infatti potrà avere una disparità massima e una minima.

Calcolo *range* di ricerca DSI

Per definire il range di ricerca dell'immagine DSI dobbiamo calcolare i valori massimi e minimi che la disparità cilindrica può assumere. Consideriamo un generico punto P' di coordinate (X', Y') sull'immagine IPM destra. Questo punto può essere la rimappatura di un punto mondo P appartenente al suolo e quindi di coordinate $(X, Y, 0)$, oppure può essere la rimappatura di punto di un ostacolo verticale, cioè con $Z \neq 0$. La possibile base di questo ostacolo si può trovare sulla retta definita dall'equazione 2.4.

Avendo definito la disparità cilindrica come uguale all'angolo α formato da questa retta e da quella corrispondente sull'immagine IPM sinistra, possiamo affermare, con semplici osservazioni geometriche sulla figura 3.1, che questo angolo è maggiore nel punto con una coordinata X' minore possibile. Quindi la disparità cilindrica massima di un un generico punto su un'immagine IPM è data dalla formula 3.9 applicata in un punto con X uguale al valore minimo X_{min} in cui la base dell'ostacolo si può trovare:

$$\Delta_{c,max} = \arctan \left(\frac{b(X_{min} - x_0)}{(X_{min} - x_0)^2 + (Y_{min} + b_d)(Y_{min} - b_s)} \right) \quad (3.16)$$

Essendo costante la disparità di un oggetto verticale sull'IPM cilindrica, la disparità massima per ogni colonna sarà costante e il suo valore può essere calcolato a

partire dalle equazioni 3.16 e 3.10

$$\Delta_{c,max} = \arctan\left(\frac{b}{\tan^2(\vartheta_d)(X_{min} - x_0) + b \tan(\vartheta_d) + (X_{min} - x_0)}\right) \quad (3.17)$$

La disparità massima non dipende dalla coordinata X'' del punto ma solo dalla coordinata ϑ_d ed essa, da semplici osservazioni geometriche sulla figura 3.1, assume il valore massimo nel punto IPM con coordinata centrale rispetto alla *baseline*:

$$Y = \frac{b_s - b_d}{2}$$

da cui:

$$\Delta_{c,max} = \arctan\left(\frac{b}{(X_{min} - x_0)}\right) \quad (3.18)$$

Abbiamo ricavato la disparità massima che ogni punto nell'immagine CIPM può assumere, ma a differenza della disparità sull'immagine IPM, la disparità minima possibile non sarà uguale a zero ma avrà valore diverso per ogni pixel dell'immagine. Riprendiamo l'immagine IPM destra del caso precedente e ipotizziamo ora che il punto $P(X', Y')$ sia la proiezione del punto mondo P con $Z = 0$. Calcoliamo ora la disparità cilindrica in questo punto dalla formula 3.16, questa sarà anche la disparità minima che il punto potrà avere essendo ad altezza zero dal suolo.

$$\Delta_{c,min} = \arctan\left(\frac{b}{\tan^2(\vartheta_d)(X'' - x_0) + b \tan(\vartheta_d) + (X'' - x_0)}\right) \quad (3.19)$$

Se si fissa la riga X'' allora la disparità minima risulta dipendere solo dalla posizione ϑ_d . Cerchiamo quindi per ogni riga della CIPM il valore minimo che la disparità può assumere. Questo valore corrisponderà alla disparità cilindrica calcolata nella posizione limite sinistra o destra dell'immagine IPM. Il valore minimo della disparità cilindrica sulla riga X'' si avrà quindi per:

$$\vartheta_1 = -\arctan\left(\frac{Y_{max} + b_d}{X'' - x_0}\right) \quad \vartheta_2 = -\arctan\left(\frac{Y_{min} + b_d}{X'' - x_0}\right)$$

$$\vartheta_{min} = \begin{cases} \vartheta_1 & \text{se } |\vartheta_1| \geq |\vartheta_2| \\ \vartheta_2 & \text{se } |\vartheta_1| < |\vartheta_2| \end{cases}$$

Calcolo errore su immagini CIPM

La generazione delle immagini CIPM impone una quantizzazione diversa dell'immagine IPM. Si vuole ora calcolare come questa quantizzazione influenzi l'errore associato al valore di disparità di ogni pixel dell'immagine CIPM. Per fare questo si deve analizzare l'immagine CIPM e calcolare per ogni suo punto il valore massimo e minimo che la disparità cilindrica in pixel (Δ') può assumere e calcolarne la differenza $D_{\Delta'}$. Dall'equazioni 3.6, 3.17 e 3.19 imponendo $b_d = b_s$ e $x_0 = 0$ si ricava:

$$\Delta'_{c,max} = \frac{\arctan\left(\frac{b}{\tan^2(\vartheta_d)X_{min} + b \tan(\vartheta_d) + X_{min}}\right) - \phi_{min}}{S}$$

$$\Delta'_{c,min} = \frac{\arctan\left(\frac{b}{\tan^2(\vartheta_d)X'' + b \tan(\vartheta_d) + X''}\right) - \phi_{min}}{S}$$

quindi:

$$D_{\Delta'} = [\Delta'_{c,max}] - [\Delta'_{c,min}] \quad (3.20)$$

dove X'' è la coordinata verticale dell'immagine CIPM. Questa coordinata, alla variazione di ϑ_d , assume valori massimi diversi sull'immagine CIPM. Si definisce quindi la funzione che permette di calcolare X''_{max} in funzione di ϑ_d :

$$X''_{max} = \begin{cases} X_{max} & \text{se } -\arctan\left(\frac{Y_{max}+b_d}{X_{max}}\right) \leq \vartheta_d \leq -\arctan\left(\frac{Y_{min}+b_d}{X_{max}}\right) \\ -\frac{Y_{max}+b_d}{\tan(\vartheta_d)} & \text{se } \vartheta_d < -\arctan\left(\frac{Y_{max}+b_d}{X_{max}}\right) \\ -\frac{Y_{min}+b_d}{\tan(\vartheta_d)} & \text{se } \vartheta_d > -\arctan\left(\frac{Y_{min}+b_d}{X_{max}}\right) \end{cases}$$

Sostituiamo ora nell'equazione 3.20 la variabile ϑ_d con il suo valore in pixel ϑ'_d , definito dall'equazione 3.4, e la variabile X'' con il suo valore quantizzato in pixel ricavato dall'equazione seguente:

$$X'' = X_{min} + \frac{X_{max} - X_{min}}{H - 1} X_q$$

con H uguale all'altezza dell'immagine CIPM.

Si è così ottenuta un'equazione che permette di calcolare il valore dell'intervallo di disparità possibile D_{Δ_c} di ciascun pixel dell'immagine in funzione della sua posizione quantizzata ϑ'_d, X_q .

Definiamo ora con E l'errore medio sulla coordinata mondo X relativa ad un ogni pixel dell'immagine:

$$E = \frac{X'' - X_{min}}{D_{\Delta_c}} \quad (3.21)$$

Questo errore dipende dalle risoluzione dell'immagine. E rappresenta quindi la quantizzazione della distanza X dalla telecamera, associata alla variazione di un'unità della disparità.

Nei grafici 3.11, sono rappresentati il range di disparità D_{Δ_c} (3.11.a e 3.11.b) e l'errore E (3.11.c e 3.11.d) dell'immagine CIPM destra di un sistema stereo con queste specifiche:

$$X_{min} = 1 \text{ m}, X_{max} = 8 \text{ m}, Y_{min} = -4 \text{ m}, Y_{max} = 4 \text{ m}, x_0 = 0, b_d = b_s = 0.5 \text{ m}$$

I grafici 3.11.a e 3.11.c si riferiscono ad immagini di risoluzione 160×160 mentre 3.11.b e 3.11.d ad immagini di dimensioni doppia, 320×320 .

Si noti che l'intervallo di disparità quantizzata aumenta all'aumentare della X dell'immagine ma in modo non uniforme su tutte le colonne ϑ' . Il valore del *range* di disparità dipende in modo lineare dalle dimensioni delle immagini CIPM scelte. Le immagini 320×320 hanno infatti D_{Δ_c} doppio rispetto a quelle di dimensione 160×160 . Conseguentemente l'errore E diminuisce in modo lineare all'aumentare delle dimensioni delle immagini. La funzione E avrà i suoi picchi sui quattro angoli dell'immagine CIPM, dove quindi si avrà meno precisione nel calcolo della distanza dei punti, e il minimo nel pixel di coordinate $X_q = 0$, e coordinata ϑ' uguale a quella del punto in cui si ha il range di disparità D_{Δ_c} massimo.

3.3 L'algoritmo

In questo capitolo verrà descritto l'algoritmo utilizzato per l'individuazione degli ostacoli a partire dall'immagine di disparità calcolata. In figura 3.12 è rappresentato

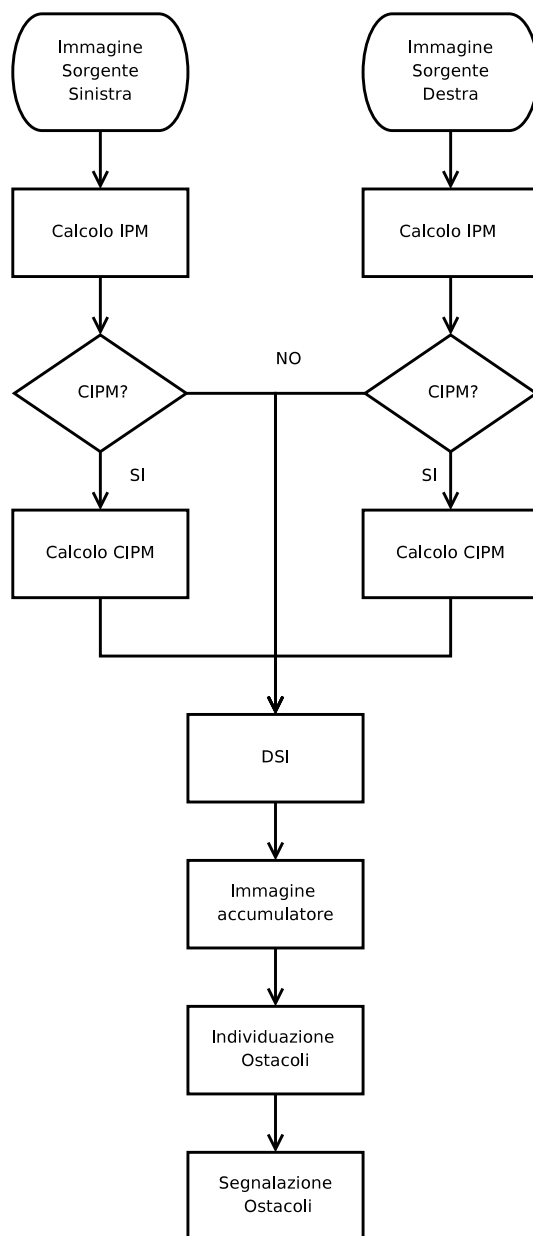


Figura 3.12: Schema a blocchi dell' algoritmo.

lo schema a blocchi dell'intero sistema. L'algoritmo implementato prende in ingresso le due immagini sorgenti, acquisite dalle telecamere poste sull'automezzo. Un grande vantaggio introdotto da questo sistema è dato dal non dovere effettuare l'allineamento delle telecamere, passo complesso e necessario per molti sistemi di visione. Le telecamere sono perciò utilizzate senza la conoscenza dei loro parametri intrinseci ed estrinseci. Gli unici parametri necessari all'algoritmo sono le coordinate delle due telecamere rispetto al mondo. Un possibile approccio analitico per la rimozione della distorsione è stato scartato per la sua notevole complessità avendo a che fare con un modello matematico delle lenti sferiche estremamente complesso.

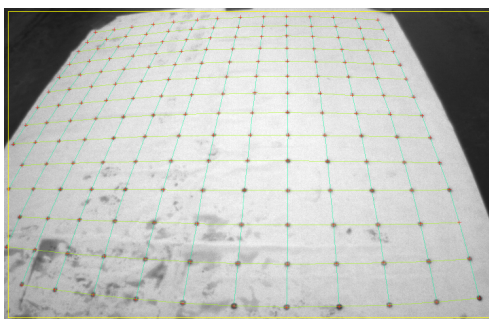
Per la rimozione della distorsione e della prospettiva è stata allora usata una strategia empirica con il calcolo *offline* di una Look Up Table. Viene per prima cosa acquisita per entrambe le telecamere un'immagine di una griglia a quadri posta sul terreno davanti al veicolo. Attraverso l'utilizzo di un'interfaccia grafica, l'operatore poi associa manualmente i punti di intersezione della griglia dell'immagine sorgente (fig. 3.13.a) ai corrispondenti punti di una griglia a scacchiera sull'immagine IPM (fig. 3.13.b). Per rimappare i restanti pixel, che non sono intersezione della griglia, viene utilizzata una funzione di interpolazione non lineare con l'uso di spline cubiche.

Conoscendo la posizione della griglia rispetto alla telecamera e assumendo che la strada sia piatta, è possibile costruire l'intera immagine IPM, rimuovendo contemporaneamente sia la distorsione sia la prospettiva [24, 26, 53]. Il risultato di questa elaborazione è una LUT che al suo interno contiene, per ogni pixel dell'immagine sorgente, la posizione dello stesso nell'immagine IPM. La LUT viene salvata per poi essere caricata in memoria all'avvio del sistema, per permettere la rimappatura *online* di tutti i fotogrammi acquisiti. La generazione della LUT quindi viene effettuata una sola volta dopo l'installazione delle telecamere.

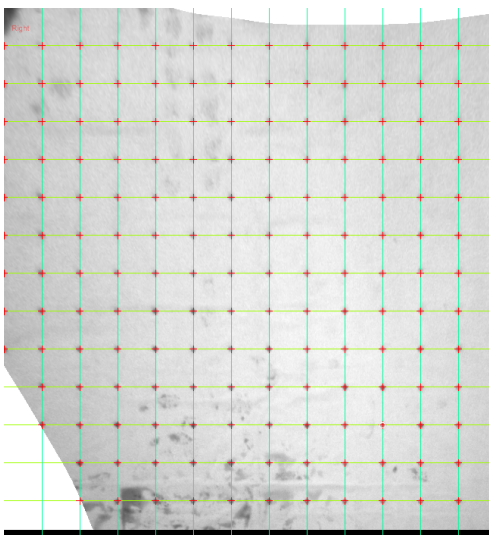
Dopo aver applicato la LUT alle immagini sorgenti per ottenere le immagini IPM, l'algoritmo continua con la generazione delle immagini di disparità su immagini IPM o CIPM. In questo ultimo caso risulta necessario generare le immagini cilindriche trasformando le immagini IPM. A partire dall'immagine di disparità è possibile effettuare una ricostruzione 3D della scena osservata e quindi individuare gli ostacoli presenti. Gli oggetti individuati saranno poi segnalati in *output* sovrapposti alle

immagini sorgenti. DSI potrà essere calcolata su immagini IPM oppure CIPM.

L'implementazione dei primi passi dell'algoritmo, compresa la generazione dell'immagine DSI sono stati trattati nei capitoli precedenti; ora si analizzeranno le fasi successive.



(a)



(b)

Figura 3.13: Mappatura manuale della griglia: immagine distorta (a), IPM (b).

3.3.1 Immagine accumulatore

L'individuazione degli ostacoli, a partire dall'immagine di disparità, avviene attraverso un'apposita immagine accumulatore. Essa ha le stesse dimensioni dell'IPM e come quest'ultima rappresenta una vista dall'alto della scena osservata. Inizialmente ogni pixel dell'immagine viene settato al valore nullo. Per tutti i pixel dell'immagine di disparità che presentano un valore di disparità valido, cioè non sconosciuto, si calcola la rispettiva posizione del mondo. Per fare ciò si utilizzeranno le formule di ricostruzione 3D calcolate nei capitoli precedenti. Se l'immagine DSI utilizzata è stata costruita a partire dalle immagini IPM si utilizzeranno le equazioni 2.6, 2.7 e 2.8; se invece l'immagine DSI deriva dalle immagini CIPM saranno usate le equazioni 3.13, 3.14, 3.15.

Avendo calcolato la posizione mondo del punto, la si può rimappare nell'immagine accumulatore andando ad aumentare il valore del pixel corrispondente a cui appartiene il punto. Effettuando questo procedimento per ogni pixel dell'immagine DSI si ottiene un'immagine in cui il valore di ogni pixel rappresenta la quantità di punti mondo con le stesse coordinate X, Y , figura 3.14.

