

ALMA MATER STUDIORUM
UNIVERSITÀ DEGLI STUDI DI BOLOGNA

Facoltà di Ingegneria

Corso di Laurea in INGEGNERIA ELETTRONICA

Tesi di Laurea in RETI LOGICHE L.A

Sviluppo di un sistema di change detection basato su informazioni 3D

Candidato:
Fabio Tonelli

Relatore:
Ing. Stefano Mattocchia

Correlatori:
Ing. Federico Tombari
Prof. Luigi di Stefano

Anno Accademico 2006/2007 - Sessione I

Alla mia famiglia

Indice

1	Introduzione	7
2	Background subtraction	9
1	Introduzione alla background subtraction	9
2	Background initialization	11
3	Limitazioni della background subtraction	13
3.1	Histogram specification	14
3	Sistemi di visione stereo	17
1	Introduzione ai sistemi stereo	17
2	Geometria di un sistema stereoscopico	18
3	Algoritmi per il matching stereo	22
4	Risultato finale degli algoritmi stereo: le mappe di disparità	23
4	Algoritmo di change detection 3D sviluppato	27
1	L'algoritmo	27
2	L'implementazione	31
2.1	Ambiente di sviluppo e settaggio delle variabili	31
2.2	Background initialization stereo e mono	32
2.3	Visualizzazione e elaborazione della sequenza	33
	<i>creaMaschera3D</i>	33
	<i>histoSpecMasked</i>	34
	<i>CreaMaschera2D</i>	35
	<i>finalOutput</i>	35
5	Risultati sperimentali	37
1	Sequenza "Laboratorio"	38

2	Sequenza “Outdoor”	45
3	Sequenza “Ufficio”	52
6	Conclusioni e sviluppi futuri	57

Capitolo 1

Introduzione

L'obiettivo di questa tesi è utilizzare con profitto dati provenienti da un sensore 3D per individuare in una scena le aree dove è avvenuto un cambiamento, quindi dove c'è stato un movimento o un'intrusione: creare cioè un algoritmo di *change detection*.

Per *change detection* si intende il rilevamento delle variazioni strutturali all'interno di una stessa scena in istanti differenti mediante l'analisi di una sequenza video; quindi un confronto tra fotogrammi, che mette in evidenza le zone in cui sono avvenuti cambiamenti.

Attualmente uno degli utilizzi principali di questa tecnologia è la video sorveglianza.

La tecnica che illustrerò nel corso di questa tesi è quella denominata *background subtraction* (o *background difference*), che consiste nella creazione di un'immagine di sfondo (*background*) che verrà poi utilizzata per effettuare la differenza col frame corrente, evidenziando come cambiati tutto ciò che è frapposto fra osservatore e background (*foreground*).

Questo metodo risulta generalmente poco affidabile con le classiche immagini 2D di una telecamera mono in presenza di disturbi non eliminabili alla sorgente (come errori di acquisizione introdotti dal rumore del sensore) o da variazioni di illuminazione della scena, in quanto la tecnica è basata sulla differenza in valore assoluto tra pixel adiacenti.

Le finalità di questa tesi sono infatti volte a risolvere molti di questi in-

convenienti, e la soluzione proposta da me è di utilizzare come input i dati 3D generati da un sistema stereoscopico, che non risentono di alcune delle problematiche dei sistemi 2D.

Nello specifico verranno sviluppati e analizzati metodi differenti e si cercherà di stabilire per via sperimentale quale sia il più efficace tra i seguenti tre approcci:

- a. Background difference con immagini 2D affiancata a un'elaborazione run-time coi dati 3D per aggiornare il background rendendolo più robusto alle variazioni di illuminazione.
- b. Background difference con mappe di disparità generate da un sistema stereo come input, e in determinate zone reputate inaffidabili l'approccio con le immagini 2D.
- c. Background difference con mappe di disparità generate da un sistema stereo come input, e solo nelle zone reputate dal sistema 3D probabilmente soggette a un cambiamento il metodo con le immagini 2D.

Nella parte finale verranno tratte le conclusioni su pregi e difetti dei possibili approcci, e si cercherà di individuare vantaggi e limitazioni di ognuno.

Capitolo 2

Background subtraction

In questo capitolo descriverò la tecnica della background subtraction, da me utilizzata nel corso della tesi, evidenziandone pregi e difetti a confronto di altre tecniche di change detection.

1 Introduzione alla background subtraction

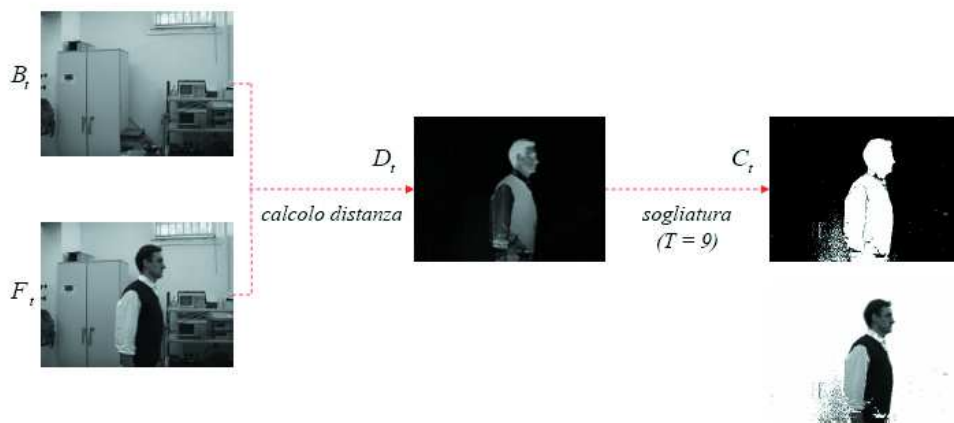


Figura 2.1: Background subtraction: B_t =Background, F_t =Frame corrente, D_t =Immagine differenza, C_t =Change mask

L'algoritmo di change detection background subtraction, analizzato anche dallo studente Luca Baldini [1], di cui possiamo vedere un esempio in Figura 2.1, richiede come ingresso una sequenza di immagini e in uscita genera

un'immagine detta “*Change Mask*”, o più semplicemente maschera, a cui è associato ad ogni pixel il valore u (“unchanged”, non cambiato) o c (“changed”, cambiato). La maschera è quindi un'immagine binaria dove i valori vengono solitamente identificati con il nero per quelli non cambiati (valore 0 nella scala di grigio) e bianco per quelli cambiati (255 nella scala di grigio).

La maschera ottenuta da questa tecnica non è basata sulla variazione temporale dei frame ma sul confronto col background, eliminando i problemi che insorgono nelle change detection basate sulla differenza tra frames adiacenti come ad esempio il caso in cui il soggetto in movimento sia di colore uniforme, cioè presenti zone a basso gradiente di livello di grigio, o sia fermo davanti al background in una posizione statica.

Quando viene effettuato il confronto tra il background e il frame attuale, nonostante la presenza di soggetti statici o poco texturati, essi vengono comunque rilevati come “estranei” al background.

La Background subtraction pur risolvendo alcune mancanze della temporal frame difference, non è esente da errori, che saranno opportunamente analizzati in seguito. Per limitare le imperfezioni, fortemente non lineari, dovute all'acquisizione delle immagini vengono effettuate operazioni di soglia e morfologiche per migliorare ulteriormente la maschera generata dall'algoritmo di background subtraction. Per avere un'idea dei miglioramenti che possono introdurre su imperfezioni di modesta entità si veda Figura 2.2.



Figura 2.2: A: Change mask senza operazioni morfologiche;

B: Dopo operazione morfologica con un'erosione e tre dilatazioni.

Per poter applicare questa tecnica è inoltre necessaria una fase iniziale di *background initialization*, dove viene appunto creato il background.

2 Background initialization

Per la generazione del background non è possibile memorizzare semplicemente il primo frame della sequenza video a causa del rumore in acquisizione della camera: gli errori introdotti dal sensore della videocamera fanno sì che il livello di grigio di ogni pixel oscilli in un certo range di valori più o meno probabili con una legge di probabilità tipicamente gaussiana. Il salvataggio di un unico fotogramma significherebbe quindi memorizzare per alcuni pixel un valore poco probabile e di conseguenza distante da quello reale, e ciò comporterebbe la presenza di errori nel background che andrebbero a compromettere irrimediabilmente la maschera.

È dunque necessario un campionamento del video per almeno alcune decine di frames, in cui è possibile costruire statistiche temporali per ogni pixel da cui fare una stima probabilistica dello sfondo.

Inoltre la possibilità della presenza di soggetti in movimento non voluti nella fase di *background initialization* rende necessario un controllo per scartare tutti quei campioni che in realtà non fanno parte dello sfondo.

Per la memorizzazione dei campioni all'interno dell'algoritmo è stata creata una variabile di tipo matriciale a tre dimensioni, con la base della stessa grandezza dell'immagine e l'altezza del valore di 256 (il numero di livelli di grigio possibili). In questa variabile, chiamata *Histogram3D*, viene memorizzato l'istogramma di livello di grigio di ogni pixel. L'istogramma di livello di grigio di un pixel (Figura 2.3) è un grafico bidimensionale discreto dove sull'asse delle ascisse si trovano i livelli di grigio(0-255), mentre su quello delle ordinate la frequenza con cui a tale pixel viene assegnato ogni livello di grigio; viene quindi effettuato un conteggio del numero di volte che al pixel vengono assegnati i vari valori di livello di grigio nell'ambito di un certo numero di campioni.

Per la creazione dei vari istogrammi si è proceduto nel modo seguente: per ogni fotogramma viene effettuato un controllo per verificare l'appartenenza

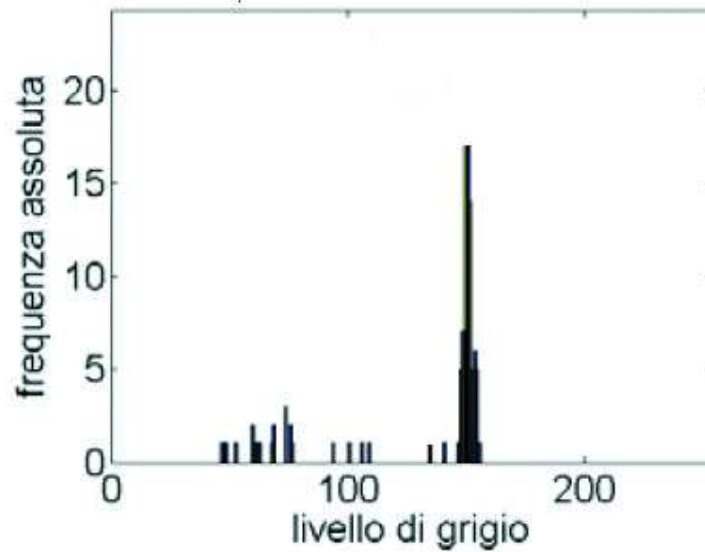


Figura 2.3: Esempio di istogramma

allo sfondo di ogni pixel, una volta verificata, in corrispondenza di ogni pixel viene incrementata la cella corrispondente al livello di grigio nella variabile `histogram3D`.

