

# UNIVERSITÀ DI PISA



FACOLTÀ DI INGEGNERIA

CORSO DI LAUREA SPECIALISTICA  
INGEGNERIA DELLE TELECOMUNICAZIONI

TESI DI LAUREA

**Tecniche di mappatura dell'ambiente subacqueo  
a mezzo di sensori ottici e acustici**

Il candidato:

Alessandro Rossini

Il relatore:

Prof. Ruggero Reggiannini

I correlatori:

Prof. Filippo Giannetti

Dott. Ovidio Salvetti

Ing. Lavinio Gualdesi

*Anno Accademico 2011 – 2012*

Ringrazio tutta la mia famiglia per il supporto duraturo negli anni,  
il Dott. Ovidio Salvetti per l'occasione a me concessa,  
i relatori per i consigli preziosi e gli spunti di riflessione,  
il personale del *SILab* per l'indispensabile compagnia.

# Sommario

Premessa.....	5
Capitolo 1 - Sensori Ottici .....	7
Introduzione.....	7
Sensori di Visione.....	10
La Videocamera: il modello ideale.....	10
La Videocamera: il dispositivo reale .....	15
La Videocamera: calibrazione .....	16
La Stereovisione e la Geometria Epipolare.....	17
Matching stereo: cenni .....	25
Capitolo 2 – Ipotesi di ricostruzione 3D subacquea.....	29
Introduzione.....	29
Laser Stripe .....	32
Stereoscopia e lama laser .....	34
Capitolo 3 - Sensori Acustici.....	57
Introduzione.....	57
Struttura di un sistema sonar attivo .....	59
Il Sidescan Sonar (SSS).....	59

Sequenza in trasmissione.....	61
Trasduttori .....	61
Ricezione .....	62
Costruzione dell'eco.....	65
Limiti e performance dei Sidescan Sonar.....	67
Risoluzione .....	67
L'equazione del sonar .....	68
Portata massima .....	69
Copertura completa.....	70
Utilizzo delle immagini SSS .....	71
Estensione nell'uso di immagini SSS .....	72
Modello di diffusione acustica .....	73
Inversione Sonar .....	76
Limiti della ricostruzione.....	79
Capitolo 4 - Conclusioni .....	80
Integrazione ottico/acustica .....	81
Bibliografia .....	84

# Premessa

L'idea che muove questa tesi di laurea nasce all'interno dell'Istituto di Scienza e Tecnologie dell'Informazione (I.S.T.I.) del C.N.R. di Pisa, più precisamente nel Laboratorio di Segnali e Immagini (SILab), ed è volta alla ricerca ed allo studio di alcune delle possibili tecniche per l'esplorazione del mondo subacqueo. Attraverso l'utilizzo di sensori sempre più sofisticati e all'analisi dei segnali da essi generati, il fondale può essere esaminato e quindi rappresentato a seconda delle necessità. Si pensi al monitoraggio dell'ecosistema marino o dei manufatti umani come le condutture posate sui fondali o ai relitti navali ma anche, ad esempio, al recupero di beni archeologici.

È infatti noto che fin dall'antichità molti mari, tra cui il Mediterraneo, erano solcati da imbarcazioni commerciali di diversa provenienza e con differenti carichi. La navigazione era un elemento fondamentale del sistema economico dei popoli dell'antico Mediterraneo e ancora oggi, seppure in minor misura, fa parte della vita economica, sociale e culturale delle comunità che si affacciano sul *mare nostrum*. È altresì noto che nel corso dei secoli, dall'età antica ai giorni nostri, numerose imbarcazioni sono affondate o hanno abbandonato il carico,

determinando così la formazione di veri e propri siti archeologici sommersi. Di pari passo alla costituzione di questi siti e depositi sottomarini, sui medesimi si è sviluppata anche un'attività illegale e continua, di individuazione e depauperamento degli stessi. Diviene dunque sempre più pressante la necessità di un censimento accurato del patrimonio culturale sommerso, unico vero deterrente alla sua continua spoliatura. Uno dei problemi principali consiste infatti nel non conoscere neppure quali siano i siti da proteggere e da tenere sotto controllo. Al massimo è nota la localizzazione di alcuni, occasionalmente scoperti e segnalati alle competenti autorità, mentre manca un'indagine capace di fornire un quadro di insieme. Lo sviluppo delle tecnologie permetterà la realizzazione di un'azione di ricerca sistematica, a basso costo, di relitti e siti archeologici anche profondi, provvedendo alla loro identificazione, documentazione ottico-acustica e classificazione, in maniera automatica o semiautomatica. Oltre ad un'aumentata conoscenza della storia locale e internazionale, ciò provocherà, attraverso la diffusione di materiale multimediale raccolto, l'aumento della conoscenza e della consapevolezza del valore non solo naturalistico, ma anche storico e archeologico dei fondali marini.

# Capitolo 1 - Sensori Ottici

## ***Introduzione***

Un approccio possibile per gli scopi di questo lavoro di tesi, porta a focalizzare l'attenzione anche sulla cosiddetta Computer Vision, nel settore dell'automazione per veicoli autonomi (*Autonomous Vehicle, AV*). Gli AV, tra cui rientrano anche gli AUV (*Autonomous Underwater Vehicle*), vengono oramai largamente utilizzati sia per eseguire dei processi di routine automatizzati per l'industria, sia per svolgere compiti particolarmente azzardati per l'uomo, o addirittura impossibili. La *Machine Vision* (visione attraverso le macchine) tenta di "donar la vista" ai veicoli, al fine di *capire* ed interagire con il loro ambiente circostante, consentendo un uso più flessibile degli stessi.

Tuttavia, anche se la velocità e la potenza dei calcolatori è in continua crescita, la Ricerca è ancora molto lontana da riuscire a processare le immagini come è in grado di fare l'uomo. Infatti, il cervello umano è in grado di analizzare una scena molto più velocemente dei chip digitali, lo fa meglio ed in tempo reale, grazie a una qualche elaborazione parallela e grazie ad un largo uso di astrazioni di tipo gerarchico (tali fenomeni complessi non sono ancora chiari e sono argomento di ricerca). Fortunatamente, la maggior parte delle applicazioni non richiede di raggiungere lo standard della sensibilità umana; si riduce così la complessità del sistema, in

funzione delle sole condizioni sufficienti contingenti. Ad esempio, spesso è sufficiente riconoscere solo la sagoma di un oggetto che scorre su un rullo trasportatore, ma ci sono circostanze in cui si richiede la conoscenza di informazioni più complesse (dimensioni, colore, velocità, distanza, ...).

Una delle abilità più utili nella nostra visione quotidiana è la capacità di stimare la distanza di un oggetto, di confrontare oggetti vicini con oggetti lontani, in modo tale da interagire col mondo che ci circonda (afferrare gli oggetti, stimarne velocità, traiettoria, etc. etc.). Per quanto banale possa sembrare questo argomento, poiché è esperienza di vita di tutti i giorni, non è altrettanto banale farne una riproduzione con macchine artificiali.

Cercando di schematizzare la visione umana in maniera semplificata, questa consiste di due sensori di visione, gli occhi, i quali lavorano in stretto legame con un interprete, il cervello. Per dotare le macchine di una capacità visiva si tenta di imitare e replicare quello di cui l'uomo è da sempre dotato.

I sensori maggiormente utilizzati sono le videocamere, in numero variabile a seconda degli scopi. Nel caso in cui vengano utilizzati più sensori si parla di tecniche *multiview*, in cui il molteplice punto di vista fornisce informazioni aggiuntive indispensabili per ottenere alcuni risultati specifici. Ad esempio, per stimare la distanza alla quale si trova un oggetto è necessario osservare la scena da almeno 2 punti di vista differenti (è quello che fanno i nostri occhi), questa tecnica è chiamata *Stereovisione* o *Visione Stereoscopica*.

La stereovisione è una delle tecniche che consente agli AV di capire il loro ambiente di lavoro in 3 dimensioni. In questo modo è possibile ricostruirne una mappa in modo che il veicolo autonomo sia in grado di operare minimizzando i rischi. La maggior parte della ricerca in questo campo può essere raggruppata in tre classi applicative principali:

1. Rilevamento degli oggetti/ostacoli:

se un veicolo autonomo si muove in un ambiente sconosciuto, deve essere in grado di riconoscerne le caratteristiche, intanto che procede, con un primo approccio che prevede l'analisi delle immagini provenienti da sensori posizionati sulla parte anteriore

*(forward looking)*. I contorni dello scenario 3D, derivati attraverso la stereovisione, possono essere usati per la navigazione del veicolo e per la localizzazione e tracciamento degli oggetti conosciuti e non, e dunque per individuare un percorso utile e per evitare collisioni. Questa operazione deve essere molto rapida, addirittura in tempo reale (*realtime, online*), per adattarsi allo scenario che evolve.

2. Esplorazione dell'ambiente sconosciuto circostante (*Survey*):

man mano che un AV procede acquisendo informazioni dai sensori di bordo, può collezionare i dati ed elaborarli anche in un secondo tempo (*offline*), per creare una mappa dell'esplorato, consultabile a posteriori. L'elaborazione *offline* ha il vantaggio di fornire risultati di solito più accurati, dato che non ci sono vincoli sui tempi di processing.

3. Sorveglianza di ambienti conosciuti (*Surveillance*):

in questo caso, i veicoli utilizzati sono dotati di un meccanismo per localizzare la loro posizione relativa all'ambiente circostante, già noto. E' importante monitorare i movimenti dell' AV e mettere in relazione queste informazioni con la stima che l'AV possiede della sua posizione. L'immagine di ciascuna videocamera viene usata per identificare gli oggetti/ostacoli confrontando l'immagine ricevuta rispetto ad una immagine di riferimento che raffigura l'ambiente di lavoro "a vuoto". Gli oggetti identificati possono essere riconosciuti da un AV, facendo uso di modelli che descrivano le feature caratteristiche.

## ***Sensori di Visione***

La scelta di un sensore di tipo “videocamera” per la realizzazione di un Sistema di Visione è motivata dai parecchi vantaggi che ne derivano; i principali risultano:

- Sensore passivo: non interferisce con l’ambiente circostante ed è necessaria poca potenza per l’uso; diversamente da ciò che accade qualora si scelgano sensori attivi, come ad esempio pattern laser o sonar.
- Dimensioni: la continua miniaturizzazione dei sensori fa sì che esistano videocamere in commercio anche molto piccole; ciò semplifica molto l’assemblamento nei dispositivi meccanici che le ospitano (veicoli autonomi di ridotte dimensioni).
- Costo: il costo di una videocamera è divenuto oramai relativamente molto contenuto, neanche lontanamente paragonabile rispetto alla maggiorparte degli altri sensori, come laser o sonar.
- Ricchezza di informazioni e facilità di interpretazione per un utente: in molte applicazioni, anche una sola videocamera fornisce informazioni sufficienti e persino intuitive, nel caso sia necessaria un’interfaccia con un utente umano.

### ***La Videocamera: il modello ideale***

Una videocamera contiene al suo interno una lente che forma una proiezione 2D della scena sul piano immagine, dove è posto il sensore (in genere CCD). Nel seguito con il termine videocamera (*camera*), piuttosto che il dispositivo fisico in sé, intenderemo la trasformazione di proiezione che vi è legata: i punti del mondo 3D (*object space*) diventano punti 2D sul piano immagine. Questo comporta la perdita di informazione della profondità della scena, infatti ad un punto sul piano immagine corrisponde una intera retta della scena reale. La conoscenza della geometria del processo di acquisizione delle immagini diventa essenziale, in modo che le informazioni provenienti dalle telecamere possano essere “fuse” con le misure provenienti da altri sensori al fine del controllo del veicolo, come sarà descritto meglio in seguito.

Per descrivere la geometria che regola il funzionamento di una videocamera, ci riferiremo al modello *pinhole* a proiezione centrale (*central projection*) [1].

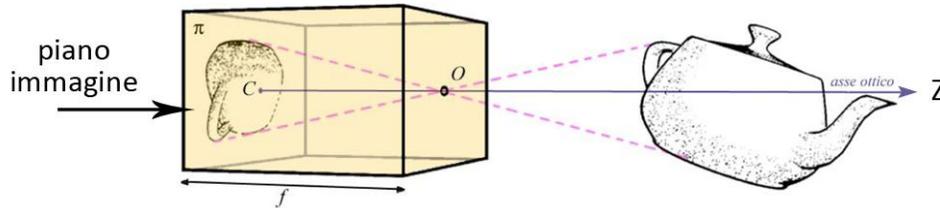


Figura 1.1. Il modello ideale "pinhole"

Il nome "pin-hole" (buco di spillo) nasce proprio dal fatto che si suppone che la luce che impressiona la lastra fotografica o il sensore di una fotocamera digitale, attraversi un foro di piccole dimensioni, tanto piccole da potersi considerare un punto, nell'accezione geometrica del termine.

Il sistema di riferimento della videocamera ha origine nel **centro di proiezione**  $O$ , detto anche centro della videocamera (talvolta chiamato  $C$ ), ed il piano individuato da  $Z = -f$ , viene chiamato **piano immagine** (*image plane* o *focal plane*), mentre  $f$  (oppure  $F$ ) viene chiamata **lunghezza focale** (*focal length*).

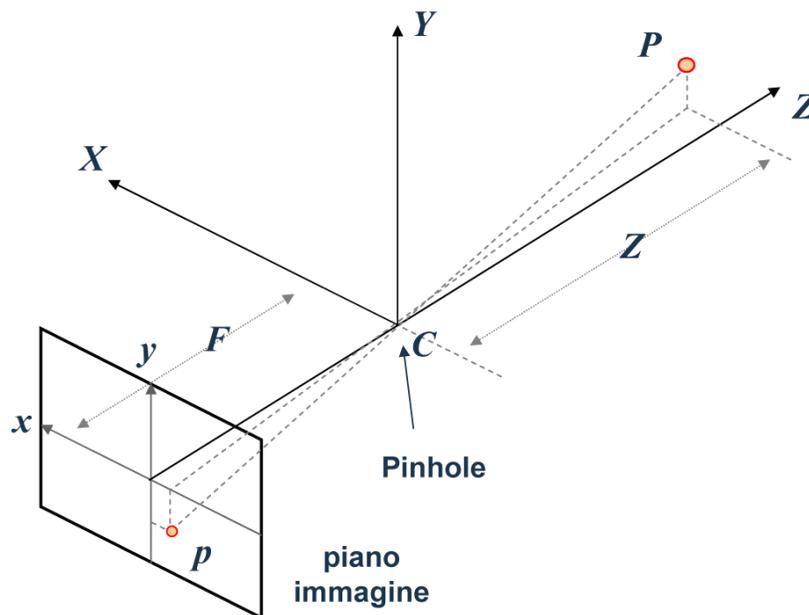


Figura 1.2. Sistema di riferimento cartesiano per il modello della videocamera

Per comodità, dato che la scena proiettata sul piano immagine ha simmetria centrale (con centro  $C$ ) rispetto a quella rilevata su un piano virtuale parallelo a distanza  $F$  rispetto all'origine degli assi, si preferisce utilizzare il modello equivalente.

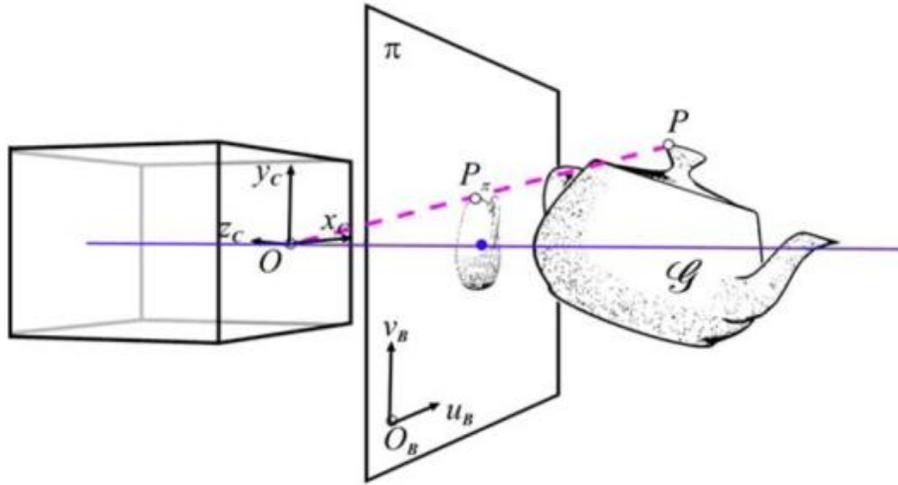


Figura 1.3. Modello pinhole equivalente

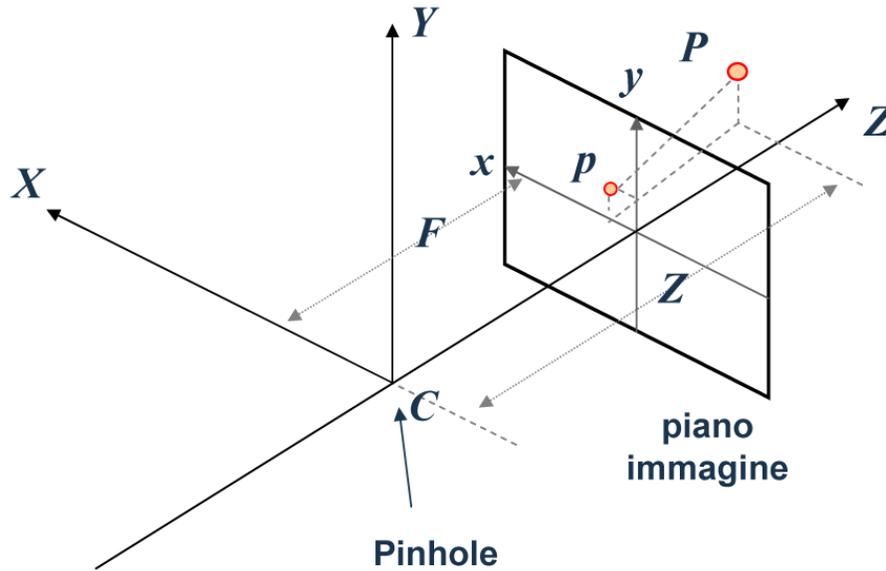


Figura 1.4. Sistema di riferimento cartesiano del modello equivalente

In riferimento alla fig. successiva, nel modello *pinhole*, un punto nello spazio  $\mathbf{P}$  di coordinate  $P = [X Y Z]^T$ , nel sistema di riferimento della videocamera (gli assi cartesiani hanno il pedice “C” come “camera”), si proietta sul piano immagine, attraverso una proiezione prospettica (*perspective projection*), alla locazione  $\bar{x} = [x y]^T$ , che corrisponde al punto di intersezione tra il piano immagine e la retta che collega il punto P al centro di proiezione.

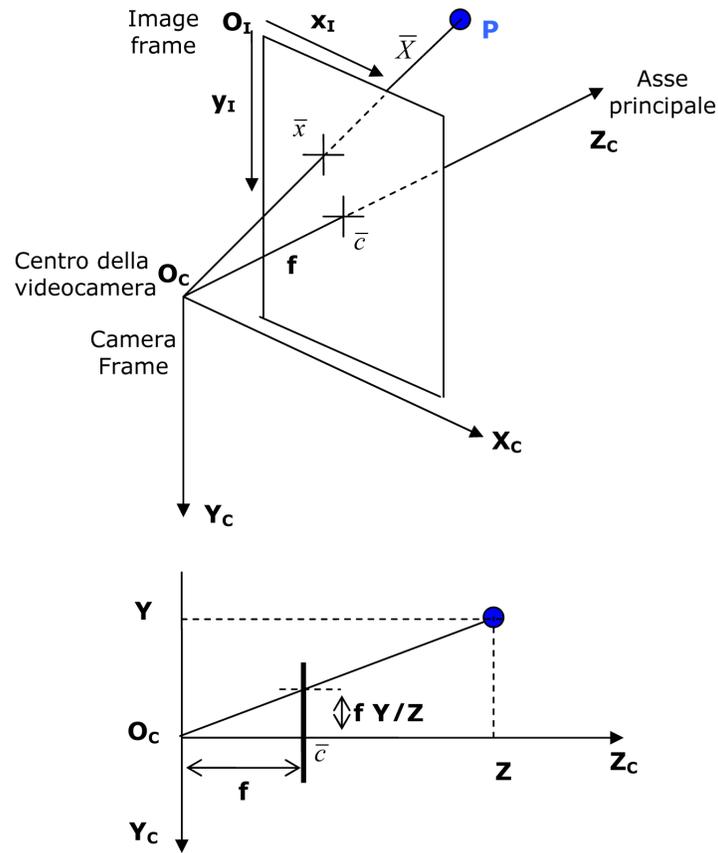


Figura 1.5. La proiezione prospettica

Si chiama asse ottico o **asse principale**, l’asse che attraversa il centro di proiezione, ortogonale al piano immagine. Questo asse interseca il piano immagine sul punto  $\bar{c} = [c_x c_y 0]^T$  (nel sistema di riferimento dell’immagine) chiamato **punto principale**.

