Università degli Studi di Padova

Dipartimento di Ingegneria dell'Informazione

Corso di Laurea Magistrale in Ingegneria Informatica

RICOSTRUZIONE 3D CON SENSORE A TEMPO DI VOLO E TELECAMERE STEREO

Relatore

Prof. Pietro Zanuttigh

Laureando

Arrigo Guizzo

Correlatore

Ing. Carlo Dal Mutto

Anno Accademico 2011-2012

Ai miei genitori,
che mi hanno sempre sostenuto.

Sommario

Questo lavoro di tesi verte alla realizzazione di un software per la ricostruzione 3D di una scena utilizzando tipi diversi di sensori. Nel caso specifico, il sistema di acquisizione è composto da una coppia di telecamere stereo e un sensore a tempo di volo. Lo scopo è fornire una ricostruzione più precisa combinando l'informazione data dai singoli sensori, cercando di sfruttare i loro vantaggi e attenuando il più possibile gli svantaggi. La scelta dei sensori gioca quindi un ruolo fondamentale: in base alle loro caratteristiche, le telecamere stereo e il sensore a tempo di volo tendono ad essere complementari, dimostrandosi una buona combinazione. Il framework di ricostruzione 3D sviluppato può essere esteso facilmente a più di due sensori.

La tesi è strutturata nei seguenti capitoli fondamentali: la descrizione della tecnologia dei sensori, compresi i loro vantaggi e svantaggi; la descrizione dei metodi di ricostruzione 3D associati ai due sensori con la relativa stima di affidabilità; l'unione delle informazioni basandosi sulla stima dell'affidabilità; infine, viene presentata l'analisi dei risultati ottenuti alla fine del progetto.

Indice

1 Introduzione **1**
- 1.1 Descrizione del progetto . . . 1
- 1.2 Sistema di acquisizione . . . 2
  - 1.2.1 Tecnologia Time-of-flight . . . 2
    - 1.2.1.1 Analisi della tecnologia . . . 3
  - 1.2.2 Sistema stereo . . . 4
  - 1.2.3 Analisi del sistema completo . . . 5
- 1.3 Letteratura sull'argomento . . . 5

2 Calcolo della disparità **7**
- 2.1 Disparità associata al sensore TOF . . . 8
- 2.2 Disparità stereo . . . 11
  - 2.2.1 Strategia di aggregazione dei costi . . . 12
  - 2.2.2 Calcolo della disparità . . . 15
  - 2.2.3 Implementazione utilizzata . . . 16
- 2.3 Stima dell'errore sulla disparità . . . 17
  - 2.3.1 Stima dell'errore per il TOF . . . 17
    - 2.3.1.1 Deviazione standard della profondità . . . 18
    - 2.3.1.2 Incertezza basata sull'ampiezza . . . 19
    - 2.3.1.3 Combinazione delle due stime di errore . . . 20
  - 2.3.2 Stima dell'errore per lo stereo . . . 20

3 Fusione delle ipotesi **23**
- 3.1 Tecnica Locally Consistent (LC) . . . 23
  - 3.1.1 Plausibilità della corrispondenza tra due punti . . . 24
  - 3.1.2 Calcolo della plausibilità per ogni disparità . . . 27
  - 3.1.3 Calcolo della disparità . . . 29
  - 3.1.4 Considerazioni sull'algoritmo . . . 29
- 3.2 Fusione pesata tramite LC . . . 30
  - 3.2.1 Plausibilità nuova per le telecamere stereo . . . 31
  - 3.2.2 Plausibilità nuova per il sensore TOF . . . 31

3.2.3 Implementazione realizzata 32

4 Analisi dei risultati **35**
4.1 Acquisizione dei dataset e parametri di test 35
4.2 Valutazione delle disparità singolarmente 38
4.2.1 Valutazione disparità associata al TOF 38
4.2.2 Valutazione disparità stereo 40
4.2.3 Osservazioni 43
4.3 Valutazione delle disparità combinate 44
4.3.1 Valutazione senza stime di errore 44
4.3.1.1 Osservazioni 48
4.3.2 Valutazione utilizzando le stime di errore 51
4.3.2.1 Stime di errore 51
4.3.2.2 Errore gaussiano ed errore esponenziale 54
4.3.2.3 Errori esponenziali 57
4.3.2.4 Considerazioni 60

5 Conclusioni **61**

Capitolo 1

Introduzione

Negli ultimi due decenni, la tecnologia ha subito un'evoluzione con ritmo esponenziale. In particolare, hardware per la visualizzazione grafica più performante e nuovi sensori per l'acquisizione di scene del mondo reale sono stati introdotti nel mercato. Inoltre, grazie alla sensibile riduzione dei costi, il bacino di utenza raggiunta è stato ampliato. Molti ambiti, compresa la computer grafica e i suoi derivati, sono stati giovati da questa evoluzione.

Unitamente al desiderio di virtualizzare la realtà, la modellazione 3D di oggetti ha conosciuto una larga diffusione: l'interesse principale è poter comunicare tutta l'informazione visiva data sia dalle geometrie che dal colore in digitale, generando quindi una copia più precisa possibile della realtà.

Questo lavoro di tesi tratta in particolare la ricostruzione tridimensionale di oggetti, a partire da una scena acquisita con due sensori di tipi diversi: un sensore a tempo di volo e due telecamere a colori in coppia stereo.

Un sistema di acquisizione di scene tridimensionali può essere utilizzato in vari ambiti: intrattenimento, robotica, architettura, arte e istruzione sono soltanto alcuni. In tutti questi, la ricostruzione tridimensionale di oggetti o dell'ambiente circostante permette di creare scene virtuali più reali possibili, di rilevare ostacoli e permettere la navigazione autonoma, di ottenere modelli di edifici o infrastrutture, di digitalizzare un'opera d'arte.

1.1 Descrizione del progetto

L'obiettivo di questo lavoro è unire le informazioni tridimensionali fornite dai sensori sfruttando la complementarietà delle loro caratteristiche. Pertanto, le prime due fasi saranno l'implementazione di due metodi di ricostruzione 3D: uno associato al sensore a tempo di volo, l'altro alle telecamere stereo.

Queste due fasi possono essere eseguite in parallelo: i due sensori infatti sono indipendenti e forniscono in parallelo i dati della scena acquisita.

La terza fase consiste nella fusione delle informazioni: una tecnica esistente è stata estesa per tenere conto della presenza di due sensori.

1.2 Sistema di acquisizione

Il sistema è composto da un sensore a tempo di volo (*Time-of-Flight* e due telecamere stereo, per brevità "TOF"). Di seguito verrà descritto il secondo componente, molto meno diffuso delle telecamere tradizionali, e viene fatta una breve analisi sui vantaggi e svantaggi di entrambi i sensori.

1.2.1 Tecnologia Time-of-flight

I sensori TOF utilizzano una tecnologia esistente da vari anni, tuttavia si sono diffusi nel mercato consumer soltanto negli ultimi tempi dopo una sensibile riduzione dei costi di produzione.

Il principio di questa tecnologia è piuttosto semplice: viene misurato il tempo che occorre ad un impulso luminoso per viaggiare da una sorgente luminosa ad un oggetto e ritornare al sensore[1], ovvero viene misurato il tempo per percorrere un tragitto pari al doppio della distanza dell'oggetto dalla telecamera. La misurazione viene eseguita in maniera indipendente per ogni pixel della telecamera, permettendo di acquisire interamente la scena inquadrata.

Il sensore TOF utilizzato in questo lavoro di tesi è lo SwissRanger SR4000, presente in figura 1.1. Gli impulsi luminosi sono segnali infrarossi inviati tramite LED e il ricevitore è una matrice di sensori CCD/CMOS. Le specifiche tecniche sono descritte dettagliatamente nel datasheet disponibile nel sito [3]; si riportano solo la distanza massima misurata di $5\,m$, la risoluzione di 176×144 pixel e il framerate di 50 frame al secondo.

Lo SwissRanger SR4000 fornisce in output quattro immagini tutte della stessa risoluzione:

- la mappa di profondità con valori espressi in metri (valori a 14 bit; l'incremento di un bit corrisponde a $0.305\,mm$ se la distanza massima è di $5\,m$);
- l'immagine di ampiezza, che contiene l'ampiezza del segnale riflesso per ogni pixel;

[1] La misura del tempo viene di fatto eseguita tramite la differenza di fase tra il segnale di invio e il segnale riflesso dall'oggetto.

Figura 1.1: SwissRanger SR4000 (immagine fornita da [3]).

- l'immagine di profondità in scala di grigi, dove la profondità viene mappata in valori a 8 bit;
- la mappa di confidenza, che indica l'attendibilità della misura per ogni pixel espressa in valori a 8bit.

Di seguito sono elencati vantaggi e limiti della tecnologia.

1.2.1.1 Analisi della tecnologia

Come probabilmente è stato intuito nella descrizione, un sensore TOF garantisce alcuni vantaggi.

- È un sistema compatto che non richiede particolari operazioni di installazione, al contrario di altri sistemi con maggiore accuratezza come gli scanner laser.
- Fornisce direttamente una mappa di profondità, al contrario di telecamere stereo per le quali devono essere eseguiti algoritmi particolari per il calcolo della disparità e la triangolazione.
- Effettua misurazioni di tutta la scena inquadrata in tempo reale; esegue tutti i calcoli utilizzando il microprocessore interno in maniera molto rapida ed efficiente, dando la possibilità di restituire solo i risultati tramite interfacce diffuse (USB o Fast Ethernet). Non richiede pertanto particolari requisiti per il computer a cui è collegato.
- È una tecnologia più robusta ai cambiamenti di luce rispetto ad altri sensori come le telecamere stereo.

Tuttavia, la tecnologia comporta anche alcuni limiti.

- La scarsa risoluzione permette di ricavare un'informazione limitata sulla geometria della scena e sulle superfici presenti.
- Per materiali poco riflettenti (stoffa) o di colore scuro non è possibile ottenere misure accurate causa il debole segnale di ritorno.
- Si possono verificare sbilanciamenti nelle rilevazioni causa errate disposizioni del sensore o angoli di inquadratura. Il "multipath error" è un esempio (figura: 1.2): nel caso di una scena con geometrie concave (due parenti incidenti), il segnale luminoso colpisce una parete, ma il segnale riflesso colpisce la seconda parete prima di essere diretto verso il sensore. In questo caso, la distanza della parete rilevata risulta maggiore rispetto a quella reale.
- La rilevazione della distanza lungo i bordi di superfici risulta imprecisa. In questi tratti il segnale luminoso viene riflesso in più direzioni e il ricevitore rileva solo una parte del segnale di ritorno; questa difficoltà fisica unita alla risoluzione limitata, costringe il sensore TOF a generare un valore di profondità unico per più particolari dell'immagine.
- Sorgenti luminose esterne come il sole possono interferire con il segnale del sensore, generando errori in fase di ricezione.

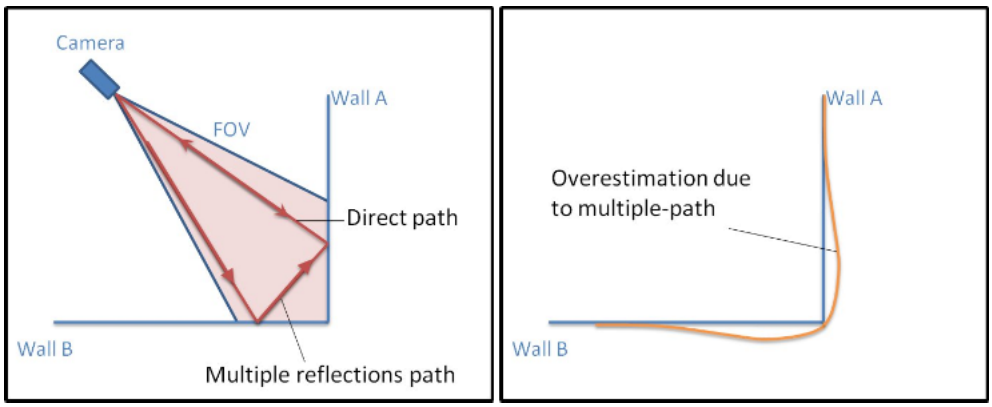


Figura 1.2: Esempio di multipath error, detto anche scattering (immagine fornita dal manuale del sensore; si veda [3]).

1.2.2 Sistema stereo

Il sistema stereo è composto da due telecamere Basler scA1000-30fc a colori con interfaccia FireWire-b. La risoluzione massima è di 1032×778 pixel e il

framerate massimo dichiarato è di 30 frame al secondo. Si rimanda al sito del costruttore per maggiori dettagli tecnici [2].

Figura 1.3: Basler scA1000-30fc (immagine fornita da [2]).

Come già ben noto, i principali vantaggi dati dall'uso di telecamere nell'ambito della visione computazionale sono l'informazione sul colore e l'alta risoluzione delle immagini.

Dal punto di vista dell'informazione 3D, la ricostruzione di scene tramite sistemi stereo è stata molto studiata e la qualità dei risultati ottenuti è molto migliorata. Tuttavia, rimangono evidenti alcune difficoltà nell'analisi di scene con scarsa texture o con pattern ripetitivi, che delineano i principali svantaggi di questi sistemi: queste scene infatti non possono essere gestite analizzando soltanto l'informazione sul colore, la quale non è discriminante per distinguere i punti corrispondenti.

1.2.3 Analisi del sistema completo

La combinazione di due tipi di sensori comporta una situazione interessante:

- La difficoltà della ricostruzione 3D in scene con scarsa texture delle telecamere viene compensata dalla maggiore robustezza del sensore TOF a queste situazioni.
- La risoluzione limitata del sensore TOF viene integrata con l'informazione fornita dalle immagini ad alta risoluzione delle telecamere.

Da queste prime considerazioni, i due sensori bilanciano in parte i loro difetti e garantiscono i rispettivi vantaggi singolarmente. Pertanto, ci si aspetta una ricostruzione 3D più precisa utilizzando assieme i due sensori.

1.3 Letteratura sull'argomento

La letteratura sull'argomento è attualmente in espansione. I sensori TOF hanno raggiunto il mercato da pochi anni e, nonostante la limitata risoluzione e la presenza di rumore, sono strumenti molto utilizzati nella ricerca.

I principali lavori che trattano l'utilizzo di più sensori per ottenere ricostruzioni 3D più accurate utilizzano gli stessi sensori per comporre il sistema di acquisizione e si possono dividere in due gruppi: migliorare la risoluzione delle mappe di profondità del sensore a tempo di volo utilizzando l'informazione del colore fornita da una telecamera oppure fondere l'informazione 3D ricavabile da un sistema stereo e il sensore TOF, come il caso in esame.

Del primo gruppo, [8, 10, 21] sono tre ottimi esempi: il primo utilizza un approccio basato su Markov Random Field che comprende le informazioni di profondità e del colore, il secondo si basa sulla segmentazione dell'immagine a colori, il terzo utilizza un approccio iterativo di raffinamento di una distribuzione 3D di probabilità utilizzando l'interpolazione bilaterale.

Del secondo gruppo, la fusione di informazioni 3D fornite dai due sensori richiede dati di input della stessa risoluzione e fanno spesso uso dei risultati del primo gruppo di lavori per ottenere mappe di disparità ad alta risoluzione associate al sensore TOF. In [17] viene proposta una tecnica semplice e real-time che consiste nel mediare le ipotesi di disparità per ogni pixel. Tecniche più ponderate e precise sono presenti in [11, 20, 23].

Capitolo 2

Calcolo della disparità

La mappa di disparità è uno degli strumenti più diffusi per la ricostruzione 3D di scene. Si basa sul concetto di punti coniugati (o corrispondenti), ovvero due punti presenti in due immagini diverse che sono la proiezione dello stesso punto della scena inquadrata. Immaginando di sovrapporre le due immagini, si definisce disparità il vettore che rappresenta la differenza di posizione dei due punti coniugati nelle due immagini. Si tratta quindi di un concetto legato ad un sistema di visione stereo. La ricerca di punti coniugati (o il calcolo della disparità) permette di eseguire la triangolazione stereoscopica e quindi stimare la profondità di ogni punto della scena inquadrata.

Il calcolo della disparità è eseguito prendendo come riferimento una delle due immagini (immagine base) e cercando i punti corrispondenti nell'altra immagine (immagine match). Grazie alla geometria epipolare, è noto che due punti coniugati giacciono sullo stesso piano individuato dai punti considerati e dai centri ottici delle due telecamere. Pertanto, è sufficiente analizzare l'intersezione del piano epipolare con l'immagine match: questa intersezione viene chiamata linea epipolare. Tuttavia, determinare questa intersezione aggiunge difficoltà al calcolo della disparità: pertanto, le immagini vengono entrambe rettificate. Questa operazione proietta entrambe le immagini nello stesso piano retina, il quale è parallelo alla baseline del sistema stereo (si veda figura 2.1). Inoltre, dopo questa trasformazione, le linee epipolari delle due immagini diventano parallele e orizzontali, semplificando il calcolo della disparità: considerando le immagini come matrici di pixel, punti corrispondenti in due immagini rettificate giacciono nella stessa riga delle immagini.

A questo punto è possibile eseguire la triangolazione stereoscopica per stimare la profondità z di ogni punto. La rettificazione delle immagini ha trasformato il sistema stereo come quello in figura 2.1: dalle equazioni prospettiche, si ottiene la seguente relazione tra la profondità z del punto

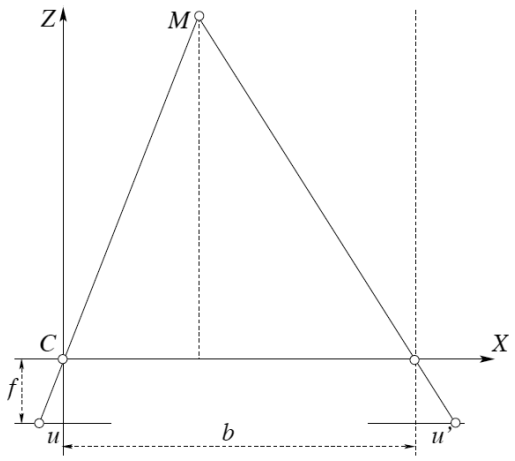


Figura 2.1: Triangolazione per un sistema stereo rettificato (immagine fornita da [9]). b è la baseline del sistema stereo, f la lunghezza focale delle camere, M è il punto reale, u e u' sono le proiezioni del punto reale nelle due immagini rispettivamente.

reale e la disparità d associata ai punti u e u' nelle due immagini rettificate:

$$z = \frac{b \cdot f}{u - u'} = \frac{b \cdot f}{d} \tag{2.1}$$

dove b è la baseline dello stereo e f la lunghezza focale delle camere (per i dettagli, si veda [9]). La disparità quindi descrive una scena 3D allo stesso modo della coordinata z, una volta noti i parametri del sistema stereo.

Come già accennato, in questo lavoro è stato utilizzato un sistema di due telecamere stereo combinato con un sensore *Time-of-Flight*, sensori che rappresentano la scena 3D in maniera differente. La disparità è stata scelta come unità di misura comune per confrontare le informazioni fornite e per applicare una strategia di combinazione delle informazioni (si veda il capitolo 3), al fine di ottenere una migliore ricostruzione 3D della scena. Pertanto, il calcolo della mappa di disparità risulta essere una delle parti centrali di questo lavoro.