Se un pixel ha valore alto significa che nel corrispondente punto mondo è presente un ostacolo verticale, con un'altezza proporzionale al valore del pixel.

Dovendo individuare solo gli ostacoli superiori ad una determinata altezza, si è deciso di eliminare tutti i punti mondo con un'altezza inferiore ad una determinata soglia, così facendo si eliminano tutti i punti mondo appartenenti al terreno e a piccoli oggetti posti su di esso. L'eliminazione di questi punti comporta anche l'eliminazione di alcuni falsi positivi dovuti ad un errato calcolo della disparità da parte nella DSI.

Analizzando le caratteristiche degli ostacoli da localizzare e considerando la forma risultante degli oggetti, si è deciso di dividere l'immagine accumulatore in zone più grandi in grado di contenere più punti mondo con diverse coordinate XY , fig. 3.15. Il mondo viene quindi diviso in mattonelle. Se si aumenta la dimensione della mattonella, più punti mondo appartenenti allo stesso ostacolo, cadranno al suo interno e quindi sarà più facile individuare la presenza di un ostacolo; ma allo stesso tempo la sua posizione mondo sarà meno precisa. In questo progetto si è scelto di utilizzare mattonelle che rappresentano una porzione di mondo dalle dimensioni 20×10 cm.

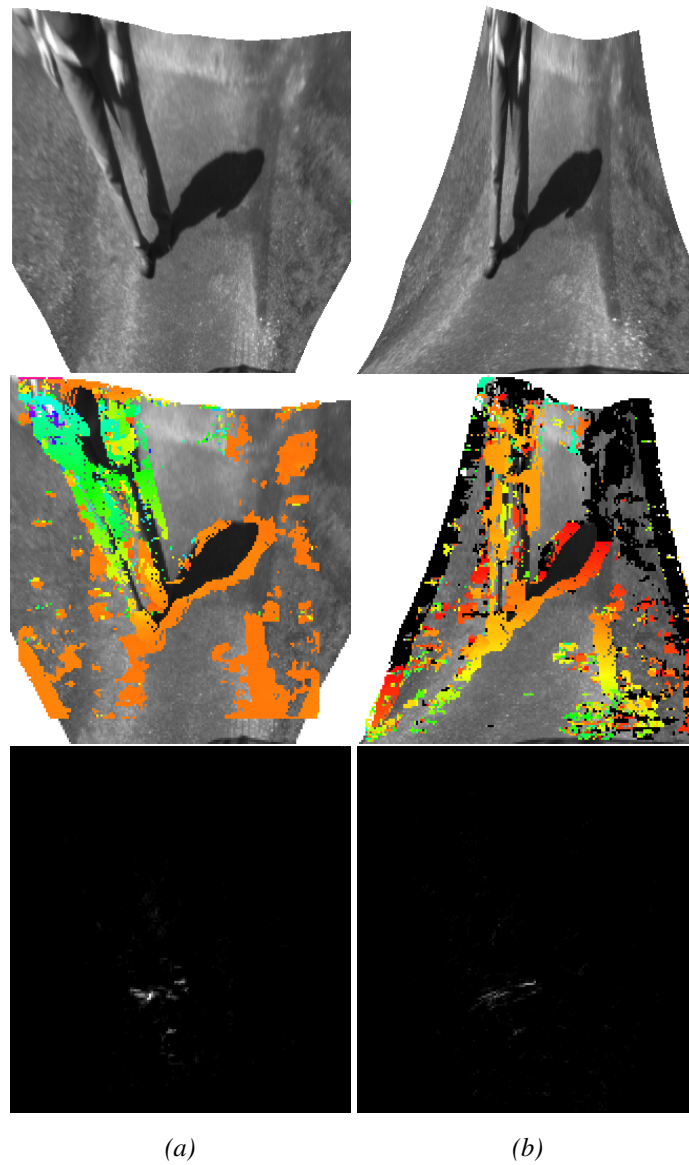


Figura 3.14: Immagini IPM, DSI e accumulatore per disparità calcolate su immagini IPM: classica (a) e cilindrica (b).

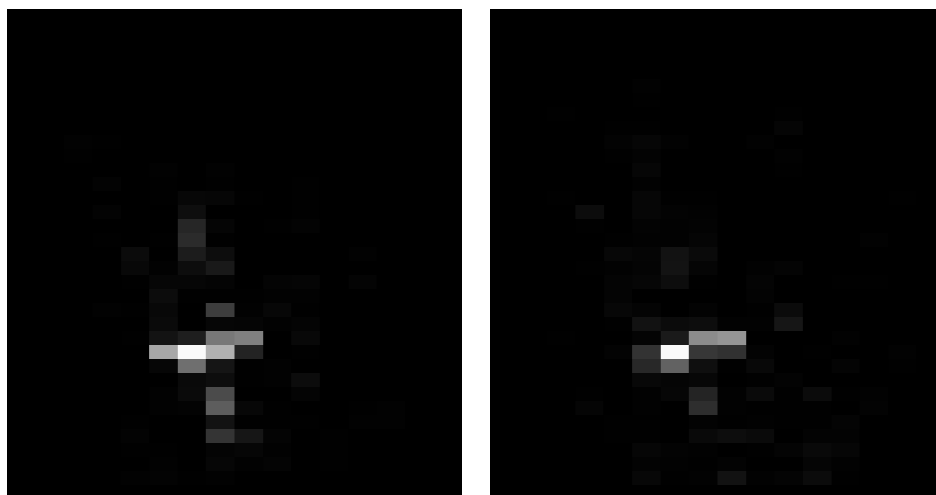


Figura 3.15: Immagine accumulatore con mattonelle di dimensione 20×10 cm generata da IPM: classica e cilindrica.

La prossima fase dell'algoritmo analizzerà l'immagine ostacoli così costruita.

3.3.2 Individuazione ostacoli

In questa fase si analizzerà l'immagine accumulatore alla ricerca di ostacoli. Per fare ciò si prende una singola mattonella dell'immagine e la si marca come ostacolo, se ha un valore superiore ad un'opportuna soglia, come raffigurato in immagine 3.16.

Il valore della soglia, che è stato individuato sperimentalmente analizzando un elevato numero di immagini, è un elemento critico per il sistema perché, agendo su di essa, si aumenta o diminuisce il numero di ostacoli individuati. Se la soglia ha un valore troppo basso, si segnalano come ostacoli un maggior numero di punti quindi si ha un'alta probabilità di segnalare correttamente tutti gli ostacoli, con il rischio però di avere falsi positivi, cioè considerare come ostacoli anche punti che non li sono. Al contrario se si tiene una soglia troppo alta si ha un minor numero di falsi positivi ma si rischia di non individuare tutti gli ostacoli reali.

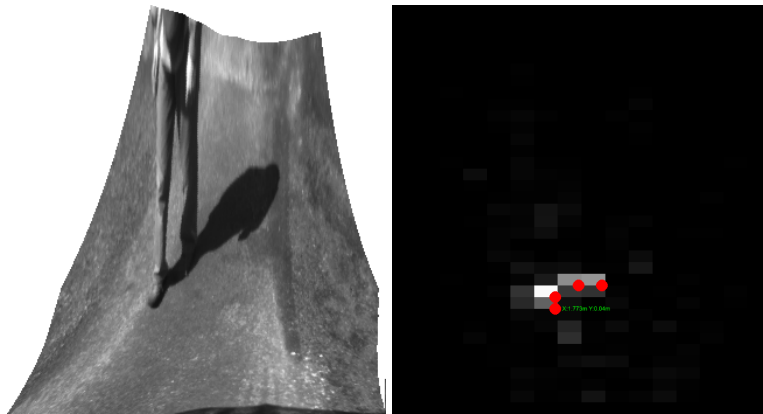


Figura 3.16: L'immagine accumulatore è rappresentata da mattonelle in toni di grigio. I pallini evidenziano le mattonelle che superano la soglia impostata e quindi segnalano la posizione di un ostacolo.



Figura 3.17: Ostacolo localizzato sull'immagine sorgente.

Si è scelto di utilizzare una soglia variabile, dipendente dalla distanza dell'ostacolo dalle telecamere. La soglia sarà maggiore nella zona vicina alla telecamera e diminuirà proporzionalmente all'aumentare della distanza. Questo perché nelle immagini, gli ostacoli più vicini alle telecamere, essendo maggiormente inquadrati, saranno rappresentati da più punti, mentre gli ostacoli lontani, essendo ripresi solo in parte, saranno rappresentati da un minor numero di pixel.

La soglia cambia anche in funzione del tipo di DSI utilizzata. Se viene usata l'immagine di disparità cilindrica è possibile utilizzare una soglia minore in quanto l'immagine CIPM ha un numero minore di pixel rispetto all'immagine IPM e in più il calcolo della disparità è più preciso avendo migliorato il *matching* fra le finestre.

Come si nota dall'immagine 3.16, è possibile ricostruire la posizione nel mondo dell'oggetto essendo nota la relazione che lega ogni singola mattonella al mondo. L'errore commesso dipende dalla quantizzazione effettuata, quindi dalle grandezze di ogni singola mattonella.

L'ultimo passo dell'algoritmo consiste nel riportare gli ostacoli individuati sull'immagine originale, per poterla mostrare al guidatore, figura 3.17. Per fare questo si applicano le formule inverse della LUT che legano le coordinate IPM all'immagine sorgente.

Sull'immagine sorgente possono essere individuati solo ostacoli presenti nella zona che inizialmente è stata mappata in entrambe le immagini IPM. Nella figura 3.17 questa zona è delimitata da una leggera linea di colore blu.

3.4 Risultati e conclusioni

In questo capitolo saranno analizzate le prestazioni e i risultati dell'algoritmo implementato, messo a confronto con un sistema preesistente, evidenziandone pregi e difetti. Verrà in primo luogo analizzato il tempo di elaborazione dell'algoritmo e poi saranno proposti attraverso immagini i risultati dell'individuazione degli ostacoli.

3.4.1 Tempi di elaborazione

Il sistema è stato testato utilizzando due calcolatori le cui caratteristiche sono indicate in tabella 3.1.

Caratteristica	PC1	PC2
Processore	Pentium 4 2.8 GHz	Pentium D 3.4 GHz
Core	1	2
Cache	512 Kb	2 Mb
RAM	2 Gb	2 Gb

Tabella 3.1: Caratteristiche tecniche dei calcolatori utilizzati.

La tabella 3.2 riporta il tempo di elaborazione dei tre algoritmi: quello utilizzato da un sistema preesistente StereoBox [19, 11, 20], l'algoritmo che applica la disparità sulle immagini IPM, infine quello che applica la disparità sulle immagini CIPM. Come si può osservare il sistema applicato sul PC1 mostra che l'algoritmo su immagini IPM ha prestazioni molto simili a quelle fornite dallo StereoBox, circa $75ms$, mentre l'algoritmo su immagini CIPM impiega più tempo di elaborazione, $90ms$. Questo fatto è dovuto all'utilizzo delle immagini di disparità nei due algoritmi implementati. Questo tempo di calcolo risulta quasi del tutto indipendente dal contenuto delle immagini ed è determinato principalmente dalle loro dimensioni e dal *range* di ricerca. Le peggiori prestazioni dell'algoritmo su immagini CIPM è dovuto alla generazione delle stesse immagini e ad un range di ricerca più ampio nell'immagine di disparità. Le prestazioni di questo algoritmo potrebbero essere migliorate se il calcolo della disparità tenesse conto di *range* di ricerca su colonne e non solo su righe.

Algoritmo	Tempo su PC1 [ms]	Tempo su PC2 [ms]
StereoBox[19]	73	35
Disparità su IPM	75	26
Disparità su CIPM	90	30

Tabella 3.2: Tempi di elaborazione

Eseguendo il sistema sul PC2, più performante del primo, notiamo un drastico calo dei tempi di elaborazione di tutti gli algoritmi, in particolare gli algoritmi implementati risultano in questo caso più veloci dello StereoBox. Questo fatto è dovuto alla struttura basata su due processori del PC2, che permette la divisione delle operazioni su due *thread*. Il codice di disparità utilizzato dal sistema, implementato utilizzando il linguaggio *assembly*, sfrutta in modo più efficiente questa struttura rispetto al sistema StereoBox.

3.4.2 Prestazioni

Il sistema è stato testato su di un elevato numero di sequenze di immagini, acquisite da un sistema stereo di telecamere montato su un automezzo usato per la compressione dell'asfalto in cantieri stradali. Le sequenze riguardano diverse scene di cantieri, con vari ostacoli fissi o in movimento, come ad esempio pedoni o altri mezzi. Le immagini sono state acquisite all'aperto durante il giorno, con varie condizioni climatiche, sole oppure cielo coperto. Questo fatto è importante perché una scena poco illuminata non permette una buona definizione dei particolari della scena, al contrario di una scena ben illuminata, che però presenta il problema delle ombre.

I test che si sono effettuati hanno riguardato l'applicazione e l'analisi di tre algoritmi di ricerca di ostacoli. Nelle immagini che saranno presentate successivamente sarà visibile il risultato finale di questi tre algoritmi mediante opportuni pallini posti sull'immagine sorgente destra, che evidenziano la base degli ostacoli rilevati e la loro relativa distanza dalla telecamera. I tre algoritmi sono dall'alto verso il basso, il sistema preesistente StereoBox, e i due algoritmi implementati da questo lavoro, cioè la disparità applicata ad immagini IPM classiche oppure cilindriche. Al fianco di queste sono proposte le immagini principali su cui operano i suddetti algoritmi: i *blob* derivati dalla differenza fra immagini IPM per lo StereoBox e le immagini accumulatore per gli algoritmi che sfruttano la disparità.

L'analisi delle sequenze hanno evidenziato ottimi risultati, in quasi tutte le situazioni, da parte degli algoritmi che usano la disparità, specialmente quello su immagini CIPM, sicuramente migliori rispetto allo StereoBox. Gli *obstacle detector* implementati permettono un calcolo più preciso della distanza degli oggetti grazie all'utilizzo

della disparità e i migliori risultati di quello derivato dalle CIPM sono dovuti alla migliore corrispondenza fra le finestre che individuano gli ostacoli. Il sistema introdotto da questo lavoro è risultato robusto e in generale più performante per quanto riguarda la non introduzione di falsi positivi, in particolare dovuti alla non omogeneità del terreno. Il sistema sviluppato è indifferente alla presenza di ombre sulla scena, mentre queste a volte influenzano il buon esito del sistema StereoBox. L'algoritmo di differenza fra immagini IPM presenta qualche problema nell'individuare ostacoli uniformi anche di grandi dimensioni; in questi casi gli *obstacle detector* implementati risultano più performanti.

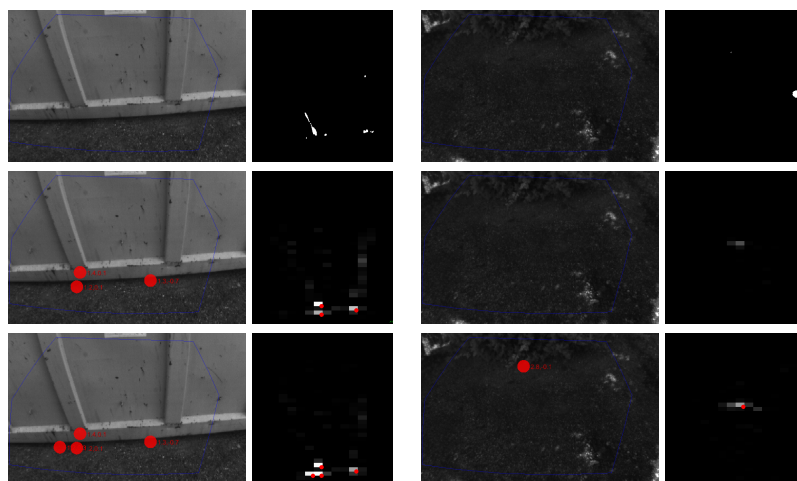
Vediamo ora tramite opportune immagini i casi principali sopra evidenziati.

In figura 3.18-a vediamo un esempio di ostacolo fisso di grosse dimensioni, caratterizzato da barre verticali. Lo StereoBox non riesce ad individuare un tale tipo di ostacolo perché la differenza fra le due IPM elimina buona parte dell'oggetto. Con la disparità invece si riesce ad evidenziare le barre verticali, evitando anche il problema dei *pattern* ripetitivi, fornendo in entrambi i casi un ottimo risultato.