La figura precedente mostra l’operazione di proiezione. Si osservano due sistemi di coordinate: il sistema di riferimento sulla videocamera di origine  $O_c$ , rispetto al quale si esprimono le coordinate 3D del punto, e il sistema di riferimento sul piano immagine di origine  $O_I$ , rispetto al quale si esprimono le coordinate in pixel. Gli assi  $X_c$  e  $X_I$  hanno stessa direzione

ma sono traslati ed anche scalati l'uno rispetto all'altro a seconda della lunghezza focale  $f$  (espressa in pixel). Lo stesso rapporto intercorre anche tra gli assi  $Y_C$  e  $Y_I$ . Per triangoli simili, come da figura, se si considera l'origine del sistema di riferimento sul piano immagine coincidente con il centro ottico, anziché sull'angolo in alto a sinistra, il punto  $P$ , ribattezzato  $\bar{X} = [X \ Y \ Z]^T$  risulta mappato sul punto  $\bar{x}$  sul piano immagine, le cui coordinate si esprimono come:

$$\bar{x} = \begin{bmatrix} f \frac{X}{Z} \\ f \frac{Y}{Z} \end{bmatrix}$$

per cui la funzione "proiezione centrale" dal mondo 3D alle coordinate sull'immagine si esprime come:

$$\mathbb{R}^3 \rightarrow \mathbb{R}^2 : \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} \rightarrow \begin{bmatrix} f \frac{X}{Z} \\ f \frac{Y}{Z} \end{bmatrix}$$

L'espressione non è del tutto corretta, in quanto bisogna considerare un offset per il punto principale:

$$\mathbb{R}^3 \rightarrow \mathbb{R}^2 : \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} \rightarrow \begin{bmatrix} f \frac{X}{Z} + c_x \\ f \frac{Y}{Z} + c_y \end{bmatrix}$$

Inoltre, quando i pixel del sensore non sono quadrati, la lunghezza focale deve essere normalizzata usando larghezza e altezza di un pixel ( $\partial x, \partial y$ ) come unità di lunghezza, ottenendo  $f_x$  e  $f_y$ . Le coordinate del punto proiettato sono in questo caso:

$$\bar{x}_{\{1,2\}} = \begin{bmatrix} f_x \frac{X}{Z} + c_x \\ f_y \frac{Y}{Z} + c_y \end{bmatrix}$$

Questo ultimo vettore rappresenta precisamente la posizione del punto proiettato sull'immagine in pixel, nel modello ideale di *pinhole* camera.

## ***La Videocamera: il dispositivo reale***

Le videocamere reali non hanno una struttura “pinhole” ma sono più complesse e utilizzano varie lenti assemblate in gruppi ottici, che introducono una certa distorsione dell’immagine. Di conseguenza la proiezione del punto  $\mathbf{P}$  del mondo 3D sul piano del sensore risulta in una posizione leggermente diversa rispetto a quella prevista dal modello ideale, che chiameremo  $\tilde{x}$ . Un modello del primo ordine della distorsione introdotta da una lente è il modello di *distorsione radiale*:

$$\left\{ \begin{array}{l} \bar{a} = \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} X/Z \\ Y/Z \end{bmatrix} \\ \bar{b} = \begin{bmatrix} b_X \\ b_Y \end{bmatrix} = \bar{a}(1 + k_c \|\bar{a}\|^2) \\ \tilde{x} = \begin{bmatrix} f_X b_X + c_X \\ f_Y b_Y + c_Y \end{bmatrix} \end{array} \right.$$

dove  $k_c$  è chiamato *coefficiente di distorsione radiale*.

E’ possibile descrivere le caratteristiche di una videocamera per mezzo di alcuni parametri che vengono suddivisi in:

- *parametri intrinseci*: sono i parametri necessari a collegare le coordinate di un pixel dell’immagine con le coordinate corrispondenti nel sistema di riferimento della camera.
- *parametri estrinseci*: sono i parametri che definiscono la posizione ed orientazione del sistema di riferimento “camera”, rispetto al riferimento “mondo”, che è supposto noto.

Alcuni di questi sono raggruppati nella cosiddetta *matrice di calibrazione* della videocamera, la quale contiene parametri propri della videocamera, quindi *intrinseci*:

$$K = \begin{bmatrix} f_X & \alpha & c_X \\ 0 & f_Y & c_Y \\ 0 & 0 & 1 \end{bmatrix}$$

dove:

- $f_X$  e  $f_Y$  sono la lunghezza focale espressa in pixel secondo la dimensione dei pixel lungo  $X_I$  e  $Y_I$  (solitamente  $f_X = f_Y = f$ )
- $c_X$  e  $c_Y$  sono le coordinate del punto principale sul frame del piano immagine

- $\alpha$  è lo *skew*, l'obliquità, ovvero l'angolo tra gli assi  $X_1$  e  $Y_1$  (ortogonali se  $\alpha = 0$ )

Non compaiono nella matrice di calibrazione ma fanno parte comunque dei parametri intrinseci di una telecamera, i parametri che rappresentano la distorsione (radiale e tangenziale).

### ***La Videocamera: calibrazione***

La procedura di calibrazione di una videocamera consiste nell'identificare i suoi *parametri intrinseci*,  $f_x, f_y, c_x, c_y, k_c$  e possibilmente anche  $\alpha$ . Un metodo standard consiste nell'acquisire un oggetto 3D noto (una scacchiera dalle dimensioni note, una scatola dalla geometria nota, etc.) e cercare il set di parametri che meglio approssimano la proiezione sul piano immagine dell'oggetto, calcolata con l'uso del modello matematico, rispetto alla sua proiezione realmente osservata. Poiché i parametri della videocamera sono calcolati a partire dalle misurazioni dell'immagine stessa, questo metodo è chiamato anche *calibrazione visuale*. Si osservi che nonostante sia nota a priori la struttura geometrica dell'oggetto 3D (ad esempio la posizione spaziale degli incroci nella scacchiera), non è affatto nota la sua posizione nello spazio reale rispetto alla videocamera ed è quindi necessario identificare anche i *parametri estrinseci*.

Risolvere questo problema significa gestire fino a 12 variabili indipendenti, ovvero 12 gradi di libertà:

- distanze focali  $f_x$  e  $f_y$ : 2 gradi di libertà
- Coordinate del punto principale  $c_x$  e  $c_y$ : 2 gradi di libertà
- Coefficiente di distorsione radiale  $k_c$ : 1 grado di libertà
- Posizione dell'oggetto 3D (punto di posa e orientamento) : 6 gradi di libertà
- Skew  $\alpha$  : 1 grado di libertà

La soluzione di questo problema consiste nello stimare il set di variabili di calibrazione che minimizzano un "errore di riproiezione" definito come la somma dei quadrati delle differenze tra la proiezione osservata dei punti noti dell'oggetto e quella prevista dal modello. Si tratta di un problema di ottimizzazione non lineare che può essere risolto con tecniche standard come

quella del “gradiente discendente”, la quale prevede l’uso di un algoritmo iterativo, che richiede una buona stima del set per l’inizializzazione. [2]

### ***La Stereovisione e la Geometria Epipolare***

Grazie all’uso coordinato di più videocamere che osservano una scena (tecniche *multiview*) è possibile ricostruire anche la terza dimensione, ovvero la profondità. A questo scopo sono sufficienti due videocamere, ciascuna delle quali funge da sensore di vista analogamente ai nostri occhi nella visione binoculare; da qui il nome di stereovisione. Il sistema delle due videocamere viene assemblato in maniera non dissimile dall’apparato visivo umano in cui la distanza tra i due “sensori” sia molto minore delle distanze in gioco nella scena osservata e le direzioni di puntamento siano parallele, ovvero i due piani immagine siano coplanari.

Aggiungendo il secondo punto di vista si sfrutta come informazione aggiuntiva la “diversità” che esiste nelle due immagini della stessa scena, osservata da due punti di vista leggermente differenti. Tanto più un oggetto sarà vicino al sistema di osservazione, tanto più la sua posizione risulterà differente nelle due immagini; viceversa un oggetto molto lontano verrà proiettato in zone corrispondenti tra i due fotogrammi acquisiti della stessa scena.



Figura 1.6. L’immagine è stata creata sovrapponendo i due punti di vista e codificandoli con colori differenti: rosso per l’immagine destra e ciano per quella sinistra

Consideriamo inizialmente un sistema stereoscopico in cui si possano fare le seguenti supposizioni [3]:

- La distanza tra le due telecamere sia fissa e nota.
- La matrice di rototraslazione tra le due videocamere sia nota.
- Le scene osservate dai due sensori non siano eccessivamente diverse l'una dall'altra
- Parametri intrinseci delle telecamere molto simili (al limite uguali).



Figura 1.7. Dispositivo stereoscopico integrato

La tipica configurazione stereo utilizzata da dispositivi integrati è quella nominata *parallel cameras*, o anche *sistema stereo "laterale"*, in cui:

- I due piani immagini sono disposti in verticale e sono coplanari nello spazio (conseguentemente gli assi ottici risultano paralleli)
- Le due telecamere hanno identica lunghezza focale e sono separate solo lungo l'asse X da una distanza  $B$  detta *Baseline*.

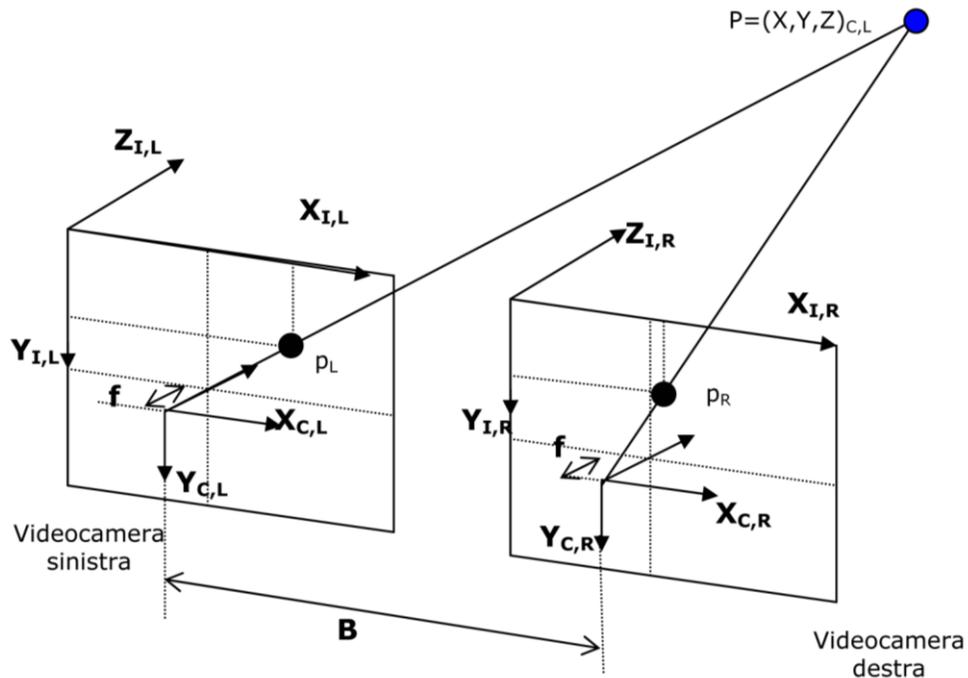


Figura 1.8. Configurazione stereo *parallel cameras*

Si chiama *coppia di punti coniugati* (oppure *omologhi*), la coppia delle due proiezioni di un medesimo punto **P** dello spazio 3D nei due piani immagine; vengono indicati con  $p_L$  e  $p_R$ .

In questo caso ideale, le proiezioni di **P**, coppia di coniugati, risultano di coordinate leggermente diverse tra le due immagini. Si faccia riferimento all'immagine precedente: per semplicità, il sistema di riferimento della realtà 3D è fatta coincidere con uno dei due sistemi di riferimento dei sistemi ottici, in questo esempio con il sinistro. Spostando semplicemente il centro del sistema di riferimento verso destra, in corrispondenza del centro ottico destro, le coordinate assolute del solito punto **P** differiranno per l'ascissa che si sposta di una quantità nota pari a  $B$ .

Quindi la coordinata verticale della proiezione del punto **P** è identica per le due immagini, e vale  $y_{I,L} = y_{I,R} = Y \frac{f}{Z}$  (supponendo per semplicità  $f_X = f_Y = f$ ); invece l'ascissa varia di poco:  $x_{I,L} = X_L \frac{f}{Z}$  e  $x_{I,R} = X_R \frac{f}{Z}$ , in cui si distingue la coordinata  $X_L$  da  $X_R$ .

La differenza delle due coordinate lungo X si chiama *disparità* (*disparity*). Le informazioni sulla profondità si ottengono facendo delle valutazioni su quest'ultima grandezza. Infatti:

$$d = x_{I,L} - x_{I,R} = (X_L - X_R) \frac{f}{Z} = B \frac{f}{Z}$$

da cui si ricava facilmente  $Z = \frac{Bf}{d}$ .

Riepilogando: dati due punti omologhi ( $p_L$  e  $p_R$ ) sulle due immagini, proiezioni del punto **P** della scena, e conoscendo i parametri  $B$  e  $f$  del sistema stereo è immediato calcolare la profondità di **P**, ovvero la sua coordinata  $Z$ . In particolare:

- $d$  grande  $\rightarrow$  punto vicino
- $d$  piccola  $\rightarrow$  punto lontano
- $d=0 \rightarrow$  punto a distanza infinita

E' possibile costruire una geometria con cui si descrivono le relazioni e i vincoli geometrici che legano le due proiezioni 2D della stessa scena 3D catturata da due fotocamere (o videocamere) con posizione e orientamento distinto. Questo tipo di geometria è detta *geometria epipolare*.

Il piano che passa per i centri delle due videocamere e il punto in esame sulla scena è chiamato *piano epipolare*. L'intersezione tra un piano immagine e il piano epipolare si chiama *linea epipolare*.

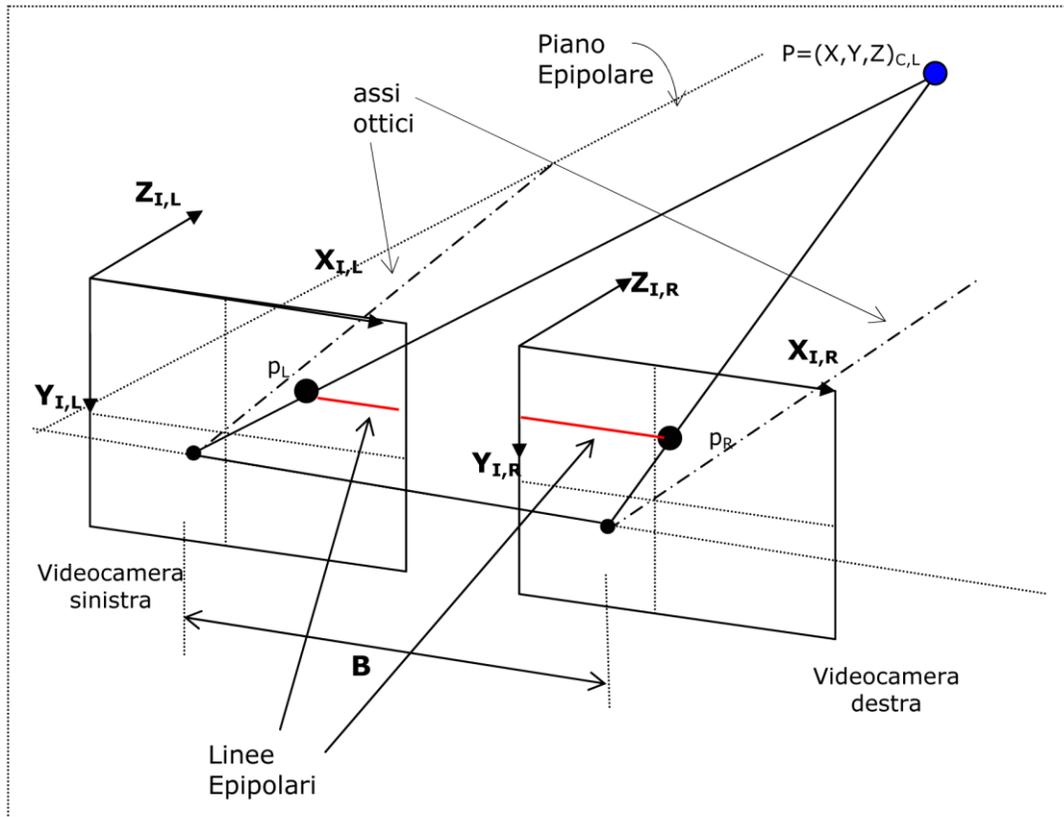


Figura 1.9. Assi ottici, Piano epipolare e Linee epipolari

La figura precedente mostra nel dettaglio il piano epipolare, in cui giacciono le rette che collegano il punto  $P$  (indicato in blu e con coordinate espresse rispetto al frame della videocamera sinistra) con i rispettivi centri delle due videocamere e la relativa baseline, e le linee epipolari, evidenziate in rosso.

In realtà, a causa della inevitabile distorsione delle immagini dovute al sistema ottico e della geometria imperfetta, è necessario considerare anche una leggera disparità verticale delle proiezioni. Definiremo più appropriatamente la disparità come la distanza tra i punti di una coppia di coniugati se le due immagini venissero sovrapposte.

La configurazione di tipo *parallel cameras* fin qui analizzata, risulta di difficile realizzazione pratica, in tutti i casi in cui si abbia a che fare con un sistema reale non ottimizzato o nei casi in

cui la configurazione stereo preveda l'uso di due videocamere indipendenti, assemblate in un prototipo artigianale.



Figura 1.10. Esempio di prototipo artigianale

Per queste ragioni si può fare riferimento ad una configurazione stereo generica, in cui:

- i due piani immagini sono disposti in verticale e non sono coplanari nello spazio
- le due telecamere hanno identica lunghezza focale e sono separate solo lungo l'asse X da una distanza di Baseline B

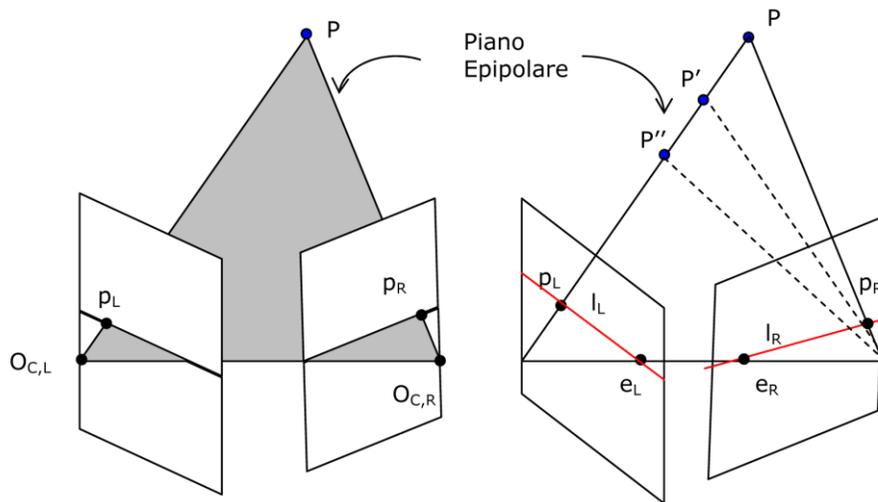


Figura 1.11. Configurazione stereo generica

Dal momento che la geometria proiettiva dipende solamente dai parametri interni alle videocamere e alla posizione relativa l'una rispetto all'altra, le definizioni di piano epipolare e di linea epipolare rimangono invariate. I centri delle videocamere, il punto nello spazio 3D  $P$  e le sue proiezioni  $p_L$  e  $p_R$ , giacciono sullo stesso piano (epipolare), per definizione.

