

UNIVERSITÀ DEGLI STUDI DI PARMA

Dottorato di Ricerca in Tecnologie dell'Informazione

XXI Ciclo

**TECNICHE IBRIDE BASATE SU
MOVIMENTO E STEREOCOPIA
PER IL RICONOSCIMENTO OSTACOLI
IN AMBIENTE AUTOMOBILISTICO**

Coordinatore:

Chiar.mo Prof. Carlo Morandi

Tutor:

Chiar.mo Prof. Alberto Broggi

Dottorando: *Paolo Medici*

Gennaio 2009

“Vision is the art of seeing what is invisible to others. ”

“La visione è l’arte di vedere cose invisibili ad altri.”

Jonathan Swift

Sommario

1	Introduzione	1
1.1	Motivazioni della ricerca	2
1.2	Il Progetto Europeo PREVENT-APALACI	4
1.3	Maggiori contributi di questa ricerca	7
1.4	Struttura della tesi	8
2	Calibrazione e dipendenza dai parametri	9
2.1	Calibrazione per veicoli	10
2.1.1	Dalla DLT alla matrice omografica	10
2.2	Rimozione della distorsione	12
2.3	Propagazione degli errori e dipendenza dai parametri	13
2.3.1	Estrazione parametri in presenza di rumore sui punti	15
2.3.2	Estrazione parametri in presenza di errori sistematici	22
2.3.3	Propagazione degli errori di calibrazione alla misura delle distanze	28
2.3.4	Effetti della distorsione	28
2.3.5	Calibrazione usando l'omografia	30
3	Inverse Perspective Mapping	39
3.1	L'equazione base dell'inversione prospettica	40
3.2	Le proprietà dell'errata riproiezione nell'IPM	42
3.3	Estensione a più camere	45

3.4	Estensione al <i>motion stereo</i>	47
3.5	Problemi del matching	48
4	Modello puntiforme del veicolo	51
4.1	Raggio di curvatura	55
4.2	Cambio di sistema di riferimento	55
4.2.1	Spostamento usando un altro sistema di riferimento	57
4.2.2	Spostamento usando dati relativi	58
5	Ground Plane Stereo IPM	61
5.1	Ground Plane Stereo	62
5.2	Calibrazione	66
5.2.1	Calibrazione con omografia e modello della lente	67
5.2.2	Calibrazione con spline	69
5.3	Implementazione	70
5.3.1	Filtro passa basso	72
5.3.2	Funzione di Accumulazione Polare	73
5.4	Risultati	75
5.4.1	Disparità su immagini IPM	78
5.5	conclusioni	80
5.5.1	Sviluppi futuri	81
6	Motion Stereo IPM	83
6.1	Stato dell'Arte	84
6.2	Problematiche	90
6.2.1	Ego-Motion	90
6.2.2	Warped Image	91
6.3	Pitch Detection	92
6.3.1	Analisi preliminare	94
6.3.2	Implementazione	96
6.4	IPM per Motion Stereo	102
6.5	Fusione dati per riconoscimento pedoni e veicoli	104

Sommario	iii
6.6 Conclusioni	106
7 Lateral Motion Stereo	107
7.1 Ricostruzione Tridimensionale e V-Disparity	107
7.1.1 Estensione al <i>Motion Stereo</i>	111
7.2 Calibrazione	112
7.3 Implementazione	112
7.4 Conclusioni	118
7.4.1 Sviluppo futuri	118
8 Discussione generale	119
8.1 Contributo di questa tesi	119
A Il modello Pin-Hole	121
A.1 Modello della telecamera	121
A.1.1 Modello della distorsione	125
A.1.2 Coordinate Mondo e Coordinate Camera	127
A.1.3 La matrice di rotazione	128
A.2 Rettificazione	130
A.2.1 Coordinate omogenee	131
A.3 Trasformazioni omografiche	132
B Matrice Pseudo-Inversa e fattorizzazione SVD	135
Bibliografia	137
Ringraziamenti	143

Elenco delle figure

1.1	Lo scenario descritto dal progetto PREVENT	5
2.1	Griglia di Calibrazione	15
2.2	Estrazione Yaw, Pitch, Roll: errore sugli angoli	17
2.3	Estrazione Yaw, Pitch, Roll: errore metrico	18
2.4	Estrazione Yaw, Pitch, Roll: dipendenza dal numero di marker	19
2.5	Estrazione Yaw, Pitch, Roll, X, Y: XY	19
2.6	Estrazione Yaw, Pitch, Roll, X, Y: angoli	20
2.7	Estrazione Yaw, Pitch, Roll, α_u , α_v	21
2.8	Confronto tra diverse funzioni usate nel minimizzatore: errore medio dell'angolo di pitch	22
2.9	Estrazione Yaw, Pitch, Roll con errore sulla coordinata Z_0 : angoli	23
2.10	Estrazione Yaw, Pitch, Roll con errore sulla coordinata Z_0 : metri	24
2.11	Estrazione Yaw, Pitch, Roll con errore sulla coordinata X_0	24
2.12	Estrazione Yaw, Pitch, Roll con errore sulla coordinata Y_0	25
2.13	Estrazione Yaw, Pitch, Roll con errore sulla coordinata u_0	26
2.14	Estrazione Yaw, Pitch, Roll con errore sulla coordinata v_0	26
2.15	Estrazione Yaw, Pitch, Roll con errore sulla coordinata α	27
2.16	Errori sulle distanze dei punti di test in caso di errore sull'orientazione	29
2.17	Errori sulle distanze dei punti di test in caso di errore sulla coordinata Z_0	30

2.18 Errori sulle distanze dei punti di test in caso di errore sul punto principale	31
2.19 Errori sulle distanze dei punti di test in caso di errore su α_u	32
2.20 Variazione della posizione dei punti immagine a causa della distorsione radiale	32
2.21 Estrazione di yaw, pitch, roll nel caso in cui i punti della griglia di calibrazione siano affetti da errore di distorsione	33
2.22 Errore sulle distanze nel mondo a seguito dell'estrazione di yaw, pitch, roll su punti della griglia affetti da errore di distorsione	33
2.23 Esempio della dispersione dei punti	34
2.24 Calibrazione con omografia in presenza di errore di individuazione dei punti	34
2.25 Calibrazione con omografia minimizzata in coordinate immagine in presenza di errore di individuazione dei punti	36
2.26 Calibrazione con omografia minimizzata in coordinate mondo in presenza di errore di individuazione dei punti	36
3.1 Errata proiezione di punti con $Z \neq 0$	42
4.1 Veicolo in spostamento e rotazione	52
4.2 Traiettoria generata dal modello del veicolo puntiforme, sovrapposizione con una mappa (<i>a</i>) e generazione di una vista dell'ambiente percorso (<i>b</i>)	54
4.3 Sistema di riferimento	55
4.4 Relazione tra i due sistemi di riferimento	55
4.5 Proiezione di \vec{p}	56
4.6 Spostamento del sistema di riferimento con dati assoluti	57
4.7 Spostamento del sistema di riferimento con dati relativi	58
5.1 Esempio di casistiche analizzate per il sistema <i>Ground Plane Stereo</i>	62
5.2 Esempio di <i>Ground Plane Stereo</i>	63
5.3 Calibrazione con omografia e modello della lente	68

5.4	Calibrazione con generazione di funzioni spline	69
5.5	Diagramma di Flusso del sistema Ground Plane Stereo IPM	70
5.6	Elaborazione IPM per immagini Stereo	71
5.7	Ground Plane Stereo	71
5.8	Costruzione incrementale dell'immagine del valor medio	73
5.9	Esempio di creazione del diagramma polare	74
5.10	Funzione di Accumulazione Polare	75
5.11	Curve ROC per il filtro di pre-processamento dell'immagine differenziale	76
5.12	Curve ROC per la dimensione della finestra nel filtro passa basso	76
5.13	Esempio di risultati prodotti dal sistema <i>Ground Plane Stereo IPM</i>	77
5.14	Esempio di ricostruzione tridimensionale usando la disparità su IPM. Coordinata X e coordinata Z	79
5.15	Esempi di installazioni del sistema <i>Ground Plane Stereo IPM</i>	80
6.1	Il veicolo fornito dal Centro Ricerche Fiat usato nel progetto	84
6.2	Esempio di ambienti analizzati per il sistema <i>Motion Stereo</i>	85
6.3	Esempio di rilevazione di ostacoli basato su <i>Warped Image</i>	88
6.4	Individuazione del punto di fuga partendo da 3 rette non affette da rumore	94
6.5	Immagine del prodotto scalare tra il gradiente e la direttrice del punto verso il punto di fuga originale.	95
6.6	Diagramma di flusso dell'algoritmo di Pitch Detection	96
6.7	Esempio del risultato della binarizzazione degli edge	97
6.8	Esempio del risultato del <i>pitch detector</i>	98
6.9	Esempio di rilevazione della coordinata v_{vp} del vanishing point sulla sequenza 33	99
6.10	Esempio di rilevazione della coordinata v_{vp} del vanishing point sulla sequenza 19	100
6.11	Esempio di rilevazione della coordinata v_{vp} del vanishing point sulla sequenza 19 usando il tracking dell'orizzonte	100

6.12	Principio base del <i>Motion Stereo</i>	101
6.13	Diagramma di Flusso del sistema Motion Stereo	102
6.14	Esempio di elaborazione del <i>Motion Stereo</i>	103
6.15	Diagramma di flusso del sistema APALACI <i>Motion Stereo</i>	104
6.16	Alcuni risultati del sistema APALACI <i>Motion Stereo</i>	105
7.1	Principio base del <i>Lateral Motion Stereo</i>	108
7.2	V-Disparity nel <i>Lateral Motion Stereo</i>	111
7.3	Diagramma di Flusso del sistema <i>Lateral Motion Stereo</i>	112
7.4	Fotogrammi di esempio nel caso <i>Lateral Motion Stereo</i>	113
7.5	Rimozione della distorsione	113
7.6	Immagine dedistorta filtrata con Sobel	114
7.7	Disparità nel <i>Lateral Motion Stereo</i>	115
7.8	Esempio di proiezione nello spazio (X,Y)	116
7.9	Esempio di individuazione ‘spazio libro’ basato su <i>Lateral Motion Stereo</i>	117
A.1	Sistema di coordinate mondo	122
A.2	Sistema di coordinate camera	122
A.3	Sistema di coordinate immagine	122

Capitolo 1

Introduzione

L'obiettivo che si propone questa tesi è quello di descrivere una classe di sistemi atti allo scopo di ricostruire l'ambiente circostante un veicolo sfruttando sia il moto che la visione monoculare e stereoscopica. La natura di tale ricostruzione tuttavia non sarà quella di essere rigorosa ma si cercherà di ottenere invece quella minima sufficiente all'interno delle specifiche del progetto. La scelta di eseguire una ricostruzione approssimata viene dalla necessità di realizzare sistemi a basso peso computazionale, cercando una elevata semplicità di installazione e limitando il numero di camere ma sfruttando, quando disponibile, anche l'informazione del movimento del veicolo.

Le sfide principali che si devono considerare nel momento in cui si sviluppa un sistema di individuazione di ostacoli in ambiente automobilistico solo che tale algoritmo deve funzionare in tempo reale con un tempo di reazione minimo, gli ostacoli devono essere riconosciuti a sufficiente distanza per avvisare il guidatore o, in maniera autonoma, eseguire azioni di contenimento dei danni da impatto. Infine le posizioni e le dimensioni orizzontali degli ostacoli devono essere stimate con cura per permettere una eventuale azione del veicolo nel caso la frenata di emergenza non fosse più sufficiente.

I principali vincoli presi in considerazione durante lo studio e lo sviluppo di questa tesi sono stati principalmente:

- il sistema deve essere installato su un veicolo (autoveicolo o veicolo commer-

ciali);

- la parte di installazione e calibrazione deve essere allo stesso tempo rigorosa ma possibilmente automatizzabile;
- l'utilizzo di un hardware limitato in modo da permettere l'abbassamento dei costi e una diffusione più capillare del sistema.

All'interno di questa tesi verranno mostrati 3 diversi sistemi, tutti basati su visione artificiale ma ognuno con un elemento di diversità:

- un sistema stereoscopico che sfrutta l'immagine prospettica inversa per individuare gli ostacoli;
- un sistema monoculare che sfrutta l'immagine prospettica inversa e il movimento;
- un sistema monoculare che simula un sistema stereoscopico attraverso l'informazione del movimento.

Questi sistemi così diversi sono stati installati e sperimentati su veicoli stradali, ed è stata valutata la loro capacità di individuare in primo luogo regioni di interesse dove la presenza di ostacoli è altamente probabile.

1.1 Motivazioni della ricerca

Le motivazioni che spingono la ricerca in questo settore si possono di fatto esprimere in maniera quantitativa. Di fatto ogni giorno in Italia si verificano in media 633 incidenti stradali, che provocano la morte di 14 persone e il ferimento di altre 893. Nel complesso, fonte ISTAT, nel solo anno 2007 sono stati rilevati 230.871 incidenti stradali, che hanno causato il decesso di 5.131 persone, mentre altre 325.850 hanno subito lesioni di diversa gravità.

Per capire dove andare ad agire in maniera più decisa è necessario entrare leggermente in dettaglio sulle dinamiche di tali incidenti. Si nota che la maggior parte

degli incidenti stradali avviene tra due o più veicoli (76,8%) e solo il 23,2% a veicoli isolati. Nell'ambito degli incidenti tra veicoli la tipologia di incidente più diffusa è lo scontro frontale-laterale (82.841 casi) con 1.371 morti e 119.800 feriti, seguita dal tamponamento, che registra 42.506 casi con 502 morti e 69.093 persone ferite. Tra gli incidenti a veicoli isolati la fuoriuscita o sbandamento del veicolo rappresenta il caso più diffuso (20.963 incidenti) con 990 morti e 26.191 feriti. L'investimento di pedone rappresenta l'8% degli incidenti: si registrano 18.368 casi in cui hanno perso la vita 571 persone e 20.937 sono rimaste ferite (fonte ISTAT).

Il costo sociale in termine di danni materiali, invalidità permanenti più o meno gravi, costi sanitari, amministrativi, giudiziari è stato stimato per il solo anno 2005 in 34.733 milioni di euro (fonte ACI) oltre, prima di tutto, all'elevato prezzo pagato in vite umane.

Attività di ricerca volte a prevenire o quantomeno attenuare i danni di tali incidenti sono sicuramente meritevoli di attenzione. Una classe di sistemi pertanto capaci di ridurre i danni da impatti, evitare le collisioni con altri veicoli e soprattutto utenti vulnerabili della strada (pedoni, ciclisti) e prevenire la fuoriuscita del veicolo dalla carreggiata potrebbero ridurre tali problematiche in maniera considerevole. Questa tesi di fatto cerca di affrontare i primi due problemi sviluppando una classe di sensori che possano fornire informazione a un pianificatore per permettere la riduzione dei danni da impatto ed eventualmente evitare collisioni.

Le motivazioni nascono anche da un discorso puramente legislativo. Infatti entro aprile 2009 sarà obbligatorio su tutti i veicoli commerciali (con peso superiore a 3,5 tonnellate) l'installazione di nuovi specchi grandangolari con un campo visivo ampliato. Già da alcuni anni infatti tali specchi sono obbligatori sui veicoli nuovi, ma ora una legge europea impone che tali ausili siano presenti su tutti i veicoli in circolazione. Sfruttando questa richiesta normativa, risulterebbe interessante sostituire a specchi esterni, sensori considerati passivi, a sistemi che avvertono attivamente l'autista della presenza di una eventuale situazione di pericolo. Oltre a questo genere di sistemi per veicoli commerciali, dal 2013 in Europa saranno infine richiesti una nuova classe di sensori denominati *Advance Emergency Braking*(AEBS) che avvertano il guidatore quando è troppo vicino al veicolo di fronte, e in determinate situazioni,

tali sistemi abbiano la possibilità di agire direttamente con una frenata di emergenza per prevenire o quantomeno ridurre le conseguenze di una collisione.

Di fatto, per questo genere di applicazioni, il sensore visivo appare particolarmente adatto, permettendo la segmentazione di scene complesse, grazie alla ricchezza delle informazioni fornite da ogni singolo fotogramma.

Per quanto riguarda il monitoraggio delle aree cieche nei veicoli, c'è da notare che lo stesso risultato si potrebbe ottenere in maniera automatica grazie all'uso di alcuni tipi di sensori come il radar (performante) o il sonar (economici), tuttavia tali oggetti tuttora in commercio presentano alcuni svantaggi.

In particolare il sonar manifesta problemi di *Cross Talk* tra i diversi dispositivi montati sullo stesso veicolo e su veicoli adiacenti. Questo problema viene risolto usando diverse frequenze o forme d'onda per ogni dispositivo, ma tale procedimento obbliga un incremento della potenza del processore del segnale per rimuovere gli eco che si vengono a formare. Contemporaneamente alcune superfici agiscono come specchi per le onde del sonar: siccome l'onda sonora incidente segue la legge della riflessione, le uniche onde riflesse verso il sonar saranno quelle perpendicolari al fronte d'onda. Il sonar poi ha una zona morta, nel senso che possiede una distanza minima entro la quale non può rilevare ostacoli a causa del fatto che quando il sonar è in emissione, il ricevitore deve ignorare ogni segnale. Questa distanza è solitamente compresa tra 15 e 35 centimetri.

Infine, oltre al discorso della sicurezza, è avvertito dall'utilizzatore abituale di autoveicoli anche il tema del comfort. Di fatto una nuova serie di *optional* ad elevato contenuto tecnologico sta per essere proposto in questi anni agli utenti finali della strada, e i sensori basati su visione, grazie al loro basso costo e alla molteplicità di informazioni che possono fornire, sono di grande interesse da parte dei costruttori di autoveicoli.

1.2 Il Progetto Europeo PREVENT-APALACI

Un discorso introduttivo va sicuramente rivolto al progetto europeo *PREVENT* e al suo sotto-progetto *APALACI Advanced Pre-crash And Longitudinal Collision mIti-*

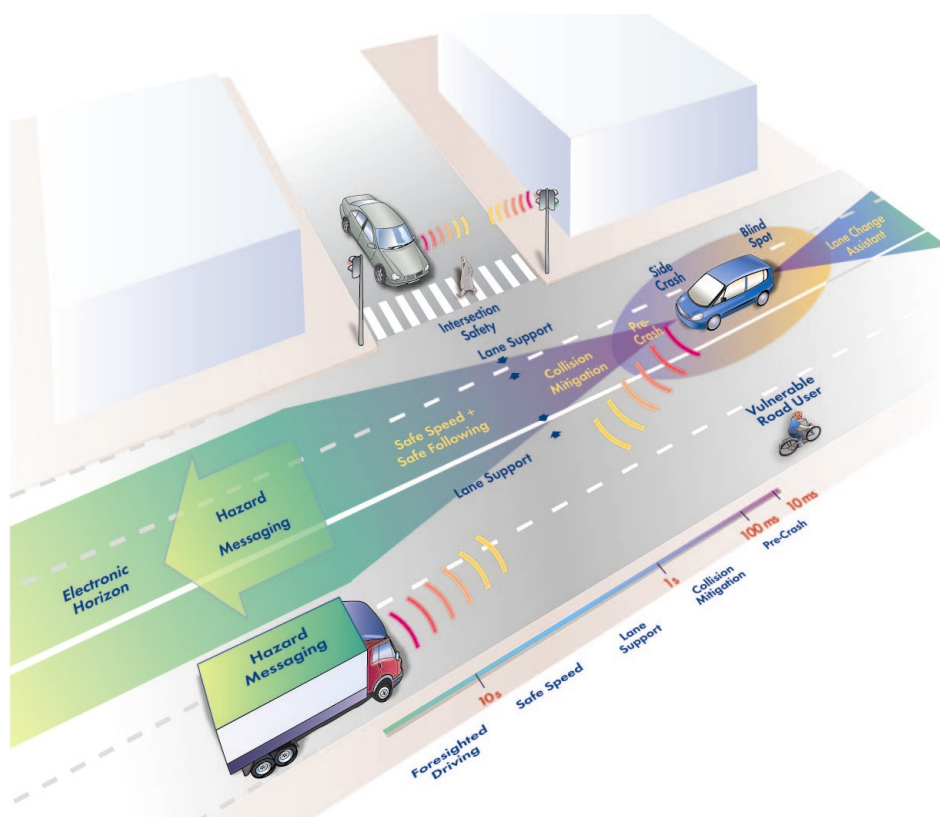


Figura 1.1: Lo scenario descritto dal progetto PREVENT



gation system with pedestrian classification ability per i quali gran parte del lavoro di questa tesi è stato sviluppato.

L'obiettivo di *PReVENT* è sviluppare una infrastruttura per aiutare i guidatori a evitare incidenti, cercando di percepire automaticamente l'ambiente intorno al veicolo e, a seconda della natura del pericolo, avvisare il conducente il prima possibile e, se esso non reagisce, assisterlo attivamente.

APALACI si propone di sviluppare una tecnologia sia di pre-collisione che di mitigazione della collisione, favorendo lo sviluppo di tecnologie innovative e robuste per la fusione di sensori. Il sistema deve mostrare di essere capace di prevenire gli incidenti a bassa velocità verso pedoni e mitigare i danni delle collisioni inevitabili agendo sui sistemi di contenimento (cinture di sicurezza, airbag) per aumentare la protezione degli occupanti della vettura.

I sistemi sviluppati devono essere pensati per poter riconoscere tra 0 m e 50 m veicoli, infrastrutture, oggetti fissi e ostacoli sensibili, usando radar, laser scanner, camere mono/stereo e sensori a ultrasuoni.

L'obiettivo di *APALACI* è quello di dimostrare che al giorno d'oggi i sistemi informatici sono già capaci di reagire più velocemente dell'utente umano e permettere l'applicazione di una nuova serie di metodologie nel campo degli ADAS ¹.

Alcuni sotto-progetti del progetto *APALACI*, con cui questa tesi risulta coinvolta, sono stati

- mitigare i danni da collisione attraverso frenata autonoma per ridurre l'energia da impatto;

¹Advanced Driver Assistance Systems

- sviluppare applicazioni in grado di preattivare le cinture di sicurezza e gli airbag per ridurre i tempi di azione;
- evitare che i veicoli commerciali pesanti possano partire quando pedoni sono presenti nell'area cieca del guidatore;
- sviluppare sensori ad alte prestazioni incluso la classificazione in tempo reale di oggetti attraverso innovative fusioni usando informazioni fornite da diverse classi di sensori.

1.3 Maggiori contributi di questa ricerca

Questa tesi recupera e dà un formalismo matematico a una classe di sistemi di riconoscimento ostacoli basati su *Inverse Perspective Mapping* il cui sviluppo è cominciato nei primi anni '90. In questa tesi viene mostrato che questa classe di sistemi è capace di rilevare ostacoli potenziali sfruttando algoritmi innovativi dal basso impatto computazionale.

I maggiori contributi di questa tesi includono:

- sviluppo di un formalismo per il riconoscimento ostacoli basato su immagini *Inverse Perspective Mapping* e la possibilità da questo di permettere una ricostruzione tridimensionale completa basata su immagine di disparità;
- sviluppo di un algoritmo capace di genere *Look Up Table* per la rimozione in un unico passaggio di distorsione e prospettiva;
- sviluppo di un riconoscitore di beccheggio del veicolo robusto, capace di rilevare le variazioni dell'angolo di *pitch* dovuti ai movimenti del veicolo;
- sfruttare un modello cinematico del veicolo basato su velocità e *yaw-rate* istantanei;
- mostrare come sia comunque possibile in alcuni casi limitati (a velocità ridotta) una ricostruzione tridimensionale della scena intorno al veicolo sfruttando solamente una camera e il movimento.

1.4 Struttura della tesi

Nel capitolo 2 verranno spiegati in dettaglio i metodi usati in questo lavoro per calibrare i sensori utilizzati e verrà dato ampio spazio alla discussione di quanto tali parametri siano ricavati correttamente mostrando, in via puramente sperimentale, la propagazione dell'errore di misura. Nel capitolo 3 vengono spiegate le proprietà della *Inverse Perspective Mapping*, trasformazione centrale in questo lavoro di tesi, con una ampia discussione sulle sue proprietà e viene mostrato il risultato notevole di poter fornire una ricostruzione tridimensionale della scena vista da una o più camere. Nel capitolo 4 viene esposto il modello approssimato del moto del veicolo che viene sfruttata all'interno dei diversi algoritmi.

In seguito si affrontano tre diverse tecniche di Visione Artificiale per il riconoscimento di ostacoli. Infatti all'interno di questo lavoro si è di fatto formato una specie di triangolo tra le parole *Stereo*, *IPM* e *Movimento*. Nel capitolo 5 viene mostrata una tecnica di riconoscimento ostacoli basata su visione Stereo e immagini IPM. Nel capitolo 6 viene mostrata invece una tecnica basata ancora su immagini IPM ma monoculare e che sfrutta il movimento della camera per individuare potenziali ostacoli. Infine nel capitolo 7 viene mostrata nuovamente una tecnica che sfrutta ancora il movimento di una sola camera, ma per simulare una visione stereoscopica.

In appendice A viene riportato il formalismo matematico del modello della *pin-hole camera* usato per proiettare i punti da mondo a immagine, e nell'appendice B la tecnica di risoluzione lineare di matrici sovradimensionate usata nel testo.

Capitolo 2

Calibrazione e dipendenza dai parametri

In tutti i sistemi di visione la calibrazione della videocamera assume un ruolo centrale per poter ottenere una ricostruzione tridimensionale accurata. Anche se il concetto di ricostruzione approssimata è centrale in questo lavoro, risulta sicuramente corretto eseguire una trattazione rigorosa della calibrazione per veicoli e, per dimensionare la precisione con cui il sistema può operare, mostrare la propagazione dell'errore dei parametri.

Essendo le tecniche di ricostruzione discusse in questa tesi molto diverse tra loro, si è cercato di uniformare comunque le tecniche di calibrazione, proponendo poi caso per caso una calibrazione specifica per il particolare compito.

Si possono distinguere in genere due tipi diversi di calibrazione: un metodo che si preoccupa semplicemente di ottenere gli elementi di una matrice di proiezione che trasferisce i singoli punti tra mondo e immagine (omografica), e un metodo in cui vengono invece estratti valori tangibili del sistema, prima i parametri intrinseci e poi quelli estrinseci (calibrazione detta globale, metrica o Euleriana) della camera.

Come si può facilmente intuire, il primo metodo fornirà risultati più accurati, ma limitati e relativi a una condizione statica, in cui l'orientazione della camera non varia rispetto alla situazione di calibrazione. Al contrario il secondo metodo permette di

usare la formula della proiezione anche quando alcuni parametri della camera siano variati rispetto alle condizioni di calibrazione nel caso in cui si venga a conoscenza del loro nuovo valore. Il secondo metodo tuttavia richiede un maggior numero di parametri estratti con tecniche differenti, parametri affetti singolarmente da errore di misura e quantizzazione.

2.1 Calibrazione per veicoli

In letteratura sono abbastanza diffuse tecniche di calibrazione metriche [1] già nei primi anni in cui il modello della *pin-hole* è stato formalizzato.

Queste discussioni sono concentrate tutte su metodi per calibrare la camera *off-line*, in laboratorio. Per motivi di installazione risulta tuttavia più pratico posizionare la camera e calibrarla direttamente sul veicolo con una lunghezza focale già impostata nelle condizioni operative, solitamente diversa dalla focale usata in laboratorio.

La trattazione del modello della camera *pin-hole* usata in questo capitolo e in generale in questo lavoro è disponibile nell'appendice A.

2.1.1 Dalla DLT alla matrice omografica

Osservando il sistema A.14 risulta evidente come basti ricavare gli 11 parametri lineari della matrice \mathbf{P} (è una matrice 3×4 ma essendo conosciuta a meno di un fattore moltiplicativo si può fissare un parametro) per avere una calibrazione del sistema implicita (ovvero dove non si conoscono i parametri interni) che tuttavia permetta di proiettare qualunque punto da coordinate mondo in coordinate immagine, e invertendo la matrice, conoscere per ogni punto dell'immagine l'equazione della retta del mondo che passa per il *pin-hole* e quel punto.

Per ottenere i parametri della matrice \mathbf{P} sono necessari 6 punti non linearmente dipendenti. Siccome il risultato è generalmente instabile usando solo 6 punti, si sfruttano tecniche come la pseudoinversa (vedi sezione B) per stimare tali valori ai minimi quadrati. Tale tecnica per calcolare la funzione prospettica si chiama DLT (*direct linear transformation*) ed è stata introdotta da Abdel-Aziz e Karara [2] nel 1971.

È possibile sviluppare una calibrazione più semplice che permetta di ricavare la proiezione di un solo piano dell'immagine piuttosto che qualunque coordinata $(X_i, Y_i, Z_i)^T$. Se l'interesse dell'analisi è un solo piano dell'immagine, per semplificare la matrice \mathbf{P} si possono porre tutti i punti che si vogliono usare per la calibrazione sul terreno. Ciò significa porre la condizione $Z_i = 0 \forall i$, che implica l'eliminazione di una colonna della matrice \mathbf{P} .

Nel caso di eliminazione della coordinata Z la matrice si riduce alla dimensione 3×3 e diventa omografica (vedi sezione A.3) in quanto permette di proiettare coordinate omogenee in coordinate omogenee e rappresenta uno spazio di \mathbb{R}^2 , nello specifico un piano (nel mondo) in un altro piano (immagine).

Partendo dall'equazione A.13 con l'ipotesi $Z_i = 0$ possiamo definire una matrice omografica \mathbf{H}_Z come

$$\lambda \begin{pmatrix} u_i \\ v_i \\ 1 \end{pmatrix} = \mathbf{A} \begin{bmatrix} r_0 & r_1 & -t_x \\ r_3 & r_4 & -t_y \\ r_6 & r_7 & -t_z \end{bmatrix} \begin{pmatrix} X_i \\ Y_i \\ 1 \end{pmatrix} = \mathbf{H}_Z \begin{pmatrix} X_i \\ Y_i \\ 1 \end{pmatrix} \quad (2.1)$$

Tale matrice permette di rettificare l'immagine, sintetizzando una visuale fronto-parallela del piano, con una trasformazione dal nome di *rettificazione ortogonale* o *Inverse Perspective Mapping* (si veda la sezione 3.1 per dettagli ulteriori).

Il sistema 2.1 della matrice \mathbf{H}_Z è un sistema, lineare, in 9 incognite più λ . Tuttavia, essendo conosciuta a meno di un fattore moltiplicativo, è possibile fissare un vincolo, per esempio $h_8 = 1$. Sotto questa semplificazione il sistema 2.1 viene svolto e i coefficienti della matrice $\mathbf{H}_Z = h_1, \dots, h_7$ possono venire espressi come:

$$\begin{bmatrix} X_i & Y_i & 1 & 0 & 0 & 0 & -X_i u_i & -Y_i u_i \\ 0 & 0 & 0 & X_i & Y_i & 1 & -X_i v_i & -Y_i v_i \end{bmatrix} \begin{pmatrix} h_0 \\ \vdots \\ h_7 \end{pmatrix} = \begin{pmatrix} u_i \\ v_i \end{pmatrix} \quad (2.2)$$

Questo è un sistema di due equazioni in 8 incognite $h_0 \dots h_7$. Ogni punto di cui si conosce la posizione nel mondo su un determinato piano e nell'immagine fornisce due vincoli, e sono richiesti almeno 4 punti per ottenere i parametri per l'omografia. Ogni vincolo in più permette di ricavare una matrice attraverso una regressione lineare ai minimi quadrati (si veda la sezione B).

2.2 Rimozione della distorsione

Siccome il modello della camera *pin-hole* non funziona con ottiche distorcenti, prima di procedere a qualunque calibrazione, risulta necessario rimuovere la distorsione della lente.

In questo caso la distorsione viene rimossa usando una tecnica del tipo *Straight lines have to be straight* [3]: una linea nell'immagine de-distorta, se il modello della distorsione della lente proposto è corretto, deve essere effettivamente una linea.

Per ottenere questo risultato viene acquisita dalla telecamera una sequenza che inquadra una griglia conosciuta in diverse pose e angolazioni. Attualmente tale griglia è formata da *marker* circolari e viene sfruttato il baricentro nell'immagine di tale *marker*. Il sistema sviluppato è indipendente da come vengono forniti i punti e si può adattare immediatamente a casi in cui la griglia è una scacchiera o a qualunque altro tipo di marcatore.

Dall'immagine di calibrazione vengono estratte le coordinate dei punti che formano la griglia.

Per determinare i coefficienti $\vec{k} = k_1, k_2, \dots, k_n$ dell'ottica gli approcci implementati sono:

1. sfruttare indipendentemente le linee della griglia orizzontali e verticali e forzare che le rette siano rette;
2. sfruttare la griglia nel suo insieme, forzando che un omografia sia una omografia.

In entrambi i casi i punti appartenenti alla griglia vengono de-distorti sfruttando dei parametri ipotetici \vec{k} e un ipotetico centro di distorsione $(x_{0,d}, y_{0,d})$ attraverso la funzione :

$$\begin{cases} x_u = (x_d - x_{0,d})(1 + 1 + k_1 r_d^2 + k_2 r_d^4 + \dots) \\ y_u = (y_d - y_{0,d})(1 + 1 + k_1 r_d^2 + k_2 r_d^4 + \dots) \end{cases} \quad (2.3)$$

dove $r_d = \sqrt{x_d^2 + y_d^2}$. La discussione riguardo la forma di questa funzione si può trovare nella sezione A.1.1. I parametri sono ipotetici nel senso che verranno valutati

da una funzione errore quanto tali valori siano prossimi alla soluzione corretta del problema o meno.

Nel caso si cerchi di minimizzare l'errore su una retta (1), i punti de-distorti dall'equazione 2.3 appartenenti alla stessa retta vengono fatti regredire linearmente ai minimi quadrati. Ogni retta fornisce un indice della bontà dei punti che l'hanno formata attraverso l'informazione sullo scarto e lo scarto complessivo, somma degli scarti di tutte le rette acquisite, è la funzione di errore usata per un minimizzatore non lineare (Nelder–Mead).

Nel caso invece si cerchi di minimizzare l'errore sulla griglia (2), dai punti ottenuti, sempre con la funzione 2.3, viene estratta una omografia (vedi 2.2) per ogni griglia acquisita. Questa volta lo scarto è la misurato sull'intera griglia, e lo scarto complessivo, somma degli scarti forniti da ogni singola griglia, è usato come funzione errore.

Infine, estratti i parametri \vec{k} e $(x_{0,d}, y_{0,d})$ della lente che minimizzano la funzione errore scelta, la funzione 2.3 viene invertita per permettere la generazione di una immagine de-distorta densa, dove a ogni punto (x_u, y_u) corrisponde un unico punto (x_d, y_d) .

2.3 Propagazione degli errori e dipendenza dai parametri

Essendo abbastanza complesso il calcolo della propagazione dell'errore nel caso della proiezione prospettica e volendo eseguire una serie di test puramente qualitativi per capire approssimativamente l'andamento dell'errore e la sua natura è stata usata una configurazione di calibrazione sintetica. È stata usato un set-up usato attualmente nel veicolo di test Grandeur i quali parametri sono mostrati in tabella 2.1. Questi parametri saranno da ritenersi i parametri esatti della camera ai quali il sistema di calibrazione dovrà tendere per ottenere i migliori risultati.

Usando questi parametri di calibrazione sono stati generati 921 punti equidistanti nel mondo 1 m lungo la direzione X e 0.25 m nella direzione Y in un area compresa tra $[-5\text{ m}, 5\text{ m}]$ e X tra $[3\text{ m}, 30\text{ m}]$ e tutti sul piano stradale $Z = 0$. Un immagine della distribuzione di questi punti è mostrata in figura 2.23. Visto che questo lavoro

α_u	0.359
α_v	0.2333
X_0	-1.78 m
Y_0	-0.235 m
X_0	1.23 m
yaw γ	0.017
pitch ϑ	0.001
roll ρ	-0.002
larghezza	640 px
altezza	480 px
u_0	320 px
v_0	240 px

Tabella 2.1: Geometria di riferimento per le prove effettuate

di tesi esamina principalmente punti solo sul piano stradale si è scelto di limitare l'analisi degli errori a solo questo specifico piano. Gli errori su eventuali punti non appartenenti a questo piano sono comunque nello stesso ordine di grandezza di quelli mostrati nella maggior parte delle prove effettuate.

Per i punti usati per la calibrazione, anche se virtuale, si è scelto di usare la medesima configurazione scelta per i veicoli del laboratorio, ovvero una griglia di geometria mostrata in figura 2.1. Tale griglia è formata da 13 coppie di marker, i cui componenti della coppia distano 0.2 m lungo l'asse Y . Nelle prove con 26 marker entrambi i componenti della coppia sono stati usati, nelle prove con 13 marker solo l'elemento destro della coppia viene considerato, mentre nella configurazione a 4 marker solo i 4 punti agli estremi della griglia vengono coinvolti nella fase di calibrazione.

In modo da valutare una specie di caso peggiore, il sistema di calibrazione viene valutato in effetti su un area molto più grande dell'area di calibrazione e perciò tenderà ad accentuare molto gli errori.