Si ottengono dunque, tanti istogrammi quanti sono i pixel, dove però il numero totale di campioni memorizzati per ogni istogramma coincide con il numero di frame in cui tale pixel è stato identificato come appartenente allo sfondo. Per i calcoli statistici è necessario conoscere tale numero, a tale scopo è stata creata un'altra variabile bidimensionale della grandezza dell'immagine dove ogni cella viene incrementata ogni volta che il pixel corrispondente ad essa viene classificato come appartenente allo sfondo e quindi memorizzato all'interno di `histogram3D`. Si ottiene quindi un contatore per ogni pixel.

A questo punto per la creazione del background, conoscendo tutti i valori che un determinato pixel ha assunto nel periodo di tempo osservato, si è scelto di effettuare la media statistica di tali valori (preferendola a mediana o moda) dividendo tutti i livelli di grigio memorizzati nel grafico per il numero totale

dei campioni.

$$media \doteq \frac{\sum_{lg=0}^{255} histogram3D[i, j, lg] \cdot lg}{numerocampioni} \quad (2.1)$$

dove [i] e [j] sono le coordinate dei pixel, mentre [lg] i livelli di grigio.

Il valore medio di (lg) del pixel (i,j) verrà assegnato al pixel (i,j) dell'immagine di background.

3 Limitazioni della background subtraction

Come già accennato in precedenza per un corretto funzionamento della Background subtraction è necessario che la scena sia pressochè statica sia dal punto di vista della conformazione dello sfondo, sia dal punto di vista dell'illuminazione. Se per esempio viene spostato un oggetto dello sfondo, nella maschera verrebbero evidenziate le zone dove era prima posizionato tale oggetto ed eventualmente dove è stato poi riposto (nel caso sia stato riposto sempre all'interno della visuale della camera). L'unica possibile modifica dell'algoritmo per fronteggiare questo problema è un aggiornamento continuo del background, in questo modo le variazioni della scena vengono assimilate dai nuovi background. Questa soluzione viene chiamata col termine *background update*, e non sarà applicata nel corso di questa tesi.

Un altro problema tipico della Background subtraction è quello del camouflage (Figura 2.6) cioè quando il soggetto in movimento presenta zone con caratteristiche simili, dal punto di vista dei livelli di grigio, allo sfondo che è andato a coprire; in tal caso il risultato della differenza può non presentare un valore sufficientemente alto da superare la soglia e quindi la zona viene identificata come non cambiata.

Per il corretto funzionamento della Background subtraction è inoltre necessaria la condizione di illuminazione stazionaria. infatti una seppur minima variazione dell'illuminazione della scena, come per esempio l'accensione di una lampada o nel caso di riprese da esterno il passare da uno stato di cielo



Figura 2.4: Frame corrente



Figura 2.5: Risultato della Background subtraction

Figura 2.6: Esempio di camouflage

nuvoloso a sole, varia sensibilmente il valore di livello di grigio di ogni pixel. Effettuando un confronto tra il background e il frame corrente tramite la differenza in valore assoluto le variazioni dei livelli di grigi vengono percepite come una variazione della scena e quindi compaiono nella maschera. Questa falsa rilevazione renderebbe inutilizzabile la change detection.

Per ovviare a questo inconveniente è stata applicata la tecnica della registrazione tonale, o *histogram specification*.

3.1 Histogram specification

Essa permette di aggiornare dinamicamente il nostro background durante l'esecuzione di una sequenza video grazie a operazioni che compie sull'istogramma dei livelli di grigio dell'immagine di sfondo, rendendolo più robusto a variazioni che interessano tutta la scena (come appunto l'illuminazione), ed è così strutturata [2]:

1. Viene estrapolato l'istogramma del background e del frame corrente.
2. Vengono confrontati gli istogrammi.
3. L'istogramma del background viene registrato tonalmente rispetto a quello del frame corrente, ottenendo una traslazione verso il bianco o il nero compatta di tutti i pixel a seconda dell'aumento o della diminuzione di illuminazione.

Un punto debole di questa tecnica è che nella creazione dell'istogramma del frame corrente possono venire considerati anche elementi del foreground, che

portano a una misurazione errata. Per ovviare a questo si può applicare la funzione di histogram specification con l'ausilio di una maschera di cambiamento, che permetta all'algoritmo di estrapolazione dell'istogramma di considerare solo i pixel sicuramente fissi, e quindi appartenenti al background. Ma come fare a creare la maschera per selezionarli? Per risolvere questo problema nel modo più robusto possibile si è deciso di integrare ai dati delle immagini, 2D, dati stereo 3D, che permettono con maggior sicurezza di individuare un pixel fisso o in movimento (ma comunque presentano altri difetti che illustrerò nel paragrafo successivo).

Capitolo 3

Sistemi di visione stereo

Nel seguente capitolo tratterò uno stato dell'arte dei sistemi stereo basati su due telecamere e del loro utilizzo nel creare mappe di disparità, ponendo particolare attenzione alla robustezza e allo stesso tempo incompletezza delle informazioni che ci forniscono. Per ulteriori e più esaustive informazioni sull'argomento si faccia riferimento al lavoro precedentemente svolto dal professor Stefano Mattocchia [3] presso ARCES, Università di Bologna, e alla tesi di laurea dello studente Kevin Gualandris [4].

1 Introduzione ai sistemi stereo

La possibilità di realizzare sensori in grado di fornire informazione tridimensionale (3D) ha sempre riscosso notevole interesse sia in ambito scientifico sia in ambito industriale per via delle innumerevoli applicazioni nelle quali tali sensori potrebbero essere utilizzati. Di particolare interesse risultano sensori tridimensionali di tipo passivo basati su telecamere tradizionali in quanto potenzialmente più economici e meno invasivi nell'ambiente in cui sono posti rispetto a sensori basati su tecnologie di tipo attivo quali time of flight, sensori basati su tecnologia laser, etc. Tra le diverse tecniche di computer vision note in letteratura e mirate alla ricostruzione della struttura tridimensionale di una scena osservata da una o più telecamere la visione stereoscopica è quella che ha riscosso la maggiore attenzione principalmente perché non impone alcun vincolo sulle caratteristiche degli oggetti presenti nella scena

(e.g. presenza o meno di oggetti in movimento, presenza o meno di particolari condizioni di illuminazione). La visione stereoscopica consente di inferire la struttura tridimensionale di una scena osservata da due o piú telecamere (nel caso specifico di due telecamere si parla di visione binoculare). Il principio alla base della visione stereoscopica, noto sin dal rinascimento, consiste in una triangolazione mirata a mettere in relazione la proiezione di un punto della scena sui due (o piú) piani immagine delle telecamere (e.g. tali punti sono denominati punti omologhi) che compongono il sistema di visione stereoscopico (stereo rig). L'individuazione dei punti omologhi, problema noto in letteratura come il problema della corrispondenza (matching stereo), consente di ottenere una grandezza denominata disparità (disparity) mediante la quale, conoscendo opportuni parametri del sistema stereoscopico, è possibile risalire alla posizione 3D del punto considerato. Il problema delle corrispondenze rimane ancora aperto e produce tuttora un'ampia attività di ricerca nonostante siano stati proposti sin dagli anni '60 innumerevoli algoritmi.

2 Geometria di un sistema stereoscopico

La trasformazione prospettica che mappa un punto dello spazio nel piano immagine di una telecamera implica la perdita dell'informazione relativa alla distanza. Questo può essere facilmente rilevato osservando la Figura 3.1 seguente nella quale due distinti punti (e.g. P e Q) nello spazio intersecati dallo stesso raggio che parte dal centro ottico O_l della telecamera corrispondono allo stesso punto (e.g. $p = q$) nel piano immagine.

Un metodo per poter risalire a quale punto dello spazio corrisponda la proiezione di un punto sul piano immagine di una telecamera consiste nell'utilizzo di due o piú telecamere. Infatti, come mostrato nella Figura 3.2 nel caso di un sistema composto da due telecamere, tra tutti punti nello spazio che giacciono sul raggio che passa per il centro ottico O_l e il punto q , proiezione di Q sul piano immagine p_l , al piú un solo punto (punto omologo) viene proiettato (e.g. q') anche sul piano immagine p_r . La determinazione dei punti omologhi consente di mettere in relazione le proiezioni dello stesso punto sui due piani immagini e di risalire, mediante una procedura denominata triangolazione,

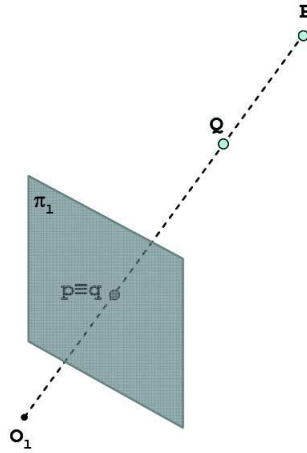


Figura 3.1: Proiezione di due distinti punti dello spazio sul piano immagine

alle coordinate dei punti dello spazio rispetto ad un sistema di riferimento opportuno. Sia dato un punto q su un piano immagine p_l , proiezione del punto Q appartenente allo spazio 3D. Per ottenere, attraverso la triangolazione, le coordinate 3D del punto nello spazio è necessario determinare (problema delle corrispondenze) il punto omologo q' nel piano immagine p_r . Tale problema, dato il punto q nel piano immagine p_l , richiederebbe una ricerca bidimensionale del punto omologo q' all'interno del piano immagine p_r .