2.1 Disparità associata al sensore TOF

Il sensore Time-of-Flight restituisce una mappa di profondità della scena, fornendo una prima ricostruzione 3D. Questa mappa potrebbe essere uti-

lizzata subito per calcolare la disparità da combinare con la disparità data dall'algoritmo stereo, tuttavia le diverse risoluzioni rendono improbabile tale scenario: la mappa di profondità è di dimensione 176×144 pixel, mentre la disparità data da un algoritmo stereo ha la stessa (alta) risoluzione delle immagini, ovvero 1032×778 pixel. Per questo motivo, è stato utilizzato un procedimento composto da tre fasi per ottenere una mappa di profondità ad alta risoluzione del sensore, dalla quale ottenere la mappa di disparità associata al sensore.

La prima fase consiste nel correggere la distorsione radiale dei punti della mappa di profondità e proiettarli nel piano immagine di una delle due telecamere stereo; senza perdita di generalità, si assuma che sia stata scelta la telecamera sinistra. La proiezione viene eseguita utilizzando la matrice dei parametri intrinseci del sensore TOF, la matrice di rototraslazione dal sistema di riferimento 3D del sensore TOF alla telecamera sinistra e la matrice dei parametri intrinseci della telecamera, tutti parametri calcolati tramite la calibrazione del sistema.

La seconda fase consiste nell'eliminare i punti occlusi: il sensore TOF e la telecamera sinistra vedono la stessa scena da punti di vista simili ma non uguali, pertanto non tutti i punti visti dal sensore TOF saranno visibili anche per la telecamera. È stato utilizzato un algoritmo basato sul calcolo di primitive triangolari costruite sui punti proiettati, al fine di quantificare la profondità di tutti i punti del piano immagine della telecamera. Di fatto, questo calcolo è stato eseguito tramite un algoritmo scanline per il filling di poligoni (filling di triangoli, nel caso in esame); contestualmente, al termine di ogni scanline è stato applicato il principio dell'algoritmo z-buffer per determinare le primitive visibili: la primitiva con profondità minima calcolata al passo precedente viene considerata visibile.

Siano P_i^T i punti 3D del sensore TOF, n il numero di punti del TOF non occlusi e proiettati, N il numero di punti del lattice Λ_l della telecamera sinistra. I punti P_i^T vengono proiettati nei punti p_i sul lattice Λ_l, occupando solo un piccolo sottoinsieme dei punti disponibili. Infatti, il rapporto tra le risoluzioni del sensore TOF e delle telecamere è pari a circa il 3%.

La terza e ultima fase consiste nella creazione della mappa di profondità ad alta risoluzione utilizzando i punti proiettati per interpolare tutti i valori. Due tra le tecniche più efficaci sono basate sulla segmentazione [10] e sull'interpolazione bilaterale congiunta [13]. La tecnica utilizzata è una combinazione delle due: combinando l'informazione data dal colore con i due filtri congiunti

dell'interpolazione bilaterale, viene fornito un ulteriore strumento per pesare il contributo dei punti nel calcolo della profondità. Infatti, i punti facenti parte della stessa regione di segmentazione contribuiranno in maniera maggiore rispetto a quelli che non ne fanno parte.

Nel dettaglio, viene innanzitutto segmentata l'immagine della telecamera sinistra secondo il metodo basato sull'algoritmo mean-shift descritto in [6]. Il risultato è una mappa che assegna ogni punto p_j del lattice Λ_l ad una regione di segmentazione $S(p_j)$. Dopodiché, viene eseguita l'interpolazione bilaterale: in ogni punto p_j del lattice viene centrata una finestra W_j di dimensione $w \times w$ contenente i punti $p_{j,k}$, $k = 1, \ldots, w^2$; tale finestra conterrà un sottoinsieme $W'_j \subset W_j$ di punti proiettati $p_{i,k}$ aventi un valore di profondità z_i, e questi ultimi saranno combinati per ottenere il valore di profondità per il punto p_j. In [13], la combinazione viene eseguita utilizzando due funzioni per pesare il contributo di ogni punto: una funzione basata sulla distanza tra il punto centrale e il punto contribuente e una funzione basata sui valori dei punti presenti nella finestra. La funzione spaziale $f_s(p_{i,k}, p_j)$ è stata modellata come una funzione Gaussiana standard in due dimensioni, con parametro la distanza euclidea tra i due punti considerati; la seconda funzione, indicata con $f_c(p_{i,k}, p_j)$, è stata anch'essa modellata come una funzione Gaussiana standard, con parametro la distanza euclidea tra i colori dei due punti considerati. L'informazione fornita dalla segmentazione viene codificata con la funzione I_{segm}:

$$I_{segm}(p_{i,k}, p_j) = \begin{cases} 1 & \text{se } S(p_{i,k}) = S(p_j) \\ 0 & \text{se } S(p_{i,k}) \neq S(p_j) \end{cases} \tag{2.2}$$

Combinando queste definizioni è possibile calcolare il valore di profondità interpolato:

$$\tilde{z}_j = \sum_{p_{i,k} \in W'_j} [f_s(p_{i,k}, p_j) \cdot I_{segm}(p_{i,k}, p_j) \cdot z_{i,k} + f_s(p_{i,k}, p_j) \cdot (1 - I_{segm}(p_{i,k}, p_j)) \cdot f_c(p_{i,k}, p_j) \cdot z_{i,k}] \tag{2.3}$$

I valori delle funzioni f_s, f_c e I_{segm} sono sempre compresi nell'intervallo [0, 1], garantendo la correttezza della definizione. Osservando la formula, la funzione spaziale viene sempre utilizzata, mentre il valore di $f_c(p_{i,k}, p_j)$ viene utilizzato soltanto se i due punti non appartengono alla stessa regione di

segmentazione. L'effetto della segmentazione è quindi assegnare il valore 1 a $f_c(p_{i,k}, p_j)$ quando $p_{i,k}$ appartiene alla stessa regione di segmentazione di p_j.

Per questi motivi, la tecnica di interpolazione combinata soffre in misura minore degli svantaggi dei metodi:

- nella segmentazione dell'immagine era possibile la formazione di artefatti, i quali minavano la precisione del risultato interpolato; tuttavia, questi artefatti si trovano spesso lungo i bordi degli oggetti, pertanto utilizzando il filtro spaziale dell'interpolazione è possibile ridurre il contributo dato da questi artefatti;
- nell'interpolazione bilaterale congiunta, i bordi nel risultato erano meno definiti, poiché i contributi esterni erano troppo elevati nonostante i due filtri utilizzati; in questo caso, viene aumentato il contributo dei punti vicini a quello interpolato, con una conseguente minore perturbazione dei contributi errati.

Una volta ottenuta la mappa di profondità interpolata, la disparità viene calcolata tramite la relazione vista in 2.1. Questa trasformazione non è eseguita nell'ambito di un sistema di visione stereo, tuttavia risulta comunque un'operazione corretta: la proiezione dei punti dal piano immagine del sensore TOF al piano immagine di una telecamera ha fornito per ogni punto della telecamera un'ipotesi di profondità z, la quale può benissimo essere tradotta in disparità tra due punti nelle immagini.

2.2 Disparità stereo

L'algoritmo utilizzato in questo lavoro è il Semi-Global Matching (SGM) [12]. Si tratta di un metodo locale basato su scanline e, tra gli algoritmi dello stesso tipo, risulta essere tra i più veloci e performanti. È bene sottolineare che un qualsiasi algoritmo stereo denso potrebbe essere utilizzato al posto di quello scelto: infatti, il sistema di ricostruzione sviluppato è indipendente dall'algoritmo utilizzato.

Secondo la review di Scharstein e Szeliski [18], quattro blocchi strutturali sono ricorrenti nella maggior parte degli algoritmi stereo: calcolo delle metriche di accoppiamento delle corrispondenze (spesso definiti in termini di costi delle

corrispondenze), strategia di aggregazione dei costi, calcolo della disparità, raffinamento della disparità. L'algoritmo SGM è stato definito utilizzando tutti e quattro i blocchi nominati; di questi, il calcolo dei costi risulta indipendente: potrebbe essere scelta una qualsiasi metrica di accoppiamento. La sua scelta è ovviamente fondamentale, dato che su di essa si basa l'intero algoritmo.

Nei paragrafi successivi sono descritte le caratteristiche principali dell'algoritmo con riferimento ai blocchi strutturali ricorrenti. Sarà considerata una coppia di immagini stereo rettificate; con immagine match si indicherà l'immagine in cui si cercano corrispondenze per i punti dell'immagine base. Con **p** si indicherà un punto dell'immagine base, mentre con **q** un punto dell'immagine match.

2.2.1 Strategia di aggregazione dei costi

La strategia di aggregazione dei costi è la caratteristica principale dell'algoritmo. La sua formulazione si basa sulla seguente definizione di energia $E(D)$ della mappa di disparità D:

$$E(D) = \sum_{\mathbf{p} \in D} \left(C\left(\mathbf{p},\, D_{\mathbf{p}}\right) + \sum_{\mathbf{p}' \in N_{\mathbf{p}}} P_1 T\left[|D_{\mathbf{p}} - D_{\mathbf{p}'}| = 1\right] + \sum_{\mathbf{p}' \in N_{\mathbf{p}}} P_2 T\left[|D_{\mathbf{p}} - D_{\mathbf{p}'}| > 1\right] \right) \tag{2.4}$$

Tale funzione è il risultato della somma, per ogni punto, di tre termini: il primo è il costo della disparità $D_{\mathbf{p}}$ per il punto **p**; il secondo termine ha lo scopo di penalizzare tutti i lievi cambiamenti di disparità all'interno della regione locale (o vicinato) $N_{\mathbf{p}}$ del punto **p**; il terzo termine invece penalizza i maggiori cambiamenti di disparità all'interno della stessa regione $N_{\mathbf{p}}$. Questa definizione è congrua con l'ipotesi di superfici lisce a tratti (*piecewise smooth*): la penalità ridotta P_1 è presente per tenere conto di superfici curve o non perfettamente piane, mentre la penalità elevata P_2 è presente per preservare le discontinuità. Per il significato delle penalità, deve essere sempre verificata la disuguaglianza $P_2 \geq P_1$, altrimenti la definizione di energia perde di validità.

Il calcolo della disparità viene quindi definito come determinare la mappa di disparità D che minimizza l'espressione 2.4, ovvero una definizione del problema tipica di un algoritmo globale. Per molte funzioni che preservano le discontinuità, come quella in esame, la minimizzazione su tutta l'immagine (in 2D) è un problema $\mathcal{NP}$-completo [19]. Al contrario, la minimizzazione lungo una dimensione dell'immagine (solitamente le righe) basata sulla programmazione dinamica viene risolta efficientemente in tempo polinomiale: tale approccio esegue una strategia di aggregazione dei costi provenienti da due orientamenti (gli unici due possibili all'interno di una riga dell'immagine), risultando molto vincolato all'unica dimensione considerata. Infatti, questa soluzione causa spesso evidenti striature, con conseguenti errori nella disparità.

Per sfruttare la velocità della programmazione dinamica e raggiungere una precisione simile ad un algoritmo globale, l'idea dell'autore è stata estendere il numero di orientamenti coinvolti nella strategia di aggregazione a tutti gli orientamenti possibili, eseguendo ad una minimizzazione locale in due dimensioni: il costo $S(\mathbf{p}, d)$ della disparità d per il punto $\mathbf{p}$ è la somma dei costi $L_r(\mathbf{p}, d)$ dei cammini in una dimensione di costo minimo lungo una direzione $\mathbf{r}$ che terminano in $\mathbf{p}$.

Il comportamento della funzione $E(D)$ viene modellato all'interno del costo dei cammini in una dimensione, definendolo come la somma del costo della disparità d per il punto $\mathbf{p}$ e del costo minimo del cammino che termina in $\mathbf{p} - \mathbf{r}$, includendo anche le penalità in maniera opportuna:

$$\begin{aligned} L'_{\mathbf{r}}(\mathbf{p}, d) = C(\mathbf{p}, d) + \min(&L'_{\mathbf{r}}(\mathbf{p} - \mathbf{r}, d), \\ &L'_{\mathbf{r}}(\mathbf{p} - \mathbf{r}, d-1) + P_1, \\ &L'_{\mathbf{r}}(\mathbf{p} - \mathbf{r}, d+1) + P_1, \\ &\min_i L'_{\mathbf{r}}(\mathbf{p} - \mathbf{r}, i) + P_2) \end{aligned} \tag{2.5}$$

In analogia con l'espressione 2.4, all'interno della funzione di minimo notiamo i due termini con penalità P_1 associati ai cammini con un lieve cambiamento di disparità (in aumento o in diminuzione rispettivamente) e il termine con penalità P_2 associato a tutti i cammini con cambiamenti significativi di disparità. Trattandosi di una somma di termini sempre positivi, il valore del costo di un cammino può crescere senza limite; tuttavia, sottraendo un termine costante (il più elevato possibile), l'aumento del valore viene limitato senza cambiare la posizione del minimo ricercato. Il costo minimo tra

i cammini che terminano in $\mathbf{p} - \mathbf{r}$ (al variare quindi della disparità) possiede le caratteristiche desiderate: è costante per tutte le ipotesi disparità e, nel peggiore dei casi, il costo aumenta di P_2 limitando in modo superiore il costo del cammino minimo $L_{\mathbf{r}} \leq C_{max} + P_2$, dove C_{max} è il massimo costo calcolato. L'espressione 2.5 diventa quindi:

$$\begin{aligned} L_{\mathbf{r}}(\mathbf{p}, d) = C(\mathbf{p}, d) + \min(& L_{\mathbf{r}}(\mathbf{p} - \mathbf{r}, d), \\ & L_{\mathbf{r}}(\mathbf{p} - \mathbf{r}, d-1) + P_1, \\ & L_{\mathbf{r}}(\mathbf{p} - \mathbf{r}, d+1) + P_1, \\ & \min_i L_{\mathbf{r}}(\mathbf{p} - \mathbf{r}, i) + P_2) - \min_k L_{\mathbf{r}}(\mathbf{p} - \mathbf{r}, k) \end{aligned} \tag{2.6}$$

Nella figura 2.2 sono mostrati due esempi con 8 e 16 direzioni per i cammini in una dimensione.

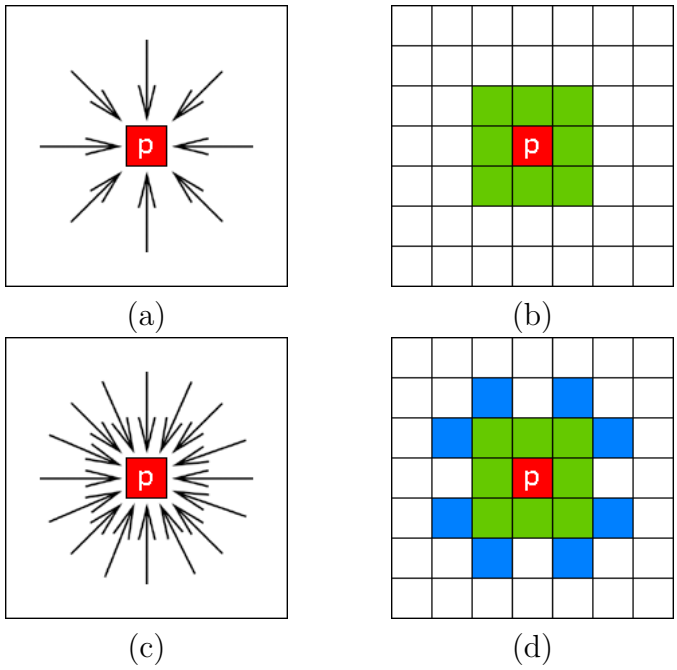


Figura 2.2: Esempio di cammini lungo 8 e 16 direzioni (rispettivamente, (a) e (c)) che terminano in $\mathbf{p}$ con relative rappresentazioni nell'immagine raster (rispettivamente, (b) e (d)). Nella figura (d) sono rappresentate in verde le prime 8 direzioni (identiche al caso (b)) e in blu le 8 direzioni rappresentate dallo spostamento verticale o orizzontale di un pixel seguito dallo spostamento in diagonale di un pixel.

Il costo $S(\mathbf{p}, d)$ della disparità d per il punto $\mathbf{p}$ viene definito come segue:

$$S(\mathbf{p}, d) = \sum_{\mathbf{r}} L_{\mathbf{r}}(\mathbf{p}, d) \tag{2.7}$$

Dalle considerazioni sul limite superiore di 2.6, il limite superiore per $S(\mathbf{p}, d)$ è dato dal numero di direzioni n_r moltiplicato per il massimo costo di un cammino di costo minimo: $S(\mathbf{p}, d) \leq n_r \cdot (C_{max} + P_2)$.

2.2.2 Calcolo della disparità

L'algoritmo prevede di calcolare due mappe di disparità: quella rispetto all'immagine base e quella rispetto all'immagine match.

In entrambi i casi, il calcolo della disparità viene eseguito secondo una strategia *winner-takes-all*. Considerando l'immagine base, la disparità del punto $\mathbf{p}$ è l'ipotesi con costo aggregato $S(\mathbf{p}, d)$ minimo:

$$D_{\mathbf{p}} = \underset{d}{\operatorname{argmin}} S(\mathbf{p}, d) \tag{2.8}$$

Nel caso dell'immagine match, la strategia è identica: la disparità del punto $\mathbf{q}$ è l'ipotesi con costo aggregato minimo. In questo caso, i costi delle ipotesi di disparità possono essere o calcolati da zero oppure assunti uguali ai costi delle ipotesi calcolati per i punti corrispondenti nell'immagine base. In questo secondo caso, considerando le due linee epipolari corrispondenti nelle due immagini, al punto $\mathbf{q}$ viene assegnata la disparità d di costo minimo per il punto $(q_x - d, q_y) = e_{mb}(\mathbf{q}, d)$ nell'immagine base:

$$D_{\mathbf{q}} = \underset{d}{\operatorname{argmin}} S(e_{mb}(\mathbf{q}, d), d) \tag{2.9}$$

dove $e_{mb}(\mathbf{q}, d)$ rappresenta la linea epipolare nell'immagine base corrispondente al punto $\mathbf{q}$.

Dopo aver calcolato le due mappe di disparità, viene eseguito il *left-right check* per determinare errori di assegnazione causati da occlusioni, garantendo inoltre l'univocità delle corrispondenze tra le due immagini.

2.2.3 Implementazione utilizzata

È stata utilizzata e riadattata l'implementazione dell'algoritmo presente in OpenCV [5], la quale utilizza 8 cammini nella fase di aggregazione dei costi. Inoltre, permette di eseguire match tra blocchi centrati nei punti di dimensione $w \times w$ con w dispari: con $w = 1$ viene eseguito il match tra pixel.

Riguardo al calcolo delle metriche di accoppiamento, è stata applicata la funzione pubblicata in [4]: si tratta di una funzione di dissomiglianza tra pixel a livello di sub-pixel. Solitamente, la dissomiglianza viene quantificata nel valore assoluto della differenza di intensità tra i due pixel ipotizzati corrispondenti. Questa funzione, invece, opera a livello di sub-pixel e calcola la differenza minima assoluta delle intensità dei due punti corrispondenti nell'intervallo di metà pixel in ogni direzione lungo la linea epipolare. Siano I_L e I_R le funzioni di intensità di due linee epipolari corrispondenti nelle due immagini, $\overline{I}_L$ e $\overline{I}_R$ le stesse due funzioni interpolate a livello di 0.5 pixel e x_L e $x_R = x_L - d$ due punti ipotizzati corrispondenti a disparità d. La dissomiglianza tra i due punti rispetto a x_L viene definita come segue:

$$\overline{d}\,(x_L,\, x_R) = \min_x \left| I_L(x_L) - \overline{I}_R(x) \right| \;,\; x \in \left\{ x_R - \frac{1}{2},\, x,\, x_R + \frac{1}{2} \right\} \tag{2.10}$$

Allo stesso modo, viene definita la dissomiglianza tra i due punti rispetto a x_R:

$$\overline{d}\,(x_R,\, x_L) = \min_x \left| \overline{I}_L(x) - I_R(x_R) \right| \;,\; x \in \left\{ x_L - \frac{1}{2},\, x_L,\, x_L + \frac{1}{2} \right\} \tag{2.11}$$

Il costo dell'ipotesi di disparità d viene definita come il minimo tra le quantità 2.10 e 2.11.