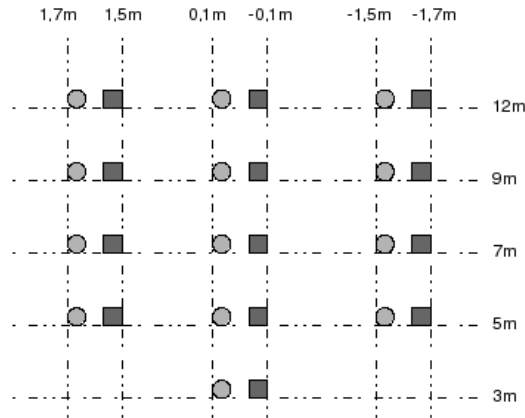


Figura 2.1: Griglia di Calibrazione

2.3.1 Estrazione parametri in presenza di rumore sui punti

Il primo esperimento sugli errori di calibrazione mostra quanto incide l'errore nel riconoscere del *marker* sulla bontà dei parametri estratti.

In questo test viene aggiunta una certa quantità di rumore uniforme a ogni *marker* e da questi punti viene chiesto a un minimizzatore non lineare (in questo caso Nelder e Mead) di fornire la configurazione di alcuni parametri della camera, che si suppongono non conosciuti, in modo da minimizzare l'errore tra i punti in coordinate mondo o immagine.

Sia pertanto un generico punto i della griglia di coordinate mondo $(X_i, Y_i, 0)^T$ non affette pertanto da errore. Il punto i viene proiettato attraverso una funzione $F_{PM}(X, Y, Z)$ (vedi equazione 3.1) in coordinate immagine usando il set-up corretto 2.1 nel punto $(u_i, v_i)^T$. Questo punto è il punto i in coordinate immagine non affetto da errore.

A questo punto infine è associato a questo punto, sempre in coordinate immagine, un punto $(\tilde{u}_i, \tilde{v}_i)^T$ a cui è stato aggiunto rumore uniforme in modo da simulare l'errore di riconoscimento.

Il punto $(\tilde{u}_i, \tilde{v}_i)^T$ viene proiettato in coordinate mondo attraverso una funzione di

IPM usando i parametri della camera questa volta ipotetici, da valutare, in un punto del tipo $(\tilde{X}_i, \tilde{Y}_i, Z_i)^\top$.

Viene fornita una stima dell'errore in coordinate mondo al minimizzatore attraverso la distanza euclidea tra i due punti:

$$e = \sum |(\tilde{X}_i, \tilde{Y}_i, Z_i)^\top - (X_i, Y_i, Z_i)^\top| \quad (2.4)$$

dove Z_i è l'altezza prevista del punto, per cui normalmente $Z_i = 0$.

In questo modo si cerca di fare in modo che i punti proiettati nell'immagine IPM siano il più vicino possibile ai punti della griglia.

È tuttavia possibile usare come funzione errore, l'errore in coordinate immagine. In questo caso vengono fatti variare i parametri della camera che influenzano la funzione F_{PM} fino a che non si trova il minimo della funzione errore

$$e = \sum |F_{PM}(X_i, Y_i, Z_i) - (\tilde{u}_i, \tilde{v}_i)^\top| \quad (2.5)$$

Si può mostrare che i minimi calcolati con i due metodi sono leggermente diversi, tuttavia il minimo ricavato dal primo metodo risulta più preciso per gli scopi di questa tesi, lavorando prevalentemente in coordinate mondo.

I risultati riportati in questa sezione mostrano come varia l'errore del parametro che si vuole estrarre, rispetto al suo valore corretto, al variare della quantità di rumore aggiunto.

Essendo il rumore casuale, è stato necessario per ogni livello di rumore generare diverse griglie (per ogni esperimento riportato sono state fatte 1000 esecuzioni) per valutare il valor medio e il massimo degli errori che si possono incontrare attraverso questo tipo di calibrazione.

Nei grafici di figura 2.2 vengono mostrati gli errori medi e massimi, ovvero le differenze in valore assoluto tra il valore ricavato dal minimizzatore degli angoli di orientazione *yaw*, *pitch*, *roll* e quello reale della camera, al variare del rumore uniforme aggiunto ai punti di calibrazione della griglia. In questo esperimento si suppone di conoscere perfettamente tutti gli altri parametri della camera e di avere solo l'incognita su i 3 angoli di orientazione in esame. Tale problema è sicuramente quello

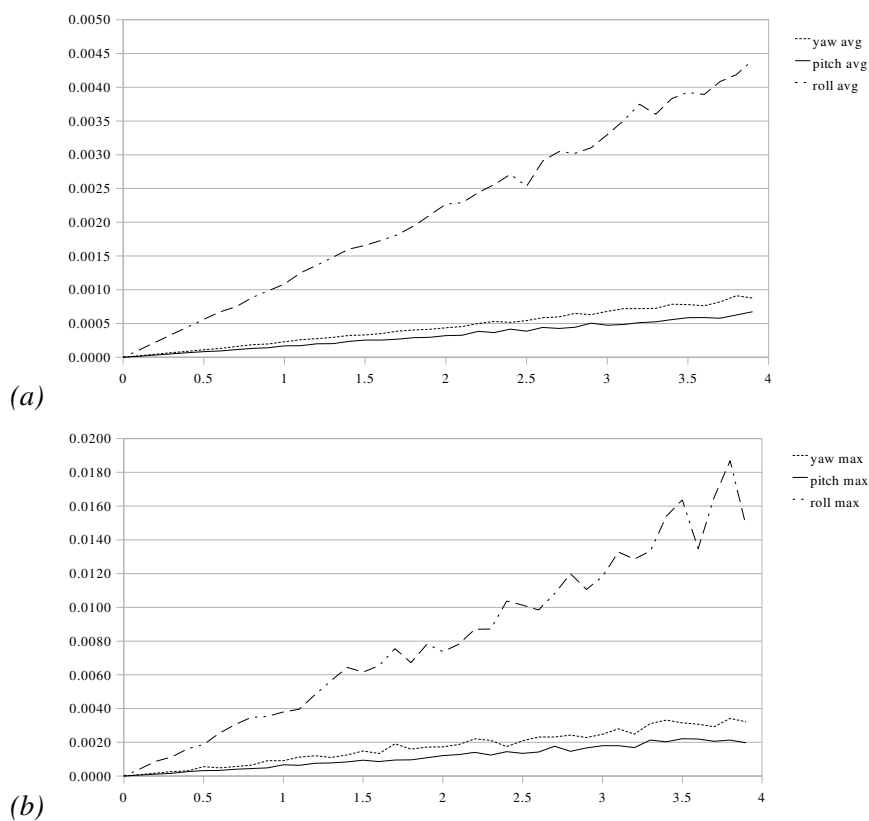


Figura 2.2: Estrazione Yaw, Pitch, Roll da 26 marker affetti da rumore: errore medio (a) e massimo (b) su tali angoli

di maggior interesse in quanto i 3 angoli di orientazione vengono sempre considerati come incognite visto che solitamente sono quelli impossibili da calibrare *offline*.

Da questo esperimento si può mostrare sia che è il *roll* il parametro più sensibile al rumore, ma che poi, dal punto di vista pratico, con un rumore massimo di 1 px (una situazione tipica), ci si può attendere la precisione con cui i parametri di calibrazione in esame inferiore a 0.004 radianti per il *roll*, 0.002 radianti per lo *yaw* e 0.001 radianti per il *pitch*, valori accettabili per applicazioni comuni. Il fatto che gli errori sui tre angoli salgono praticamente insieme è dovuto al fatto che il minimizzatore

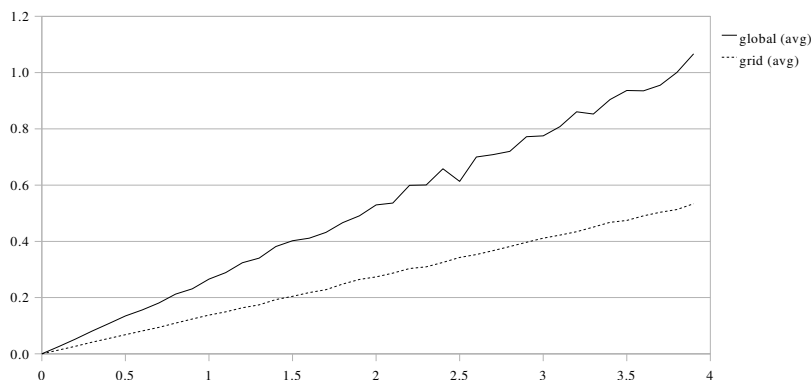


Figura 2.3: Estrazione Yaw, Pitch, Roll: errore in metri risultante sui punti usati per la stima (linea tratteggiata) ed errore conseguente nei punti di test (linea piena)

nel valutare un certo angolo cerca di compensare la discrepanza dei valori facendo variare gli altri due angoli di conseguenza.

Nel grafico 2.3 viene mostrato cosa rappresenta fisicamente l'errore sugli angoli, ovvero come l'errore dei parametri si propaga sui 26 punti usati per calibrare e sui 921 punti usati per la stima. Siccome l'utente avverte solo attraverso l'errore sui punti della griglia se la calibrazione è andata a buon fine o meno, il rapporto tra queste due curve diviene importante. In maniera abbastanza spannometrica, nel caso mostrato in figura 2.3 se l'utente trova un errore medio sulla griglia di un certo valore, fornito dall'equazione 2.4, potrà stimare qualitativamente che l'errore medio sui punti dell'area di test sarà quasi il doppio (precisamente un fattore medio di 1.93). Il valor medio dell'errore nei punti di controllo (l'inclinazione della retta del valor medio) è circa 0.26 metri per ogni pixel di rumore.

Anche se basterebbero solo 3 marker per ricavare il valore dei tre angoli *yaw*, *pitch* e *roll*, usare più elementi di calibrazione diminuisce il contributo di rumore del singolo punto. La figura 2.4 mostra come il numero di marker nella griglia di calibrazione influenza la propagazione dell'errore di rilevamento. Usando la griglia completa da 26 marker, quella con solo il marker destro a 13 elementi, o usando solo i 4 marker ai quattro estremi della griglia, l'errore sui singoli punti viene rigettato in

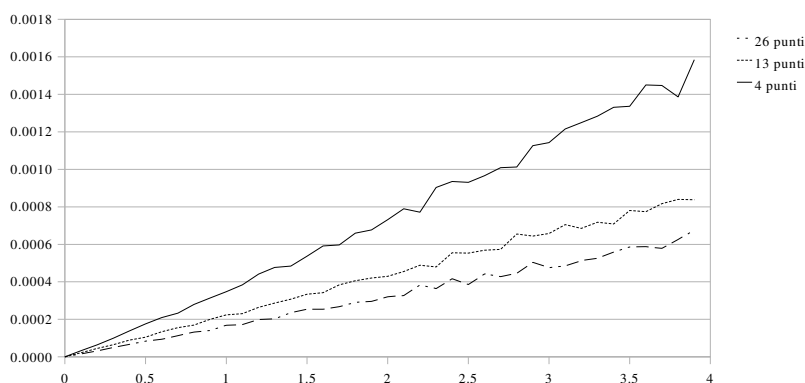


Figura 2.4: Estrazione Yaw, Pitch, Roll: Errore medio del pitch al variare del numero di marker della griglia

maniera differente. Man mano che il numero di marker aumenta infatti l'errore diminuisce. È da notare che la particolare geometria della griglia, formata da 13 coppie di marker vicini, di fatto quasi linearmente dipendenti, mostra come il passaggio da 13 marker a 26 non diminuisca in proporzione l'errore.

Man mano che aumentano il numero di incognite ovviamente l'errore sulle singole aumenta, in quando aumenta la possibilità di trovare dei minimi locali validi. I grafici 2.5 mostrano l'errore in metri di individuazione delle coordinate X_0 e Y_0

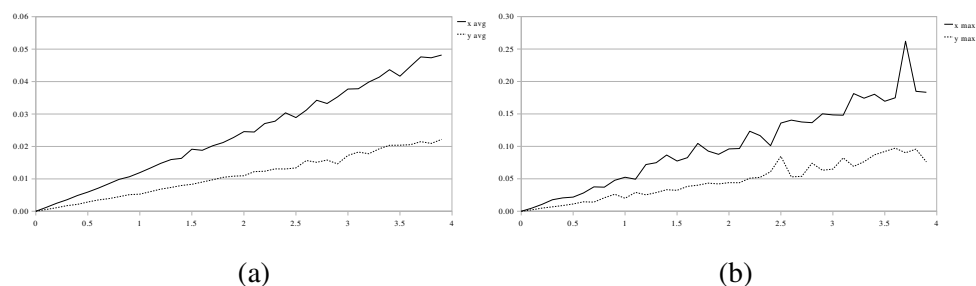


Figura 2.5: Estrazione Yaw, Pitch, Roll, X, Y da 26 marker affetti da rumore: errore medio (a) e massimo (b) sulla posizione del pin-hole (X_0, Y_0)

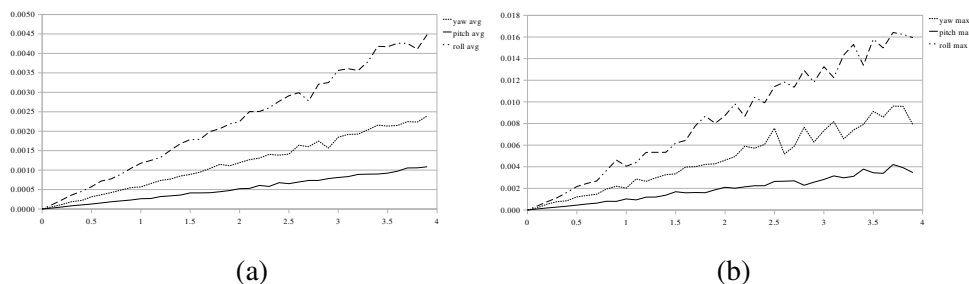


Figura 2.6: Estrazione Yaw, Pitch, Roll, X, Y da 26 marker affetti da rumore: errore medio (a) e massimo (b) sulla determinazione degli angoli

del pin-hole. Se di fatto l'altezza Z_0 del fuoco sia facilmente misurabile con filo a piombo, con precisione anche buona, e non varia a seconda di come il veicolo viene parcheggiato sulla griglia, non si può dire lo stesso delle altre due componenti.

La componente più affetta da errore è quella lungo l'asse X e si può stimare che il valore fornito da questo sistema (limitatamente come sempre alla sola configurazione in esame) sia vicino al valore corretto di un valore di circa 0.05 m se si stima l'errore di individuazione dei marker in 1 px.

Contemporaneamente alla determinazione della posizione del pin-hole lungo le direttrici X e Y viene anche rilevata ancora l'orientazione della camera attraverso gli angoli *yaw*, *pitch* e *roll*. Negli altri due grafici di questo esperimento in figura 2.6 vengono riportate le differenze in valore assoluto medie e massime tra gli angoli estratti e quelli reali e si può fare un confronto qualitativo con i grafici 2.2 rispettivamente. Se l'errore sull'angolo di *roll* non cambia sensibilmente ma cresce invece quello su *pitch* e soprattutto quello sullo *yaw*. La motivazione di questo fenomeno va sempre cercata nel tentativo di compensare gli errori sulle altre componenti, questa volta X e Y .

La nuova serie di grafici di figura 2.7 mostrano l'andamento dell'errore quando si vuole ricavare insieme l'orientazione della camera la sua apertura, orizzontale e verticale, dando sempre come conosciuti perfettamente i restanti parametri di calibrazione. Questo metodo è di fatto usato normalmente quando si riesca a misurare esattamente la posizione della camera nel mondo ma non si conosce la sua focale,

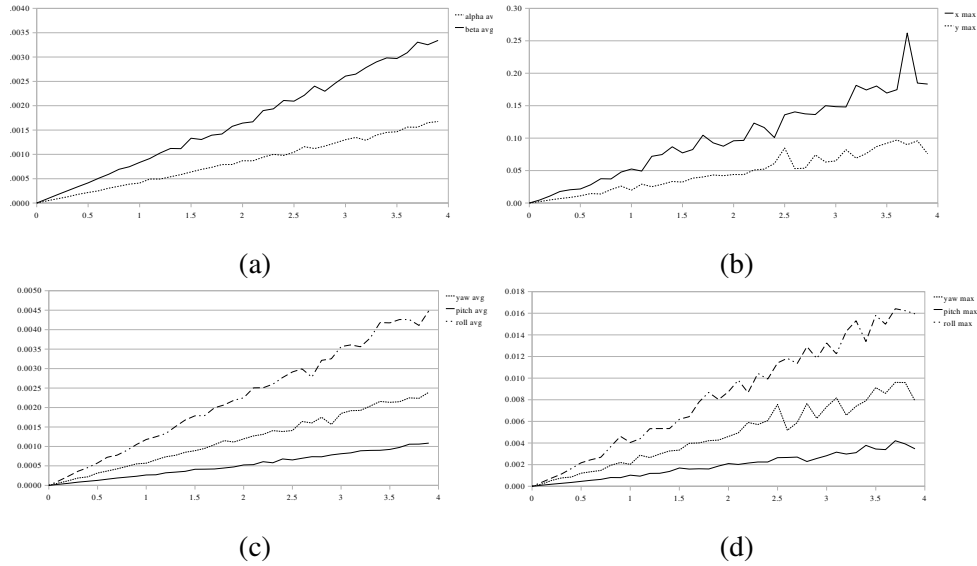


Figura 2.7: Estrazione Yaw, Pitch, Roll, α_u , α_v da punti 26 marker affetti da rumore: errore medio (a) e massimo (b) sull'apertura, ed errori medio (c) e massimo (d) sull'orientazione

oltre all'orientazione. Va fatto notare che in questo esperimento non è stata fatta nessuna ipotesi sul fatto che il pixel del sensore fosse quadrato e perciò non è stato messo nessun vincolo tra α_u e α_v , lasciati dunque liberi di variare indipendentemente.

C'è comunque da far presente che i parametri intrinseci sono comunque invarianti rispetto all'orientazione della camera e possono venire rilevati con altre tecniche [4] o comunque minimizzati tra diverse esecuzioni dell'algoritmo stesso.

Tutti questi esperimenti sono stati fatti usando come funzione errore per il minimizzatore l'equazione 2.4. È possibile mostrare come l'equazione 2.5 che minimizza l'errore in coordinate immagine fornisce infine risultati migliori per quanto riguarda la determinazione dei singoli parametri. Gli errori dopotutto sono in coordinate immagine e in questo modo i parametri affetti da errore non vengono proiettati da una trasformazione prospettica inversa in coordinate mondo ma si limitano solamente ad essere un riferimento. In figura 2.8 sono mostrati i confronti dell'errore medio

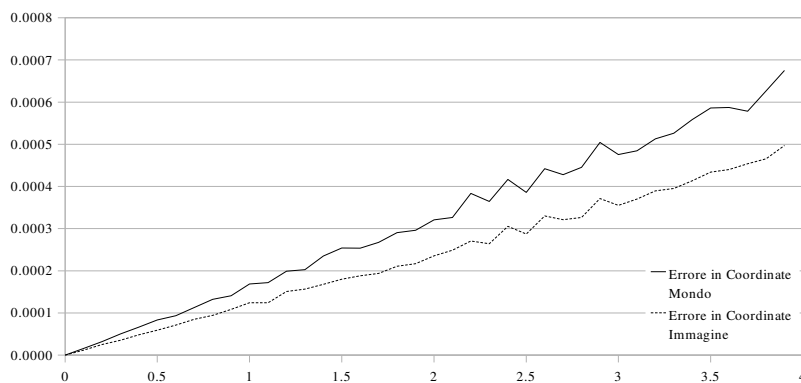


Figura 2.8: Confronto tra diverse funzioni usate nel minimizzatore: errore medio dell'angolo di pitch

dell'angolo di pitch usando come funzione errore quella in coordinate mondo 2.4 e immagine 2.5.

Un discorso opposto si farà nel caso si usi per calibrare l'omografia come si vedrà nella sezione 2.3.5. A seconda se si voglia ottenere un'immagine prospettica, o un'immagine prospettica inversa, si può scegliere una tecnica oppure l'altra.

È infine interessante far notare come se il punto di partenza per il minimizzatore non fosse vicino a quello corretto i risultati tendono a convergere a soluzioni più errate. Questo non avviene per l'estrazione dei soli angoli di orientamento (*yaw*, *pitch*, *roll*) ma si propone man mano che aumentano le incognite da minimizzare. Negli sviluppi futuri verranno studiati diversi tipi di minimizzatore, più robusti, come per esempio Levenberg–Marquardt, per verificare se tale comportamento secondario possa essere attenuato.

2.3.2 Estrazione parametri in presenza di errori sistematici

Un'altra classe di esperimenti interessanti consiste nell'esaminare come il sistema di estrazione dei parametri si comporti in presenza di un errore su uno dei parametri, non considerati incognite ma considerati attendibili. Capita infatti che i parametri

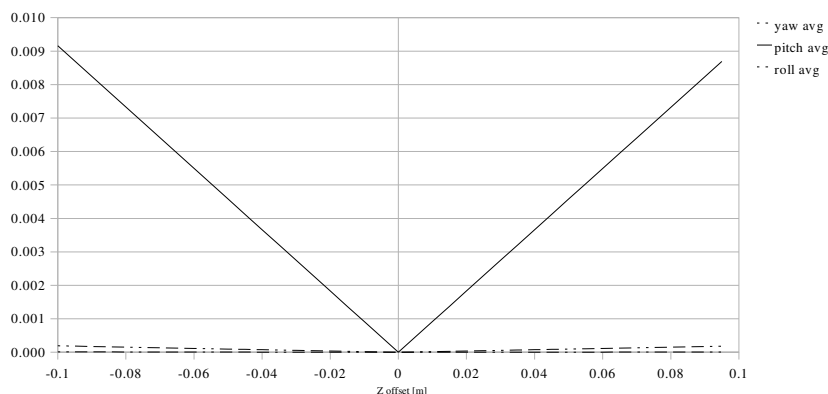


Figura 2.9: Estrazione Yaw, Pitch, Roll con errore sulla coordinata Z_0 : valori di errore negli angoli

che si reputano affidabili siano comunque misurati con precisione limitata o spesso in maniera approssimata a causa della particolare geometria del problema. Risulta pertanto interessante capire con quanta precisione i diversi parametri vanno stimati per poter avere una precisione sufficiente alla soluzione del problema in esame.

In questa serie di esperimenti verranno estratti nuovamente *yaw*, *pitch* e *roll* usando la tecnica mostrata nella sezione precedente, usando come punti i i punti della griglia non affetti da errore, con tutti i parametri della camera, tranne quelli di orientazione, dati per conosciuti esattamente tranne uno che verrà fatto variare all'interno di un intervallo intorno al suo valore corretto.

Il grafico 2.9 mostra l'andamento dell'estrazione di *yaw*, *pitch* e *roll* quando si commette un errore di misura sulla coordinata Z_0 del *pin-hole*. In tal caso sono stati fatti esperimenti nei quali alla misura corretta dell'altezza dal suolo Z_0 è stato aggiunto un valore compreso tra $[-0.1\text{ m}, 0.1\text{ m}]$. Dal grafico l'unico angolo ad essere influenzato da questo tipo di errore è quello di *pitch*. Il sistema cerca di controbilanciare all'errore su Z_0 modificando l'angolo di *pitch*, tuttavia, come mostrato dal grafico 2.10, il risultato è una compensazione solo apparente e si commettono errori nella successiva determinazione delle distanze nel mondo rappresentate dai 921 punti di esame.

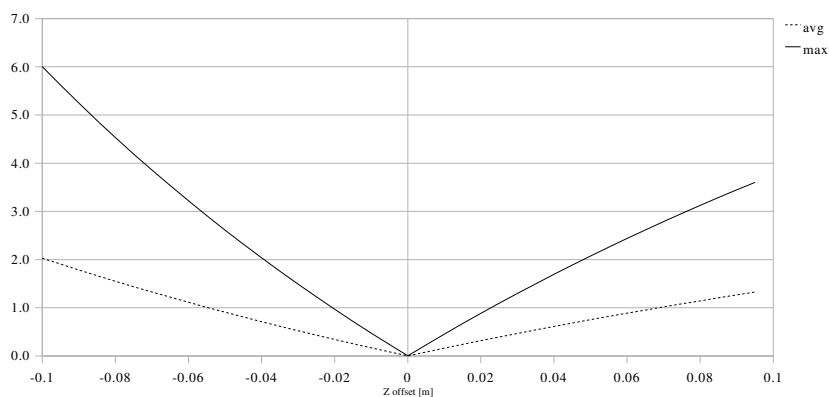


Figura 2.10: Estrazione Yaw, Pitch, Roll con errore sulla coordinata Z_0 : errore in metri, medio e massimo, nel determinare le distanze finali con il set point trovato

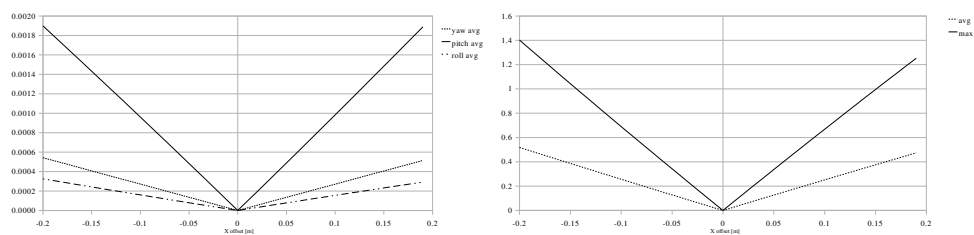


Figura 2.11: Estrazione Yaw, Pitch, Roll con errore sulla coordinata X_0 : errore nel determinare gli angoli (a) e le distanze finali con il set point trovato (b)

(a)

(b)

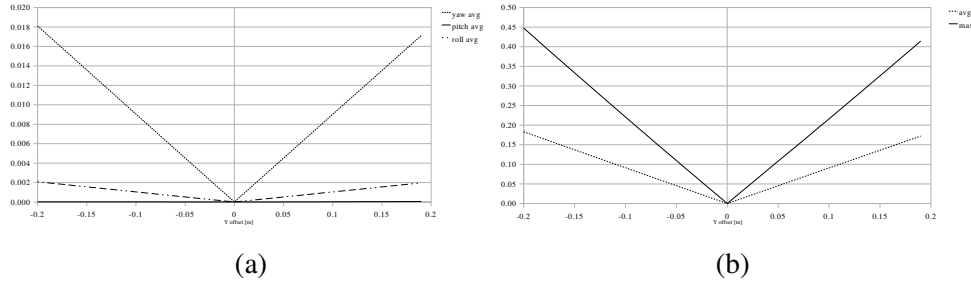


Figura 2.12: Estrazione Yaw, Pitch, Roll con errore sulla coordinata Y_0 : valori di errore negli angoli (a) ed errore nel determinare le distanze finali con il set point trovato (b)

Le coppie di grafici 2.11 mostrano invece gli errori di determinazione degli angoli di orientazione e l'errore sui punti d'esame nel caso venga aggiunto un offset sulla coordinata X_0 . L'intervallo di variazione stavolta è stato di $[-0.2 m, 0.2 m]$ e se si guarda l'entità dell'errore si vede che è molto inferiore di un ordine di grandezza rispetto al caso precedente.

Le coppie di grafici in figura 2.12 mostrano gli errori di determinazione degli angoli di orientazione e l'errore sui punti d'esame nel caso venga aggiunto un offset sull'ultima coordinata del *pin-hole*, ovvero Y_0 . Questa volta l'angolo di *yaw* risulta coinvolto nel cercare di compensare l'errore di misura. Se l'errore sull'angolo è abbastanza elevato, l'errore finale sulle coordinate mondo ne risente limitatamente rispetto al caso di X_0 .

Uno dei parametri di più difficile determinazione nelle camere risulta essere il Punto Principale. Solitamente per convenienza si pone tale punto nel centro geometrico del sensore lasciando al produttore della camera la richiesta di fare in modo che l'ottica proietti il fuoco sull'immagine (u_0, v_0) in un punto il più prossimo possibile al centro.

Il grafico 2.13.a mostra come gli angoli di orientazione risultino modificati da un errore di determinazione di u_0 , ovvero con che quantità il lasciare il centro ottico a metà della larghezza immagine possa produrre risultati errati. Siccome u_0 influenza il calcolo della focale k_u e questo parametro si suppone conosciuto correttamente, è

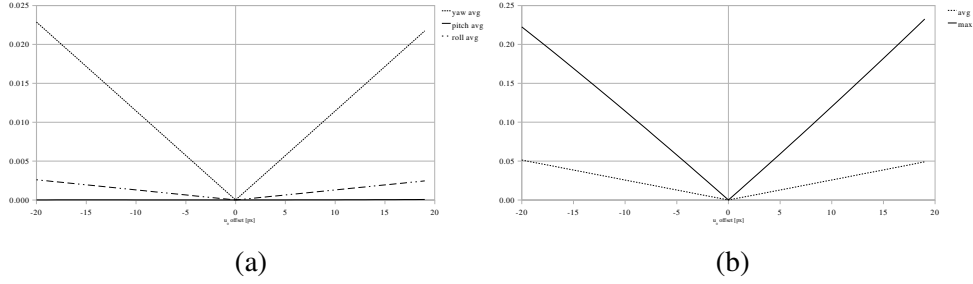


Figura 2.13: Estrazione Yaw, Pitch, Roll con errore sulla coordinata u_0 : valori di errore negli angoli (a) ed errore nel determinare le distanze finali con il set point trovato (b)

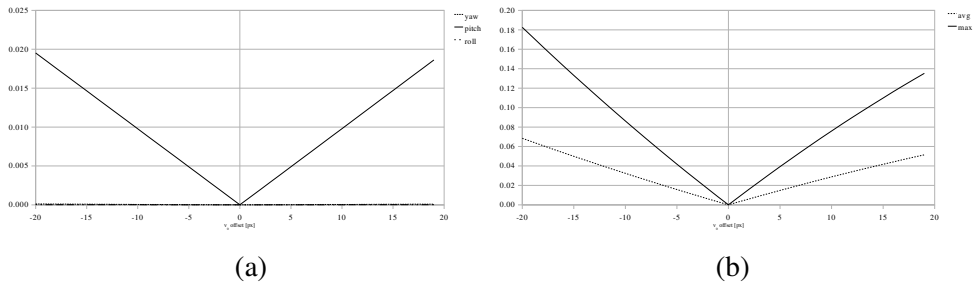


Figura 2.14: Estrazione Yaw, Pitch, Roll con errore sulla coordinata v_0 : valori di errore negli angoli (a) ed errore nel determinare le distanze finali con il set point trovato (b)

stato variato α_u in modo da lasciare k_u costante:

$$\tan \alpha'_u = \tan \alpha_u \frac{u_0 + offset}{u_0} \quad (2.6)$$

La variazione di u_0 pregiudica principalmente lo *yaw* e il conseguente errore in coordinate mondo è mostrato nel grafico 2.13.b.

L'altra coordinata del Punto Principale, v_0 , viene modificata allo stesso modo e i risultati sono mostrati nel grafico 2.14.a. In questo caso è stato reso costante k_v agendo su α_v :

$$\tan \alpha'_v = \tan \alpha_v \frac{v_0 + offset}{v_0} \quad (2.7)$$

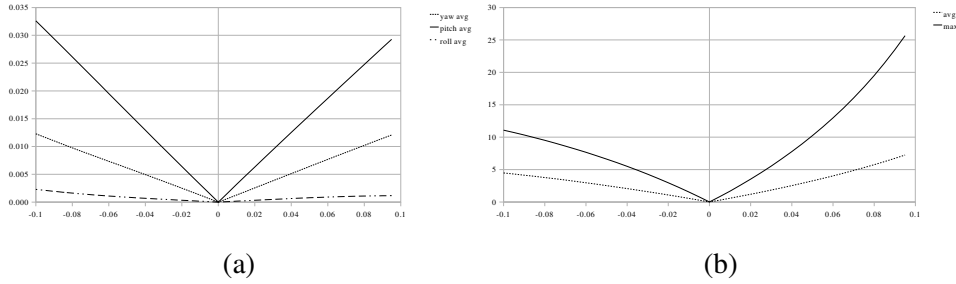


Figura 2.15: Estrazione Yaw, Pitch, Roll con errore sulla coordinata α con focale costante: valori di errore negli angoli (a) e nel determinare le distanze finali con il set point trovato

Siccome il setup della camera è particolare (poco *roll*, camera praticamente orizzontale parallela al terreno), la modifica alla coordinata v_0 influenza solamente il *pitch*. L'errore sul minimo trovato è mostrato infine nel grafico 2.14.b.

L'ultima coppia di grafici di figura 2.15 mostra l'errore di misura sull'apertura. Per semplicità viene mostrato il caso in cui il rapporto tra la tangente delle aperture sia costante, ovvero si conosca precisamente la geometria del sensore ma non ci sia adeguata precisione nell'aver determinato la lunghezza focale. In questo caso viene modificata la componente α_u di un certo *offset* e α_v cambia secondo la regola

$$\tan \alpha'_v = \tan \alpha_v \frac{\tan(\alpha_u + \text{offset})}{\tan \alpha_u} \quad (2.8)$$

In questo caso, mostrato in figura 2.15.a, risultano coinvolti tutti gli angoli in quantità differenti. L'intervallo di variazione è comunque molto grande rispetto alla condizione corretta e infatti i numeri mostrati nel grafico 2.15.b sono sicuramente più elevati degli altri. In questo caso il minimo sugli angoli non riesce a bilanciare l'errore introdotto da una errata misura di una delle componenti del sistema, e di conseguenza l'utente si accorge più facilmente che sui punti della griglia il minimizzatore non trova un minimo valido e può controllare se i parametri misurati siano effettivamente affidabili.

2.3.3 Propagazione degli errori di calibrazione alla misura delle distanze

La serie di grafici commentata in questa sezione esamina in dettaglio le variazioni delle coordinate prodotti da un errore di misura su uno dei parametri della camera nel caso in cui l'errore non venga compensato in alcun modo dai restanti parametri. I grafici mostrano gli errori medi e massimi sulla serie di 921 punti di controllo corretti e quelli ottenuti da una riproiezione in coordinate mondo errata usando parametri mal condizionati.

Nella serie di grafici di figura 2.16 viene mostrato tale errore (medio e massimo) nel caso vengano misurati in maniera errata a turno uno degli angoli di orientazione della camera. Nel caso del solo *pitch* il grafico non è simmetrico, ma a causa del setup della camera, più i punti si avvicinano all'orizzonte più l'errore cresce in maniera esponenziale.

Nel grafico 2.17 viene presentato il caso in cui venga misurata erroneamente l'altezza del pin-hole Z_0 . Nei grafici di figura 2.18 la posizione del *principal point* e infine nel 2.19 la focale, intesa come errore su α_u ma con il rapporto tra le tangenti con α_v costante, usando l'equazione 2.8.

2.3.4 Effetti della distorsione

Questa sezione affronta brevemente gli effetti che ha la distorsione dell'ottica sulla calibrazione. Siccome l'equazione della pin-hole camera non è valida in presenza di distorsione verrà discusso che fenomeni incorrono quando la distorsione è bassa e capire quando è possibile comunque trascurarla. Le curve in figura 2.20 mostrano la variazione in modulo della posizione nell'immagine dei 921 punti di controllo quando è presente una distorsione radiale approssimabile a quella polinomiale con un solo parametro k_1 . Nella configurazione degli esperimenti, l'errore introdotto dalla distorsione può essere paragonato a quello introdotto da un rumore uniforme della stessa ampiezza. Siccome il parametro di distorsione modifica anche le aperture sono stati applicati fattori di scala orizzontali e verticali all'immagine in modo che l'apertura finale rimanga comunque la stessa.