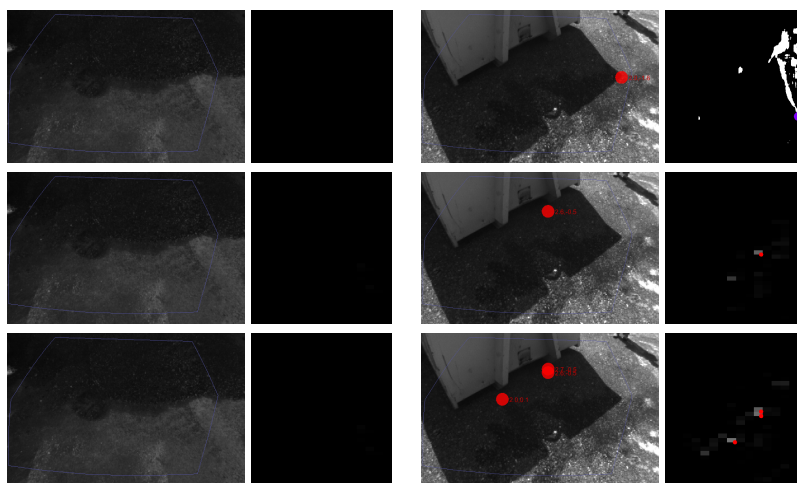
L'ostacolo in figura 3.18-b è una siepe con poco contenuto informativo visto la sua ripetitività. Come prevedibile il sistema StereoBox non è in grado di individuare un tale tipo di ostacolo, mentre l'uso della disparità permette la sua rilevazione, come si vede nelle immagini accumulate. La differente scelta delle soglie e la maggior quantità di punti sull'immagine che utilizzano le immagini IPM cilindriche fanno sì che solo in queste l'ostacolo sia poi correttamente segnalato.

Tutti gli algoritmi qui proposti non sono in grado di individuare un ammasso di catrame come quello della figura 3.18-c. Anche se questo ostacolo ha un'altezza non trascurabile la totale assenza di informazioni rende impossibile il suo riconoscimento. Notiamo che anche l'occhio umano da queste immagini fatica ad individuare una simile situazione.

Il sistema proposto non viene disturbato in modo particolare dalla presenza di ombre. In questo caso, l'ostacolo in figura 3.18-d, ha davanti a sé la propria ombra. Lo StereoBox viene ingannato da questa ombra, e oltre a non riconoscere l'ostacolo ne segnala erroneamente la posizione. Il sistema proposto, invece, riesce ad individuare l'ostacolo individuando principalmente le barre poste su di esso. Se queste



(a) Ostacolo con barre verticali (b) Ostacolo fisso uniforme



(c) Catrame (d) Ostacolo fisso con ombra

Figura 3.18: Esempi di output in diverse situazioni: (a) Ostacolo fisso con evidenti barre verticali, (b) ,ostacolo fisso uniforme (c) catrame , (d) ostacolo fisso con ombra.

non fossero presenti non si riuscirebbe ad individuare l'oggetto perché perfettamente uniforme.

In figura 3.19-e un altro esempio di come il sistema proposto riconosce in modo più preciso la posizione dell'ostacolo, non essendo influenzato dall'ombra che lo stesso ostacolo crea davanti a se. Lo StereoBox invece sbaglia a segnalare la posizione.

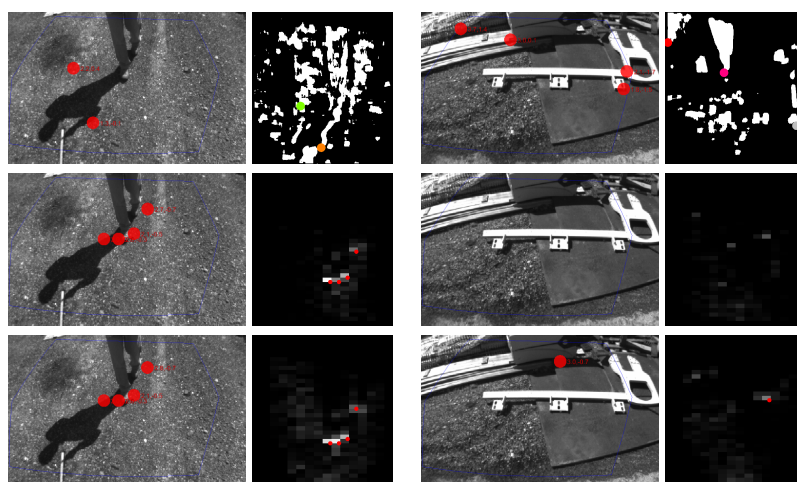
L'immagine accumulatore derivante dall'immagine di disparità segnala solo punti con un'altezza superiore ad una determinata soglia. Dei vari ostacoli presenti nella figura 3.19-f solo quello più alto viene rilevato mentre gli altri sono volutamente ignorati perché troppo bassi. L'algoritmo che usa le immagini IPM classiche non riesce però a trovare tale oggetto. Lo StereoBox invece trova tutti gli ostacoli anche quelli più bassi.

In figura 3.19-g, due pedoni sono individuati in modo opposto dal sistema StereoBox e dagli algoritmi che usano la disparità. Questi ultimi riconoscono correttamente l'ostacolo di sinistra ma non sono in grado di individuare il pedone destro perché quasi completamente esterno all'area di ricerca. Al contrario, il sistema preesistente commette un errore di localizzazione sul pedone sinistro, segnalandolo più lontano, ma riesce ad individuare correttamente il pedone destro.

In questa scena, figura 3.19-h, è simulato il lavoro di due operai. I sistemi proposti funzionano correttamente individuando la base delle persone, mentre il sistema preesistente commette qualche errore, segnalando gli ostacoli ad una distanza superiore a quella reale.

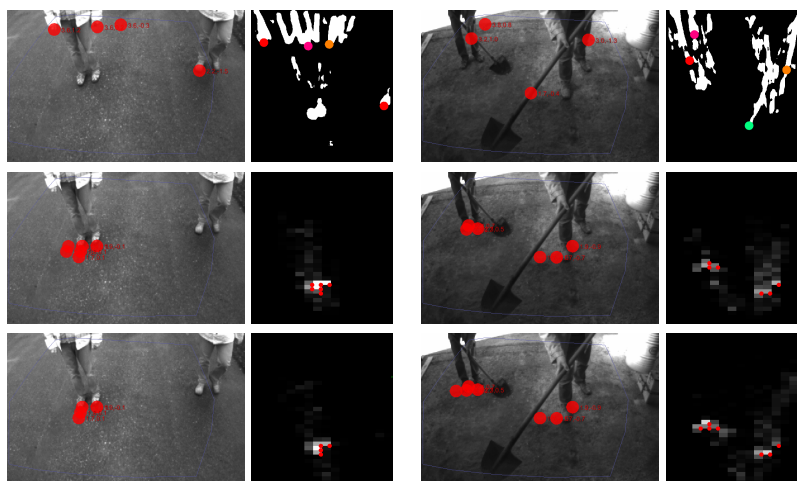
3.4.3 Conclusioni

Questa tesi ha avuto come oggetto lo sviluppo di un sistema per l'individuazione di ostacoli in cantieri stradali mediante la visione artificiale. Il sistema è stato progettato per essere installato su un automezzo operante in cantieri stradali a basse velocità, ed è in grado di individuare ostacoli a breve distanza e segnalare la loro presenza al guidatore tramite segnale visivo o acustico. Per fare questo sono utilizzate due telecamere stereo con ottiche grandangolari allineate in modo arbitrario: una semplice calibrazione del sistema mediante una griglia permette l'eliminazione della



(e) Pedone con ombra

(f) Ostacoli vari



(g) Due pedoni

(h) Operai al lavoro

Figura 3.19: Ulteriori esempi di output in diverse situazioni: (e) pedone con ombra, (f) ostacoli vari, (g) due pedoni, (h) operai al lavoro.

prospettiva e della distorsione generando nuove immagini IPM, raffiguranti una vista dall'alto della scena. Questo lavoro è partito dallo studio e dall'analisi di un algoritmo stereoscopico analizzato precedentemente nel Cap. 2.2. Dopo aver constatato le problematiche ad esso legate si è implementato un nuovo algoritmo di individuazione ostacoli più accurato e più preciso nel rilevare e localizzare ostacoli presenti in cantieri stradali. Questo studio infatti ha reso possibile la ricostruzione tridimensionale del mondo sfruttando la differente deformazione che la trasformazione IPM applica ad ostacoli verticali. In particolare l'analisi dell'algoritmo di generazione dell'immagine di disparità ha fornito l'idea per migliorare le sue prestazioni mediante la generazione di nuove immagini chiamate CIPM (Cylindric Inverse Perspective Mapping) con un'opportuna trasformazione di coordinate. Nell'immagine CIPM gli ostacoli verticali sono rimappati in modo da subire la stessa deformazione indipendentemente dalla posizione della telecamera da cui le immagini sono state generate. L'algoritmo di disparità ha un funzionamento migliore perché la ricerca dei punti omologhi è più performante.

Durante i test effettuati il sistema ha fornito buoni risultati, evidenziando prestazioni migliori rispetto ad un sistema preesistente [19], soprattutto nella precisione della distanza degli ostacoli e nel riconoscere strutture verticali anche con poche caratteristiche evidenti. Un confronto fra l'algoritmo che utilizza la disparità su immagini IPM e quelle CIPM dimostra come l'approccio sviluppato in questo sistema introduca innovazione e vantaggi dal punto di vista della precisione.

Capitolo 4

Sistema di identificazione ostacoli per veicoli autonomi

L'identificazione ostacoli tramite visione stereoscopica è un compito fondamentale della visione artificiale ed ha dato luogo ad una grande quantità di ricerca negli anni, specialmente nel campo della robotica dei veicoli. Le informazioni fornite da questa classe di algoritmi è usata sia nei sistemi di assistenza alla guida sia nei veicoli autonomi, per questo la qualità dei risultati e il tempo di elaborazione sono critici, in quanto una identificazione sbagliata o un ritardo possono avere serie conseguenze. L'algoritmo di identificazione ostacoli sviluppato e descritto in questa tesi è stato testato intensivamente durante VIAC, VisLab Intercontinental Autonomous Challenge [3, 14, 18, 7, 8]. Questo test ha offerto un'occasione unica per incontrare, lungo le strade che attraversano due continenti, una serie di differenti scenari con le più diverse condizioni ambientali. I dati raccolti durante la spedizione sono anche diventati un *benchmark* di riferimento per successivi miglioramenti dell'algoritmo.

4.1 Introduzione

Metodi veloci per ottenere una mappa di disparità accurata e densa sono diventati sempre più comuni [33, 28, 30], e di grande interesse scientifico, così come l'estra-

zione di informazioni utili da questa grande quantità di dati. Dare una definizione di un ostacolo, senza tener conto della sua complessità algoritmica, non è un compito facile. Un'opinione è di determinare una superficie dominante [58] e considerare come ostacolo qualsiasi cosa che si stacca da essa: comunque esistono situazioni dove questo approccio fallisce, ad esempio in ambienti molto affollati e confusi dove non si ha una chiara visione del manto stradale o dei bordi laterali della strada. Se non è possibile utilizzare nessuna assunzione sull'ambiente esterno, come nel caso della spedizione VIAC, è saggio e sicuro considerare la fisica e la meccanica del veicolo autonomo (ad esempio l'altezza, la larghezza e la massima pendenza affrontabile), per identificare l'area che non può essere attraversata. Il punto debole di questo approccio è la pesantezza computazionale che limita la quantità di punti che possono essere gestiti: per ovviare a questo in questa tesi si è sviluppato uno schema per l'elaborazione parallela che permette un'esecuzione *real-time* a 10 Hz utilizzando una piattaforma hardware commerciale.

4.1.1 Configurazione Hardware

Le immagini usate per la ricostruzione stereoscopica sono acquisite alla risoluzione di 752×480 pixel da una coppia di telecamere IEEE1394-A equipaggiate con lenti da 4mm. Le telecamere sono installate sotto il pannello solare, come mostrato in Fig. 4.1-a e la loro sincronizzazione è garantita da un segnale di *trigger* hardware. L'elaborazione è realizzata su un computer con un processore Intel® Core™ 2 Quad Q9100 @ 2.26 GHz e 4 GB di RAM posizionato nel vano posteriore del veicolo (Fig. 4.1-b).

4.1.2 La spedizione

Durante la spedizione VIAC i veicoli hanno attraversato ambienti molto diversi fra di loro, a partire dalle autostrade in aperta campagna in Ungheria fino al traffico frenetico delle città di Mosca e Shanghai, dai 2900 metri di altezza del passo di montagna di Lanquan ai cantieri stradali delle lunghissime autostrade in costruzione in Cina (Fig. 4.2). Anche le condizioni climatiche cambiavano drammaticamente raggiun-



(a)



(b)

Figura 4.1: Evidenziati in rosso, i componenti hardware del sistema stereo: (a) vista frontale delle telecamere stereo, e (b) l'unità di elaborazione nel baule.

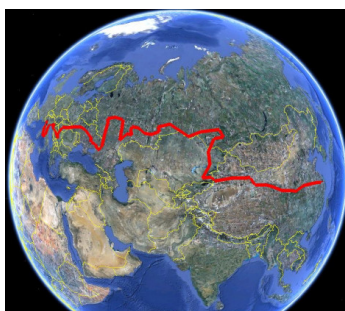


Figura 4.2: Il percorso seguito dai veicoli durante VIAC

gendo gli estremi opposti: dalla torrida estate in Ucraina, con una temperatura media durante il giorno di 45°C , al freddo settembre in Russia, dalla pioggia in Cina alla

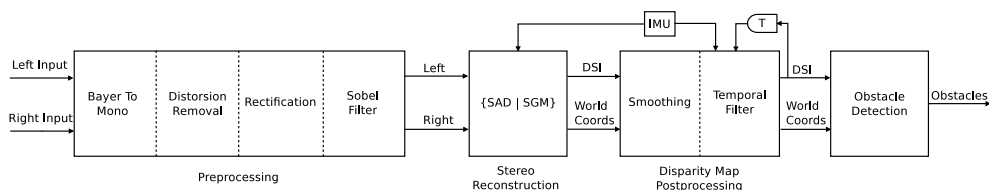


Figura 4.3: Architettura del sistema.

neve sui passi montani. Non tutte le situazioni che i veicoli hanno incontrato erano state previste ma questo era proprio lo scopo del test: sviluppare un algoritmo il più robusto possibile, valutarne le prestazioni e migliorarlo dopo aver raccolto i dati degli scenari più disparati e critici.

4.2 Algoritmi

L'elaborazione dell'algoritmo avviene seguendo i passi presentati nella Fig. 4.3:

- *Elaborazione a basso livello* — le immagini acquisite dalle telecamere con il *pattern* di Bayer sono convertite in immagini in toni di grigi, la distorsione delle lenti viene rimossa [26], e la coppia stereo rettificata. Un filtro di Sobel verticale è usato per migliorare la successiva fase di corrispondenze [35];
- *Generazione della mappa di disparità* — viene effettuata la ricostruzione tridimensionale della scena; sono stati testati due approcci diversi, un algoritmo di Semi-Global Matching (SGM) [33] e uno basato sulla correlazione, come descritto nel capitolo 4.2.1;
- *Applicazione di filtri sulla mappa di disparità* — sono applicati per aumentare ulteriormente la qualità della mappa;
- *Individuazione ostacoli* — effettiva fase dell'*obstacle detection*.

4.2.1 Generazione mappa di disparità

Data la stringente pianificazione dello sviluppo del progetto per la spedizione VIAC, la generazione della immagine Disparity Space Image (DSI) è stata implementata espandendo l'algoritmo basato sulla correlazione di finestra Sum of Absolute Difference (SAD), tecnica di correlazione già descritta in [27]. Come sarà riportato nel Cap. 4.3, i risultati ottenuti si sono dimostrati sufficienti; nonostante questo, nei mesi successivi a VIAC è stata ulteriormente sviluppata un'efficiente implementazione dell'algoritmo SGM [15], producendo risultati ottimi rispetto al precedente sistema.