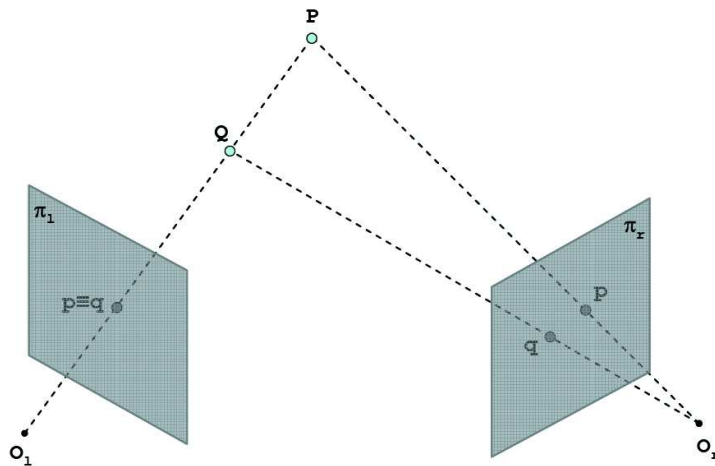


Figura 3.2: Principio alla base di un sistema stereoscopico

In realtà, sfruttando una particolare caratteristica della geometria del sistema

stereoscopico, è possibile effettuare la ricerca del punto omologo in uno spazio monodimensionale. Infatti, come mostrato nella Figura 3.3, gli omologhi di tutti i punti dello spazio che potrebbero risultare proiezione nello stesso punto q del piano immagine p_l , (e.g. punto p proiezione di P o lo stesso punto q proiezione di Q) giacciono sulla retta generata dall'intersezione tra il piano immagine p_r e il piano (denominato piano epipolare) che passa per la retta $O_l Q_P$ e i due centri ottici O_l e O_r . Tale vincolo, denominato vincolo epipolare, consente di limitare lo spazio di ricerca dei punti omologhi ad un segmento di retta semplificando considerevolmente il problema delle corrispondenze sia da un punto di vista della complessità algoritmica sia per quanto concerne la correttezza della soluzione. Si fa notare che il problema delle corrispondenze non necessariamente ha soluzione: infatti a causa della diversa posizione delle telecamere che compongono un sistema di visione stereoscopico nello spazio è possibile che un punto non risulti proiettato su tutti i piani immagini delle telecamere. In tal caso il problema delle corrispondenze non ha soluzione e non è possibile determinare la distanza del punto esaminato dalle telecamere (occlusioni).

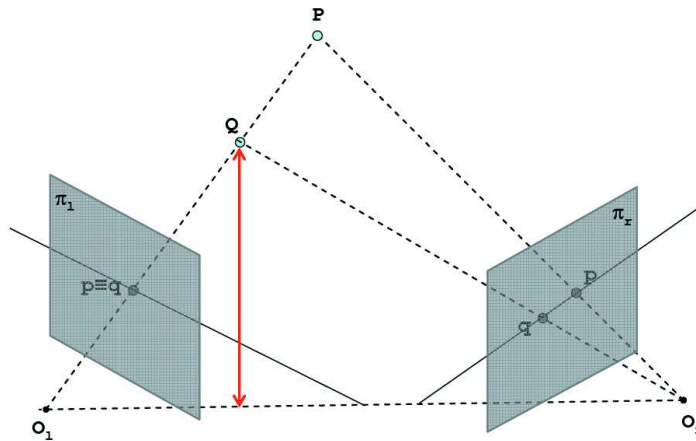


Figura 3.3: Vincolo epipolare

Un sistema di visione stereoscopico è completamente caratterizzato mediante i parametri intrinseci ed estrinseci. I primi consentono di definire la trasformazione che mappa un punto dello spazio 3D nelle coordinate del piano immagine di ogni telecamera e risultano le coordinate relative al piano imma-

gine del principal point (punto di intersezione tra il piano immagine e la retta ortogonale al piano immagine stesso passante per il centro ottico), la distanza focale, ed eventualmente altri parametri che descrivono altre caratteristiche del sensore (distorsione delle lenti, forma dei pixels, etc). I parametri estrinseci rappresentano le posizioni di ogni telecamera rispetto ad una sistema di riferimento noto.

La determinazione dei parametri intrinseci ed estrinseci, ottenuta mediante la procedura di calibrazione, consente quindi di descrivere completamente il sistema stereoscopico ed in particolare di inferire informazioni relative alle coordinate dei punti nello spazio mediante la triangolazione di punti omologhi (Figura 3.4).

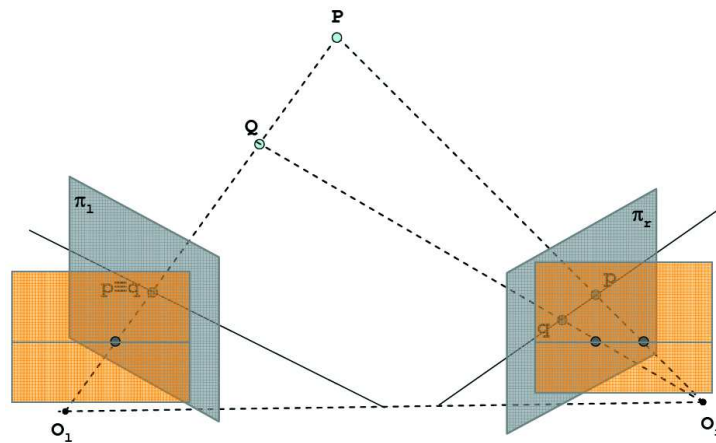


Figura 3.4: Immagini in forma standard

L'obiettivo della calibrazione è rettificare le immagini stereoscopiche provenienti dal dispositivo di acquisizione in modo che le righe di entrambe le immagini (scanline) vadano a corrispondere alle linee epipolari. La rettificazione consente anche di ridurre i problemi legati alla distorsione provocata dalle ottiche, e di ottenere immagini con la stessa distanza focale. A titolo di esempio, è mostrato nelle due figure seguenti l'effetto della rettificazione di immagini acquisite da un sistema stereoscopico alla risoluzione di 640x480 pixel. In Figura 3.5 sono visualizzate le immagini acquisite dal sistema stereo mentre in Figura 3.6 è mostrato il risultato della rettificazione ottenuta mediante i parametri stimati con la procedura di calibrazione. Le immagini

fanno parte di una sequenza video acquisita presso la facoltà di Ingegneria dell'università di Bologna, e reperibili integralmente in [6].



Figura 3.5: Coppia di immagini acquisite da un sistema stereo



Figura 3.6: Le stesse immagini stereo rettificate

3 Algoritmi per il matching stereo

Le immagini rettificate in uscita da un sistema stereoscopico devono quindi essere elaborate via software per poter ottenere informazioni 3D, e alla fine della loro manipolazione otteniamo le mappe di disparità, che verranno analizzate nel dettaglio nel paragrafo seguente.

Per fare il matching stereo vi sono innumerevoli algoritmi, molti dei quali sono analizzati dettagliatamente nell'articolo [7]; è però possibile in prima istanza classificarli in tre categorie principali: algoritmi di tipo feature-based, algoritmi di tipo denso-locali e denso-globali.

Quello di nostro interesse nel corso di questa tesi è di tipo denso-locale, e si chiama *Single Matching Phase* (SMP)[5]. Questo algoritmo da un lato è molto veloce, tanto da poter elaborare immagini stereo in tempo reale, dall'altro ha un limite legato al fatto che non necessariamente è in grado di produrre mappe di disparità completamente dense (e.g. non è garantito che per ogni punto dell'immagine sia possibile stabilire un valore di disparità).

Questo avviene principalmente per due motivi: a causa di regioni uniformi e a causa di occlusioni. Nel primo caso non è possibile determinare informazioni di disparità in regioni che presentano intensità costanti o comunque pochi elementi distintivi a causa dell'elevata ambiguità nella soluzione del problema delle corrispondenze. Nel secondo caso, a causa della geometria del sistema stereoscopico, in presenza di punti adiacenti posti a differente distanza dalle telecamere risulta che alcuni punti siano visibili in una immagine ma non nell'altra dello stereo-pair: in casi come questo non è possibile risolvere il problema delle corrispondenze.

4 Risultato finale degli algoritmi stereo: le mappe di disparità

L'output degli algoritmi di matching stereo sono le mappe di disparità. Tali mappe, codificate come immagini, memorizzano al loro interno informazioni relative alla distanza dei punti della scena acquisita da due o più sensori di immagine. E' convenzione diffusa rappresentare nelle mappe i punti con un valore di disparità valido con un colore sempre più chiaro man mano che la disparità cresce (punto vicino) e più scuro man mano che la disparità decresce (punto lontano). I punti ambigui in cui non è stato possibile calcolare attendibilmente la disparità (punti "*non matching*" o NM) a causa di regioni uniformi o di occlusioni si colorano di bianco.



Figura 3.7: Immagine Tsukuba e sua mappa di disparità ground truth

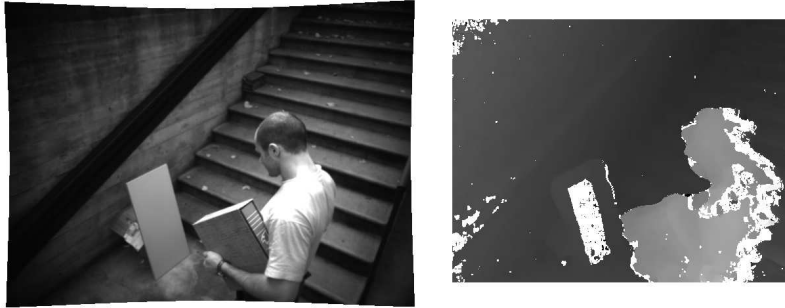


Figura 3.8: Mappa di disparità reale. Si noti come per la regione uniforme al centro del pannello appoggiato al muro non sia stato possibile determinare la disparità. Si osservi inoltre che in prossimità dei contorni le mappe di disparità presentano numerose occlusioni.

A titolo di esempio si riportano in Figura 3.7 e Figura 3.8 rispettivamente una mappa di disparità ideale (*ground truth*) e quella generata dall'algoritmo SMP con a fianco le rispettive immagini di riferimento. Sono rappresentati in bianco i punti per cui non è stato possibile risolvere il problema delle corrispondenze mentre con i livelli di grigio si codificano le informazioni relative alla disparità determinata.

Il grafico in Figura 3.9 rappresenta un semplice esempio relativo alla geometria di un sistema stereo: due immagini dello stesso punto P sono prese da differenti punti di vista. Sono evidenziati parametri intrinseci come la distanza fra i sensori ottici, detta baseline, b , e la lunghezza focale delle lenti f .