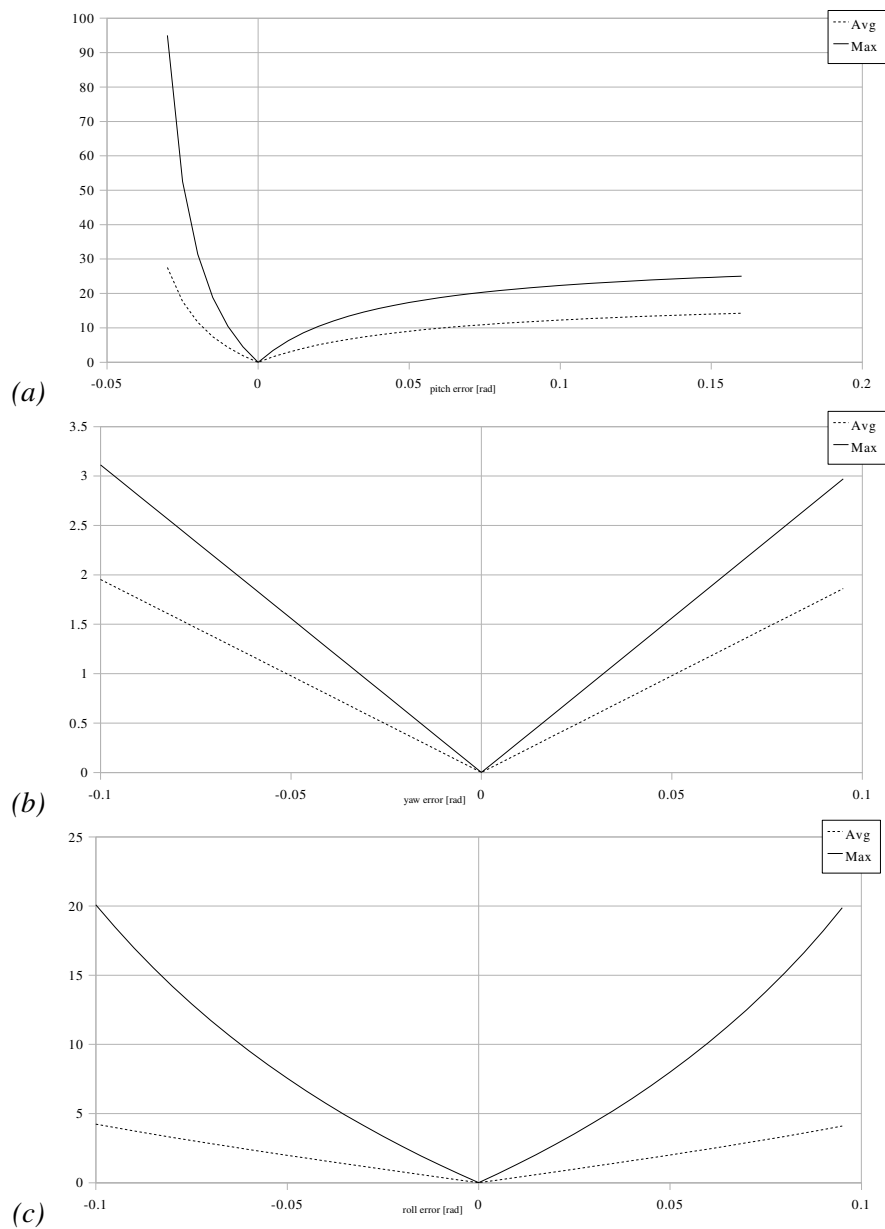


Figura 2.16: Errori sulle distanze dei punti di test in caso di errore sul pitch (a), yaw (b), roll (c)

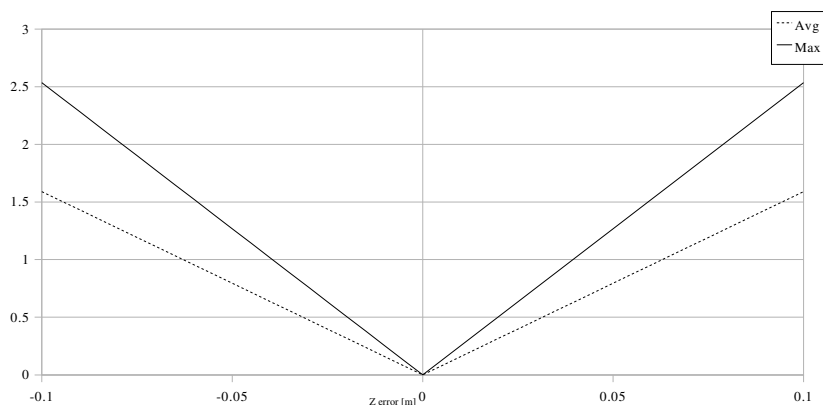


Figura 2.17: Errori sulle distanze dei punti di test in caso di errore sulla coordinata Z_0

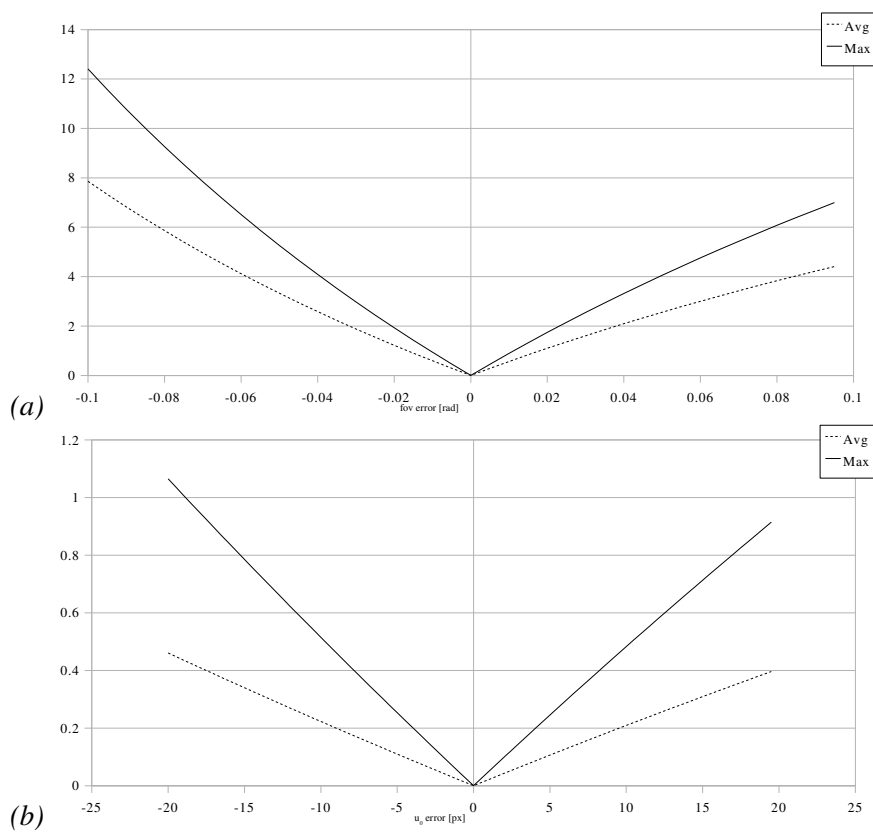
Per riferimento il grafico 2.21 mostra l'errore di estrazione dei parametri partendo da una griglia dove i punti si conoscono perfettamente ma sono affetti da errore di distorsione non corretto, e il grafico associato 2.22 mostra gli errori metrici dei punti di controllo a seguito della compensazione dell'errore con il nuovo set point della camera.

2.3.5 Calibrazione usando l'omografia

Per particolari compiti come nel caso di elaborazione di immagini IPM, una alternativa alla calibrazione Euleriana è usare direttamente la matrice omografica 2.1. Come il minimizzatore anche l'omografia è influenzata dalla presenza di rumore sui punti della griglia di calibrazione, e se con la calibrazione Euleriana i parametri sono vincolati a quelli supposti esatti, nella calibrazione attraverso la matrice di proiezione ogni elemento è un parametro con una sua dinamica.

Partendo dagli n punti della griglia viene generata una matrice di trasformazione omografica che minimizza, a scelta, l'errore in coordinate mondo o in coordinate immagine.

Per esempio le curve presentate in figura 2.24 mostrano gli errori massimi ri-

Figura 2.18: Errori sulle distanze dei punti di test in caso di errore su u_0 (a) e v_0 (b)

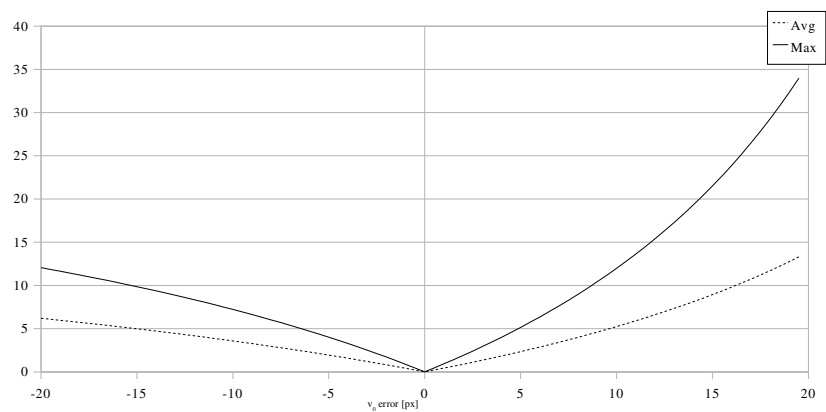
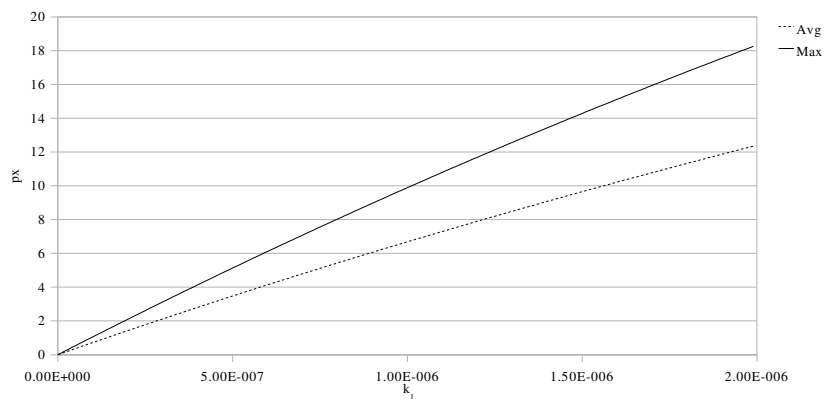
Figura 2.19: Errori sulle distanze dei punti di test in caso di errore su α_u 

Figura 2.20: Variazione della posizione dei punti immagine a causa della distorsione radiale

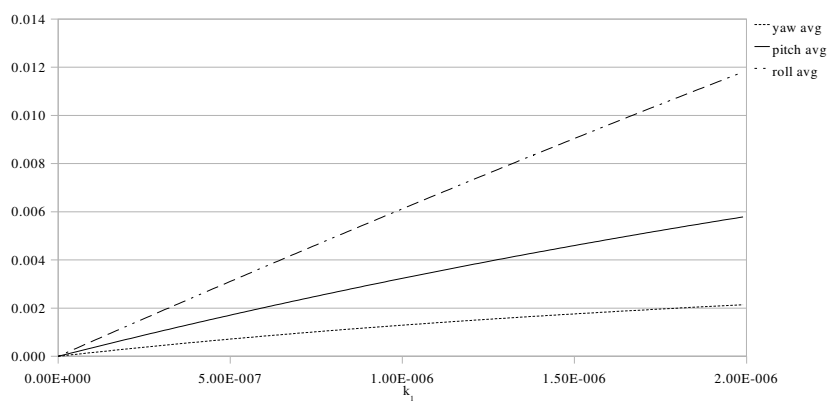


Figura 2.21: Estrazione di yaw, pitch, roll nel caso in cui i punti della griglia di calibrazione siano affetti da errore di distorsione

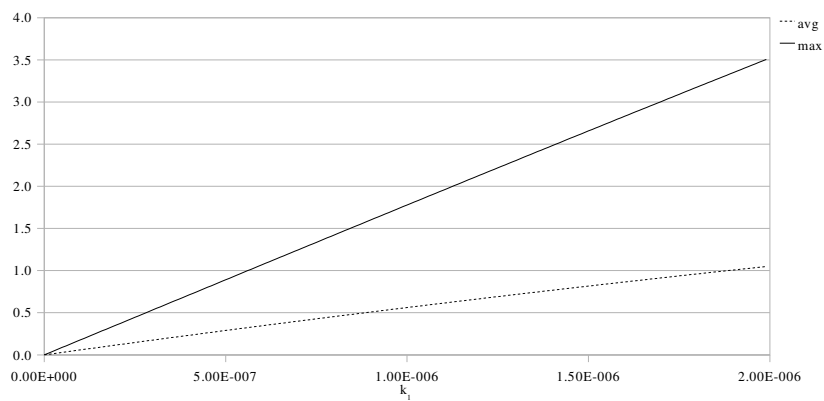


Figura 2.22: Errore sulle distanze nel mondo a seguito dell'estrazione di yaw, pitch, roll su punti della griglia affetti da errore di distorsione

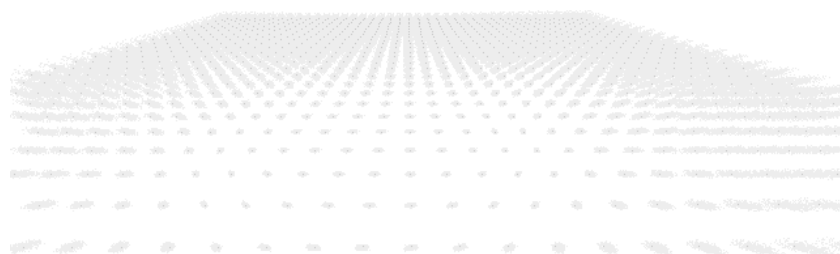


Figura 2.23: Esempio della dispersione dei punti: in nero i punti di controllo corretti, in grigio quelli generati dalle fasi di test sull'omografia

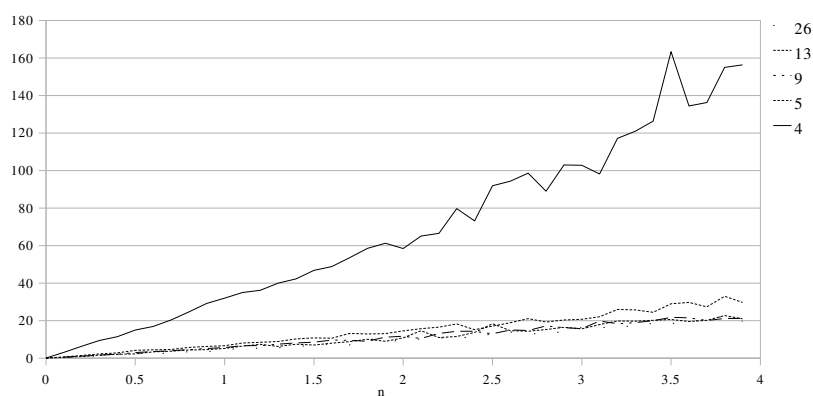


Figura 2.24: Calibrazione con omografia in presenza di errore di individuazione dei punti: errore massimo sull'area di controllo al variare del numero di punti usati nella calibrazione

5	3.77
9	2.78
13	2.28
26	2.03

Tabella 2.2: Rapporto tra l'errore massimo sui punti di prova con l'errore massimo sui punti di calibrazione

scontrati nei punti di controllo in coordinate immagine quando per calibrare vengono usati un numero variabile n di punti della griglia. Come è prevedibile più il numero di punti aumenta più il rumore viene filtrato. La condizione con 4 soli punti è la minima necessaria per realizzare una omografia, ma è interessante notare che poi da già da 5 punti in sù la componente di rumore si attenua sensibilmente.

In figura 2.23 sono mostrati i 921 punti di controllo (in nero) non affetti da errore, e la proiezione errata che subiscono nelle diverse esecuzioni di questo esperimento.

Come nel caso della calibrazione Euleriana è possibile stimare dall'errore sul residuo dei punti coinvolti dalla calibrazione, quanto questi siano funzionali a una calibrazione. La tabella 2.2 mostra, al variare del numero di elementi della griglia, i rapporti tra gli errori massimi (in pixel) riscontrati sui punti di controllo rispetto all'errore massimo trovato sui punti di calibrazione. Più il numero di elementi della griglia aumenta, anche se l'errore massimo non diminuisce sensibilmente, aumenta la confidenza che si hanno dei punti e risulta più facile capire se i punti selezionati sono un buon esempio di piano di calibrazione.

Per confronto con il caso della calibrazione Euleriana in figura 2.25 vengono mostrati gli errori in metri dei punti così valutati. L'andamento dell'errore medio è compreso tra 0.7 e 0.9 metri su pixel di rumore (molto più elevato rispetto allo 0.26 della calibrazione dei soli parametri di yaw, pitch, roll).

Risulta tuttavia possibile minimizzare la funzione errore dell'omografia non sulle coordinate dei pixel, ma sulle coordinate metriche. A causa della particolare forma della matrice 2.2, questa volta l'errore si riduce sensibilmente come mostrato in figura 2.26. In tal caso la propagazione dell'errore dal rumore sul pixel al valor medio

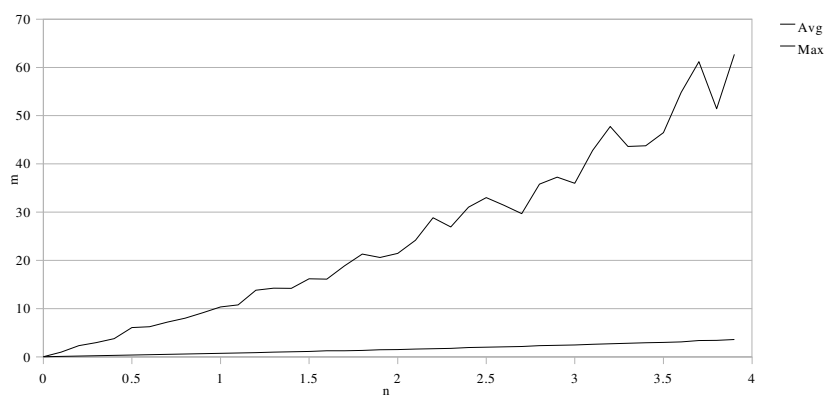


Figura 2.25: Calibrazione con omografia minimizzata in coordinate immagine in presenza di errore di individuazione dei punti: errore medio e massimo sull'area in metri

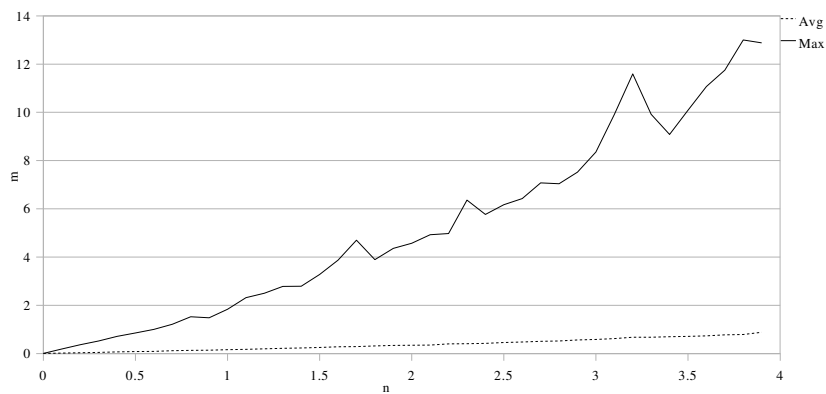


Figura 2.26: Calibrazione con omografia minimizzata in coordinate mondo in presenza di errore di individuazione dei punti: errore medio e massimo sull'area in metri

sui punti di test è compreso tra 0.16 e 0.21 metri per ogni pixel di rumore, questa volta inferiore anche alla calibrazione dei soli parametri di *yaw*, *pitch*, *roll*, anche i casi peggiori (in cui tutti i punti della griglia presentano l'errore nella medesima direzione, rafforzandosi, invece di annullarsi) risultano con errore più elevato rispetto al caso euclideo.

Capitolo 3

Inverse Perspective Mapping

Siccome buona parte delle tecniche sviluppate in questo lavoro di tesi basano il loro funzionamento sul concetto di *Inverse Perspective Mapping*, come prima fase della discussione verranno introdotte le peculiarità di tale trasformazione.

Il termine *Inverse Perspective Mapping* (IPM) compare nel 1990 introdotto da Mallot [5]. Citando Mallot

“Il termine *inverse perspective mapping* non corrisponde a una vera inversione della proiezione prospettiva, la quale cosa è matematicamente impossibile, ma denota invece un'inversione sotto il vincolo aggiuntivo che i punti riproiettati debbano stare su un piano orizzontale.”

L'idea base dell'IPM è quella di riproiettare lo spazio immagine acquisito dalla camera reale usando una camera virtuale posta ad altezza elevata che osserva perpendicolarmente il piano di proiezione. Tale trasformazione infatti è conosciuta in letteratura anche come *Bird Eye View* ‘vista a volo d'uccello’. Come fa notare lo stesso Mallot è impossibile eseguire una riproiezione esatta della visuale, visto che ogni punto dell'immagine è conosciuto a meno di una sua componente. Viene fissata pertanto una delle incognite (in questo caso l'altezza del punto dal suolo supposta costante) e si applica la riproiezione sotto questo vincolo aggiuntivo.

È da notare come tutti i punti che violino l'ipotesi di piano costante verranno conseguentemente proiettati in una coordinata mondo errata, acquisendo di fatto er-

roneamente la coordinata del punto del piano che sottendono. Siccome in diversi studi viene considerato come ostacolo tutto ciò che non giace sul piano stradale, risulta evidente come una trasformazione che proietta erroneamente ostacoli definiti proprio in questo modo può essere utilizzata per rilevarli.

In questo capitolo verranno discusse le proprietà di questa trasformazione per introdurre le applicazioni pratiche discusse poi nei capitoli successivi.

3.1 L'equazione base dell'inversione prospettica

Per introdurre le proprietà dell'IPM, bisogna fare una osservazione preliminare.

Consideriamo la funzione di proiezione della *pin-hole camera* scritta in forma esplicita:

$$F_{pm}(X, Y, Z) = \begin{pmatrix} k_u \frac{r_1(X - X_0) + r_2(Y - Y_0) + r_3(Z - Z_0)}{r_7(X - X_0) + r_8(Y - Y_0) + r_9(Z - Z_0)} + u_0 \\ k_v \frac{r_4(X - X_0) + r_5(Y - Y_0) + r_6(Z - Z_0)}{r_7(X - X_0) + r_8(Y - Y_0) + r_9(Z - Z_0)} + v_0 \end{pmatrix} \quad (3.1)$$

dove $(X, Y, Z)^T$ è un generico punto in coordinate mondo, e i dettagli sui restanti parametri sono discussi in appendice A. È ben visibile come tale funzione proietti punti da uno spazio di \mathbb{R}^3 in uno spazio \mathbb{R}^2 e per questo motivo non è direttamente invertibile. Se tuttavia fissiamo una delle 3 componenti (X, Y, Z) la funzione diventa di \mathbb{R}^2 e risulta invertibile.

Definiamo dunque l'immagine IPM come l'immagine rappresentante per ogni suo punto (X, Y) la rispettiva coordinata mondo e a cui è associata a tutta l'immagine una coordinata Z costante.

Teorema 3.1.1. *La trasformazione di proiezione prospettica di un generico piano è una trasformazione omografica.*

Dimostrazione. L'equazione della *pin-hole camera* ovvero l'espressione della proiezione prospettica in coordinate omogenee (vedi equazione A.13), relativa a un

generico piano $Z = Z_p$ costante, si scrive come

$$\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \mathbf{A} \begin{bmatrix} \mathbf{R} & \vec{t}_0 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z_p \\ 1 \end{bmatrix} \quad (3.2)$$

Siccome il vettore dei punti $(X, Y, Z_p)^T$ possiede una componente costante Z_p e moltiplica una matrice, è possibile eseguire la seguente trasformazione

$$\mathbf{A} \begin{bmatrix} X \\ Y \\ k \end{bmatrix} + \vec{b} = \mathbf{C} \begin{bmatrix} X \\ Y \\ 1 \end{bmatrix} \quad (3.3)$$

dove $\mathbf{C} = \begin{bmatrix} \vec{a}_1 & \vec{a}_2 & k\vec{a}_3 + \vec{b} \end{bmatrix}$.

La trasformazione che proietta i punti del piano (X, Y) in coordinate mondo sui punti immagine (u, v) può essere scritta pertanto

$$\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \mathbf{H}_Z \begin{bmatrix} X \\ Y \\ 1 \end{bmatrix} \quad (3.4)$$

avendo definito

$$\mathbf{H}_Z = \mathbf{A} \cdot \mathbf{R}_Z \quad (3.5)$$

dove la matrice \mathbf{R}_Z è formata dall'unione della matrice di rotazione con il vettore di traslazione $\vec{t}_0 = (t_x, t_y, t_z)^T$ in coordinate camera (si veda la definizione A.11) sfruttando l'equazione 3.3.

La trasformazione 3.4 è di fatto una trasformazione che proietta coordinate omogenee in coordinate omogenee ed è una trasformazione omografica.

Si può esplicitare ulteriormente \mathbf{R}_Z , come

$$\mathbf{R}_Z = \begin{bmatrix} r_0 & r_1 & r_2 Z_p - t_x \\ r_3 & r_4 & r_5 Z_p - t_y \\ r_6 & r_7 & r_8 Z_p - t_z \end{bmatrix} \quad (3.6)$$

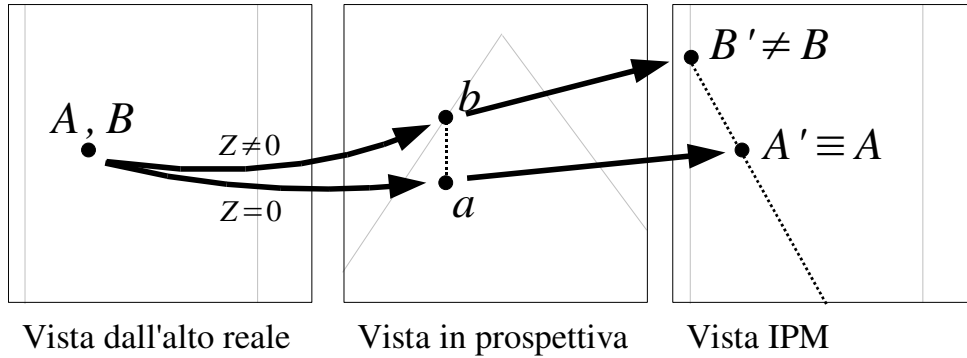


Figura 3.1: Errata proiezione di punti con $Z \neq 0$

e, siccome normalmente quando si lavora su immagini IPM $Z_p = 0$, la scrittura diventa

$$\mathbf{R}_Z = \begin{bmatrix} r_0 & r_1 & -t_x \\ r_3 & r_4 & -t_y \\ r_6 & r_7 & -t_z \end{bmatrix} \quad (3.7)$$

I parametri r_0, \dots, r_7 sono disponibili in forma esplicita nell'equazione A.17. \square

3.2 Le proprietà dell'errata riproiezione nell'IPM

In questa sezione vengono esaminati in dettaglio quali legami sussistono tra i punti sull'immagine IPM e i corrispondenti punti reali in coordinate mondo.

Si osservi la figura 3.1. Siano $\mathbf{A} = (X, Y, 0)^T$ e $\mathbf{B} = (X, Y, Z)^T$, con $Z \neq 0$, due punti tridimensionali. \mathbf{A} appartiene al piano $Z = 0$ mentre \mathbf{B} si può considerare un generico ostacolo esterno al piano. I due punti tridimensionali \mathbf{A} e \mathbf{B} verranno proiettati su una immagine bidimensionale attraverso la funzione prospettica rispettivamente nei punti \mathbf{a} e \mathbf{b} in coordinate immagine. Ricordiamo nuovamente che il punto \mathbf{a} rappresenta sempre un punto sul terreno, mentre \mathbf{b} rappresenta un punto a una altezza variabile. Siccome la proiezione supposta è generica, i due punti immagine non

sono coincidenti. I punti, dopo la proiezione, di fatto perdono l'informazione su una dimensione spaziale.

Se tali punti, dall'immagine in prospettiva, vengono trasferiti in una vista dall'alto, attraverso la funzione di *Inverse Perspective Mapping*, la loro reale altezza rimane ignota e la funzione, non possedendo ulteriori informazioni, calcolerà la loro posizione virtuale nell'immagine IPM applicando l'ipotesi che ogni punto abbia $Z = 0$ costante.

I due punti **A** e **B** saranno riproiettati perciò in due punti **A'** e **B'**. Se i due punti avevano le coordinate (X, Y) uguali in origine nel mondo, solo il punto **A**, effettivamente sul terreno, viene proiettato nella coordinata corretta in quanto soddisfa l'ipotesi $Z = 0$, mentre il punto verticale **B** finisce in un punto con coordinate (X', Y') diverse dalle sue originali. In sintesi, dopo la proiezione IPM, $\mathbf{A}' \cong \mathbf{A}$ e $\mathbf{B}' \neq \mathbf{B}$.

L'obiettivo di questa sezione è mostrare la relazione che intercorre tra due punti riproiettati **A'** e **B'**, e mostrare che è possibile ricavare il valore Z associato al punto **B'**.

Senza perdita di generalità dichiariamo $(X_0, Y_0, Z_0)^T = (0, 0, h)^T$, dove h è l'altezza della camera dal suolo. Allo stesso modo lavoriamo in coordinate camera normalizzate e sottintendiamo i parametri intrinseci k_u, k_v, u_0, v_0 , perchè si elidono nel passaggio da prospettiva a prospettiva inversa.

Sotto queste semplificazioni il punto (X, Y, Z) viene proiettato nel punto (u, v) secondo l'equazione

$$\begin{aligned} u &= \frac{r_0X + r_1Y + r_2(Z - h)}{r_6X + r_7Y + r_8(Z - h)} \\ v &= \frac{r_3X + r_4Y + r_5(Z - h)}{r_6X + r_7Y + r_8(Z - h)} \end{aligned} \quad (3.8)$$

I punti **a** e **b** sono di fatto generati attraverso l'equazione 3.8.

Invertendo 3.8 si ottiene l'equazione per la trasformazione prospettica inversa:

$$\begin{aligned} X &= \frac{r_0u + r_3v + r_6}{r_2u + r_5v + r_8}(Z - h) \\ Y &= \frac{r_1u + r_4v + r_7}{r_2u + r_5v + r_8}(Z - h) \end{aligned} \quad (3.9)$$

dove Z è la presunta altezza del punto da proiettare (perciò normalmente $Z = 0$), (u, v) è il punto sull'immagine in prospettiva, e infine (X, Y, Z) sono le coordinate mondo del punto in esame sotto questa ipotesi.

Usando l'equazione 3.9, l'espressione del punto $\mathbf{B}' = (X', Y', 0)$ proveniente dal punto immagine \mathbf{b} diventa pertanto:

$$\begin{aligned} X' &= -\frac{r_1u + r_4v + r_7}{r_3u + r_6v + r_9}h \\ Y' &= -\frac{r_2u + r_5v + r_8}{r_3u + r_6v + r_9}h \end{aligned} \quad (3.10)$$

avendo applicato l'ipotesi $Z = 0$.

Se si unisce l'equazione 3.8 con l'equazione 3.10, si sviluppa il sistema risultante, si ottiene la trasformazione omografica:

$$\begin{aligned} X' &= -h \frac{X(r_0r_0 + r_3r_3 + r_6r_6) + Y(r_0r_1 + r_3r_4 + r_6r_7) + (Z-h)(r_0r_2 + r_3r_5 + r_6r_8)}{X(r_2r_0 + r_5r_3 + r_8r_6) + Y(r_2r_1 + r_5r_4 + r_8r_7) + (Z-h)(r_2r_2 + r_5r_5 + r_8r_8)} \\ Y' &= -h \frac{X(r_1r_0 + r_4r_3 + r_7r_6) + Y(r_1r_1 + r_4r_4 + r_7r_7) + (Z-h)(r_1r_2 + r_4r_5 + r_7r_8)}{X(r_2r_0 + r_5r_3 + r_8r_6) + Y(r_2r_1 + r_5r_4 + r_8r_7) + (Z-h)(r_2r_2 + r_5r_5 + r_8r_8)} \end{aligned} \quad (3.11)$$

Siccome $r_0 \dots r_8$ sono gli elementi di una matrice di rotazione (vedi espressione A.15) è possibile sfruttare le proprietà che tale matrice possiede. Infatti le riga e le colonna della matrice \mathbf{R} sono vettori ortonormali, ne consegue che il modulo di ogni riga e di ogni colonna di tale matrice vale 1 e il prodotto scalare di due righe o di due colonne da sempre valore nullo.

La trasformazione 3.11 si semplifica e il risultato notevole è:

$$\begin{aligned} X' &= -h \frac{X}{Z-h} \\ Y' &= -h \frac{Y}{Z-h} \end{aligned} \quad (3.12)$$

da cui, rimuovendo la semplificazioni introdotte in precedenza sulla posizione del *pin-hole*, il risultato generale diventa

$$\begin{aligned} X' &= -Z_0 \frac{X - X_0}{Z - Z_0} + X_0 \\ Y' &= -Z_0 \frac{Y - Y_0}{Z - Z_0} + Y_0 \end{aligned} \quad (3.13)$$

In questa relazione $(X_0, Y_0, Z_0)^\top$ sono le coordinate del *pin-hole* della camera, $(X, Y, Z)^\top$ la posizione del punto in coordinate mondo, e $(X', Y', 0)^\top$ le coordinate del medesimo punto nell'immagine IPM.

Siccome sono scomparsi tutti parametri che dipendono dell'orientazione della camera questa relazione vale per qualunque sistema di riferimento, indipendentemente dalla calibrazione e dal modo in cui l'IPM è stata calcolata. Tale immagine è pertanto indipendente dalla rotazione con cui la camera osserva la scena, e in generale, dal tipo di camera usata. Se due camere diverse, con ottica diversa, osservano la medesima scena dalla stessa posizione del mondo, ovvero dalla stessa coordinata del *pin-hole*, l'immagine IPM ottenuta sarà sempre la stessa. Tale considerazione risulta importante in questo lavoro di tesi perché consente di calcolare l'immagine IPM attraverso la tecnica che si preferisce sapendo infatti che il risultato sarà comunque unico.

Dalla relazione 3.13 si può ricavare la relazione che lega le coordinate dell'immagine tra loro:

$$\frac{X' - X_0}{X - X_0} = \frac{Y' - Y_0}{Y - Y_0} \quad (3.14)$$

mostrando come tale relazione sia di tipo lineare.

Il coefficiente angolare di tale retta che congiunge il punto *A* con il punto *B* vale effettivamente

$$m' = \frac{X' - X_0}{Y' - Y_0} = \frac{X - X_0}{Y - Y_0} \quad (3.15)$$

e tutti gli oggetti verticali convergono verso la proiezione del pin-hole (X_0, Y_0) sul piano.

È da ricordare che nel testo sono state usate le stesse unità di misura sia per X , Y , Z che per i punti provenienti sicuramente da un immagine IPM come X' , Y' . È ovvio che quando si lavora con immagini, le coordinate (X', Y') devono essere trasferite da uno spazio metrico a un'area di memoria in pixel, e per fare questo bisogna tenere conto di un qualche fattore moltiplicativo e di un eventuale *offset*. Per semplicità, in questo testo verrà sottintesa questa conversione e verranno espressi (X', Y') come se fossero metri.

3.3 Estensione a più camere

Esaminiamo ora cosa si può dire nel caso in cui ci siano due o più camere che osservano la medesima area da punti differenti. Tale procedimento sarà la base poi della

tecnica di *Ground Plane Stereo IPM* descritta nel capitolo successivo.

Prendiamo il caso di due camere allineate, alla stessa altezza, ma separate orizzontalmente sull'asse Y da una *baseline* di lunghezza b . Per semplicità i pin-hole delle due camere saranno $x_1 = x_2 = 0$, $y_1 = \frac{b}{2}$, $y_2 = -\frac{b}{2}$, e $z_1 = z_2 = h$.

Un generico punto mondo $(X, Y, Z)^T$, secondo la relazione 3.13, verrà proiettato nelle immagini IPM prodotte dalle due camere in

$$\begin{aligned} X'_1 = X'_2 &= -h \frac{X}{Z-h} = X' \\ Y'_1 &= -h \frac{Y - \frac{b}{2}}{Z-h} + \frac{b}{2} \\ Y'_2 &= -h \frac{Y + \frac{b}{2}}{Z-h} - \frac{b}{2} \end{aligned} \quad (3.16)$$

dove è ben visibile che le due coordinate combaciano solamente quando $Z = 0$. È da notare che in questa particolare condizione di camere allineate, i punti nelle due immagini IPM avranno sempre la stessa coordinata X' .

Risolvendo il sistema 3.16 è possibile ricavare la posizione della coordinata $(X, Y, Z)^T$ in una immagine IPM (per esempio la 2) in funzione delle coordinate sull'altra immagine IPM (esempio la 1):

$$Y'_2 = Y'_1 - b \left(\frac{h}{Z-h} + 1 \right) = Y'_1 + b \frac{Z}{h-Z} \quad (3.17)$$

data comunque la conoscenza di Z .

Nel caso tipico in cui $Z < h$ e $b > 0$ il secondo pezzo dell'equazione 3.17 è positivo e $Y'_2 \geq Y'_1$. In questo modo è possibile, dato un potenziale punto (X'_1, Y'_1) su una immagine IPM, verificare per quale altezza incognita del punto Z vengono soddisfatti dei vincoli fotogrammetrici con il punto (X'_2, Y'_2) dell'altra immagine.

L'equazione 3.17 può essere invertita, e permette di calcolare l'altezza Z del punto data la disparità $\Delta = Y'_2 - Y'_1$:

$$Z = h \left(1 - \frac{b}{\Delta + b} \right) = h \frac{\Delta}{\Delta + b} \quad (3.18)$$

Anche in questo caso la disparità Δ anche se espressa nelle coordinate mondo è normalmente misurata in pixel sottintendendo nuovamente il fattore di scala.

L'espressione 3.18 fornisce anche il risultato notevole secondo il quale punti alla stessa altezza Z nel mondo, possiedono la stessa disparità Δ nel confronto tra immagini IPM stereo.

Quando la disparità scende fino a $\Delta = 0$ correttamente Z si annulla, mentre man mano che la disparità cresce l'equazione tende alla singolarità a $Z = h$.