In termini di efficienza e di velocità di esecuzione dell'algoritmo, sono state utilizzate le istruzioni SIMD (*Single Instruction, Multiple Data*) per garantire una maggiore velocità di esecuzione dell'algoritmo: la struttura rigida e regolare, assieme alle semplici operazioni da eseguire, permette di eseguire in parallelo la stessa operazione per più dati contemporaneamente, situazione ben gestita da questo insieme di istruzioni.

2.3 Stima dell'errore sulla disparità

Le mappe di disparità ottenute tramite i due metodi possono contenere alcuni errori di assegnazione. Questi sono presenti principalmente per due motivi:

- l'algoritmo di calcolo può commettere errori;
- la creazione dei dati di input da parte dei sensori può essere affetta da rumore (illuminazione, rumore termico), il quale può perturbare la stima della disparità.

Per ognuno dei due metodi sono state quindi sviluppate delle funzioni il cui scopo è determinare l'affidabilità della disparità. Tutti i metodi forniscono una stima dell'errore della profondità espresso in metri.

2.3.1 Stima dell'errore per il TOF

Come descritto nel paragrafo 1.2.1, l'acquisizione tramite sensore TOF restituisce quattro dati in output: la mappa di profondità in metri, l'immagine di profondità in scala di grigi a 8 bit, l'immagine di ampiezza del segnale e la mappa di confidenza del segnale.

Tra questi dati, la mappa di profondità in metri e l'immagine di ampiezza del segnale sono stati scelti come parametri per la stima dell'errore. L'immagine di profondità e la mappa di confidenza sono state scartate perché i loro valori non sono ritenuti validi per questo calcolo:

- l'immagine di profondità viene creata trasformando il valore della profondità rilevato in metri in un valore a 8 bit, perdendo precisione;
- dal manuale del sensore TOF [3], la confidenza è un valore stimato sulla base di alcuni dati e, causa l'eventuale presenza di rumore in uno o più di questi ultimi, potrebbe essere inaccurato; inoltre, l'ampiezza del segnale nella fase di ritorno è uno dei dati utilizzati per stimarla, fatto che suggerisce di utilizzare direttamente l'ampiezza stessa.

Di seguito sono approfonditi gli indici di errore sviluppati divisi in due gruppi, in base alla modalità di calcolo. La stima dell'errore per un punto g della mappa di disparità associata al sensore TOF sarà indicata con $\Delta^T(g)$.

2.3.1.1 Deviazione standard della profondità

Il primo gruppo di indici si basa sul calcolo della deviazione standard della profondità in un intorno di un punto; ognuno di questi è strutturato come segue:

1. a partire dalla mappa di profondità interpolata ad alta risoluzione, ogni punto viene proiettato nella mappa di profondità originale del sensore TOF;
2. se la proiezione è all'interno della mappa, al punto viene associata la misura di errore.

Riguardo al punto 1, la proiezione viene eseguita utilizzando le matrici dei parametri intrinseci dei sensori e le matrici di rotazione da un piano focale all'altro; nel passaggio dal piano immagine della mappa interpolata al piano focale, è stato utilizzato il valore interpolato della profondità.

Riguardo l'associazione della misura al punto 2, è stata eseguita secondo uno dei seguenti possibili metodi.

Interpolazione. Per ogni punto della mappa di profondità originale, viene calcolata la deviazione standard della profondità all'interno di una finestra di dimensione dispari centrata nel punto; per i punti la cui finestra non è contenuta completamente nell'immagine, viene assunta deviazione standard nulla. Dopo questa operazione, è possibile eseguire la stima dell'errore.

La proiezione del punto avrà coordinate frazionarie e si troverà in un intorno delimitato da 4 punti (con coordinate intere) del piano immagine del sensore; la misura di errore assegnata alla proiezione corrisponde all'interpolazione bilineare delle deviazioni standard dei quattro punti.

Deviazione standard su finestra pari. In questo caso, la proiezione del punto viene utilizzata come centro di una finestra di dimensione pari; la misura di errore coincide con la deviazione standard della profondità calcolata all'interno della finestra.

Deviazione standard su finestra dispari. La proiezione del punto viene approssimata con il punto del piano immagine ad essa più vicino; la

misura di errore coincide con la deviazione standard calcolata su una finestra centrata sul punto approssimato.

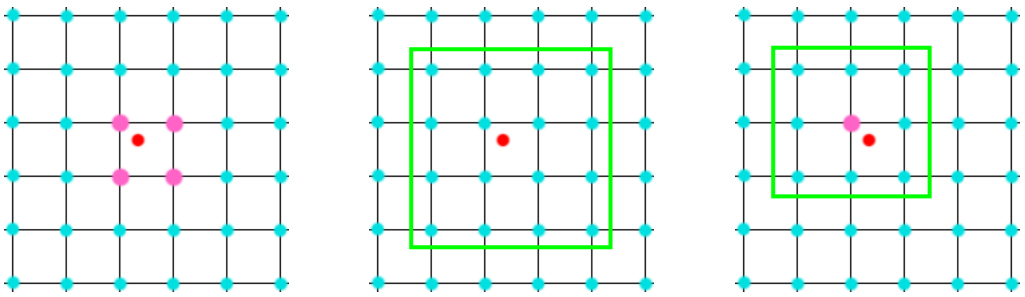


Figura 2.3: Raffigurazioni dei tre approcci per il calcolo della deviazione standard σ per il punto proiettato (punto rosso): (a) σ equivale all'interpolazione bilineare dei valori σ_i associati ai quattro punti vicini (punti viola); (b) σ equivale alla deviazione standard dei valori presenti nella finestra pari; (c) il punto proiettato viene approssimato con il punto più vicino (punto viola) e σ equivale alla deviazione standard dei valori presenti nella finestra dispari.

I due indici basati su finestra hanno in comune una caratteristica: vi saranno gruppi di punti della mappa ad alta risoluzione proiettati nello stesso punto nel piano immagine del sensore TOF e, quindi, che avranno lo stesso errore. Si tratta di una conseguenza inevitabile data l'evidente differenza tra le risoluzioni della camera e del sensore a tempo di volo.

2.3.1.2 Incertezza basata sull'ampiezza

Un altro parametro distintivo riguardo la qualità della misura è l'ampiezza misurata in fase di ricezione. Controllare questo parametro risulta utile, ad esempio, per verificare le misure di profondità lungo i bordi di oggetti dato che, come già detto in 1.2.1.1, il segnale infrarosso inviato viene riflesso verso molte direzioni e solo una piccola parte ritorna al sensore.

Basandosi su questa osservazione, è stata creata una tabella di look-up, che associa un'incertezza della misura di profondità ad ogni valore di ampiezza che il sensore può rilevare. Le misurazioni di incertezza sono state eseguite in maniera sperimentale: sono stati acquisiti in successione un insieme di frame ed è stata calcolata la deviazione standard della profondità per ogni possibile valore di ampiezza del segnale.

Questa stima di errore incorpora la presenza di rumore termico dovuto al funzionamento del sensore e del rumore gaussiano presente in fase di creazione dell'output del sensore.

2.3.1.3 Combinazione delle due stime di errore

Per semplicità ed uniformità, risulta utile associare una sola misura di errore per ogni sensore alla disparità in esame. Pertanto, è necessaria una strategia di combinazione dei due errori appena definiti.

Entrambe le misure sono deviazioni standard della profondità e, quindi, risultano grandezze direttamente confrontabili. Come metodi di combinazione, sono stati scelti i più semplici possibili e con bassissimo costo computazionale supplementare: massimo o somma o media aritmetica delle due grandezze.

2.3.2 Stima dell'errore per lo stereo

La stima dell'errore per la disparità stereo è ovviamente dipendente dall'algoritmo utilizzato. Tuttavia, la soluzione proposta è valida per vari algoritmi: è sufficiente che siano basati su strategie di aggregazione di costo.

Secondo la struttura dell'algoritmo SGM [12], la miglior ipotesi di disparità d all'interno dell'intervallo $[d_{min}, d_{max}]$ per un punto $\mathbf{p}$ è quella con costo $S(\mathbf{p}, d)$ minore. Siano quindi d_o l'ipotesi ottima, d' la seconda migliore ipotesi, $S(\mathbf{p}, d_o)$ e $S(\mathbf{p}, d')$ i costi associati; vale la relazione $S(\mathbf{p}, d_o) \leq S(\mathbf{p}, d')$.

Intuitivamente, l'errore potrebbe essere dato dalla differenza tra le due disparità: maggiore è la differenza, maggiore sarà l'errore. Tuttavia, questo criterio non trae vantaggio dall'informazione messa a disposizione dalla struttura dell'algoritmo. Combinando l'errore intuitivo con l'osservazione che l'ipotesi con costo minore viene considerata ottima, possiamo ottenere il seguente indicatore di errore:

$$\Delta'^{S} = (d_o - d') \cdot \frac{C(d_o)}{C(d')} \tag{2.12}$$

Con questa funzione di errore è possibile codificare contemporaneamente l'errore tra le disparità candidate e la sicurezza sulla correttezza della disparità

ottima d_o: in presenza di un rapporto tra i costi molto basso, l'algoritmo ha identificato un minimo assoluto molto accentuato nella funzione $C(\mathbf{p}, d_o)$ al variare di d, fatto che determina con alta probabilità la correttezza di d_o.

Questo indicatore è espresso in termini di disparità, mentre nel caso del sensore TOF l'errore è espresso come profondità in metri. La trasformazione dell'indicatore per la disparità stereo avviene applicando la relazione tra profondità e disparità prima della moltiplicazione tra costi, diventando come segue:

$$\Delta^S = \left(\frac{b \cdot f}{d_o} - \frac{b \cdot f}{d'} \right) \cdot \frac{C(d_o)}{C(d')} \tag{2.13}$$

dove b è la baseline dello stereo e f la lunghezza focale. Per indicare la stima dell'errore associata ad un punto g della mapap di disparità ottenuta dall'algoritmo stereo sarà utilizzata la notazione $\Delta^S(g)$.

In figura 2.4 vi è il grafico dell'andamento della funzione. Da notare che per alte differenze di profondità può esserci un basso errore (tonalità blu).

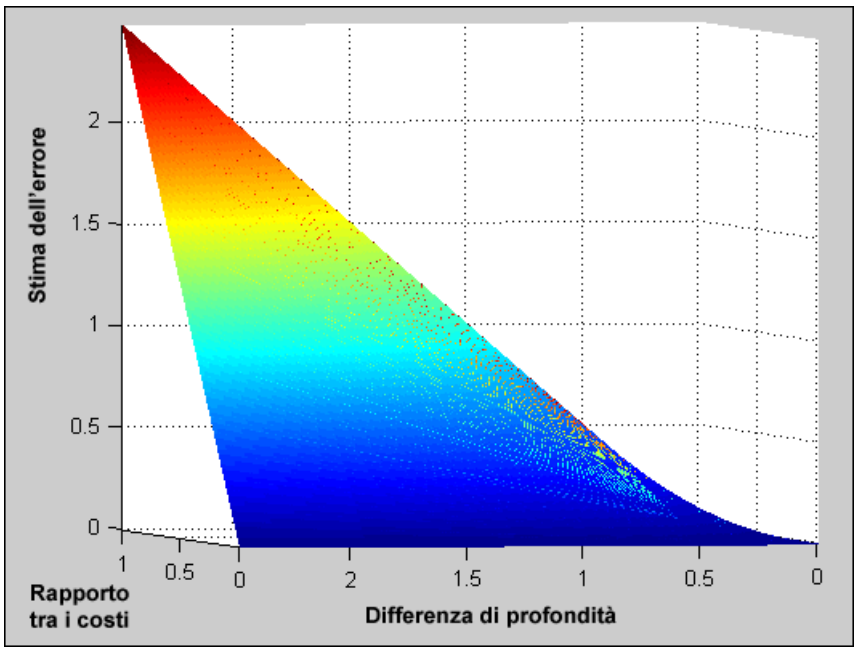


Figura 2.4: Grafico della funzione di errore 2.13 per i seguenti parametri: $d_o, d' \in [16, 224]$, baseline $= 0.176803$, lunghezza focale $= 856.311$.

Capitolo 3

Fusione delle ipotesi

Le due mappe di disparità calcolate forniscono due ipotesi di ricostruzione 3D. Considerate singolarmente, queste possono essere confrontate per stabilire quale dei due sensori fornisca il risultato migliore. Tuttavia, risulta più interessante cercare di combinarle: dai vantaggi dei due sensori è auspicabile una fusione delle due mappe di disparità, dove le ipotesi probabilmente errate date da un sensore vengano sostituite da ipotesi più plausibili fornite dall'altro e viceversa.

3.1 Tecnica Locally Consistent (LC)

La maggior parte degli algoritmi stereo basati su strategie locali calcolano la disparità di un punto f in base alle ipotesi sostenute dai punti presenti in una finestra centrata in f detta supporto di f. Si tratta quindi di calcolare un valore che sia compatibile con le ipotesi all'interno della finestra.

Al contrario, la tecnica LC considera il calcolo della disparità da un altro punto di vista: utilizzando una finestra di dimensione $m \times n$, uno stesso punto g viene incluso in molte finestre centrate sui punti vicini ($m \times n$ finestre). Assumendo l'ipotesi di superfici lisce a tratti, la disparità del punto g viene assunta, per ognuna delle finestre, pari a quella del punto in cui è centrato il supporto: è quindi possibile avere $m \times n$ diverse ipotesi di disparità per uno stesso punto g.

Combinando questa idea con l'utilizzo di vincoli di spaziali e fotometrici all'interno del supporto, è possibile calcolare una plausibilità per ogni ipotesi di disparità. Si tratta di una forma di ottimizzazione globale applicata in ambito locale (all'interno della finestra).

Di seguito viene descritto l'algoritmo. Lo scopo iniziale è il miglioramento della mappa di disparità fornita da un algoritmo stereo: pertanto, si assume che sia stata assegnata un'ipotesi di disparità per tutti i punti delle immagini (evitando le occlusioni). Sarà considerata la situazione composta da una coppia di immagini stereo rettificate (reference R e target T) e la disparità variabile in un intervallo predefinito $[d_{min}, d_{max}]$.

3.1.1 Plausibilità della corrispondenza tra due punti

La prima parte è il calcolo della funzione di plausibilità. Data una certa ipotesi di disparità d, si consideri la seguente situazione: il punto $f \in R$ con supporto S_f, un punto $g \in S_f$, i punti nell'immagine target univocamente determinati come $f' = f - d$ e $g' = g - d$, il supporto $S_{f'}$ di f'. La plausibilità dei punti g e g' rappresenta l'occorrenza congiunta dei seguenti eventi (figura 3.1).

$\mathbf{E^R_{f\,g}}$ **:** Questo evento codifica l'opinione che il punto g abbia la stessa disparità di f (oppure molto vicina). La plausibilità di tale evento è legata alla distanza tra f e g e la prossimità di colore tra i due punti. Per il primo parametro, si considera a priori che, minore è la distanza di f da g, maggiore è la rilevanza dell'ipotesi di f nel calcolo della disparità per g. Per il secondo, poiché la scena contiene cambiamenti graduali di colore, la prossimità di colore risulta necessariamente legata alla plausibilità.

$\mathbf{E^T_{f'\,g'}}$ **:** Questo evento codifica l'opinione che il punto g' abbia la stessa disparità di f' (oppure molto vicina). Tale evento risulta simile al precedente e la sua plausibilità viene caratterizzata dagli stessi parametri.

$\mathbf{E^{RT}_{g\,g'}(d)}$ **:** Questo evento codifica l'opinione che i punti $g \in S_f$ e $g' \in S_{f'}$ abbiano la stessa disparità (d e $-d$ rispettivamente). Tale evento è quindi caratterizzato dalla prossimità di colore tra i due punti (g e g' sono lo stesso punto in due piani immagine diversi solo per un'operazione di traslazione) e dall'ipotesi di disparità d.

Siano $\Delta^{\psi}_{f\,g}$ e $\Delta^{\psi}_{f'\,g'}$ due funzioni che calcolano la prossimità di colore dato un determinato spazio di colore, rispettivamente, tra f e g e tra f' e

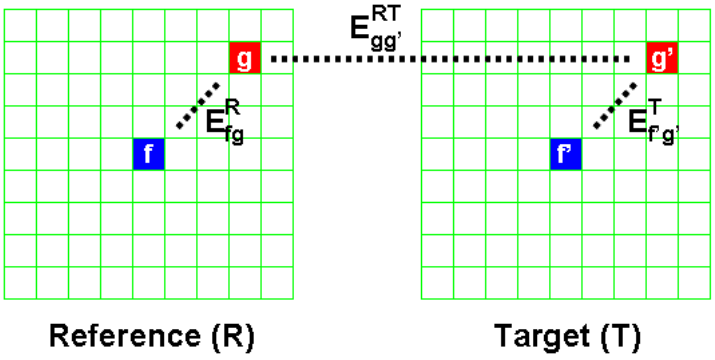


Figura 3.1: Situazione delineata per il calcolo della funzione di plausibilità con i tre eventi $\mathbf{E^R_{f\,g}}$, $\mathbf{E^T_{f'\,g'}}$ e $\mathbf{E^{RT}_{g\,g'}(d)}$ (immagine fornita da [14]).

g'; sia $\Delta^{\omega}_{g\,g'}$ una funzione che calcola la prossimità di colore tra g e g'. La plausibilità viene definita come la probabilità congiunta dei tre eventi $E^R_{f\,g}$, $E^T_{f'\,g'}$ e $E^{RT}_{g\,g'}(d)$ dati i valori di prossimità spaziale e di colore:

$$P\left(E^R_{f\,g},\, E^T_{f'\,g'},\, E^{RT}_{g\,g'}(d) \,\middle|\, \Delta^{\psi}_{f\,g},\, \Delta^{\psi}_{f'\,g'},\, \Delta^{\omega}_{g\,g'}\right)$$

Per il teorema di Bayes, vale la seguente formulazione in probabilità a priori (*prior probability*) e verosimiglianza (*likelihood*):

$$\begin{aligned} P\left(E^R_{f\,g},\, E^T_{f'\,g'},\, E^{RT}_{g\,g'}(d) \,\middle|\, \Delta^{\psi}_{f\,g},\, \Delta^{\psi}_{f'\,g'},\, \Delta^{\omega}_{g\,g'}\right) \propto \\ P_P\left(E^R_{f\,g},\, E^T_{f'\,g'},\, E^{RT}_{g\,g'}(d)\right) \cdot \\ P_L\left(\Delta^{\psi}_{f\,g},\, \Delta^{\psi}_{f'\,g'},\, \Delta^{\omega}_{g\,g'} \,\middle|\, E^R_{f\,g},\, E^T_{f'\,g'},\, E^{RT}_{g\,g'}(d)\right) \end{aligned} \tag{3.1}$$

Assumendo l'indipendenza tra gli eventi $E^R_{f\,g}$, $E^T_{f'\,g'}$ e $E^{RT}_{g\,g'}(d)$, si ottiene quanto segue:

$$\begin{aligned} P\left(E^R_{f\,g},\, E^T_{f'\,g'},\, E^{RT}_{g\,g'}(d) \,\middle|\, \Delta^{\psi}_{f\,g},\, \Delta^{\psi}_{f\,g},\, \Delta^{\omega}_{f\,g}\right) \propto \\ P_P\left(E^R_{f\,g}\right) \cdot P_L\left(\Delta^{\psi}_{f\,g} \,\middle|\, E^R_{f\,g}\right) \cdot \\ P_P\left(E^T_{f'\,g'}\right) \cdot P_L\left(\Delta^{\psi}_{f'\,g'} \,\middle|\, E^T_{f'\,g'}\right) \cdot \\ P_P\left(E^{RT}_{g\,g'}(d)\right) \cdot P_L\left(\Delta^{\omega}_{g\,g'} \,\middle|\, E^{RT}_{g\,g'}(d)\right) \end{aligned} \tag{3.2}$$

Riguardo alle probabilità a priori, $P_P\left(E^R_{f\,g}\right)$ e $P_P\left(E^T_{f'\,g'}\right)$ rappresentano un vincolo spaziale e sono caratterizzate in maniera gaussiana come segue:

$$P_P\left(E^R_{f\,g}\right) = e^{-\frac{\Delta_{f\,g}}{\gamma_s}} \tag{3.3}$$

$$P_P\left(E^T_{f'\,g'}\right) = e^{-\frac{\Delta_{f'\,g'}}{\gamma_s}} \tag{3.4}$$

dove $\Delta_{a\,b}$ rappresenta la distanza euclidea tra i due punti a e b, mentre γ_s è un parametro che controlla il vincolo di distanza spaziale. Riguardo a $P_P\left(E^{RT}_{g\,g'}(d)\right)$, non vi è alcuna conoscenza a priori sulla disposizione dei punti g e g' nelle immagini, ovvero sulla disparità d: pertanto viene stimata in maniera uniforme, senza influire sulla plausibilità.