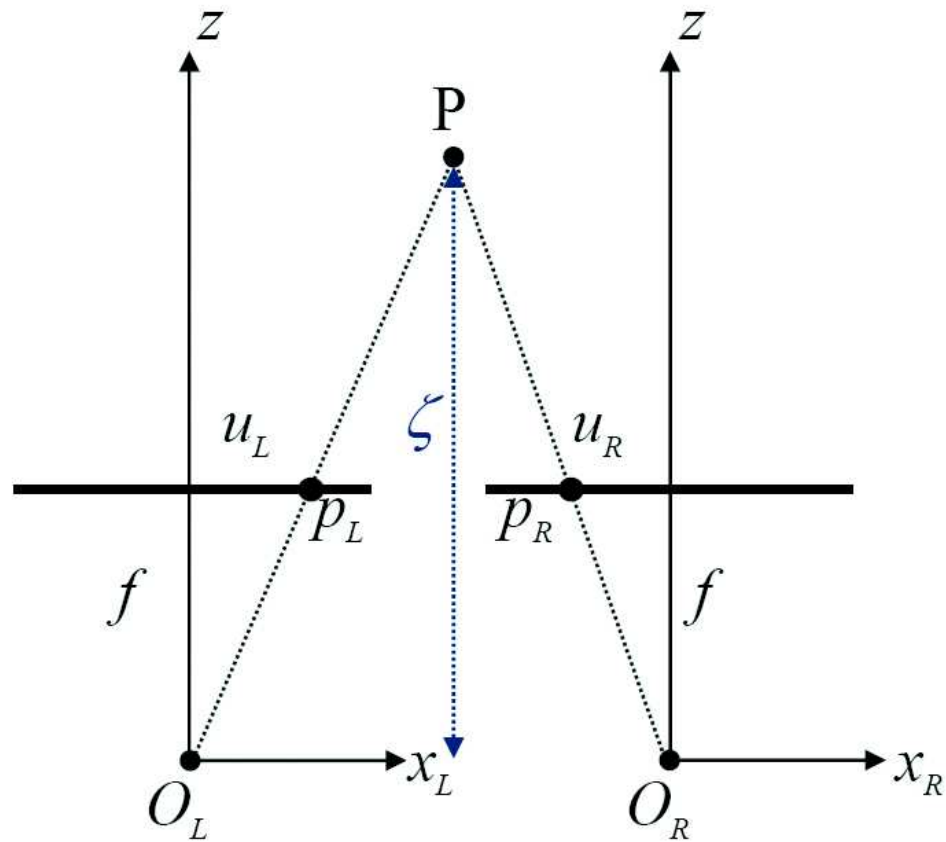


Figura 3.9: Le grandezze considerate nel calcolo

Normalmente si settano le telecamere in modo da avere le immagini ottenute complanari, anche se questo non è mai perfettamente possibile ma lo si può compensare con la rettificazione. Sotto questa condizione la differenza fra u_L e u_R viene detta disparità ed è direttamente correlata con la distanza ξ dell'oggetto normale rispetto al piano delle immagini. La relazione è:

$$\xi = \frac{b \cdot f}{d} \quad \text{dove} \quad d = u_L - u_r \quad (3.1)$$

Osservando la 3.1 e la Figura 3.10 si può intuire che ci sono limiti di vicinanza e lontananza oltre i quali la telecamera non può processare correttamente le immagini degli oggetti. Per avere una idea più precisa facciamo un esempio numerico: con una baseline di 8cm , pixel di $14\mu\text{m}$ e lunghezza focale delle

lenti di $6.3mm$ abbiamo che la minima distanza è mezzo metro con una disparità di $70pixel$, mentre la distanza massima oltre la quale l'immagine risultante non avverte variazioni è di circa $35metri$.

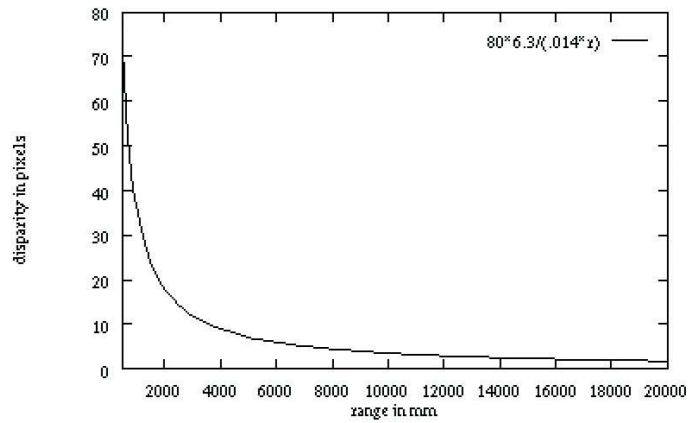


Figura 3.10: Relazione fra disparità e distanza in mm

Capitolo 4

Algoritmo di change detection 3D sviluppato

Alla luce dei concetti mostrati fino a questo punto viene presentata la parte sperimentale della mia tesi. Nel capitolo seguente verrà descritto nel dettaglio l'algoritmo di change detection 3D sviluppato da me, e analizzata la sua implementazione in C.

1 L'algoritmo

L'algoritmo sviluppato, schematizzato in Figura 4.1, prende in ingresso una sequenza di mappe di disparità e la rispettiva sequenza di immagini usate come riferimento per il calcolo della disparità. Dati questi presupposti un punto dell'immagine corrente corrisponde univocamente al punto rappresentato alla medesima coordinata nella mappa di disparità.

La prima operazione che compie è la generazione del background di entrambe le sequenze, che verrà successivamente utilizzato per l'operazione di background subtraction. Si è scelto fra i vari metodi possibili il più semplice, cioè immagazzinare e fare la media statistica delle prime immagini della scena, che devono essere senza soggetti in movimento, pena un'errata valutazione dello sfondo. Inoltre nella sequenza di disparità è molto importante individuare e segnalare le zone inaffidabili dello sfondo, in particolare quelle caratterizza-

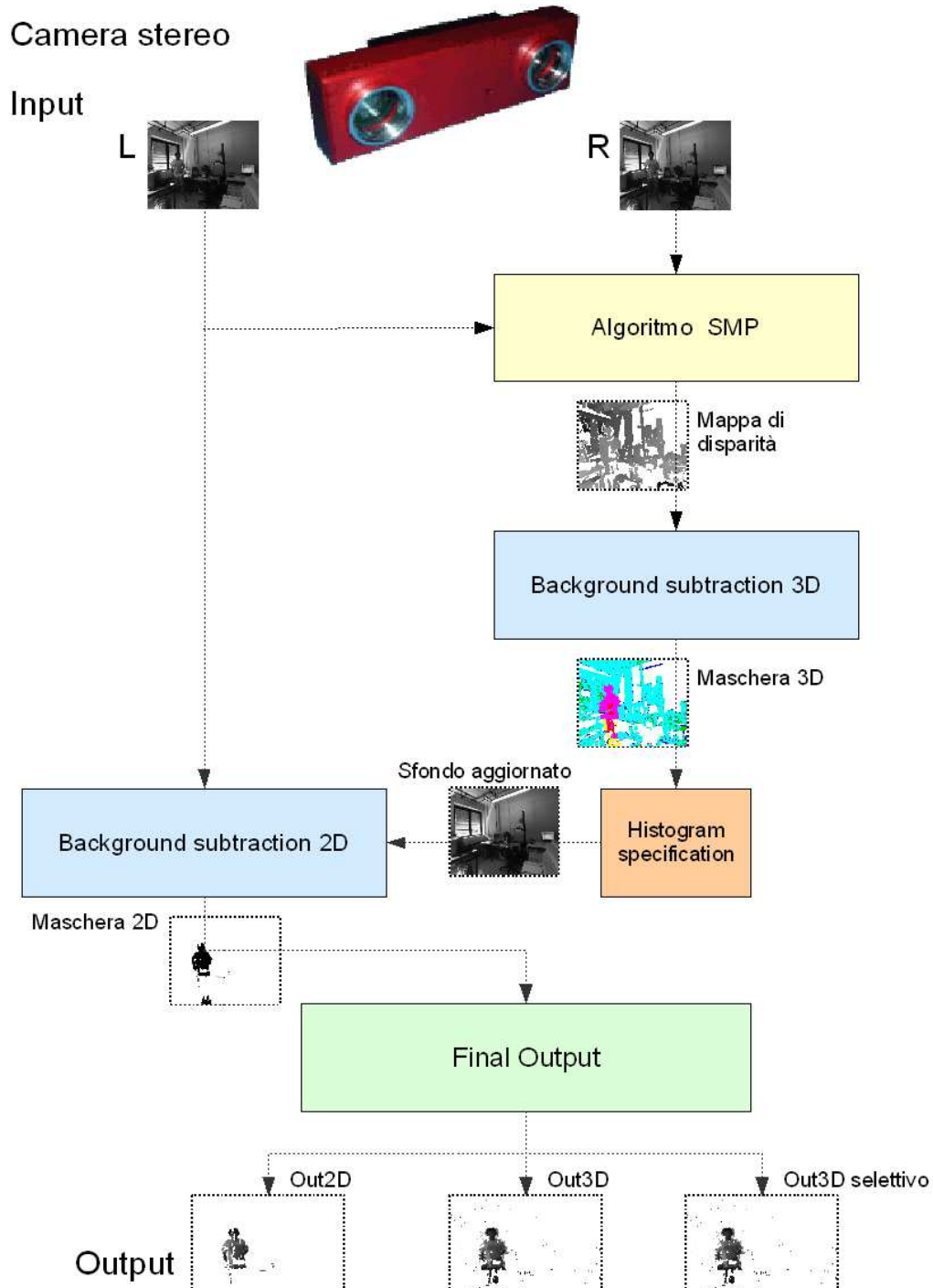


Figura 4.1: Visualizzazione grafica dell'algoritmo realizzato nel corso della tesi

te da punti che cambiano il proprio valore di disparità anche in assenza di movimento o da punti che passano da una situazione di disparità valida a una di non matching (NM) apparentemente senza motivo, che verranno rispettivamente colorate di verde e di blu scuro e non saranno più considerate nelle elaborazioni successive. L'esistenza di queste imperfezioni è dovuta al rumore, il cui effetto provoca un errore nella rilevazione della disparità dell'algoritmo SMP.