Analogamente a Z , si può infine ricavare la coordinata X :

$$X = X' \frac{b}{\Delta + b} \quad (3.19)$$

Dall'equazione 3.18 si può ricavare una relazione che lega la distanza tra le camere alla sensibilità di individuazione di un ostacolo.

Sia s il fattore di scala tra coordinate immagine e mondo. Poiché un elemento a una determinata altezza Z dal suolo abbia almeno la disparità di un pixel, ovvero $\Delta = s$, sull'IPM la *baseline* deve essere

$$\frac{b}{s} > \frac{h - Z}{Z} \quad (3.20)$$

siccome s è la disparità in metri di un pixel ma anche il fattore di scala, $\frac{b}{s}$ è in effetti il valore della *baseline* espresso in pixel.

3.4 Estensione al *motion stereo*

Esaminiamo ora il caso in cui le due camere siano sempre allineate ma traslate lungo la coordinata X nello spazio e, nel caso particolare del *motion stereo*, nel tempo.

Definiamo questa volta $x_1 = 0$, $x_2 = d$ e $y_1 = y_2 = 0$, e $z_1 = z_2 = h$ e le equazioni 3.13 diventano:

$$\begin{aligned} X'_1 &= -h \frac{X}{Z - h} \\ X'_2 &= -h \frac{X - d}{Z - h} + d \quad rl \\ Y'_1 = Y'_2 = Y' &= -h \frac{Y}{Z - h} \end{aligned} \quad (3.21)$$

Da questa relazione si possono fare le medesime considerazioni viste in precedenza. Di fatto questa volta i punti sono a Y' costante e la variazione è espressa lungo

X' :

$$X'_1 = X'_2 - d \left(\frac{h}{Z-h} + 1 \right) = X'_2 + d \frac{Z}{h-Z} \quad (3.22)$$

da cui la disparità $\Delta = X'_2 - X'_1$ vale:

$$Z = h \frac{\Delta}{\Delta + d} \quad (3.23)$$

3.5 Problemi del matching

La trasformazione IPM non ha carattere simile, perciò oggetti che in prospettiva erano pressapoco simili, riproiettati possono risultare notevolmente differenti (si veda per esempio l'immagine 5.6).

Scorrendo la superficie di un oggetto lungo la coordinata Y , mantenendo X costante, si ottiene nell'immagine IPM uno spostamento sulla coordinata X' costante e Y' che varia lungo l'asse delle Y . Se le due camere che inquadrano la scena sono alla stessa altezza, esaminando l'equazione 3.17, lungo l'asse delle Y tra le due immagini IPM c'è idealmente un matching perfetto, a meno dell'*offset* dovuto alla disparità.

La variazione di forma dell'oggetto inquadrato è invece colpa dello spostamento che ha questo lungo il suo asse Z che provoca una variazione di disparità (equazione 3.17) lungo la direzione X nell'immagine IPM, nonostante la coordinata X dell'oggetto sia costante.

Sotto queste considerazioni è necessario che la finestra usata per calcolare la disparità sia il più estesa possibile lungo le Y e la più stretta possibile lungo l'asse X .

Se si osservano le due immagini IPM tuttavia si nota come gli ostacoli siano totalmente simili se confrontati lungo rette polari che passano per la proiezione del *pin-hole* e per il punto (X, Y) .

Prendiamo perciò in considerazione una proiezione IPM polare, ovvero

$$\begin{aligned} X &= X_0 + \rho \cos \vartheta \\ Y &= Y_0 + \rho \sin \vartheta \end{aligned} \quad (3.24)$$

dove $X^2 + Y^2 = \rho^2$.

In questo nuovo spazio immagine (ρ, ϑ) gli ostacoli sono proiettati correttamente lungo la direzione ρ , tuttavia si perdono i vantaggi della disparità.

L'equazione 3.13 può essere tuttavia trasformata in una proiezione semi-polare cercando di conservare sia la coordinata verticale ma anche la possibilità di analizzare l'immagine a ϑ costante, ovvero proiettando un punto di una immagine (u, v) in un punto di una immagine (X', ϑ) .

La coppia di equazioni 3.13 si possono riscrivere in forma semi-polare come

$$\begin{aligned} \tan \vartheta &= \frac{X' - X_0}{Y' - Y_0} = \frac{X - X_0}{Y - Y_0} \\ X' &= -h \frac{X - x_0}{Z - h} + x_0 \end{aligned} \quad (3.25)$$

La disparità ora si misura lungo l'asse ϑ ma è riportabile al caso originale attraverso l'equazione:

$$\Delta = Y_2' - Y_1' = X_1' \tan \vartheta_1 - X_2' \tan \vartheta_2 = X'(\tan \vartheta_1 - \tan \vartheta_2) - b \quad (3.26)$$

Questa immagine ha la proprietà di eseguire una trasformazione simile di oggetti nel mondo a X costante tra diverse immagini IPM prese da punti di vista differenti e permettere in questo modo di fare un mapping robusto tra le immagini. Come contro di tale tecnica, nel confronto di immagini generate dalla trasformazione 3.25, il terreno non ha più disparità nulla.

Capitolo 4

Modello puntiforme del veicolo

Questo capitolo si pone l'obiettivo di introdurre un modello semplificato del moto di un veicolo, conosciuta la velocità del veicolo stesso e la variazione dell'angolo di imbardata (*yaw-rate*). Il valore di *yaw-rate* è fornito da un sensore giroscopico solidale con il veicolo, possibilmente installato presso il centro di massa del veicolo stesso.

Sotto queste ipotesi è possibile approssimare il veicolo a un punto materiale, localizzato nel punto di massa.

La richiesta di sfruttare dati inerziali, come si vedrà nei capitoli del *motion-stereo*, nasce dal momento in cui si posseggano i dati odometrici del veicolo, imprecisi e a bassa risoluzione, ma tuttavia non si voglia usare l'individuazione dell'*ego-motion* per stimare il movimento del veicolo stesso.

Siano pertanto conosciuti la *velocità* v e lo *yaw-rate* γ del veicolo.

Istantaneamente per il veicolo valgono le relazioni del moto rettilineo uniforme:

$$\begin{aligned}\Gamma(t) &= \Gamma_0 + \gamma t \\ s(t) &= vt\end{aligned}\tag{4.1}$$

dove Γ è l'angolo di *yaw* assoluto, rispetto al valore relativo fornito dal sensore, e s è lo spazio percorso nell'istante di tempo (vedi figura 4.1). Γ_0 è l'angolo che il veicolo assumeva rispetto al sistema di riferimento al tempo $t = 0$. Questo modello suppone che per $t < T$, dove T sia non troppo grande, i valore di γ e v rimangano pressochè

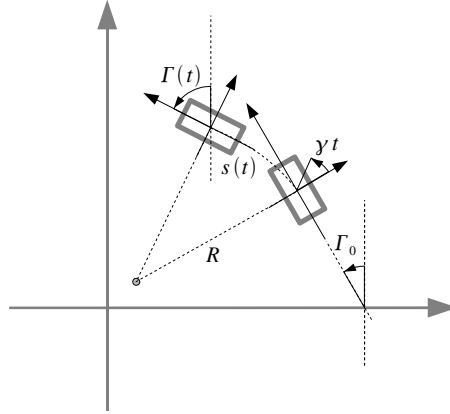


Figura 4.1: Veicolo in spostamento e rotazione

costanti. Dal punto di vista pratico T è il tempo che intercorre tra due letture dei parametri o dei fotogrammi ed è nell'ordine delle decine di millisecondi.

Se ci poniamo in un sistema di riferimento cartesiano (X crescente verso destra, Y crescente verso l'alto), da 4.1, per piccoli valori di γ , si può fare un'approssimazione (dettata anche dall'esperienza) per ricavare la variazione della posizione del veicolo:

$$\begin{aligned}\Delta x &= vT \cos(\Gamma_0) \\ \Delta y &= vT \sin(\Gamma_0)\end{aligned}\quad (4.2)$$

Tuttavia per grandi valori di γ questi valori tendono a differire dalla realtà.

Riportando le equazioni 4.2 in forma differenziale, calcolate per piccoli spostamenti di s ovvero di t , si ottiene che

$$\begin{aligned}ds &= v dt \\ dx &= \cos \Gamma(t) ds = v \cos(\Gamma_0 + \gamma t) dt \\ dy &= \sin \Gamma(t) ds = v \sin(\Gamma_0 + \gamma t) dt\end{aligned}\quad (4.3)$$

Integrando le equazioni 4.3 nel tempo tra 0 e T (tempo tra due fotogrammi dentro

il quale si considerano γ e v costanti) si ottengono le relazioni:

$$\begin{aligned}\Delta x &= 2\frac{v}{\gamma}\sin\left(\frac{\gamma}{2}T\right)\cos\left(\Gamma_0 + \frac{\gamma}{2}T\right) \\ \Delta y &= 2\frac{v}{\gamma}\sin\left(\frac{\gamma}{2}T\right)\sin\left(\Gamma_0 + \frac{\gamma}{2}T\right) \\ \Gamma_1 &= \Gamma_0 + \gamma T\end{aligned}\quad (4.4)$$

che convergono alle 4.2 quando $\gamma \rightarrow 0$.

È da notare che il vettore $(\Delta x, \Delta y)$ non ha orientazione Γ come invece accadeva nel caso 4.2.

Si definisce il fattore comune S come

$$S = 2\frac{v}{\gamma}\sin\left(\frac{\gamma}{2}T\right) \stackrel{\gamma \rightarrow 0}{\simeq} vT \quad (4.5)$$

in maniera da scrivere le espressioni 4.2 e 4.4 allo stesso modo:

$$\begin{aligned}\Delta x &= S\cos\left(\Gamma_0 + \frac{\gamma}{2}T\right) \\ \Delta y &= S\sin\left(\Gamma_0 + \frac{\gamma}{2}T\right)\end{aligned}\quad (4.6)$$

Altrimenti, ponendo $\Gamma_0 = 0$ (veicolo rivolto inizialmente verso destra), la funzione 4.4 si può scrivere in maniera più compatta anche come

$$\begin{aligned}\Delta x &= \frac{v}{\gamma}\sin(\gamma T) \\ \Delta y &= \frac{v}{\gamma}(1 - \cos(\gamma T))\end{aligned}\quad (4.7)$$

dove si vede molto bene come il veicolo compia archi di circonferenza. La condizione $\Gamma_0 = 0$ vuole anche dire che il veicolo è rivolto nella direzione X crescente, con l'asse Y crescente alla sua sinistra.

Usando l'equazione 4.6 è possibile ricavare la traiettoria del veicolo, ovvero ricavare (x_i, y_i) in coordinate mondo, conoscendo le informazioni inerziali v_i e γ_i .

In figura 4.2 è possibile vedere alcuni risultati del sistema in alcune sequenze di test del sistema di *Motion Stereo* con in particolare la sovrapposizione con una mappa aerea della zona in cui sono stati eseguiti i test.

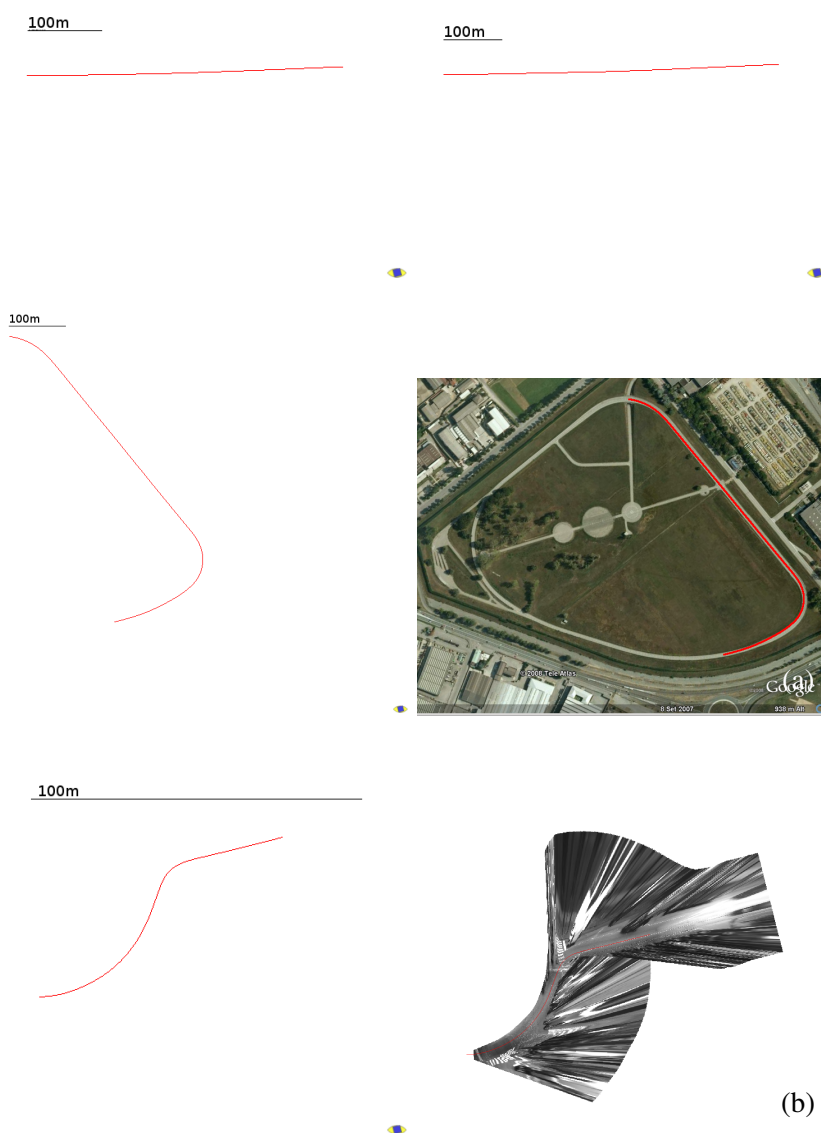


Figura 4.2: Traiettoria generata dal modello del veicolo puntiforme, sovrapposizione con una mappa (a) e generazione di una vista dell'ambiente percorso (b)

4.1 Raggio di curvatura

Definiamo il vettore posizione $\mathbf{r}(T)$ come

$$\mathbf{r} = \frac{v}{\gamma} \sin(\gamma T) \hat{\mathbf{x}} + \frac{v}{\gamma} (1 - \cos(\gamma T)) \hat{\mathbf{y}} \quad (4.8)$$

Il vettore tangente è definito come

$$\begin{aligned} \hat{\mathbf{T}} &= \frac{d\mathbf{r}}{dT} \\ &= v \cos(\gamma T) \hat{\mathbf{x}} + v \sin(\gamma T) \hat{\mathbf{y}} \end{aligned} \quad (4.9)$$

pertanto la curvatura è

$$R = \frac{v}{\gamma} \quad (4.10)$$

costante, e il cui centro di rotazione sarà

$$\begin{aligned} x_r &= x_0 - R \sin(\Gamma_0) \\ y_r &= y_0 + R \cos(\Gamma_0) \end{aligned} \quad (4.11)$$

4.2 Cambio di sistema di riferimento

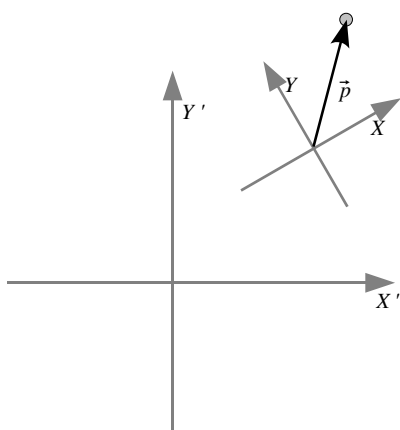


Figura 4.3: Sistema di riferimento

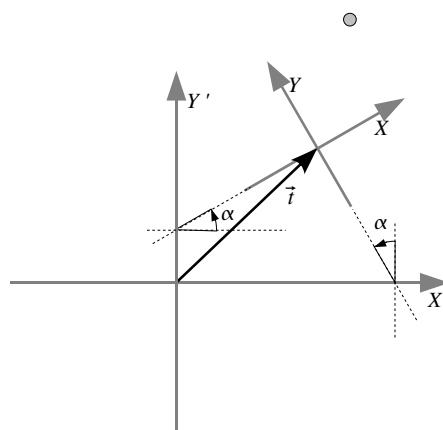


Figura 4.4: Relazione tra i due sistemi di riferimento

In questa sezione vengono riportate le formule usate poi nel capitolo del *Motion Stereo* per proiettare i punti dell'immagine IPM acquisiti in un tempo precedente sui rispettivi punti nell'immagine attuale. È da notare che quando si fa riferimento a coordinate veicolo e coordinate mondo queste sono tra di loro delle isometrie, infatti conservano la distanza tra punti, e il passaggio da una coordinata all'altra avviene solo per traslazioni e rotazioni, ma non attraverso cambiamenti di scala.

In figura 4.3 il punto p ha una coordinata diversa nei due sistemi di riferimento di figura. La differenza tra i centri dei due sistemi di riferimento è definita in \vec{t} e la rotazione dall'angolo α (si veda la figura 4.4). Queste considerazioni si possono fare quando si possiede la posizione e l'orientazione assoluta del veicolo nel mondo (è il caso per esempio dell'utilizzo di un sensore GPS).

Per calcolare quanto vale \vec{p} nel secondo sistema di riferimento, ovvero \vec{p}' , proiettiamo sugli assi del primo sistema di riferimento tale vettore, ottenendo le coordinate (p_x, p_y) (vedi figura 4.5). Per via puramente geometrica il punto \vec{p}' si costruisce $\vec{p}' = \vec{t} + \vec{p}_x + \vec{p}_y$. Bisogna pertanto calcolare il valore dei vettori (i valori delle loro componenti) \vec{p}_x e \vec{p}_y nel secondo sistema di riferimento.

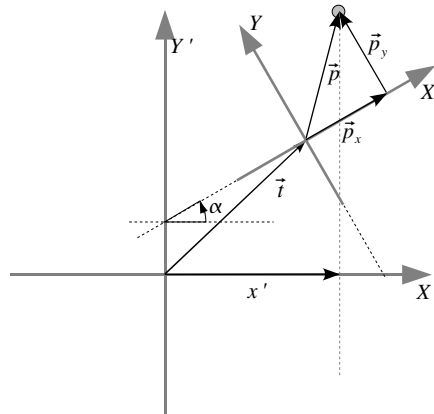


Figura 4.5: Proiezione di \vec{p}

Pertanto le coordinate di $\vec{p} = (x, y)$ nel secondo sistema di riferimento si possono

scrivere (in forma matriciale):

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} \cos \alpha & -\sin \alpha & t_x \\ \sin \alpha & \cos \alpha & t_y \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \quad (4.12)$$

e la trasformazione inversa si può scrivere come:

$$\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} \cos \alpha & \sin \alpha & -t_x \cos \alpha - t_y \sin \alpha \\ -\sin \alpha & \cos \alpha & +t_x \sin \alpha - t_y \cos \alpha \end{bmatrix} \begin{bmatrix} x' \\ y' \\ 1 \end{bmatrix} \quad (4.13)$$

4.2.1 Spostamento usando un altro sistema di riferimento

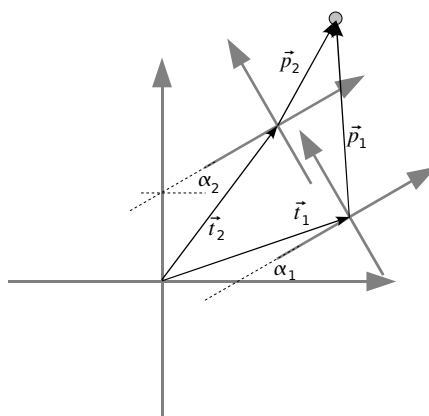


Figura 4.6: Spostamento del sistema di riferimento con dati assoluti

Uno dei problemi è calcolare lo spostamento che hanno avuto i punti del proprio sistema di riferimento (veicolo), usando tuttavia coordinate del sistema di riferimento secondo (mondo), come in figura 4.6.

Siano $\vec{p}_1 = (x_1, y_1)$ e $\vec{p}_2 = (x_2, y_2)$, espressi nelle coordinate nei rispettivi sistemi di riferimento, due vettori che puntano allo stesso punto in coordinate mondo.

Dalla composizione delle due trasformazioni si ricava, usando le formule di prostaferesi, la nuova coordinata nel nuovo sistema di riferimento (veicolo) di un punto

in coordinate mondo, la quale vale

$$\begin{bmatrix} x_2 \\ y_2 \end{bmatrix} = \begin{bmatrix} \cos(\Delta\alpha) & \sin(\Delta\alpha) & -\cos(\alpha_2)\Delta t_x - \sin(\alpha_2)\Delta t_y \\ -\sin(\Delta\alpha) & \cos(\Delta\alpha) & -\sin(\alpha_2)\Delta t_x - \cos(\alpha_2)\Delta t_y \end{bmatrix} \begin{bmatrix} x_1 \\ y_1 \\ 1 \end{bmatrix} \quad (4.14)$$

dove $\Delta\alpha = \alpha_2 - \alpha_1$ è la variazione di angolo e $\Delta\vec{t} = \vec{t}_2 - \vec{t}_1$ il vettore spostamento calcolati nel sistema di riferimento esterno.

4.2.2 Spostamento usando dati relativi

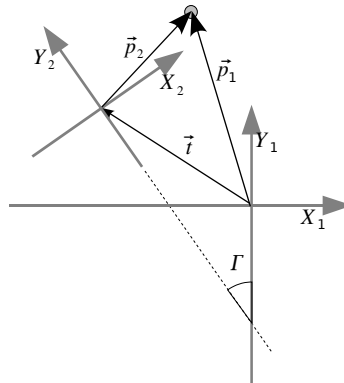


Figura 4.7: Spostamento del sistema di riferimento con dati relativi

Nel caso si possiedano dati relativi e non assoluti il procedimento è compatibile. Si può senza perdita di generalità considerare il caso in cui il sistema 1 sia il riferimento e calcolare dove si trova, secondo le equazioni 4.6, il veicolo nell'istante 2. La situazione è mostrata in figura 4.7.

Pertanto (x_1, y_1) è la posizione del medesimo punto nel mondo all'istante di tempo T_0 nel primo sistema di riferimento, (x_2, y_2) è la posizione del punto all'istante T_1 nel secondo sistema di riferimento.

Come è facile vedere si tratta della stessa situazione espressa dall'equazione 4.13

con la considerazione $\Gamma = \alpha = \gamma t$ (variazione di angolo tra i due istanti) :

$$\begin{aligned}x_2 &= \cos(\Gamma)x_1 + \sin(\Gamma)y_1 - t_x \cos(\Gamma) - t_y \sin(\Gamma) \\y_2 &= -\sin(\Gamma)x_1 + \cos(\Gamma)y_1 + t_x \sin(\Gamma) - t_y \cos(\Gamma)\end{aligned}\tag{4.15}$$

e l'inversa:

$$\begin{aligned}x_1 &= t_x + x_2 \cos(\Gamma) - y_2 \sin(\Gamma) \\y_1 &= t_y + x_2 \sin(\Gamma) + y_2 \cos(\Gamma)\end{aligned}\tag{4.16}$$

Infine si può usare l'equazione 4.7 per determinare \vec{l} .

Capitolo 5

Ground Plane Stereo IPM

La spinta iniziale per questa tecnica viene dal progetto *Apalaci-PreVENT*. All'interno di tale progetto è deciso di sviluppare una tecnica di riconoscimento ostacoli basata su una coppia di telecamere che osservino l'ampia zona frontale di un veicolo commerciale. L'analisi doveva essere limitata a condizioni in cui il veicolo era fermo o a velocità comunque limitate (inferiori a 30 km/h). Tale sistema doveva avere un basso peso computazionale ed essere semplice da calibrare.

L'utilizzo di due o più camere risulta praticamente un obbligo nel momento in cui si voglia ottenere una ricostruzione tridimensionale della scena affidabile mantenendo comunque i tempi computazionali contenuti. La visione stereo, rispetto alla sola tecnica basata su movimento che si andrà a discutere nel capitolo successivo, risulta più affidabile e permette di usare ipotesi più forti.

Nel momento in cui si sceglie di utilizzare più di una camera il problema della calibrazione tende ad aumentare di complessità perché, oltre a dover ricavare i parametri intrinseci ed estrinseci di ogni camera, risulta necessario conoscere la posizione precisa di ogni camera rispetto all'altra. Oltre a questa considerazione, nel caso pratico di voler analizzare un'ampia area di fronte a un veicolo, è richiesta una camera con lunghezza focale corta a cui corrisponde solitamente una distorsione dell'ottica. Nasce pertanto il problema di gestire contemporaneamente un certo numero di parametri, ognuno dei quali risulta affetto da seppur minimo errore.

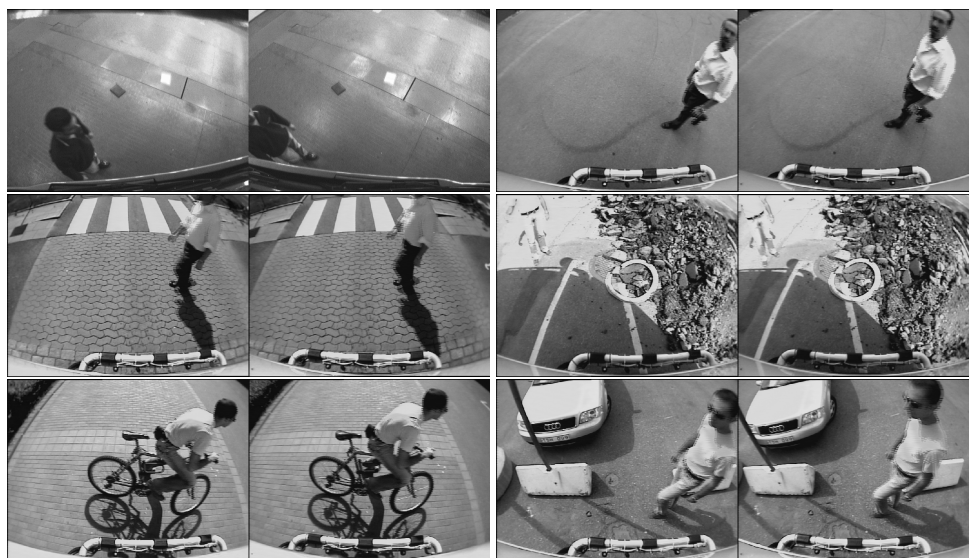


Figura 5.1: Esempio di casistiche analizzate per il sistema *Ground Plane Stereo*

Sotto queste ipotesi un classico sistema stereo basato su camere allineate e uso della disparità per ricostruzione tridimensionale e individuazione degli ostacoli risulta abbastanza complesso a causa principalmente dei problemi di calibrazione.

5.1 Ground Plane Stereo

I primi studi sulla stereoscopia fanno riferimento a tecniche chiamate di *Ground Plane Stereo* introdotte da Mallot [6]. La tecnica del *Ground Plane Stereo* o *Ground Plane Obstacle Detection* (GPOD) calcola (in generale *offline*) per ogni punto di un immagine un vettore *disparità* che permette di proiettare il punto di quell'immagine sul corrispondente punto prodotto dall'altra camera della coppia stereo, sotto la condizione che si stia inquadrando un piano. Dopo questa fase di calibrazione, se l'immagine di una camera, proiettata attraverso il campo vettoriale, risulta diversa da quella dell'altra camera siamo in presenza di un ostacolo.

In letteratura una certa serie di studi si concentrano su questa classe di sistemi in

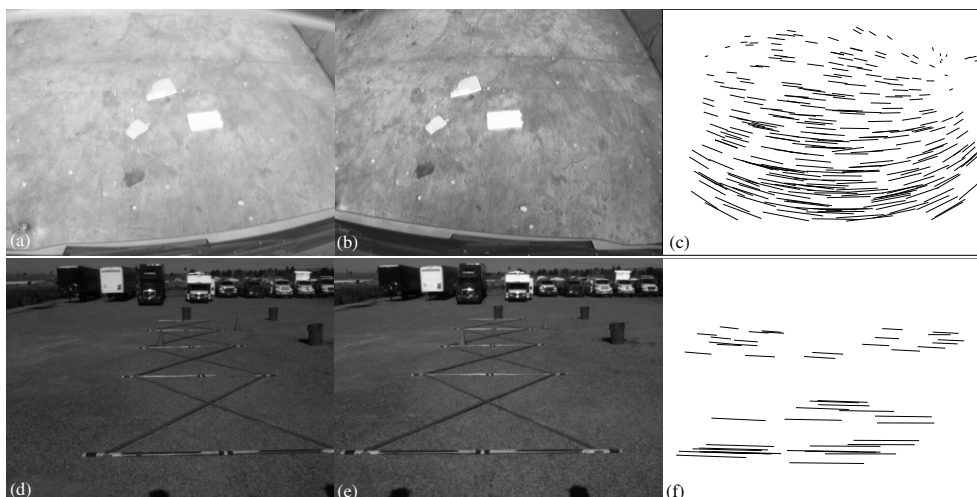


Figura 5.2: Esempio di *Ground Plane Stereo*: (a,d) Immagine sinistra, (b,e) immagine destra e (c,f) flusso ottico del terreno

grado di rilevare la presenza di ostacoli tra due immagini, anche distorte, attraverso la tecnica di *Ground Plane Stereo*. In particolare vengono affrontate in dettaglio le tecniche per modellare il campo vettoriale del flusso ottico che permette di proiettare un punto di un'immagine sull'altra nelle ipotesi che esso sia un piano. Tali funzioni spaziano dall'omografia [7, 8] per casi non distorti, a casi più generali che usano reti neurali [6, 9, 10].

Nel caso in cui le camere non siano distorte la funzione che proietta i punti di una immagine sull'altra è in effetti una omografia (vedi il risultato per esempio 5.3), mentre in generale è una funzione che permette di associare punti di un'immagine su quelli dell'altra. Indipendentemente da come è rappresentata la funzione tale procedimento si può fare in maniera automatica usando i moderni sistemi di individuazione di *feature*. Alcuni risultati di *matching* di *feature* tra coppie stereo e del corrispondente flusso ottico generato sono mostrati nella figura 5.2 usando come riconoscitore SURF [11]. In questo caso, non essendo densa, bisogna modellare in qualche modo la funzione vettoriale complessiva in modo da permettere la proiezione di ogni punto dell'immagine. Il vantaggio è che questa calibrazione si può ottenere del tutto in

maniera automatica senza la necessità di una griglia di calibrazione particolare, ma di contro non rimane disponibile nessuna informazione sulla posizione degli ostacoli nel mondo.

Nonostante tali potenzialità il *Ground Plane Stereo* è una tecnica poco utilizzata in letteratura, dove spesso viene preferito lo stereo classico basato su disparità in quanto può fornire informazioni importanti sulla tridimensionalità degli oggetti.

In questo lavoro si andrà invece a introdurre una tecnica innovativa di *Ground Plane Stereo* basata su IPM e verrà proposto un metodo per realizzare una ricostruzione tridimensionale basata su tale immagine. L'idea di base di sfruttare l'IPM per fare confronti su immagini stereo è ispirata dai lavori di Broggi e altri [12].

Il *Ground Plane Stereo* IPM si basa sull'effetto che la proiezione prospettica inversa di un piano inquadrato da punti di vista differenti deve essere sempre uguale, indipendentemente da come viene osservata la scena (si veda la dissertazione nel capitolo 3). Le uniche differenze che possono intercorrere sono effettivamente gli ostacoli verticali presenti nella scena. Partendo da questa considerazione è possibile realizzare un sistema stereo confrontando la 'proiezioni inverse' della stessa regione del piano prese da camere poste arbitrariamente in posizioni differenti. Il fatto di porre le camere in posizioni arbitrarie semplifica molto la fase di installazione e ne permette l'utilizzo in maniera diffusa su diversi veicoli dove possono intercorrere infatti limitazioni strutturali e aerodinamiche.

Se si considera un'immagine in prospettiva, *pixel* in zone diverse dell'immagine sottendono superfici dello spazio differenti in estensione. La regione di piano tridimensionale rappresentata da un *pixel* per esempio nella parte bassa dell'immagine ha un'estensione notevolmente inferiore di quella di un *pixel* vicino all'orizzonte. Sotto questa considerazione risulta difficile, a valle del confronto tra due immagini attraverso una tecnica di *Ground Plane Stereo* tradizionale, fornire un'unica soglia per la loro differenza in quanto ostacoli lontani richiedono di fatto soglie più basse di ostacoli vicini.

Il vantaggio invece dell'approccio qui mostrato basato sull'immagine IPM, sta proprio nel fatto che dopo la fase di rimozione della proiezione prospettica si lavora sempre in coordinate mondo e ogni punto di tale immagine rappresenta per

costruzione una parte dello spazio alla stessa risoluzione.

La calibrazione di questo approccio rimane infine più semplice rispetto a quella necessaria nello stereo classico basato su disparità e, come vedremo, tutti gli errori di calibrazione sono puramente locali e non intaccano globalmente l'immagine.

Infatti, invece di dover considerare fasi di calibrazione complicate dove a ogni fase è possibile introdurre rumore sui parametri estratti, nella tecnica di *Ground Plane Stereo IPM* bastano 4 punti per poter fare una proiezione prospettica inversa.

Consideriamo per semplicità il caso in cui la distorsione non è presente o trascurabile il campo vettoriale è rappresentabile da una trasformazione omografica.

Esplicitando il modello della pin-hole camera per le entrambe le camere

$$\begin{aligned} \begin{bmatrix} u_l \\ v_l \\ 1 \end{bmatrix} &= \mathbf{A}_l \mathbf{R}_l \left(\begin{bmatrix} X \\ Y \\ 0 \end{bmatrix} - \vec{t}_l \right) \\ \begin{bmatrix} u_r \\ v_r \\ 1 \end{bmatrix} &= \mathbf{A}_r \mathbf{R}_r \left(\begin{bmatrix} X \\ Y \\ 0 \end{bmatrix} - \vec{t}_r \right) \end{aligned} \quad (5.1)$$

dove si è indicato con $(X, Y, 0)$ il generico punto del mondo appartenere al terreno.

Dal sistema 5.1 è possibile unire tra loro \mathbf{AR} e incorporare il vettore \vec{t} con le considerazioni fatte per ottenere la matrice 3.4:

$$\begin{aligned} \begin{bmatrix} u_l \\ v_l \\ 1 \end{bmatrix} &= \mathbf{H}_l \begin{bmatrix} X \\ Y \\ 1 \end{bmatrix} \\ \begin{bmatrix} u_r \\ v_r \\ 1 \end{bmatrix} &= \mathbf{H}_r \begin{bmatrix} X \\ Y \\ 1 \end{bmatrix} \end{aligned} \quad (5.2)$$

dove viene sottolineato nuovamente che ogni trasformazioni prospettica inversa di un piano è un'omografia.

Risulta immediato dimostrare che la combinazione delle due viste è

$$\begin{bmatrix} u_l \\ v_l \\ 1 \end{bmatrix} = \mathbf{H}_{rl} \begin{bmatrix} u_r \\ v_r \\ 1 \end{bmatrix} \quad (5.3)$$

avendo definito $\mathbf{H}_{rl} = \mathbf{H}_l \mathbf{H}_r^{-1}$, ovvero che proiettare lo stesso piano da una immagine in prospettiva all'altra è anch'essa un'omografia.

I metodi basati su *Ground Plane Stereo* tradizionale usano la relazione 5.3 e la matrice \mathbf{H}_{rl} di fatto può essere generata automaticamente. Nel *Ground Plane Stereo IPM* viene invece usata la relazione 5.2, la quale obbliga alla determinazione di almeno una matrice in maniera esplicita, tuttavia conserva l'informazione sulla tridimensionalità del mondo.

5.2 Calibrazione

Andiamo in questa sezione a discutere delle tecniche utilizzate per generare le viste IPM per il *Ground Plane Stereo* nel caso in cui si usino ottiche a corta focale, quasi grandangolari.

Nel caso di ottiche distorte la relazione 5.2 non è più valida, tuttavia è sempre possibile stimare una generica funzione anche solo locale (approssimando i fenomeni non lineari) che permette di proiettare un punto da mondo a immagine e viceversa.

Tale procedimento viene eseguito completamente *offline* e tale funzione si ipotizza che rimanga costante, ovvero si suppone che i parametri delle camere non variano nel tempo. Ipotizziamo pertanto di metterci nelle condizioni in cui il veicolo sia praticamente fermo con le camere esenti da vibrazioni o variazioni degli angoli di calibrazione.

In questa situazione risulta conveniente realizzare una mappa che permetta di proiettare ogni punto dell'immagine della camera, pertanto affetta da distorsione e prospettiva, su un punto dell'immagine IPM. Per realizzare ciò si può disporre di una griglia di calibrazione sufficientemente densa e associare a ogni punto sulla griglia acquisito dalla camera un punto sull'immagine IPM.

Associati alcuni i punti tra l'immagine prospettica e l'immagine IPM, le corrispondenze dei punti residui andranno estrapolate attraverso una tecnica opportuna.