Riguardo alle verosimiglianze, si basano sulle seguenti ipotesi: utilizzo di superfici lambertiane e presenza di rumore gaussiano indipendente e identicamente distribuito nella formazione dell'immagine. Sia $I(p)$ un vettore che rappresenta l'intensità di colore del punto p; le funzioni di prossimità di colore sono definite come la distanza euclidea tra i vettori di intensità del colore dei due punti. Assumendo lo spazio di colore RGB (l'intensità $I(p)$ è quindi caratterizzata da $I_R(p)$, $I_G(p)$ e $I_B(p)$), si ottengono le funzioni di prossimità del colore associate ai tre eventi:

$$\Delta^{\psi}_{f\,g} = \sqrt{\sum_{c\in R,G,B}\left(I_c(f) - I_c(g)\right)^2} \tag{3.5}$$

$$\Delta^{\psi}_{f'\,g'} = \sqrt{\sum_{c\in R,G,B}\left(I_c(f') - I_c(g')\right)^2} \tag{3.6}$$

$$\Delta^{\omega}_{g\,g'} = \sqrt{\sum_{c\in R,G,B}\left(I_c(g) - I_c(g')\right)^2} \tag{3.7}$$

Anche per le verosimiglianze, la plausibilità viene modellata in maniera gaussiana:

$$P_L\left(\Delta^{\psi}_{f\,g}\,\middle|\,E^R_{f\,g}\right) = e^{-\frac{\Delta^{\psi}_{f\,g}}{\gamma_c}} \tag{3.8}$$

$$P_L\left(\Delta^{\psi}_{f'\,g'}\,\middle|E^{T}_{f'\,g'}\right) = e^{-\frac{\Delta^{\psi}_{f'\,g'}}{\gamma_c}} \tag{3.9}$$

$$P_L\left(\Delta^{\omega}_{g\,g'}\,\middle|E^{RT}_{g\,g'}(d)\right) = e^{-\frac{\Delta^{\omega}_{g\,g'}}{\gamma_t}} \tag{3.10}$$

dove γ_c e γ_t sono due parametri che regolano i vincoli di prossimità di colore.

La plausibilità viene quindi espressa nel seguente modo:

$$P\left(E^{R}_{f\,g},\, E^{T}_{f'\,g'},\, E^{RT}_{g\,g'}(d)\,\middle|\Delta^{\psi}_{f\,g},\, \Delta^{\psi}_{f\,g},\, \Delta^{\omega}_{f\,g}\right) \propto e^{-\frac{\Delta_{f\,g}}{\gamma_s}} \cdot e^{-\frac{\Delta^{\psi}_{f\,g}}{\gamma_c}} \cdot e^{-\frac{\Delta_{f'\,g'}}{\gamma_s}} \cdot e^{-\frac{\Delta^{\psi}_{f'\,g'}}{\gamma_c}} \cdot e^{-\frac{\Delta^{\omega}_{g\,g'}}{\gamma_t}} \tag{3.11}$$

L'equazione 3.11 rappresenta sia la plausibilità del punto g rispetto a f per la disparità d, sia la plausibilità del punto g' rispetto a f' per la disparità $-d$. Di seguito, l'equazione 3.11 sarà rappresentata dal simbolo $P^{R}_{f\rightarrow g}(d) = P^{T}_{f'\rightarrow g'}(-d)$.

3.1.2 Calcolo della plausibilità per ogni disparità

Il calcolo della funzione di plausibilità viene eseguito per ogni punto $f \in R$ avente un'ipotesi di disparità d: tutti i punti $g \in S_f$ saranno assunti a disparità d e riceveranno una plausibilità rispetto al punto f per tale ipotesi di disparità. Senza perdita di generalità, si assume che $\forall f \in R : d \in [d_{min},\, d_{max}]$[1].

Pertanto, ogni punto $g \in R$ riceverà una plausibilità da un determinato insieme di punti, ovvero tutti quelli all'interno di una finestra centrata su di esso della stessa dimensione del supporto S_f. Questo insieme di punti forma il supporto attivo S_g di g. Da ogni punto di tale supporto, g può ricevere una plausibilità per un'ipotesi di disparità potenzialmente sempre diversa, pur sempre compresa nell'intervallo $[d_{min}, d_{max}]$.

Contemporaneamente, verrà calcolata la plausibilità rispetto ai punti $f' = f - d$ e questa verrà propagata ai punti $g' \in S_{f'}$. Tuttavia, la propagazione delle plausibilità verso i punti dell'immagine target non risulta

[1]Se non sono conosciuti gli estremi dell'intervallo, è sufficiente scorrere la mappa di disparità in cerca del valore minimo e massimo.

uniforme come per i punti dell'immagine reference: poiché i punti f' sono determinati dalla disparità d, vi saranno alcuni punti g' che non riceveranno plausibilità da punti presenti nel loro supporto attivo poiché sono punti senza disparità (occlusi ad esempio) e, quindi, senza una plausibilità da propagare. Questo caso particolare può essere risolto osservando che, per un punto f' che dovrebbe propagare una plausibilità, non fare effettivamente parte del supporto attivo di alcuni punti è esso stesso un chiaro segnale della sua scarsa plausibilità e, pertanto, può essere considerato a plausibilità nulla.

In entrambe le immagini di reference e target vengono quindi calcolate le plausibilità accumulate per ogni disparità d assegnata da un algoritmo stereo:

$$\Omega'^{R}(g \,|\, d) = \sum_{i \in S_g} P^{R}_{i \rightharpoonup g}(d) \tag{3.12}$$

$$\Omega'^{T}(g' \,|\, -d) = \sum_{i' \in S_{g'}} P^{T}_{i' \rightharpoonup g'}(-d) \tag{3.13}$$

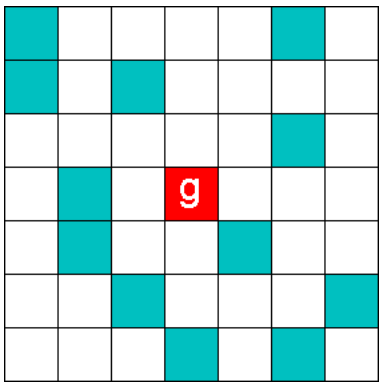


Figura 3.2: Esempio di accumulazione della plausibilità $\Omega'^{R}(g \,|\, d)$: è rappresentato il supporto attivo di g di dimensione 7×7; i pixel evidenziati in azzurro hanno tutti la stessa disparità d e propagano la loro plausibilità verso g. A tutti gli altri punti è stata associata una disparità diversa, pertanto propagano una plausibilità nulla per la disparità d.

Per rendere questi valori più simili a probabilità e per poterli confrontare, i valori calcolati in 3.12 e 3.13 vengono normalizzati per la somma delle plausibilità del rispettivo punto g e g':

$$\Omega^R(g \mid d) = \frac{\Omega'^R(g \mid d)}{\sum_{i=d_{min}}^{d_{max}} \Omega'^R(g \mid i)} \tag{3.14}$$

$$\Omega^T(g' \mid -d) = \frac{\Omega'^T(g' \mid -d)}{\sum_{i=d_{min}}^{d_{max}} \Omega'^T(g' \mid -i)} \tag{3.15}$$

Le equazioni 3.14 e 3.15 rappresentano le plausibilità accumulate normalizzate per i punti g e g' rispettivamente per la stessa disparità d.

3.1.3 Calcolo della disparità

Avendo calcolato tutte le plausibilità accumulate, è ora possibile calcolare la mappa di disparità. L'approccio scelto è quello proposto da [14] tramite la cross-validazione delle plausibilità accumulate: in questo modo, il risultato è più robusto e fornisce una mappa di disparità densa, cosa non sempre garantita da altre tecniche come il left-right check.

La cross-validazione è stata calcolata rispetto all'immagine di reference:

$$\Omega^{RT}(g \mid d) = \Omega^R(g \mid d) \cdot \Omega^T(g - d \mid -d) \tag{3.16}$$

La disparità è quindi data dalla seguente semplice strategia locale:

$$\overline{d} = \underset{d \in [d_{min}, d_{max}]}{\operatorname{argmax}} \; \Omega^{RT}(g \mid d) \tag{3.17}$$

3.1.4 Considerazioni sull'algoritmo

Come è stato premesso, la tecnica Locally Consistent è una strategia di miglioramento di un'ipotesi di disparità calcolata su due immagini stereo. Risulta quindi fondamentale sottolineare che dipende strettamente dall'ipotesi di disparità fornita: ipotesi non valide, come per esempio su punti occlusi, potrebbero perturbare le plausibilità accumulate. Pertanto, è decisamente consigliato eseguire un left-right check per eliminare ipotesi errate di plausibilità prima di applicare la tecnica.

Un vantaggio dato da LC, invece, è che la mappa di disparità iniziale non deve essere densa: infatti, aumentando la dimensione del supporto attivo, è possibile includere più punti dai quali ricevere una plausibilità e garantire una mappa di disparità densa come risultato al prezzo di un maggior tempo di esecuzione.

Una terza osservazione riguarda la funzione di plausibilità, la quale è definita sull'informazione del colore e dalla distanza tra i punti considerati. Si tratta di una definizione in accordo con gli algoritmi stereo e rende la tecnica adatta sicuramente per il miglioramento di mappe di disparità date da algoritmi stereo; tuttavia, senza una coppia di immagini stereo a colori, la tecnica non è applicabile: sarebbe necessario cambiare la funzione di plausibilità radicalmente. Una ulteriore osservazione è la determinazione empirica dei parametri γ_s, γ_c e γ_t, che aggiunge un livello di incertezza all'algoritmo.

3.2 Fusione pesata tramite LC

Come era stata presentata in [14], LC è una tecnica per migliorare la mappa di disparità data da un algoritmo stereo, propagando più valori di plausibilità per le ipotesi di disparità considerate migliori.

L'estensione del framework LC al caso in cui vi sia, per ogni punto, più di un'ipotesi di disparità risulta piuttosto immediata: ogni punto permette di propagare tante plausibilità quante le ipotesi di disparità fornite dai sensori utilizzati. In questo caso, i sensori utilizzati sono due ed ogni punto potrà propagare due, una o nessuna plausibilità. Questa strategia di fusione delle mappe di disparità date dalle telecamere stereo e dal sensore TOF è in accordo con le caratteristiche dei sensori: in regioni con texture molto uniforme, l'algoritmo stereo tende a fornire ipotesi di disparità errate, le quali saranno caratterizzate da una plausibilità più bassa rispetto a quella delle ipotesi date dal ToF (il quale risulta preciso in queste situazioni); al contrario, in regioni dove il sensore ToF non fornisce ipotesi valide (oggetti poco riflettenti o regioni scure dell'immagine), verrà propagata la disparità data dall'algoritmo stereo.

Così facendo, le plausibilità calcolate per le disparità date dai sensori saranno considerate alla stessa maniera. Tuttavia, è possibile che un sensore fornisca errate ipotesi di disparità per una regione della mappa, perturbando

la plausibilità accumulata per i punti nella regione. Risulta quindi opportuno dare un peso alle plausibilità dei sensori, che indichi l'affidabilità dell'ipotesi di disparità sulla quale si basa la funzione di plausibilità. L'idea è stata di utilizzare le stime di affidabilità elencate nel paragrafo 2.3: aggiungendo un termine nella funzione di plausibilità, è stato possibile fornire un peso che definisse la quota di ogni plausibilità da propagare. In caso di alta affidabilità, la quota si avvicina a 1, mentre in caso di bassa affidabilità si avvicina allo 0.

3.2.1 Plausibilità nuova per le telecamere stereo

Ricordando l'equazione 3.11, la plausibilità è data da cinque termini esponenziali. L'inclusione dell'affidabilità potrebbe essere eseguita moltiplicando la plausibilità nota per un nuovo termine, anch'esso esponenziale per uniformità. Per la disparità fornita dall'algoritmo stereo, la nuova funzione di plausibilità diventerebbe quindi:

$$P^R_{f \rightharpoonup g}(d) = P^T_{f' \rightharpoonup g'}(-d) = e^{-\frac{\Delta_{f\,g}}{\gamma_s}} \cdot e^{-\frac{\Delta^{\psi}_{f\,g}}{\gamma_c}} \cdot e^{-\frac{\Delta_{f'\,g'}}{\gamma_s}} \cdot e^{-\frac{\Delta^{\psi}_{f'\,g'}}{\gamma_c}} \cdot e^{-\frac{\Delta^{\omega}_{g\,g'}}{\gamma_c}} \cdot e^{-\frac{\Delta^{S}(f)}{\gamma_{es}}} \tag{3.18}$$

dove $\Delta^S(f)$ indica la stima dell'errore per la disparità associata al punto f, mentre γ_{es} è un parametro che controlla l'incidenza della stima dell'errore. Come per i parametri γ_s, γ_c e γ_t in [14], anche γ_{es} viene determinato empiricamente.

3.2.2 Plausibilità nuova per il sensore TOF

Nel caso del sensore TOF, la stima dell'errore è stata definita come la deviazione standard della profondità originata da due fonti differenti: la mappa di profondità e una tabella sperimentale che associa un'incertezza sulla profondità ad ogni valore dell'immagine di ampiezza. L'affidabilità viene sempre inclusa moltiplicando la plausibilità per un termine, il quale può essere espresso concettualmente in due forme:

1. un semplice termine moltiplicativo, come nel caso dell'affidabilità dell'algoritmo stereo;

2. un termine moltiplicativo basato su una distribuzione di probabilità gaussiana avente deviazione standard uguale a quella della stima dell'errore.

Nel primo caso, la funzione di plausibilità applicata per la disparità data dal sensore TOF diventerebbe:

$$P^R_{f\rightharpoonup g}(d) = P^T_{f'\rightharpoonup g'}(-d) = e^{-\frac{\Delta_{f\,g}}{\gamma_s}} \cdot e^{-\frac{\Delta^{\psi}_{f\,g}}{\gamma_c}} \cdot e^{-\frac{\Delta_{f'\,g'}}{\gamma_s}} \cdot e^{-\frac{\Delta^{\psi}_{f'\,g'}}{\gamma_c}} \cdot e^{-\frac{\Delta^{\omega}_{g\,g'}}{\gamma_c}} \cdot e^{-\frac{\Delta^T(f)}{\gamma_{et}}} \tag{3.19}$$

dove $\Delta^T(f)$ indica la stima dell'errore per la disparità del punto f, mentre γ_{et} è un parametro che controlla l'incidenza della stima dell'errore. Anche in questo caso, il parametro γ_{et} viene determinato empiricamente.

Nel secondo caso, per utilizzare correttamente la distribuzione gaussiana, è necessario calcolare la media assieme alla deviazione standard, la quale viene ottenuta contemporaneamente dal calcolo della deviazione standard.

È quindi possibile applicare l'errore basato su distribuzione gaussiana alla plausibilità come segue:

$$P^R_{f\rightharpoonup g}(d) = P^T_{f'\rightharpoonup g'}(-d) = e^{-\frac{\Delta_{f\,g}}{\gamma_s}} \cdot e^{-\frac{\Delta^{\psi}_{f\,g}}{\gamma_c}} \cdot e^{-\frac{\Delta_{f'\,g'}}{\gamma_s}} \cdot e^{-\frac{\Delta^{\psi}_{f'\,g'}}{\gamma_c}} \cdot e^{-\frac{\Delta^{\omega}_{g\,g'}}{\gamma_c}} \cdot e^{-\frac{(z-\mu)^2}{2\sigma^2}} \tag{3.20}$$

dove $z = (b \cdot f)/d$ è il valore di profondità associato alla disparità d, μ è la media della profondità all'interno della finestra del metodo di errore (si veda 2.3.1) e $\sigma = \Delta^T(f)$ è la stima dell'errore del sensore TOF.

3.2.3 Implementazione realizzata

L'algoritmo è stato implementato come presentato in [14]: si tratta della versione originale dell'algoritmo, poiché in un articolo più recente (si veda [15]) sono stati presentati dei suggerimenti per migliorare l'efficienza al costo di una leggera perdita di precisione. In questo lavoro è stata preferita la versione originale per poter valutare al meglio la precisione data dalla combinazione delle informazioni.

L'implementazione ripercorre perfettamente tutta la descrizione dell'algoritmo: sono calcolate innanzitutto tutte le plausibilità accumulate, dopodiché vengono normalizzate e, infine, viene applicata la cross-validazione. La disparità viene calcolata scegliendo l'ipotesi con maggiore plausibilità.

Nel dettaglio, la memoria richiesta è proporzionale alla dimensione delle immagini e all'intervallo di valori della disparità: si tratta quindi di un elevato requisito per immagini ad alta risoluzione o per intervalli di disparità numerosi. Il calcolo della plausibilità viene eseguito con complessità $\mathcal{O}\left(w \cdot h \cdot (2r+1)^2\right)$: per ogni punto dell'immagine di dimensione $w \times h$, viene considerato il suo supporto attivo di raggio r e, per ogni punto al suo interno, sono calcolati al più due valori di plausibilità con un numero costante di operazioni. Il calcolo della cross-validazione e della disparità vengono eseguiti con complessità $\mathcal{O}\left(w \cdot h \cdot (d_max - d_min)\right)$: per ogni punto dell'immagine di dimensione $w \times h$ sono considerate tutte le ipotesi di disparità nell'intervallo $[d_{min},\, d_{max}]$ e vengono eseguite un numero costante di operazioni (moltiplicazioni per la cross-validazione, massimo per il calcolo della disparità).

La realizzazione non è la più efficiente e non è real-time, tuttavia è stata applicata una parallelizzazione implicita tramite le direttive OpenMP [1] per ottenere un iniziale fattore di speed-up. Il calcolo delle plausibilità risulta essere la fase computazionalmente più pesante, che si presta perfettamente alla parallelizzazione. Un'implementazione più efficiente impiegherebbe le istruzioni SIMD, come nel caso dell'algoritmo stereo. Inoltre, il calcolo delle plausibilità potrebbe essere eseguito a fasce sulle immagini stereo e senza mantenere in memoria tutte le plausibilità accumulate, con una conseguente minore allocazione di memoria.