In seguito viene effettuata un'operazione di background subtraction per ognuna delle due sequenze, e gli output saranno maschere di cambiamento. La background difference della sequenza 2D indicherà nella maschera i punti rilevati come cambiati rispetto allo sfondo mediante un'opportuna operazione di differenza con sogliatura tra pixel corrente e rispettivo pixel nel background, mentre quella 3D permetterà di scoprire anche ulteriori informazioni sulla natura del cambiamento di ogni singolo punto.

Infatti analogamente alla maschera generata dalle immagini correnti in scala di grigi (*maschera2D*, nella maschera in uscita alla background subtraction delle mappe di disparità (*maschera3D*) sono segnalati i punti a disparità valida che hanno mantenuto invariato il proprio valore (colorati in azzurro), o lo hanno cambiato superando una certa soglia (colorati in fucsia). I punti inaffidabili già trovati nella fase di background generation vengono risegnalati (colori verde e blu), e rimangono i punti caratterizzati o nel background o nel foreground da una situazione di NM.

Nello sviluppo dell'algoritmo si è deciso di dividerli in tre categorie, e di rappresentarle tutte con colori differenti nella maschera 3D.

Essi sono i punti che hanno un valore NM sia nello sfondo che nell'immagine della mappa di disparità corrente (colorati di bianco), che mentre nello sfondo hanno un valore NM nell'immagine della mappa di disparità corrente presentano un valore valido (colorati di rosso) e, viceversa, che passano da una condizione di valore valido ad una di NM (colorati in giallo).

Come già accennato in precedenza il sistema 3D ci fornisce dati molto robusti e affidabili, infatti i punti azzurri appartengono con alto grado di affidabilità

allo sfondo, mentre quelli fucsia, salvo leggere imperfezioni dovute al rumore, al foreground. Per quanto riguarda gli altri invece, se da un lato per quelli blu e verdi non possiamo dire nulla sulla loro provenienza, dall'altro i gialli e i rossi possono essere sia dovuti a movimento che a occlusioni, e quindi ci danno solo l'informazione di un potenziale cambiamento.

Come già spiegato nel Capitolo 2 Sezione 3, una variazione di illuminazione nella scena rischia di compromettere l'output della background subtraction 2D, perchè va a modificare il livello di grigio dei punti appartenenti allo sfondo. Si è quindi deciso di introdurre una funzione di histogram specification, implementata dallo studente Errico De Lisi, per aggiornare in tempo reale lo sfondo a queste possibili variazioni. Oltre alle immagini di sfondo e frame corrente questa funzione richiede in ingresso anche una maschera di cambiamento, come supporto per poter essere applicata solo ai punti sicuramente appartenenti al background dell'immagine corrente. A questo scopo si è deciso di creare questa maschera coi soli punti azzurri, considerati cioè come non cambiati, della maschera3D.

In quest'ottica dunque l'operazione di change detection sulle informazioni 3D serve principalmente a determinare una maschera di punti appartenenti al background da utilizzare ai fini della registrazione tonale.

In seguito l'output della funzione di histogram specification (che è lo sfondo aggiornato mediante la registrazione tonale), viene passato frame dopo frame alla funzione che calcola la background subtraction 2D.

Dopo l'acquisizione di tutti i dati l'algoritmo genera tre output finali, ognuno dei quali tenta di sfruttare diversamente in base alle situazioni i punti forti dei dati 2D e 3D :

Output2D assegna direttamente ai punti della maschera 2D i valori dell'immagine corrente di riferimento, e quindi utilizza le informazioni delle mappe di disparità solo nella fase di registrazione tonale dello sfondo

Output3D considera come non cambiati i punti della maschera 3D con la differenza dei valori tra sfondo e immagine corrente di riferimento sotto soglia (azzurri) e come cambiati quelli con la differenza superiore alla soglia prestabilita (fucsia). Tutti i pixel rimanenti vengono riempiti con i valori rispettivi di output2D

Output3D selettivo è invece uguale all'approccio precedente per quanto riguarda i punti azzurri e fucsia del 3D, con la differenza che riempie coi dati di output2D solo le zone indicate dalla maschera 3D come probabilmente soggette ad una variazione: quelle che passano da una situazione di valore di disparità valido ad uno NM o viceversa tra sfondo e immagine corrente mono (rossi o gialli).

Il risultato finale sono tre approcci differenti di change detection, ognuno con vantaggi e svantaggi che verranno analizzati in seguito.

2 L'implementazione

2.1 Ambiente di sviluppo e settaggio delle variabili

L'ambiente di sviluppo utilizzato è Microsoft visual studio 2005, in particolare il suo compilatore di C++. Per gestire le immagini si è fatto riferimento alla libreria opensource OpenCv, reperibile gratuitamente in [8].

Questa libreria tratta le immagini con tipo di dato `IplImage`, uno struct che ha per campi i valori di larghezza, altezza, numero di canali, ... e tutto quanto possa descrivere in modo esaustivo l'immagine specificata. Inoltre un campo molto importante è `imageData`, un puntatore che contiene le informazioni sui valori dei pixel.

Gli input saranno sequenze di immagini da file precedentemente acquisite (mappe di disparità e immagini mono rettificate) che aperte in sequenza frame per frame daranno origine a un flusso video. Per fare questo è stata creata la funzione `creaStringa` che avuto in ingresso nome dell'immagine, l'estensione e il numero univoco che la contraddistingue permette di aprirle in sequenza.

Una volta aperta, l'immagine corrente selezionata da file viene ritagliata per eliminare le cornici monocromatiche generate tipicamente dagli algoritmi densi locali, che andrebbero a compromettere le misurazioni successive per la registrazione tonale, e salvata in `IplImage *orig` se è una mappa di disparità, o in `IplImage *orig_rect` se immagine rettificata.

A questo punto vengono allocate delle variabili dinamiche, per l'esattezza `Histogram3D` (seguita dai suffissi che indicano l'immagine di disparità o quella mono in scala di grigi) e `counter`. La prima è di tipo doppio puntatore a `BYTE`, un tipo di dato definito come `unsigned char` (8 bit), e permette di salvare come matrice i primi n fotogrammi necessari alla fase di inizializzazione del background, che in seguito dovranno quindi essere sottoposti all'operazione di media. `counterNM` invece, come suggerisce il nome, è un contatore che incrementa il valore nella cella corrispondente a un determinato pixel ogni qual volta questo risulti NM. È fondamentale nella creazione del background delle mappe di disparità perché permette di individuare, grazie a una soglia prestabilita, l'inaffidabilità di certi punti che anche in assenza di movimento passano alternativamente da disparità valida a NM, e quindi porterebbero a una misurazione errata. L'esistenza di questi punti, situati principalmente nella zona di confine tra punti a validi e NM, è giustificata dai limiti degli algoritmi per la rilevazione della disparità.

2.2 Background initialization stereo e mono

Una volta settate tutte le dimensioni delle immagini e allocate le variabili inizia l'elaborazione vera e propria, il cui primo passo è la creazione dello sfondo.

Vengono salvate con un ciclo `for` le prime n immagini mono sicuramente ferme in `Histogram3D_rect`, e in seguito con la funzione `media_rect` ne viene calcolata la media, che verrà assunta come background. Per le immagini stereo la procedura è analoga, solo che è necessario un controllo aggiuntivo, per permettere di capire se le parti della scena individuate da un singolo pixel sono valide, NM o inaffidabili. Per fare questo si sono scelte delle soglie, $P1$ e $P2$. Inoltre, essendo i valori della disparità discreti, un punto che si dovesse trovare proprio sulla soglia in cui l'algoritmo SMP va a calcolare

la disparità potrebbe essere identificato in differenti frames a distanze non uguali, e quindi mostrare un'elevata varianza nei valori di grigio che assume (Figura 4.2). Per eliminare alla radice questo inconveniente quindi i punti del background a varianza troppo elevata vengono filtrati (Figura 4.3).



Figura 4.2: Visualizzazione grafica delle soglie per lo sfondo 3D

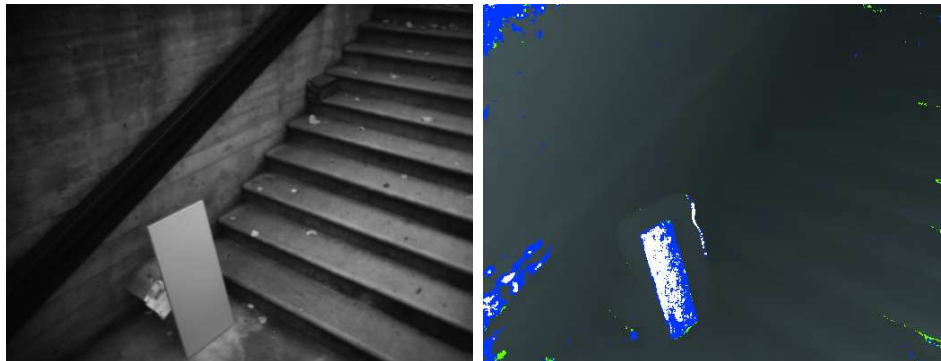


Figura 4.3: Sfondo 2D e Sfondo 3D: in Blu i punti inaffidabili, in verde quelli a varianza elevata

2.3 Visualizzazione e elaborazione della sequenza

A questo punto le due sequenze vengono analizzate e confrontate con lo sfondo appena creato, frame per frame. Per fare ciò ci si serve di quattro funzioni, il cuore vero e proprio dell'algoritmo, che vengono chiamate in questa sequenza: *creaMaschera3D*, *histoSpecMasked*, *creaMaschera2D*, *finalOutput*; analizziamole più nel dettaglio.

creaMaschera3D

La prima funzione si occupa di creare due maschere della Background difference tra lo sfondo 3D e le mappe di disparità: prende infatti in ingresso

entrambe queste immagini, e in uscita genera *binaryMask* e *maschera3D*, che permettono di distinguere tutte le tipologie di punti presenti nell'immagine a cui avevamo accennato prima.

maschera3D è un'immagine a colori di tipo BGR e segnala:

1. punti inaffidabili presenti nello sfondo
2. punti a varianza troppo elevata presenti nello sfondo
3. punti NM dello sfondo che risultano a disparità valida nell'immagine corrente
4. punti a disparità valida dello sfondo che risultano a disparità valida, ma con un valore di grigio differente, nell'immagine corrente
5. punti a disparità valida dello sfondo che risultano NM nell'immagine corrente
6. punti a disparità valida dello sfondo che risultano a disparità valida con un valore di grigio uguale entro una certa soglia nell'immagine corrente
7. punti NM dello sfondo che risultano NM nell'immagine corrente

binaryMask è invece un'immagine binaria, che può assumere cioè due soli valori per ogni pixel: bianco o nero. In particolare il colore bianco identifica i soli pixel del tipo 7, cioè sicuramente fissi, e mi servirà come supporto per applicare correttamente la prossima funzione.

histoSpecMasked

Questa funzione prende in ingresso due immagini mono in bianco e nero, l'immagine di background e l'immagine corrente, ed una mappa, *binaryMask*, generata dalle informazioni stereo e che identifica i punti sicuramente fermi nell'immagine; essa è fondamentale perché funge da supporto all'istogramma specification, che viene svolta proprio in questa funzione. La maschera infatti impedisce alla parte successiva dell'algoritmo di considerare nell'istogramma dell'immagine corrente le parti relative a cambiamenti strutturali della scena. Se così non fosse l'operazione di registrazione tonale, che aggiorna il background fino a fargli assumere un istogramma il più possibile simile a quello dell'immagine corrente, in presenza di soggetti in movimento di colore differente dallo sfondo porterebbe a visibili (e difficilmente eliminabili)

errori anche con lievissime variazioni di illuminazione. Applicata con questo accorgimento invece la tecnica della registrazione tonale rende il sistema molto robusto a variazioni di illuminazione diffuse, e genera un'immagine che rappresenta il background aggiornato.