Analizzando a ritroso: se il punto dell'immagine sinistra  $p_L$  viene proiettato all'indietro nello spazio 3D, questo corrisponde ad un intero raggio, definito tra il centro della telecamera sinistra e il punto stesso, appartenente quindi al piano epipolare. Tale raggio di punti si proietta come una linea  $l_R$  nel secondo piano immagine (piano immagine destro): la proiezione di  $P$  sulla seconda immagine (immagine destra) è un punto che appartiene sicuramente a tale linea epipolare.

Si chiama *epipolo* (in inglese *epipole*) l'intersezione della linea che congiunge i centri delle videocamere (*baseline*) con il piano immagine. Accade che tutte le linee epipolari si intersecano all'epipolo del piano immagine, ovvero ruotano intorno all'epipolo al variare del punto. Allo stesso modo, l'epipolo è la proiezione del centro ottico dell'altra videocamera (non è detto che rientri nel fotogramma, ma sarà parte del piano su cui esso giace). L'epipolo costituisce anche il punto di fuga della direzione di traslazione della baseline.

Ciascun piano epipolare interseca il piano dell'immagine sinistra e il piano dell'immagine destra lungo le linee epipolari e definisce le corrispondenze tra le linee.

Grazie alla possibilità di poter caratterizzare algebricamente la geometria che regola le proiezioni di un punto nello spazio 3D sui due piani immagini, ovvero la Geometria Epipolare, è possibile calcolare la distanza del punto  $P$  dai piani delle due videocamere.

Osservando i due fotogrammi della stessa scena, destro e sinistro, un problema non banale è quello di identificare le coppie di punti omologhi: è il *problema delle corrispondenze*. [4]



Figura 1.12. Alcune corrispondenze stereo

Infatti un punto  $p_L$  sul fotogramma sinistro potrebbe avere come omologo un qualunque punto  $p_R$  sul fotogramma destro, a meno che non si pongano dei vincoli sulla ricerca che possono essere di tipo geometrico, logico, strutturale, etc. Potrebbe essere opportuno limitare il più possibile l'intervallo di disparità al fine di considerare solo la porzione di spazio 3D entro cui giacciono gli oggetti di interesse. Inoltre la disparità varia con continuità quasi ovunque e potrebbe essere utile identificare e valutare le zone in cui ciò non accade (ad esempio i bordi degli oggetti). Un'altra considerazione sta nel fatto che ciascun punto di un'immagine può avere al più un punto corrispondente nell'altra immagine.

La *geometria epipolare* viene in aiuto per semplificare il problema delle corrispondenze in quanto è possibile ridurre la ricerca del punto omologo ad un problema monodimensionale. Infatti, invece di cercare il punto omologo nello spazio bidimensionale dell'immagine corrispondente, si sfrutta il *vincolo epipolare* sapendo che il punto giace sulla *linea epipolare*, precedentemente descritta.

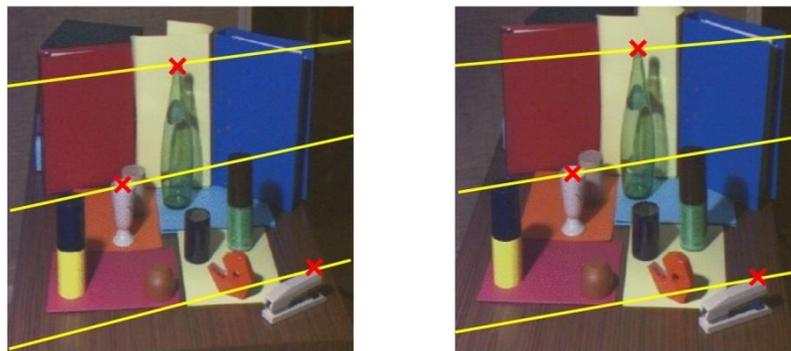


Figura 1.13. I punti omologhi giacciono sulle linee epipolari

Le stesse considerazioni fatte per la ricerca nello spazio bidimensionale valgono anche nella ricerca monodimensionale.

Il passo successivo per ridurre ulteriormente la complessità della ricerca consiste nella cosiddetta *rettificazione delle immagini*. La rettificazione è una procedura che, mediante l'uso dei parametri intrinseci ed estrinseci ottenuti mediante calibrazione, mira a trasformare le immagini stereoscopiche, provenienti dal dispositivo di acquisizione, in modo che risultino

soddisfatti alcuni vincoli. Tra questi, la rettificazione consente di trasformare le immagini in *forma standard*: nel caso di sistemi binoculari questo assicura che dato un punto in un'immagine il suo omologo possa essere rintracciato sulla stessa riga dell'altra immagine consentendo una notevole riduzione dei calcoli ed una maggiore affidabilità nella soluzione del problema delle corrispondenze.

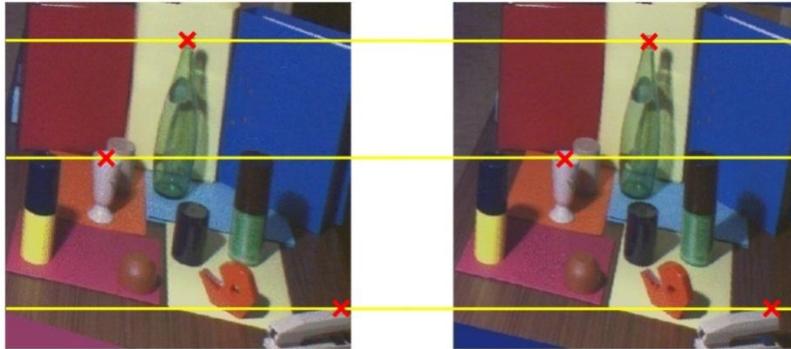


Figura 1.14. Linee epipolari rettificate

Mediante la rettificazione è possibile trasformare le immagini in modo che i punti omologhi di una riga (*scanline*) dell'immagine possano essere ricercati nella corrispondente riga dell'altra immagine.

E' possibile dunque ottenere una configurazione stereoscopica equivalente a quella standard (*parallel cameras*, o sistema stereo laterale) in cui le rette epipolari coniugate sono orizzontali e collineari, applicando a ciascuna delle due immagini una trasformazione (omografia). La rettificazione consente anche di ridurre i problemi legati alla distorsione provocata dalle ottiche e di ottenere immagini con la stessa distanza focale.

Una volta identificate le coppie di punti omologhi, mediante una qualunque tecnica di corrispondenza (*matching stereo*), è possibile calcolarne la posizione nello spazio 3D, come visto in precedenza, con una triangolazione.

## ***Matching stereo: cenni***

La bibliografia relativa alla visione stereoscopica è estremamente vasta ed è possibile in prima istanza classificare gli algoritmi di matching stereo in due categorie principali [4, 5]:

- algoritmi di tipo *feature-based*
- algoritmi di tipo denso (*dense stereo*).

I primi consentono di ottenere informazioni di disparità (o mappe di disparità) per un numero limitato di punti delle immagini per i quali sono identificate features particolari come linee, segmenti, angoli. Tali algoritmi risultano efficienti dal punto di vista computazionale, per via del ridotto numero di features identificabili nelle immagini rispetto al numero totale di punti, ed anche estremamente affidabili perché le features estratte dalle immagini generalmente sono intrinsecamente distintive e non pongono problemi di ambiguità nella soluzione del problema delle corrispondenze. Tali algoritmi però risultano attualmente poco utilizzati perché le mappe di disparità risultano limitate ai soli punti che presentano features distintive.

Contrariamente, gli algoritmi di tipo denso mirano a generare mappe di disparità per ogni punto dell'immagine ed è possibile classificare tali algoritmi in due principali categorie: algoritmi di tipo locale (*local*) e algoritmi di tipo globale (*global*).

Gli algoritmi di tipo locale cercano i punti omologhi singolarmente ed indipendentemente gli uni dagli altri sfruttando generalmente un supporto locale (*local support*) nell'intorno di ogni punto esaminato; tale area nell'intorno di ogni punto è denominata anche finestra di correlazione (*correlation window*). Utilizzando un supporto locale prestabilito, tali algoritmi implicitamente assumono che al suo interno la disparità sia limitata anche se tale ipotesi, non essendo sempre verificata, conduce ad una discrepanza tra gli oggetti della scena e le mappe di disparità in prossimità di punti posti a differente distanza dalle telecamere, lontani fra loro. E' importante osservare che esistono anche algoritmi locali che adattano il proprio supporto in base alle caratteristiche locali delle immagini mediante i quali è possibile ottenere eccellenti

risultati ad un costo computazionale tipicamente molto più alto rispetto agli algoritmi locali standard.

Generalmente gli algoritmi di tipo locale standard hanno una struttura molto regolare dal punto di vista computazionale e possono essere implementati efficacemente sia sfruttando schemi di calcolo ricorsivi sia utilizzando istruzioni di tipo SIMD (Single Instruction Multiple Data) disponibili attualmente in quasi tutte le architetture ed anche negli acceleratori grafici di ultima generazione. Per questo motivo tali algoritmi possono essere attualmente utilizzati in applicazioni di tipo real-time.

Oltre alla assunzione sulla costanza della disparità all'interno della finestra di correlazione, gli algoritmi di tipo locale hanno un limite legato al fatto che non necessariamente sono in grado di produrre mappe di disparità completamente dense (non è garantito che per ogni punto dell'immagine sia possibile stabilire un valore di disparità). Questo avviene principalmente per tre motivi:

- a causa di regioni uniformi,
- di occlusioni,
- di pattern ripetitivi.

Nel primo caso non è possibile determinare informazioni di disparità in regione che presentano intensità costante o comunque pochi elementi distintivi a causa dell'elevata ambiguità nella soluzione del problema delle corrispondenze.



Figura 1.15. Ricerca ambigua del punto omologo causata da zone uniformi [4]

Nel secondo caso, a causa della geometria del sistema stereoscopico, in presenza di punti adiacenti posti a differente distanza dalle telecamere, risulta che alcuni punti siano visibili in un'immagine ma non nell'altra accoppiata: in casi come questo non è possibile risolvere il problema delle corrispondenze.

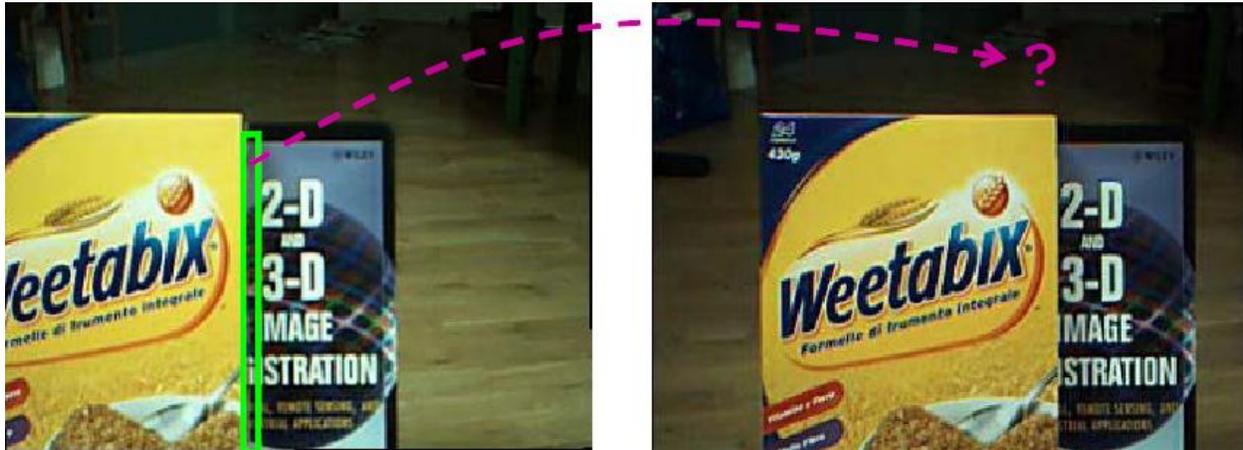


Figura 1.16. Corrispondenza impossibile a causa delle occlusioni ottiche [4]

Nel terzo caso l'ambiguità potrebbe nascere dal fatto che un punto faccia parte di un pattern ripetitivo e si ripropone nell'immagine corrispondente più di una volta, inserito nello stesso contesto. In questo caso il contenuto della finestra di correlazione è molto simile in varie parti della scena.

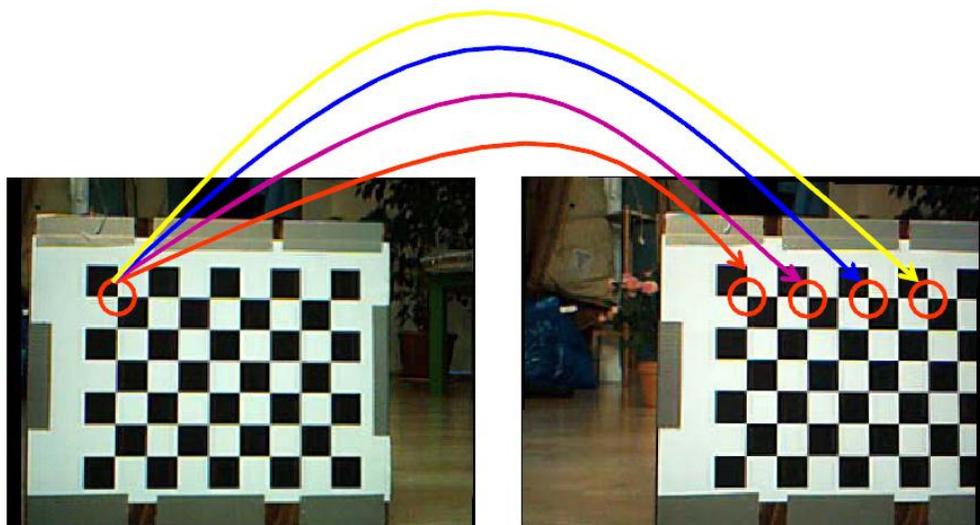


Figura 1.17. Corrispondenza ambigua a causa di pattern ripetitivi [4]

Diversamente dagli algoritmi di tipo locale, gli algoritmi densi di tipo globale non utilizzano tipicamente alcun supporto locale ma, facendo alcune ipotesi sulla natura degli oggetti, ricercano un assegnamento di disparità tale che venga minimizzata una funzione costo globale (*global cost function*). La maggior differenza tra tali algoritmi è proprio nel metodo con il quale viene minimizzata la funzione costo (*simulated annealing, probabilistic diffusion, graph-cuts*).

Tali algoritmi, in particolare quelli basati su *graph-cuts* consentono di ottenere mappe di disparità decisamente migliori rispetto agli algoritmi di tipo local ma ad un costo computazione estremamente elevato che li rende attualmente non utilizzabili in applicazioni real-time. Inoltre, anche se alcuni algoritmi, sfruttando il metodo della minimizzazione della funzione di costo, sono in grado di risolvere il problema delle occlusioni, queste possono comunque causare mappe di disparità non completamente dense, analogamente agli algoritmi di tipo local.

## Capitolo 2 – Ipotesi di ricostruzione 3D subacquea

### *Introduzione*

La stereoscopia funziona grazie all'uso di videocamere, ovvero sensori ottici passivi che rilevano la luminosità della scena. L'ipotesi intrinseca è che la scena sia illuminata a sufficienza, in modo tale che parte della luce riflessa dagli oggetti raggiunga i sensori in questione. In ambiente subacqueo non è sempre possibile fare quest'ipotesi, infatti quando la profondità e la torbidità dell'acqua sono elevate è necessario l'utilizzo di illuminatori.

Anche il tipo di illuminazione presente può influire in maniera importante sulla ricostruzione tridimensionale. E' preferibile avere una luminosità diffusa ed uniforme piuttosto che un singolo punto luce che crea forti contrasti ed ombre scure che possono mascherare parte del contenuto della scena, difficilmente ricostruibile. Una situazione critica potrebbe essere la presenza di una sorgente luminosa nell'immagine che satura i sensori, rendendo di fatto indistinguibili i vari punti. Questo potrebbe essere causato, ad esempio, da un riflesso dell'illuminatore su una superficie particolarmente riflettente, ad esempio metallica.

I sensori ottici possono acquisire e quindi digitalizzare, un certo range dinamico di intensità luminosa, che viene rappresentata numericamente con un numero finito di bit; solitamente 8 bit, quindi 256 livelli possibili di intensità, dall'assenza di luce (nero) alla totalità (bianco).



Figura 2.1. Una tipica scena subacquea

Quando la luce che raggiunge il sensore eccede questi limiti, non è possibile registrarne le variazioni e si perdono così informazioni sulla scena.

Inoltre in condizioni di acque torbide, i livelli possibili ottenuti dalla digitalizzazione potrebbero non essere sufficienti per rilevare variazioni significative di luminosità, in condizioni di scarso contrasto.

Nell'immagine precedente si nota come già a 2,5 metri di profondità, in acque poco torbide, i colori siano pressochè indistinguibili e la visibilità si riduca notevolmente all'aumentare della distanza. Questa situazione è simile a quella sperimentata in atmosfera con condizioni meteorologiche di nebbia o foschia.

Strumenti di miglioramento dell'immagine come l'aumento del contrasto, ad esempio per mezzo dell'equalizzazione dell'istogramma, non risolverebbero completamente il problema, in quanto non possono tener conto della variazione della distanza dal punto di osservazione, ma agiscono indistintamente su tutto il fotogramma.

Per risolvere questo problema sono stati proposte alcune soluzioni come ad esempio quella di Yoav Y. Schechner and Nir Karpel [6], basata sull'acquisizione della scena mediante l'utilizzo di filtri polarizzatori, in diversi stati di polarizzazione. Tali immagini servono per inizializzare un algoritmo in grado di stimare gli effetti della torbidità, e quindi ridurli.



Figura 2.2. Sistema di videocamere aggiunto di filtri polarizzatori

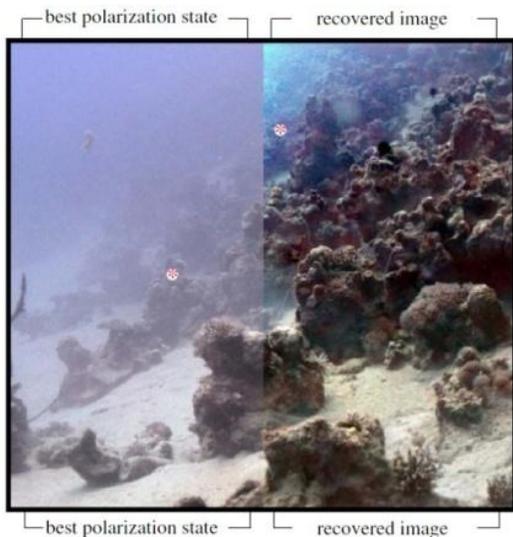


Figura 2.3. Confronto tra l'immagine originale e quella elaborata. Il risultato di tali tecniche è interessante poiché aumenta notevolmente la visibilità con la distanza, rendendo il contrasto della scena nettamente migliore e più uniforme.

L'uso di soluzioni simili, anche se interessanti, non è compatibile con l'obiettivo che muove questa ricerca poiché i veicoli che dovrebbero ospitare questi dispositivi devono soddisfare alcuni requisiti, tra cui si ricorda la totale autonomia operativa che implica l'utilizzo di apparecchi dal basso assorbimento elettrico e dalle ridotte dimensioni, in grado di essere alloggiati su un veicolo affusolato, ma anche dal costo contenuto.

## *Laser Stripe*

Al fine di cogliere più particolari dalla scena osservata e di aggirare il problema della eventuale torbidità si possono usare particolari proiettori laser che, invece di illuminare globalmente la scena e diffondere luce in ogni direzione, concentrano l'energia luminosa in direzioni particolari, ad esempio lungo una semplice linea. E' possibile anche l'utilizzo congiunto di proiettori laser e di illuminatori standard (ad esempio a LED) laddove le condizioni di visibilità lo permettano.