Partendo da una delle funzioni di 5.2, è possibile scrivere la relazione prospettica di una camera a cui viene applicata un generica funzione di distorsione radiale $f(r)$:

$$\begin{cases} \frac{h_0 * x + h_1 * y + h_2}{h_6 * x + h_7 * y + h_8} = i_u = (i_d - i_0)f(r_d^2) + i_0 = g_i(i_d, j_d) \\ \frac{h_3 * x + h_4 * y + h_5}{h_6 * x + h_7 * y + h_8} = j_u = (j_d - j_0)f(r_d^2) + j_0 = g_j(i_d, j_d) \end{cases} \quad (5.4)$$

Tale formalismo vuole volutamente mostrare la non esistenza di una forma chiusa e univoca per l'espressione di f e ciò che ne consegue.

Per ovviare a questa mancanza di conoscenza del sistema si possono percorrere diverse strade. Si può ricorrere all'espansione di Taylor della funzione $f(r)$ e ad approssimare l'equazione a una certa derivata, o altrimenti eseguire molteplici espansioni in serie locali una per ogni punto della griglia.

Genericamente l'unica cosa che si può infatti è che esisterà una generica funzione $\vec{p}(x,y)$ che trasforma punti da immagine IPM a immagine in prospettiva:

$$\begin{cases} p_u(x,y) = u_d \\ p_v(x,y) = v_d \end{cases} \quad (5.5)$$

dove si è scelta la sintassi di ricavare (u_d, v_d) in funzione di (x,y) in modo da avere un'immagine IPM densa.

Al sistema è possibile fornire un elenco di corrispondenze tra punti immagine con punti IPM che soddisfino dei vincoli del tipo:

$$\begin{cases} p_u(x_i, y_i) = u_{i,d} \\ p_v(x_i, y_i) = v_{i,d} \end{cases} \quad (5.6)$$

Sfruttando questi vincoli, la generica funzione \vec{p} può venire ricavata e infine quantizzata spazialmente nei valori (x,y) generando una *Look Up Table*.

5.2.1 Calibrazione con omografia e modello della lente

Allo stesso modo della tecnica discussa nella sezione 2.2 è possibile estrarre in un solo passaggio la funzione di distorsione e una trasformazione omografica, sfruttando

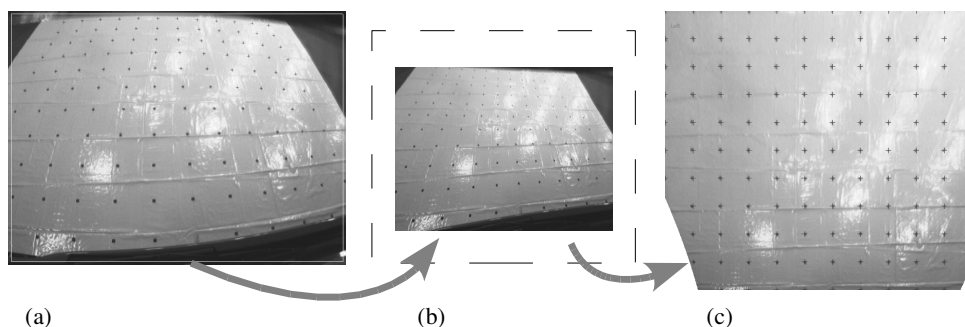


Figura 5.3: Calibrazione con omografia e modello della lente: griglia in prospettiva (a), griglia dedistorta (b) e griglia nell'immagine IPM (c)

l'idea che punti su un piano, anche se in prospettiva, devono rimanere punti di un piano con un certo legame tra loro.

Attraverso un minimizzatore non lineare viene valutato un potenziale modello parametrico della lente. Con questo modello i punti della griglia in coordinate immagine $(u_{i,d}, v_{i,d})$ vengono trasformati in una forma pi o meno dedistorta (u'_i, v'_i) . Usando le coordinate di tali punti e le corrispondenti nell'immagine IPM (x_i, y_i) è possibile, usando l'equazione 2.2, ricavare una trasformazione omografica che permetta di trasformare le une alle altre. Essendo il numero di punti i maggiori di 4 la trasformazione omografica non sarà perfetta, ma sarà una regressione ai minimi quadrati.

Il calcolo successivo dello scarto tra i punti reali nell'immagine IPM e quelli prodotti dall'omografia è in effetti un indice di quanto i punti generati dal modello della lente siano effettivamente dedistorti bene. Tale scarto è usato come funzione errore per il minimizzatore che eventualmente proporrà, grazie ad esso, un nuovo modello per la lente.

Alla fine della fase di minimizzazione viene fornito il modello della lente e la matrice della trasformazione omografica che permette di generare la *Look Up Table* (figura 5.3).

I parametri della lente estratti in questo modo non rappresentano in effetti il mi-

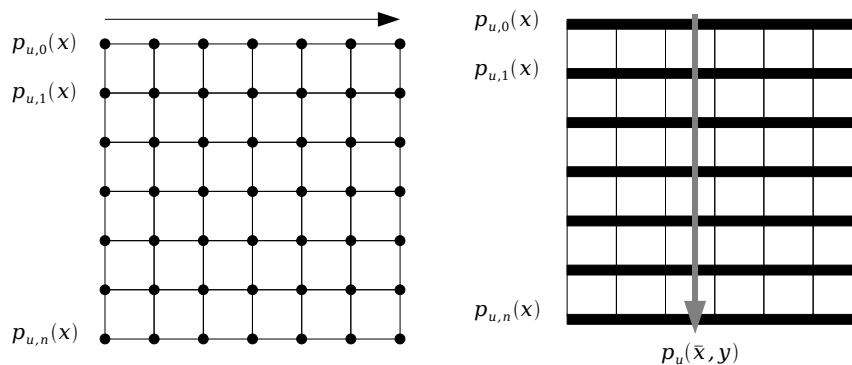


Figura 5.4: Calibrazione con generazione di funzioni spline: prima vengono generate le spline sui punti conosciuti orizzontali, e poi per tutti i punti in verticale

gior modello della distorsione, in quanto questo non è lo scopo che tale sistema si prefigge, ma rappresentano invece la miglior trasformazione radiale che si può applicare ai punti immagine in modo che generi la miglior omografia.

5.2.2 Calibrazione con spline

Il metodo descritto in precedenza ricava una regressione ai minimi quadrati dei punti della griglia e perciò non viene forzato che i punti della trasformazione finale soddisfino esattamente i vincoli imposti (l'esatta corrispondenza tra i punti indicati e quelli ottenuti dalla funzione).

Per conservare questi vincoli e per mantenere la continuità della funzione $\vec{p}(x, y)$ sono utilizzate funzioni polinomiali a tratti, nel caso specifico spline lineari e cubiche.

La funzione $p_u(x, y)$ e $p_v(x, y)$ diventano funzioni polinomiali a tratti, che interpolano i punti dati, mantenendo fisso il grado del polinomio, ma dividendo l'intervallo in numero pari al numero di elementi della griglia.

Questo metodo richiede che lungo una direzione, ad esempio la y , i punti della griglia siano spazati in maniera uniforme. Partendo dai vincoli 5.6, vengono generate le funzioni $p_{u,y_k}(x)$ e $p_{v,y_k}(x)$, tante funzioni quanti sono gli elementi della griglia lun-

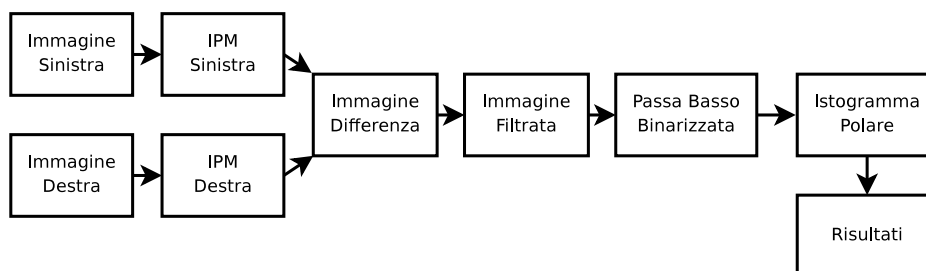


Figura 5.5: Diagramma di Flusso del sistema Ground Plane Stereo IPM

go l'asse y . Interpolando queste funzioni, vengono infine generate $p_u(x,y)$ e $p_v(x,y)$ (figura 5.4).

5.3 Implementazione

Il diagramma di flusso dell'algoritmo è mostrato in figura 5.5. I principi base dell'elaborazione mostrati in un caso ideale sono presentati in figura 5.6.

Nella prima fase a entrambe le immagini acquisite dalla camera (figure 5.6.a e 5.6.b) è rimossa la prospettiva e la distorsione (figure 5.6.c e 5.6.d) in un solo passaggio attraverso l'applicazione di *Look Up Table* generate durante la fase di calibrazione. Risultati su immagini reali sono visibili nelle figure di immagine 5.7.

Inquadrando la stessa scena da punti di vista differenti, le zone che presentano variazioni di luminosità nelle due immagini IPM sono dovute principalmente a ostacoli verticali o rumore (esempio piccoli errori di calibrazione o riflessi). Per evidenziare queste particolari aree viene fatta la differenza punto a punto tra queste due immagini IPM (risultato in figura 5.6.e). Nel caso le immagini in ingresso siano a colori vengono sommate le differenze nelle componenti RGB per ottenere un unico valore di differenza per ogni punto.

Il sistema per come è concepito non è molto robusto al rumore, in quanto le *Look Up Table* sono prodotte in condizioni di riposo e potrebbero subire variazioni temporanee a causa di vibrazioni. Nel caso in cui le variazioni siano comunque piccole, è

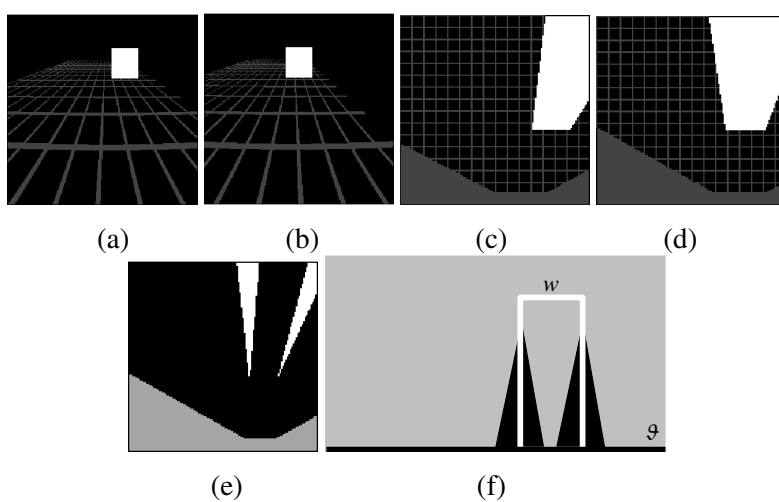


Figura 5.6: Caso Ideale: Immagini sorgenti (a)(b), vista IPM (c)(d), loro immagine differenza (e) e istogramma polare (f)

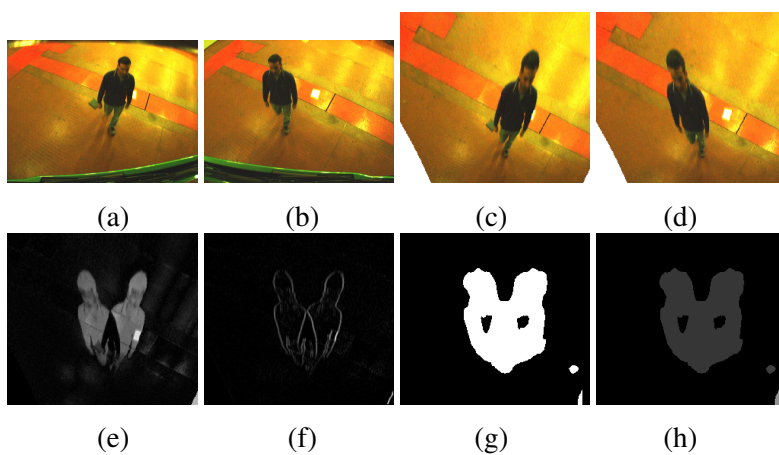


Figura 5.7: Ground Plane Stereo IPM: Immagini sorgenti (a)(b), vista IPM (c)(d), Immagine differenza (e), filtrata per rimuovere il rumore (f), immagine binaria (g) ed etichettata (h)

possibile procedere a un filtraggio locale per accentuare alcune componenti e filtrare via rumore.

Esperimenti su filtri che possono venire applicati in cascata sono

- *Minimo*, che assegna al pixel il valore minimo dei punti nell'intorno 3×3 ;
- *Sobel*, che rimuove la componente continua dell'immagine con un filtro derivativo.

Il Filtro *Minimo* rigetta buona parte del rumore ma mantiene comunque informazione sulle differenze quando queste occupano un'area maggiore dell'intorno. Il Filtro di *Sobel* permette di rigettare i casi in cui tra le due immagini ci sia una differenza di luminosità globale, a causa di una diversa illuminazione della scena da parte dei due punti di vista.

Le prestazioni di questi filtri sono mostrate nei grafici in figura 5.11.

5.3.1 Filtro passa basso

Per poter filtrare le zone con basse differenze, ma allo stesso tempo filtrare passa basso spazialmente l'immagine viene introdotto un particolare filtro.

Tale filtro calcola il valor medio dell'immagine differenza in un intorno rettangolare e fornisce come risultato un'immagine binaria nel caso in cui tale valor medio sia sopra una determinata soglia ξ

Sia D l'immagine differenza da processare e sia A_i un intorno del punto i in cui vogliamo calcolare l'immagine binaria T_i

$$\forall i \in T \quad m_i = \frac{\sum_{j \in A_i} D_j}{s} \quad T_i = \begin{cases} 0 & \text{se } m_i < \xi \\ 1 & \text{se } m_i > \xi \end{cases} \quad (5.7)$$

dove con $s = \sum_A 1$ è indicata il fattore di normalizzazione che equivale alla dimensione in pixel dei vari insiemi A_i .

Tale funzione è stata scelta perché oltre a fornire un risultato adatto allo scopo, nel caso in cui gli intorni A_i siano di forma rettangolare l'immagine risultante T è

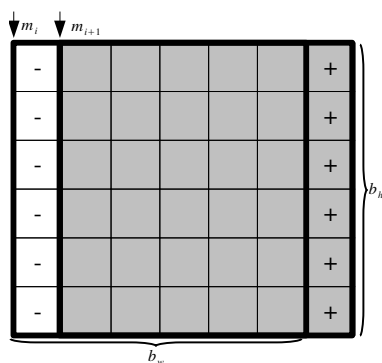


Figura 5.8: Costruzione incrementale dell'immagine del valor medio

ottenibile in maniera efficiente. Infatti sia $b_w \times b_h$ la dimensione degli intorno A_i . Ne consegue che il numero di operazioni necessarie per il passaggio da m_i a m_{i+1} è costante e vale $2b_h$ somme (vedi figura 5.8).

5.3.2 Funzione di Accumulazione Polare

Il sistema ora dispone di un immagine binaria che rappresenta indicativamente gli ostacoli presenti nella scena. Per permettere una rimozione ulteriore del rumore e localizzare con più precisione tali ostacoli viene presa in esame una proprietà delle immagini IPM.

Nell'equazione 3.13 veniva mostrato che ogni ostacolo verticale sull'immagine IPM è una retta passante per il *pin-hole* e per la proiezione sul terreno dell'ostacolo stesso. Nel caso stereo, in cui i *pin-hole* sono due e tra queste due immagini IPM viene fatta una differenza, si crea una struttura tipicamente a V, dove la base di tale struttura sull'immagine IPM è effettivamente la coordinata (X, Y) dell'ostacolo. Siccome gli ostacoli sono oggetti tipicamente verticali, sull'immagine IPM saranno rappresentati da strutture che convergono verso il rispettivo pin-hole.

Per sfruttare questo fenomeno, viene definita la Funzione di Accumulazione Polare *Polar Accumulation Function* (PAF) dove gli ostacoli vengono contati a seconda dell'angolo che formano in coordinate polari (figura 5.9). Il punto nell'immagine

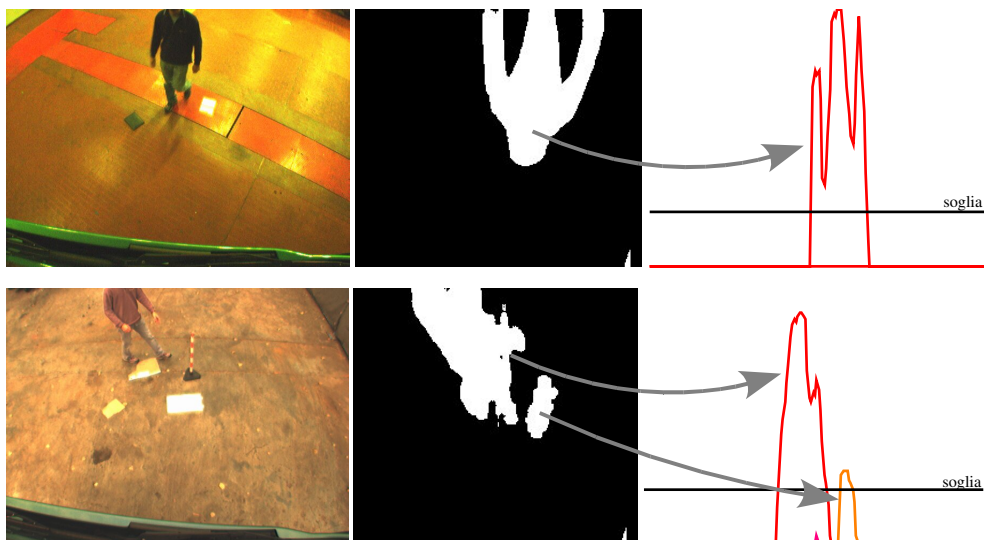


Figura 5.9: Esempio di creazione del diagramma polare

IPM (X, Y) , accesso nell'immagine binaria, contribuisce alla coordinata ϑ secondo l'espressione

$$\vartheta = \tan^{-1} \frac{Y - Y_0}{X - X_0} \quad (5.8)$$

dove (X_0, Y_0) è la proiezione di uno dei *pin-hole* sotto il quale viene osservata la scena. L'istogramma viene infine normalizzato considerando il numero effettivo di punti che cadono lungo quella direttrice.

Viene generato un istogramma per ogni label e con questo istogramma viene valutato se tale oggetto possa essere considerato un ostacolo oppure meno. Vengono eseguite soglie sul massimo e sulla larghezza in gradi w (figura 5.10). I label che passano la selezione saranno riportati al sistema e le coordinate IPM del punto più vicino al pin-hole (in figura k) vengono usate per indicare la posizione nel mondo e nell'immagine di quell'ostacolo.

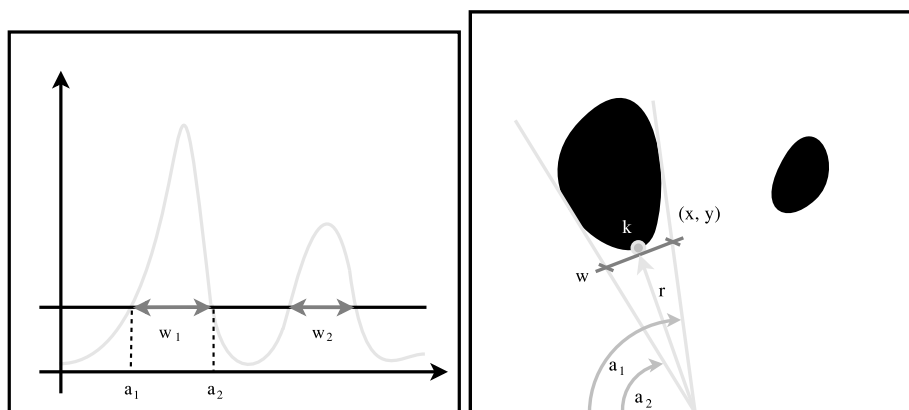


Figura 5.10: Funzione di Accumulazione Polare: (a) Esempio di soglia sulla PAF, (b) delimitazioni del blob

5.4 Risultati

Nelle figure 5.11 e 5.12 sono mostrate delle curve ROC per quanto riguarda le prestazioni dei filtri nell'elaborazione di due fotogrammi presi in due situazioni diverse. L'asse delle 'detezioni corrette' (cd) rappresenta il numero di punti dell'immagine IPM correttamente individuati come ostacolo, mentre sull'asse dei 'falsi positivi' (fp) è indicata la quantità (normalizzata a 1) dell'immagine IPM occupata da pixel segnati come ostacolo sul suolo.

La figura 5.11 mostrano gli andamenti delle curve quando si sceglie di non usare nessun filtro sull'immagine differenza, si sceglie di usare solo il filtro di sobel, o usare sia il filtro di sobel che il filtro di minimo locale.

La figura 5.12 mostrano gli andamenti delle curve all'aumentare della dimensione della finestra usata per il filtro passa basso.

Immagini sui risultati dell'algoritmo sono mostrati in figura 5.13. Il sistema ha dato capacità di riconoscere gli ostacoli presenti di fronte al veicolo, indipendentemente dalla loro forma, orientazione e posizione all'interno dell'immagine. A causa delle soglie dell'istogramma polare indipendenti dalla posizione del mondo dell'ostacolo, il sistema si dimostra cauto nel riconoscere ostacoli parzialmente esterni

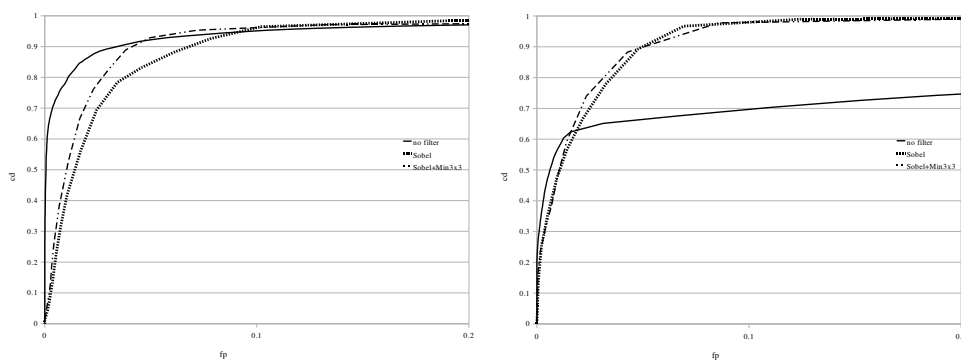


Figura 5.11: Curve ROC per i filtro di pre-processamento dell'immagine differenza

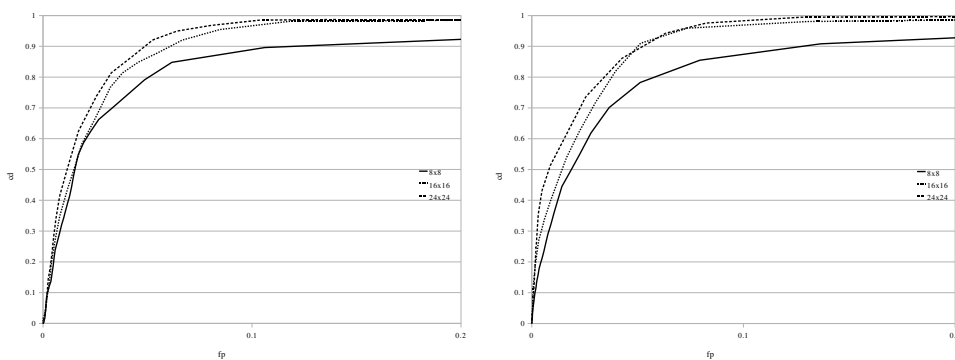


Figura 5.12: Curve ROC per la dimensione della finestra nel filtro passa basso

Figura 5.13: Esempio di risultati prodotti dal sistema *Ground Plane Stereo IPM*

Architettura	tempo di elaborazione
Pentium4 a 3.2 Ghz	18 ms
Via EPIA EN15000 a 1.5 GHz	80 ms
TI DM642 a 600 Mhz	48 ms
Intel XScale 80200 a 400Mhz	180 ms

Tabella 5.1: Prestazioni del sistema di *Ground Plane Stereo IPM*

all'immagine situati sui bordi di essa.

La posizione dell'ostacolo è sempre fornita con un elevato grado di precisione e la stima dell'occupazione dell'istogramma polare permette di stimare anche l'estensione dell'ostacolo.

Per quanto riguarda il peso computazionale, le prestazioni del sistema sono mostrate in tabella 5.1. L'algoritmo è stato portato con successo su diverse architetture, tra cui DSP e ARM9.

5.4.1 Disparità su immagini IPM

Il sistema mostrato finora ha prestazioni computazionali molto elevate, ma non permette un riconoscimento preciso della posizione dell'ostacolo nel mondo. Sfruttando le considerazioni fatte nel capitolo 3 e in particolare delle equazioni 3.18 e 3.19 risulta possibile una ricostruzione tridimensionale basata su immagini IPM.

In questo caso dalle immagini IPM viene generata l'immagine di disparità e con questa e le equazioni sopracitate è possibile fornire una rappresentazione dell'area circostante (figura 5.14).

L'immagine disparità permette un'analisi più dettagliata dell'immagine in quanto permette di indicare per ogni punto le sue coordinate nel mondo. Questo è sicuramente più raffinato del semplice blob fornito dall'algoritmo base, a scapito di un aumento della capacità computazionale. Di fatto, per migliorare le prestazioni del sistema, l'analisi della disparità può essere fatta solamente all'interno dell'area indicata dai blob del *Ground Plane Stereo IPM*.

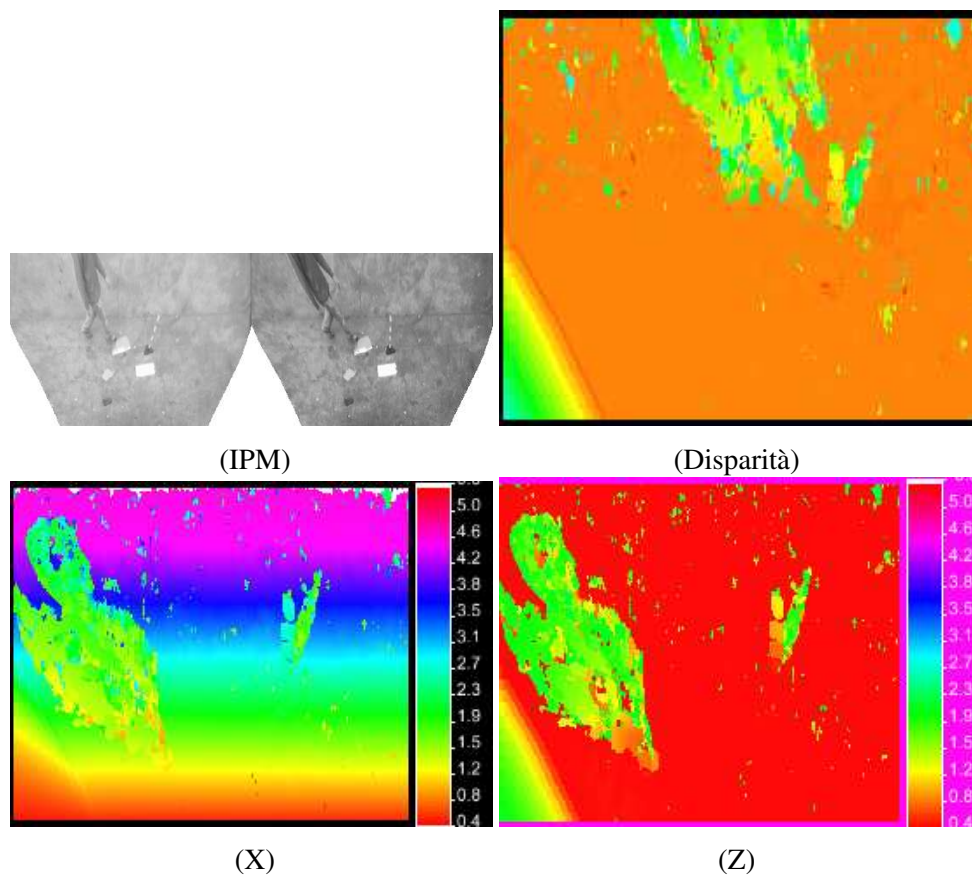


Figura 5.14: Esempio di ricostruzione tridimensionale usando la disparità su IPM. Coordinata X e coordinata Z



Figura 5.15: Esempi di installazioni del sistema *Ground Plane Stereo IPM*. Il riquadro rosso indica dove è stata installata la coppia di videocamere

La disparità deve essere calcolata su una finestra larga e bassa in modo da ridurre gli effetti di distorsione che l'immagine IPM esercita sugli ostacoli verticali.

5.5 conclusioni

In questo capitolo è stato presentato un algoritmo di riconoscimento ostacoli basato sull'analisi congiunta di coppia di immagini IPM stereo. Le immagini IPM sono generate grazie all'utilizzo di *Look Up Table* generate *offline* con una griglia di calibrazione.

Il sistema è stato installato su diversi veicoli di prova (figura 5.15) a diverse altezze (3.0 m, 2.0 m, 1.5 m) e con aree di interesse differenti (rispettivamente 10 m, 7 m e 15 m).

La particolarità del sistema è che non obbliga l'installazione delle camere in maniera particolare, con il solo vincolo che solo l'area inquadrata da entrambe le camere è quella utilizzata nel riconoscimento ostacoli. Questo permette di semplificare la fase di setup del sistema e non complica in maniera eccessiva la fase di calibrazione.

Una peculiarità di utilizzare funzioni locali per generare la *Look Up Table* fa

in modo che errori sui punti lontani non vadano a influenzare i punti localmente. Allo stesso modo usando questa tecnica è possibile generare immagini IPM anche di *superfici non piane* per applicazioni diverse da quelle automobilistiche dove le camere possono essere installate in maniera fissa e osservare una scena di dinamica complessa.

Grazie all'analisi degli ostacoli nell'immagine IPM, la precisione di riconoscimento è costante in tutto lo spazio, a differenza di una tecnica basata su disparità classica.

Il sistema è in grado di segnalare quale porzione dell'immagine contiene un potenziale ostacolo e fornire la coordinata del punto di contatto con il terreno. Per estendere questo comportamento al caso di una rappresentazione tridimensionale è stato infine proposto un metodo per estrarre la tridimensionalità dell'immagine da viste IPM.

5.5.1 Sviluppi futuri

Gli sviluppi che si stanno facendo su questa tecnica sono principalmente

- utilizzo di due istogrammi di accumulazione polare, uno per ogni *pin-hole* della scena ed eseguire in questo modo valutazioni separate su entrambi;
- la disparità su immagini IPM semi-polari (si veda l'equazione 3.24) che permette di ridurre la deformazione degli ostacoli e migliorare il confronto tra le due immagini.

Uno dei problemi tuttora aperti rimane quello di gestire le variazioni degli angoli delle camere, a causa del beccheggio istantaneo per esempio, e riuscire a modificare di conseguenza le *Look Up Table* che generano l'IPM.

Capitolo 6

Motion Stereo IPM

Un'altra delle applicazioni da sviluppare all'interno del progetto *Apalaci-PreVENT* consisteva in un riconoscitore di ostacoli e successiva loro classificazione in ambiente stradale. Veniva in particolare richiesta la possibilità di individuare sia veicoli che pedoni, con l'unico vincolo progettuale quello di usare una camera monocromatica e gli altri sensori messi a disposizione.

Essendo la camera parte di un sistema multiapplicazione, solo una parte del sensore (640×300 pixel) era disponibile per questa elaborazione.

La piattaforma di elaborazione sulla quale il sistema doveva venir eseguito era un Pentium4 a 2.0 Ghz. Sul veicolo su cui l'algoritmo è stato sviluppato era disponibile un Radar a 24 Ghz, un giroscopio capace di fornire la variazione di imbardata del veicolo e l'informazione sulla velocità fornita dal veicolo stesso, tuttavia con precisione limitata a 1 km/h.

Sotto questi vincoli architettureali e a causa delle peculiarità dell'hardware, per non appesantire troppo l'elaborazione, si è scelto di sviluppare un metodo per estrarre dall'immagine regioni di interesse nelle quali, attraverso ulteriori tecniche, estrarre e classificare eventuali ostacoli.

È da ribadire l'algoritmo qui sviluppato di *Motion Stereo* si avvicina in effetti più a una tecnica di individuazione di spazio libero piuttosto che di ricostruzione tridimensionale, in quanto l'interesse non era quello di realizzare una ricostruzione



Figura 6.1: Il veicolo fornito dal Centro Ricerche Fiat usato nel progetto

tridimensionale precisa dell'area frontale ma solo indicare le zone potenzialmente occupate da ostacoli verticali.

La tipologia di ambienti esaminata per il problema è rappresentata da una scelta di fotogrammi in figura 6.2. Il sistema si propone di estrarre, principalmente pedoni, in immagini proveniente da ambienti urbani, extraurbani e in strade ad alta percorribilità. Oltre alle immagini sono stati acquisite anche le informazioni inerziali provenienti dall'odometria del veicolo per la velocità e da un sensore giroscopico per l'imbardata del veicolo.

6.1 Stato dell'Arte

L'obiettivo di questo lavoro era di cercare di sviluppare una tecnica innovativa per il riconoscimento ostacoli. L'attenzione pertanto era caduta sull'utilizzo del *Motion Stereo* su immagini IPM. Visto che studi sull'utilizzo della coppia IPM e moto come base per riconoscimento di ostacoli non sono diffusi in letteratura, lo studio preliminare a questo lavoro si è basato generalmente su sistemi di riconoscimento ostacoli con una sola camera che sfruttano l'informazione sul movimento, con particolare attenzione all'ambito automobilistico. Nel capitolo 5 si è già parlato in dettaglio della tecnica del *Ground Plane Stereo* e perciò diversi aspetti di questa discussione possono essere considerati sfondo per questo argomento.



Figura 6.2: Esempio di ambienti analizzati per il sistema *Motion Stereo*

Diversi studi mostrano come nella visione umana l'informazione inerziale, fornita principalmente dal sistema vestibolare, è usata, oltre che direttamente per la navigazione, orientamento e l'equilibrio, anche per analizzare il mondo che ci circonda, ovvero per permettere al cervello di eseguire confronti nel tempo oltre che nello spazio tra le immagini percepite.

Lobo e Dias [13] partono da questa idea nel discutere un sistema basato su informazioni inerziali (accelerometri e giroscopi) uniti a un sistema di visione stereo, anche se in questo caso l'inclinometro è usato per stimare il piano del terreno e segmentare i punti immagine di conseguenza. Tale utilizzo combinato della visione stereoscopica e del movimento fa parte delle tecniche chiamate di *Temporal Stereo*.

C'è da notare che diverse ricerche affrontano in generale il problema senza far ricorso a sensori inerziali e usando una sola camera in due principali ambiti di ricerca: quello della *Structure From Motion*, che si occupa di ricostruzione tridimensionale dove l'oggetto è in movimento rispetto alla camera, e quello dell'*ego-motion*, dove viene ricostruita la traiettoria eseguita dalla camera nel mondo. Tecnicamente non esiste esplicita differenza tra il concetto di *Motion Stereo* e quello di *Structure From Motion*, anche se nel secondo è contemplato anche il caso in cui sia l'oggetto in esame a muoversi e la camera a rimanere ferma.

Entrambi questi campi sono abbastanza ampi da trattare, tuttavia l'analisi dello stato dell'arte si può limitare all'ambito automobilistico in cui sono espresse il maggior numero di problematiche. Tra le svariate problematiche di fatto quella centrale negli approcci monoscopici è certamente il fatto di non poter separare il moto dovuto al movimento del veicolo dal moto degli eventuali altri veicoli, e questa difficoltà intrinseca al problema è comune in tutti gli approcci esaminati.

Le prime implementazioni del *motion stereo* cominciano ad essere discusse nei primi anni '90. Bhanu e altri [14] usano l'informazione del sensore inerziale per stimare il *Focus of Expansion* e con questo valutare se le *feature* tracciate siano nella posizione predetta da tale movimento.

In ambito prettamente più automobilistico, Kuhnert [15] presenta un lavoro di riconoscimento della distanza di ostacoli in strade ad alta percorribilità dove anche qui un sensore inerziale è usato per fornire lo stato di ricerca iniziale della soluzione.

Huber e Graefe [16] partendo dall'ipotesi di riuscire a prendere due punti del medesimo oggetto e ipotizzando che tali punti siano alla stessa distanza dalla camera, riescono a calcolare una relazione tra la variazione nel tempo della distanza in pixel e la distanza in metri dalla camera, e a fornire in forma esplicita il tempo di collisione con l'oggetto.

Sempre in ambito automobilistico esistono applicazioni particolari dove l'ambiente è rigorosamente statico, e si possono applicare tecniche pure di *Structure From Motion* come nel caso di ricostruzione dell'ambiente circostante per applicazioni di parcheggio intelligente ed assistiti [17, 18].

La scelta di usare tracciare *feature* tra le immagini temporalmente differenti è solitamente una scelta meno costosa dal punto di vista computazionale. Inoltre le immagini di ambienti stradali contengono larghe zone omogenee che forniscono solamente poca informazione per stimare contemporaneamente moto e ostacoli.