Per la generazione dell'immagine Disparity Space Image D , l'algoritmo Semi-Global Matching effettua un procedimento per la minimizzazione dell'energia. La funzione energia $E(D)$ che deve essere minimizzata globalmente è composta da due termini: il costo del *pixel-wise matching* $E_{data}(D)$ e il vincolo di uniformità $E_{smooth}(D)$:

$$E(D) = E_{data}(D) + E_{smooth}(D) \quad (4.1)$$

Il termine $E_{data}(D)$ è la somma di tutti i costi di *matching* dei pixel C per ogni disparità D :

$$E_{data}(D) = \sum_{\mathbf{p}} C(\mathbf{p}, D_{\mathbf{p}}) \quad (4.2)$$

Invece di utilizzare la correlazione come nella funzione di *pixel-wise matching*, come è stato descritto in [33], si è utilizzata la distanza di Hamming [31] della trasformata di Census [61] su una finestra 5×5 presa attorno a \mathbf{p} . Questo metodo produce infatti, risultati simili [36, 28] riducendo il carico computazionale complessivo.

Il termine E_{smooth} aggiunge una piccola penalità P_1 a tutti i pixel \mathbf{q} nel vicinato $N_{\mathbf{p}}$, per i quali la disparità varia di uno da \mathbf{p} , ed una penalità più grande P_2 se la differenza è maggiore:

$$E_{smooth} = \sum_{\mathbf{q} \in N_{\mathbf{p}}} P_1 \mathbf{T}[|D_{\mathbf{p}} - D_{\mathbf{q}}| = 1] + \sum_{\mathbf{q} \in N_{\mathbf{p}}} P_2 \mathbf{T}[|D_{\mathbf{p}} - D_{\mathbf{q}}| > 1] \quad (4.3)$$

con

$$\mathbf{T}[x] = \begin{cases} 1 & \text{se } x \text{ vero} \\ 0 & \text{altrimenti} \end{cases} \quad (4.4)$$

La minimizzazione di $E(D)$ è un problema NP completo e l'algoritmo SGM l'approssima calcolando i valori di $E(D)$ lungo 8 rette equispaziate passanti per tutti i pixel vicini, usando strumenti di programmazione dinamica. I costi L'_r di ogni percorso \mathbf{r} sono aggregati come descritto in in Eq. 4.5 per ogni pixel \mathbf{p} e disparità d :

$$\begin{aligned} L'_r(\mathbf{p}, d) &= C(\mathbf{p}, d) + \min(L'_r(\mathbf{p} - \mathbf{r}, d), \\ &L'_r(\mathbf{p} - \mathbf{r}, d - 1) + P_1, L'_r(\mathbf{p} - \mathbf{r}, d + 1) + P_1, \\ &\min_i L'_r(\mathbf{p} - \mathbf{r}, i) + P_2) \end{aligned} \quad (4.5)$$

Il valore finale di disparità per ogni pixel è quindi determinato da una strategia *Winner-Takes-All* (WTA) applicata ai valori di L'_r . Per aumentare ulteriormente la qualità dei risultati, viene eseguita un'interpolazione sub-pixel in aggiunta ad un filtro mediano e ad un controllo di consistenza sinistra-destra.

È stato quindi introdotto uno schema di elaborazione SIMD *multi-threading*, affinché si possa completamente sfruttare l'efficienza dell'elaborazione parallela delle moderne CPU con multiprocessore e raggiungere frequenze di elaborazione *real-time*. Nell'algoritmo SGM, la fase che occupa più tempo è l'aggregazione dei percorsi, dato che questa deve essere realizzata per ogni pixel, disparità e percorso. Per velocizzare il procedimento, per ogni percorso i pixel vengono divisi in vari tratti indipendenti processati in parallelo; inoltre solo il valore di aggregazione è salvato in memoria, mentre i dati locali che servono per l'elaborazione incrementale sono mantenuti all'interno dei registri della CPU¹. Infine, quando viene calcolato il risultato dell'Eq. 4.5 è utilizzato il set di istruzioni Intel[®] SSE, generando contemporaneamente 16 valori di disparità alla volta.

¹Sono supportati solo range di ricerca di disparità fino a 128; oltre questa soglia non ci sono abbastanza registri CPU disponibili.

Algoritmo	Piattaforma Hardware	Dim. Immagine [px]	Disp. Max [px]	Tempo [ms]	Banda di disparità [s ⁻¹ × 10 ⁶]
Gehrig EC-VW10 [28]	Intel® Core™ i7 975 EX @ 3.3 GHz	640 × 320	128	224	117
Hirschmüller ISVC10 [34]	NVIDIA® GeForce™ 8800 Ultra	640 × 480 320 × 240	128 64	238 76	165 64
Gehrig IC-VS09 [29]	Xilinx® Virtex-4 FX140	2 × 340 × 200	64	40	218
Nedevschi IV10 [30]	NVIDIA® GeForce™ GTX 200	512 × 383	56	19	578
Questa Tesi	Intel® Core™ i7 920 @ 3.20 GHz	640 × 320	128	27	970

Tabella 4.1: Prestazioni dell'algoritmo SGM

La tabella 4.1 presenta un confronto con lo stato dell'arte per l'implementazione dell'algoritmo SGM. Le condizioni di *testing* differiscono significativamente fra di loro, così l'ultima colonna contiene la banda di disparità ottenuta misurata in numero di valori calcolati per secondo; le piattaforme hardware risultano altrettanto eterogenee, sia per quanto riguarda l'architettura sia per le prestazioni, ad esempio la scheda grafica usata in [30] è il doppio più veloce di quella usata in [34]. L'approccio sviluppato gira circa otto volte più veloce rispetto alla miglior implementazione su CPU o hardware simile, e due volte più forte della migliore implementazione su GPU, la quale oltretutto usa solo quattro percorsi di aggregazione ed un singolo coefficiente P per calcolare E_{smooth} .

4.2.2 Filtri applicati all'immagine DSI

Al termine della generazione della mappa di disparità, questa viene ulteriormente processata per eliminare possibili falsi valori; questo procedimento risulta particolarmente utile quando si utilizza l'algoritmo stereo basato su correlazione, in quanto i minimi locali probabilmente introducono più rumore.

Un primo filtro, chiamato filtro di *smoothing*, analizza una finestra 3×3 attorno ad ogni elemento dell'immagine DSI, controllando che il valore di disparità sia simile ad un numero sufficiente di vicini, e lo marca come non valido nel caso che questa condizione non sia soddisfatta. Dopo questo, ai pixel non validi che hanno un vicinato con una varianza inferiore ad una soglia prefissata, viene assegnato il valore medio degli elementi circostanti.

Un secondo filtro usa le informazioni provenienti da un sensore inerziale IMU montato sul veicolo, per calcolare la traiettoria del veicolo fra l'acquisizione precedente e quella corrente [21], controllando la consistenza della corrispondente mappa di disparità. Al tempo T ogni punto $p_i(u, v, d)$ della mappa di profondità D_T viene proiettato nel corrispondente punto mondo $p_w(x_w, y_w, z_w)$ [32], il quale viene nuovamente riproiettato nelle coordinate DSI, utilizzando le posizioni e le orientazioni delle telecamere di un sistema stereo virtuale, rispetto al sistema di riferimento corrente, al tempo $T - 1$. Questo processo genera la mappa di profondità D_{T-1}^T , che può essere direttamente comparata a D_{T-1} , calcolata al tempo $T - 1$. Ogni pixel di D_{T-1}^T viene analizzato, e considerato valido solo se almeno uno dei pixel nell'intorno 3×3 nell'immagine D_{T-1} ha un valore di disparità abbastanza simile. Come si può vedere nella figure 4.4 questo tipo di filtro è efficace nel cancellare il rumore casuale, visto la scarsa probabilità che questo rimanga consistente fra frame diversi, ma per la sua stessa costruzione conduce altresì alla soppressione degli oggetti che si sono mossi velocemente nella scena. Per questo motivo si deve prestare particolare attenzione nella regolazione della soglia di rimozione, che dovrà essere fissata tenendo presente la massima velocità relativa degli oggetti che si vogliono individuare.

4.2.3 Individuazione ostacoli

Siccome non si possiede nessuna conoscenza a priori sulla qualità delle infrastrutture della strada e sul tipo di traffico che si sta analizzando, per sviluppare l'algoritmo di individuazione ostacoli è stato utilizzato un approccio precedentemente descritto in [45]. Questa tecnica definisce un criterio per raggruppare punti in ostacoli (*cluster*), basandosi sulla loro posizione nello spazio e sulle caratteristiche fisiche del veicolo autonomo, vale a dire la sua altezza H_{max} , l'altezza minima degli ostacoli da rilevare

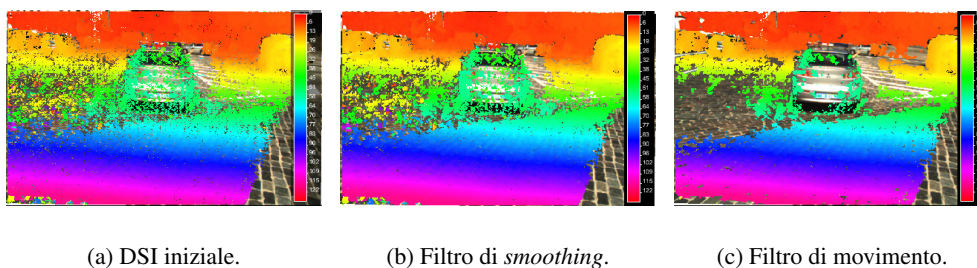


Figura 4.4: Filtri sull'immagine di disparità; i colori codificano i valori di disparità. Il filtro di *smoothing* rimuove i valori spuri ed isolati e riempie i buchi nelle aree uniformi, mentre il filtro di movimento rimuove i pixel inconsistenti, come molti punti al centro dell'immagine. In questo caso i filtraggi riescono a sopprimere la maggior parte del rumore dovuto ad un disallineamento istantaneo del sistema stereo causato da un dosso sulla strada.

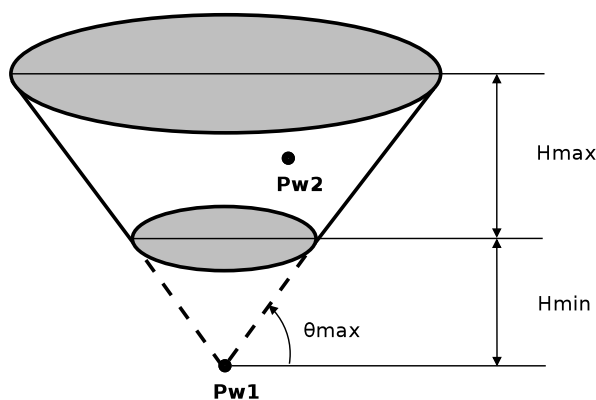


Figura 4.5: Criterio di compatibilità usato per la generazione dei *cluster*. Il punto p_{w2} è considerato un ostacolo siccome la sua posizione relativa a p_{w1} porta ad un percorso che non è attraversabile dal veicolo.

H_{min} e la massima pendenza percorribile θ_{max} . Dati due punti p_{w1} e p_{w2} questi vincoli sono usati per definire un cono troncato nello spazio, con il vertice corrispondente a p_{w1} , come rappresentato in Fig. 4.5: se p_{w2} cade all'interno del cono è etichettato come ostacolo, e p_{w1} e p_{w2} sono definiti *compatibili*.

Per ridurre il numero di confronti da effettuare, il controllo sulla compatibilità viene effettuato in coordinate immagine, al posto che in quello mondo: questo si ottiene riproiettando il cono troncato dal mondo all'immagine usando le informazioni sulla calibrazione delle camere. Effettuando questa operazione si ottiene approssimativamente un trapezio, sul quale viene controllato se i punti della disparità contenuti in esso rispettano i vincoli imposti. Ripetendo questa operazione per ogni possibile trapezio contenuto nell'immagine DSI, iterando dal basso all'alto e da sinistra a destra, è possibile unire correttamente tutti gli ostacoli in *cluster*. Durante la fase di *clustering* ad ogni regione è assegnata un'etichetta; per assicurare che un singolo ostacolo non sia diviso in più regioni adiacenti è stata adottata questa strategia: siano l_{w1} e l_{w2} rispettivamente le etichette associate ai punti p_{w1} e p_{w2} , quando

- se $l_{w1} \neq \text{definita}$ e $l_{w2} = \text{non definita}$, $l_{w2} \leftarrow l_{w1}$
- altrimenti se $l_{w1} = \text{definita}$ e $l_{w2} \neq \text{non definita}$, $l_{w1} \leftarrow l_{w2}$
- altrimenti se $l_{w1} \neq l_{w2}$ tutti i punti con etichetta l_{w2} , più p_{w2} , sono rietichettati come l_{w1}

Il risultato finale è salvato in due rappresentazioni duali:

- un'immagine, avente la stessa risoluzione della mappa di profondità, dove ogni valore di pixel corrisponde ad un'etichetta ID;
- un vettore di regioni, ognuno contenente la lista dei punti che appartengono a quella determinata regione.

Usando le coordinate DSI il carico computazionale associato a questa fase di *clustering* è considerevole, specialmente quando si utilizza la mappa di disparità densa prodotta dall'algoritmo di *matching* SGM. Per sfruttare completamente l'efficienza dell'elaborazione parallela della piattaforma hardware designata, è stato introdotto

un'ulteriore schema di analisi *multi-thread* e multi-risoluzione. L'immagine DSI originale, insieme ai vettori di punti 3D associati ad essa, viene divisa in N immagini, ognuna contenente solo i punti corrispondenti ad un prefinito raggio di distanze lungo l'asse X , come illustrato in Fig. 4.6.

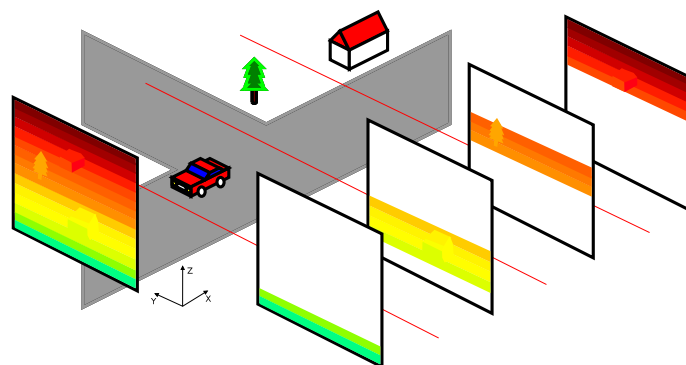


Figura 4.6: Partizione dell'immagine DSI. Sono create diverse mappe di profondità, ognuna contenente una porzione dei punti del mondo, poi le mappe sono analizzate in parallelo sui diversi processori della CPU.



(a) DSI iniziale.

(b) Etichette prima dell'unione.

(c) Etichette unite.

Figura 4.7: Etichettatura degli ostacoli in un tratto di strada in salita. I primi punti sono uniti secondo la loro posizione relativa, in parallelo su più processori, poi le etichette sono unite e filtrate per produrre la lista degli ostacoli finali. Da notare che malgrado la pendenza che il veicolo sta affrontando, non vi è nessuna individuazione sbagliata, nonostante non sia effettuata nessuna modellazione esplicita del terreno.

Le mappe di disparità sono poi scalate per ridurre ulteriormente il numero di confronti da effettuare: idealmente si potrebbe ottenere una risoluzione spaziale costante [46], ma in pratica risulta conveniente definire una singola risoluzione per ogni striscia. Utilizzando il setup hardware descritto nel Cap. 4.1.1, la partizione dello spazio è stata sperimentalmente scelta nel seguente modo:

- 0 – 5 m → sottocampionato 4×
- 5 – 15 m → sottocampionato 2×
- 15 – 30 m → piena risoluzione
- 30 m – ∞ → piena risoluzione

Dopo questo passo, ogni striscia viene processata in parallelo, generando N gruppi di etichette; questi sono poi ricopiati in una singola mappa con la risoluzione massima, riscaldando i *cluster* che erano stati ridimensionati, controllando che i punti estrapolati risultino sufficientemente simili a quelli originali, per evitare di introdurre artefatti vicino alle discontinuità della disparità.

Un ulteriore passaggio risulta necessario per assicurare che tutte le etichette siano correttamente unite, visto che un ostacolo posto nella zona di confine risulterebbe ancora diviso in due etichette distinte, come mostrato in Fig. 4.7. Perciò, ogni punto in ogni *cluster* viene controllato per determinare se il valore all'interno dell'immagine delle etichette corrisponda a quello salvato nei vettori delle regioni, e nel caso siano diversi, i due set di punti sono uniti insieme così come nell'immagine delle etichette.

4.3 Risultati

In questo capitolo sono presentati i risultati ottenuti e i tempi di elaborazione di entrambi gli algoritmi proposti, utilizzando come riferimento alcune immagini acquisite durante il viaggio VIAC.