Tra i dispositivi commerciali si può citare come esempio il "SeaStripe" della Trittech International, un generatore di linea laser per ROV/AUV, per capire le peculiarità di questi strumenti [7]:



Figura 2.4. L'illuminatore laser "SeaStripe" della Trittech International

Dal datasheet del dispositivo vengono elencate le seguenti caratteristiche:

- ottimizzato per la penetrazione in acque marine
- proiettore ad elevato angolo di apertura
- intensità della linea uniforme con cut-off ben definito
- sistema a stato solido compatto e robusto
- utilizzo fino a 4000 metri di profondità

Il laser genera un fascio di luce verde a 532 nanometri, filtrato con una lente di tipo Powell che produce un raggio a "ventaglio" di 64° (in acqua) dai contorni ben definiti e può essere usato come profilo di riferimento a elevato contrasto sul fondo marino.

Un'immagine esplicativa, proposta nel datasheet, evidenzia il vantaggio nell'uso della linea laser per identificare dettagli altrimenti difficilmente distinguibili, come i raggi di una ruota di bicicletta adagiata sul fondo:

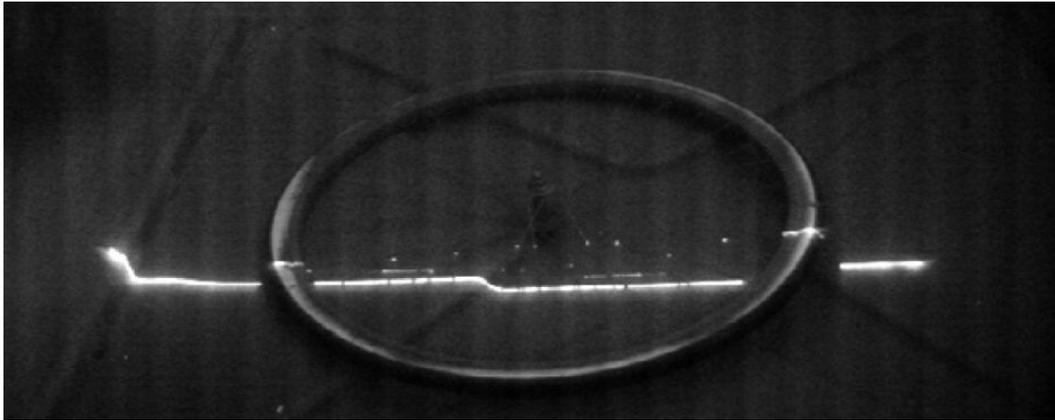


Figura 2.5. Immagine dal datasheet. Evidenza dei particolari rilevati

La luce monocromatica generata da questo laser è facilmente distinguibile anche nel caso in cui si utilizzino illuminatori aggiuntivi o nel caso in cui l'ambiente sia illuminato da luce naturale.

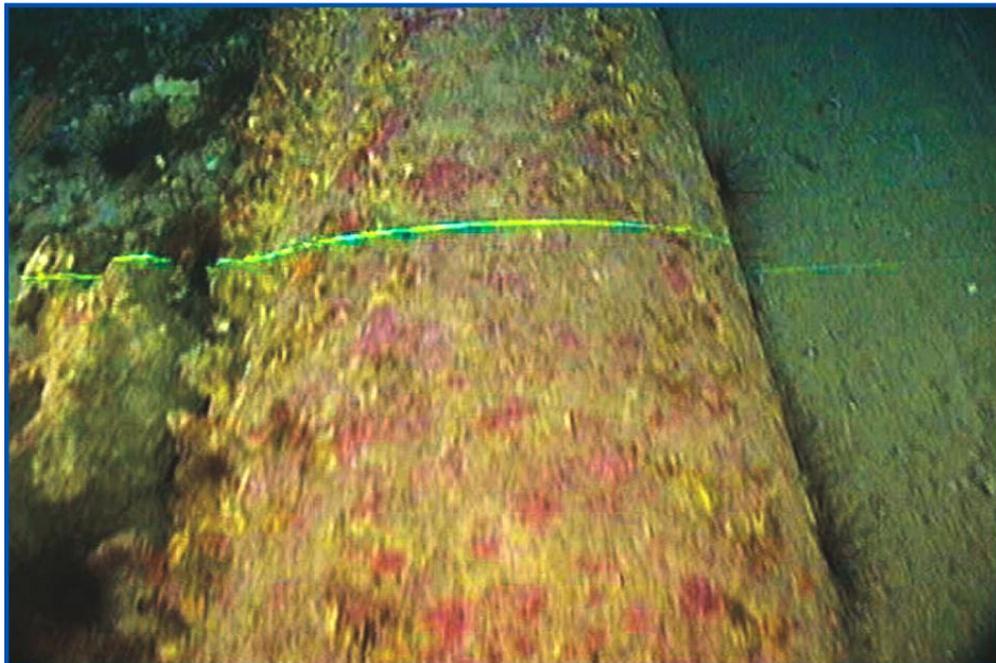


Figura 2.6. Immagine dal datasheet. Linea su fondo illuminato.

Altre specifiche dal datasheet:

<b>Laser specification</b>	Class 3R, 20 mW at 532 nm wavelength (green) Swath angle 64 degrees in water
<b>Power</b>	11- 26 V DC 3 W max. No warm-up period
<b>Dimensions</b>	292 mm x 80 mm diam
<b>Weight in air</b>	1.7 kg
<b>Weight in water</b>	0.9 kg
<b>Depth rating</b>	4000 m

Da questi dati si può osservare anche che le dimensioni fisiche sono ridotte e che il consumo energetico è contenuto.

L'utilizzo di dispositivi relativamente semplici ed economici per l'esplorazione del fondo marino rende necessaria una qualche elaborazione a posteriori per estrarre le informazioni utili, e tali tecniche si diversificano in base alla sensoristica di bordo dei natanti adoperati ma anche in base al tempo di processing necessario ai calcolatori per ottenere le informazioni cercate.

### ***Stereoscopia e lama laser***

Un'idea che unisce le necessità alle specifiche potrebbe risiedere nell'utilizzo congiunto dell'illuminazione mediante lama laser con la tecnica della stereoscopia. Questo modo operativo potrebbe essere giustificato da uno scenario subacqueo in cui le condizioni di visibilità non siano ottimali, in cui la profondità o la torbidità non permettano alla luce naturale di filtrare fino alla zona in esame e si renda necessario l'uso di illuminatori artificiali. Laddove poi la torbidità fosse particolarmente evidente anche l'uso di illuminatori a luce diffusa (come i LED) potrebbe restituire un'immagine eccessivamente degradata.

Come già osservato, il problema principale nella risoluzione della tecnica stereoscopica è quello del *matching stereo*, ovvero della localizzazione delle coppie di punti corrispondenti fra

le due immagini coniugate (destra e sinistra) e l'uso della lama laser può ridurre la complessità della ricerca.

Al fine di giustificare questa idea è stato assemblato un setup per simulare un semplice sistema stereoscopico (montato su un natante, ad esempio un AUV che procede all'esplorazione), utilizzando una lama laser e due fotocamere digitali di uso comune. L'esperimento è condotto in aria ma è rivolto alla semplificazione del *matching stereo* che per natura del sistema stereoscopico si presenta anche in acqua.

Si suppone che, in un sistema ottimizzato, le videocamere stiano osservando una scena che evolve, man mano che l'AUV naviga ad una certa quota sopra il fondo marino, con una velocità nota, registrando due flussi video sincronizzati dai quali è possibile estrarre le coppie di fotogrammi corrispondenti. La sincronizzazione deve garantire che nell'istante di tempo in cui viene catturato il fotogramma destro, venga catturato anche il sinistro, in modo che la scena sia esattamente la stessa, ipotesi fondamentale per la ricerca delle corrispondenze.



Figura 2.7. Oggetto di test per la ricostruzione 3D tramite stereoscopia e lama laser

Nell'esperimento condotto, il bersaglio illuminato dalla lama laser è un piccolo soprammobile a forma di anfora, scelto semplicemente per richiamare alla mente l'idea che muove questa ricerca.

Il sistema assemblato, per quanto rudimentale, serve a prendere conoscenza della tecnica stereoscopica e non per simulare realisticamente quanto accadrebbe in acqua.

Si osservi che tutto il setup è statico, a differenza di quanto accadrebbe se fosse montato su un veicolo, ma l'azione del moto è simulata spostando l'anfora manualmente a passi di 1mm, usando un foglio di carta bianca come supporto. Sono state necessarie 74 coppie di scatti (destro e sinistro) per coprire l'intera anfora.

Come si vede dall'immagine successiva, le condizioni di illuminazione della stanza sono casuali, ovvero non si è cercato di ottimizzare la presenza della linea laser proiettata (ad esempio scurendo le luci naturali e artificiali, rendendo la stanza buia). Questo perché le condizioni di illuminazione in cui si troverebbe ad operare un veicolo autonomo non sono prevedibili e, di certo, non saranno ottime. Si dovrà quindi cercare un metodo per evidenziare la presenza della linea laser.

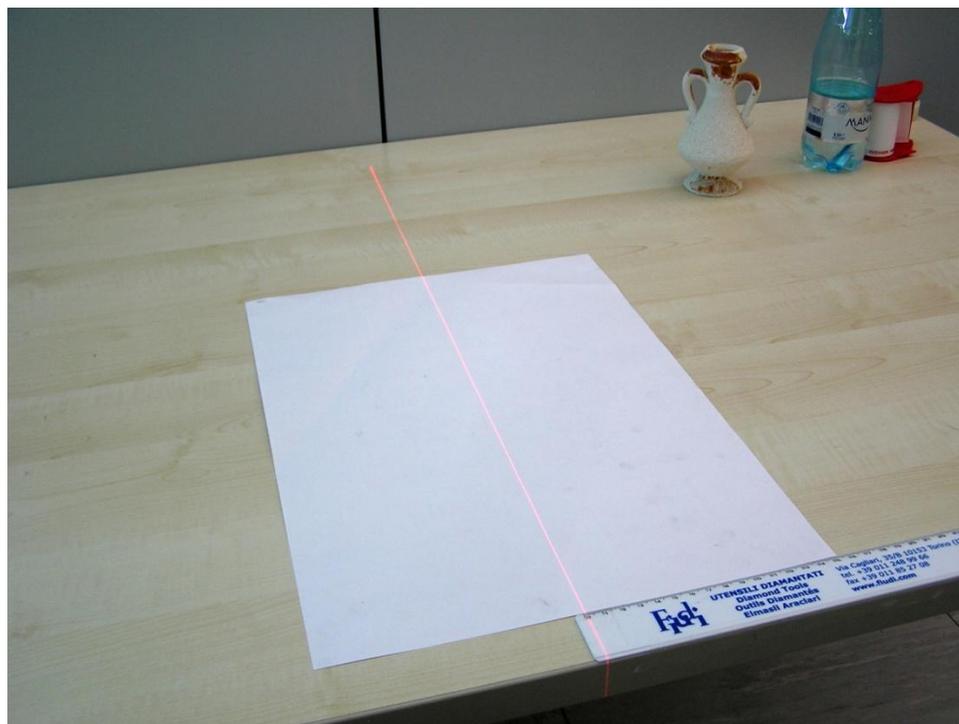


Figura 2.8. Condizioni operative dell'esperimento condotto

Le due fotocamere usate sono diverse l'una dall'altra ed in particolare si tratta di una Olympus C-750 UltraZoom e di una Panasonic Lumix DMC-FZ50. Montate su due cavalletti, sono rimaste nella stessa posizione per tutta la durata delle acquisizioni, inquadrando l'anfora illuminata dalla linea laser, al centro del fotogramma.



Figura 2.9. Fotocamere utilizzate

La dimensione delle immagini utilizzate è di 1024 x 768 pixel; nonostante la risoluzione delle due fotocamere si potesse spingere ben oltre, si è scelto di utilizzare una dimensione standard vicina a quella delle telecamere tradizionali (il formato video HD prevede una risoluzione di 1280 x 720 e il formato Full HD di 1920 x 1080 pixel). Quindi le immagini acquisite sono state opportunamente ridimensionate.

Le ottiche delle due fotocamere non sono fisse (dalla distanza focale nota) e l'escursione della distanza focale non è graduata adeguatamente, per cui non è stato possibile assemblare il sistema conoscendo le grandezze geometriche a priori. Inoltre, al fine di evitare le forti distorsioni introdotte dalle lenti nell'uso grandangolare si è scelto di posizionare lo zoom in una posizione intermedia. Dai dati EXIF delle immagini acquisite è stato comunque possibile risalire alla distanza focale effettivamente utilizzata che risulta di 12mm per la Olympus e 17mm per la Panasonic.

Una volta posizionato il sistema si è resa necessaria la calibrazione delle due fotocamere e del sistema stereoscopico stesso (calcolo dei parametri intrinseci ed estrinseci). Per ovviare a questo problema è stato fatto uso di un toolbox sviluppato in MatLab da Jean-Yves Bouguet e chiamato per l'appunto *Camera Calibration Toolbox for Matlab* [8]. Il metodo prevede l'uso di un pattern geometrico dalle dimensioni note ed in particolare è stato utilizzato un pattern a scacchiera i cui quadrati hanno un lato di 30mm. Il pattern è stato stampato e incollato su un piano di cartone rigido.

Sono state acquisite 24 pose da entrambe le fotocamere necessarie alla procedura di calibrazione che richiedeva la identificazione manuale degli spigoli più esterni della scacchiera in ogni fotogramma acquisito. Il metodo suggerisce particolare attenzione nella scelta del primo angolo selezionato nel caso in cui si voglia calibrare un sistema stereo, per il successivo calcolo della posizione reciproca delle fotocamere (parametri estrinseci). Questo è un mosaico dei 24 scatti di calibrazione effettuati dalla fotocamera destra:

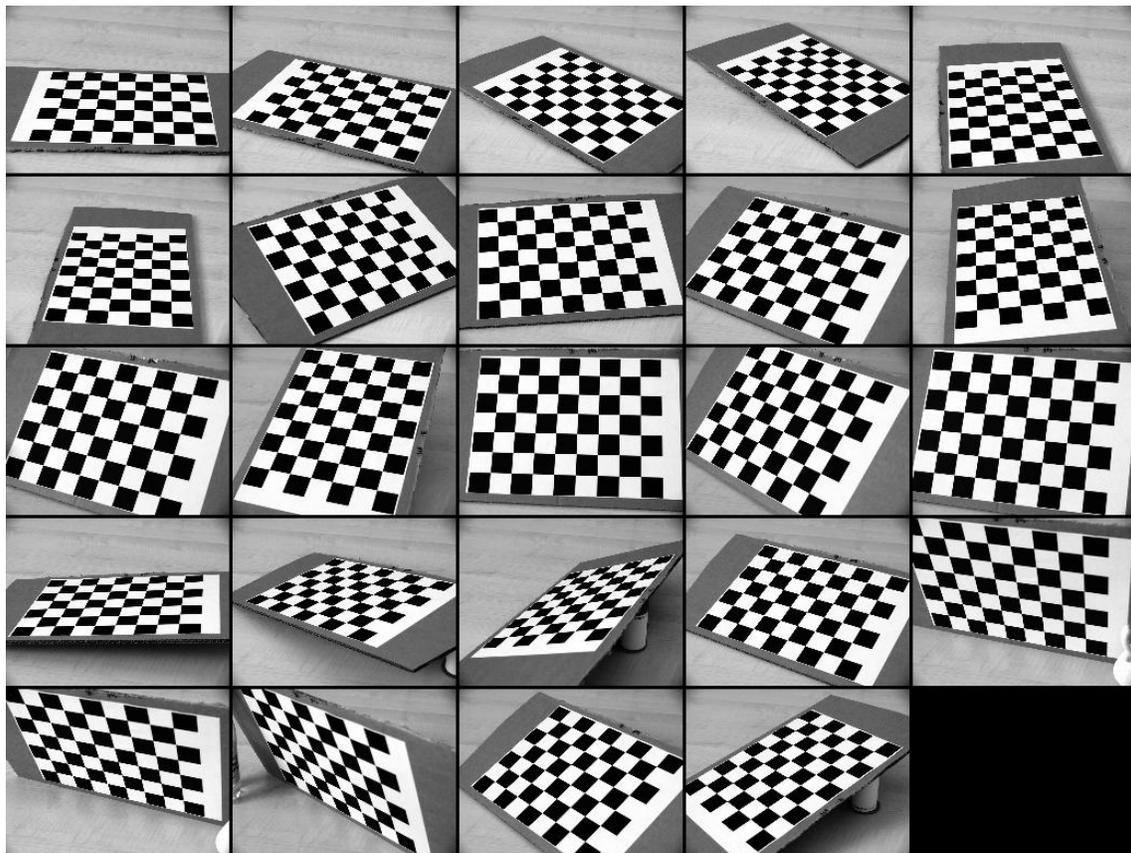


Figura 2.10. Mosaico delle 24 pose scattate dall'ottica destra, usate per la calibrazione

Il toolbox richiedeva poi di inserire la dimensione del lato del quadrato come informazione a priori e procedeva con l'identificazione degli angoli in maniera autonoma per ogni immagine, richiedendo se il risultato ottenuto fosse accurato. E' possibile infatti che l'algoritmo usato non riesca ad identificare gli angoli dei quadrati in condizioni di forti distorsioni introdotte dall'ottica, nel qual caso si suggerisce di inserire un valore iniziale del coefficiente di distorsione radiale  $k_c$  per modificare l'immagine e procedere con l'identificazione degli angoli. Dopo l'estrazione degli angoli viene automaticamente generato un file di dati matlab (calib\_data.mat) che contiene tutte le informazioni raccolte durante questa fase (coordinate dell'immagine, coordinate 3D della griglia di punti estratta, dimensioni della griglia, etc.) e vengono mostrati i parametri calcolati.