CreaMaschera2D

CreaMaschera2D è molto semplice e calcola la background difference mediante la funzione *cvAbsDiff*, alla quale poi si applica una soglia, facendo una differenza assoluta pixel a pixel tra sfondo aggiornato da *histoSpecMasked* e immagini corrente mono. In uscita dà una maschera con valore nero per i pixel ritenuti in movimento, sopra soglia, e bianco per quelli sotto soglia.

finalOutput

Le informazioni 2D e 3D immagazzinate nelle due maschere vengono poi date insieme all'immagine corrente come input alla funzione *finalOutput*, dove si sceglie come utilizzarle per generare l'output finale del sistema. In particolare nel corso di questa tesi si è scelto di seguire diversi approcci, basati sul 3D e su 2D opportunamente correlati, cercando di coprire in parte i difetti di entrambe le tipologie di acquisizione dati.

Per la scelta nello specifico dei tre approcci si faccia riferimento al Capitolo 4 Sezione 1.

Capitolo 5

Risultati sperimentali

Al fine di valutare le prestazioni del sistema sviluppato, con particolare riferimento alla robustezza rispetto alle variazioni di illuminazione, in questo capitolo saranno esaminati i fotogrammi più significativi ed i relativi output di tre sequenze video, denominate nell'ordine *Laboratorio*, *Outdoor*, *Ufficio*. Mentre *Outdoor* è ambientata all'esterno le altre sono state acquisite nel corso della tesi in locali interni. In particolare *Laboratorio* presenta un'illuminazione prevalentemente solare, quindi molto stabile e uniforme, mentre *Ufficio* ha solo luce al neon.

1 Sequenza “Laboratorio”

Questa sequenza è stata ripresa nel Laboratorio di visione artificiale presso il DEIS, Università di Bologna.

É in un ambiente interno, con una diffusa e intensa illuminazione solare ed è soggetta ad una forte e diffusa variazione di illuminazione.

É interessante notare che in assenza di variazioni, di illuminazione o generate da soggetti in movimento, il metodo più efficace risulta essere l’approccio 2D (Figura 5.2 immagine di destra), che non segnala falsi cambiamenti, mentre entrambi i 3D (Figura 5.3) sono soggetti a un certo rumore di fondo (comunque facilmente eliminabile con poche operazioni di morfologia).

Il passaggio di un soggetto estraneo fintanto che le condizioni di illuminazione rimangono invariate viene rilevato abbastanza bene da tutti gli output. Infatti nonostante in certi frames sia presente molto camouflage nell’output 2D (Figura 5.6 immagine di destra) o qualora le mappe di disparità presentino molti punti NM sulla figura estranea allo sfondo (Figura 5.9 immagine di destra) tutti gli output danno buoni risultati.

Si noti inoltre la precisione con cui maschera2D rileva i bordi delle sagome in movimento.

In presenza di una sostanziosa variazione di illuminazione, come mostrato nella sequenza di immagini di Figura 5.16, abbiamo invece una pessima risposta da parte dell’output 2D, che nonostante l’histogram specification commette molti errori rilevando cambiamenti in zone rimaste invariate.

Negli stessi frames è già più evidente la robustezza dei dati 3D. Mentre output 3D risente delle variazioni di illuminazione solo nelle poche zone dello sfondo dove non si avevano valori validi di disparità, output 3D selettivo è ancora più robusto. É interessante notare però che la variazione di illuminazione ha reso NM molte zone che prima avevano un valore valido di disparità, e addirittura ha fatto cambiare il valore della disparità di altre fino a superare la soglia per la rilevazione del movimento.

A titolo di esempio si è anche voluto analizzare come sarebbe stata la scena senza la registrazione tonale dello sfondo (Figura 5.20). In questo caso l’output coi soli dati 2D sarebbe stato disastroso.

Frame 18, Figura 5.4

In questo frame vediamo il solo sfondo della scena. Quindi non c’è nessun movimento, e come si può notare l’output ottimale è quello semplicemente 2D, mentre entrambi i 3D sono soggetti a un lieve rumore, comunque eliminabile con poche operazioni di morfologia.

Frame 44, Figura 5.8

È possibile notare al centro sinistra dell’immagine una persona estranea alla scena. In questo caso mentre l’algoritmo 2D è soggetto a un certo camouflage entrambi i 3D sono più efficienti nel rilevare l’intera sagoma.

Frame 114, Figura 5.12

Il soggetto a sinistra, vicino al sensore della camera, risulta per la maggior parte NM nella mappa di disparità, e inoltre si trova in una delle zone poco affidabili dello sfondo. Per questo motivo gli output 3D sono meno soddisfacenti del 2D.

Frame 159, Figura 5.16

In questo caso è presente un’intensa variazione di illuminazione, che nonostante l’histogram specification causa svariati errori al 2D: sono infatti visibili molti aloni, non facilmente eliminabili con operazioni di morfologia. Al contrario i 3D sono più robusti, in particolar modo quello selettivo.

Frame 173, Figura 5.20

Nelle immagini si è messa a confronto la risposta del sistema applicando una background subtraction puramente 2D all’immagine corrente, senza cioè l’utilizzo della funzione di histogram specification, con i tre output dell’algoritmo.

Figura 5.1: Immagine corrente e mappa di disparità del frame 18



Figura 5.2: Maschera 3D a colori e Output 2D

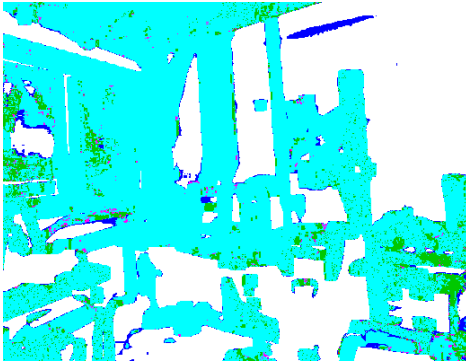


Figura 5.3: Output 3D e Output 3D selettivo



Figura 5.4: FRAME 18, sfondo

Figura 5.5: Immagine corrente e mappa di disparità del frame 18



Figura 5.6: Maschera 3D a colori e Output 2D



Figura 5.7: Output 3D e Output 3D selettivo



Figura 5.8: FRAME 44, soggetto estraneo

Figura 5.9: Immagine corrente e mappa di disparità del frame 114



Figura 5.10: Maschera 3D a colori e Output 2D



Figura 5.11: Output 3D e Output 3D selettivo



Figura 5.12: FRAME 114, soggetto con ampie parti NM nella mappa di disparità

Figura 5.13: Immagine corrente e mappa di disparità del frame 159



Figura 5.14: Maschera 3D a colori e Output 2D



Figura 5.15: Output 3D e Output 3D selettivo



Figura 5.16: FRAME 159, intensa variazione di illuminazione

Figura 5.17: Immagine corrente e sfondo non aggiornato con l'Histogram Specification del frame 173



Figura 5.18: Maschera 2D e maschera di cambiamento 2D se non avessi usato Histogram Specification

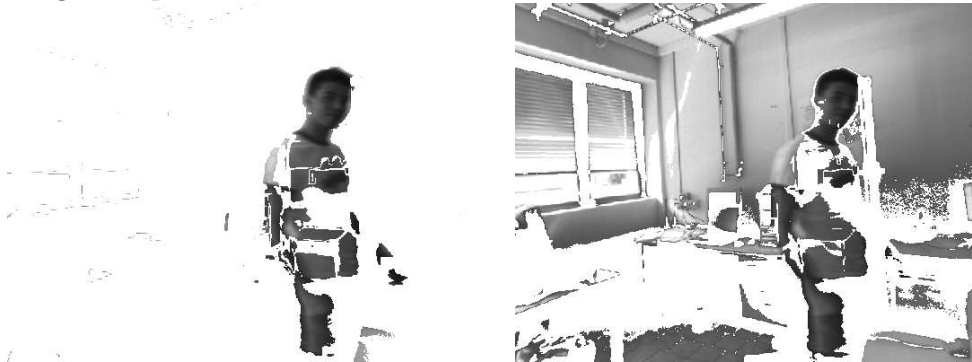


Figura 5.19: Output 3D e Output 3D selettivo



Figura 5.20: FRAME 173, intensa variazione di illuminazione con e senza Histogram Specification

2 Sequenza “Outdoor”

Video acquisito in esterno dal quale risultano mappe di disparità molto dense, con poche zone NM grazie allo sfondo opaco, con elevato contenuto di segnale e illuminazione solare uniforme e stabile. Nella scena i soggetti camminando producono ampie ombre.

Questa sequenza è ambientata in una condizione ideale per il rilevamento corretto delle informazioni 3D, e infatti tra tutte quelle analizzate è quella che dà senza dubbio i migliori risultati.

Le mappe di disparità sono molto stabili, e infatti nello sfondo non vi sono quasi per niente imperfezioni dovute al rumore (Figura 5.23). I soggetti via via presenti nel foreground sono rilevati molto bene dal sistema, nonostante siano presenti varie zone di camouflage nell’output 2D, dovute alle molte tonalità di grigio delle scale sullo sfondo, e nonostante i contorni siano poco definiti in entrambi gli approci 3D (Figura 5.34 immagine di destra, Figura 5.35).