4.3.1 Prestazioni

Per produrre una stima quantitativa delle prestazioni degli algoritmi, si sono utilizzate due delle tre metriche presentate in [56] su di una sequenza contenente mille frame acquisita nel centro di Kiev (Ucraina) approssimativamente alle 14:00 del 5 Agosto 2010. Sebbene questi dati rappresentino solo una piccola frazione dell'intero viaggio sono piuttosto rappresentativi delle tipiche condizioni incontrate nelle città trafficate.

Il primo parametro estrapolato è la percentuale di false corrispondenze $m_{fc} = N_{fc}/N$. Dove N_{fc} è il numero di punti dell'immagine DSI che riproiettati nel mondo si vengono a trovare all'interno di un volume definito dal piano della strada, dall'altezza e dalla larghezza del veicolo autonomo e da un intervallo di tempo di sicurezza di 1 s moltiplicato per la velocità del veicolo stesso; mentre N è il numero totale di punti validi nell'immagine DSI. Da notare che il valore di N_{fc} con un volume così definito dovrebbe teoricamente essere zero, in quanto con il veicolo in movimento nessun ostacolo dovrebbe essere presente di fronte al veicolo.

Il secondo parametro rappresenta la misura della posizione laterale del veicolo che precede il veicolo autonomo, definito come $m_{lp} = |lp_{measure} - lp_{groundtruth}|$, con il *ground truth* generato dalla misura diretta effettuata con un sensore LIDAR montato sul veicolo autonomo.

La tabella 4.2 contiene i valori calcolati per entrambi gli algoritmi, SAD ed SGM; come riferimento, sono stati inseriti nella tabella anche i valori riportati in [56], anche se sono riferiti ad un differente *dataset*.

Algoritmi	Questa Tesi		Steingrube ICVS09 [56]	
	m_{fc} [%]	m_{lp} [m]	m_{fc} [%]	m_{lp} [m]
SAD	3.0	0.15	1.02	0.13
SGM	0.0153	0.11	0.98	0.11

Tabella 4.2: Prestazioni dei sistemi

I risultati sono molto simili; solo nel caso del valore m_{fc} con la ricostruzione SGM cambiano significativamente, probabilmente per la differente quantità di dati

su cui sono stati calcolati.

4.3.2 Tempi di elaborazione

Gli algoritmi sono stati testati usando la stessa sequenza descritta precedentemente nel Cap. 4.3.1 su due diversi hardware: Il primo è un PC desktop con un processore Intel® Core™ i7 920 @ 3.20 GHz e 6 GB RAM, mentre il secondo è il PC industriale installato sui veicoli VIAC, descritto nel Cap. 4.1.1. La tabella 4.3 contiene i tempi di elaborazione suddivisi per ognuna delle operazioni svolte nello schema degli algoritmi, e permette un confronto reciproco dei due approcci. Le immagini sono state scalate alla risoluzione di 500×320 pixel, mantenendo il tempo totale di elaborazione del singolo frame sotto i $100ms$ in tutti i casi tranne uno.

Alcuni esempi degli output dell'algoritmo sono presentati nelle Fig. 4.8.

Fasi dell'algo- ritmo	Tempo di Elaborazione [ms]			
	Intel® Core™ i7 920		Intel® Core™ 2 Quad Q9100	
	SGM	SAD	SGM	SAD
Preproc.	2.9	2.9	4.9	5.0
DSI	21.5	5.8	60.2	7.8
Filtri DSI.	2.8	2.8	6.5	6.7
Obstacle det.	32.1	25.2	59.2	54.1
Totale	59.3	36.7	130.8	73.6

Tabella 4.3: Tempi di elaborazione

4.3.3 Conclusioni

Il sistema di individuazione ostacoli presentato è stato impiegato con successo durante la spedizione VIAC fronteggiando efficacemente una grande varietà di scenari; il sistema è stato perfino in grado di affrontare positivamente la presenza di ripidi

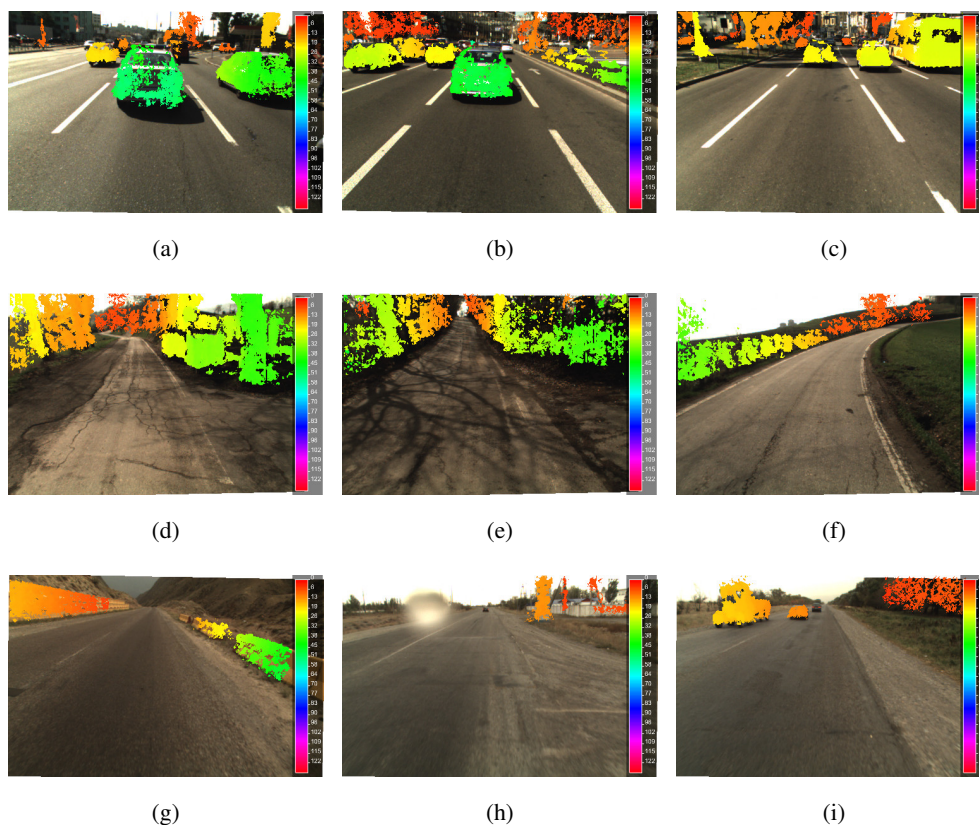


Figura 4.8: Alcuni esempi di outputs in differenti scenari. a) – (c) una trafficata autostrada a Kiev, (d) – (f) strade di campagna con alberi e strada in salita, (g) una montagna deserta in Kazakistan, (h) una goccia di pioggia sulla camera destra ed (i) un autocarro nella corsia opposta.

dislivelli e strade sterrate. Alla fine della spedizione l'algoritmo è stato ulteriormente rafforzato con l'introduzione dell'implementazione del più performante algoritmo SGM per la generazione della mappa di disparità. Inoltre, sebbene la calibrazione rimanga sempre un fattore critico, in quanto una corretta individuazione della posizione degli ostacoli dipende strettamente dalla conoscenza delle coordinate e dalle inclinazioni delle telecamere, i filtri descritti nel Cap. 4.2.2 riducono notevolmente

gli errori dovuti alle vibrazioni.

Un possibile sviluppo futuro per ottenere una miglior segmentazione degli ostacoli, potrebbe essere ottenuto dall'applicazione del flusso ottico; tuttavia andrebbe sviluppata una efficiente implementazione per permettere un'esecuzione in tempo reale anche con l'utilizzo di hardware convenzionali.

Conclusioni

In questa tesi di dottorato si sono sviluppati quattro nuovi sistemi di visione stereoscopica per la ricostruzione tridimensionale della scena inquadrata. L'ambito di applicabilità di tali sistemi ha riguardato principalmente due scenari, quello industriale e quello *automotive*. La particolare differenza fra questi due scenari è data dalla posizione del sistema di visione, in quanto nel campo industriale si sono utilizzate telecamere fisse mentre per quello *automotive* telecamere mobili solidali al veicolo. Caratteristica comune dei diversi algoritmi di visione presentati è la necessità di un'esecuzione in tempo reale e quindi l'obbligo di utilizzare tecniche per ottenere la mappa di disparità della scena in grado di fornire elevate prestazioni senza tuttavia degradare i risultati. Sono stati infatti sviluppati diversi tipi di algoritmi, sia locali che globali, per ottenere l'immagine di disparità. Una volta ottenuta la mappa di disparità si è sviluppato un sistema di individuazione ostacoli per ogni sistema, tenendo in considerazione le necessità del progetto. Le varie tecniche che sono state applicate si possono sommariamente dividere in due filoni dipendenti dalla costruzione algoritmica dell'immagine di disparità e dal metodo di calibrazione del sistema. Oltre all'approccio classico basato su di una calibrazione di tutti i parametri della testa stereoscopica e al successivo algoritmo di generazione della DSI, è stato introdotto un innovativo metodo che permette di applicare gli stessi algoritmi per la costruzione della disparità ad immagini trasformate IPM e CIPM ottenute tramite una semplice calibrazione. Questi due metodi sono stati applicati sia nel campo industriale, quindi con telecamere fisse, sia in quello *automotive* con telecamere in movimento.

Le maggiori problematiche nei campi industriale ed *automotive* (sicurezza, video-

sorveglianza e guida autonoma) sono state effettivamente coperte dai progetti in cui si sono sviluppati i sistemi presentati in questa tesi di dottorato.

Il sistema di visione stereoscopica per la sicurezza sul lavoro presentato in questa tesi ha dimostrato di poter essere utilizzato come unico sistema di sicurezza per il controllo di aree di lavorazione adiacenti a macchine a controllo numerico per la lavorazione del legno. Il sistema è stato installato e testato in diverse situazioni e si è dimostrato robusto ai test eseguiti. Il sistema di visione fornisce vantaggi che altri di sensori non possono garantire: funzionamento anche in caso di ambienti polverosi, rilevamento di oggetti che non toccano il terreno, basso impatto strutturale nella scena. Rispetto ad altri sistemi di visione, inoltre i due sistemi stereo installati ai lati opposti dell'area da vigilare garantiscono la ridondanza dei risultati e l'esclusione di possibili zone occluse, garantendo quindi l'affidabilità del sistema. La presenza di procedure di autodiagnosi che segnalano l'insorgere di possibili problemi dovuti a guasti hardware o errori software dovuti alla perdita della calibrazione, l'utilizzo di hardware industriali certificati e una scrittura del codice seguendo le norme di sicurezza imposte dalla legge, hanno reso possibile una certificazione ufficiale di sicurezza e se ne prevede una produzione industriale nei prossimi anni.

Il secondo sistema proposto è un sistema stereo per individuare ed estrarre le traiettorie dei pedoni in prossimità di accessi con porte scorrevoli per permettere l'apertura automatica delle porte. Il sistema comanda l'apertura delle porte scorrevoli solo se il pedone cammina in direzione delle stesse, mentre nel caso in cui la traiettoria del pedone sia parallela alla porta, il sistema evita l'apertura interpretando l'intenzione della persona di volere solo passare e non attraversare il varco. Il sistema proposto risolve il problema delle aperture non volute utilizzando due telecamere, installate sopra la porta, ed un elaboratore che permette la ricostruzione tridimensionale della scena tramite la costruzione della mappa di disparità. L'innovazione e la robustezza di questo sistema è data dall'implementazione di un nuovo metodo per l'estrazione dell'immagine di disparità. Infatti rispetto ad un classico sistema di visione stereoscopico viene effettuata una calibrazione basata sull'implementazione di una LUT per la rimozione dell'effetto di prospettiva e della distorsione. Questo porta alla generazione di nuove immagini IPM (Inverse Perspective Mapping), raffiguranti

una vista dall'alto della scena. Lo studio delle proprietà dell'immagine di disparità generate dalle suddette immagini ha reso possibile la ricostruzione tridimensionale del mondo sfruttando la differente deformazione che la trasformazione IPM applica ad ostacoli verticali. Il sistema ha dimostrato di lavorare correttamente e fornire ottimi risultati in diverse situazioni.

Il terzo sistema sviluppato si è occupato dell'individuazione di ostacoli in cantieri stradali mediante la visione artificiale. Il sistema è stato progettato per essere installato su un automezzo operante in cantieri stradali a basse velocità, ed è in grado di individuare ostacoli a breve distanza e segnalare la loro presenza al guidatore tramite segnale visivo o acustico. Questo lavoro è partito dallo studio e dall'analisi di un algoritmo stereoscopico analizzato precedentemente nel Cap. 2.2. Dopo aver constatato le problematiche ad esso legate si è implementato un nuovo algoritmo di individuazione ostacoli più accurato e più preciso nel rilevare e localizzare ostacoli presenti in cantieri stradali. In particolare l'analisi dell'algoritmo di generazione dell'immagine di disparità ha fornito l'idea per migliorare le sue prestazioni mediante la generazione di nuove immagini chiamate CIPM (Cylindric Inverse Perspective Mapping) con un'opportuna trasformazione di coordinate. Nell'immagine CIPM gli ostacoli verticali sono rimappati in modo da subire la stessa deformazione indipendentemente dalla posizione della telecamera da cui le immagini sono state generate. L'algoritmo di disparità ha un funzionamento migliore perché la ricerca dei punti omologhi è più performante. Durante i test effettuati il sistema ha fornito buoni risultati, evidenziando prestazioni migliori rispetto a sistemi presenti in letteratura, soprattutto nella precisione della distanza degli ostacoli e nel riconoscere strutture verticali anche con poche caratteristiche evidenti. Un confronto fra l'algoritmo che utilizza la disparità su immagini IPM e quelle CIPM dimostra come l'approccio sviluppato in questo sistema introduca innovazione e vantaggi dal punto di vista della precisione.

L'ultimo sistema di individuazione ostacoli presentato è stato impiegato con successo durante la spedizione VIAC (VisLab Intercontinental Autonomous Challenge), fronteggiando efficacemente una grande varietà di scenari; il sistema è stato perfino in grado di affrontare positivamente la presenza di ripidi dislivelli e strade sterrate. Alla fine della spedizione l'algoritmo è stato ulteriormente rafforzato con l'introduzione

dell'implementazione del più performante algoritmo SGM (Semi-Global Matching) per la generazione della mappa di disparità. Infatti in questo lavoro è stata data una notevole importanza all'implementazione e allo studio di algoritmi real-time in grado di abbassare i tempi di elaborazione rispetto a lavori già presenti in letteratura. Inoltre, sebbene la calibrazione rimanga sempre un fattore critico, in quanto una corretta individuazione della posizione degli ostacoli dipende strettamente dalla conoscenza delle coordinate e dalle inclinazioni delle telecamere, i filtri introdotti riducono notevolmente gli errori dovuti alle vibrazioni.

In conclusione si può affermare che la visione stereoscopica e la ricostruzione tridimensionale applicata a differenti scenari, anche dalle finalità più diverse, rappresenta un ottimo strumento per risolvere problemi complessi fornendo buoni risultati, superiori o comunque comparabili ad approcci basati su altre tecnologie.

Appendice A

Visione stereoscopica: Disparità

In un sistema di visione artificiale l'utilizzo di due telecamere che osservano la stessa scena, porta a notevoli vantaggi rispetto all'uso di una singola camera. Avendo infatti a disposizione due immagini della stessa scena riprese da due punti di vista leggermente diversi è possibile ottenere una visione stereoscopica in grado di fornire una rappresentazione tridimensionale del mondo.

La stessa vista umana sfrutta questo principio per dare profondità alle immagini bidimensionale che i nostri occhi percepiscono. Se non avessimo due occhi avremmo notevoli difficoltà nel capire le distanze di ogni singolo oggetto.

La base di ogni algoritmo di visione stereo sta nell'individuare la posizione di uno stesso oggetto nelle immagini provenienti dalle due telecamere. Conoscendo la posizione e l'orientazione delle telecamere è poi possibile risalire alla posizione dell'oggetto nel mondo.

La differenza di posizione fra le due immagini di uno stesso oggetto prende il nome di disparità.

Per comprendere i principi di funzionamento della visione stereoscopica e della disparità bisogna risalire alle ragioni fisiche del fenomeno. Per fare ciò si deve per prima cosa analizzare il modello geometrico che approssima il processo di formazione dell'immagine all'interno di una telecamera.