#### Parametri fotocamera destra:

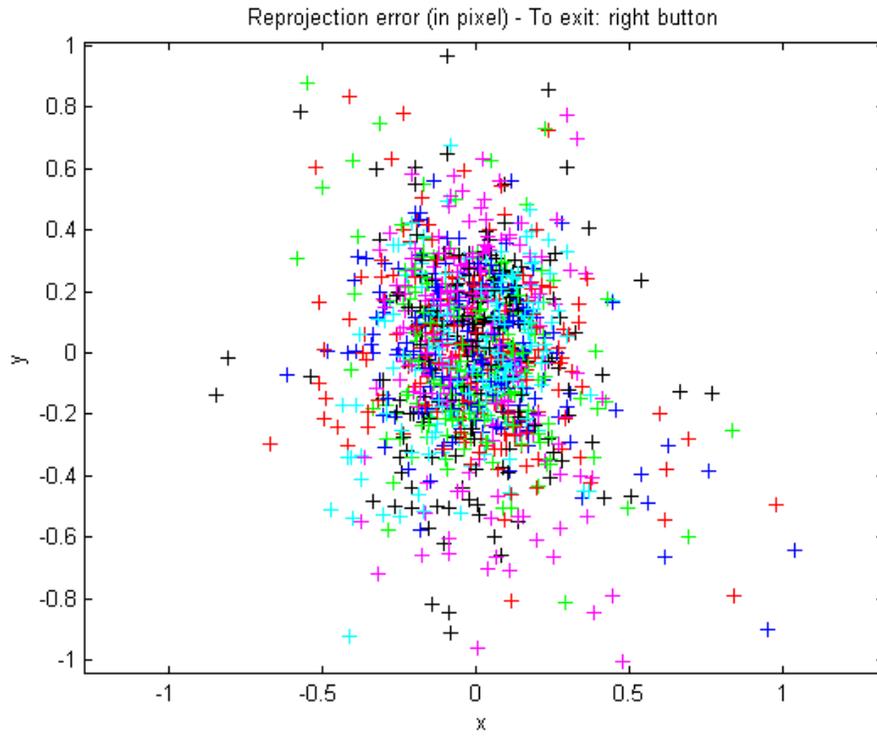
```
Focal Length:    fc = [ 2305.37836  2323.41976 ] ± [ 6.03179  6.95081 ]
Principal point: cc = [ 428.58980  415.74752 ] ± [ 14.73099  10.81176 ]
Skew:           alpha_c = [ 0.00000 ] ± [ 0.00000 ]
                => angle of pixel axes = 90.00000 ± 0.00000 degrees
Distortion:     kc = [ -0.10628  0.35833  -0.00116  -0.00661  0.00000 ]
                ± [ 0.03427  0.66801  0.00119  0.00173  0.00000 ]
Pixel error:    err = [ 0.21092  0.29894 ]
```

#### Parametri fotocamera sinistra:

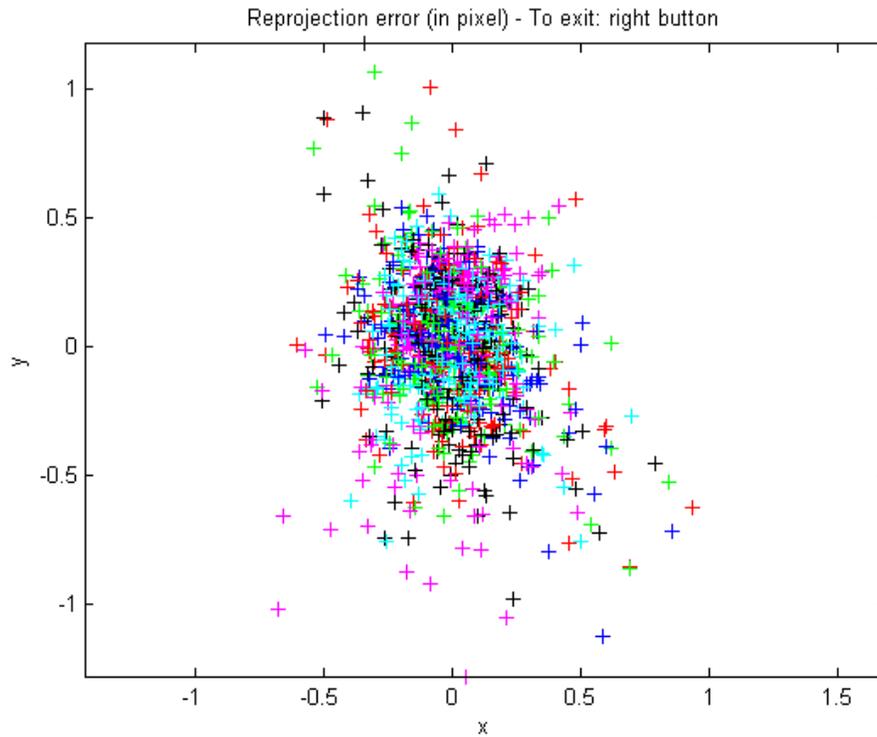
```
Focal Length:    fc = [ 2300.19337  2317.83104 ] ± [ 5.15349  6.05193 ]
Principal point: cc = [ 510.38638  420.72005 ] ± [ 12.55580  8.59118 ]
Skew:           alpha_c = [ 0.00000 ] ± [ 0.00000 ]
                => angle of pixel axes = 90.00000 ± 0.00000 degrees
Distortion:     kc = [ -0.21083  0.23322  -0.00384  0.00088  0.00000 ]
                ± [ 0.02807  0.58023  0.00089  0.00089  0.00000 ]
Pixel error:    err = [ 0.20245  0.28538 ]
```

Il *Camera Calibration Toolbox for Matlab* fornisce anche utili strumenti per l'analisi dei risultati ottenuti. In primo luogo si può osservare l'errore di riproiezione (*reprojection error*) graficato come una nuvola di punti. I punti della griglia acquisita (di ognuna delle 24 immagine di calibrazione) sono codificati con un colore differente.

Errore di riproiezione per la fotocamera destra:

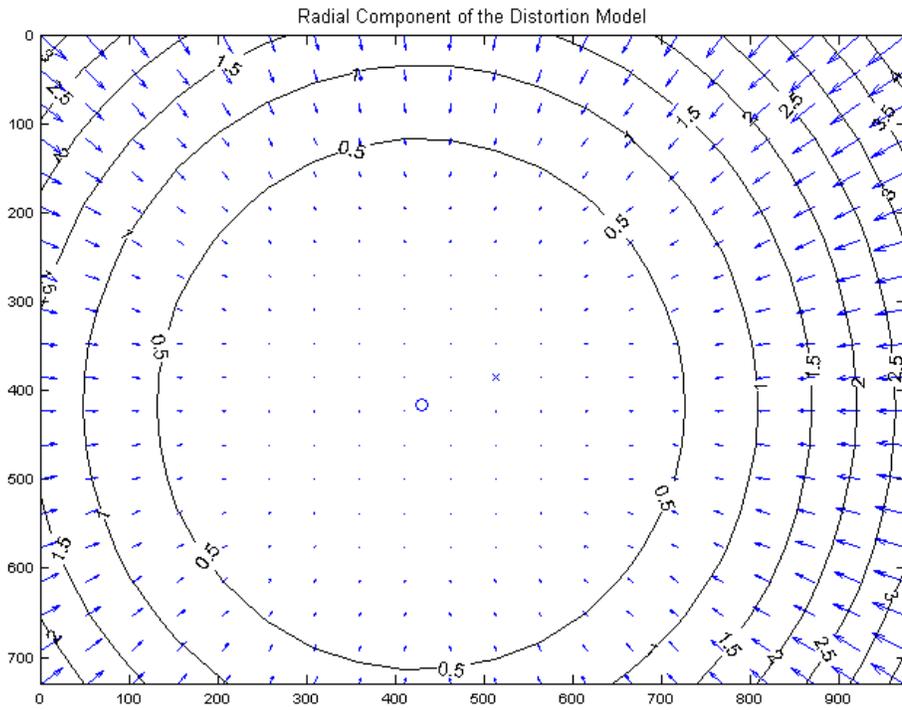


Errore di riproiezione per la fotocamera sinistra:

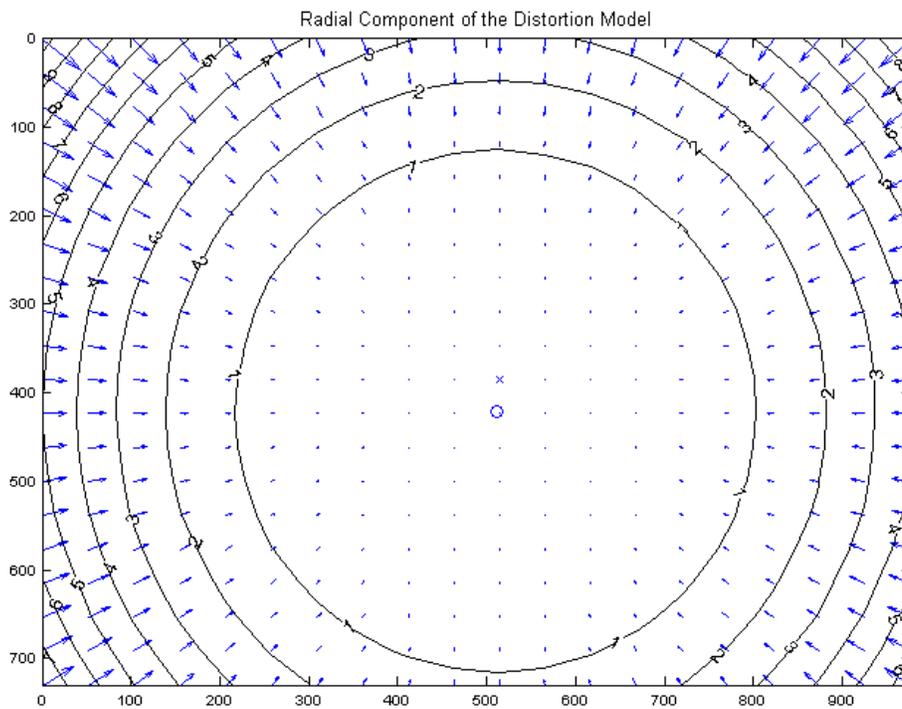


Per ognuna delle ottiche è fornito il modello di distorsione radiale, tangenziale e globale.

Distorsione radiale per l'ottica destra:



Distorsione radiale per l'ottica sinistra:



Per ogni immagine di calibrazione, il *Camera Calibration Toolbox for Matlab*, una volta identificati i punti della griglia, è in grado di generare un sistema di riferimento locale (O, X, Y, Z) in cui l'origine si trova sul primo angolo (identificato manualmente) e l'asse Z risulta ortogonale al piano su cui giace la scacchiera. E' quindi possibile calcolare quale sia la disposizione dei punti della griglia rispetto al sistema di riferimento della fotocamera (O<sub>c</sub>, X<sub>c</sub>, Y<sub>c</sub>, Z<sub>c</sub>) secondo una trasformazione lineare (rototraslazione), utilizzando i *parametri estrinseci* calcolati per ogni fotogramma di calibrazione (una matrice di rotazione 3x3 ed un vettore di traslazione 3x1).

Sia **P** un punto nello spazio di coordinate  $XX=[X;Y;Z]$  rispetto al sistema di riferimento della griglia; siano  $XX_c=[X_c;Y_c;Z_c]$  le coordinate di **P** nel sistema di riferimento della fotocamera. La relazione che esiste tra  $XX$  e  $XX_c$  è un'equazione di moto rigido:

$$XX_c = Rc\_1 * XX + Tc\_1$$

In particolare, il vettore di traslazione  $Tc\_1$  è la coordinata dell'origine del pattern a scacchiera nel sistema di riferimento della fotocamera e la terza colonna della matrice  $Rc\_1$  è il vettore normale al piano contenente la griglia di punti sempre nel sistema di riferimento fotocamera.

Poiché ogni posa della scacchiera è stata acquisita da due fotocamere è possibile procedere alla calibrazione del sistema stereo, calcolando la posizione reciproca delle due fotocamere e completando il calcolo dei *parametri estrinseci*.

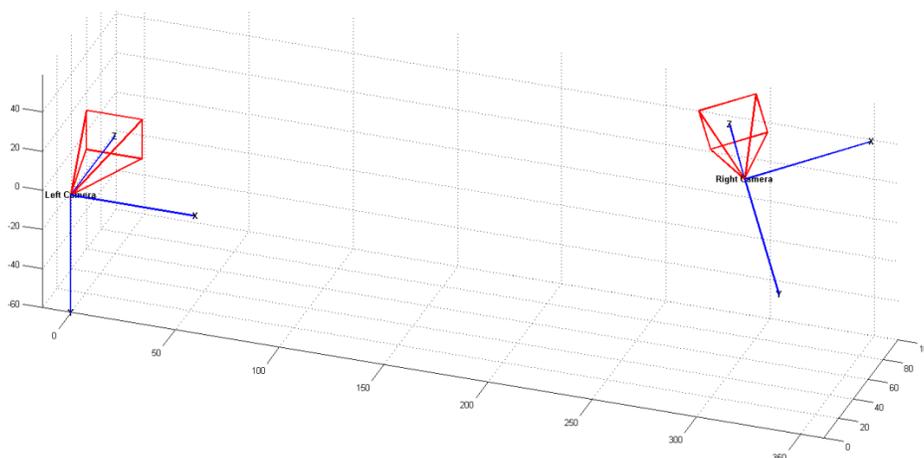


Figura 2.11. Posizione reciproca ricostruita delle due fotocamere

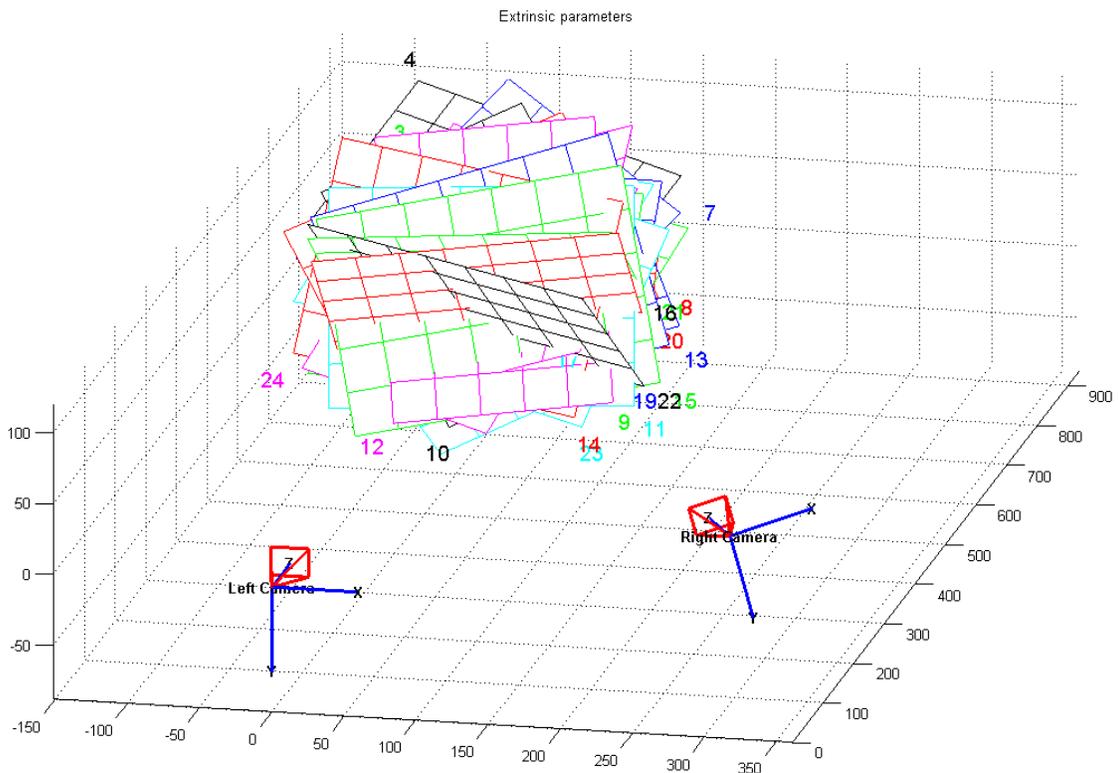
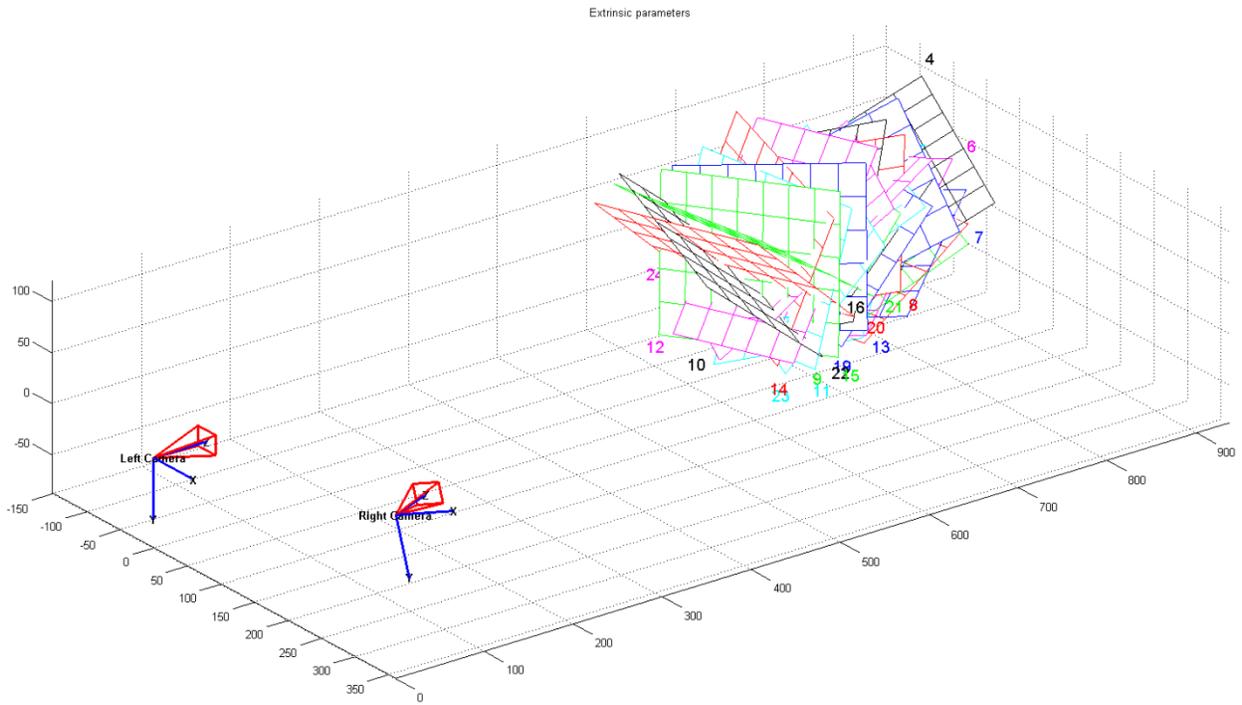
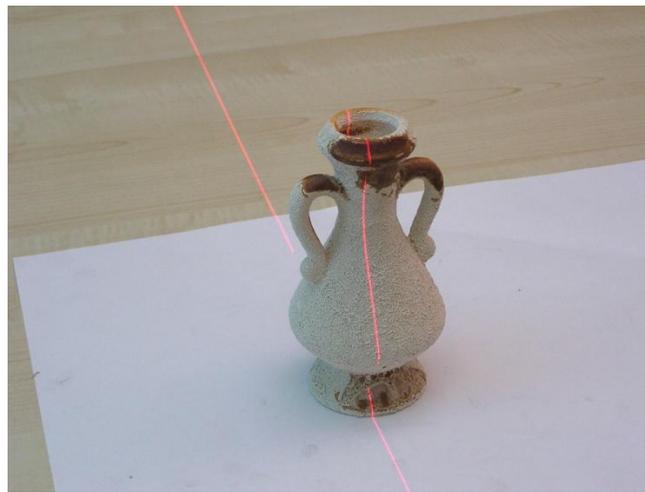


Figura 2.12. Due viste del sistema stereoscopico ricostruito con le 24 pose di calibrazione

Una volta stimati i parametri intrinseci e quelli estrinseci è possibile procedere alla rettificazione delle coppie di immagini acquisite, correggendo le distorsioni introdotte dai gruppi ottici, portando le immagini alla stessa distanza focale (fattore di scala o di zoom) e portando il sistema alla forma standard, ricercando i punti omologhi sulle righe corrispondenti.

Prima di operare alla rettificazione, si cerca un modo per aumentare la visibilità della lama laser proiettata sull'anfora. Lo scopo è infatti l'identificazione della sola posizione su ogni riga della coppia di immagini della lama laser proiettata. Saranno quindi ignorate le informazioni sull'intensità luminosa o sul colore originale per la ricostruzione del modello tridimensionale, ma queste risultano utili per il pre-processing necessario.

Per poter mettere in evidenza la riga laser proiettata, si è scelto di utilizzare l'informazione (nota a priori) sul colore. Infatti si acquisiscono immagini a colori, composte quindi dai 3 canali R, G e B, ognuno dei quali è considerabile come un'immagine in scala di grigio.



Canale R

Canale G

Canale B

Figura 2.13. Scomposizione di un fotogramma nei 3 canali colore (R,G e B)

Essendo rossa, la lama è più visibile sul canale del rosso e poco o nulla sui canali del verde e del blu. Invece di utilizzare il solo canale del rosso si è scelto di comporre una nuova immagine miscelando i 3 canali:  $I = 1 \cdot R - 0.5 \cdot G - 0.5 \cdot B$ . L'idea è quella di sottrarre luminosità nelle zone di scarsa rilevanza; i tre coefficienti possono essere ottimizzati.

In questo modo si ottiene un'immagine in scala di grigio, massimizzando l'intensità luminosa della lama e scurendo drasticamente tutto il resto (nell'ipotesi in cui nell'immagine non ci siano porzioni dello stesso colore della lama).