La zona più critica del background per il 3D risulta essere la mensola posizionata ai piedi delle scale, molto liscia e di colore uniforme rispetto al resto della sfondo. Se questa viene coperta da un’ombra mentre output 3D spesso commette un errore rilevandola come cambiata, output 3D selettivo ne viene poco influenzato (Figura 5.31).

Un aspetto molto critico per l’approccio 2D sono le ombre dei soggetti in movimento, che non riescono a essere totalmente filtrate dall’istogram specification. Inoltre l’approccio 2D commette delle false rilevazioni quando un soggetto fuori campo si va e frapporre tra il sole e la scena, generando un’ombra molto estesa in una zona grande circa un quarto della superficie totale dell’immagine (Figura 5.37, Figura 5.38). In questo caso confrontando gli output 2D in presenza e in assenza di registrazione tonale se da una parte notiamo un miglioramento nell’output registrato tonalmente dall’altra, essendo la variazione di illuminazione locale e molto intensa, le modifiche all’istogramma dello sfondo lo vanno a correggere in modo errato. Al contrario in presenza di ombre gli approcci 3D si rivelano molto robusti (Figura 5.39).

Frame 10, Figura 5.24

Le caratteristiche della scena permettono di impostare parametri selettivi nei confronti del rumore senza perdere troppa informazione, e quindi anche col 3D riusciamo a ottenere uno sfondo quasi esente da imperfezioni.

Frame 35, Figura 5.28

Il soggetto estraneo viene correttamente rilevato. Si noti però il leggero camouflage del 2D e i bordi poco precisi dei 3D.

Frame 128, Figura 5.32

In questo esempio la persona viene rilevata con molto camouflage dal 2D, che inoltre considera l'ombra come movimento. Al contrario per entrambi i 3D la sagoma risulta completamente piena, e per di più nel 3D selettivo l'ombra è completamente filtrata.

Frame 211, Figura 5.36

Analogo al precedente frame, vediamo che in questo caso l'approccio 3D selettivo è molto preciso, il 3D normale sbaglia nella zona di NM corrispondente alla mensola e il 2D ha un'output poco soddisfacente.

Frame 181, Figura 5.40

Un soggetto fuori campo si è frapposto tra il sole e la scena: non vi è alcun soggetto estraneo, ma solo una variazione locale di illuminazione. Si noti che nella Figura 5.38, a sinistra, l'Histogram Specification se da un lato minimizza di molto l'ombra, dall'altro è fonte di errore in basso a destra.

Figura 5.21: Immagine corrente e mappa di disparità del frame 10

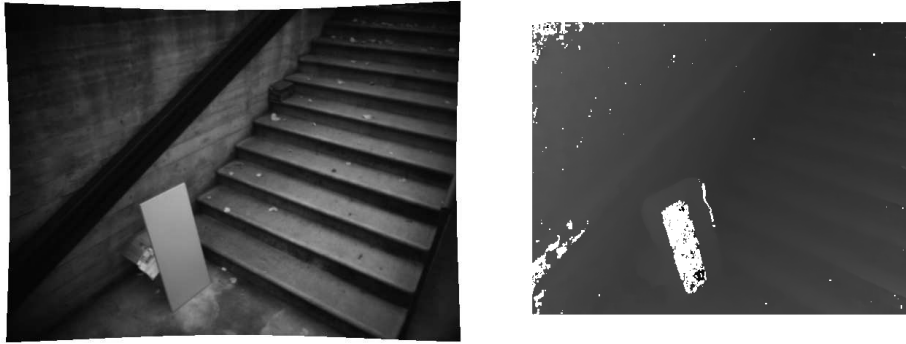


Figura 5.22: Maschera 3D a colori e Output 2D

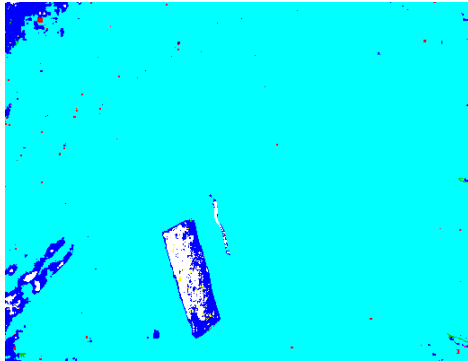


Figura 5.23: Output 3D e Output 3D selettivo

Figura 5.24: FRAME 10, solo sfondo

Figura 5.25: Immagine corrente e mappa di disparità del frame 35



Figura 5.26: Maschera 3D a colori e Output 2D

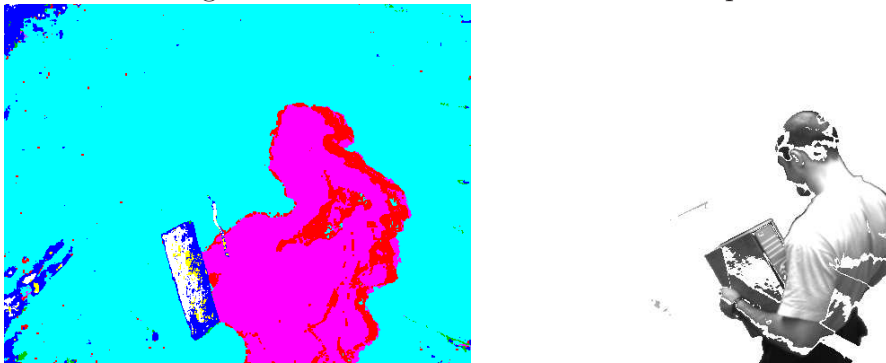


Figura 5.27: Output 3D e Output 3D selettivo



Figura 5.28: FRAME 35, soggetto estraneo

Figura 5.29: Immagine corrente e mappa di disparità del frame 128



Figura 5.30: Maschera 3D a colori e Output 2D

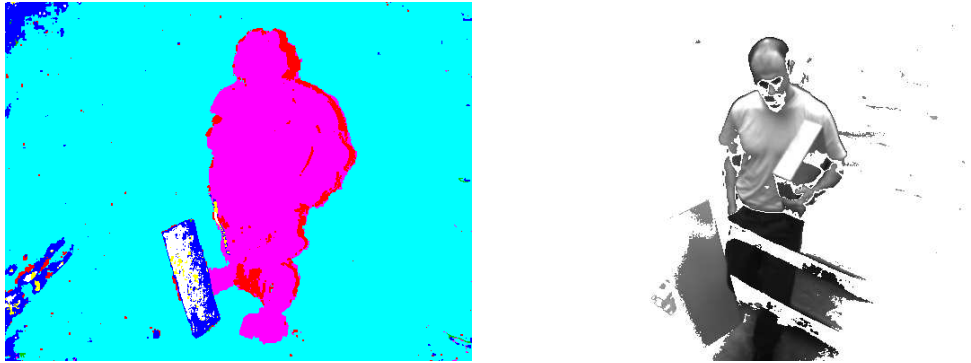


Figura 5.31: Output 3D e Output 3D selettivo



Figura 5.32: FRAME 128, soggetto estraneo con ombra

Figura 5.33: Immagine corrente e mappa di disparità del frame 211

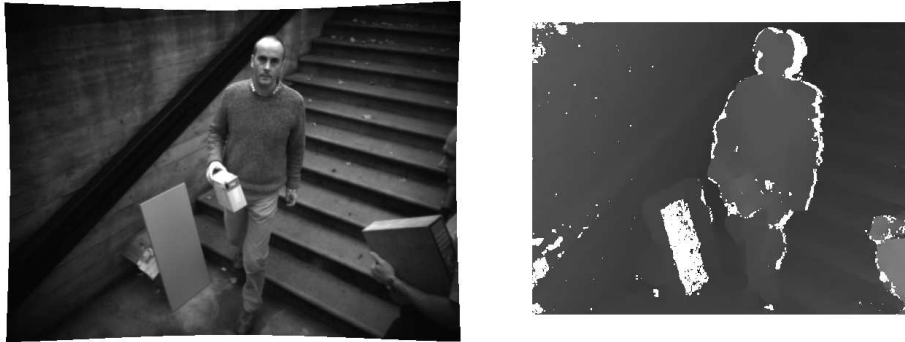


Figura 5.34: Maschera 3D a colori e Output 2D

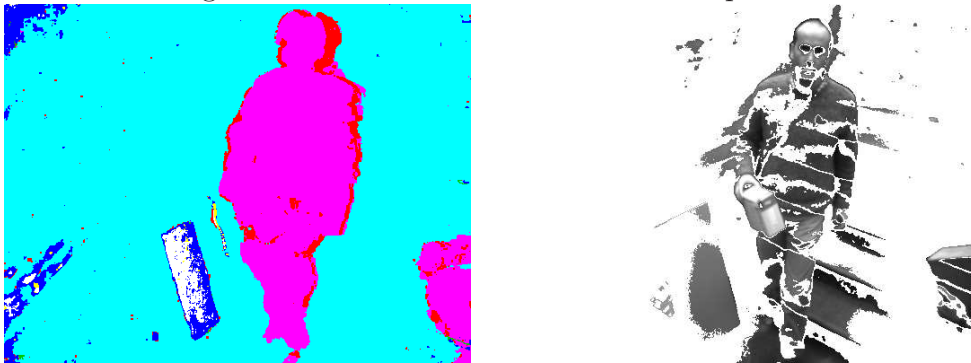


Figura 5.35: Output 3D e Output 3D selettivo



Figura 5.36: FRAME 211, soggetto estraneo con ombra 2

Figura 5.37: Immagine corrente e sfondo non aggiornato del frame 181



Figura 5.38: Maschera 2D e Output 2D



Figura 5.39: Output 3D e Output 3D selettivo

Figura 5.40: FRAME 181, variazione di illuminazione senza alcuna intrusione

3 Sequenza “Ufficio”

La terza ed ultima sequenza è ambientata in un ufficio del DEIS, ed è caratterizzata dalla sola illuminazione elettrica al neon. In questa scena si è cercato di simulare una serie di variazioni di intensità luminosa agendo sulle sorgenti di illuminazione.

Data la bassa intensità di illuminazione della scena risulta molto ridotta la profondità dei colori, e quindi il camouflage risulta molto più accentuato negli output 2D (Figura 5.46 immagine di destra). Inoltre anche la qualità dell'illuminazione, totalmente elettrica al neon, influisce negativamente facendo rilevare nelle mappe di disparità i soggetti in movimento con ampie zone di NM e molti valori di disparità dello sfondo ambigui, ovvero con varianza molto elevata (Figura 5.47).