Capitolo 4

Analisi dei risultati

Per valutare le prestazioni del sistema precedentemente descritto, è stato implementato un software per eseguire test al variare dei parametri fondamentali. La sua implementazione è stata progettata per permettere l'aggiunta in futuro di altri sensori, senza doversi fermare ai due utilizzati in questo lavoro. Non è infatti escluso l'utilizzo di altri tipi di sensori o di un numero superiore di sensori per poter garantire sempre maggiore precisione.

Tutti i test sono stati eseguiti su scene statiche, questo perché la sincronizzazione del sistema di acquisizione non permette di raggiungere elevati framerate e perché il software non esegue i calcoli in real-time.

Il principale parametro di valutazione è stato l'errore quadratico medio (*mean squared error*, MSE), dato dalla somma dei quadrati delle differenze punto a punto tra la mappa di disparità calcolata e quella del ground truth. La valutazione ha considerato soltanto i punti non occlusi presenti nel ground truth.

4.1 Acquisizione dei dataset e parametri di test

La composizione del sistema prevede il posizionamento del sensore TOF al centro tra le due telecamere stereo con gli obiettivi quasi allineati. Il sistema è stato calibrato con il metodo descritto in [7], con un errore di calibrazione

di circa $5\,mm$.

Sono state acquisite 5 scene differenti per confrontare i metodi di calcolo della disparità e per verificare i miglioramenti dati dalla fusione. L'ambiente ripreso era stato predisposto per evitare il più possibile interferenze luminose dall'esterno. L'acquisizione è stata effettuata tenendo a mente l'ipotesi di superfici lisce a tratti (*piecewise smooth*) e le caratteristiche dei sensori: sono presenti scene con materiali poco riflettenti (svantaggiose per il sensore TOF), scene con poca texture (svantaggiose per il sistema stereo) e scene più complesse dove i risultati dei due sensori dovrebbero essere combinati per ottenere mappe di disparità più precise.

Il *ground truth* dei dataset è stato ricavato secondo il metodo descritto in [22]: è stato predisposto un sistema spacetime, il quale ha acquisito ed integrato 600 immagini per ottenere una mappa di disparità accurata.

Per quanto riguarda i test, le immagini sono state innanzitutto antidistorte e rettificate utilizzando i parametri forniti dalla calibrazione. Le figure di esempio mostrate durante l'analisi non rappresentano l'intera scena: sono state tagliate per evidenziare i soggetti inquadrati e gli eventuali errori.

(a) Immagine acquisita. (b) Immagine antidistorta e rettificata.

Figura 4.1: Esempio di antidistorsione e rettificazione dell'immagine acquisita dalla telecamera sinistra per il dataset 1.

(a) Dataset 1.

(b) Dataset 2.

(c) Dataset 3.

(d) Dataset 4.

(e) Dataset 5.

Figura 4.2: Regioni inquadrate per i 5 dataset; immagini acquisite dal punto di vista della telecamera sinistra.

4.2 Valutazione delle disparità singolarmente

4.2.1 Valutazione disparità associata al TOF

Come già detto in 2.1, la disparità associata al sensore TOF è il risultato di un'operazione di interpolazione. Il principale parametro in questo caso è la dimensione della finestra di interpolazione: nella figura 4.3 si può notare come cambia il MSE della mappa di disparità, per ognuno dei dataset, al variare del raggio della finestra di interpolazione. Dalla dimensione 21 (raggio 10) in poi, si può notare la saturazione del parametro poiché non vi sono più miglioramenti. Pertanto, è stato deciso di utilizzare questo valore per i test successivi. Per le due funzioni Gaussiane dell'interpolazione bilaterale congiunta sono stati utilizzati i seguenti due valori: $\sigma_s = 10$ per f_s, $\sigma_c = 5$ per f_c.

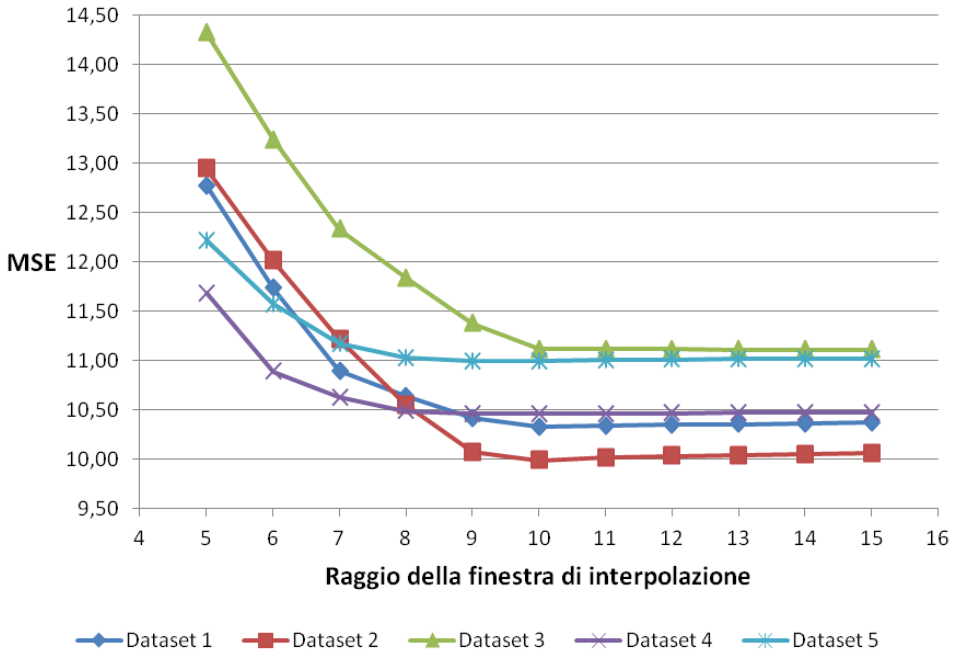
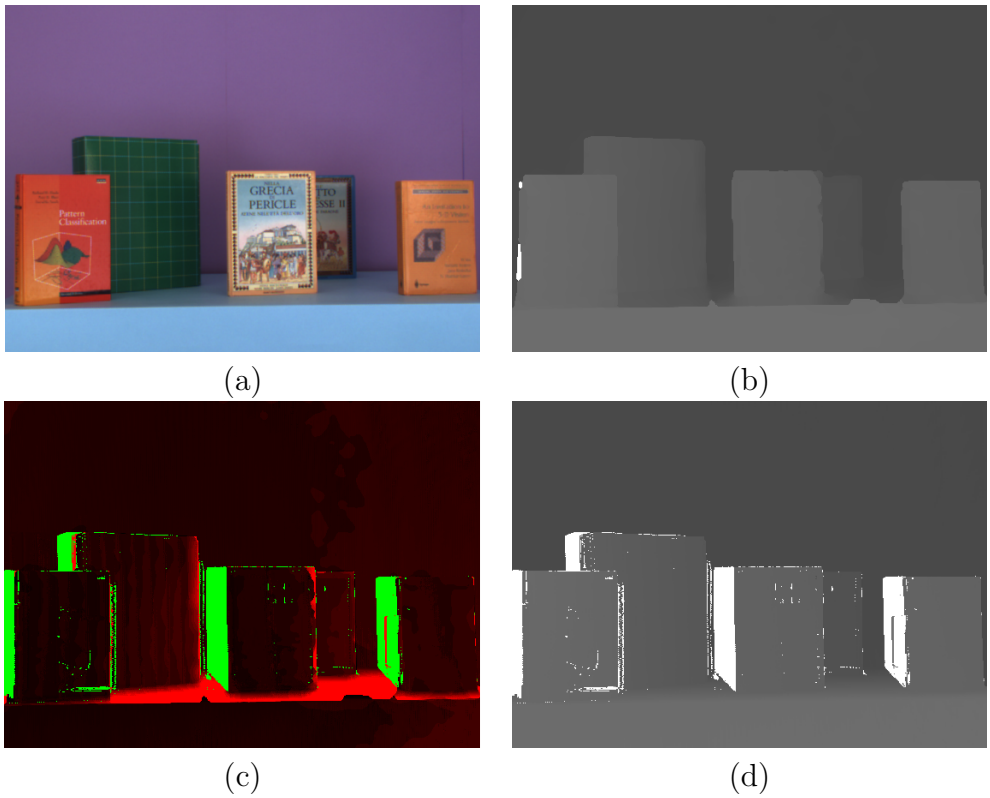


Figura 4.3: Variazione dell'errore MSE per ogni dataset al variare del raggio della finestra di interpolazione.

Nelle figure 4.4 e 4.5 possiamo notare le immagini che evidenziano le regioni con maggiore errore di disparità. Le intensità di rosso sono state moltiplicate per 5 per rendere l'errore ben visibile.



(a) (b)

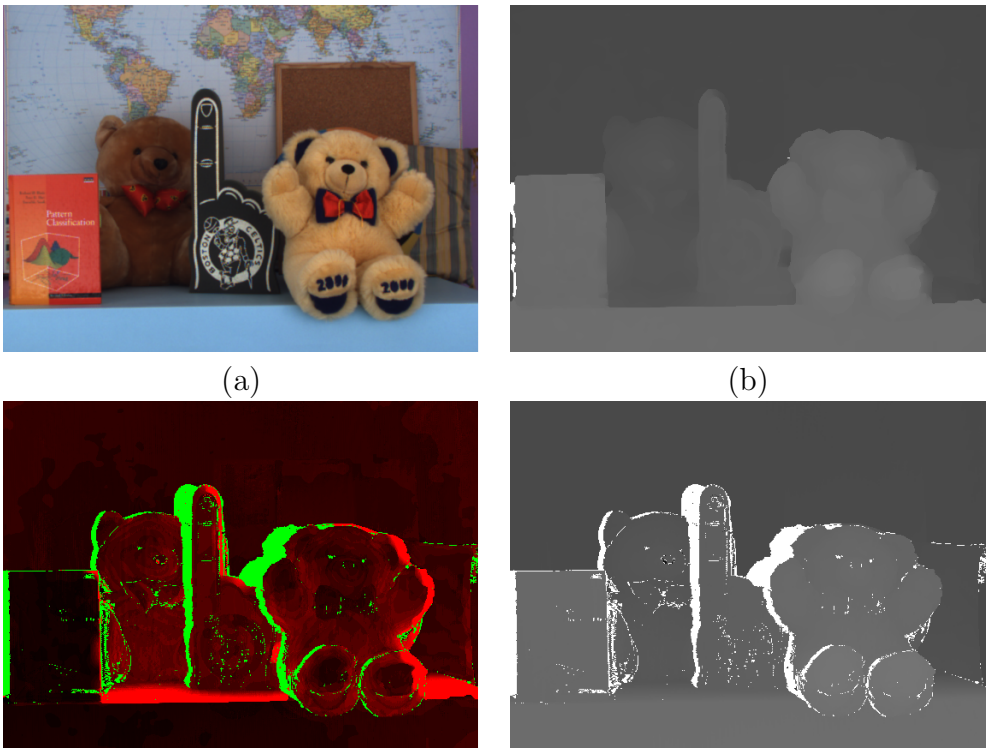
(c) (d)

Figura 4.4: Dataset 2: (a) immagine della telecamera sinistra; (b) mappa di disparità calcolata (i punti bianchi indicano disparità non assegnata); (c) errore della mappa di disparità, dove i punti rossi indicano l'errore moltiplicato per un fattore 5 e i punti verdi indicano aree occluse non considerate nel calcolo dell'errore; (d) ground truth.

I due casi presentati rappresentano due opposti. Il dataset 2 rispetta perfettamente l'ipotesi di superfici lisce a tratti e il sensore presenta difficoltà lungo i bordi degli oggetti, come già accennato, e sul tavolo, il quale è quasi parallelo alla direzione degli impulsi infrarossi del sensore, rendendo difficile la misura.

Il dataset 5, invece, contiene una scena con geometrie più complesse e con materiali meno riflettenti (la stoffa, la mano nera di gomma): anche in questo caso l'errore è presente lungo i bordi degli oggetti e risulta più marcato causa l'irregolarità dei bordi (si veda l'orsacchiotto a destra); tuttavia è presente anche sugli orsacchiotti e sulla mano nera in maniera diffusa e uniforme.

È degno di nota un effetto collaterale dell'interpolazione della mappa di



(a) (b)

(c) (d)

Figura 4.5: Dataset 5: (a) immagine della telecamera sinistra; (b) mappa di disparità calcolata (i punti bianchi indicano disparità non assegnata); (c) errore della mappa di disparità, dove i punti rossi indicano l'errore moltiplicato per un fattore 5 e i punti verdi indicano aree occluse non considerate nel calcolo dell'errore; (d) ground truth.

profondità: questa operazione genera valori di profondità (e, successivamente, anche di disparità) anche per punti appartenenti ad aree occluse, che per definizione non possono avere punti corrispondenti (si vedano le immagini (b) delle figure 4.4 e 4.5).

4.2.2 Valutazione disparità stereo

Nel caso della disparità stereo, innanzitutto è stato stabilito a priori un intervallo di disparità da 48 a 224 pixel, per un totale di 176 ipotesi possibili; il valore minimo è stato scelto causa la creazione dei cuscinetti neri in fase di rettificazione delle immagini (si veda figura 4.1).

I parametri su cui è stata posta maggior attenzione sono le penalità P_1 e P_2; riguardo alla dimensione del blocco, la dimensione pari a 5 si è dimostrata ottimale. I valori delle penalità sono stati scelti in maniera sperimentale: per P_1 è stato scelto sempre un valore proporzionale al numero di canali delle immagini e al quadrato della dimensione del blocco, mentre per P_2 è stato scelto sempre un multiplo di P_1. Le migliori performance sono state raggiunte con i seguenti valori: $P_1 = 1200$, $P_2 = 4 \cdot P1 = 4800$.

Nelle figure 4.6 e 4.7 sono presentati i risultati dell'algoritmo per gli stessi due casi analizzati nel paragrafo precedente.

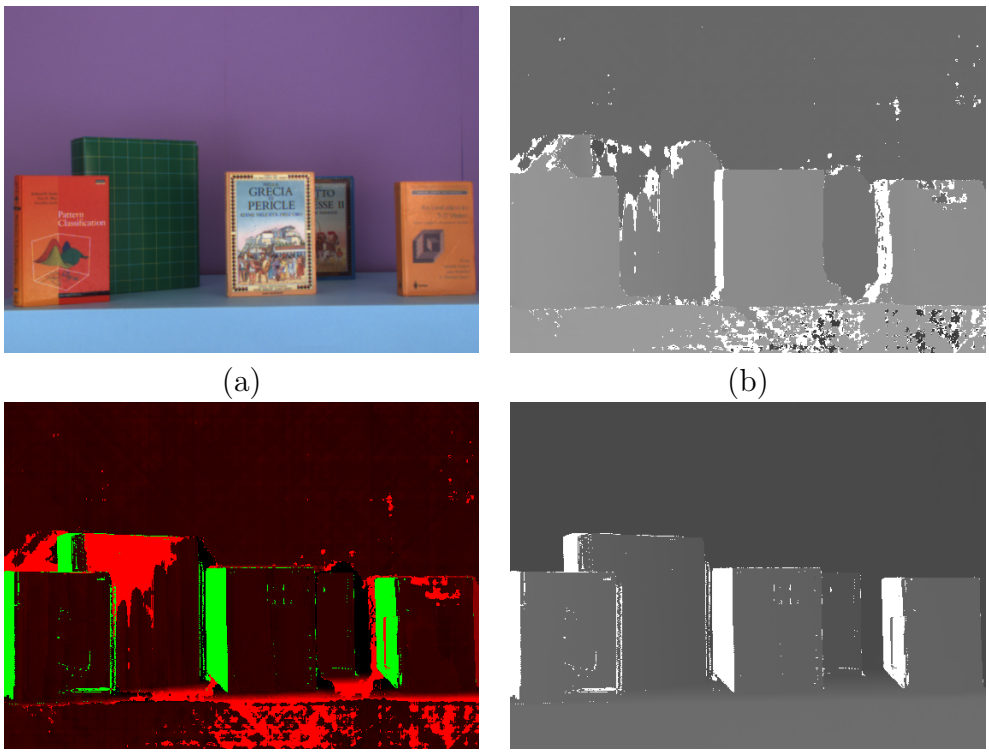


Figura 4.6: Dataset 2: (a) immagine della telecamera sinistra; (b) mappa di disparità calcolata (i punti bianchi indicano disparità non assegnata); (c) errore della mappa di disparità, dove i punti rossi indicano l'errore moltiplicato per un fattore 5 e i punti verdi indicano aree occluse non considerate nel calcolo dell'errore; (d) ground truth.

Anche per l'algoritmo stereo, i risultati ottenuti per le due scene sono piuttosto differenti. Il dataset 2 contiene molte aree senza texture, che creano molte difficoltà nella ricerca di punti corrispondenti. Inoltre, poiché l'algoritmo

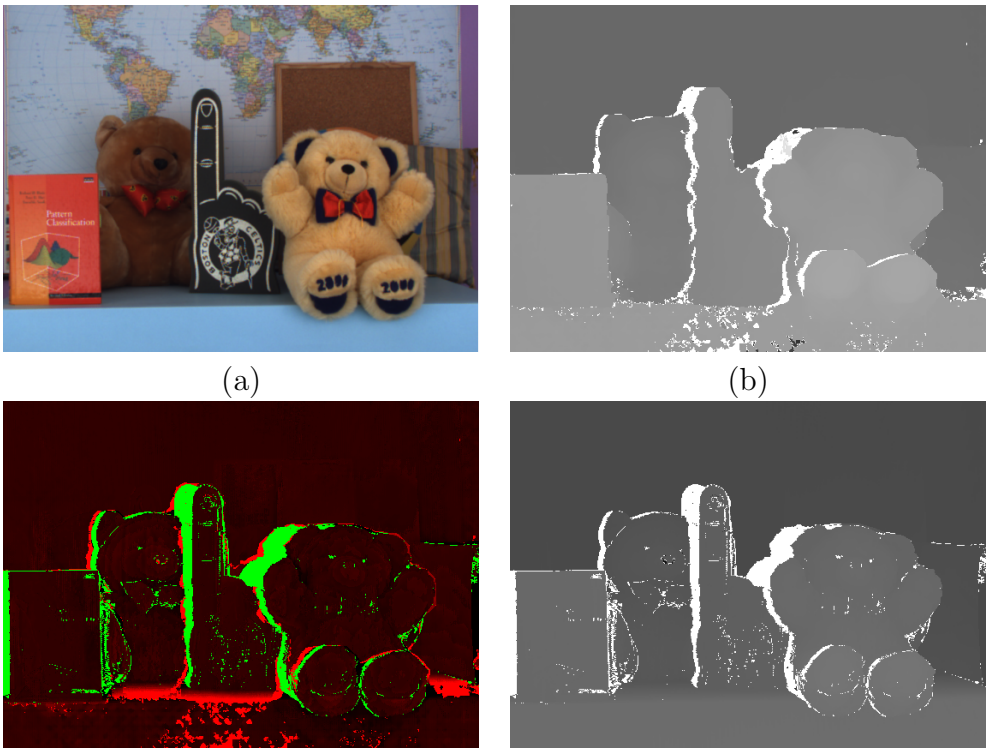


Figura 4.7: Dataset 5: (a) immagine della telecamera sinistra; (b) mappa di disparità calcolata (i punti bianchi indicano disparità non assegnata); (c) errore della mappa di disparità, dove i punti rossi indicano l'errore moltiplicato per un fattore 5 e i punti verdi indicano aree occluse non considerate nel calcolo dell'errore; (d) ground truth.

si basa sull'informazione del colore, oggetti con lo stesso colore (ad esempio, i due libri parzialmente arancioni agli estremi dell'immagine) possono generare disparità errate oppure mancate assegnazioni causa la violazione del vincolo di univocità individuata nel *left-right check*. Tuttavia, i bordi degli oggetti sono ben delineati: in tali aree infatti la texture risulta molto descrittiva e il calcolo per scanline si rivela preciso.