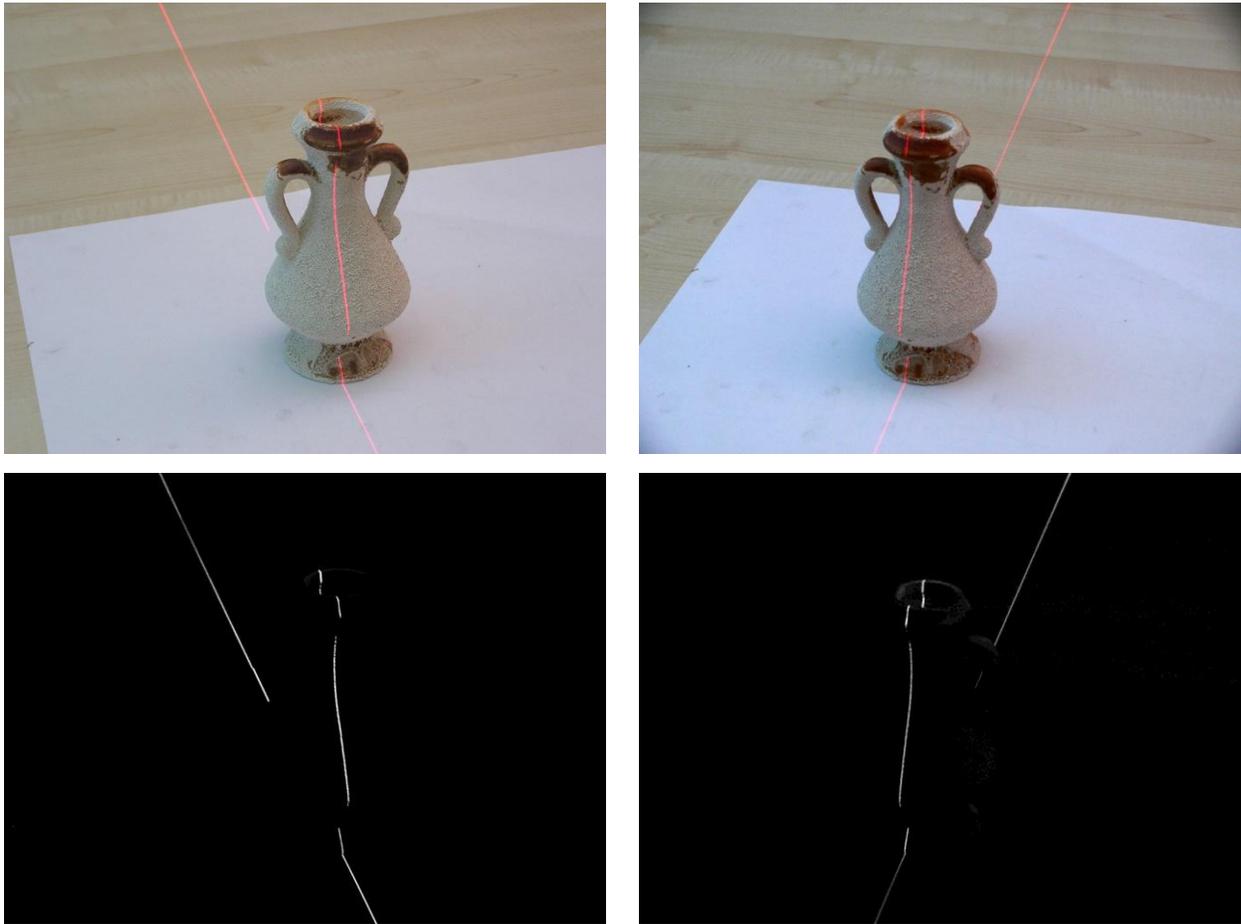


Figura 2.14. Processamento di una coppia di fotogrammi. Mescelamento dei canali colore

Nelle immagini risultanti in scala di grigio, oltre alla lama, rimangono presenti molte zone di intensità luminosa non trascurabile. Si rende necessario l'utilizzo di strategie aggiuntive per l'identificazione univoca dei pixel di interesse.

Rettificando le immagini si ottiene una nuova collezione di coppie in cui i punti omologhi giacciono sulla stessa riga.



Figura 2.15. Coppia di fotogrammi rettificati

Come si vede dagli ultimi fotogrammi, la rettificazione introduce delle zone bianche negli angoli dopo la rotazione delle immagini. Questo diventa un problema per la ricerca della linea utile, la lama laser, ed è necessario cancellare queste zone, o riducendo la zona osservata ad una sottoporzione del fotogramma o costruendo una maschera ad hoc.

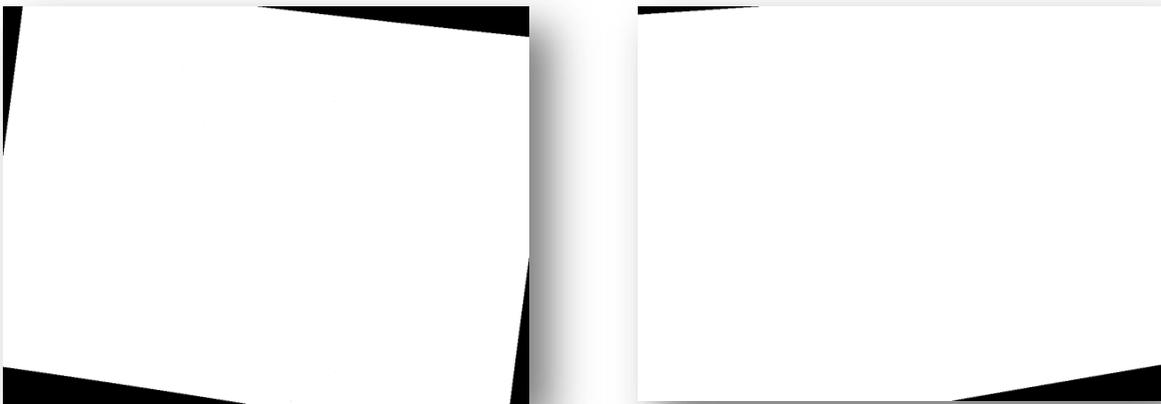


Figura 2.16. Coppia di maschere per cancellare le zone bianche

E' inoltre comodo eliminare la proiezione della lama laddove non è strettamente necessario ricostruire il modello come ad esempio sul piano del tavolo. A tale scopo si è fatto uso di una coppia di immagini a vuoto, in cui è stata rimossa l'anfora.

Una volta fatte le dovute correzioni e semplificazioni, rimangono evidenti i soli punti utili al calcolo della disparità. Il problema del matching stereo è completamente risolto avendo imposto l'esistenza di un solo punto per ogni riga, e la soluzione è univoca, quando presente. L'obiettivo da raggiungere adesso è l'identificazione dell'ascissa alla quale, su ogni riga, si trova la lama laser.

La riga proiettata ha una propria larghezza che risulta essere di più di un pixel nell'immagine acquisita ed inoltre l'andamento della sua intensità lungo una direzione ortogonale ha tipicamente una forma a campana. Esistono diversi algoritmi per la determinazione del picco di un segnale, con accuratezza subpixel, ed uno studio è stato fatto da D.K.Naidu e R.B.Fisher [9] nel quale risulta essere un buon metodo, robusto in presenza di rumore e di elevata efficienza computazionale, quello proposto da Blais *et al.*.

L'algoritmo di determinazione della "linea dei centri" della lama laser è basato sulla differenziazione del segnale e sulla successiva riduzione del rumore ad alta frequenza, mediante l'uso di un filtro a risposta impulsiva finita (FIR). In particolare, considerando che le linee laser hanno un'orientazione prevalentemente verticale, il segnale filtrato nel pixel di coordinate  $(x, y)$  è ottenuto mediante un operatore di ordine  $2n$  che fornisce il segnale filtrato:

$$I_f = \sum_{i=-n}^n -\text{sign}(x) \cdot I(x + i, y)$$

dove  $I(x, y)$  è l'intensità del livello di grigio dell'immagine di partenza.

Il filtro in esame (con  $n = 3$ ) ha una risposta impulsiva così definita:

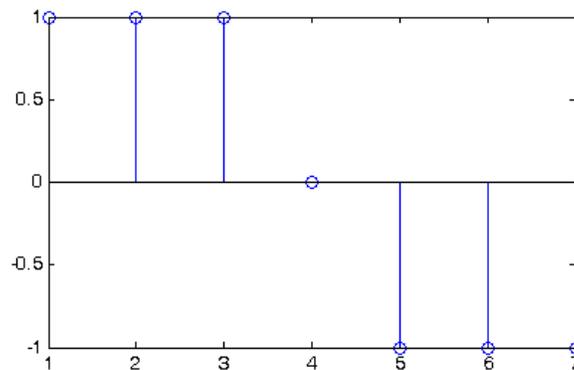


Figura 2.17. Risposta impulsiva del filtro usato per la rivelazione della lama laser

L'immagine filtrata ha il seguente aspetto:

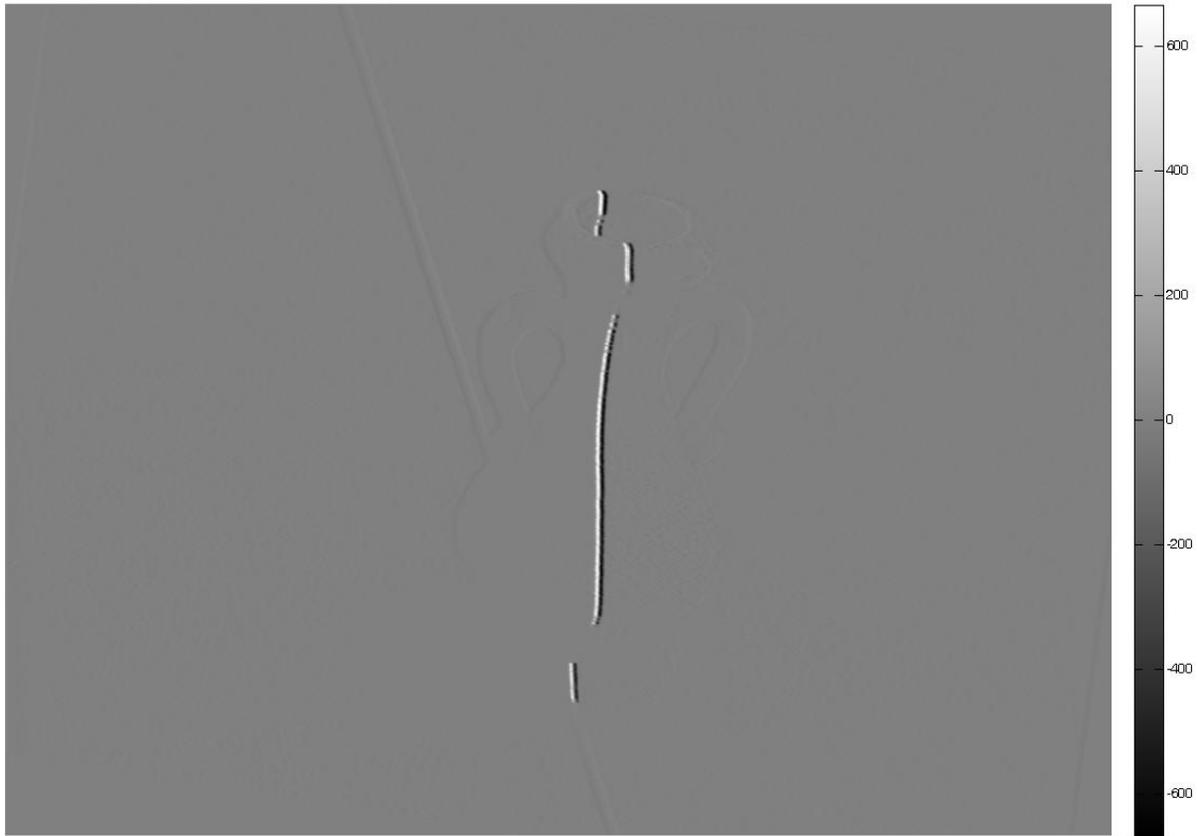


Figura 2.18. Un fotogramma processato col filtro FIR  $I_f$

Analizzando una singola riga (di ordinata  $y_c$ ) in presenza della lama laser, si identifica rapidamente la presenza della linea osservando un massimo assoluto seguito da un minimo nell'intensità della luminosità.

Facendo l'analisi di un singolo fotogramma, ed uno studio riga per riga, vediamo nel dettaglio cosa accade; nell'immagine seguente viene mostrata l'intensità luminosa della riga 186 (in cui è presente la lama laser) di un fotogramma non ancora filtrato:

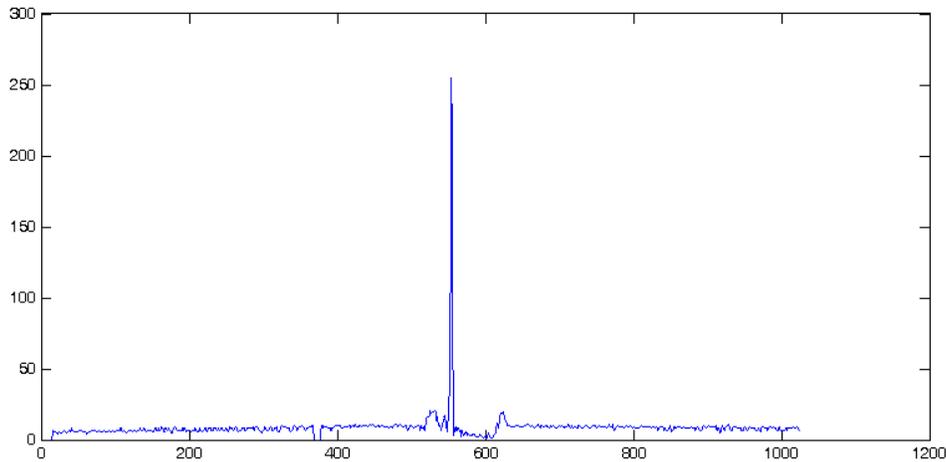


Figura 2.19. Il picco di segnale in presenza di lama laser

Si può osservare il picco del segnale in prossimità della lama laser immerso in un sottofondo “rumoroso” in cui possono eventualmente comparire altri picchi non voluti, la cui intensità dovrebbe comunque essere inferiore a quella del laser.

Dopo il filtraggio, l'intensità luminosa si presenta così:

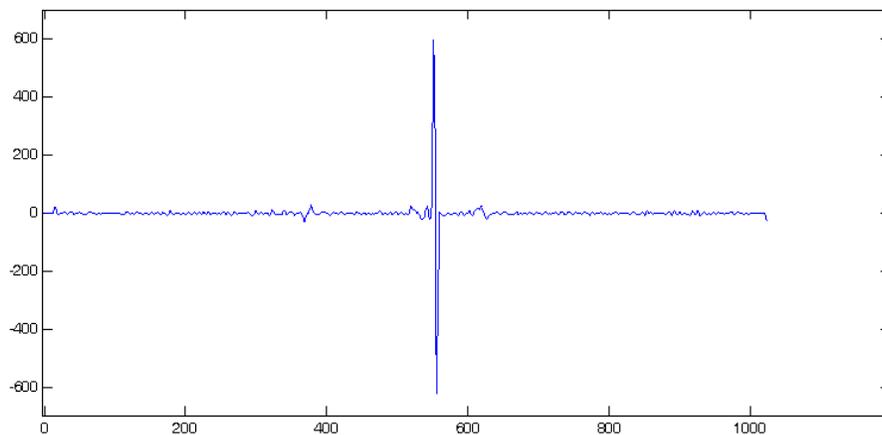


Figura 2.20. Andamento del segnale filtrato in presenza di lama laser

Il filtro ha un comportamento derivativo e si osserva infatti che il segnale non è più strettamente positivo come quello non filtrato, ma ha valori anche negativi, inoltre varia mediamente intorno allo zero. L'intensità nel punto di massimo (o di minimo) è molto alta rispetto ai picchi secondari e questo rende possibile l'utilizzo di una soglia per discernere quale sia il valore di interesse.

Osserviamo cosa accade intorno al picco della singola riga dell'immagine non filtrata: si può facilmente distinguere la forma a campana dell'intensità lunimosa della lama laser.

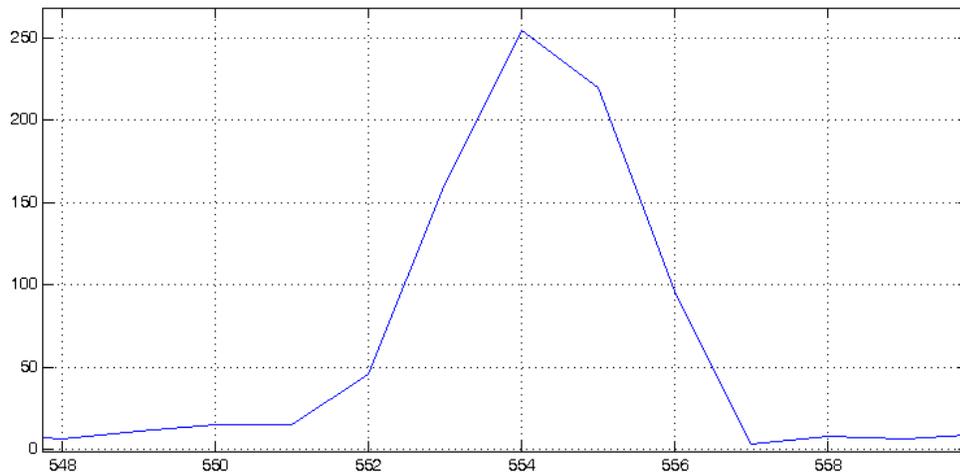


Figura 2.21. Ingrandimento intorno al massimo del segnale

Nell'immagine filtrata si osserva il comportamento derivativo del filtro:

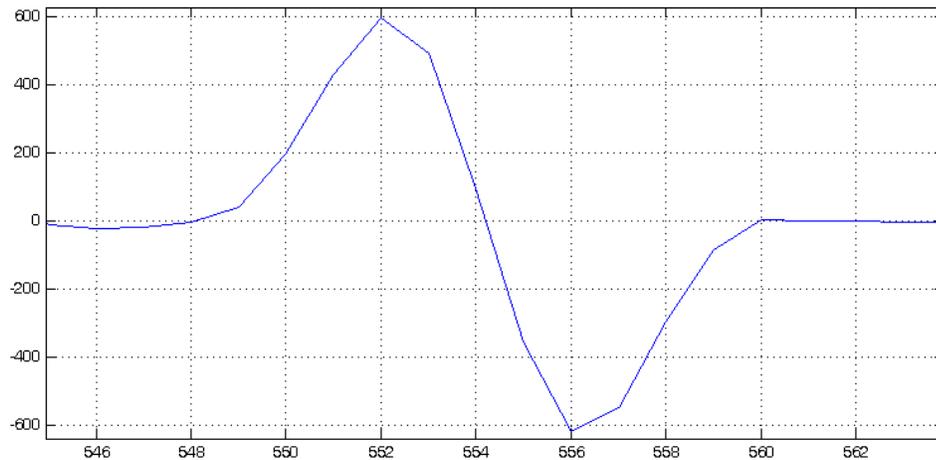


Figura 2.22. Ingrandimento intorno al passaggio per lo zero utile

La ricerca del picco potrebbe fermarsi alla scelta del pixel più luminoso di ogni riga ma questa è un'approssimazione che non fornisce un risultato accurato per almeno due motivi. Supponendo l'andamento a campana, non è detto che questo punto sia realmente il suo picco. In secondo luogo, il successivo calcolo della disparità fornirà risultati discreti, multipli del passo di un pixel. Si cerca invece di identificare l'attraversamento per lo zero del segnale filtrato, mediante interpolazione lineare, al quale corrisponde il picco supposto.

Sia  $y_c$  l'ordinata della riga in esame e  $x_c$  l'ascissa del pixel immediatamente prima dell'attraversamento dello zero del segnale filtrato. E' possibile calcolare con accuratezza subpixel la coordinata orizzontale per interpolazione lineare, conoscendo i valori dell'intensità di grigio nel punto  $I_f(x_c, y_c)$  e  $I_f(x_c + 1, y_c)$ . Lo scostamento  $\delta$  dalla coordinata  $x_c$  sarà:

$$\delta = \frac{I_f(x_c, y_c)}{I_f(x_c, y_c) - I_f(x_c + 1, y_c)}$$

e dunque la coordinata del picco del segnale, con accuratezza subpixel, sarà  $(x_c + \delta, y_c)$ .

Ripetendo questa operazione per ogni riga di ogni fotogramma si ha a disposizione la posizione dei "centri" della lama laser, dove presente, e si può procedere al calcolo della disparità. Per ogni coppia di fotogrammi (destra e sinistra) è adesso possibile ricostruire la posizione della lama nello spazio ed infine, unendo le singole ricostruzioni dei profili, si può dare forma al modello 3D dell'oggetto in esame.