Per questi motivi si è scelto di utilizzare soglie poco restrittive per non perdere troppi punti validi nelle mappe di disparità, ma con la conseguenza di avere entrambi gli output 3D molto soggetti a rumore, come è ben visibile in Figura 5.43). Comunque in questo caso risulta semplice migliorare l'output con alcune semplici operazioni di morfologia.

In presenza di ombre o variazioni di illuminazione però i risultati non sono del tutto soddisfacenti. Oltre ai noti problemi dell'output 2D, che rileva come cambiate zone dello sfondo solo diversamente illuminate (Figura 5.51 immagine di sinistra), vengono alla luce alcuni difetti del sistema stereo. Avendo abbassato la soglia ci accorgiamo che il valore di disparità delle zone sottoposte a illuminazione maggiormente diretta è cambiato a sufficienza per essere rilevato dalla maschera3D, e di conseguenza segnalato dai rispettivi output 3D (Figura 5.51 immagine di destra, Figura 5.50). Si osserva che questi punti sono in numero limitato, e tipicamente associati ai contorni degli oggetti (regioni caratterizzate da variazioni di disparità), a regioni uniformi o caratterizzate da effetti di riflessione ottica come le finestre sullo sfondo. Inoltre, si osserva che molti di questi punti potrebbero essere localizzati in fase di inizializzazione del sistema e conseguentemente impostati come punti

inaffidabili agendo senza soggetti che non appartengano al background variando l'intensità luminosa entro i livelli ipotizzabili per la scena.

Frame 5, Figura 5.44

Essendo l'illuminazione solo elettrica al neon, peraltro non troppo intensa, nel rilevare i dati stereo il sensore introduce molto rumore, e quindi si sono dovuti usare dei parametri e delle soglie poco restrittive, per non perdere troppa informazione. La conseguenza più evidente è lo sfondo molto sporco, anche in assenza di movimento.

Frame 58, Figura 5.48

É ben visibile il soggetto nella parte centro sinistra dell'immagine, ma con ampie zone di camouflage. Questo problema si ripercuote anche negli output 3D, che non riescono a riempire i buchi NM delle mappe di disparità. In particolare il 3D selettivo se da un lato elimina l'ombra ai piedi del soggetto, ne va però a limare ulteriormente parti corrette della sagoma.

Frame 149, Figura 5.52

Anche in questo caso tutti i tre approcci si rivelano critici. La variazione di illuminazione, peraltro accentuata dal vetro sullo sfondo, non viene eliminata dall'histogram specification, e il camouflage sulla gamba del soggetto non è trascurabile. Anche gli output basati sulle mappe di disparità non sono esenti da imperfezioni, infatti in alcune zone dello sfondo è addirittura cambiato il valore di disparità. In questo caso l'approccio 3D selettivo è il peggiore dei tre, dato che il soggetto si trova frapposto a una zona dello sfondo in gran parte inaffidabile o NM.

Figura 5.41: Immagine corrente e mappa di disparità del frame 5



Figura 5.42: Maschera 3D a colori e Output 2D



Figura 5.43: Output 3D e Output 3D selettivo



Figura 5.44: FRAME 5, solo sfondo

Figura 5.45: Immagine corrente e mappa di disparità del frame 63



Figura 5.46: Maschera 3D a colori e Output 2D

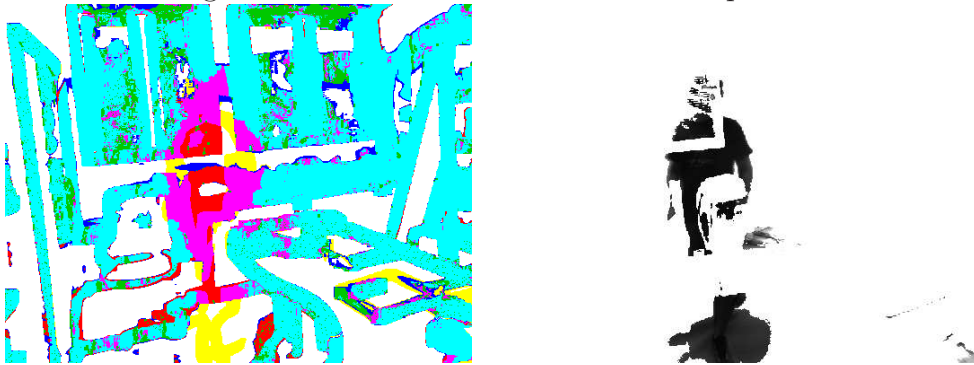


Figura 5.47: Output 3D e Output 3D selettivo



Figura 5.48: FRAME 58, soggetto 2

Figura 5.49: Immagine corrente e mappa di disparità del frame 149



Figura 5.50: Maschera 3D a colori e Output 2D

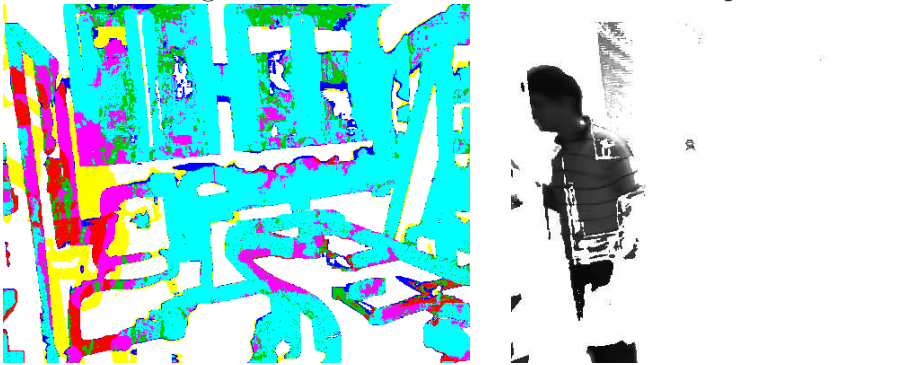


Figura 5.51: Output 3D e Output 3D selettivo



Figura 5.52: FRAME 149, soggetto 2

Capitolo 6

Conclusioni e sviluppi futuri

Sulla base dell'esperienza acquisita nel corso di questa tesi abbiamo visto che ai fini della change detection in presenza di variazioni di illuminazione i dati 3D possiedono delle caratteristiche molto interessanti.

Infatti i dati 2D, che si basano solo sull'intensità di colore dei singoli punti analizzati, in presenza di variazioni di illuminazione spesso interpretano una zona illuminata diversamente come un cambiamento della scena. Contrariamente gli approcci 3D basandosi sulla disparità riducono significativamente questa eventualità.

Ciononostante, le mappe di disparità presentano alcuni problemi intrinseci. Alcuni sono eliminabili facilmente, come il rumore da cui possono spesso essere caratterizzate, e altri solo con particolari accorgimenti. Vi sono infatti le zone di non matching, dove non abbiamo informazione alcuna sulla disparità, ed il passaggio di una di queste zone ad un valore valido di disparità (o viceversa il passaggio di una da disparità valida a non matching) non sempre è dovuto ad un cambiamento della scena.

Output 3D e output 3D selettivo sono due approcci differenti nati dall'esigenza di riempire le zone non matching, dove non abbiamo alcuna informazione 3D. Da un lato il secondo si è rivelato senza dubbio il migliore in presenza di mappe di disparità molto dense, perchè considera come potenzialmente variati, e quindi gli attribuisce il valore dell'output 2D, solo i punti che hanno cambiato la propria condizione da validi a NM e viceversa. Dall'altro se queste sono poco dense e molto rumorose output 3D si è dimostrato molto

più efficace, in quanto considera come potenzialmente cambiati più punti di output 3D selettivo, anche quelli che mantengono una condizione di NM o sono punti dello sfondo considerati non affidabili.

In situazioni senza particolari variazioni di illuminazione l'approccio migliore si è dimostrato il 2D, che a parte alcune zone caratterizzate dal fenomeno del camouflage, è molto precisa nel rilevare sagome estranee allo sfondo e non è quasi per niente soggetta al rumore. Dobbiamo però sempre ricordare che uno dei primi passaggi per generare l'output 2D consiste nella registrazione tonale, che ha come supporto i punti individuati come sicuramente fermi dalla maschera in uscita alla background subtraction 3D.

Per gli sviluppi futuri si potrà sicuramente cercare di dare una veste grafica all'algoritmo, in modo da rendere molto più semplice il suo utilizzo da parte di qualsiasi utente anche in tempo reale, collegandosi ad una telecamera stereo e andando ad analizzare scene run time. Inoltre a mio avviso sarebbe molto interessante, dato il costo esoso delle moderne telecamere stereo, tentare di fissare e calibrare opportunamente due telecamere mono, anche due webcam, in modo da ottenere un sensore 3D a basso costo, e in seguito dare le immagini e le mappe in loro uscita come input all'intero sistema.

Per concludere uno sviluppo futuro sarà sicuramente l'utilizzo di algoritmi differenti da SMP per il matching stereo, in particolar modo di tipo semi-globale. Questi generano mappe di disparità molto più dense e precise degli algoritmi locali anche nei contorni degli oggetti, al costo di una più lunga e complessa elaborazione software. Dati questi presupposti non potranno che valorizzare i pregi degli approcci 3D.

Inoltre anche l'utilizzo di telecamere a colori potrà sicuramente portare molti vantaggi a livello di change detection, e sarà prossimamente testato.

Bibliografia

- [1] L. Baldini (2007): *Sviluppo di un algoritmo di change detection basato su matching*
- [2] R.C. Gonzales, R.E. Woods (1993): *Digital image processing* ed. Addison Wesley
- [3] S. Mattocchia : *Introduzione alla visione stereo*
www.vision.deis.unibo.it/smatt
- [4] K. Gualandris (2007): *Tesi di laurea*
- [5] L. Di Stefano, M. Marchionni, S. Mattocchia (2004) *A Fast Area-Based Stereo Matching Algorithm* Image and Vision Computing, Vol. 22, No. 12, pp 983-1005, Elsevier
- [6] Sito web di Stefano Mattocchia presso Labvisione, università di Bologna
<http://labvisione.deis.unibo.it/smattocchia/stereo.htm>
- [7] D. Scharstein and R. Szelinsky.(2001) *A Taxonomy and Evaluation of Dense Two-Frame Stereo Correspondence Algorithms*
<http://cat.middlebury.edu/stereo>
- [8] www.sourceforge.net