Riguardo al dataset 5, la presenza di molti oggetti genera maggiore texture, dando all'algoritmo maggiore informazione per stabilire corrispondenze: infatti, il MSE in questo caso è decisamente inferiore, così come la presenza di aree rosse nell'immagine (c) della figura 4.7. Anche in questo caso, sono generate mancate assegnazioni soprattutto nell'area azzurra con scarsa texture. Riguardo ai bordi, la precisione risulta minore: il contorno interno degli oggetti

è stato calcolato correttamente, al contrario del contorno esterno. Lungo le aree occluse, l'algoritmo non riesce ad assegnare valori di disparità per i punti adiacenti ad esse: la presenza di ombre comporta una minor discriminazione del colore, con conseguenti difficoltà nel trovare corrispondenze corrette.

4.2.3 Osservazioni

A confronto, i due metodi si comportano come era stato previsto nella descrizione dei sensori: il sensore a tempo di volo risulta piuttosto preciso, tuttavia genera un errore uniforme su superfici poco riflettenti ed un errore più marcato lungo i bordi degli oggetti; mentre il sistema stereo risulta più preciso nelle zone con texture e lungo i bordi degli oggetti (più sono netti, più è preciso), ma non riesce a gestire aree ripetitive o a tinta unita.

Possiamo notare come la disparità data dal sensore TOF sia generalmente più precisa rispetto al sistema stereo (tabella 4.1): l'errore dovuto ai materiali meno riflettenti o di colore scuro è limitato, mentre per l'algoritmo stereo vi sono molti punti con disparità non assegnata causa la scarsa presenza di texture.

Tuttavia, l'errore MSE indica che, per entrambi i sensori, la disparità per entrambi i sensori varia di circa ± 10 pixel: a causa del rapporto inversamente proporzionale tra distanza e disparità, si traduce in un errore di distanza marcato per valori piccoli di disparità, viceversa per valori elevati di disparità l'errore di distanza diminuisce. In figura 4.8 viene presentato un esempio per diversi valori di disparità con i relativi valori massimi e minimi avendo un MSE di 10 pixel.

	Dataset 1	Dataset 2	Dataset 3	Dataset 4	Dataset 5
MSE TOF	10.3274	9.99054	11.1169	10.4589	10.9976
MSE Stereo	12.9005	13.9302	11.6744	11.0265	10.1698

Tabella 4.1: Valori più bassi di MSE per le mappe di disparità ottenute dal sensore TOF e dall'algoritmo stereo.

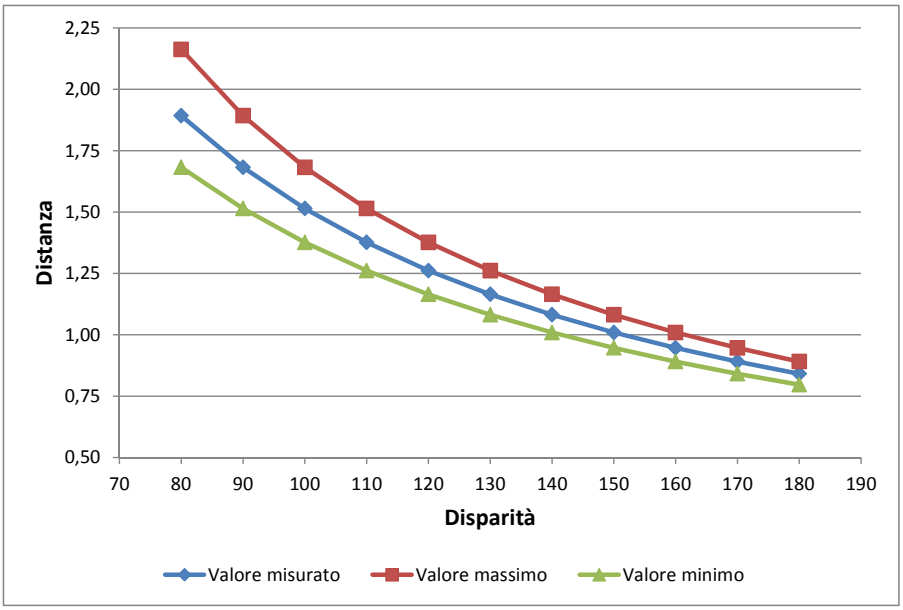


Figura 4.8: Valori massimi e minimi della distanza in metri avendo un errore MSE di disparità pari a 10 pixel.

4.3 Valutazione delle disparità combinate

4.3.1 Valutazione senza stime di errore

Entrambe le mappe di disparità hanno presentato errori in regioni diverse dell'immagine a seconda del sensore. Ci si aspetta che sia possibile unire le due mappe di disparità per ottenerne una più precisa: questa operazione viene attuata applicando la tecnica LC.

I parametri fondamentali di LC sono il raggio r del supporto e i tre coefficienti γ della plausibilità. Riguardo ai tre coefficienti, sono stati assunti pari a $\gamma_s = 14$, $\gamma_c = 15$ e $\gamma_t = 53$. In figura 4.9 è presente un grafico che evidenzia l'andamento del MSE al variare del raggio del supporto.

Possiamo notare come la dimensione del raggio saturi, così come era successo per il raggio della finestra di interpolazione per il sensore TOF. Per i tutti dataset utilizzati, dopo il valore 10 i miglioramenti sono esigui; nei test successivi, è stato utilizzato questo valore per il raggio del supporto.

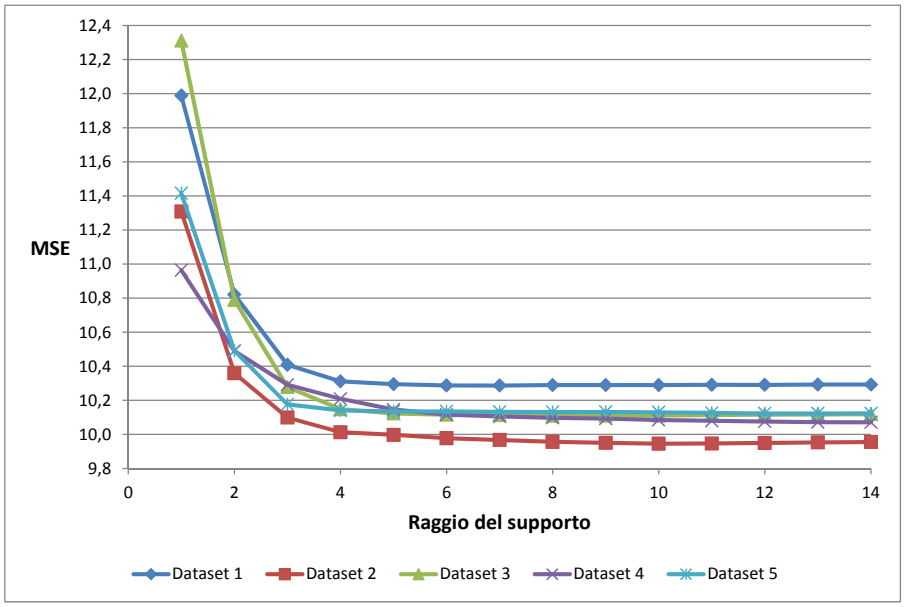


Figura 4.9: Andamento del MSE al variare del raggio del supporto di LC.

Nelle figure 4.10 e 4.11 sono analizzati gli stessi due casi visti precedentemente. Le mappe di disparità associate ai due sensori sono state calcolate con gli stessi parametri elencati nei paragrafi 4.2.1 e 4.2.2.

Osservando le immagini di errore del dataset 2 (figura 4.10), le due mappe di disparità iniziali sono state fuse con parziale successo: i bordi degli oggetti sono più delineati e l'errore presente sulla superficie del tavolo è diminuito. Tuttavia, la fusione ha propagato anche alcuni errori nella parte sinistra: nel caso del raccoglitore verde, la disparità data dal sensore TOF risultava molto precisa, tuttavia è stata in parte propagata quella errata dello stereo; lo stesso è accaduto sopra il primo libro da sinistra.

Nel caso del dataset 5 (figura 4.11), i bordi sono sempre meglio delineati dopo l'utilizzo di LC ed è diminuito l'errore uniforme presente su tutte le superfici. Anche in questo caso, alcuni errori delle mappe di disparità originali sono stati propagati, ma in misura quasi impercettibile: si nota leggermente sul bordo destro della mano nera e all'interno degli orsacchiotti.

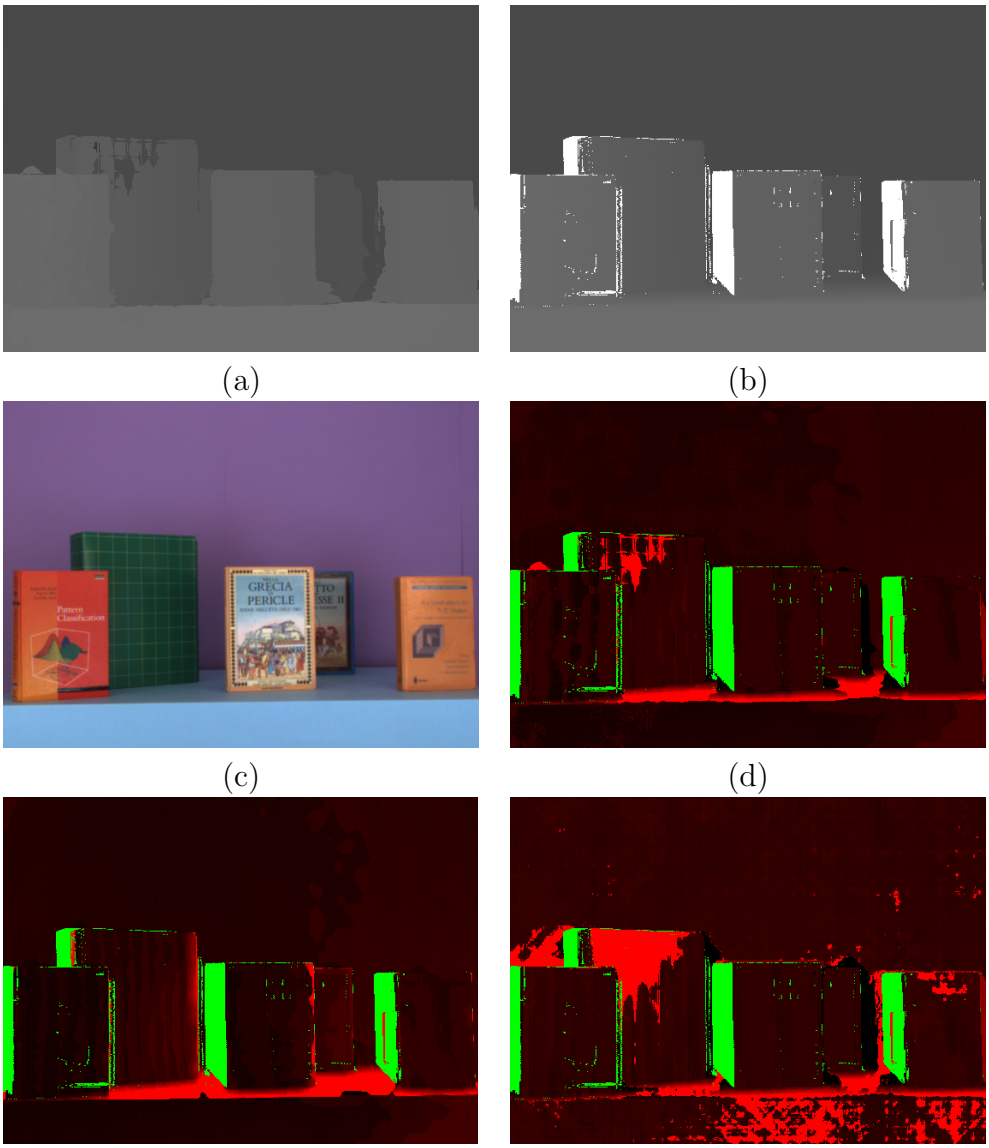


Figura 4.10: Dataset 2: (a) mappa di disparità calcolata con LC (i punti bianchi indicano disparità non assegnate); (b) ground truth (i punti bianchi indicano disparità non assegnate); (c) immagine della telecamera sinistra. Le immagini (d), (e), (f) sono errori delle mappe di disparità calcolate, dove i punti rossi indicano l'errore moltiplicato per un fattore 5 e i punti verdi indicano aree occluse non considerate nell'errore: (d) LC; (e) sensore TOF; (f) sistema stereo.

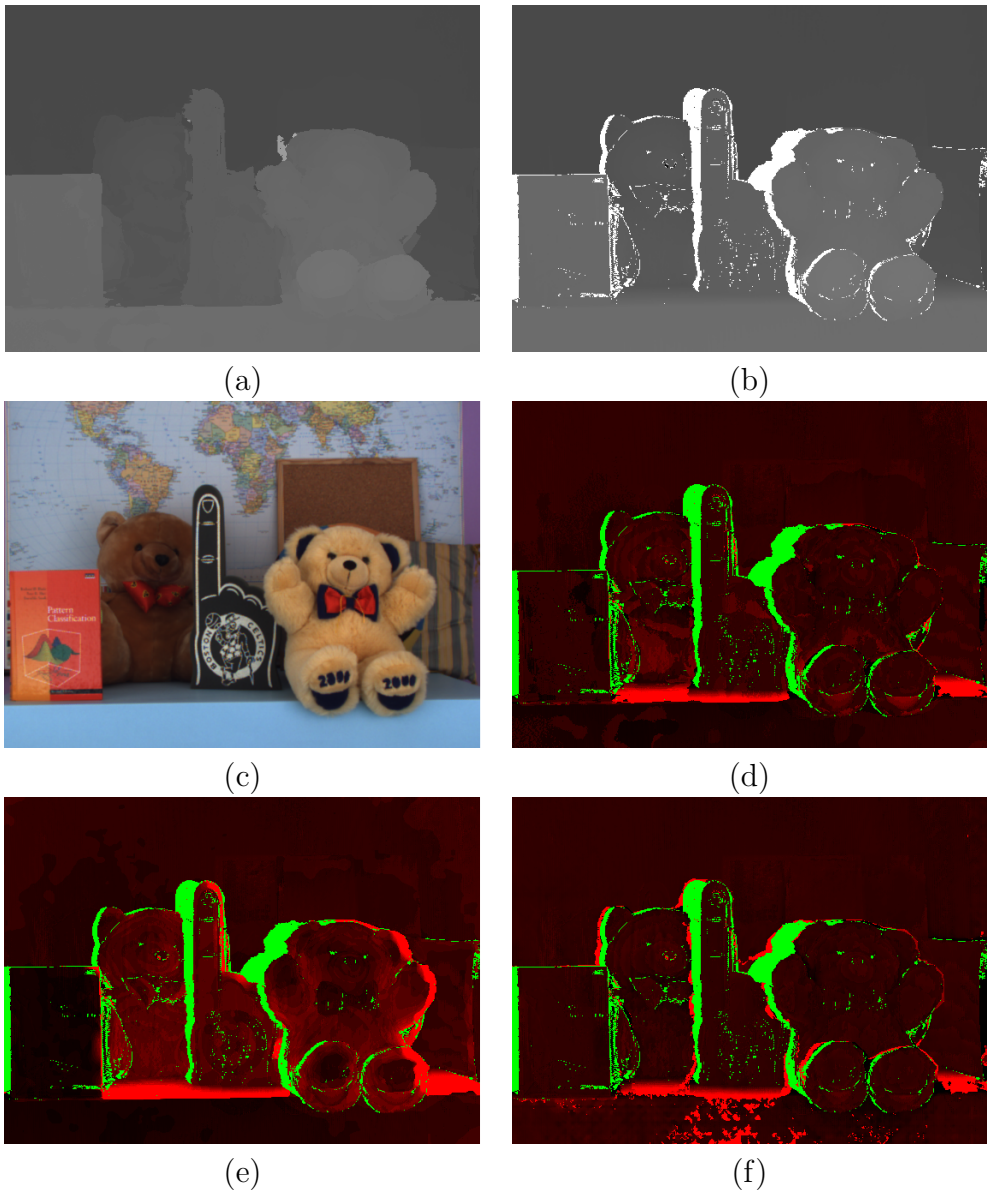


Figura 4.11: Dataset 5: (a) mappa di disparità calcolata con LC (i punti bianchi indicano disparità non assegnate); (b) ground truth (i punti bianchi indicano disparità non assegnate); (c) immagine della telecamera sinistra. Le immagini (d), (e), (f) sono errori delle mappe di disparità calcolate, dove i punti rossi indicano l'errore moltiplicato per un fattore 5 e i punti verdi indicano aree occluse non considerate nell'errore: (d) LC; (e) sensore TOF; (f) sistema stereo.

4.3.1.1 Osservazioni

In tabella 4.2 sono elencati i valori di MSE ottenuti. Analizzando i dataset (figura 4.2), i primi due rappresentano scene regolari e lineari con una presenza maggiore o minore di texture, mentre i restanti tre dataset rappresentano scene con geometrie più complesse da ricostruire e con più occlusioni tra oggetti. I risultati indicano che quando il sensore TOF fornisce un risultato piuttosto corretto, l'algoritmo stereo agisce solo come una conferma, lasciando pressoché invariato il risultato dell'altro sensore; quando il sensore TOF riprende scene più complesse, l'algoritmo stereo fornisce un contributo più significativo.

	Dataset 1	Dataset 2	Dataset 3	Dataset 4	Dataset 5
MSE TOF	10.3274	9.99054	11.1169	10.4589	10.9976
MSE Stereo	12.9005	13.9302	11.6744	11.0265	10.1698
MSE LC	10.3733	10.0065	10.3113	10.1333	10.1568
Var. TOF	0.44%	0.16%	−7.25%	-3.11%	-7.65%
Var. Stereo	−19.59%	−28.17%	−11.68%	−8.10%	−0.13%

Tabella 4.2: Valori più bassi di MSE per le mappe di disparità ottenute dal sensore TOF, dall'algoritmo stereo e da LC; per l'algoritmo stereo, il calcolo del MSE non considera eventuali punti senza disparità (sono meno del 4.5%). Le ultime due righe indicano, per ogni dataset, le variazioni percentuali di MSE dell'algoritmo LC rispetto all'errore del sensore indicato.

Come già accennato, l'interpolazione della mappa di profondità per il sensore TOF assegna valori di profondità a tutti i punti, generando successivi valori di disparità anche a punti appartenenti ad aree occluse: nel caso di LC, questo particolare potrebbe aver perturbato il risultato, poiché sono state propagate ipotesi errate di disparità. Per l'algoritmo stereo, le occlusioni vengono generalmente eliminate grazie al *left-right check*, tuttavia per il sensore TOF non è possibile eseguire questa tecnica. Per mitigare in parte questo effetto collaterale, è stato posto un limite inferiore al numero di punti già aventi un valore di profondità contenuti dalla finestra di interpolazione: se il numero è inferiore, l'interpolazione per il punto in esame non viene eseguita. Questo valore sarà identificato con il simbolo N_{occ}.