A.1 Il modello *pinhole camera*

Il modello *pinhole camera* è l'approssimazione del modo in cui si svolge il processo di formazione delle immagini, fig. A.1. Il modello utilizza il principio di funzionamento della macchina fotografica: si ipotizzi un parallelepipedo con un foro infinitesimale su un lato, detto **punto di fuoco** o **pinhole**, attraverso il quale passano tutti i raggi luminosi provenienti dal mondo. I raggi vanno a ricreare sul lato opposto al foro all'interno della scatola, chiamato **piano immagine**, un'immagine rovesciata della scena inquadrata. I punti che possono essere proiettati nell'immagine fanno parte di una piramide infinita che ha come vertice il **pinhole**. La retta perpendicolare al piano immagine, passante per il *pinhole*, è detto **asse ottico** [39].

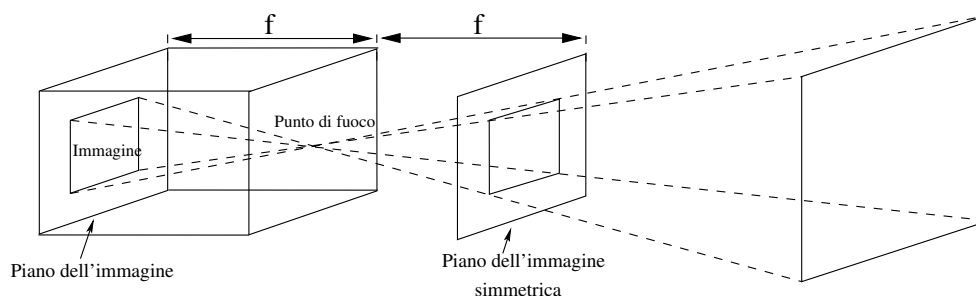


Figura A.1: Il modello *pinhole camera*.

Per ricondursi al caso reale sono usate delle lenti ottiche, che sono in grado di focalizzare i raggi luminosi in un solo punto, come supposto dal modello. Con l'utilizzo delle lenti è anche possibile variare la lunghezza focale, la distanza fra il *piano immagine* ed il *pinhole*, f nella figura. Aumentando questo valore si ingrandisce l'immagine acquisita a discapito di un restringimento del campo visivo, stringendosi la piramide.

L'immagine acquisita, come già detto, risulta essere rovesciata. Essa può essere raddrizzata attraverso un processo hardware (interno alla telecamera) o software. Considerando il piano dell'immagine, simmetricamente ad esso rispetto al *pinhole* si

individua un altro piano, detto **piano dell'immagine simmetrica**, su cui è proiettata un'immagine virtuale corrispondente all'immagine reale raddrizzata.

A.2 La trasformazione prospettica

Avendo definito il modello fisico di acquisizione delle immagini si possono ora analizzare le funzioni geometriche utilizzate per la creazione delle immagini. La trasformazione prospettica [51] è appunto quella funzione che proietta i punti del mondo (3D) sul piano immagine (2D).

Essa dipende dal modello fisico di acquisizione, *pinhole camera* nel nostro caso, e dalla posizione e dall'orientamento dei sistemi di riferimento del mondo (X,Y,Z) e dell'immagine (x,y,z) .

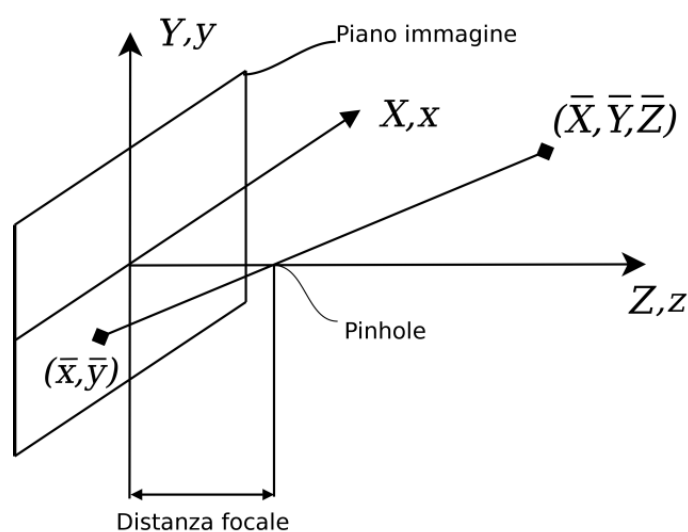


Figura A.2: Il sistema di riferimento mondo e immagine.

Come mostra la figura A.2, si suppone inizialmente che il piano immagine sia allineato al piano XY del mondo e che l'asse ottico coincida con l'asse Z [51].

Inoltre la trasformazione può essere applicata soltanto ai punti con una Z maggiore della distanza focale f , in pratica i punti fuori dalla telecamera.

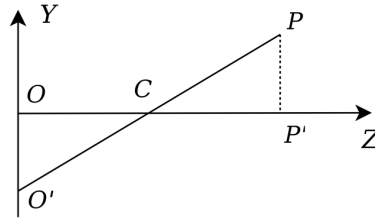


Figura A.3: La trasformazione prospettica con il modello *pinhole camera*.

Dato un punto P di coordinate mondo $(\bar{X}, \bar{Y}, \bar{Z})$, si può facilmente mettere in relazione [51] quest'ultime con le coordinate immagine, osservando che i triangoli $O'\hat{O}C$ e $C\hat{P}P'$, raffigurati in figura A.3, sono simili per costruzione:

$$\frac{-\overline{PP'}}{\overline{CP'}} = \frac{\overline{OO'}}{\overline{OC}} \quad \text{cioè} \quad \frac{-\bar{Y}}{\bar{Z}-f} = \frac{\bar{y}}{f} \quad \text{e} \quad \frac{-\bar{X}}{\bar{Z}-f} = \frac{\bar{x}}{f} \quad (\text{A.1})$$

da cui

$$\bar{x} = \frac{f\bar{X}}{f-\bar{Z}} \quad \bar{y} = \frac{f\bar{Y}}{f-\bar{Z}} \quad (\text{A.2})$$

Introduciamo ora il concetto di coordinate omogenee per riscrivere in formato matriciale le equazione A.2.

Le coordinate omogenee [39] di un punto $(x, y, z) \in \mathbb{R}^3$ sono una qualsiasi quadrupla di numeri (x_o, y_o, z_o, w_o) che soddisfino:

$$x = \frac{x_o}{w_o}, \quad y = \frac{y_o}{w_o}, \quad z = \frac{z_o}{w_o} \quad (\text{A.3})$$

L'impiego di tali coordinate, ad esempio nelle trasformazioni di rototraslazione di un punto, permette di scrivere espressioni più compatte nelle quali non compaiono somme tra vettori. Esse permettono di descrivere alcuni sistemi non lineari con semplici

calcoli matriciali, al prezzo di introdurre una nuova coordinata e quindi passare da \mathbb{R}^3 a \mathbb{R}^4 .

Ora che è stato introdotto il concetto di coordinate omogenee, si ricavano facilmente le equazioni della trasformazione prospettica **P** A.4, in forma matriciale [51]:

$$\mathbf{c} = \mathbf{P}\mathbf{w} \quad \text{con} \quad \mathbf{c} = \begin{bmatrix} k\bar{x} \\ k\bar{y} \\ k\bar{z} \\ k \end{bmatrix} \quad \mathbf{P} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & -\frac{1}{f} & 1 \end{bmatrix} \quad \mathbf{w} = \begin{bmatrix} K\bar{X} \\ K\bar{Y} \\ K\bar{Z} \\ K \end{bmatrix} \quad (\text{A.4})$$

Dove **c** sono le coordinate omogenee del punto nell'immagine e **w** quelle nel mondo. Per ottenere le coordinate dell'immagine occorre quindi svolgere il prodotto **Pw** e dividere le prime tre coordinate omogenee per la quarta:

$$\mathbf{c} = \mathbf{P}\mathbf{w} = \begin{bmatrix} \bar{x} \\ \bar{y} \\ \bar{z} \end{bmatrix} = \begin{bmatrix} \frac{f\bar{X}}{f-\bar{Z}} \\ \frac{f\bar{Y}}{f-\bar{Z}} \\ \frac{f\bar{Z}}{f-\bar{Z}} \end{bmatrix} \quad (\text{A.5})$$

Nell'equazione A.5 si nota che le coordinate \bar{x} e \bar{y} assumono un valore sensato mentre la coordinata \bar{z} è priva di senso, in quanto per costruzione ogni punto sul piano immagine ha $\bar{z} = 0$. Invece dalla A.5 si ha $\bar{z} = \frac{f\bar{Z}}{f-\bar{Z}} < 0$ avendo posto $\bar{z} > f$.

Proprio per questo motivo la trasformazione prospettica non può essere una funzione invertibile, per cui dato un punto immagine non si riesce a risalire ad un solo punto nel mondo. Ogni singolo punto dell'immagine rappresenta infatti la proiezione di tutti i punti mondo che si trovano sulla retta passante per lo stesso punto immagine $(\bar{x}, \bar{y}, 0)$ e il *pinhole* $(0, 0, f)$. L'equazione di questa retta si può ricavare dall'equazione A.5:

$$X = \frac{\bar{x}}{f}(f - Z) \quad Y = \frac{\bar{y}}{f}(f - Z) \quad (\text{A.6})$$

La funzione di trasformazione prospettica è non lineare, come si può osservare dall'equazione A.5, a causa della presenza della coordinata Z al denominatore.

L'equazione A.5 è stata ricavata ipotizzando che i sistemi di riferimento mondo e immagine siano coincidenti. Questo vincolo può essere eliminato applicando le opportune matrici di rototraslazione.

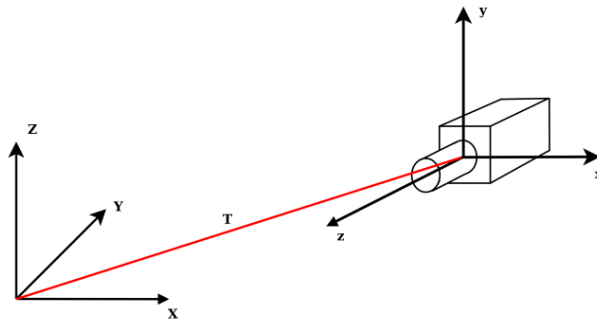


Figura A.4: Il sistema di riferimento immagine (x,y,z) e mondo (X,Y,Z) .

Se i due sistemi di riferimento hanno la stessa orientazione e sono solo traslati, fig. A.4, è sufficiente definire una matrice di traslazione \mathbf{T} , A.7:

$$\mathbf{T} = \begin{bmatrix} 1 & 0 & 0 & -X'_T \\ 0 & 1 & 0 & -Y'_T \\ 0 & 0 & 1 & -Z'_T \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (\text{A.7})$$

Dove (X'_T, Y'_T, Z'_T) sono la posizione nelle coordinate mondo della telecamera [51].

È possibile riscrivere l'equazione A.4 nella forma:

$$\mathbf{c} = \mathbf{PTw} \quad (\text{A.8})$$

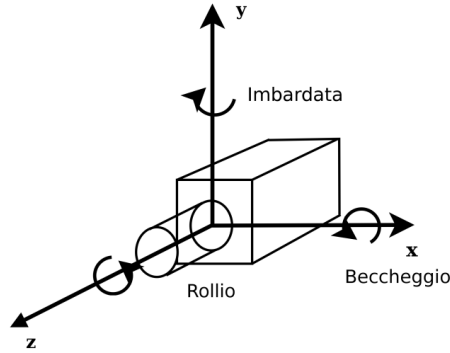


Figura A.5: L'orientamento della telecamera: beccheggio (ϕ), rollio (ψ) e imbardata (θ).

In generale i due sistemi di riferimento possono presentare anche diverse orientazioni. Questo fattore viene espresso tramite l'utilizzo degli angoli di *yaw*, *pitch* e *roll* (imbardata, beccheggio, rollio), fig. A.5.

Questi angoli vengono modellati attraverso le opportune matrici di rotazione [39]:

$$\mathbf{c} = \mathbf{P}\mathbf{R}_{\psi\phi\theta}\mathbf{T}\mathbf{w} = \mathbf{P}\mathbf{R}_{\psi}\mathbf{R}_{\phi}\mathbf{R}_{\theta}\mathbf{T}\mathbf{w}$$

$$\mathbf{R}_{\psi} = \begin{bmatrix} \cos \psi & -\sin \psi & 0 & 0 \\ \sin \psi & \cos \psi & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

$$\mathbf{R}_{\phi} = \begin{bmatrix} \cos \phi & 0 & \sin \phi & 0 \\ 0 & 1 & 0 & 0 \\ -\sin \phi & 0 & \cos \phi & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

$$\mathbf{R}_{\theta} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos \theta & -\sin \theta & 0 \\ 0 & \sin \theta & \cos \theta & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (\text{A.9})$$

Queste matrici di rotazione sono costruite utilizzando i parametri estrinseci della telecamera. La funzione di trasformazione prospettica finale deve tenere conto anche dei parametri intrinseci della telecamera (lunghezza focale, distorsione, apertura), perciò all'equazione A.9 deve essere applicata un'ulteriore matrice di trasformazione \mathbf{K} , che

tenga conto di questi parametri:

$$\mathbf{c} = \mathbf{KPR}_{\psi\phi\theta}\mathbf{T}\mathbf{w} = \mathbf{AR}_{\psi\phi\theta}\mathbf{T}\mathbf{w} \quad (\text{A.10})$$

A.3 Ricostruzione delle coordinate mondo

La ricostruzione delle coordinate mondo di un punto immagine, come abbiamo visto in precedenza, non è possibile usando una sola telecamera. Infatti allo stesso punto immagine appartengono tutti i punti mondo sulla retta passante dal *pinhole* e per il punto stesso. La soluzione al problema sta nella visione stereoscopica. Avendo a disposizione un'ulteriore telecamera che riprende la stessa scena si può calcolare l'intersezione fra le relative rette di proiezione, risalendo così alle coordinate mondo del punto immagine (fig. A.6).

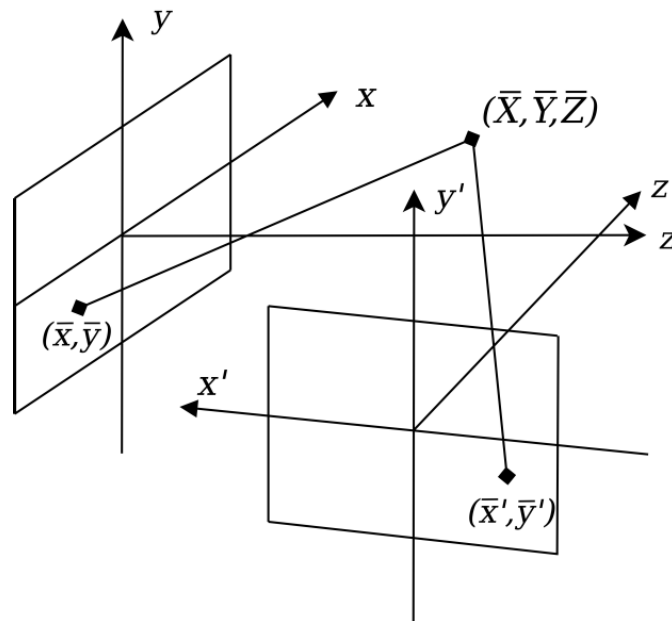


Figura A.6: La triangolazione stereo.

Utilizzando quest'approccio è teoricamente possibile risalire alle coordinate di qualsiasi punto immagine inquadrato da entrambe le telecamere cioè può essere ricostruita solo la scena comune. In realtà, tenendo conto degli errori dovuti alla calibrazione, è possibile che le due rette siano sghembe. Se le due rette non hanno intersezione è quindi impossibile risalire alla coordinata mondo del punto. L'intersezione quindi deve essere calcolata in modo approssimato, aumentando così l'errore sulle coordinate [6]. Altro fattore da considerare sta nel fatto che bisogna conoscere le coordinate immagine dello stesso punto mondo proiettato in entrambe le immagini (punti omologhi). Un leggero errore nella scelta del punto porta quindi ad un ulteriore errore nella costruzione della retta e quindi nell'intersezione fra le due.

Le coordinate mondo calcolate sono misure indirette e come già detto affette da errore. Quest'errore di norma aumenta con la distanza dell'oggetto, è quindi minimo ed accettabile per oggetti vicini mentre può risultare enorme ed intollerabile per oggetti lontani.

In conclusione possiamo affermare che la ricostruzione 3D di un punto mondo è possibile solo se si conoscono i punti omologhi e i parametri intrinseci ed estrinseci della telecamere [17].