Un esempio di utilizzo di *feature* per il riconoscimento del piano dominante dell'immagine (il terreno) senza l'utilizzo di informazioni inerziali è mostrato da Zhou e Li [19] i quali usano una sola camera e sfruttano solamente il movimento avvertito attraverso le immagini, senza sensori inerziali. Viene sfruttata la considerazione che un piano comunque rimarrà un piano dopo lo spostamento della camera. Siccome la trasformazione omografica tra viste prospettiche è l'unica trasformazione che permette di trasferire piani in piani solo i punti che soddisferanno tale vincolo apparterranno al terreno. In questo lavoro, usando una semplice tecnica di tracciamento dei *corner di Harris* [20] e usando come funzione di confronto la semplice SAD (Sum of Absolute Difference) su una finestra nell'intorno della *feature*, viene mostrato come sia possibile generare una lista di corrispondenze tra due fotogrammi acquisiti dalla stessa camera in istanti di tempo differente. Usando un metodo che massimizza il numero di coppie (per esempio RANSAC) di punti che soddisfano il modello di un piano è possibile estrarre il piano dominante dell'immagine (tecnica definita di *H-Clustering*). Questa trasformazione omografica può essere usata poi per classificare i punti che non appartengono a questo piano come eventuali ostacoli. Un problema di questo approccio è che se il movimento della camera è una pura rotazione (o comunque la traslazione è trascurabile) tutte le *feature* soddisfano comunque un

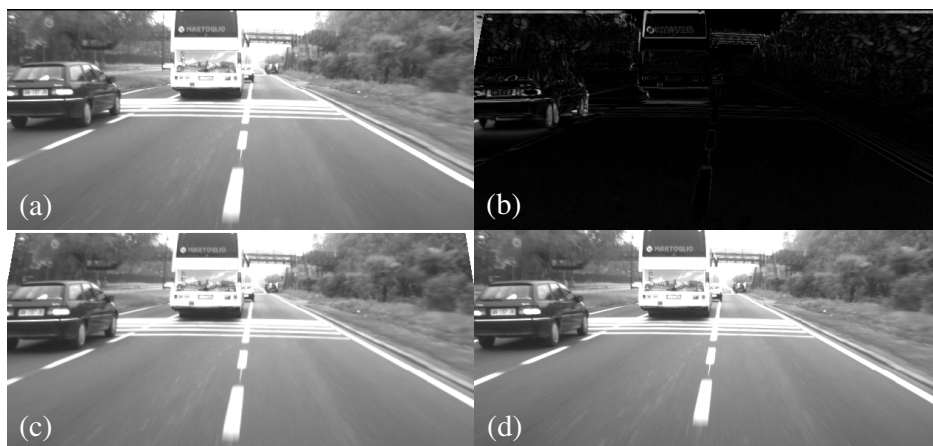


Figura 6.3: Esempio di rilevazione di ostacoli basato su *Warped Image*: il fotogramma precedentemente acquisito (*d*) viene trasformato usando i dati inerziali (*c*) e confrontato con il fotogramma corrente (*a*) ottenendo un immagine differenza (*b*)

vincolo omografico (un esempio lampante infatti è il caso di immagini panoramiche) e non è possibile classificare i punti in base a questa tecnica. L'utilizzo in questo caso di un sensore inerziale permetterebbe di scartare alcune soluzioni errate dei piani proposti, e cercare la trasformazione solo tra quelle permesse dall'informazione sul movimento.

Questo lavoro di tesi è stato ispirato principalmente da due articoli [21, 22] entrambi basati sul concetto di *warped image*. L'idea di fondo è quella di riuscire a proiettare il fotogramma acquisito in un tempo precedente sul frame corrente attraverso la conoscenza del moto del veicolo. Tuttavia siccome non si ha nessuna conoscenza della tridimensionalità dell'immagine che si vuole riproiettare, uno degli approcci possibili è quello di fissare un vincolo (ad esempio $Z = 0$) e riproiettare tutto come se fosse piano conosciuto, fissando un ipotesi di mondo piatto (indicata come *flat-world assumption*). In tal modo la regione dell'immagine appartenente a questo piano combacia tra l'immagine corrente e quella precedente riproiettata, mentre tutto ciò che non appartiene al piano (ostacoli verticali) viene piazzato in maniera errata e provoca

una differenza nel confronto (figura 6.3).

È da notare che tali tecniche sono limitate a quella parte dello spazio sotto l'orizzonte siccome qualsiasi punto sopra tale linea non potrà appartenere sicuramente al piano in esame.

Nel lavoro di Batavia e altri [22], usando le informazioni odometriche del moto del veicolo, vengono proiettati solamente i bordi, essendo strutture dense di informazione, dall'immagine acquisita al fotogramma precedente alla posizione predetta al fotogramma corrente sotto l'ipotesi che tali punti appartengano al piano stradale. Il crollo di questa ipotesi, ovvero che il bordo sia parte di un ostacolo verticale, fa fallire il confronto ed è indice della presenza di un ostacolo. Uno dei problemi in ambiente stradale che verrà discusso in seguito, la variazione dell'imbardata del veicolo, viene filtrata usando una tecnica di tracking verticale degli stessi bordi, chiamata flusso ottico 1-dimensionale.

Aubert e altri [21] propongono un approccio per il *motion stereo* basato totalmente sulla *warped image*. In questo caso le informazioni sul moto del veicolo sono fornite dall'odometria per la velocità e da un giroscopio per le rotazioni, entrambi sensori normalmente presenti su un veicolo commerciale (in particolare il sensore di *yaw-rate* è sfruttato dall'ESP). In questo caso le regioni che non soddisfano il confronto tra l'immagine corrente e l'immagine riproiettata sono le potenziali regioni verticali. Su tale regioni viene implementata una specie di V-disparity [23] per individuare in maniera robusta gli ostacoli. Essendo una riproiezione perfetta del piano stradale la V-Disparity della strada rappresenta una retta verticale, e tutto ciò che si discosta da tale retta risulta essere un ostacolo. In questo lavoro le problematiche relative al beccheggio del veicolo dovrebbero risultare filtrate, in quanto la variazione del *pitch* provoca una trasformazione della retta del profilo della strada che da verticale si dovrebbe trasformare in una retta inclinata.

Wedel e altri [24, 25] hanno proposto un sistema per riconoscimento di ostacoli in ambiente stradale usando una variante della *warped image* e sfruttando un sensore inerziale per conoscere lo spostamento del veicolo nel tempo. In questo articolo, nelle diverse regioni di interesse in cui l'immagine viene decomposta, viene ricavato il flusso ottico e da questo viene ricavato un fattore di scala rappresentante la trasforma-

zione che ogni regione ha subito tra il fotogramma precedente e quello attuale. Con questo fattore risulta possibile fare un'ipotesi sulla distanza di un eventuale ostacolo, e su questa ipotesi la regione del fotogramma precedente viene trasformata e vengono generate due immagini: una con l'ipotesi di essere terreno e una se è effettivamente un ostacolo alla distanza ipotetica. La regione viene in seguito classificata a seconda della trasformazione più simile al fotogramma attuale. In questo articolo non vengono volutamente considerati i fenomeni istantanei di variazione dell'orientazione della camera in quanto viene stimato che tale contributo al fattore di scala sia tendenzialmente trascurabile.

Recentemente Simond [26] ha proposto un riconoscimento di spazio libero in ambiente urbano, che sfrutta in effetti immagini IPM. Tale sistema non richiede alcun sensore inerziale ma sfrutta parzialmente l'informazione prodotta da una coppia stereo di immagini per sviluppare una sorta di *ego-motion*. La vista IPM viene generata in questo articolo rettificando lo spazio dell'immagine compreso tra due linee della corsia. I fenomeni di beccheggio in questo modo vengono totalmente filtrati, a scapito tuttavia di una conoscenza precisa delle distanze metriche. Gli ostacoli sono individuati studiando l'angolazione dei bordi dell'immagine IPM, se convergenti o meno verso la proiezione del pin-hole.

6.2 Problematiche

In questa sezione vengono discusse le problematiche inerenti al *Motion Stereo* con particolare interesse per l'ambito automobilistico. Verranno presentate possibili tecniche che affrontano il medesimo problema e ne verranno discussi benefici e controindicazioni.

6.2.1 Ego-Motion

Ci si prefigge l'obiettivo di valutare l'utilità delle informazioni inerziali e chiedersi se non è possibile ricavare tale informazione direttamente dalle immagini.

Per rilevazione dell'*ego-motion* si intende una serie di tecniche atte a ricavare il movimento che ha subito la camera all'interno di una sequenza di fotogrammi. Se

la tecnica è ampiamente studiata in letteratura, il problema fondamentale dell'*ego-motion* in ambiente stradale è la bassa quantità di *feature* da inseguire appartenenti al piano stradale, e viceversa l'alta quantità di punti appartenenti a oggetti in movimento.

Stein e altri [27] affrontano il problema della rilevazione del moto in ambito automobilistico, analizzando il flusso ottico in maniera robusta su base probabilistica. Tuttavia nelle conclusioni dell'articolo viene giustificato l'uso di una qualche informazione inerziale del movimento del veicolo, in quanto non c'è sicurezza di trovare sempre *feature* nella zona della strada.

Suzuki e Kanade [28] usano le informazioni inerziali (velocità, yaw-rate, pitch-rate) e le fondono attraverso il filtro di Kalman con le informazioni del flusso ottico proveniente dall'immagine. In questo modo viene fornita una stima robusta del movimento del veicolo e può essere una buona base per un riconoscitore di ostacoli.

Nister [29], senza tuttavia esigenze di *real-time* e in ambiente ristretto, sfruttando il tracciamento delle *feature* di Harris, calcola da queste la posa relativa delle camere tra i fotogrammi. Nella parte dei risultati infine mostra il confronto tra l'*ego-motion* calcolato con una (mono) o due camere (stereo) e l'angolo di *yaw* fornito da un sensore inerziale. Entrambi questi due metodi sono affetti da deriva, ma viene mostrato come dalle immagini si possa fornire in effetti una precisione superiore al sensore.

Esperimenti fatti sull'*ego-motion* sulle immagini IPM, usando l'informazione inerziale come partenza per la ricerca, hanno tuttavia un cattivo riconoscimento del minimo a causa della bassa quantità di texture nell'area inquadrata.

6.2.2 Warped Image

L'utilizzo della *warped image* obbliga ad analizzare obbligatoriamente due fotogrammi, quello corrente e uno passato riproiettato al tempo corrente.

Per affrontare questo problema bisogna effettuare una scelta arbitraria tra tutti i fotogrammi acquisiti. La scelta deve tenere conto che se vengono scelti due fotogrammi troppo vicini temporalmente le differenze tra i due potrebbero essere della stessa livello del rumore, mentre se vengono scelti due fotogrammi troppo distanti temporalmente c'è il rischio che non ci sia invece informazione in comune. Normalmente,

sfruttando l'informazione sulla velocità del veicolo, vengono scelti fotogrammi tra loro a una distanza pressapoco costante nel mondo o alternativamente analizzati più fotogrammi alla volta con un aumento sensibile dei tempi di computazione.

Rimangono tuttavia le considerazioni fatte nel capitolo precedente, ovvero che la semplice differenza su immagini in prospettiva non permette comunque una soglia univoca su tutta la superficie per determinare gli eventuali ostacoli.

6.3 Pitch Detection

Risulta importante, nell'ambito del *Motion Stereo* riuscire a stimare l'orientazione istantanea della camera. Per poter eseguire anche la semplice sottrazione tra le immagini acquisite a tempi differenti, risulta importante conoscere, oltre al movimento del veicolo, eventuali variazioni degli angoli di beccheggio (*pitch*), imbardata (*yaw*) e rollio (*roll*).

Supponendo di poter disporre di un sensore di *yaw-rate* e trascurando l'angolo di *roll* in quanto solitamente poco influente, l'unico angolo che risulta critico è quello di *pitch*. Per poter estrarre l'angolo di *pitch* una tecnica è quella di ricavare dall'immagine il punto di fuga (*Vanishing Point*).

Se nell'equazione prospettica 3.1 si manda a infinito la coordinata X si ottiene la coordinate del punto di fuga (u_{vp}, v_{vp}) :

$$\begin{aligned} (u_{vp}, v_{vp}) &= \lim_{x \rightarrow \infty} F_{pm}(x, y, z) \\ &= \left(u_0 + k_u \frac{r_0}{r_6}, v_0 + k_v \frac{r_3}{r_6} \right) \\ &= \left(u_0 - k_u \frac{\tan \gamma}{\cos \vartheta}, v_0 - k_v \tan \vartheta \right) \end{aligned} \quad (6.1)$$

Pertanto conosciuti i parametri intrinseci della camera, ovvero il punto principale (u_0, v_0) e l'apertura, è possibile, individuato il punto di fuga, estrarre l'angolo di *pitch* ϑ :

$$\tan \vartheta = -\frac{v_{vp} - v_0}{k_v} \quad (6.2)$$

È da notare che la coordinata verticale v_{vp} del *vanishing point* è funzione dei parametri intrinseci (apertura, punto principale) e della sola imbardata.

C'è da sottolineare che la coordinata v_{vp} risulta essere costante su tutta la linea dell'orizzonte nel solo caso l'angolo di rollio e imbardata nulli.

Infatti l'equazione della linea dell'orizzonte in generale vale

$$k_v(r_3r_7 - r_4r_6)u + k_u(r_6r_1 - r_7r_0)v + k_u k_v(r_4r_0 - r_3r_1) = 0 \quad (6.3)$$

dove per semplicità di scrittura si è posto ($u_0 = v_0 = 0$). Nel caso di angoli di yaw e roll vicini a zero (si veda per esempio 7.1) tale retta è effettivamente una retta orizzontale.

Non è una assunzione debole supporre ciò, in quanto tale angoli sono supposti praticamente costanti ed in fase di installazione è possibile calibrarli, o comunque rettificare in seguito l'immagine, in modo che siano nulli.

Il problema di estrarre il *pitch* può pertanto diventare un problema di estrazione del punto di fuga.

Rasmussen e altri [30, 31] trovano il *vanishing point*, in ambiente *offroad*, analizzando l'orientazione locale della texture del terreno. Sfruttando l'elevata tessitura in ambiente fuoristrada, ogni tessera del terreno permette di fornire una retta passante per il punto di fuga. L'intersezione di tali rette fornisce una stima della posizione di tale punto.

In ambiente urbano invece Simond e Rives [32] trovano il Vanishing Point come intersezione dei bordi risultanti da un filtro di Canny.

Un confronto di prestazioni tra il *pitch* ricavato dal *vanishing point* e quello di un sensore inerziale può essere trovata in [33]. In questo caso il *vanishing point* viene trovato come intersezione delle linee bianche poste ai lati dei bordi della carreggiata.

Tale approccio, ovvero lo sfruttare principalmente le linee della carreggiata, è utilizzato anche in questo lavoro.

Il *Vanishing Point* è di fatto definito come l'intersezione della proiezione prospettica di linee che sono parallele nel mondo, e, dopo la proiezione, non sono parallele nel piano immagine. Le linee della carreggiata, nella maggior parte delle condizioni stradali, soddisfano appieno questa definizione. Tuttavia capita spesso che le linee

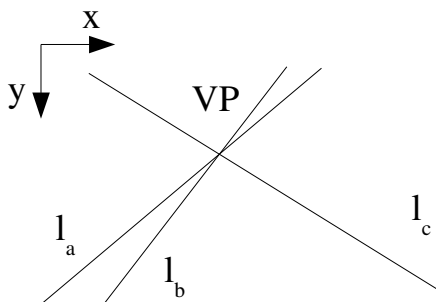


Figura 6.4: Individuazione del punto di fuga partendo da 3 rette non affette da rumore

della strada siano mancanti o poco marcate e torna utili poter sfruttare anche altre strutture costruite dall'uomo che possano fornire indicazione sulla posizione del punto di fuga.

In questa sezione andremo a parlare di un *pitch detector* che sfrutta i diversi artefatti umani (linee della carreggiata, bordi dei marciapiedi, guard rail) per fornire una stima affidabile del punto di fuga della scena.

6.3.1 Analisi preliminare

Per riconoscere sia linee della strada che altre strutture è necessario agire localmente sui bordi dell'immagine.

Per far ciò, viene applicato all'immagine un filtro convolutivo capace di estrarre il gradiente dell'immagine, nelle due componenti spaziali (orizzontale e verticale), come per esempio il filtro di Sobel.

Sia pertanto \mathbf{A} una generica immagine e

$$\mathbf{G}_x = \begin{bmatrix} -1 & 0 & +1 \\ -2 & 0 & +2 \\ -1 & 0 & +1 \end{bmatrix} * \mathbf{A} \quad \text{e} \quad \mathbf{G}_y = \begin{bmatrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ +1 & +2 & +1 \end{bmatrix} * \mathbf{A} \quad (6.4)$$

le immagini generate dall'operatore convoluzione con un filtro 3×3 . Viene preso come sistema di riferimento quello delle x crescenti verso destra e le y crescenti verso il basso. In questo modo è possibile associare a ogni punto dell'immagine \mathbf{A}

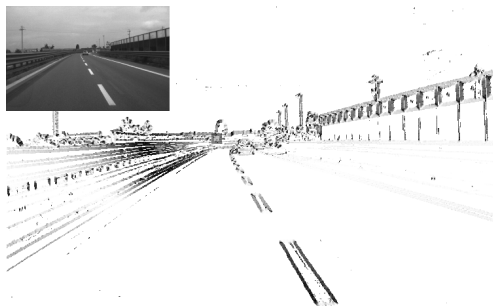


Figura 6.5: Immagine del prodotto scalare tra il gradiente e la direttrice del punto verso il punto di fuga originale. In scuro i punti interessanti dove il gradiente è perpendicolare alla direttrice.

un vettore gradiente $\vec{g}(x, y) = (G_x(x, y), G_y(x, y))$, la cui direzione è la direzione di massima variazione dell'intensità luminosa.

L'idea di base è quella di supporre che il beccheggio istantaneo non faccia cambiare molto la posizione del *vanishing point*, sia rispetto alla posizione di riposo che quella del fotogramma precedente. Per ogni punto dell'immagine è possibile capire se rappresenta un bordo che punta in direzione del *vanishing point* previsto o meno, ad esempio valutando se il prodotto scalare tra il vettore gradiente e il vettore che congiunge il punto di fuga sia inferiore a una soglia data. Un risultato di questo genere di elaborazione è mostrato in figura 6.5.

Come si intuisce, l'immagine risulta abbastanza rumorosa ed obbliga comunque diverse elaborazioni per ogni pixel.

Per sfruttare l'informazione sui punti vicini e ridurre il peso computazione si è scelto un approccio leggermente diverso, mostrato nella sezione successiva.

Recentemente sono usciti articoli a prova del fatto che l'approccio adottato è degno di interesse. Per esempio Nieto e altri [34] nel 2007 hanno prodotto immagini IPM stabilizzate per veicoli usando l'informazione del *vanishing point*. In questo articolo il punto di fuga è stato determinato con l'intersezione delle linee della carreggiata ricavate usando Hough e in seguito filtrato passa basso temporalmente.

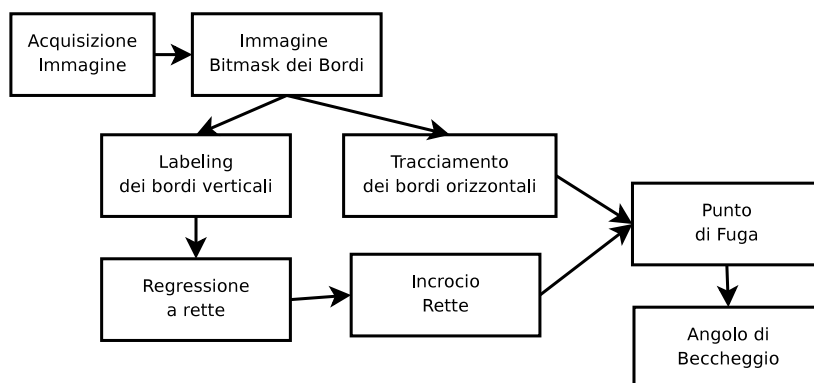


Figura 6.6: Diagramma di flusso dell' algoritmo di Pitch Detection

6.3.2 Implementazione

Il diagramma di flusso dell' algoritmo è mostrato in figura 6.6.

Per poter considerare localmente la morfologia dell' immagine bisogna fare una prima considerazione. Se una retta nel mondo viene proiettata in una immagine in prospettiva, dove il contributo della distorsione dell' ottica è trascurabile, rimane ancora una retta. Di conseguenza il bordo continuo della retta nell' immagine avrà la stessa fase per tutta la sua lunghezza.

Per risolvere il problema della rumorosità, tenere basso il peso computazionale e allo stesso tempo sfruttare l' informazione sulla località dei punti, la fase di sobel è stata quantizzata lungo 8 direzioni. Per ottenere questo risultato ogni pixel dell' immagine dei bordi è stato generato nel modo seguente

$$\begin{array}{l}
 1 \quad g_x > \xi \\
 2 \quad g_x < -\xi \\
 4 \quad g_y > \xi \\
 8 \quad g_y < -\xi
 \end{array} \tag{6.5}$$

Un esempio di tale elaborazione, in falsi colori è mostrata in figura 6.7.

Attraverso l' utilizzo di questa immagine i punti che rappresentano bordi forti vengono coinvolti in una fase di *labeling*. È possibile definire un area rettangolare,

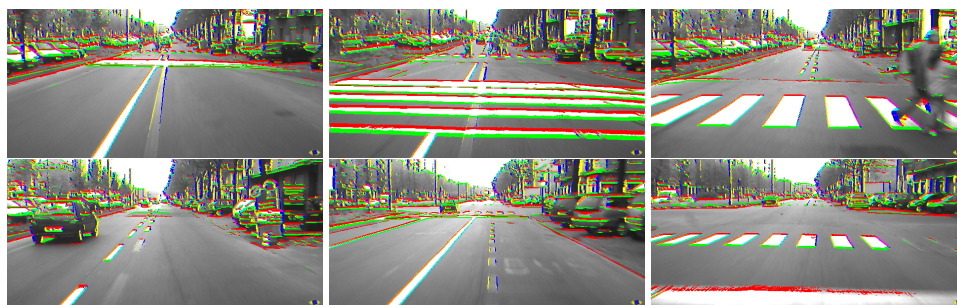


Figura 6.7: Esempio del risultato della binarizzazione degli edge

sottoparte dell'immagine originale situata sotto la linea dell'orizzonte fino al bordo inferiore dell'immagine, che rappresenta la parte dell'immagine che contiene la maggior parte dei punti adatti allo studio del punto di fuga. Dentro quest'area, partendo dal bordo inferiore vengono presi i punti attivi nell'immagine quantizzata e seguiti dal basso verso l'alto passando da una riga dell'immagine a quella superiore prendendo punti contigui dallo stesso segno del gradiente verticale. Sfruttando il segno i gradienti verso l'alto e verso il basso non vengono etichettati insieme.

I *label* così ottenuti vengono trasformati in rette usando una regressione lineare ai minimi quadrati, minimizzando gli scarti orizzontali (non vengono infatti prese in considerazione rette totalmente orizzontali).

Per poter introdurre un indice sull'affidabilità, ogni retta viene memorizzata in una lista con associato il numero di pixel dell'immagine che hanno portato alla sua creazione. Oltre all'informazione sulla dimensione del *label* viene anche memorizzato il segno del gradiente associato alla retta.

A questo punto le rette con lo stesso segno vengono confrontate a due a due, e quelle vicine vengono unite e viene generata una nuova linea attraverso un processo di media pesata usando i coefficienti di affidabilità delle due rette generatrici. La nuova retta avrà un indice di confidenza pari alla somma degli indici delle rette generatrici.

Le rette così ottenute vengono ordinate per indice di affidabilità e vengono tenute le n migliori che comunque abbiano una confidenza minima. Attraverso i risultati

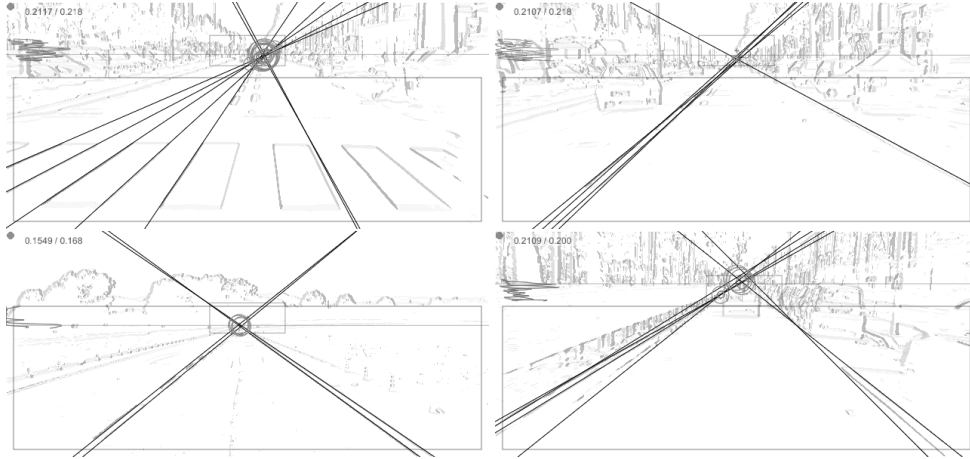


Figura 6.8: Esempio del risultato del *pitch detector*

sperimentali è stato scelto $n = 7$ e una confidenza di almeno 40 pixel.

Tra le rette risultanti vengono calcolati tutti gli incroci, e se il punto di incrocio è in un'area nell'intorno del punto di fuga a riposo tale punto viene memorizzato e gli viene associato un peso pari al minimo valore di confidenza delle rette generatrici. Viene associato a questo punto la somma delle confidenze dei singoli punti che lo hanno generato. Per tenere conto che un punto generato dall'incrocio di due rette che formano un piccolo angolo è più soggetto a rumore, rispetto a un punto generato da rette più lontane, viene aggiunto un peso funzione dell'angolo tra le due rette.

Un risultato del sistema è mostrato in figura 6.8. In questa figura sono visibili, in nero, le linee rilevate e le aree di interesse coinvolte nell'elaborazione.

Tra tutti gli incroci generati viene estratto un punto come media pesata dei punti dell'elenco denominato punto di fuga istantaneo \vec{v}_i e, per migliorare la stabilità, tale punto viene filtrato passa basso temporalmente:

$$\vec{v}_{t,n+1} = \vec{v}_i \alpha + \vec{v}_{t,n} (1 - \alpha) \quad (6.6)$$

È stato scelto come compromesso tra reattività e stabilità $\alpha = 0.25$.

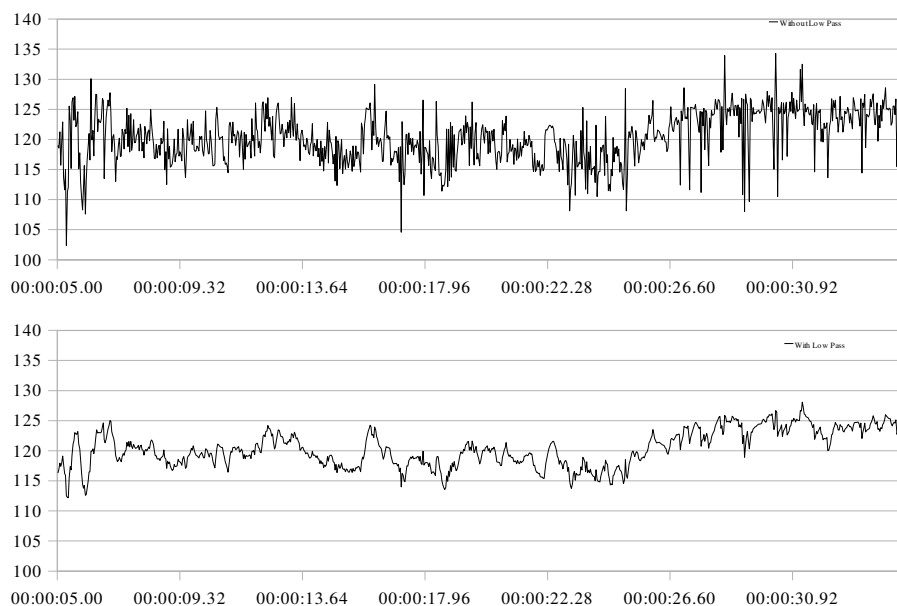


Figura 6.9: Esempio di rilevazione della coordinata v_{vp} del vanishing point sulla sequenza 33

Dal punto di fuga \vec{v}_t viene infine estratto l'angolo di beccheggio attraverso l'equazione 6.2.

Nelle figure 6.10 e 6.9 sono mostrati l'andamento dell'angolo di beccheggio misurati dal sistema con e senza il filtro passa basso temporale. In figura 6.11 viene mostrato il confronto tra i risultati ottenuti con il metodo basato sui bordi e quello che si sarebbe ottenuto con il solo tracciamento dell'orizzonte. Si vede in questo caso come comunque il sistema basato sul tracking dei bordi all'orizzonte sia comunque affidabile (gli andamenti sono di fatto simili) ma presenti tuttavia un problema di deriva non trascurabile.

Può capitare tuttavia che non vengano estratte abbastanza rette dall'immagine e di conseguenza la confidenza del punto di fuga finale sia bassa. In questo caso il sistema passa a una modalità compatibile dove vengono inseguiti i bordi dell'orizzonte. Viene nuovamente sfruttata l'immagine dei bordi generata in precedenza 6.5, ma que-

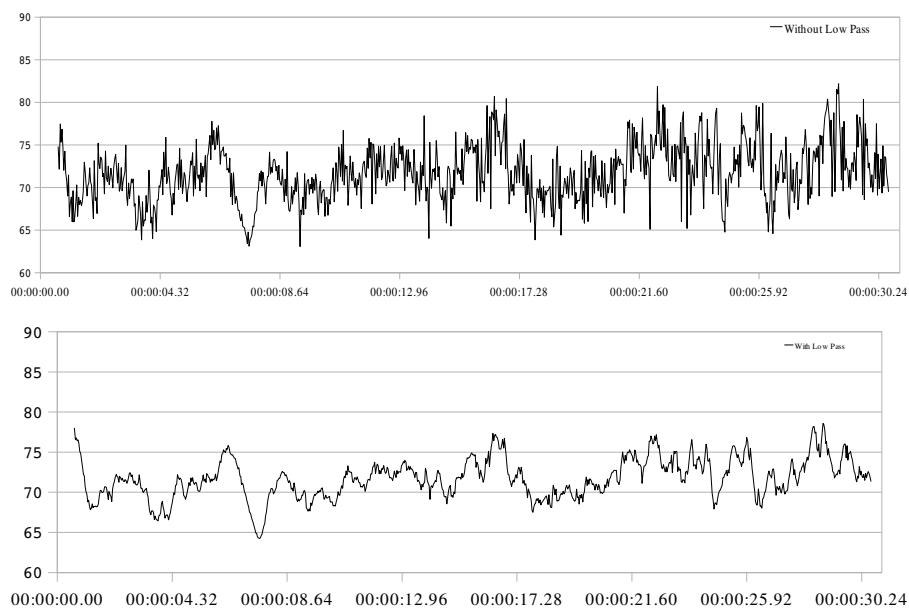


Figura 6.10: Esempio di rilevazione della coordinata v_{vp} del vanishing point sulla sequenza 19

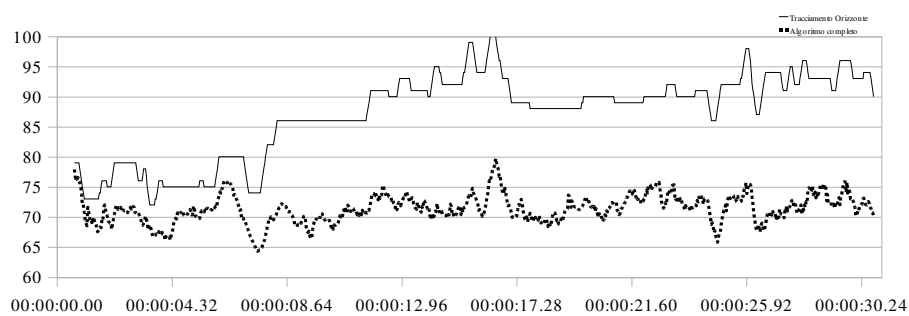
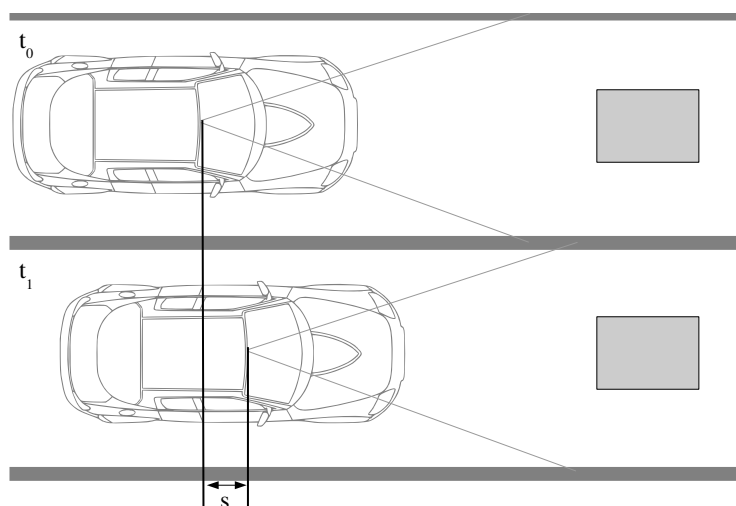


Figura 6.11: Esempio di rilevazione della coordinata v_{vp} del vanishing point sulla sequenza 19 usando il tracking dell'orizzonte

Figura 6.12: Principio base del *Motion Stereo*

sta volta vengono presi in considerazione solamente i bordi orizzontali, perciò i bit generati da gradienti verticali elevati. Non vengono usati i bit dei bordi verticali in quanto sarebbero poco indicativi nell'inseguire oscillazioni verticali dell'immagine.

Per ogni riga dell'immagine nell'intorno del punto di fuga a riposo vengono creati due istogrammi: uno sommando il numero di pixel associati a bordi positivi e uno relativo a bordi negativi. Tali istogrammi vengono confrontati con quelli generati nel fotogramma precedente e viene cercato lo scostamento verticale che minimizza la differenza in valore assoluto tra i due istogrammi.

Tale indice viene sommato all'ultima posizione predetta del *vanishing point* e usato per misurare il beccheggio.

Se la posizione del punto di fuga fornita dall'incrocio delle linee è assoluta, l'informazione prodotta dal tracciamento dei bordi è affetta da deriva e va in effetti utilizzata solo in mancanza di *feature* significative.

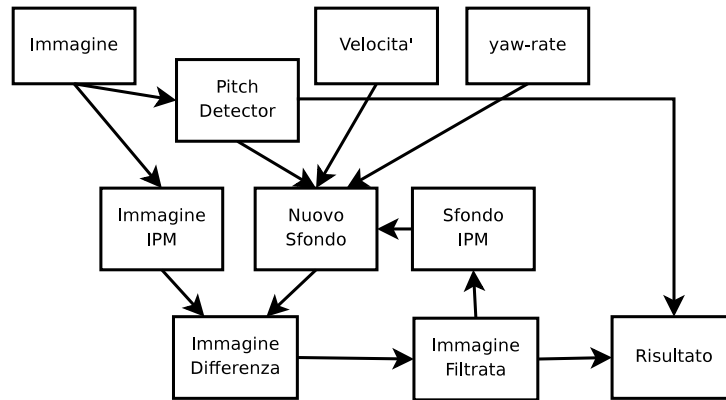


Figura 6.13: Diagramma di Flusso del sistema Motion Stereo

6.4 IPM per Motion Stereo

In questa sezione viene discusso infine l'algoritmo di *Motion Stereo IPM* che permette di rilevare zone occupate da ostacoli sfruttando una sola camera e l'informazione sul movimento. Il diagramma di flusso dell'algoritmo è mostrato in figura 6.13.

Definiamo un'immagine IPM accumulatore A della storia regressa dell'ambiente incontrato. A causa della sua natura di storia, questa immagine sarà disponibile solo dal secondo fotogramma in poi, mentre nel primo potrà essere posta esattamente uguale all'immagine IPM della scena.

L'immagine A rappresentativa della storia passata viene rototraslata in accordo con i dati inerziali. Dall'equazione 4.6 vengono estratti i parametri $(\Delta x, \Delta y)$ dello spostamento che ha subito il mezzo e la variazione dell'angolo $\Gamma = \gamma T$.

Per ottenere una nuova immagine IPM accumulatore densa, i punti dell'immagine IPM vengono trasformati attraverso la trasformazione isometrica inversa:

$$\begin{bmatrix} x_a \\ y_a \end{bmatrix} = \begin{bmatrix} \cos \Gamma & -\sin \Gamma \\ \sin \Gamma & \cos \Gamma \end{bmatrix} \begin{bmatrix} x_b - \Delta x \\ y_b - \Delta y \end{bmatrix} \quad (6.7)$$

dove (x_a, y_a) è il punto nell'immagine accumulatore A nel frame precedente e (x_b, y_b) è il rispettivo punto sulla nuova immagine accumulatore B . Il sistema di riferimento di (x_b, y_b) è fatto in modo che $(0, 0)$ sia la proiezione nell'immagine IPM del *pin-hole*.