Un valore di riferimento è dato dal seguente rapporto:

$$N_{occ} = \#\text{ punti finestra interpolazione} \times \frac{\text{risoluzione TOF}}{\text{risoluzione telecamera}}$$

Con i parametri del sistema di acquisizione e il valore 10 per il raggio della finestra di interpolazione, si ottiene:

$$N_{occ} = (10+1)^2 \times \frac{176 \times 144}{1032 \times 778} = 13.92 \approx 14$$

Nella figura 4.12 è mostrato l'andamento dell'errore MSE della mappa di disparità calcolata da LC al variare di N_{occ}. Si può notare come il valore di questo parametro incida sulla precisione della mappa di disparità, soprattutto per i dataset 3 e 5. La differenza principale tra i due gruppi di dataset è la combinazione di due fattori: una geometria più complessa e la presenza di determinati materiali e colori. Per i dataset 1 e 2, la combinazione presente è a favore del sensore; anche nel caso del dataset 4 la combinazione è a favore: la maggior parte degli oggetti sono regolari e di materiale adatto alle misurazioni del sensore TOF (soltanto l'orsacchiotto incide negativamente sulla precisione). Per i dataset 3 e 5 invece la combinazione è decisamente sfavorevole: tonalità di colore più scure, materiali soprattutto poco riflettenti e maggiori ombreggiature causa occlusioni mettono in evidenza la scarsa risoluzione del sensore e la difficoltà nel fornire misure precise. In figura 4.13 vi è il confronto fra due mappe di disparità associate al sensore con diverso valore per N_{occ}.

Assegnando il valore 18 a questo nuovo parametro, il cambiamento della mappa di disparità del sensore TOF determina i valori di MSE in tabella 4.3.

	Dataset 1	Dataset 2	Dataset 3	Dataset 4	Dataset 5
MSE (1)	10.3733	10.0065	10.3113	10.1333	10.1568
MSE (18)	10.3416	9.9554	10.0883	10.0996	10.0884
Var. %	-0.31%	-0.51%	-2.16%	-0.33%	-0.67%

Tabella 4.3: Valori di MSE della mappa di disparità fornita da LC variando il parametro $N_{occ} \in \{1, 18\}$ (i valori sono messi fra parentesi). La variazione percentuale indica il miglioramento rispetto al calcolo con $N_{occ} = 1$.

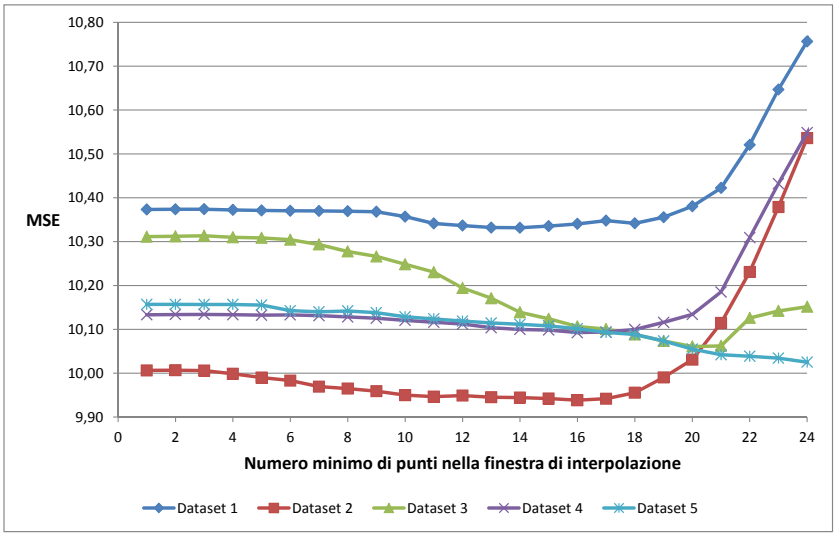
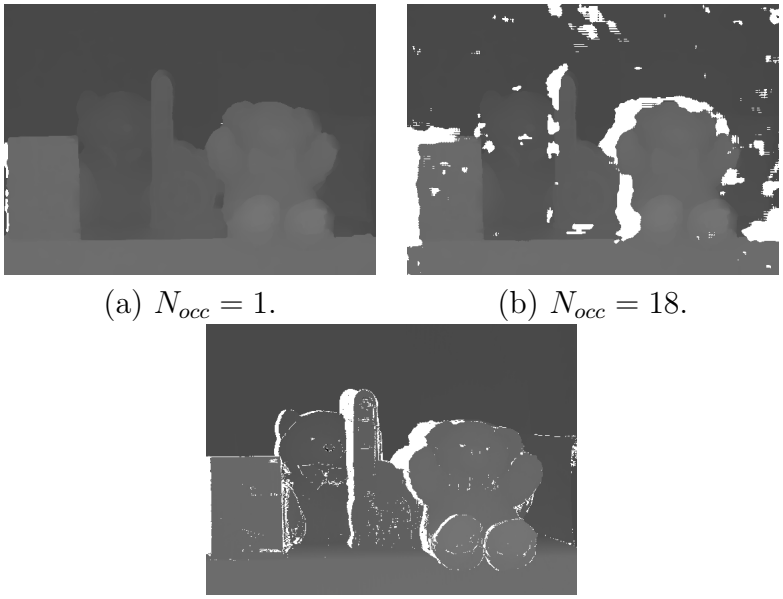


Figura 4.12: Andamento del MSE per la mappa di disparità calcolata con LC al variare del parametro N_{occ} per la mappa di disparità del sensore TOF. $N_{occ} = 18$ determina un errore MSE medio minore.



(a) $N_{occ} = 1$. (b) $N_{occ} = 18$.

(c) Ground truth.

Figura 4.13: Dataset 5: esempio di mappe di disparità calcolate per il sensore TOF variando il valore di N_{occ}. Si può notare la maggiore preservazione delle occlusioni (punti di colore bianco).

4.3.2 Valutazione utilizzando le stime di errore

La fusione delle due mappe di disparità è stata eseguita considerandole entrambe equamente plausibili fino ad ora. Tuttavia, questa assunzione può causare la propagazione di errori nel risultato, come è successo per il dataset 5 (figura 4.11).

Come risulta evidente dagli svantaggi dei sensori, le ipotesi di disparità fornite non possono essere considerate equamente plausibili. Risulta necessario pesare i contributi dei due sensori: un metodo per ottenere questa funzionalità consiste nell'utilizzare le stime di errore precedentemente definite. Nel paragrafo 3.2 la funzione di plausibilità è stata ridefinita per adattare LC a questi nuovi parametri.

Per i test seguenti è stato assunto $N_{occ} = 1$. Come già detto, questo parametro agisce come un filtro sul numero di punti che possono avere una disparità ricavata dalla misura del sensore TOF, tuttavia può scartare alcuni punti che potrebbero essere interpolati correttamente o con pochissimo errore. L'utilizzo della stima dell'errore permette di raggiungere un risultato migliore, essendo più precisa della semplice sogliatura.

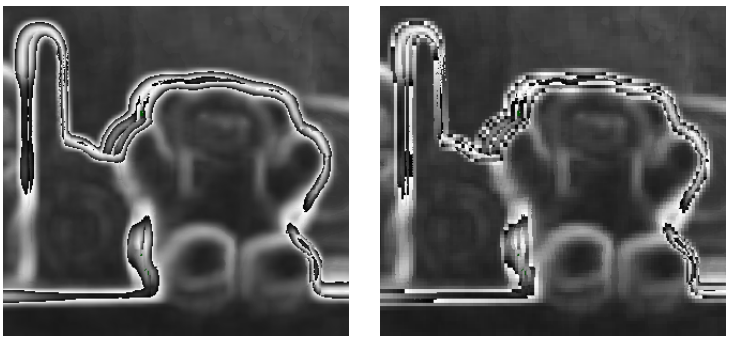
Come visto nel paragrafo 2.3.1, vi sono 3 parametri per la stima dell'errore per il sensore TOF: il metodo di calcolo per la deviazione standard della profondità, la dimensione della finestra e la fusione tra l'incertezza dell'ampiezza e la deviazione standard della profondità. Saranno presentate soltanto le differenze visive più evidenti degli errori. Riguardo al sistema stereo, la stima dell'errore non necessità di parametri.

4.3.2.1 Stime di errore

Nelle figure seguenti sono presentati alcuni particolari del dataset 5, al fine di evidenziare alcune differenze visive della stima di errore per il sensore TOF al variare dei parametri visti.

Riguardo il metodo di calcolo della deviazione standard della profondità (figura 4.14), i metodi con finestra pari e dispari forniscono risultati simili con l'errore distribuito "a patch" poiché vi sono gruppi di punti con lo stesso errore causa la riproiezione; mentre il metodo basato su interpolazione bilineare determina una distribuzione dell'errore senza particolari artefatti. Si può

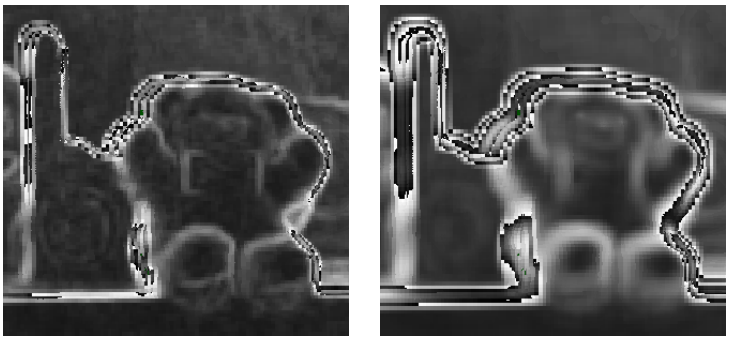
notare in entrambi i casi la concentrazione dell'errore lungo i bordi degli oggetti.



(a) Interpolazione. (b) Finestra dispari.

Figura 4.14: Stime dell'errore Δ^T a confronto calcolate con finestra di dimensione 5 e fusione tramite somma.

Analizzando la figura 4.15, finestre di dimensione minore comportano una stima di errore più localizzata, con picchi più evidenti lungo i bordi; nel caso di finestre di dimensione maggiore, invece, i valori di errore risultano più diffusi e uniformi. Statisticamente, una finestra di dimensione minore garantisce valori di stima più plausibili: una varianza elevata in una piccola regione della mappa di disparità può essere una prova evidente della presenza di errori di disparità.

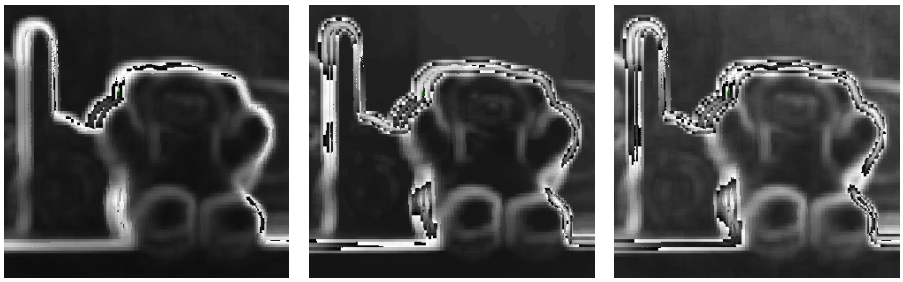


(a) Dimensione 3. (b) Dimensione 7.

Figura 4.15: Stime dell'errore Δ^T a confronto calcolate con il metodo della finestra dispari e fusione tramite somma.

La fusione degli errori (figura 4.16), invece, comporta effetti differenti: la media aritmetica e la somma generano valori simili a meno di un fattore di scala. L'operazione di massimo, invece, mette in evidenza i picchi delle stime

di errore; inoltre, per definizione, può approssimare le altre due operazioni di fusione: a seconda dei due valori confrontati, se la distanza in valore assoluto tende a zero, il massimo approssima la media; se la distanza in valore assoluto tende ad uno dei due valori, allora il massimo approssima la somma.



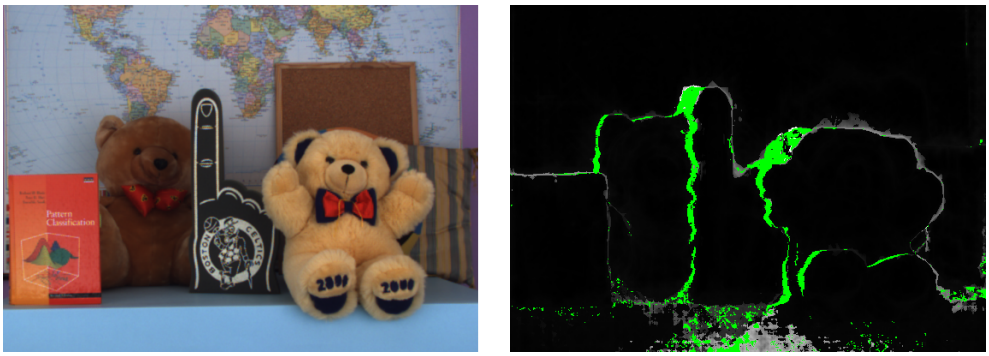
(a) Media aritmetica. (b) Massimo. (c) Somma.

Figura 4.16: Stime dell'errore Δ^T a confronto calcolate con il metodo della finestra dispari e finestra di dimensione 3. L'operazione di massimo genera valori intermedi rispetto alle altre due operazioni di fusione.

Riguardo alla stima dell'errore per l'algoritmo stereo, nelle figure 4.17 e 4.18 viene presentata un'immagine visiva per i dataset 2 e 5. La funzione utilizzata conferisce una stima coerente con le regioni di errore per l'algoritmo: regioni con scarsa texture (dataset 2) o con ombre e colori poco definiti (dataset 5) conseguono un errore maggiore.

(a) (b)

Figura 4.17: Stima dell'errore per l'algoritmo stereo. Dataset 2: (a) immagine della telecamera sinistra; (b) stima dell'errore. I punti verdi indicano punti senza disparità e, quindi, senza stima di errore.



(a) (b)

Figura 4.18: Stima dell'errore per l'algoritmo stereo. Dataset 5: (a) immagine della telecamera sinistra; (b) stima dell'errore. I punti verdi indicano punti senza disparità e, quindi, senza stima di errore.

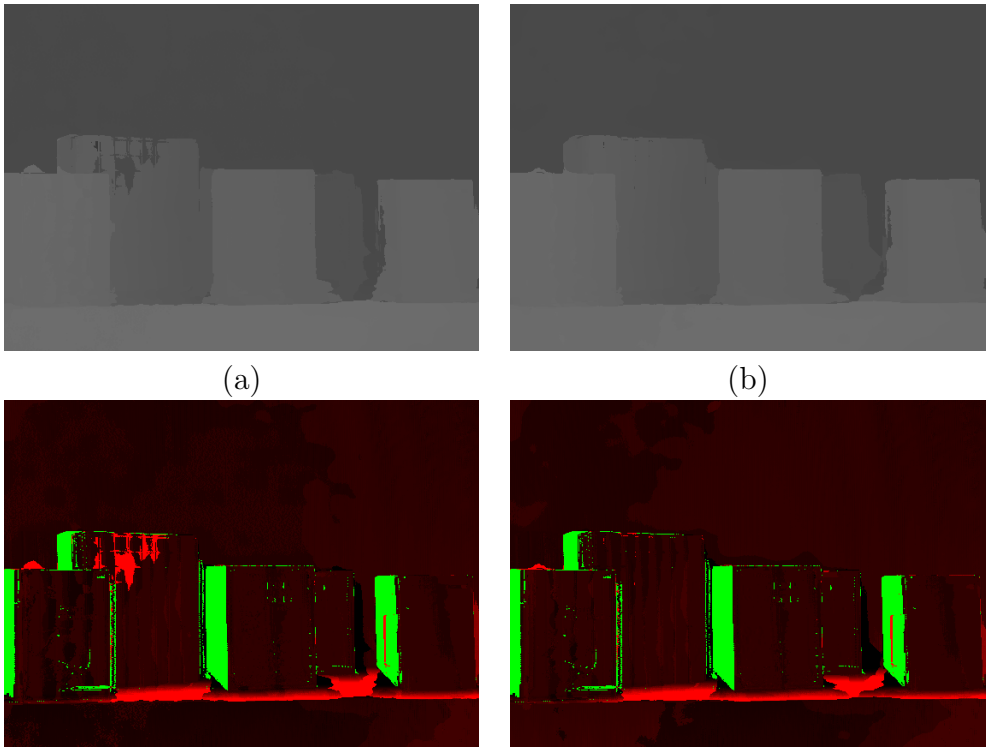
4.3.2.2 Errore gaussiano ed errore esponenziale

Il primo gruppo di test è stato eseguito utilizzando, per il sensore TOF, la plausibilità comprendente il termine moltiplicativo basato su una distribuzione gaussiana. Di fatto, questa stima di errore quantifica il rumore casuale presente nelle misure del sensore: la presenza di un errore elevato in regioni senza cambiamenti di disparità può indicare la scarsa ripetitività delle misure del sensore TOF, assieme all'elevato rumore in fase di acquisizione.

Nelle figure 4.19 e 4.20 sono presentati risultati ottenuti con LC utilizzando le nuove funzioni di plausibilità.

I miglioramenti più evidenti si notano sul dataset 2. Infatti, l'errore presente nella regione di sinistra è stato corretto completamente. L'errore presente sul tavolo azzurro, invece, è stato leggermente diminuito. Considerando il dataset 5, l'errore uniforme è mediamente diminuito: la funzione gaussiana ha agito come un filtro di riduzione del rumore dovuto alle misurazioni del sensore TOF.

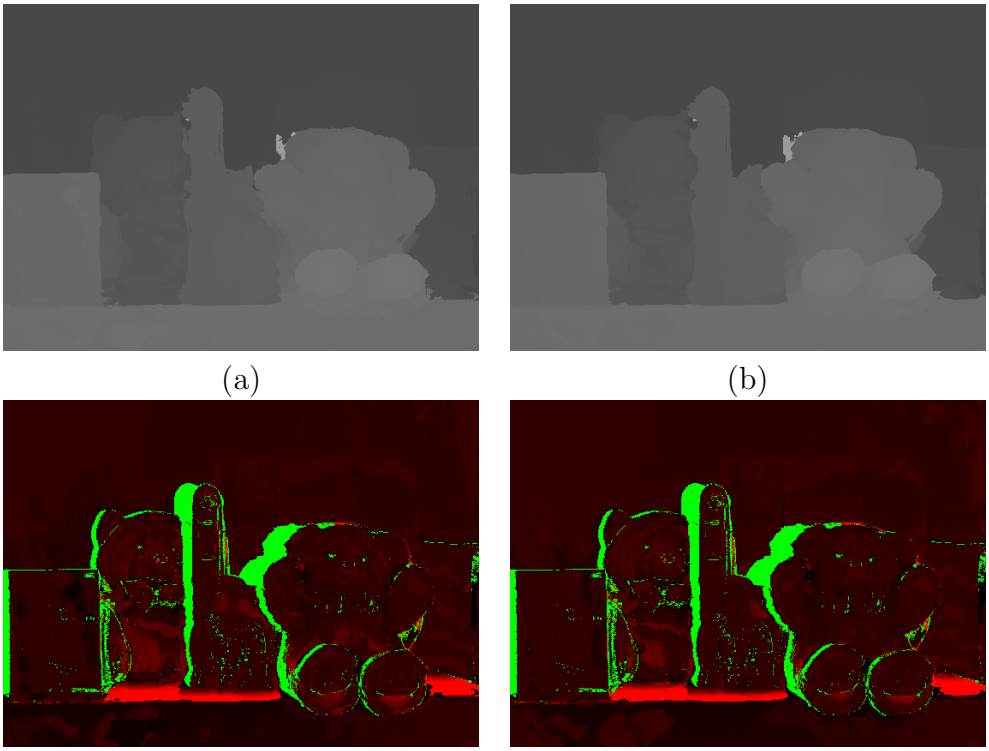
Il risultato migliore è stato ottenuto per $\gamma_{es} = 0.135$ e per una stima di errore per il sensore TOF calcolata con il metodo della finestra dispari di dimensione 3 e fusione delle due incertezze tramite massimo.