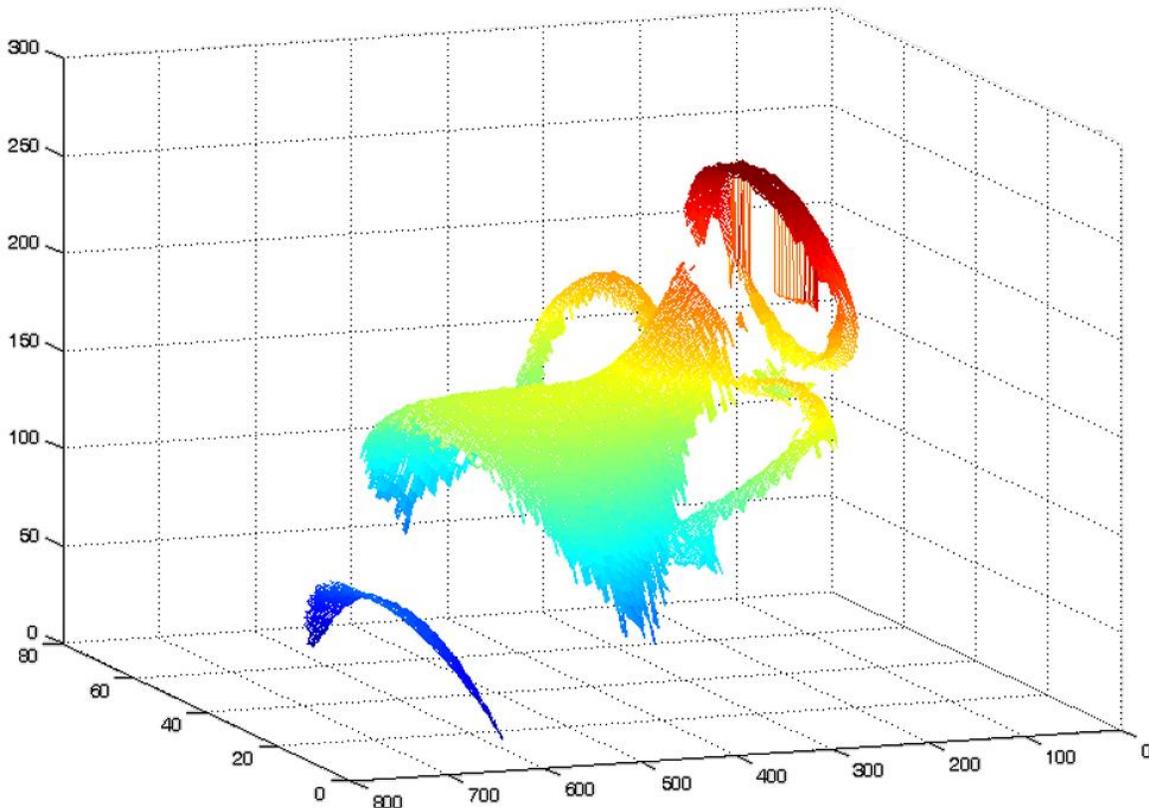
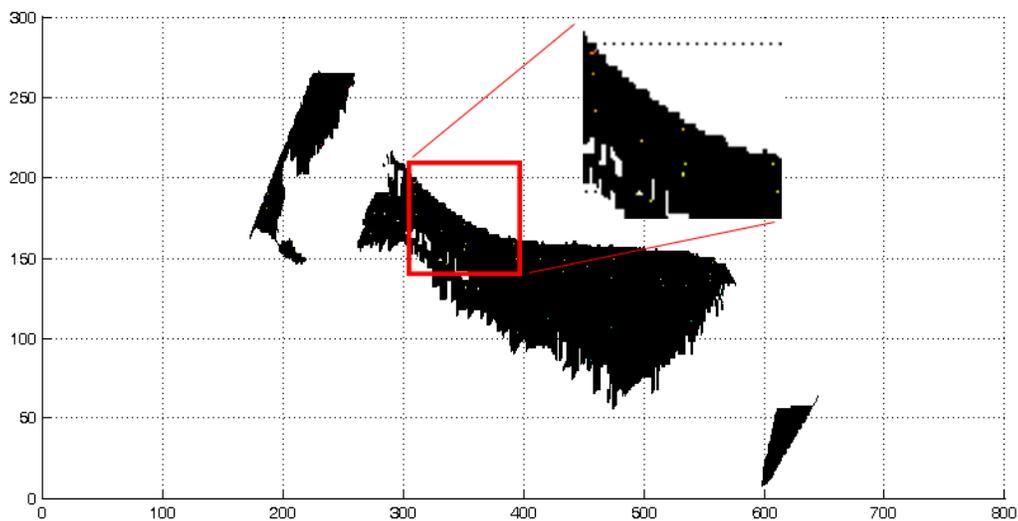


Figura 2.23. Ricostruzione della nuvola di punti densa 3D

Questa operazione porta alla creazione di una nuvola densa di punti in uno spazio 3D, la quale descrive la superficie osservata dal punto di vista frontale in cui sono posizionate le camere.

E' possibile fare un confronto con i punti che si otterrebbero se non si valutasse la posizione del picco con accuratezza sub-pixel. Osservando le due nuvole di punti lateralmente è evidente la quantizzazione introdotta nella ricostruzione scarsamente accurata.

Accuratezza pixel:



Accuratezza sub-pixel:

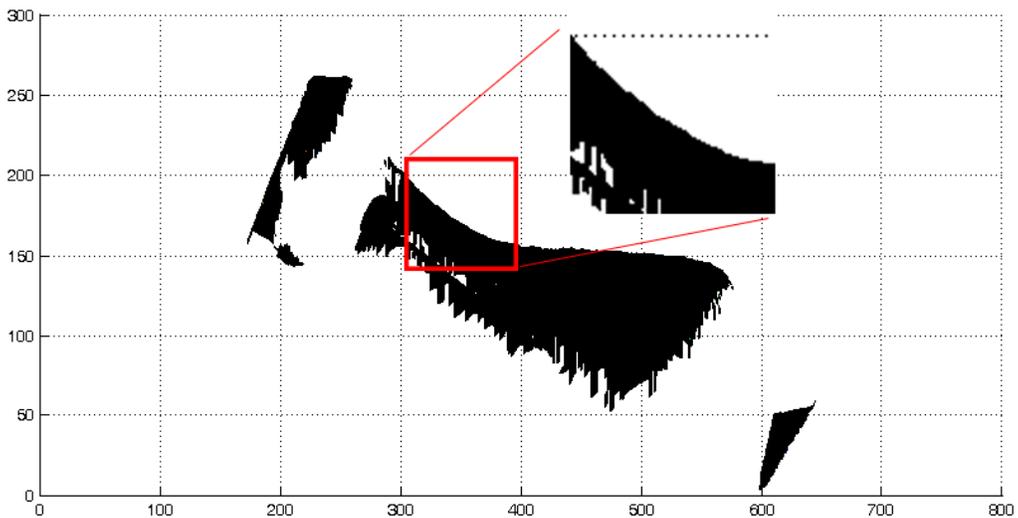


Figura 2.24. Confronto tra la ricostruzione semplice e quella con accuratezza subpixel

Osservando nel dettaglio la ricostruzione senza accuratezza sub-pixel si nota l'effetto quantizzazione della terza dimensione ricostruita:

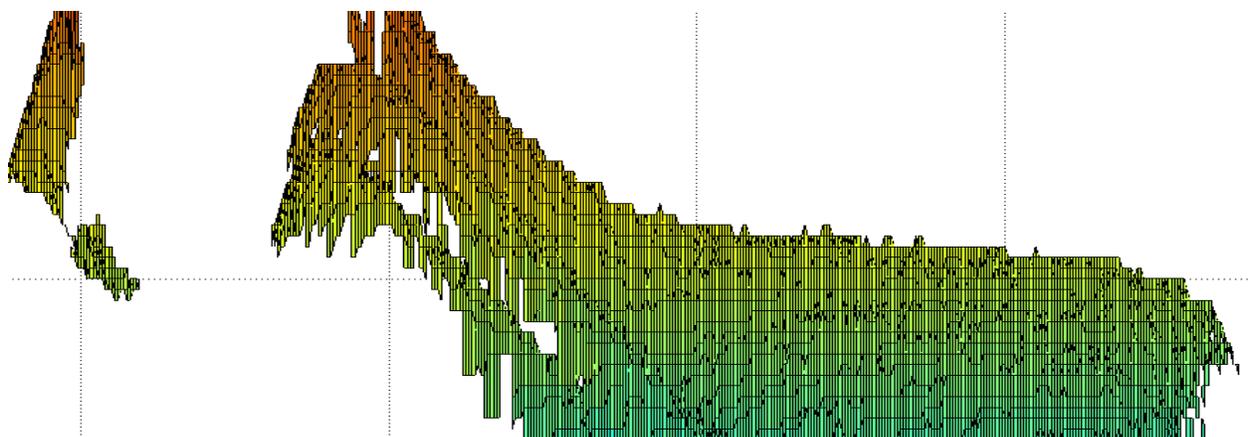


Figura 2.25. Dettaglio di ricostruzione senza accuratezza subpixel. Si nota l'effetto "quantizzazione"

Avendo adesso a disposizione la nuvola di punti con accuratezza sub-pixel è possibile costruire una mesh poligonale che rappresenta la superficie dell'oggetto ricostruito. Una mesh poligonale non è altro che una collezione di vertici, di spigoli e di facce tra loro connesse, che rendono possibile anche la visualizzazione e la navigazione in realtime con software di modellazione 3D e l'eventuale rendering grafico, con l'applicazione di texture che rendono leggibile e verosimile il modello ricostruito.

Si pone adesso il problema della generazione delle informazioni minime mancanti alla creazione della mesh, ovvero della generazione delle facce poligonali.

La scelta più semplice è la faccia triangolare. Con uno script apposito in MatLab si è fatto in modo che, dati 3 punti adiacenti della nuvola densa, si costruisca una faccia piana triangolare. Si è scelto di utilizzare un formato comune a molti software dedicati, come il formato "obj" in cui è sufficiente elencare vertici e facce per descrivere integralmente l'oggetto grafico. Le facce sono orientate (l'ordinamento dei 3 punti costituenti è significativo), ovvero hanno una normale uscente. Questo serve al software di elaborazione grafica per il rendering dell'illuminazione in realtime, trattando in maniera differente la parte esterna da quella interna.

Per la visualizzazione realtime del modello si è usato il software open source “MeshLab” sviluppato nel laboratorio di Visual Computing dell’ISTI-CNR [10]. Questo software nasce con l’intento di fornire utilità nel trattamento di mesh triangolari dense, che nascono appunto dall’acquisizione di oggetti reali tramite scanning 3D, implementando anche dei tools per la modifica, la pulizia, il risanamento, l’ispezione e il rendering dei dati acquisiti.

L’immagine seguente mostra uno screenshot della mesh poligonale grezza aperta col software suddetto, in cui non si è fatta nessuna modifica sostanziale.

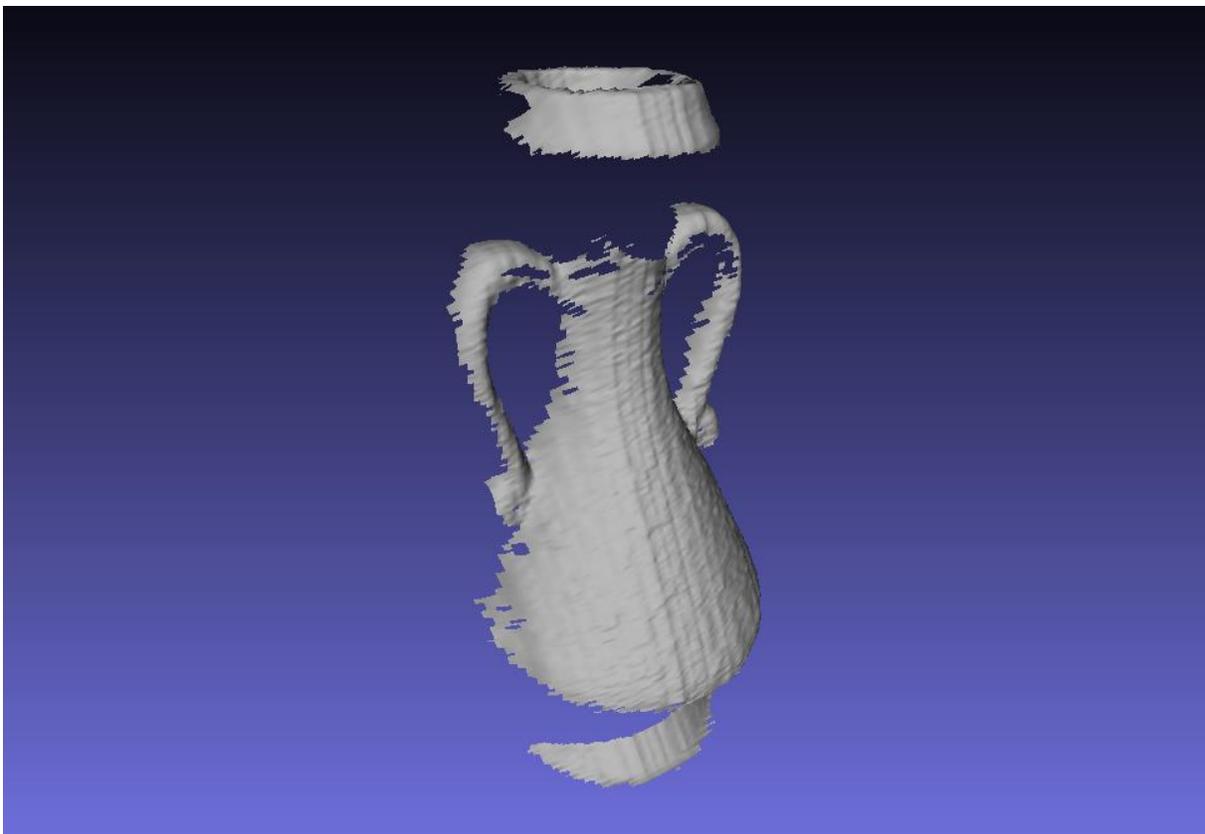


Figura 2.26. La nuvola densa di punti e le facce ricostruite e visualizzate in Meshlab

Si può osservare una disomogeneità nella superficie della pancia dell’anfora ricostruita che può essere attribuita alla scarsa precisione nella raccolta dei dati, avendo utilizzato un sistema allestito come prototipo. Si ricorda che questo test era mirato allo studio della tecnica e dei suoi limiti, e non all’ottimizzazione del risultato.

Utilizzando i tools per l'affinamento della mesh poligonale è possibile limitare le imprecisioni ed ottenere superfici più o meno levigate, a seconda delle necessità, con eventuale perdita dei dettagli ma preservando le forme ed i contorni.



Figura 2.27. La mesh poligonale è stata affinata usando un tool di Meshlab dedicato

Altri tipi di tool possono risanare la mesh andando a chiudere i buchi le cui dimensioni rientrano entro un certo limite, modificando sostanzialmente la nuvola di punti acquisita, ma garantendo continuità nella ricostruzione delle superfici più morbide.

L'ultimo passo per una ricostruzione visuale è l'utilizzo di texture per ricoprire la superficie del modello. E' possibile associare una texture idonea per il tipo di superficie supposta, utilizzando una collezione già esistente, oppure ricavare le texture dalle acquisizioni, quando possibile, per garantire maggior realismo.

Nell'immagine seguente è stata aggiunta una texture sfruttando direttamente una delle immagini memorizzate durante l'acquisizione:



Figura 2.28. Ricostruzione finale del modello 3D dell'anfora di prova, con l'aggiunta di texture

In questa ricostruzione si possono osservare chiare lacune dovute all'impossibilità di rilevare la lama laser da parte delle ottiche. Questo è dovuto a due fattori principali. Per prima cosa la lama laser non è stata proiettata in maniera ottimale e porzioni dell'oggetto (testa e pancia) oscuravano la sua proiezione sulle zone sottostanti. Sarebbe stato sufficiente proiettare il laser da un'angolazione differente per ottenere risultati migliori. In secondo luogo c'è il problema delle oclusioni: ogni ottica vede una scena leggermente diversa dall'altra ed in qualche punto la lama laser è visibile da un'ottica soltanto. Questo purtroppo è inevitabile e dev'essere considerato come un limite di questa tecnica.

# Capitolo 3 - Sensori Acustici

## *Introduzione*

Un modo alternativo per l'analisi del fondo marino si basa sull'utilizzo di sensori acustici, più semplicemente noti come SONAR. Tra le forme di radiazione subacquea conosciute, il suono si propaga in mare nella maniera migliore. Infatti, nelle acque più o meno torbide e ricche di sali minerali, sia la luce che le onde elettromagnetiche subiscono un'attenuazione di gran lunga superiore rispetto a questa forma di energia meccanica, meglio nota come suono. Proprio per questa semplicità relativa di propagazione, i campi di applicazione dell'acustica sottomarina sono molteplici e molto diversi tra loro.

I dispositivi che si possono incontrare nella generazione o nella rivelazione di segnali acustici sottomarini sono i costituenti di un sistema complesso, conosciuto come sistema sonar. La parola sonar è un acronimo derivante dall'espressione inglese "SOund NAVigation and Ranging" e questo ci fornisce già alcune informazioni sull'utilizzo possibile dell'acustica sottomarina. L'impiego del suono a cui siamo più abituati è indubbiamente la comunicazione (verbale e non) e sembra apparentemente strano che si possa utilizzare il suono anche per l'orientamento e per la localizzazione. Il fatto però non deve stupire, basti pensare che un

sistema equivalente a quello artificiale sviluppato dall'uomo, è usato in natura da alcuni mammiferi, come ad esempio dai pipistrelli o dai cetacei marini, ed è chiamato appunto biosonar. Si osserva che gli habitat naturali di questi animali rendono difficile, se non impossibile, utilizzare la luce come fonte di informazioni per l'orientamento.

In mare, grazie al suono, tramite l'ecolocalizzazione è possibile localizzare, identificare, stimare la natura degli oggetti che ci circondano, stimare la profondità del fondale o la presenza di pericolose asperità. Per fare questo è necessario produrre un suono di una certa intensità nell'ambiente ed ascoltarne le eventuali eco prodotte dai cosiddetti "ostacoli" circostanti. Il modo in cui l'eco ritorna al punto di partenza fornisce importanti informazioni da cui è possibile estrapolare le caratteristiche dell'ostacolo su cui è stato riflesso il suono, come ad esempio la distanza, la posizione spaziale, una stima della dimensione, etc. Si potrebbe dire, con un certo grado di semplificazione, che l'acustica sottomarina gioca lo stesso ruolo che il radar e le onde radio in generale giocano nell'atmosfera e nello spazio libero, nonostante le performance siano molto più basse. Le onde acustiche subacquee sono principalmente usate per [11]:

- *Identificazione e localizzazione* di ostacoli o bersagli; questa è la prima funzione di un sistema sonar, principalmente per applicazioni militari ma anche per la navigazione e per la pesca;
- *Misura* delle caratteristiche dell'ambiente marino (topografia del fondale, organismi viventi, correnti e strutture idrologiche, etc.) o *stima della posizione e della velocità* di un oggetto in movimento;
- *Trasmissione* di segnali, che possono essere dati acquisiti da strumenti scientifici, messaggi scambiati tra semoventi marini (ad esempio comandi operativi verso un dispositivo autonomo come un AUV)

I dispositivi acustici sottomarini sono, per la maggior parte dei casi, *sistemi attivi*, cioè trasmettono un segnale caratteristico e controllato (quindi noto) che sarà successivamente riflesso da un bersaglio o trasmesso direttamente ad un ricevitore. Viceversa i *sistemi passivi* sono progettati per intercettare le onde acustiche generate e irradiate dal bersaglio stesso.

## ***Struttura di un sistema sonar attivo***

Per identificare e localizzare un particolare bersaglio, il sonar invia segnali acustici di elevata potenza in una direzione predefinita a mezzo di un trasduttore acustico. I segnali si propagano nel mare verso il cosiddetto bersaglio che rifletterà parte dell'energia acustica nuovamente verso il sonar; lungo il loro percorso bidirezionale, questi saranno attenuati, deformati e mixati con la riverberazione ed il rumore ambientale. I segnali saranno dunque ricevuti per mezzo di una *antenna*, possibilmente un array costituito da diversi trasduttori (idrofoni), la quale uscita è combinata per ottenere un certo diagramma di direttività. Dopo una fase di preamplificazione, digitalizzazione e filtraggio, i segnali entrano in una catena di processing dedicata. Considerando i limiti imposti dall'ambiente fisico sottomarino, il design dei segnali ed il loro processing influenza direttamente le performance dei ricevitori sonar. L'identificazione nel rumore e la localizzazione spaziale sono rese possibili dall'utilizzo di antenne direzionali, il quale scopo è di incrementare l'efficienza (in trasmissione) ed il rapporto segnale/rumore (in ricezione) e di stimare gli angoli d'arrivo usando la loro stessa direttività o la selettività spaziale.