A.4 La disparità

La **disparità** è definita come la distanza orizzontale, espressa in pixel, tra due punti omologhi. In un sistema stereo la disparità è un parametro calcolato in modo diretto sulle immagini acquisite. Nelle immagini A.7 sono evidenziate le disparità di diversi oggetti posti a diverse distanze e posizioni.

Dalla semplice analisi di queste immagini si deducono importanti considerazioni sulla disparità:

- gli oggetti verticali hanno la stessa disparità (ad es. colonne nelle fig. A.7.a);
- la disparità diminuisce all'aumentare della distanza dalla telecamera;
- elementi alla stessa distanza presentano stessa disparità (ombra nelle fig. A.7.b).

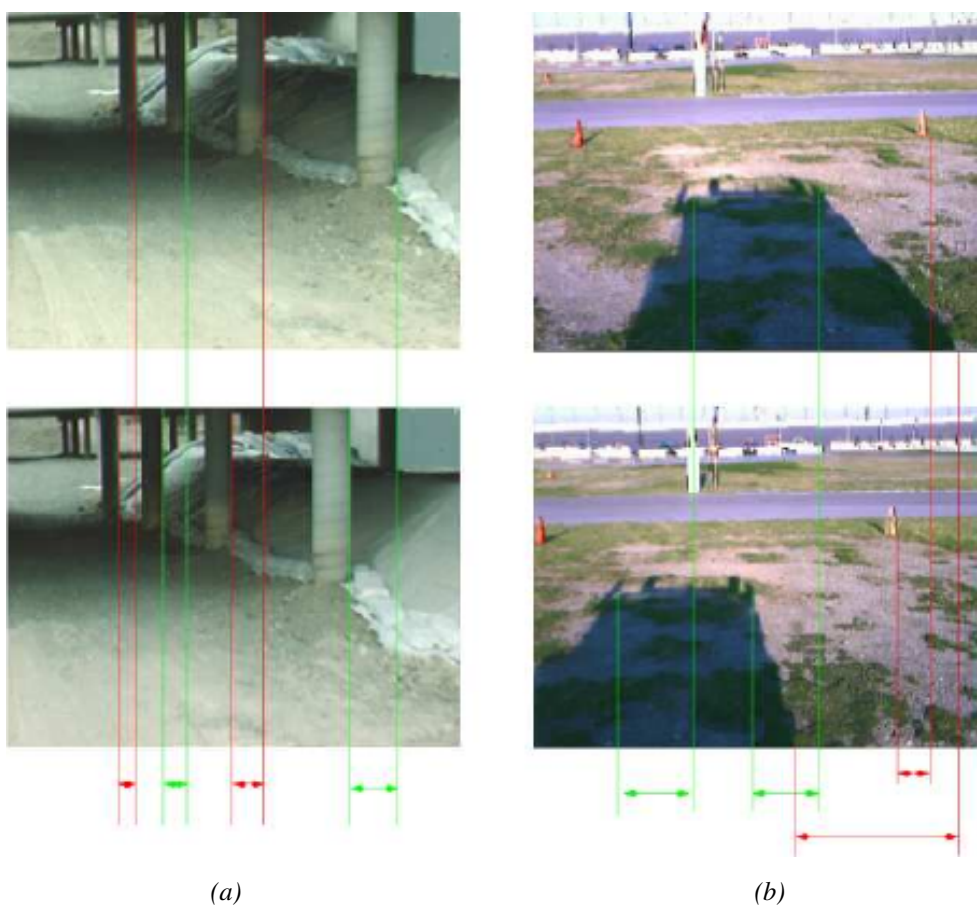


Figura A.7: Proprietà della disparità al variare delle distanze degli ostacoli. Le immagini superiori sono state acquisite con la telecamera sinistra, mentre quelle inferiori con la destra.

Con queste semplici considerazioni si può già affermare che:

Note le caratteristiche del sistema di acquisizione e la disparità con cui un oggetto appare, è possibile ricavare la sua posizione nel mondo e viceversa [62].

La disparità è una misurazione diretta e meno soggetta ad errori rispetto alle coordinate mondo. Per questo la disparità rappresenta un metodo alternativo per la rappresentazione tridimensionale del mondo, soprattutto per gli oggetti distanti dalla telecamera.

Cerchiamo di dimostrare le affermazioni precedenti analizzando il modello di telecamere stereo, schematizzato in fig. A.8.

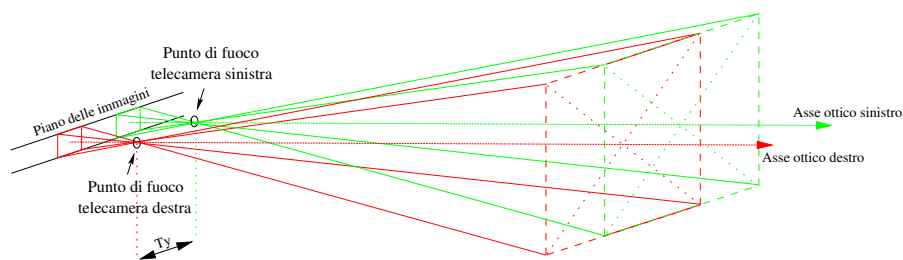
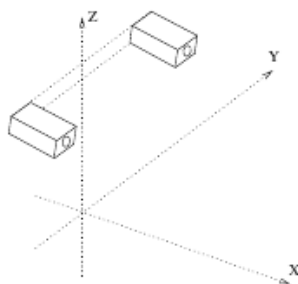


Figura A.8: Sistema di telecamere stereo.

In figura sono evidenziati i rispettivi **assi ottici** delle telecamere, che sono la perpendicolare al piano immagine passante per il **punto fuoco**. Nel modello si è supposto che i due piani immagini siano complanari, per questo i loro bordi superiori e inferiori risultano allineati. La distanza T_y rappresenta la **baseline**, cioè la differenza di posizione lungo l'asse Y delle telecamere, fig. A.9.

La **baseline** in un sistema stereo in ambito *automotive*, può assumere un valore che varia da pochi centimetri a qualche metro.

La figura A.10 rappresenta la vista dall'alto del modello precedente. Su questo sono evidenziate le proiezioni di un punto del mondo, sul piano immagine simmetrico delle due immagini, ottenute sfruttando la retta congiungente il punto mondo e il rispettivo *pinhole*.

Figura A.9: *Baseline*.

Osservazione 1 *Punti del mondo reale posti su uno stesso piano parallelo al piano delle immagini presentano la stessa disparità [16].*

L'esattezza di quest'enunciato può essere dimostrata geometricamente verificando che la disparità, ovvero la distanza relativa tra i punti sull'immagine simmetrica di uno stesso punto mondo, non cambia al variare della posizione del punto mondo lungo l'asse Z e Y . È chiaro che muovendo il punto mondo lungo l'asse Z (perpendicolare al foglio), non comporta nessun cambiamento della figura A.10, non modificando quindi la disparità. Analizziamo ora il caso in cui il punto mondo si stia spostando lungo la linea tratteggiata parallela al piano immagine simmetrico. Si considerino il triangolo di vertice punto mondo, punto fuoco destro e sinistro, e quello di vertice punto mondo, punto immagine simmetrica destra e sinistra. I due triangoli sono evidentemente simili avendo la base parallela e il vertice e l'angolo opposto uguali. La similitudine tra i due triangoli rimane anche variando il punto mondo lungo la linea del piano parallelo. Le due basi risultano quindi sempre proporzionali, cioè il rapporto tra la distanza fra i due punti immagine e T_y rimane sempre costante. Essendo T_y costante per definizione ne risulta che anche la disparità rimane sempre costante.

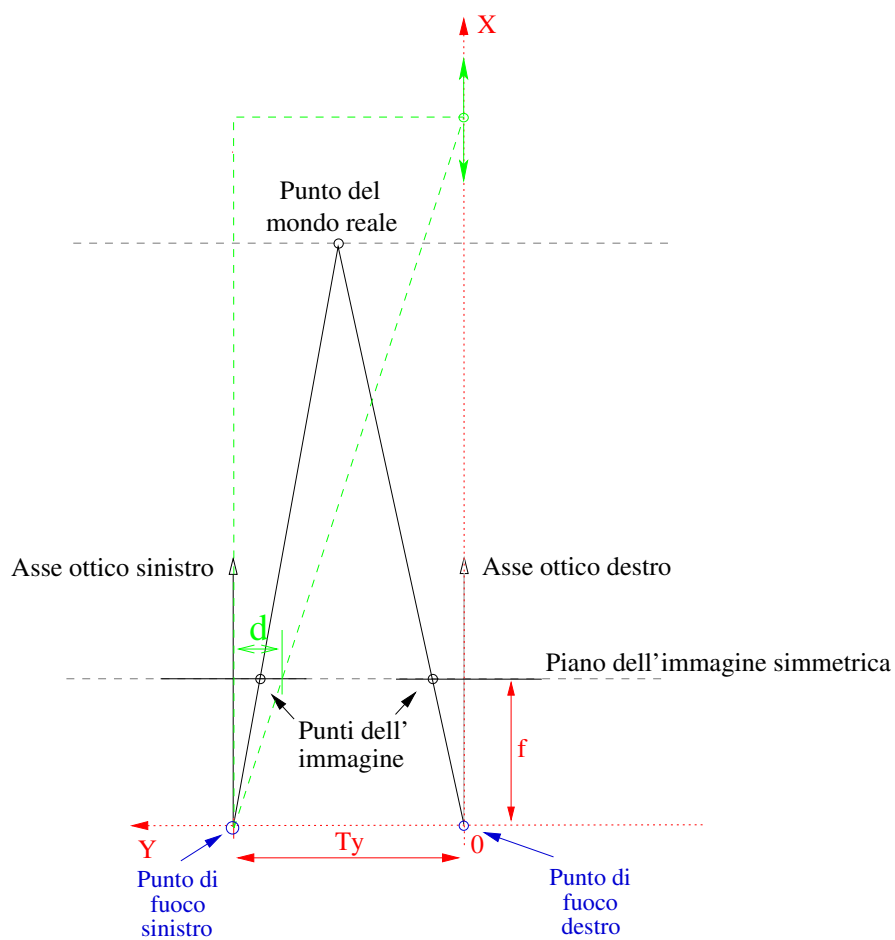


Figura A.10: La proiezione di un punto del mondo reale sul piano delle immagini.

Osservazione 2 *La disparità con la quale un punto del mondo reale è proiettato nelle due immagini è inversamente proporzionale alla distanza tra il punto reale e il piano delle immagini simmetriche (prolungato all'infinito), e può essere espressa secondo la formula [22]:*

$$d = \frac{fT_y}{X} \quad (\text{A.11})$$

dove T_y è la distanza tra i punti di fuoco delle due telecamere, f la lunghezza focale e X la distanza del punto reale dal piano delle immagini.

Partendo sempre dalla figura A.10 si prende in esame il movimento del punto mondo sull'asse X coincidente all'asse ottico dell'immagine destra. Si individua il triangolo, evidenziato nella figura dalla linea tratteggiata verde, e il triangolo con vertice basso posto sul punto di fuoco sinistro e con i vertici della base individuati dall'intersezione del piano immagine simmetrico rispettivamente con l'asse ottico sinistro e la retta obliqua che congiunge il punto del mondo al fuoco sinistro. I due triangoli sono simili per le stesse considerazioni fatte in precedenza e rimangono simili al muoversi del punto mondo. La base del triangolo più piccolo posta sul piano immagine, risulta essere per costruzione la disparità del punto mondo, perciò utilizzando le proporzioni sui triangoli si ha:

$$X : T_y = f : d \quad (\text{A.12})$$

da cui si ricava l'equazione A.11 dell'osservazione.

A.5 La ricerca dei punti omologhi

Condizione necessaria per il calcolo della disparità fra due immagini è quella di individuare nell'immagine sinistra o destra, la proiezione di un punto mondo di cui si conosce la proiezione sull'altra immagine. La ricerca dei punti omologhi è quindi un passo fondamentale e critico dal quale dipende la precisione della misurazione della disparità. La localizzazione dei punti omologhi è anche computazionalmente onerosa, per ogni punto di un'immagine occorre ricercare il suo omologo sull'immagine acquisita dall'altra telecamera. Conoscendo le disposizioni e l'orientamento

delle telecamere è possibile semplificare questi calcoli diminuendo significativamente la zona di ricerca e conseguentemente il tempo di calcolo. In figura A.11 è rappresentato un punto nel mondo reale e la sua proiezione sul piano simmetrico dell'immagine sinistra e destra. Il vincolo sulla complanarità dei piani immagine imposto precedentemente è stato rilassato.

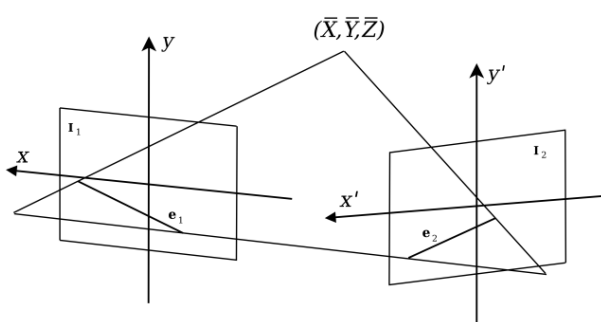


Figura A.11: Il piano e la retta epipolare.

Si definisce **piano epipolare** il piano passante per il punto mondo e per i due *pin-hole*. La sua proiezione sul piano immagine è detta **retta epipolare**. Si può osservare che, dato un punto P_1 appartenente all'immagine I_1 , il suo omologo sull'immagine I_2 sarà vincolato a trovarsi sulla retta epipolare e_2 .

L'introduzione del vincolo di epipolarità consente di ridurre la complessità della ricerca da bidimensionale ad unidimensionale. Un'ulteriore semplificazione si può ottenere nel caso in cui le rette epipolari fossero perfettamente orizzontali, ciò si verifica quando le immagini vengono rettificate oppure quando l'angolo di rollio delle telecamere è nullo e l'angolo di beccheggio uguale per entrambe.

Notiamo che non tutti i punti del mondo possono essere osservati da entrambe le telecamere rendendo perciò impossibile il calcolo della disparità. Ad esempio un oggetto può essere visto da una telecamera mentre può essere coperto da un altro oggetto dal punto di vista dell'altra telecamera. Questo fenomeno prende il nome di **occlusione** e rende impossibile il calcolo della disparità per gli oggetti occlusi. Per l'individuazione dei punti omologhi devono essere scelti punti con determinate carat-

teristiche. Infatti zone dello stesso colore sono prive di informazione ed è impossibile individuare i punti omologhi.

A.6 L'immagine di disparità

L'immagine di disparità, o *disparity space image* (DSI [49, 60, 38, 25]), è un'immagine i cui pixel hanno un valore pari alla propria disparità (figura A.12). L'immagine di disparità utilizza un approccio di tipo locale e fornisce un ottimo strumento per l'individuazione degli ostacoli [62].

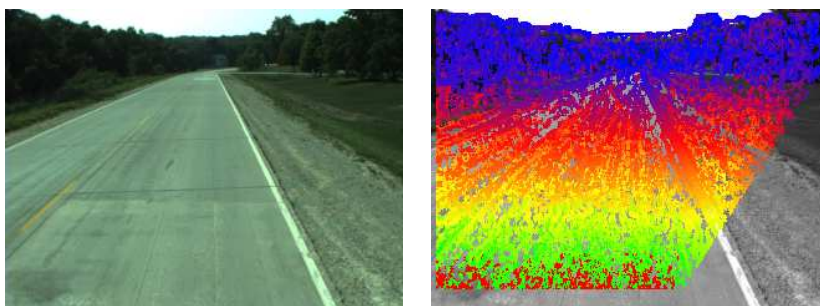


Figura A.12: Immagine destra ed immagine di disparità.

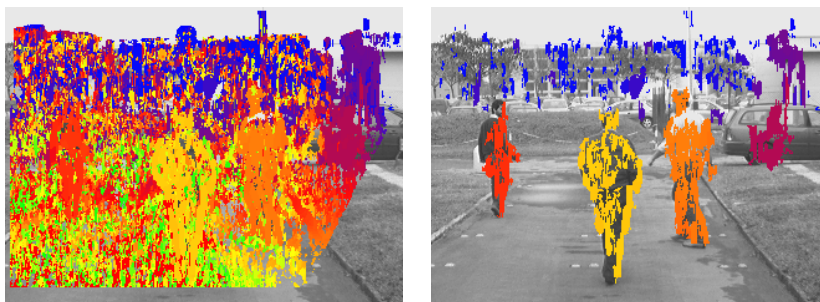


Figura A.13: L'immagine di disparità e gli ostacoli individuati su di essa.