(a) (b)

(c) (d)

Figura 4.19: Dataset 2: (a) mappa di disparità calcolata senza stime di errore; (b) mappa di disparità calcolata con stime di errore; (c) errore della mappa di disparità calcolata senza stime di errore; (d) errore della mappa di disparità calcolata con stime di errore. Nelle immagini (c) e (d) i punti rossi indicano l'errore moltiplicato per un fattore 5 e i punti verdi indicano aree occluse non considerate nel calcolo dell'errore.



(a) (b)

(c) (d)

Figura 4.20: Dataset 5: (a) mappa di disparità calcolata senza stime di errore; (b) mappa di disparità calcolata con stime di errore; (c) errore della mappa di disparità calcolata senza stime di errore; (d) errore della mappa di disparità calcolata con stime di errore. Nelle immagini (c) e (d) i punti rossi indicano l'errore moltiplicato per un fattore 5 e i punti verdi indicano aree occluse non considerate nel calcolo dell'errore.

4.3.2.3 Errori esponenziali

Questo gruppo di test, invece, è stato eseguito utilizzando il termine moltiplicativo esponenziale per entrambe le stime di errore nella formula della plausibilità.

Nelle figure 4.21 e 4.22 sono presenti i risultati dei test per gli stessi due dataset.

Nel caso del dataset 2, l'utilizzo della stima di errore ha ridotto l'errore sulla zona del tavolo ed ha corretto un'errata propagazione di disparità: l'errore presente sui primi due libri da sinistra è stato sensibilmente ridotto passando dall'immagine (c) all'immagine (d). Riguardo al dataset 5, si nota soprattutto la riduzione di errore sul tavolo azzurro; in misura minore è stato ridotto anche l'errore uniforme presente sulle superfici è stato ridotto.

I risultati migliori sono stati ottenuti per $\gamma_{et} = 0.006$ e $\gamma_{es} = 0.004$: i due parametri sono piuttosto simili in valore, tuttavia la differenza indica che i valori del sensore TOF sono ritenuti più attendibili dell'algoritmo stereo, come ci si aspettava mediamente. Gli ulteriori parametri per il calcolo della stima di errore del sensore TOF sono i seguenti: metodo della finestra dispari di dimensione 5 e fusione delle due incertezze tramite somma.

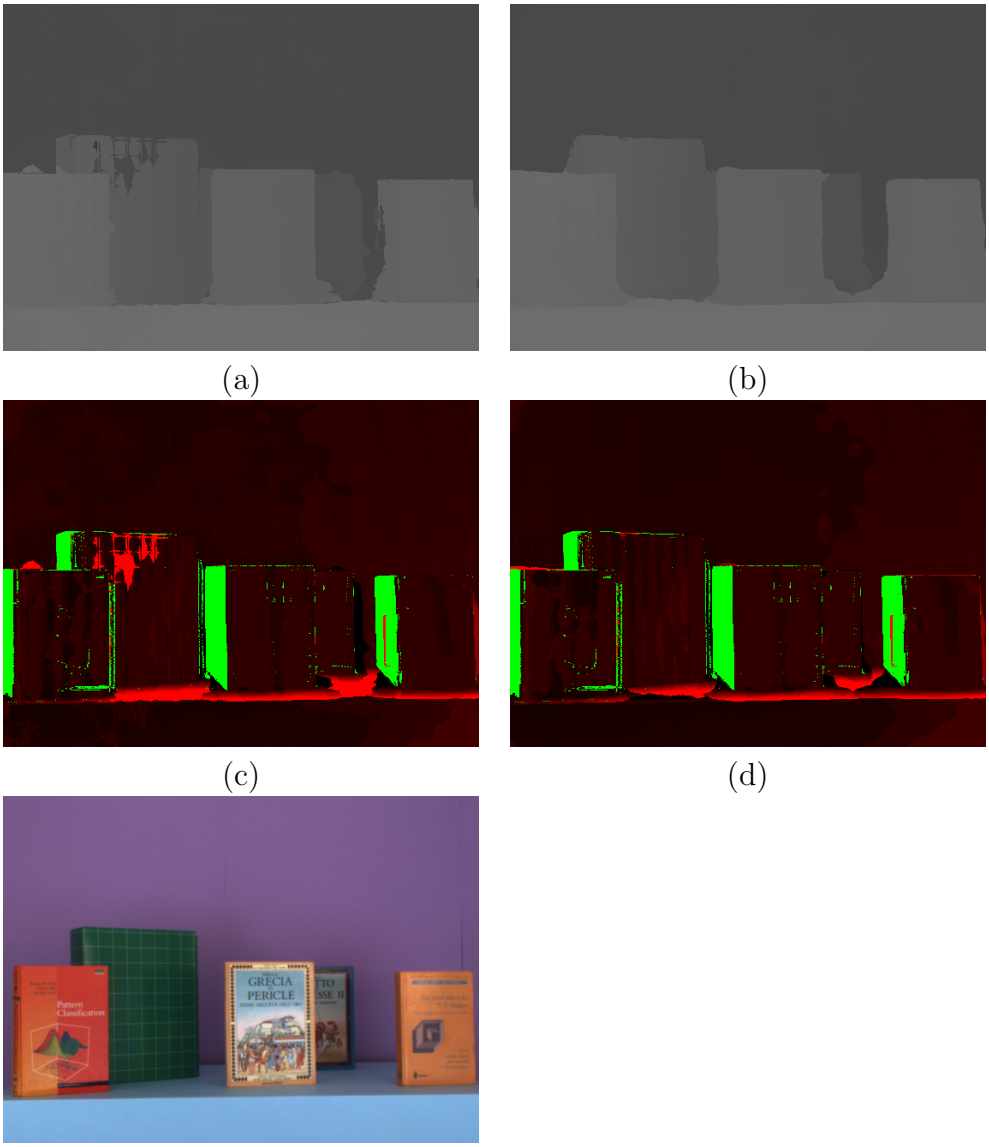
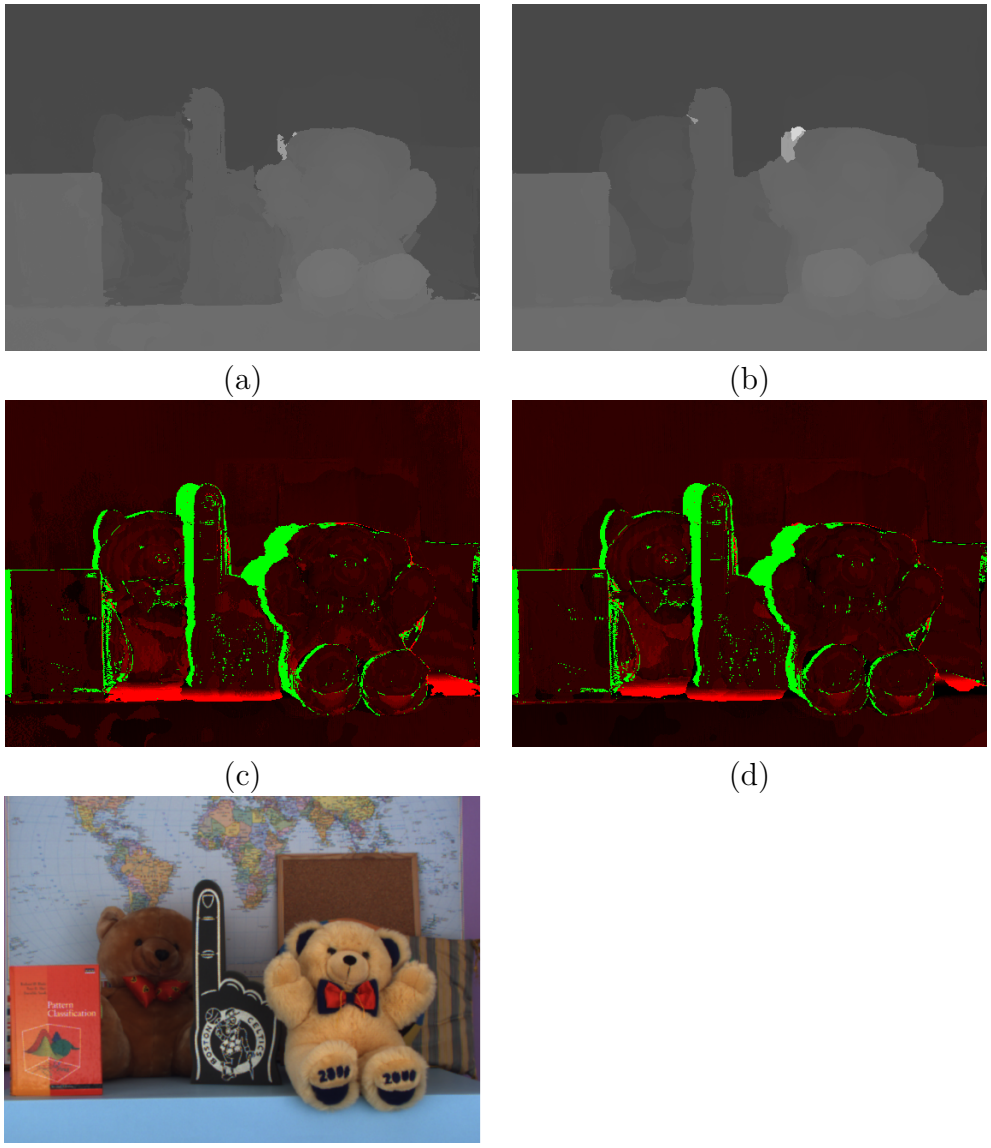


Figura 4.21: Dataset 2: (a) mappa di disparità calcolata senza stime di errore; (b) mappa di disparità calcolata con stime di errore; (c) errore per la mappa di disparità in (a); (d) errore per la mappa di disparità in (b); (e) immagine della telecamera sinistra. Nelle immagini (c) e (d) i punti rossi indicano l'errore moltiplicato per un fattore 5 e i punti verdi indicano aree occluse non considerate nel calcolo dell'errore.



(a) (b) (c) (d) (e)

Figura 4.22: Dataset 5: (a) mappa di disparità calcolata senza stime di errore; (b) mappa di disparità calcolata con stime di errore; (c) errore per la mappa di disparità in (a); (d) errore per la mappa di disparità in (b); (e) immagine della telecamera sinistra. Nelle immagini (c) e (d) i punti rossi indicano l'errore moltiplicato per un fattore 5 e i punti verdi indicano aree occluse non considerate nel calcolo dell'errore.

4.3.2.4 Considerazioni

Entrambi i gruppi di test hanno evidenziato miglioramenti significativi nella riduzione dell'errore. Il secondo metodo basato su due funzioni esponenziali caratterizzate rispettivamente da γ_{et} e γ_{es} ha fornito risultati più accurati considerando l'errore MSE (tabella 4.4). L'unico vantaggio della stima di errore basata sulla distribuzione gaussiana per il sensore TOF rispetto alla sua alternativa è l'assenza del calcolo di γ_{et}, facilitando l'utilizzo dell'algoritmo.

	Dataset 1	Dataset 2	Dataset 3	Dataset 4	Dataset 5
MSE TOF	10.3274	9.99054	11.1169	10.4589	10.9976
MSE Stereo	12.9005	13.9302	11.6744	11.0265	10.1698
MSE LC	10.3733	10.0065	10.3113	10.1333	10.1568
MSE LC (a)	10.2336	9.9153	10.1915	9.9181	10.1662
Var. %	-1.35%	-0.91%	-1.16%	-2.12%	0.09%
MSE LC (b)	9.9003	9.4723	9.7556	9.6488	9.9561
Var. %	-4.56%	-5.34%	-5.39%	-4.78%	-1.98%

Tabella 4.4: Valori più bassi di MSE per tutte le mappe di disparità. Il caso (a) è stato calcolato utilizzando la stima di errore gaussiana per il sensore TOF e con funzione esponenziale di parametro $\gamma_{es} = 0.135$ per l'algoritmo stereo; il caso (b), invece, è stato calcolando utilizzando le stime di errore con funzioni esponenziali per entrambi i sensori con parametri rispettivamente $\gamma_{et} = 0.006$ e $\gamma_{es} = 0.004$. Le variazioni percentuali sono calcolate rispetto all'errore MSE dato da LC senza utilizzare le stime di errore.

Capitolo 5

Conclusioni

La ricostruzione tridimensionale è un ambito di ricerca molto studiato e seguito da molte aziende. Questo lavoro ha fornito un ulteriore contributo alla letteratura già esistente: è stato sviluppato e valutato un sistema per la ricostruzione tridimensionale di una scena a partire dai dati forniti da due diversi tipi di sensori, ovvero un sensore a tempo di volo e due telecamere stereo. A corredo, è stato implementato un software per la valutazione delle prestazioni del sistema.

Il miglioramento delle prestazioni combinando i sensori rispetto all'utilizzo singolo ha messo in evidenza le loro caratteristiche e la complementarietà delle stesse.

Avanzamenti futuri nel progetto comprendono sicuramente il miglioramento dell'implementazione software per eseguire le operazioni in tempo reale. A questo scopo, sarà necessario utilizzare hardware più performante per eseguire le operazioni richieste: il GPU computing sarà sicuramente uno strumento adatto. Inoltre, la definizione di una plausibilità ad hoc per i dati del sensore a tempo di volo potrebbe fornire ulteriori miglioramenti nelle prestazioni del sistema.

Dal punto di vista della ricerca, una diversa combinazione di sensori potrebbe garantire risultati simili o migliori. In commercio sono presenti altri sensori basati su tecnologie simili a quella *Time-of-Flight*, in particolare Microsoft Kinect [16].

Ringraziamenti

Un ringraziamento particolare è rivolto ai miei genitori, Emilio e Maria Pia, che mi hanno sostenuto sempre e in ogni modo durante questi anni di studi. Ringrazio anche i colleghi universitari per l'amicizia e l'aiuto nello studio. Un ringraziamento anche agli amici di lunga data Rita, Matteo e Roberto, organizzatori di serate nonché di scherzi e festeggiamenti per questo traguardo.

Ultimo, ma non meno importante, un ringraziamento al relatore Pietro Zanuttigh per questa occasione di crescita culturale, per la disponibilità e per la costante presenza durante lo sviluppo di questo lavoro.

Bibliografia

[1] Openmp. URL http://openmp.org/.

[2] Basler AG, . URL http://www.baslerweb.com/.

[3] MESA Imaging AG, . URL http://www.mesa-imaging.ch/.

[4] Stan Birchfield and Carlo Tomasi. Depth discontinuities by pixel-to-pixel stereo. *Int. J. Comput. Vision*, 35(3):269–293, December 1999. ISSN 0920-5691. URL http://dx.doi.org/10.1023/A:1008160311296.

[5] G. Bradski. The OpenCV Library. *Dr. Dobb's Journal of Software Tools*, 2000. URL http://opencv.willowgarage.com.

[6] Dorin Comaniciu and Peter Meer. Mean shift: A robust approach toward feature space analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(5):603–619, May 2002. ISSN 0162-8828. URL http://dx.doi.org/10.1109/34.1000236.

[7] Carlo Dal Mutto, Pietro Zanuttigh, and Guido Maria Cortelazzo. A probabilistic approach to tof and stereo data fusion. In *3DPVT*, Paris, France, May 2010.

[8] J. Diebel and S. Thrun. An application of markov random fields to range sensing. In *Proceedings of Conference on Neural Information Processing Systems (NIPS)*, Cambridge, MA, 2005. MIT Press.

[9] Andrea Fusiello. *Visione Computazionale - appunti delle lezioni.* pubblicato a cura dell'autore, Giugno 2008.

[10] Valeria Garro, Carlo Dal Mutto, Pietro Zanuttigh, and Guido M. Cortelazzo. A novel interpolation scheme for range data with side information. In *Proceedings of the 2009 Conference for Visual Media Production*, CVMP '09, pages 52–60, Washington, DC, USA,

2009. IEEE Computer Society. ISBN 978-0-7695-3893-8. URL http://dx.doi.org/10.1109/CVMP.2009.26.

[11] Sigurjon Arni Gudmundsson, Henrik Aanaes, and Rasmus Larsen. Fusion of stereo vision and timeofflight imaging for improved 3d estimation. *Int. J. Intell. Syst. Technol. Appl.*, 5(3/4):425–433, November 2008. ISSN 1740-8865. doi: 10.1504/IJISTA.2008.021305. URL http://dx.doi.org/10.1504/IJISTA.2008.021305.

[12] Heiko Hirschmuller. Stereo processing by semiglobal matching and mutual information. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 30(2):328–341, February 2008. ISSN 0162-8828. URL http://dx.doi.org/10.1109/TPAMI.2007.1166.

[13] Johannes Kopf, Michael F. Cohen, Dani Lischinski, and Matt Uyttendaele. Joint bilateral upsampling. In *ACM SIGGRAPH 2007 papers*, SIGGRAPH '07, New York, NY, USA, 2007. ACM. URL http://doi.acm.org/10.1145/1275808.1276497.

[14] Stefano Mattoccia. A locally global approach to stereo correspondence. In *IEEE Workshop on 3D Digital Imaging and Modeling (3DIM2009)*, pages 1763–1770. IEEE, October 3-4, 2009.

[15] Stefano Mattoccia. Fast locally consistent dense stereo on multicore. In *Sixth IEEE Embedded Computer Vision Workshop (ECVW2010)*. IEEE, June 13, 2010.

[16] Microsoft. URL http://www.xbox.com/en-US/Kinect.

[17] Andreas Nüchter, Oliver Wulf, and Kai Lingemann. Fusion of stereo-camera and pmd-camera data for real-time suited precise, 2006.

[18] Daniel Scharstein and Richard Szeliski. A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. *Int. J. Comput. Vision*, 47(1-3):7–42, April 2002. ISSN 0920-5691. URL http://dx.doi.org/10.1023/A:1014573219977.

[19] Olga Veksler. *Efficient graph-based energy minimization methods in computer vision*. PhD thesis, Ithaca, NY, USA, 1999. AAI9939932.

[20] Qingxiong Yang, Kar han Tan, Bruce Culbertson, and John Apostolopoulos. Fusion of active and passive sensors for fast 3d capture, .

[21] Qingxiong Yang, Ruigang Yang, James Davis, and David Nistér. Spatial-depth super resolution for range images. In *In CVPR, 2007. 8*, .

[22] Li Zhang, Brian Curless, and Steven M. Seitz. Spacetime stereo: Shape recovery for dynamic scenes. In *CVPR (2)*, pages 367–374. IEEE Computer Society, 2003. ISBN 0-7695-1900-8. URL `http://dblp.uni-trier.de/db/conf/cvpr/cvpr2003-2.html#ZhangCS03`.

[23] Jiejie Zhu, Liang Wang, Ruigang Yang, James E. Davis, and Zhigeng pan. Reliability fusion of time-of-flight depth and stereo geometry for high quality depth maps. *IEEE Trans. Pattern Anal. Mach. Intell.*, 33(7): 1400–1414, July 2011. ISSN 0162-8828. doi: 10.1109/TPAMI.2010.172. URL `http://dx.doi.org/10.1109/TPAMI.2010.172`.