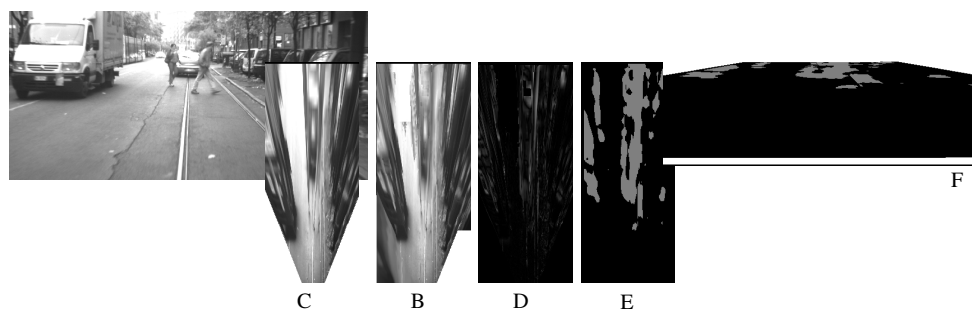


Figura 6.14: Esempio di elaborazione del *Motion Stereo*

L'immagine prospettica C acquisita nel fotogramma corrente viene convertita in immagine IPM sfruttando l'informazione del pitch istantaneo del veicolo (come indicato nella sezione 6.3). Le due immagini B e C inquadrano la stessa porzione del mondo ma da punti di vista differenti: C dal punto di vista attuale, B da un punto di vista passato. Viene generata l'immagine differenza D e tale immagine viene filtrata passa basso con lo stesso filtro descritto in 5.3.1 per ottenere un'immagine binarizzata E .

A questo punto l'immagine binaria E , rappresentate i potenziali ostacoli nella vista IPM, può essere proiettata sull'immagine in prospettiva F e usata per scopi di riconoscimento ostacoli o di individuazione di *free space* (figura 6.14).

Nel frattempo l'immagine binaria E viene usata per aggiornare l'accumulatore. La nuova immagine accumulatore A' viene infatti costruita nel modo seguente

$$a'_i = \begin{cases} \alpha b_i + (1 - \alpha)c_i & \text{se } e_i \text{ è un ostacolo} \\ c_i & \text{se } e_i \text{ non è un ostacolo} \end{cases} \quad (6.8)$$

dove $\alpha = 0.75$ è un parametro modificabile per variare la reattività al cambiamento dell'immagine. In questo modo solo i pixel che il sistema reputa 'spazio libero' finiscono direttamente nella nuova immagine accumulatore, mentre i punti segnalati come ostacoli restano praticamente immutati (il parametro α è elevato). In questo modo l'immagine accumulatore A assume il significato di immagine che modella

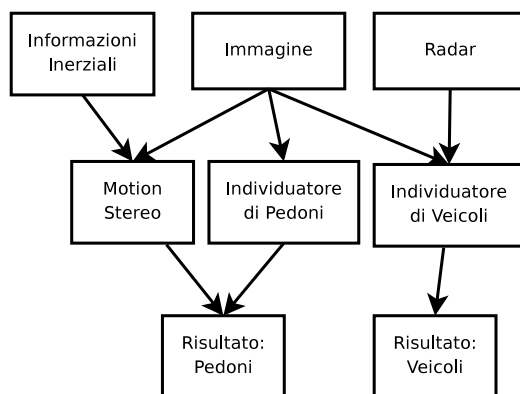


Figura 6.15: Diagramma di flusso del sistema APALACI *Motion Stereo*

il solo piano stradale attraverso dal veicolo nel tempo, nel quale piano gli eventuali ostacoli hanno un peso percentualmente inferiore.

Questa tecnica permette contemporaneamente, quando il veicolo è fermo, di trasformarsi in una tecnica di *background subtraction* ed è perciò in grado di riconoscere anche eventuali oggetti in movimento quando il veicolo è in sosta.

6.5 Fusione dati per riconoscimento pedoni e veicoli

All'interno del progetto APALACI la tecnica di *motion stereo* è stata usata congiuntamente a una tecnica di riconoscimento di pedoni basati sulla simmetria [35, 36]. Tale tecnica se usata solo su immagini monoscopiche producono una quantità elevata di falsi positivi e di fatto viene affiancata da un sistema stereoscopico per validare nelle due immagini della coppia stereo i risultati.

Tale tecnica infatti individua nell'immagine tutte quelle zone che presentano una simmetria tipica dei pedoni (una condizione di elevata simmetria ma non totale) e tende a segnalare anche zone in effetti esenti da ostacoli.

Nel caso del progetto pertanto stata sfruttata l'informazione proveniente dalla tecnica di motion stereo per filtrare questi potenziali pedoni. Per fare questo l'area fornita dal sistema di individuazione pedoni viene sovrapposta con l'area fornita dal



Figura 6.16: Alcuni risultati del sistema APALACI *Motion Stereo*: in verde i pedoni e in arancione i veicoli.

Algoritmo	tempo di elaborazione
Pitch detector	4 ms
Motion Stereo	43 ms
Vehicle detection	1 ms
Pedestrian Detection	83 ms

Tabella 6.1: Prestazioni del sistema di *Motion Stereo IPM*

motion stereo in prospettiva (immagine XX). È giusto notare che viene considerata solo l'area coinvolta nell'analisi, pertanto tutta la parte dell'immagine sopra l'orizzonte è totalmente esclusa. Siccome un pedone non coprirà totalmente la parte dell'immagine coinvolta si è fissata una soglia (il 50% della copertura in pixel) per decidere se la regione fornita è un potenziale pedone o meno.

Insieme alla tecnica di individuazione pedoni basata su simmetrie e *motion stereo* il sistema sfrutta anche l'informazione proveniente da un radar a 24Ghz per individuare la posizione di potenziali veicoli sui quali in seguito la visione va a processare l'immagine attraverso una tecnica ancora basata sulle simmetrie [37, 38].

Il diagramma di flusso del sistema è mostrato in figura 6.15 e i risultati in figura 6.16. Le prestazioni delle diverse parti in cui è composto il sistema sono mostrate in tabella 6.1. È possibile conoscere ulteriori informazioni da [39].

6.6 Conclusioni

In questo capitolo è stata presentata una tecnica di individuazione di ostacoli in ambito automobilistico basata su visione monoculare e movimento. Il sistema è in grado, analizzando viste IPM temporalmente differenti, di indicarne le differenze rappresentate dai potenziali ostacoli.

Per ottenere questo risultato è stato sviluppato un modello matematico del movimento del veicolo e usato per rototraslare le immagini IPM. Per permettere di ottenere una proiezione prospettica inversa stabile è stato necessario implementare anche un algoritmo di *Pitch Detection*, in grado di rilevare il beccheggio istantaneo del veicolo.

L'immagine IPM permette di analizzare la scena a risoluzione costante permettendo di utilizzare un'unica soglia per tutta la superficie. Il fattore di scala di tale immagine è funzione della quantità di imprecisioni che il sistema raccoglie nelle sue diverse fasi (errori di calibrazione, errori odometrici, errori nel *pitch detector*). Gli esperimenti fatti, con una risoluzione di 0.1 m/pixel, mostrano un buon compromesso tra riconoscimento ostacoli e reiezione del rumore. Il sistema presenta ottime rilevazioni in ostacoli verticali contrastati con lo sfondo e ostacoli in movimento.

Questa tecnica può essere considerata un riconoscitore di 'spazio libero' e permette di fornire regioni di interesse per altri algoritmi. Sono stati mostrati i risultati di questa tecnica unita ad altre tecniche di classificazione ostacoli basati su simmetrie.

Capitolo 7

Lateral Motion Stereo

Il triangolo viene chiuso, parlando di una tecnica di ricostruzione tridimensionale che fa uso di una sola camera ma nuovamente ne sfrutta il movimento del veicolo per simulare una visione stereoscopica. Utilizzo di una sola camera, questa stavolta orientata trasversalmente al moto del veicolo, permettere di realizzare uno stereo basato ancora sul moto (figura 7.1).

Esistono diversi studi dove si fa esplicitamente riferimento a sequenze di immagini laterali al movimento [40], e viene fatto riferimento anche al caso in cui vengano considerate anche più di una coppia di immagini alla volta [41]. Normalmente tuttavia sono pure estensioni di algoritmi di stereoscopia al caso del moto e in questi caso la *baseline* è supposta conosciuta o costante e non intervengono altre problematiche.

In condizioni ideali, ovvero un mondo statico e in totale assenza di vibrazioni, la configurazione di un sistema *Lateral Motion Stereo* è completamente indistinguibile da una coppia di camere stereoscopiche normali e ne segue completamente l'implementazione.

7.1 Ricostruzione Tridimensionale e V-Disparity

Esaminiamo le equazioni base della ricostruzione tridimensionale per adattare al caso del *Motion Stereo* Laterale.

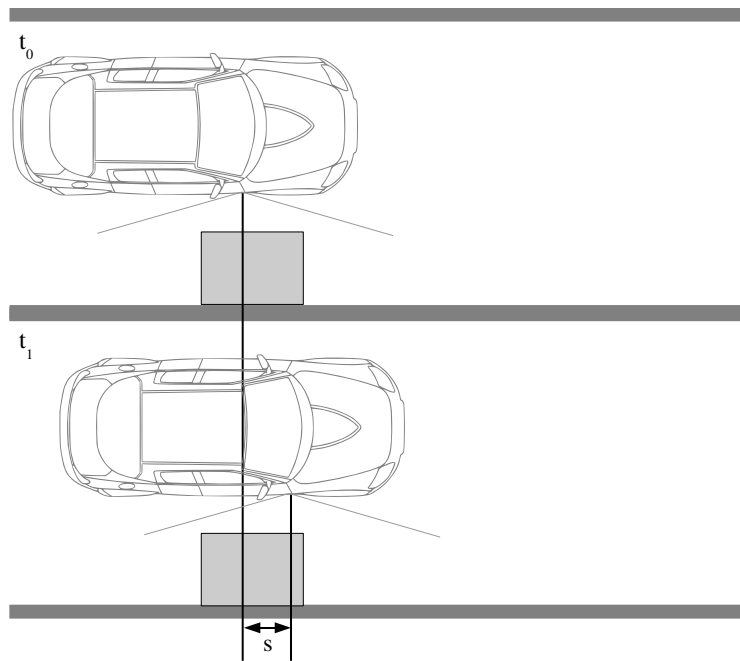


Figura 7.1: Principio base del *Lateral Motion Stereo*

Condizione necessaria per usare un approccio basato sulla *V-Disparity* è considerare due camere perfettamente calibrate con gli stessi parametri intrinseci, gli angoli di yaw e roll nulli e lo stesso angolo di pitch. I *pin-hole* di tali camere devono vivere sullo stesso asse Y nel mondo.

È da notare che le variazioni dei solo parametri intrinseci e rotazioni si possono sempre ottenere rettificando le immagini di partenza se queste sono affette da parametri diversi da quelli imposti (vedi sezione A.2). In queste particolari condizioni la coordinata v del medesimo punto mondo nelle due immagini risulta la stessa. La matrice \mathbf{R} risulta infatti:

$$\mathbf{R} = \begin{bmatrix} 0 & -1 & 0 \\ -\sin \vartheta & 0 & -\cos \vartheta \\ \cos \vartheta & 0 & -\sin \vartheta \end{bmatrix} \quad (7.1)$$

e la matrice di proiezione complessiva si può esprimere come:

$$\mathbf{H} = \begin{bmatrix} u_0 \cos \vartheta & -k_u & -u_0 \sin \vartheta \\ -k_v \sin \vartheta + v_0 \cos \vartheta & 0 & -k_v \cos \vartheta - v_0 \sin \vartheta \\ \cos \vartheta & 0 & -\sin \vartheta \end{bmatrix} \quad (7.2)$$

Sia pertanto (x, y, z) la coordinata di un punto nel mondo, (u_a, v) la coordinata di tale punto sull'immagine della camera A , (u_b, v) la coordinata sull'immagine della camera B , $(x_{0,a}, y_{0,a}, z_{0,a})$ la posizione del *pin-hole* A nel mondo e $(x_{0,b}, y_{0,b}, z_{0,b})$ la posizione del *pin-hole* B .

La coordinata u dell'ostacolo si può scrivere come:

$$\begin{aligned} u_a &= u_{0,a} - \frac{k_{u,a}(y-y_{0,a})}{\cos \vartheta(x-x_{0,a}) - \sin \vartheta(z-z_{0,a})} \\ u_b &= u_{0,b} - \frac{k_{u,b}(y-y_{0,b})}{\cos \vartheta(x-x_{0,b}) - \sin \vartheta(z-z_{0,b})} \end{aligned} \quad (7.3)$$

Imposte le condizioni $x_{0,a} = x_{0,b} = x_0$ e $z_{0,a} = z_{0,b} = z_0$, $y_{0,b} - y_{0,a} = b$ (avendo definito b *baseline* la distanza tra le due camere) e uguali $k_u, k_v, u_0, v_0, \vartheta$ la matrice di proiezione tra le due immagini risulta la medesima e, valutando l'equazione 7.3, l'unica differenza tra u_a e u_b risulta essere il numeratore. Definendo la differenza $d = u_a - u_b$ come *disparità*, si può scrivere

$$d = u_a - u_b = \frac{k_u b}{\cos \vartheta(x - x_0) - \sin \vartheta(z - z_0)} \quad (7.4)$$

e usando la relazione 7.3 si ottiene il risultato notevole

$$u_a = u_0 - d \frac{y - y_a}{b} \quad (7.5)$$

da cui si ricava che

$$y = b \frac{u_a - u_0}{d} + y_a \quad (7.6)$$

Nel caso in cui le camere siano posizionate in modo corretto, l'unico parametro di calibrazione che incide sulla coordinata y risulta essere pertanto b .

$$v - v_0 = -\frac{k_v}{bk_u} (\sin \vartheta (x - x_0) + \cos \vartheta (z - z_0)) d \quad (7.7)$$

Da cui il sistema di equazioni:

$$\begin{aligned} \cos \vartheta (x - x_0) - \sin \vartheta (z - z_0) &= \frac{bk_u}{d} \\ \sin \vartheta (x - x_0) + \cos \vartheta (z - z_0) &= -\frac{v - v_0}{k_v} \frac{bk_u}{d} \end{aligned} \quad (7.8)$$

Calcoliamo quanto vale la relazione della disparità dalla coordinata v ricavando il valore di x dalla seconda e sostituendolo nella prima delle 7.8:

$$\begin{aligned} x - x_0 &= \tan \vartheta (z - z_0) + \frac{k_u}{d \cos \vartheta} b \\ v - v_0 &= -k_v \tan \vartheta - d \frac{k_v}{k_u} \frac{z - z_0}{b \cos \vartheta} \end{aligned} \quad (7.9)$$

Dalla prima delle equazioni 7.9, si vede che l'espressione della disparità dipende solamente dalla distanza x se l'altezza z è fissata (suolo), e dalla seconda la disparità d cresce linearmente con la coordinata v seguendo un coefficiente angolare noto

$$d = \cos \vartheta \frac{b}{z_0} (v - v_{d=0}) \quad (7.10)$$

nel caso classico in cui $k_u \approx k_v$ (pixel quadrato). Il punto di disparità nulla $v_{d=0}$, sopra menzionato, si trova in

$$v_{d=0} = v_0 - k_v \tan \vartheta \quad (7.11)$$

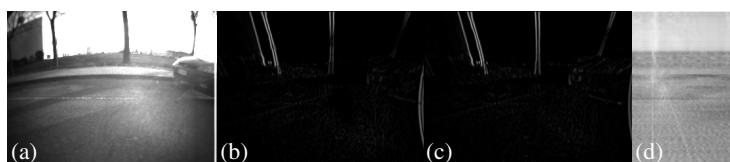


Figura 7.2: V-Disparity nel *Lateral Motion Stereo*: immagine acquisita (a), immagine preprocessata (b), immagine preprocessata precedente (c), V-Disparity (d)

e dipende solo dall'apertura verticale e dal *pitch*. Come ci si può aspettare, tale coordinata è la stessa coordinata del *vanishing point* (si confronti con l'equazione 6.2).

7.1.1 Estensione al *Motion Stereo*

Nel caso del 'motion stereo laterale' i parametri intrinseci sono uguali per costruzione, essendo la stessa camera che acquisisce entrambi i fotogrammi. Ipotizzando che la camera si muova longitudinalmente in maniera ideale l'orientazione della camera rimane costante nel tempo, e in ogni caso attraverso la calibrazione è possibile porre, almeno nel caso statico, gli angoli di *yaw* e *roll* a zero.

In queste condizioni le equazioni della disparità risultano soddisfatte e permettono una ricostruzione tridimensionale dell'ambiente circostante quando la *baseline* b risulta perfettamente conosciuta.

È da notare che tale valore può essere ricavato attraverso l'analisi dell'immagine di V-Disparity (esempio in figura 7.2). Da questa immagine è possibile estrarre la retta rappresentante l'andamento della disparità sul terreno e usando l'equazione 7.10 ricavare la baseline.

Questo approccio risulta tuttavia non sempre possibile in quanto il piano stradale spesso non è inquadrato. In tal caso l'informazione dell'odometria del veicolo risulta indispensabile.

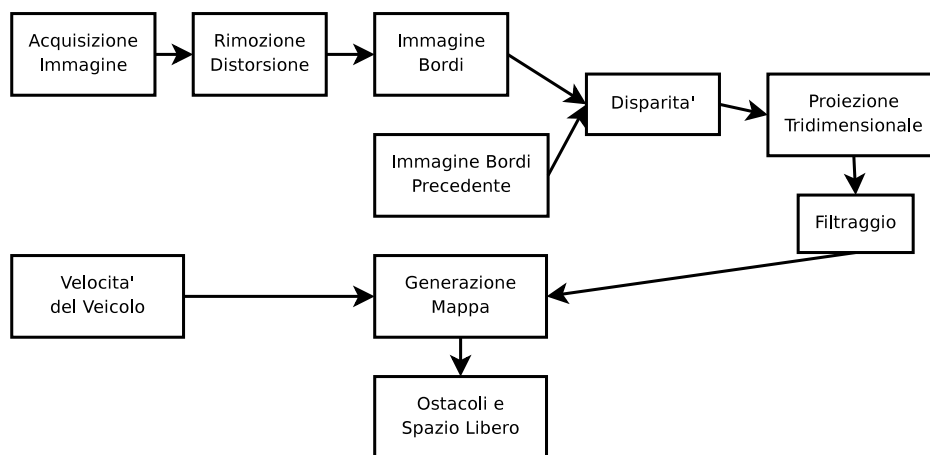


Figura 7.3: Diagramma di Flusso del sistema *Lateral Motion Stereo*

7.2 Calibrazione

La camera viene installata lateralmente sul veicolo e orientata in modo da osservare una ampia scena di lato del veicolo. L'installazione migliore per la camera è quella in cui il sensore è ruotato di 90 gradi rispetto all'orientazione normale, in modo da massimizzare l'apertura verticale, interessante per l'analisi del terreno, e l'informazione proveniente da una elevata apertura orizzontale è comunque affetta dalle occlusioni dei veicoli.

Sfruttando una tecnica delle tecniche discusse in sezione 2.2 viene calcolata una *Look Up Table* per rimuovere la distorsione.

Rimossa la distorsione risulta possibile estrarre i parametri intrinseci usando un metodo proposto da Zhang [4] e infine estrarre i parametri estrinseci attraverso una calibrazione Euleriana (vedi sezione 2).

7.3 Implementazione

Il diagramma di flusso del sistema è visibile in figura 7.3.

A causa della particolare lunghezza focale richiesta, l'immagine acquisita (figu-



Figura 7.4: Fotogrammi di esempio nel caso *Lateral Motion Stereo*



Figura 7.5: Rimozione della distorsione: a sinistra l'immagine acquisita e a destra l'immagine dedistorta

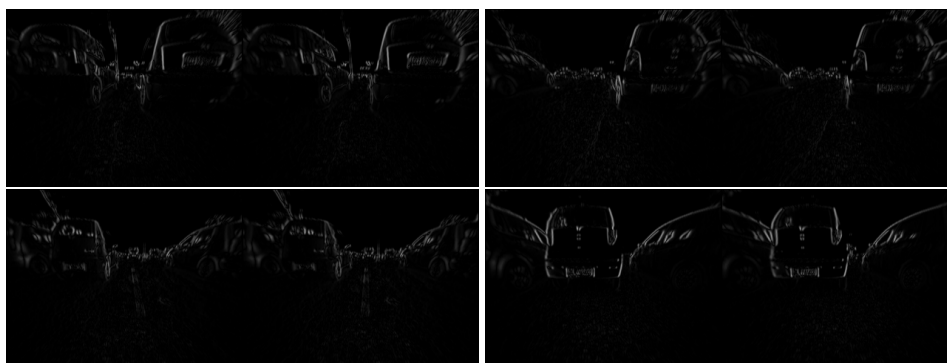


Figura 7.6: Immagine dedistorta filtrata con Sobel: l'immagine dal fotogramma corrente (sinistra) e l'immagine dal fotogramma precedente (destra)

ra 7.4) è affetta da distorsione abbastanza accentuata. La distorsione viene rimossa grazie a una *Look Up Table* generata *offline* usata anche per calibrare l'immagine stessa. I risultati si possono vedere in figura 7.5.

L'immagine dedistorta viene trasformata attraverso un filtro derivativo verticale per accentuare i bordi (figure 7.6). L'utilizzo di un filtro derivativo permette di accentuare le zone di discontinuità e migliorare l'individuazione della disparità nel passo successivo dell'algoritmo.

Come ingresso per la ricostruzione tridimensionale sono disponibili due immagini di bordi: una ottenuta dal fotogramma corrente, e una proveniente da un fotogramma precedente. Il fotogramma precedente viene scelto in maniera tale che tra le due immagini temporali il veicolo si sia mosso di uno spazio minimo (negli esperimenti è stata usata una *baseline* minima di 0.1 m).

Per generare l'immagine di disparità infine si è utilizzato un algoritmo già disponibile (per dettagli vedere Broggi e altri [42]). Confrontando con una finestra mobile le due immagini della coppia stereo virtuale viene generata l'immagine di disparità (figura 7.7). Il valore di ogni elemento di quest'immagine rappresenta il numero di pixel (orizzontali) che permettono la trasformazione del corrispondente punto sull'immagine sinistra, nell'immagine destra. La mappa di disparità generata con questa



Figura 7.7: Disparità nel *Lateral Motion Stereo*: in scuro i punti a bassa e in chiaro i punti ad alta disparità

tecnica è non densa: esistono dei punti per i quali non è stata trovata abbastanza confidenza tra le finestre nelle due immagini. Questo accade nelle zone a basso contenuto informativo e in presenza di occlusioni.

Dall'immagine di disparità infine vengono estratte le coordinate mondo sfruttando una triangolazione tra le rette epipolari. Queste coordinate, a differenza di quelle generate attraverso la tecnica IPM, sono le effettive coordinate del punto del mondo.

Invertendo l'equazione A.12 si ottiene una retta che passa per il *pin-hole*:

$$\begin{aligned}\vec{p}_l &= \vec{t}_l + \lambda (\mathbf{A}_l \mathbf{R}_l)^{-1} (u_l, v_l, 1)^\top = \vec{t}_l + \lambda \vec{v}_l \\ \vec{p}_r &= \vec{t}_r + \lambda (\mathbf{A}_r \mathbf{R}_r)^{-1} (u_r, v_r, 1)^\top = \vec{t}_r + \lambda \vec{v}_r\end{aligned}\quad (7.12)$$

dove \vec{p}_l, \vec{p}_r sono le due rette epipolari e \vec{t}_l, \vec{t}_r le coordinate del due *pin-hole*.

Tali due rette possono incrociarsi, essere parallele o, nel caso generale di camere non perfettamente calibrate, due rette sghembe. Per ricavare un punto di incontro unico per tutte queste casistiche viene usata la distanza tra due rette e fornita come posizione mondo il valor medio del segmento che unisce i punti più vicini delle due rette.

A ogni punto con disparità valida è possibile pertanto associare un punto in coordinate mondo (X, Y, Z) . Vengono filtrati i punti con Z inferiore a una certa soglia per

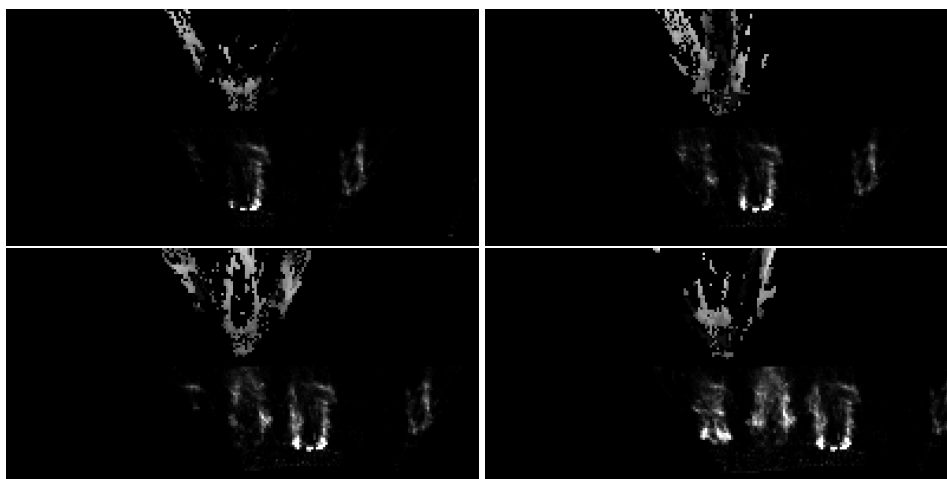


Figura 7.8: Esempio di proiezione nello spazio (X,Y) : mappa delle altezze (alto), mappa del numero di punti attivati (basso)

ridurre il rumore e per scartare comunque valori non interessanti. I punti rimanenti vengono proiettati in uno spazio (X,Y) quantizzato (gli esperimenti mostrati sono stati generati con valori di quantizzazione di 0.1 m). Per ogni elemento di questo spazio viene memorizzata l'altezza massima Z rilevata e il numero di punti che cadono al suo interno maggiori della soglia.

Di fatto solo una piccola fetta dell'immagine fornisce informazione utile per la navigazione, considerando anche la stretta fetta del mondo inquadrata. Contemporaneamente perciò un'immagine (accumulatore) della storia pregressa del veicolo viene rototraslata in accordo con il movimento del veicolo stesso e ai suoi punti vengono aggiunti quelli attualmente prodotti (figura 7.8).

A questo punto è possibile disporre di una mappa delle zone verticali incontrate dal veicolo al suo passaggio utilizzabile in svariate applicazioni.



Figura 7.9: Esempio di individuazione ‘spazio libro’ basato su *Lateral Motion Stereo*

7.4 Conclusioni

In questo capitolo è stato mostrato una ulteriore modalità di individuazioni di ostacoli basata su movimento e una stereoscopia (in questo caso virtuale).

Una applicazione pratica per il sistema *Lateral Motion Stereo* potrebbe essere quella di individuare ostacoli e spazi liberi in un area di parcheggio per segnalare, ad esempio, al conducente eventuali posti liberi. In figura 7.9 sono presentati alcuni risultati per questo genere di applicazione. Nella parte in basso di ogni figura è mostrata la mappa della distribuzione degli ostacoli nella quale vengono cercate aree rettangolari senza ostacoli. Nella parte alta queste eventuali aree trovate sono proiettate nell'immagine.

Tale sistema è basato su disparità tra due immagini temporalmente differenti simulando un movimento lungo l'asse X del veicolo, asse Y delle camere. Nel caso il veicolo non si stia muovendo esattamente in linea retta, ma stia eseguendo una rotazione, nei casi in cui questa non sia intensa è possibile approssimarla ancora con uno spostamento lineare e limitarsi solamente a rototraslare la mappa degli ostacoli.

7.4.1 Sviluppo futuri

Attualmente il sistema ha dimostrato la capacità di fornire una semplice rappresentazione dell'ambiente circostante in condizioni statiche. Risulta interessante studiare il sistema nel caso ci siano ostacoli in movimento. In tal caso gli oggetti in movimento andrebbero studiati per capirne la dinamica e se non pericolosi, filtrati via.

È stato presentato un metodo semplice per la fusione delle informazioni pregresse. Diverse tecniche possono venire a questo punto analizzate per fondere le informazioni prodotte dai fotogrammi precedenti con quelli attuali.

Nel caso le informazioni odometriche non fossero disponibili, o fossero rumorose, è possibile capire il movimento che ha subito il veicolo analizzando l'immagine della V -Disparity, come mostrato nell'equazione 7.10. Alternativamente si può usare comunque questa informazione per correggere le oscillazioni che il veicolo può aver istantaneamente subito.

Capitolo 8

Discussione generale

Questo lavoro di tesi discute del problema dello sviluppo di una classe di sensori basati su visione capaci di percepire un ambiente altamente dinamico e complesso come quello stradale. Nel testo sono stati discussi diversi approcci che hanno dimostrato di essere capaci di generare mappe ‘approssimative’ del mondo che circonda il veicolo.

Sono mappe ‘approssimative’ in quanto non necessariamente inaccurate, ma sicuramente non dense, senza una informazione sulla tridimensionalità capaci però di avvertire il sistema della presenza di un eventuale pericolo al quale deve prestare in seguito attenzione, per classificarlo in maniera più precisa e verificare se le ipotesi di presenza di ostacoli fossero effettivamente corrette.

Queste mappe, eventualmente raffinate, sono state usate per allertare il guidatore del veicolo o fornite a sistemi intelligenti di guida automatica per pianificare il moto del veicolo stesso.

8.1 Contributo di questa tesi

I tre approcci presentati mostrano una combinazione tra le tecniche di *Motion Stereo*, *Ground Plane Stereo* e *Inverse Perspective Mapping* capaci di elaborare velocemente l’informazione dell’immagine.

Indubbiamente il sistema *Ground Plane Stereo IPM* è quello che richiede i più

bassi tempi computazionali e fornisce una rappresentazione accurata dell'immagine, richiedendo contemporaneamente limitate richieste di calibrazione e una elevata libertà di installazione. La variante basata su disparità anche se richiede un'analisi ulteriore permette di fornire una ricostruzione tridimensionale della scena partendo sempre da immagini IPM. Tale sistema è stato utilizzato per riconoscere ostacoli nell'angolo cieco di veicoli commerciali, allertando con segnali visivi, acustici e agendo sulle cinture di sicurezza il guidatore di un eventuale pericolo. Allo stesso modo è stato usato in fusione con altri sistemi per individuare ostacoli vicini e consentire la navigazione di un veicolo autonomo.

Il sistema *Motion Stereo IPM* richiede invece una sola camera per funzionare ed è in grado comunque di individuare qualitativamente sia regioni dell'immagine in movimento che ostacoli verticali. Tale tecnica può essere usata per fornire lo spazio libero davanti a un veicolo che può usare tale informazione per pianificare il moto. L'algoritmo di *Motion Stereo IPM*, unito a un riconoscitore di pedoni e di veicoli basati anch'essi su visione, costituisce un sistema capace di individuare e classificare ostacoli pericolosi nell'intorno del veicolo.

Tale sistema ha richiesto lo sviluppo di un algoritmo di *Pitch Detector* capace di rilevare il beccheggio istantaneo del veicolo. Attualmente una versione di questa tecnica di *Pitch Detector* viene anche usata per generare le immagini IPM in un algoritmo di *Lane Departure Warning*, capace di riconoscere le linee della carreggiata e avvisare quando si stanno oltrepassando inavvertitamente.

Allo stesso modo l'immagine IPM così generata, sfruttando il modello di moto del veicolo, può essere usata per produrre una mappa del mondo che il veicolo attraversa (si veda l'immagine 4.2).

Come ultimo sistema viene mostrato come è possibile generare comunque una ricostruzione tridimensionale affidabile usando una sola camera rivolta però perpendicolarmente al moto del veicolo. Nell'algoritmo di *Lateral Motion Stereo* le immagini acquisite da una camera vengono confrontate temporalmente usando un algoritmo standard che ne calcola la disparità. Ottenute le coordinate mondo fornite dal fotogramma corrente queste vengono fuse con quelle contenute nella storia passata e permettono di fornire una mappa dell'ambiente attraversato dal veicolo.

Appendice A

Il modello Pin-Hole

Questa appendice affronta dal punto di vista analitico il modello *Pin-Hole* e come, secondo questo modello, avviene la formazione dell'immagine mostrando le relazioni che intercorrono tra i punti nel mondo e i punti nell'immagine.

Per parlare del modalità attraverso le quali l'informazione tridimensionale del mondo giunge impressa su un generico sensore bisogna fare una prima considerazione. Non esistono infatti regole che impongono una struttura definita a un qualsiasi dispositivo per l'acquisizione di immagini, anche se per motivi pratici e storici (la camera oscura, che introduce il modello *pin-hole*, si conosce dal rinascimento) è possibile rappresentare la quasi totalità dei dispositivi come formati da un piano su cui si forma l'immagine e da un fuoco.

Il modello che si andrà a introdurre perciò seguirà questa semplice struttura, e il formalismo utilizzato in questa tesi segue quando possibile quello introdotto da Faugeras [43].

A.1 Modello della telecamera

Nella discussione del modello vengono coinvolti 3 sistemi di coordinate diversi:

- mondo $(X, Y, Z)^T$, ovvero le coordinate effettive degli oggetti (figura A.1);

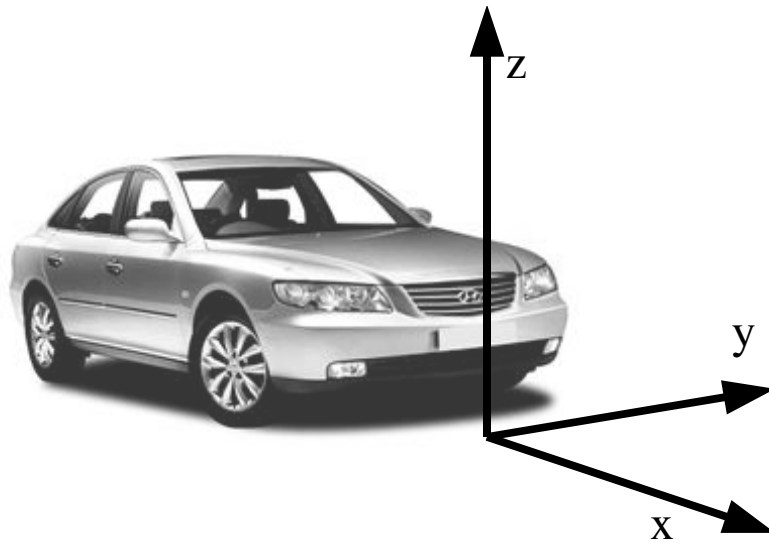


Figura A.1: Sistema di coordinate mondo

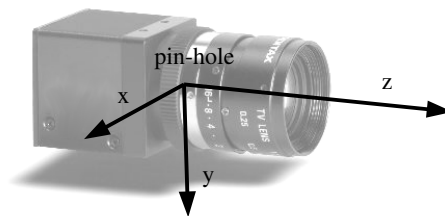


Figura A.2: Sistema di coordinate camera

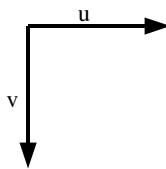


Figura A.3: Sistema di coordinate immagine

- camera $(x, y, z)^T$, ovvero le coordinate dove il pin-hole è il centro del sistema e orientata di conseguenza (figura A.2);
- immagine $(u, v)^T$, ovvero quelle di un piano bidimensionale dove viene acquisita o mostrata l'immagine (figura A.3).

Basandosi sui rapporti tra triangoli simili, triangoli generati dai raggi ottici, si può scrivere l'equazione che permette di proiettare un generico punto in coordinate della camera $(x_i, y_i, z_i)^T$ in coordinate immagine $(u_i, v_i)^T$:

$$\begin{bmatrix} \tilde{u}_i \\ \tilde{v}_i \end{bmatrix} = \frac{f}{z_i} \begin{bmatrix} x_i \\ y_i \end{bmatrix} \quad (\text{A.1})$$

dove f è la distanza focale, distanza tra il *pin-hole* e il piano immagine.

È da precisare che le coordinate $(x_i, y_i, z_i)^T$, espresse in coordinate camera, in questo testo seguono la regola *della mano sinistra* (molto usata in computer grafica), contrapposta alla regola *della mano destra* (più usata in applicazioni robotiche) che spesso viene usata per esprimere le coordinate mondo, come nelle applicazioni qui discusse.

Il punto $(\tilde{u}_i, \tilde{v}_i)^T$ non rappresenta un punto dell'immagine, ma è solo una coordinata intermedia di comodo. È necessaria una ulteriore trasformazione per ottenere le coordinate immagine:

$$\begin{bmatrix} u_i \\ v_i \end{bmatrix} = \begin{bmatrix} D_u \tilde{u}_i \\ D_v \tilde{v}_i \end{bmatrix} + \begin{bmatrix} u_0 \\ v_0 \end{bmatrix} \quad (\text{A.2})$$

In questa ultima equazione i parametri (u_0, v_0) tengono conto dello scostamento dell'origine delle coordinate sugli schermi video. Tale punto viene chiamato Punto Principale (*Principal Point*), ed è quel punto sul piano immagine che è alla base della perpendicolare del centro della lente o, in maniera più corretta, del *Pin-Hole* (Atkinson, 1996). D_u e D_v sono invece fattori di conversione tra le unità di misura del mondo e quelle immagine (come la conversione tra metri e pixel), e tengono conto dei diversi fattori di scala nel caso in cui il pixel del sensore non sia perfettamente quadrato.