La costruzione dell'immagine di disparità avviene abbinando ciascun pixel dell'immagine destra con quelli dell'immagine sinistra, presenti sulla corrispondente retta epipolare, scegliendo come valore di disparità quello con correlazione maggiore. Se il sistema di acquisizione utilizza telecamere allineate, con angoli di beccheggio identici, e di rollio nullo la ricerca viene semplicemente effettuata su rette epipolari orizzontali, semplificando notevolmente i calcoli.

L'abbinamento non può essere fatto utilizzando solo il singolo pixel ma richiede un confronto tra finestre di punti. Per ogni pixel viene individuata una finestra e nell'altra immagine verrà appunto cercata una finestra delle stesse dimensioni e con le stesse caratteristiche. La scelta della finestra di ricerca cambia notevolmente il risultato dell'immagine di disparità. Finestre piccole sono più adatte per l'individuazione di piccoli oggetti ma potrebbero presentare false associazioni proprio dovute alla scarsità di informazioni presenti nella finestra. Mentre finestre troppo grandi non riescono ad associare elementi di piccole dimensioni.

Per problemi dovuti alla mancanza di informazione di alcune zone, se si calcola direttamente la DSI sulle immagini acquisite, non si ottengono risultati soddisfacenti. Viene per questo motivo applicato un filtro di Sobel su entrambe le immagini per esaltarne i bordi verticali. Le DSI risultano perciò sparse perché oltre all'assenza di informazioni alcune zone possono anche risultare occluse. Alle finestre che non trovano riscontro viene associato un valore predefinito di disparità sconosciuta.

Ad ogni valore di disparità è associato nella DSI un determinato colore a cui corrisponde una determinata distanza.

Analizzando un'immagine di disparità, fig. A.13, è possibile individuare gli ostacoli presenti nella scena [42, 41, 40, 16]. Infatti tutti i pixel di un oggetto verticale presentano la stessa disparità. Una ricerca di zone verticali connesse con la stessa disparità rappresenta un semplice *obstacle detector* per ostacoli verticale, fig. A.13.

Bibliografia

- [1] Rapporto annuale INAIL 2011. 2012-07-10.
- [2] EN IEC 61508: Sicurezza funzionale dei sistemi di sicurezza elettrici, elettronici ed elettronici programmabili (E/E/PE).
- [3] VIAC <http://viac.vislabs.it>.
- [4] 2008. UNI EN ISO 13849-1: Sicurezza del macchinario - Parti dei sistemi di comando legate alla sicurezza - Parte 1: Principi generali per la progettazione.
- [5] Y. Benezeth, P. Jodoin, B. Emile, H. Laurent, and C. Rosenberger. Review and evaluation of commonly-implemented background subtraction algorithms. In *ICPR 2008 19th International Conference on Pattern Recognition*, pages 1–4, Dec. 2008.
- [6] M. Bertozzi, E. Binelli, A. Broggi, and M. Del Rose. Stereo Vision-based approaches for Pedestrian Detection. In *Procs. Intl. IEEE Wks. on Object Tracking and Classification in and Beyond the Visible Spectrum*, San Diego, USA, June 2005.
- [7] M. Bertozzi, L. Bombini, A. Broggi, M. Buzzoni, E. Cardarelli, S. Cattani, P. Cerri, A. Coati, S. Debattisti, A. Falzoni, R. I. Fedriga, M. Felisa, L. Gatti, A. Giacomazzo, P. Grisleri, M. C. Laghi, L. Mazzei, P. Medici, M. Panciroli, P. P. Porta, P. Zani, and P. Versari. VIAC: an Out of Ordinary Experiment. In *Procs. IEEE Intelligent Vehicles Symposium 2011*, pages 175–180, Baden Baden, Germany, June 2011. ISSN: 1931-0587.

-
- [8] M. Bertozzi, A. Broggi, E. Cardarelli, R. I. Fedriga, L. Mazzei, and P. P. Porta. VIAC Expedition Toward Autonomous Mobility. *Robotics and Automation Magazine*, 18(3):120–124, Sept. 2011. ISSN: 1070-9932.
- [9] M. Bertozzi, A. Broggi, and A. Fascioli. An extension to the Inverse Perspective Mapping to handle non-flat roads. In *Procs. IEEE Intelligent Vehicles Symposium '98*, pages 305–310, Stuttgart, Germany, Oct. 1998.
- [10] M. Bertozzi, A. Broggi, and A. Fascioli. Stereo Inverse Perspective Mapping: Theory and Applications. *Image and Vision Computing Journal*, 8(16):585–590, 1998.
- [11] M. Bertozzi, A. Broggi, P. Medici, P. P. Porta, and A. Sjögren. Stereo Vision-Based Start-Inhibit for Heavy Goods Vehicles. In *Procs. IEEE Intelligent Vehicles Symposium 2006*, pages 350–355, Tokyo, Japan, June 2006.
- [12] L. Bombini, A. Bontempi, M. Buzzoni, and C. Terenzi. Sistema e procedimento di supervisione, July 2012. Italian Patent Office, RI2012A000036, Publication date: 2012-07-09.
- [13] L. Bombini, M. Buzzoni, M. Felisa, and P. Medici. Sistema per il Controllo di Porte Automatiche, Mar. 2010. CCIAA di Milano, n. domanda MI2010A000460.
- [14] A. Broggi, L. Bombini, C. Stefano, P. Cerri, and R. I. Fedriga. Sensing requirements for a 13,000 km intercontinental autonomous drive. In *Procs. IEEE Intelligent Vehicles Symposium 2010*, pages 500–505, La Jolla, CA, USA, June 2010.
- [15] A. Broggi, M. Buzzoni, M. Felisa, and P. Zani. Stereo obstacle detection in challenging environments: the VIAC experience. In *Procs. IEEE/RSJ Intl. Conf. on Intelligent Robots and Systems*, pages 1599–1604, San Francisco, California, USA, Sept. 2011.

- [16] A. Broggi, C. Caraffi, R. I. Fedriga, and P. Grisleri. Obstacle Detection with Stereo Vision for off-road Vehicle Navigation. In *Procs. Intl. IEEE Wks. on Machine Vision for Intelligent Vehicles*, San Diego, USA, June 2005.
- [17] A. Broggi, C. Caraffi, P. P. Porta, and P. Zani. The Single Frame Stereo Vision System for Reliable Obstacle Detection used during the 2005 Darpa Grand Challenge on TerraMax. In *Procs. IEEE Intl. Conf. on Intelligent Transportation Systems 2006*, pages 745–752, Toronto, Canada, Sept. 2006.
- [18] A. Broggi, P. Medici, E. Cardarelli, P. Cerri, A. Giacomazzo, and N. Finardi. Development of the control system for the VisLab Intercontinental Autonomous Challenge. In *Procs. IEEE Intelligent Transportation Systems 2010*, pages 635–640, 2010. ISSN: 2153-0009.
- [19] A. Broggi, P. Medici, and P. P. Porta. StereoBox: a Robust and Efficient Solution for Automotive Short Range Obstacle Detection. *EURASIP Journal on Embedded Systems – Special Issue on Embedded Systems for Intelligent Vehicles*, June 2007. ISSN 1687-3955.
- [20] A. Broggi, P. Medici, P. P. Porta, and M. Posterli. Short Range Sensing for Intelligent Vehicle: the Experience of the DARPA Grand Challenge. In *Procs. 15th World Congress on ITS*, New York, USA, Nov. 2008.
- [21] A. Broggi, P. Medici, P. Zani, A. Coati, and M. Panciroli. Autonomous vehicles control in the VisLab Intercontinental Autonomous Challenge. *Annual Reviews in Control*, 36(1):161–171, 2012. ISSN: 1367-5788.
- [22] C. Caraffi. Stabilizzazione di immagini ed individuazione di ostacoli mediante visione artificiale stereo in ambienti non strutturati. Master’s thesis, Università degli Studi di Parma - Facoltà di Ingegneria, Oct. 2004.
- [23] L. Christensson, L. Holmqvist, and H. Larne. Sensor arrangements, systems and method in relation to automatic door openers, June 2004. European Patent Office, Publication nr. EP1633950A1, Publication date: 2006-03-15.

- [24] D. Claus and A. Fitzgibbon. A rational function lens distortion model for general cameras. In *Procs. Conf. on Computer Vision and Pattern Recognition*, volume 1, pages 213–219, San Diego, USA, June 2005.
- [25] I. J. Cox, S. L. Hingorani, and S. B. Rao. Fast obstacle detection for urban traffic situations. *Comput. Vis. Image Understand*, 63:542–567, 1996.
- [26] F. Devernay and O. D. Faugeras. Straight Lines have to be Straight. *Machine Vision Application*, 13(1):14–24, 2001.
- [27] M. Felisa and P. Zani. Incremental Disparity Space Image computation for automotive applications. In *Procs. IEEE/RSJ Intl. Conf. on Intelligent Robots and Systems*, St.Louis, Missouri, USA, Oct. 2009.
- [28] S. Gehrig and C. Rabe. Real-time semi-global matching on the cpu. In *ECVW10*, pages 85–92, 2010.
- [29] S. K. Gehrig, F. Eberli, and T. Meyer. A real-time low-power stereo vision engine using semi-global matching. In *Proceedings of the 7th International Conference on Computer Vision Systems: Computer Vision Systems*, ICVimages/Capitolo4S '09, pages 134–143, Berlin, Heidelberg, 2009. Springer-Verlag.
- [30] I. Haller, C. Pantilie, F. Oniga, and S. Nedeveschi. Real-time semi-global dense stereo solution with improved sub-pixel accuracy. In *Procs. IEEE Intelligent Vehicles Symposium 2010*, pages 369–376, San Diego, CA, USA, June 2010.
- [31] R. W. Hamming. Error detecting and error correcting codes. *Bell System Technical Journal*, 29(2):147–160, 1950.
- [32] R. I. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, ISBN: 0521540518, second edition, 2004.
- [33] H. Hirschmüller. Accurate and Efficient Stereo Processing by Semi-Global Matching and Mutual Information. In *Intl. Conf. on Computer Vision and Pattern*

- Recognition*, volume 2, pages 807–814, San Diego, CA, USA, June 2005. IEEE Computer Society.
- [34] H. Hirschmüller and I. Ernst. Mutual information based semi-global stereo matching on the gpu. In *ISVC (1) 08*, pages 228–239, 2008.
- [35] H. Hirschmüller and S. Gehrig. Stereo matching in the presence of sub-pixel calibration errors. In *Intl. Conf. on Computer Vision and Pattern Recognition*, pages 437–444, Miami, FL, USA, 2009.
- [36] H. Hirschmüller and D. Scharstein. Evaluation of stereo matching costs on images with radiometric differences. *PAMI*, 31(9):1582–1599, September 2009.
- [37] K. Hoshide, M. Tobita, and A. Sato. Automatic door apparatus, 2004. European Patent Office, Publication nr. EP1681424A1, Publication date: 2011-07-26.
- [38] S. S. Intille and A. F. Bobick. A maximum likelihood stereo algorithm. In *Proc. ECCV'94, Stockholm, Sweden*, pages 179–186, 1994.
- [39] B. Jahne. *Digital Image Processing*. Springer, Berlin, Germany, fifth edition, Sept. 2002.
- [40] R. Labayrade and D. Aubert. Robust and fast stereovision based road obstacles detection for driving safety assistance. In *Machine Vehicle Application 2002, Nara, Japan*, Dec. 2002.
- [41] R. Labayrade and D. Aubert. In-vehicle obstacles detection and characterization by stereovision. In *IEEE In Vehicle Cognitive Computer Vision Systems, Graz, Austria*, pages 13–19, Apr. 2003.
- [42] R. Labayrade, D. Aubert, and J.-P. Tarel. Real Time Obstacle Detection in Stereo Vision on non Flat Road Geometry through “V-Disparity” Representation. In *Procs. IEEE Intelligent Vehicles Symposium 2002, Paris, France, June 2002*.
- [43] B. Lo and S. Velastin. Automatic congestion detection system for underground platforms. In *Proceedings of 2001 International Symposium on Intelligent Multimedia, Video and Speech Processing*, pages 158–161, 2001.

- [44] H. A. Mallot, H. H. Bülthoff, J. J. Little, and S. Bohrer. Inverse perspective mapping simplifies optical flow computation and obstacle detection. *Biological Cybernetics*, 64:177–185, 1991.
- [45] R. Manduchi, A. Castano, A. Talukder, and L. Matthies. Obstacle detection and terrain classification for autonomous off-road navigation. *Auton. Robots*, 18(1):81–102, 2005.
- [46] S. Nedevschi, R. Danescu, R. Schmidt, and T. Graf. High accuracy stereovision system for far distance obstacle detection. In *Procs. IEEE Intelligent Vehicles Symposium 2004*, Parma, Italy, June 2004.
- [47] S. Nichani and D. Schatz. Method and apparatus for monitoring a passageway using 3D images, Dec. 2003. United States Patent Office, Patent nr. 7397929, Publication date: 2008-07-8.
- [48] N. R. D. Norman R. Draper, H. Smith. *Applied Regression Analysis*. Wiley-Interscience, Milano, Italy, third edition, 1998.
- [49] Y. Ohta and T. Kanade. Stereo by intra- and inter-scanline search using dynamic programming. *IEEE Trans. Pattern Anal. Machine Intell.*, 7, 1985.
- [50] A. Prati, I. Mikic, M. M. Trivedi, and R. Cucchiara. Detecting moving shadows: Algorithms and evaluation. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 25(7):918–923, July 2003.
- [51] W. K. Pratt. *Digital Image Processing*. Addison-Wesley, Milano, Italy, third edition, May 2001.
- [52] J. Ratcliffe. Video surveillance of public places. *U.S. Dept. of Justice, Office of Community Oriented Policing Services, Washington, D.C.*, 2005.
- [53] Roger Y. Tsai. A Versatile Camera Calibration Technique for High-accuracy 3D Machine Vision Metrology Using off-the-shelf TV Cameras and Lenses. *IEEE Journal of Robotics and Automation*, 3:323–344, Aug. 1987.

-
- [54] M. Sai and K. Tsutsumi. Automatic door opening and closing system, 1995. United States Patent Office, Publication nr. 5583405, Publication date: 1996-12-10.
- [55] V. A. Spinelli. Automatic door assembly with video imaging device, Dec. 2010. United States Patent Office, Patent nr. 7042492, Publication date: 2006-05-09.
- [56] P. Steingrube, S. K. Gehrig, and U. Franke. Performance evaluation of stereo algorithms for automotive applications. In *Proceedings of the 7th International Conference on Computer Vision Systems, ICVS '09*, pages 285–294, Berlin, Heidelberg, 2009. Springer-Verlag.
- [57] Z. Wang and W. C. Kastilahn. System and method for detecting, tracking and counting human objects of interest, June 2011. United States Patent Office, Patent nr. 8238607, Publication date: 2012-08-07.
- [58] A. Wedel, H. Badino, C. Rabe, H. Loose, U. Franke, and D. Cremers. B-spline modeling of road surfaces with an application to free-space estimation. *Intelligent Transportation Systems, IEEE Transactions on*, 10(4):572–583, 2009.
- [59] C. Wöhler, W. Progscha, L. Krüger, D. Döttling, and M. Wendler. Method and device for safeguarding a hazardous area, June 2010. United States Patent Office, Patent nr. 7729511, Publication date: 2005-06-01.
- [60] Y. Yang, A. Yuille, and J. Lu. Local, global, and multilevel stereo matching. In *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, pages 274–279, 1993.
- [61] R. Zabih and J. Woodfill. Non-parametric local transforms for computing visual correspondence. In *ECCV (2)*, pages 151–158, 1994.
- [62] P. Zani. Algoritmi ottimizzati per la localizzazione di ostacoli in ambienti non strutturati mediante visione stereo. Master's thesis, Università degli Studi di Parma - Facoltà di Ingegneria, Sept. 2005.

Ringraziamenti

Un sincero ringraziamento ad Alberto per avermi dato questa splendida possibilità.
Grazie a tutti i ragazzi del VisLab, con i quali ho condiviso in questi anni esperienze lavorative ma soprattutto bei momenti di vita.

Grazie a tutti gli amici, con i quali ci si diverte sempre come un tempo.

Un affettuoso ringraziamento a mia sorella, a mio papà ed a tutta la famiglia.

Grazie di cuore a Serena, l'amore della mia vita.

La dedica di questa tesi ed un caro pensiero a mia mamma.

“Il ricordo non ha aggettivi, solo emozioni.”