In mancanza di poter ottenere dal produttore della camera informazioni accurate su f , D_u e D_v si tende ad accorpare queste tre variabili in due nuove, i cui valori sono ottenibili in maniera empirica.

Infatti si possono definire

$$\begin{aligned} k_u &= D_u f = \frac{u_0}{\tan \alpha_u} \\ k_v &= D_v f = \frac{v_0}{\tan \alpha_v} \end{aligned} \quad (\text{A.3})$$

dove α e β sono la semi ampiezza dell'apertura della camera, orizzontale e verticale rispettivamente, nel caso in cui il punto principale sia esattamente al centro del sensore. Nel caso in cui il punto principale sia spostato α_u rappresenta l'apertura orizzontale tra il lato sinistro e la coordinata u_0 , mentre α_v rappresenta l'apertura verticale dal lato superiore alla coordinata v_0 .

In mancanza di distorsione e considerando che le camere moderne hanno una geometria del pixel praticamente quadrato, k_u e k_v tendono ad assumere valori molto simili la cui differenza è inferiore al 1

Il sistema A.1 non rappresenta un vincolo lineare tra i parametri. Il poter lavorare con sistemi lineari semplifica il formalismo matematico e permette di usare strumenti matematici potenti adatti a tali sistemi. Sotto queste considerazioni è possibile riscrivere l'equazione A.1, aggiungendo un vincolo esplicito, per poter rappresentare in forma di sistema tale equazione:

$$\begin{bmatrix} \lambda & u_i \\ \lambda & v_i \\ \lambda & \end{bmatrix} = \mathbf{A} \begin{bmatrix} x_i \\ y_i \\ z_i \end{bmatrix} \quad (\text{A.4})$$

che infatti risolta fornisce $\lambda = z_i$. Questo formalismo permette di rendere implicita la divisione per la coordinata z . Il sistema infatti risulta ben definito quando la terza coordinata assume valore 1, ovvero dividendo le prime due coordinate per la terza. Nella sezione A.2.1 verranno mostrate le proprietà di questo genere di coordinate, dette omogenee.

È possibile ora introdurre una matrice \mathbf{A} definita come:

$$\mathbf{A} = \begin{bmatrix} \frac{u_0}{\tan \alpha_u} & k_\gamma & u_0 \\ 0 & \frac{v_0}{\tan \alpha_v} & v_0 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} k_u & k_\gamma & u_0 \\ 0 & k_v & v_0 \\ 0 & 0 & 1 \end{bmatrix} \quad (\text{A.5})$$

Tale matrice, non dipendendo da fattori che non siano altri che quelli della camera stessa, è detta matrice dei fattori intrinseci. k_γ , lo *skew factor*, viene normalmente posto a zero $k_\gamma = 0$ per camere normali e rimane solo per motivi storici.

L'inversa della matrice A.5 esiste sempre e vale:

$$\mathbf{A}^{-1} = \begin{bmatrix} \frac{1}{k_u} & 0 & -\frac{u_0}{k_u} \\ 0 & \frac{1}{k_v} & -\frac{v_0}{k_v} \\ 0 & 0 & 1 \end{bmatrix} \quad (\text{A.6})$$

A.1.1 Modello della distorsione

Nelle espressioni appena introdotte non viene tenuto conto dei contributi dovuti alla distorsione della lente. In generale questi contributi si dividono in radiali (diretti lungo la direttrice che unisce il punto in esame al centro ottico di distorsione) o tangenziali (che sono perpendicolari alla direttrice). In generale i contributi tangenziali (e altri contributi qui non citati) sono trascurabili. La distorsione radiale invece è sempre presente, e man mano che vengono usate ottiche con focale corta in generale (non è ovviamente vero sempre) il peso di questa distorsione aumenta. Il discorso non è generale in quanto la distorsione è dovuta a non idealità nella fase di produzione della lente e ottenere una lente con focale corta che non distorca è possibile ma estremamente costoso.

La quasi totalità delle telecamere commerciale e in generale qualsiasi lente devia dal modello della pin-home camera, in particolare nella condizione di ampi angoli di visuale. Tali distorsioni generano una distorsione non lineare difficilmente modellizzabile. Calibrare e correggere la distorsione geometrica è un prerequisito per ricostruire

in 3D la scena del mondo. Questo passo risulta necessario in quanto i produttori di lenti e camere non danno normalmente informazioni geometriche accurate.

Siccome ogni ottica è formata da un certo numero di lenti diverse, e ogni lente segue un procedimento diverso di produzione, l'unica cosa che si può dire sull'equazione della distorsione radiale è che questa è non lineare. Per livelli di distorsione non eccessivi (lontani per esempio da un ottica *Fish-Eye* che ha un modello particolare in quanto degenera) l'equazione viene spesso approssimata attraverso i primi termini del suo sviluppo in serie di Taylor.

Il punto ideale (inteso come se fosse preso con una lente ideale non distorcente) (x_u, y_u) è in relazione con il punto immagine distorto (inteso come punto effettivamente acquisito con la lente reale e non ideale) (x_d, y_d) attraverso una funzione del solo raggio r_d , distanza tra il punto e il centro di distorsione (che generalmente è diverso dal punto principale e di conseguenza dal centro geometrico dell'immagine).

La generica funzione della distorsione radiale può essere definita come rapporto tra la distanza ideale non distorta r_u e la distanza reale distorta r_d , ovvero

$$\frac{r_u}{r_d} = f_d(r_d) \quad (\text{A.7})$$

ed essendo una funzione incognita, viene espansa attraverso lo sviluppo in Serie di Taylor [1]:

$$f_d(r) = 1 + k_1 r^2 + k_2 r^4 + k_3 r^6 + \dots \quad (\text{A.8})$$

La presenza delle sole potenze multiple di 2 è dovuta alla parità della funzione f_d .

Il punto ideale (x_u, y_u) si ottiene di conseguenza dal punto distorto attraverso la relazione

$$\begin{cases} x_u = x_d f_d(r_d) \\ y_u = y_d f_d(r_d) \end{cases} \quad (\text{A.9})$$

L'equazione A.8 ha 3 proprietà:

- È radialmente simmetrica rispetto al centro di distorsione;
- È una funzione continua;
- La funzione inversa è una funzione dispari.

In diverse occasioni l'equazione inversa della A.8 risulta necessaria per l'elaborazione dell'immagine per poter ottenere un'immagine finale densa. In questi casi ottenere i coefficienti dell'equazione inversa fornisce risultati migliori, ma si perde la semplificazione delle potenze pari.

A.1.2 Coordinate Mondo e Coordinate Camera

Per lavorare con immagini risulta necessario passare da un sistema di riferimento solidale con la camera, dove il punto $(0,0,0)$ coincide con il fuoco del sistema, a un sistema di riferimento più generico, che meglio si adatta alle esigenze dell'utilizzatore, dove la camera è piazzata in qualche punto del mondo, e orientata rispetto ad esso in un certo modo arbitrario.

Partendo dall'equazione A.4 per ottenere l'equazione definitiva della *pin-hole camera* bisogna applicare una matrice di rotazione, e un certo *offset* per considerare la posizione assoluta della camera nel mondo. Per motivi storici viene anche cambiato nome e direzione degli assi, per passare dal sistema di riferimento camera al sistema di riferimento mondo. In coordinate camera, la coordinata z rappresenta la profondità, mentre in coordinate mondo la coordinata Z rappresenti l'altezza e il piano (X, Y) è un piano cartesiano a tutti gli effetti.

Sia $(X_i, Y_i, Z_i)^T$ un generico punto in coordinate mondo, e $(x_i, y_i, z_i)^T$ il medesimo punto in coordinate camera.

La relazione che lega questi punti è

$$\begin{bmatrix} x_i \\ y_i \\ z_i \end{bmatrix} = \mathbf{R} \left(\begin{bmatrix} X_i \\ Y_i \\ Z_i \end{bmatrix} - \begin{bmatrix} X_0 \\ Y_0 \\ Z_0 \end{bmatrix} \right) = \mathbf{R} \begin{bmatrix} X_i \\ Y_i \\ Z_i \end{bmatrix} - \vec{t}_0 \quad (\text{A.10})$$

dove $(X_0, Y_0, Z_0)^T$ sono le coordinate mondo del *pin-hole* e \mathbf{R} è una matrice 3×3 che tiene conto delle rotazioni tra coordinate mondo e coordinate camera e, limitatamente al *framework* utilizzato, è anche una matrice di permutazione che permette di applicare una di nome e segno alle coordinate. È da notare che il vettore

$$\vec{t}_0 = \mathbf{R}(X_0, Y_0, Z_0)^T \quad (\text{A.11})$$

rappresenta la posizione dell'origine del mondo nelle coordinate camera.

È ora possibile scrivere l'equazione che proietta un generico punto in coordinate mondo $(X_i, Y_i, Z_i)^T$ in un punto corrispondente in coordinate immagine $(u_i, v_i)^T$:

$$\lambda \begin{bmatrix} u_i \\ v_i \\ 1 \end{bmatrix} = \mathbf{A}\mathbf{R} \begin{bmatrix} X_i - X_0 \\ Y_i - Y_0 \\ Z_i - Z_0 \end{bmatrix} \quad (\text{A.12})$$

o usando il vettore \vec{t}_0 si ottiene in forma estremamente compatta:

$$\lambda \begin{bmatrix} u_i \\ v_i \\ 1 \end{bmatrix} = \mathbf{A} \begin{bmatrix} \mathbf{R} & \vec{t}_0 \end{bmatrix} \begin{bmatrix} X_i \\ Y_i \\ Z_i \\ 1 \end{bmatrix} \quad (\text{A.13})$$

dove $\begin{bmatrix} \mathbf{R} & \vec{t}_0 \end{bmatrix}$ è una matrice 3×4 .

Da questa equazione risulta abbastanza evidente come ad ogni punto dell'immagine (u_i, v_i) siano associati infiniti punti del mondo (X_i, Y_i, Z_i) , che vivono su una retta al variare del parametro λ .

Definendo $\mathbf{P} = \mathbf{A}[\mathbf{R}t]$, la matrice 3×4 di proiezione complessiva, e sottintendendo λ , si ottiene infine la più generica equazione che permette di rappresentare ogni proiezione:

$$\begin{bmatrix} u_i \\ v_i \\ 1 \end{bmatrix} = \mathbf{P} \begin{bmatrix} X_i \\ Y_i \\ Z_i \\ 1 \end{bmatrix} \quad (\text{A.14})$$

A.1.3 La matrice di rotazione

La generica matrice di rotazione si può esprimere come

$$\mathbf{R} = \begin{bmatrix} r_0 & r_1 & r_2 \\ r_3 & r_4 & r_5 \\ r_6 & r_7 & r_8 \end{bmatrix} \quad (\text{A.15})$$

usando l'indicizzazione tipica degli *array C* e non delle matrici, per motivi di compattezza del formalismo e per semplificare la scrittura successiva del codice.

La matrice di rotazione ha la proprietà di non modificare le distanze e le aree, in quanto $\det(\mathbf{R}) = 1$. L'inversa di una matrice di traslazione è la sua trasposta.

Per queste ragioni, risulta possibile, conoscendo la sottomatrice 2×2 superiore sinistra, ricavare tutti gli altri elementi della matrice stessa, basandosi sul fatto che ogni riga o colonna deve avere norma unitaria (in quanto basi ortonormali dello spazio). Allo stesso modo il prodotto scalare tra due vettori riga o due vettori colonna deve dare valore nullo, in quanto ortogonali tra di loro.

Dati tali vincoli, esistono due soluzioni esatte: Una di queste è:

$$\mathbf{R} = \begin{bmatrix} r_0 & r_1 & (1 - r_0^2 - r_1^2)^{\frac{1}{2}} \\ r_3 & r_4 & s(1 - r_3^2 - r_4^2)^{\frac{1}{2}} \\ (1 - r_0^2 - r_3^2)^{\frac{1}{2}} & s(1 - r_1^2 - r_4^2)^{\frac{1}{2}} & (r_0^2 + r_1^2 + r_3^2 + r_4^2 - 1)^{\frac{1}{2}} \end{bmatrix} \quad (\text{A.16})$$

dove $s = \text{sgn}(r_1 r_4 + r_2 r_5)$.

L'altra soluzione ha i segni invertiti.

Conoscendo infine due vettori colonna della matrice \vec{r}_1, \vec{r}_2 è possibile determinare la terza base come prodotto vettoriale dei due.

Il framework **GOLD** utilizzato dal laboratorio di Visione Artificiale dell'Università di Parma usa la seguente matrice di rotazione (rotopermutazione):

$$\mathbf{R} = \begin{bmatrix} -\sin \gamma & -\cos \rho \cos \gamma & -\sin \rho \cos \gamma \\ -\cos \gamma \sin \vartheta & \sin \rho \cos \vartheta + \cos \rho \sin \gamma \sin \vartheta & -\cos \rho \cos \vartheta + \sin \rho \sin \gamma \sin \vartheta \\ \cos \gamma \cos \vartheta & \sin \rho \sin \vartheta - \cos \rho \sin \gamma \cos \vartheta & -\cos \rho \sin \vartheta - \sin \rho \sin \gamma \cos \vartheta \end{bmatrix} \quad (\text{A.17})$$

dove ϑ è l'angolo di *pitch*, γ è l'angolo di *yaw* e ρ è l'angolo di *roll*. La matrice \mathbf{R} descritta in A.17 è di fatto anche una matrice di cambiamento di base, trasformando il sistema di riferimento da coordinate mondo, nel sistema in coordinate camera.

Per ottenere questa matrice sono state composte in cascata le seguenti matrici di rotazione.

Matrice di rotazione *pitch* ϑ (asse Y):

$$\mathbf{R}_{\vartheta} = \begin{bmatrix} \cos \vartheta & 0 & -\sin \vartheta \\ 0 & 1 & 0 \\ \sin \vartheta & 0 & \cos \vartheta \end{bmatrix} \quad (\text{A.18})$$

Matrice di rotazione *yaw* γ (asse Z):

$$\mathbf{R}_{\gamma} = \begin{bmatrix} \cos \gamma & -\sin \gamma & 0 \\ \sin \gamma & \cos \gamma & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad (\text{A.19})$$

Matrice di rotazione *roll* ρ^* (asse X):

$$\mathbf{R}_{\rho} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos \rho & \sin \rho \\ 0 & -\sin \rho & \cos \rho \end{bmatrix} \quad (\text{A.20})$$

Infine la matrice di permutazione che esegue il cambiamento di base:

$$\mathbf{P}_0 = \begin{bmatrix} 0 & -1 & 0 \\ 0 & 0 & -1 \\ 1 & 0 & 0 \end{bmatrix} \quad (\text{A.21})$$

La matrice di permutazione trasforma il sistema di riferimento mondo in sistema di riferimento camera (X crescente verso destra, Y crescente verso il basso, Z la profondità).

Usando queste matrici, la matrice di rotazione \mathbf{R} si può scrivere:

$$\mathbf{R} = \mathbf{P}_0 \cdot \mathbf{R}_{\vartheta} \cdot \mathbf{R}_{\gamma} \cdot \mathbf{R}_{\rho} \quad (\text{A.22})$$

A.2 Rettificazione

Questa ultima parte di questa appendice affronta dal punto di vista teorico il problema della rettificazione, ovvero come trasformare una immagine acquisita da una camera con certe caratteristiche nell'immagine prodotta da una camera da caratteristiche diverse.

Consideriamo l'equazione A.12. Supponiamo di voler trasformare l'immagine prodotta con i parametri \mathbf{A}_1 e \mathbf{R}_1 , in una nuova immagine di parametri \mathbf{A}_2 e \mathbf{R}_2 . Invertendo l'equazione A.12 e proiettando nelle coordinate della nuova immagine si ottiene

$$\begin{bmatrix} u_{1,i} \\ v_{1,i} \\ 1 \end{bmatrix} = \mathbf{A}_1 \mathbf{R}_1 \mathbf{R}_2^{-1} \mathbf{A}_2^{-1} \begin{bmatrix} u_{2,i} \\ v_{2,i} \\ 1 \end{bmatrix} \quad (\text{A.23})$$

che è in effetti una trasformazione omografica $\mathbf{H}_{2,1} = \mathbf{A}_1 \mathbf{R}_1 \mathbf{R}_2^{-1} \mathbf{A}_2^{-1}$.

Tale trasformazione permette al costo di una omografia di cambiare gli angoli di orientazione della camera e tutti i parametri intrinseci (ovvero apertura e posizione del punto principale).

A.2.1 Coordinate omogenee

Chiameremo coordinate omogenee di un punto $P = (x, y)$ del piano una qualsiasi terna ordinata (X, Y, W) di numeri reali tali che $W \neq 0$, $\frac{X}{W} = x$ e $\frac{Y}{W} = y$. Allo stesso modo coordinate omogenee di un punto $P = (x, y, z)$ saranno una quaterna di numeri (X, Y, Z, W) tali che $W \neq 0$ e $\frac{X}{W} = x$, $\frac{Y}{W} = y$ e $\frac{Z}{W} = z$.

Risulta ben chiaro che il punto espresso in coordinate omogenee (X, Y, Z, W) equivale al punto *reale* $(\frac{X}{W}, \frac{Y}{W}, \frac{Z}{W})$.

Le coordinate omogenee hanno le seguenti proprietà:

- Le coordinate omogenee sono definite a meno di un coefficiente di proporzionalità. Ad esempio, la terna $(x, y, 1)$ e ogni suo multiplo r , ovvero $(x, y, 1) = (rx, ry, r)$, sono coordinate omogenee dello stesso punto dello spazio (x, y) ;
- I punti in coordinate omogenee con coordinata W nulla sono detti **impropri**, e non hanno nessun significato geometrico nello spazio cartesiano, ma possono rappresentare un punto all'infinito, nella direzione del vettore tridimensionale (X, Y, Z) .

In coordinate omogenee c'è pertanto distinzione tra vettore ($W = 0$) e punto ($W \neq 0$), cosa che non accade con le coordinate euclidee.

Le coordinate omogenee permettono di rappresentare punti all'infinito, e consentono di esprimere tutte le trasformazioni di coordinate in forma matriciale. L'insieme costituito da tutte le quaterne (terne) non nulle forma uno spazio proiettivo tridimensionale (bidimensionale).

L'uso di coordinate omogenee è usato in *computer graphics* per il fatto non banale che qualunque trasformazione affine è rappresentabile con un prodotto tra matrice, e lo è perfino la stessa proiezione prospettica.

A.3 Trasformazioni omografiche

Dati due piani distinti Π_1 e Π_2 si dice che sono riferiti a una trasformazione omografica (*homographic transformation*) quando esiste una corrispondenza biunivoca tale che:

- ad ogni punto o a ogni rette di Π_1 corrisponde un solo punto e una sola retta di Π_2
- ad ogni fascio di rette di Π_1 corrisponde un fascio proiettivo su Π_2

L'omografia (la trasformazione omografica) deve essere rappresentata da equazioni del tipo:

$$\begin{aligned} u' &= \frac{m_0 * u + m_1 * v + m_2}{m_6 * u + m_7 * v + m_8} \\ v' &= \frac{m_3 * u + m_4 * v + m_5}{m_6 * u + m_7 * v + m_8} \end{aligned} \quad (\text{A.24})$$

dove (u, v) sono coordinate dei punti appartenenti al piano Π_1 , mentre (u', v') sono punti del piano Π_2 . Il sistema è sovradimensionato (i parametri per una determinata trasformazione non sono univoci), pertanto normalmente si normalizzano, ponendo $m_8 = 1$, e i parametri da determinare si riducono a 8.

Per la sua particolare forma tale trasformazione è facilmente descrivibile in forma di coordinate omogenee (A.2.1):

$$\begin{bmatrix} u' \\ v' \\ w' \end{bmatrix} = \mathbf{H}_{ij}^{\Pi} \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} \quad (\text{A.25})$$

Viene definita matrice omografica \mathbf{H}_{ij}^{Π} la matrice che converte punti omogenei \vec{x}_i appartenenti al piano Π dell'immagine i in punti \vec{x}_j omogenei dell'immagine j con la relazione

$$\vec{x}_j = \mathbf{H}_{ij}^{\Pi} \vec{x}_i \quad (\text{A.26})$$

Per mantenere il riferimento a un array in C la matrice \mathbf{H}_{ij}^{Π} 3×3 viene espressa usando i coefficienti $m_0 \dots m_8$ piuttosto che la classica sintassi per indicare gli elementi della matrice.

Se i due piani sono paralleli ($m_6 = 0 \wedge m_7 = 0$) la trasformazione è detta affine (*affine transformation*) ed è rappresentata dalle equazioni:

$$\begin{aligned} u' &= m_0 * u + m_1 * v + m_2 \\ v' &= m_3 * u + m_4 * v + m_5 \end{aligned} \quad (\text{A.27})$$

Un modo per realizzare le trasformazioni prospettiche, trascurando la separazione tra parametri intrinseci ed estrinseci (e la loro determinazione in forma esplicita), è determinare i soli 8 coefficienti della matrice omografica A.24 (per esempio attraverso il metodo dei minimi quadrati, un modo per ricavare i coefficienti si trova nell'equazione 2.2). Attenzione che tale trasformazione vale solamente per i punti del piano che si è andato a rimappare.

L'inversa della trasformazione A.24 (*unnormalized inverse homographic matrix*) è una trasformazione omografica anche essa:

$$\begin{aligned} u &= \frac{(m_5 m_7 - m_4 m_8) u' + (m_1 m_8 - m_2 m_7) v' + m_4 m_2 - m_1 m_5}{(m_4 m_6 - m_3 m_7) u' + (m_0 m_7 - m_1 m_6) v' + m_1 m_3 - m_4 m_0} \\ v &= \frac{(m_3 m_8 - m_5 m_6) u' + (m_2 m_6 - m_0 m_8) v' + m_0 m_5 - m_2 m_3}{(m_4 m_6 - m_3 m_7) u' + (m_0 m_7 - m_1 m_6) v' + m_1 m_3 - m_4 m_0} \end{aligned} \quad (\text{A.28})$$

Appendice B

Matrice Pseudo-Inversa e fattorizzazione SVD

È fatto assai consueto nei problemi di ingegneria dover affrontare sistemi lineari con più dati che incognite. Se è possibile fare l'ipotesi che tutti questi dati siano *inliers*, una approssimazione che minimizza l'errore ai minimi quadrati della soluzione è solitamente la scelta corretta. La matrice pseudo-inversa fornisce una approssimazione ai minimi quadrati del sistema originale.

Consideriamo un sistema lineare *sovradimensionato*

$$\mathbf{A}x = y \tag{B.1}$$

dove \mathbf{A} è una matrice rettangolare $m \times n$, ed essendo sovradimensionato $m > n$. Tale matrice, essendo rettangolare, non ammette inversa.

Dato un vettore x , possibile soluzione del problema, è sempre possibile calcolare un indice dell'errore che questa soluzione comporterebbe se fosse scelta:

$$e(x) = \|\mathbf{A}x - y\|^2 \tag{B.2}$$

Ricavare la soluzione ai minimi quadrati del problema significa minimizzare la funzione $e(x)$ attraverso il metodo dei minimi quadrati, individuando il punto x in cui il gradiente della funzione si annulla.

Senza coinvolgere le derivate, si può dimostrare che la soluzione x che minimizza la funzione $e(x)$ esiste e vale

$$\begin{aligned} \mathbf{A}x &= y \\ \mathbf{A}^T \mathbf{A}x &= \mathbf{A}^T y \\ x &= (\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T y \end{aligned} \quad (\text{B.3})$$

Per costruzione x è ovviamente una soluzione del sistema B.1, ed è anche il vettore che minimizza la funzione B.2. Viene indicata con \mathbf{A}^+ la matrice pseudoinversa (*pseudoinverse matrix*) di \mathbf{A} , la quale vale

$$\mathbf{A}^+ = (\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T \quad (\text{B.4})$$

Questa soluzione del sistema è detta pseudoinversa di Moore-Penrose.

Alcune proprietà di questa matrice sono

- La pseudoinversa di una matrice esiste se esiste l'inversa di $\mathbf{A}^T \mathbf{A}$.
- La pseudoinversa di una matrice quadrata coincide con la sua inversa.
- La pseudoinversa di una matrice, se esiste, è unica.

Se è possibile assegnare alle equazioni del sistema pesi diversi, questi possono essere rappresentati in una matrice diagonale di preconditionamento della matrice \mathbf{A} .

La matrice pseudoinversa si può calcolare utilizzando la *Singular Value Decomposition* (SVD). Sia $\mathbf{A} = \mathbf{U}\mathbf{S}\mathbf{V}^*$ la Singular Value Decomposition di \mathbf{A} , dunque si può dimostrare che $\mathbf{A}^+ = \mathbf{V}\mathbf{S}^+\mathbf{U}^*$ dove la pseudoinversa di una matrice diagonale è sempre una matrice diagonale con i valori reciproci. Se infine il sistema è omogeneo $\mathbf{A}x = 0$ la SVD è una tecnica molto rapida per calcolare la soluzione ai minimi quadrati del problema, visto che le basi del *kernel* di \mathbf{A} sono le colonne di \mathbf{V} associate ai valori nulli della matrice diagonale. Nel caso di sistema sovradimensionato le colonne di \mathbf{V} rappresentano la base che minimizza la norma di $\mathbf{A}x$.

Per concludere la trattazione, l'ultimo metodo per ottenere la soluzione cercata (chiamato delle *normal equations*, ovvero equazioni perpendicolari) è, ricordando i passaggi sopra, risolvere con la tecnica di risoluzione che si preferisce il sistema $n \times n$ del tipo $\mathbf{A}^T \mathbf{A}x = \mathbf{A}^T y$.

Bibliografia

- [1] 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, August 1987.
- [2] Y.I. Abdel-Aziz and H.M. Karara. Direct linear transformation from comparator coordinates into object space coordinates in close-range photogrammetry. In *Proc. ASP/UI Symp. on Close-Range Photogrammetry*, pages 1–18, Urbana, Illinois, January 1971.
- [3] Frédéric Devernay and Olivier Faugeras. Straight lines have to be straight: automatic calibration and removal of distortion from scenes of structured environments. *Mach. Vision Appl.*, 13(1):14–24, 2001.
- [4] Zhengyou Zhang. A flexible new technique for camera calibration. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(11):1330–1334, 2000. Available from World Wide Web: <http://research.microsoft.com/~zhang/Calib/>.
- [5] 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.
- [6] H. A. Mallot, E. Schulze, and K. Storjohann. Neural network strategies for robot navigation. In *Proc. of nEuro'88*, 1988.

- [7] J. Gaspar, J. Santos-Victor, , and J. Sentieiro. Ground plane obstacle detection with a stereo vision system. In *Proceedings of the International Workshop on Intelligent Robotic Systems*, Grenoble, France, 1994.
- [8] Fuxing Li and Michael Brady. Dynamic calibration of an active vision system to compute the ground plane transformation. In *ACCV '98: Proceedings of the Third Asian Conference on Computer Vision-Volume I*, pages 120–127, London, UK, 1997. Springer-Verlag.
- [9] Stuart Cornell, John Porrill, and John E W Mayhew. Ground plane obstacle detection under a variable camera geometry using a predictive stereo matcher. In *BMVC92*, pages 548–559, 1992.
- [10] Y. Shao, J. E. W. Mayhew, and S. D. Hippisley-Cox. Ground plane obstacle detection of stereo vision under variable camera geometry using neural nets. In *BMVC '95: Proceedings of the 1995 British conference on Machine vision (Vol. 1)*, pages 217–226, Surrey, UK, UK, 1995. BMVA Press.
- [11] Herbert Bay, Tinne Tuytelaars, and L. Van Gool. Surf: Speeded up robust features. In *9th European Conference on Computer Vision*, Graz Austria, May 2006.
- [12] Massimo Bertozzi, Alberto Broggi, and Alessandra Fascioli. Stereo Inverse Perspective Mapping: Theory and Applications. *Image and Vision Computing Journal*, 8(16):585–590, 1998.
- [13] J. Lobo and J. Dias. Ground plane detection using visual and inertial data fusion. *Intelligent Robots and Systems, 1998. Proceedings., 1998 IEEE/RSJ International Conference on*, 2:912–917 vol.2, Oct 1998.
- [14] B. Bhanu, B. Roberts, and J. Ming. Inertial navigation sensor integrated motion analysis for obstacle detection. *Robotics and Automation, 1990. Proceedings., 1990 IEEE International Conference on*, pages 954–959 vol.2, May 1990.

- [15] K.-D. Kuhnert. Quantitative motion stereo for high speed car driving. *Intelligent Robots and Systems, 1992., Proceedings of the 1992 IEEE/RSJ International Conference on*, 3:2147–2154, Jul 1992.
- [16] J. Huber and V. Graefe. Motion stereo for mobile robots. *Industrial Electronics, 1993. Conference Proceedings, ISIE'93 - Budapest., IEEE International Symposium on*, pages 263–270, 1993.
- [17] K. Fintzel, R. Bendahan, C. Vestri, S. Bougnoux, and T. Kakinami. 3d parking assistant system. *Intelligent Vehicles Symposium, 2004 IEEE*, pages 881–886, June 2004.
- [18] C. Vestri, S. Bougnoux, R. Bendahan, K. Fintzel, S. Wybo, F. Abad, and T. Kakinami. Evaluation of a vision-based parking assistance system. *Intelligent Transportation Systems, 2005. Proceedings. 2005 IEEE*, pages 131–135, Sept. 2005.
- [19] Jin Zhou and Baoxin Li. Homography-based ground detection for a mobile robot platform using a single camera. *Robotics and Automation, 2006. ICRA 2006. Proceedings 2006 IEEE International Conference on*, pages 4100–4105, May 2006.
- [20] C. J. Harris and M. Stephans. A combined corner and edge detector. *4th Alvey Vision Conference Manchester*, pages 147–151, 1988.
- [21] R. Alix, F. Le Coat, and D. Aubert. Flat world homography for non-flat world on-road obstacle detection. In *Procs. IEEE Intelligent Vehicles Symposium 2003*, pages 310–315, Columbus, USA, June 2003.
- [22] Parag H. Batavia, Dean A. Pomerleau, and Charles E. Thorpe. Overtaking Vehicle Detection using Implicit Optical Flow. In *Procs. IEEE Intl. Conf. on Intelligent Transportation Systems'97*, pages 729–734, Boston, USA, November 1997.

- [23] R. Labayrade, D. Aubert, and J.-P. Tarel. Real time obstacle detection on non flat road geometry through ‘v-disparity’ representation. In *Proceedings of IEEE Intelligent Vehicle Symposium*, volume 2, pages 646–651, Versailles, France, 2002. <http://perso.lcpc.fr/tarel.jean-philippe/publis/iv02.html>.
- [24] A. Wedel, U. Franke, J. Klappstein, T. Brox, and D. Cremers. Realtime depth estimation and obstacle detection from monocular video. In K. Franke et al., editor, *Pattern Recognition (Proc. DAGM)*, volume 4174 of *LNCS*, pages 475–484, Berlin, Germany, September 2006. Springer.
- [25] A. Wedel, T. Schoenemann, T. Brox, and D. Cremers. Warpcut - fast obstacle segmentation in monocular video. In *Pattern Recognition (Proc. DAGM)*, LNCS, Heidelberg, Germany, September 2007. Springer.
- [26] Nicolas Simond and Michel Parent. Obstacle detection from ipm and superhomography. *Intelligent Robots and Systems, 2007. IROS 2007. IEEE/RSJ International Conference on*, pages 4283–4288, 29 2007-Nov. 2 2007.
- [27] G.P. Stein, O. Mano, and A. Shashua. A robust method for computing vehicle ego-motion. *Intelligent Vehicles Symposium, 2000. IV 2000. Proceedings of the IEEE*, pages 362–368, 2000.
- [28] T. Suzuki and T. Kanade. Measurement of vehicle motion and orientation using optical flow. In *Intelligent Transportation Systems, 1999. Proceedings. 1999 IEEE/IEEJ/JSAI International Conference on*, pages 25–30, 1999. Available from World Wide Web: <http://dx.doi.org/10.1109/ITSC.1999.821022>.
- [29] David Nistér, Oleg Naroditsky, and James Bergen. Visual odometry for ground vehicle applications. *Journal of Field Robotics*, 23:2006, 2006.
- [30] Christopher Rasmussen. Texture-based vanishing point voting for road shape estimation. In *British Machine Vision Conference*, 2004. Available from World Wide Web: <http://vision.cis.udel.edu/pubs/2004/Ras04a>.

- [31] Christopher Rasmussen and Thommen Korah. On-vehicle and aerial texture analysis for vision-based desert road following. In *CVPR '05: Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05) - Workshops*, page 66, Washington, DC, USA, 2005. IEEE Computer Society.
- [32] N. Simond and P. Rives. Homography from a vanishing point in urban scenes. *Intelligent Robots and Systems, 2003. (IROS 2003). Proceedings. 2003 IEEE/RSJ International Conference on*, 1:1005–1010 vol.1, Oct. 2003.
- [33] T. Suttorp and T. Bucher. Robust vanishing point estimation for driver assistance. *Intelligent Transportation Systems Conference, 2006. ITSC '06. IEEE*, pages 1550–1555, Sept. 2006.
- [34] M. Nieto, L. Salgado, F. Jaureguizar, and J. Cabrera. Stabilization of inverse perspective mapping images based on robust vanishing point estimation. *Intelligent Vehicles Symposium, 2007 IEEE*, pages 315–320, June 2007.
- [35] Massimo Bertozzi, Alberto Broggi, Alessandra Fascioli, and Massimiliano Sechi. Shape-based Pedestrian Detection. In *Procs. IEEE Intelligent Vehicles Symposium 2000*, pages 215–220, Detroit, USA, October 2000.
- [36] Alberto Broggi, Michael Del Rose, Alessandra Fascioli, Isabella Fedriga, and Amos Tibaldi. Stereo-based Preprocessing for Human Shape Localization in Unstructured Environments. In *Procs. IEEE Intelligent Vehicles Symposium 2003*, pages 410–415, Columbus, USA, June 2003.
- [37] Alberto Broggi, Pietro Cerri, and Pier Claudio Antonello. Multi-Resolution Vehicle Detection using Artificial Vision. In *Procs. IEEE Intelligent Vehicles Symposium 2004*, pages 310–314, Parma, Italy, June 2004.
- [38] Luca Bombini, Pietro Cerri, Paolo Medici, and Giancarlo Alessandretti. Radar-Vision Fusion for Vehicle Detection. In *Procs. Intl. Workshop on Intelligent Transportation*, pages 65–70, Hamburg, Germany, March 2006.

- [39] Massimo Bertozzi, Luca Bombini, Pietro Cerri, Paolo Medici, Pier Claudio Antonello, and Maurizio Miglietta. Obstacle Detection and Classification fusing Radar and Vision. In *Procs. IEEE Intelligent Vehicles Symposium 2008*, pages 608–613, Eindhoven, Netherlands, June 2008.
- [40] Ja Seong Ku, Kyoung Mu Lee, and Sang Uk Lee. Multi-image matching for a general motion stereo camera model. *Image Processing, 1998. ICIP 98. Proceedings. 1998 International Conference on*, 2:608–612 vol.2, Oct 1998.
- [41] M.S. Park, M. Mozerov, D.Y. Kim, K.S. Roh, and T.S. Chui. Object shape recovery in lateral navigation system using motion stereo technique. *Multisensor Fusion and Integration for Intelligent Systems, 1999. MFI '99. Proceedings. 1999 IEEE/SICE/RSJ International Conference on*, pages 273–278, 1999.
- [42] Alberto Broggi, Claudio Caraffi, Pier Paolo Porta, and Paolo 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, September 2006.
- [43] Olivier Faugeras. *Three-dimensional computer vision: a geometric viewpoint*. MIT Press, Cambridge, MA, USA, 1993.

Ringraziamenti

Prima di tutti vorrei ringraziare il mio supervisore, Alberto Broggi, per la fiducia, le idee, i suggerimenti, gli incoraggiamenti in questo periodo (presente, passato e futuro) che ho trascorso al VisLab di Parma. Allo stesso modo va un doveroso ringraziamento a Massimo Bertozzi per i contributi dati al progetto *PREVENT*.

Mi ritengo fortunato a poter lavorare in un team così dinamico che mira a un livello di eccellenza. Sono da ringraziare tutti i miei colleghi per le continue occasioni di discussione che hanno portato costanti miglioramenti nel mio lavoro. Questo gruppo, proprio in questi stessi anni, è cresciuto in maniera incredibile, sia per livello tecnologico ma anche per numero di persone, diventando un punto di riferimento a livello mondiale per la Visione Artificiale nel campo Automobilistico.

Un ringraziamento particolare a tutti i componenti (sono innumerevoli) di *APALACI*, i quali hanno sia permesso il finanziamento di queste ricerche, che data un supporto materiale di questo lavoro. In particolare un ringraziamento a Agneta Sjögren e Jonas Nilsson (VTEC), Maurizio Miglietta e Pier Claudio Antonello (CRF).

Infine ringrazio i parenti e amici che in tutti questi anni non hanno fatto mai mancare il loro supporto non in termini tecnici ma soprattutto umani.