## ***Il Sidescan Sonar (SSS)***

I *sonar a scansione laterale* o *Sidescan Sonar* (detti anche SSS) sono i principali strumenti di visualizzazione acustica, e forniscono una *immagine acustica* del fondale marino. Solitamente sono installati su una piattaforma trainata o su semoventi autonomi come gli AUV, che operano vicino al fondale. Questa configurazione rende possibile l'operatività in condizioni di buona stabilità e in buone condizioni di rumore. Il sidescan sonar "sonorizza" il fondale per mezzo di due antenne laterali con direttività orizzontale estremamente ridotta (tipicamente qualche frazione di grado). Il segnale scatterato dal fondo, registrato nel tempo, riproduce la struttura delle piccole irregolarità dell'interfaccia.

Il principio di funzionamento di questa tecnica di mappatura acustica è significativo grazie alla sua semplicità: si trasmette uno stretto raggio acustico ad incidenza radente che intercetta il fondale in una striscia sottile che si allarga all'aumentare della distanza.

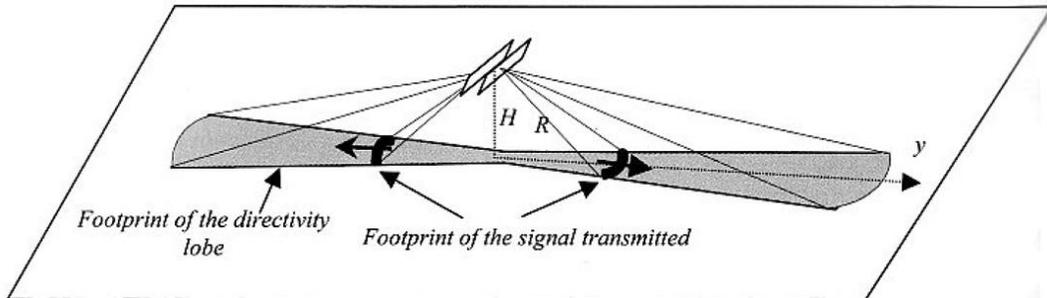


Figura 3.1. geometria della insonificazione del sidesca sonar

All'interno di questa striscia, il segnale trasmesso molto corto delimita l'area sonorizzata istante per istante, dalle dimensioni molto piccole, e si propaga, spazzolando l'intera area coperta. Il segnale ricevuto nel tempo rappresenta la riflettività del fondale lungo la striscia ed in particolare identifica la presenza di irregolarità ed ostacoli. Poiché il segnale è irradiato lateralmente il sistema prende il nome di sonar a scansione laterale, o Sidescan Sonar. Ogni successiva acquisizione temporale è aggiunta alla precedente, linea per linea, ricreando un'immagine fedele del fondo marino.

Le frequenze tipicamente utilizzate sono dell'ordine del centinaio di KHz, con durata dell'impulso temporale più corta possibile, tipicamente 0.1 ms o meno.

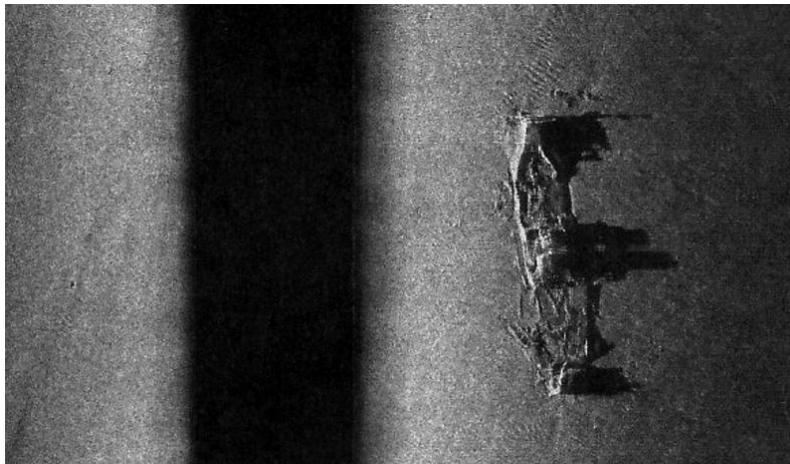


Figura 3.2.  
Immagine prodotta da un Sidescan sonar Klein 5400 (frequenza operativa 455 KHz). Si osservano le ombre prodotte dall'ostacolo. Il relitto è una nave da carico adagiata su un fondale sabbioso all'ingresso della Baia di Brest.

## ***Sequenza in trasmissione***

Il ritardo inter-ping per i sidescan sonar obbedisce a una semplice regola: si trasmette un nuovo ping dopo che l'eco del precedente è stato completamente ricevuto (ovvero il suo livello è diminuito sufficientemente). Il ritardo tra due trasmissioni è quindi:

$$T_R > \frac{2R_{max}}{c}$$

dove  $R_{max}$  è la massima portata obliqua raggiungibile; poiché il sonar lavora ad angoli di incidenza radente la portata obliqua è approssimativamente uguale alla distanza laterale orizzontale.

## ***Trasduttori***

Il sistema è basato su antenne rettangolari molto lunghe (relativamente alla lunghezza d'onda acustica) con un diagramma di irradiazione risultante:

- Molto aperto sul piano verticale (diverse decine di gradi), con l'asse del lobo principale un po' inclinato verso il basso per poter sonificare fino a grande distanza (si raggiungono comunemente angoli fino a 80/85° dalla verticale), per evitare ritorni eco dalla superficie del mare e per limitare la sonorizzazione lungo la verticale stessa.
- Molto stretta sul piano orizzontale (meno di 1° per ottenere la massima risoluzione spaziale). Questa performance di direttività cruciale è migliorata dal fatto che il solito array è usato sia in trasmissione che in ricezione, diminuendo quindi l'apertura del fascio di un fattore  $\sqrt{2}$  ed il livello dei lobi secondari (in dB) di un fattore 2.



Figura 3.3. Esempio di dispositivo commerciale:  
SeaKing ROV/AUV Sidescan Sonar System, Tritech International [12]

Queste antenne sono installate su entrambi i lati del semovente marino. La sua caratteristica idrodinamica deve assicurare una buona stabilità nel movimento. Poiché le frequenze in uso sono generalmente elevate (qualche centinaio di KHz), si raggiungono buoni diagrammi di direttività solo con l'uso di antenne di lunghezza ragionevole. Si consideri che con bande di lavoro di qualche decina di KHz si raggiungono buone risoluzioni (di qualche centimetro). Le distanze ottenibili si limitano a un centinaio di metri. Tipicamente non è richiesto nessun tipo di array signal processing, poiché la geometria dell'antenna è sufficiente ad assicurare le suddette performance di direttività; comunquem in qualche caso di fascio molto stretto e frequenze operative molto elevate, si può pretrattare il segnale.

### ***Ricezione***

La struttura base del ricevitore è simile a quella dei sonar cosiddetti single-beam. La catena di processing è duplicata su entrambi i lati del dispositivo. Una tipica struttura per le operazioni di ricezione di primo livello è la seguente:

- TVG (time varying gain): equalizzazione del livello dell'eco ricevuta
- Demodulazione e filtraggio
- Conversione analogico/digitale (A/D)
- Processing di ricezione non coerente: filtraggio passabanda e rivelazione dell'involuppo (pochi sistemi utilizzano segnali modulati e compressione di impulso)
- Compensazione del guadagno di processing
- Memorizzazione dei dati digitalizzati

In seconda istanza, le eco di ritorno sono registrate e visualizzate come funzioni del tempo: per ogni ping trasmesso, l'intensità scatterata è graficata come una linea di pixel perpendicolare alla rotta del semovente. Sono possibili diverse modalità:

- Come primo approccio, si visualizzano semplicemente una serie di eco successive, registrate nel tempo. L'immagine grezza costruita in questo modo permette il controllo dei dati acquisiti (ed in molti casi ha una qualità figurativa sorprendentemente buona). Comunque questa è una rappresentazione fisicamente povera, insufficiente ad ottenere un'immagine geometricamente accurata del fondale.
- Si presentano problemi di natura geometrica quando si prova a recuperare un'immagine non distorta del fondo, usando queste registrazioni temporali del segnale ricevuto. Poiché il tempo di ricezione e la lunghezza orizzontale (e laterale) sul fondo non sono proporzionali, campioni equidistanti nel tempo non corrispondono a campioni regolari del fondale. Per sostituire i campioni temporali in maniera corretta nello spazio, deve essere applicata una correzione geometrica. Se il fondale è piatto, questa è una semplice relazione trigonometrica: usando la distanza obliqua  $R = ct/2$ , relativa alla distanza dal fondo  $H$  vale semplicemente che  $R^2 = y^2 + H^2$ . L'altezza  $H$  è determinata dalla prima eco sottostante, e si ottiene:

$$y = \sqrt{\frac{c^2 t^2}{4} - H^2}$$

- Quando il fondale non è piatto, la correzione geometrica è più complicata, poiché la formula precedente diventa

$$y = \sqrt{\frac{c^2 t^2}{4} - H(y)^2}$$

Questo richiede un'ipotesi a priori della topografia (ad esempio una pendenza regolare, piuttosto che un fondale piatto), o una conoscenza preliminare della topografia locale, o anche la misura (e registrazione) della batimetria (grazie ad altri sensori sonar dedicati, o al SSS stesso).

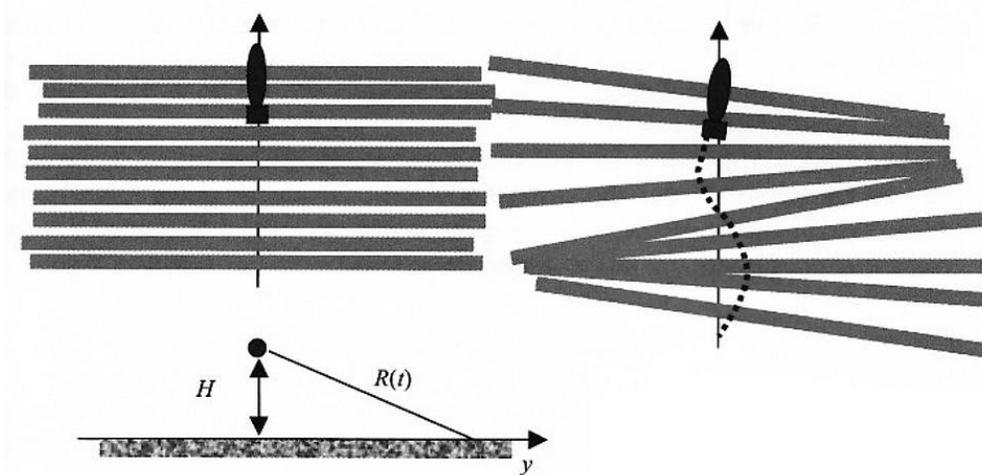


Figura 3.4. Modi differenti di ricostruzione geometrica dell'immagine acustica [11]

- Infine, se il natante è equipaggiato con sensori ausiliari di navigazione e controllo del moto, il suo movimento può essere compensato ed i pixel dell'immagine sonar possono essere rilocate in un preciso fotogramma geografico, in cui si possono aggiungere diverse tracce sonar; la risultante è chiamata *mosaicatura di immagini sonar*.

## Costruzione dell'eco

La generazione spazio-temporale del segnale ricevuto da un sidescan sonar è mostrata nella figura successiva. Il segnale trasmesso si propaga inizialmente nella colonna d'acqua (parte A), ed il sonar riceve quindi solamente rumore di fondo ed occasionalmente alcune eco da bersagli a mezz'acqua (bollicine, pesci, etc.). L'eco effettiva del fondale prende forma solo quando l'impulso colpisce il *nadir* del sonar (parte B). Questa riflessione crea inizialmente una prima eco molto intensa (poiché l'intensità scatterata è massima e la perdita per trasmissione è minima). Questa non può essere utilizzata per l'*imaging* stesso ma è molto utile per stimare l'altezza istantanea del sonar dal fondo.

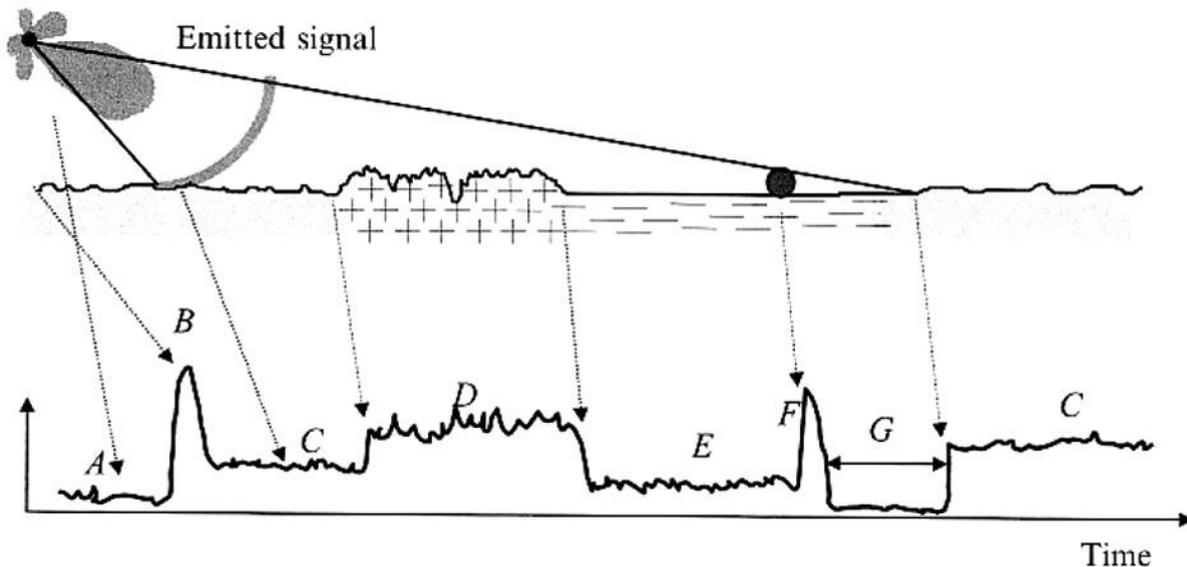


Figura 3.5. Generazione dell'eco: (A) rumore e reverbero nella colonna d'acqua; (B) eco di primo ritorno; (C) area sabbiosa; (D) rocce; (E) fango; (F) eco del bersaglio; (G) ombra del bersaglio [11].

Man mano che si propaga nel tempo, il segnale esplora zone angolari ancora vicine alla verticale. Questa prima parte, con alte riflettività e risoluzione spaziale orizzontale povera, restituisce generalmente una pessima qualità grafica.

Infine, il segnale raggiunge l'incidenza obliqua e radente, diventando quindi utile alla costruzione dell'immagine acustica. Una volta corrette le perdite per trasmissione, il suo livello

medio dipenderà dal tipo di fondale locale; a causa della sua incidenza radente e delle alte risoluzioni, il livello del segnale scatterato seguirà la dettagliata struttura dell'interfaccia con il fondale marino.

Un effetto d'interesse è la formazione di *ombre acustiche* sul fondo. Se è presente un ostacolo grande abbastanza da intercettare buona parte del segnale nel settore angolare trasmesso, questo impedirà lo scatteramento del fondale associato normalmente con quegli angoli.

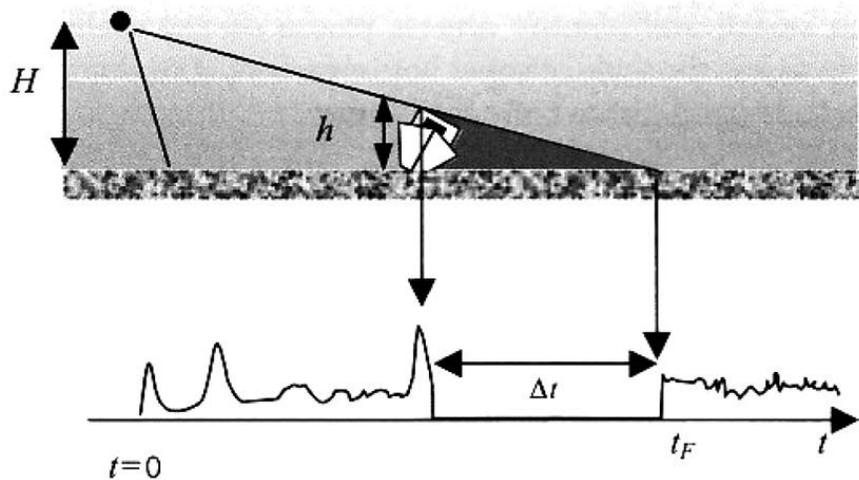


Figura 3.6. Descrizione geometrica della proiezione di un'ombra sonar [11]

L'eco ricevuta sarà quindi molto bassa, per una durata che dipende dall'angolo di elevazione e dall'altezza dell'ostacolo che agisce da maschera. Questo crea nell'immagine sonar un'ombra commisurata alla forma dell'ostacolo stesso. La sua analisi fornisce una stima della geometria dell'ostacolo e questo è molto interessante in tutte le applicazioni che ricercano ed identificano oggetti sul fondale (ad esempio mine, relitti, manufatti umani, etc.) oppure per valutare la scala di un rilievo sottomarino.

La durata dell'ombra acustica permette di valutare l'altezza dell'ostacolo sopra il fondo:

$$\frac{h}{\Delta t} = \frac{H}{t_f} \Rightarrow h = H \frac{\Delta t}{t_f}$$

dove  $t_f$  è il tempo misurato dall'istante di trasmissione del ping acustico.

## Limiti e performance dei Sidescan Sonar

### Risoluzione

La risoluzione lungo la traccia sonar è data proiettando la lunghezza equivalente del segnale usato; uno sviluppo esatto porta all'espressione accurata seguente ed alla sua approssimazione ad angoli obliqui:

$$\delta y = H \tan \theta \left[ \sqrt{1 + \frac{cT \cos \theta}{H \sin^2 \theta}} - 1 \right]$$

$$\delta y \approx \frac{cT}{2 \sin \theta}$$

Ad incidenza radente, per  $\theta \rightarrow \pi/2$ ,  $\delta y \rightarrow cT/2$  e la risoluzione diventa esattamente la risoluzione nominale del segnale. Quando  $\theta \rightarrow 0$ , la formula esatta diventa  $\delta y \rightarrow \sqrt{HcT}$  che è la risoluzione di un sonar di tipo single-beam con impulso lungo.

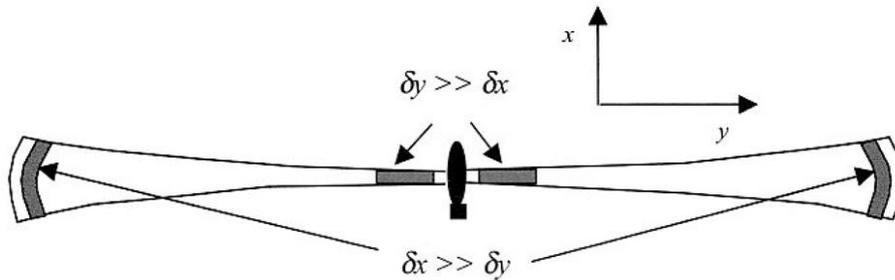


Figura 3.7. Risoluzione del Sidescan Sonar [11]

La risoluzione ortogonale alla traccia si definisce con la proiezione dell'apertura del fascio  $\delta x = R \phi$ , dove  $R$  è la distanza obliqua del sonar dal fondo e  $\phi$  è l'angolo di apertura del lobo di direttività sul piano orizzontale. Come nel caso dei sonar di tipo single-beam, la risoluzione peggiora proporzionalmente con la distanza percorsa dal segnale in propagazione.

Ne consegue che la risoluzione di un sidescan sonar è disomogenea sia lungo la traccia che ortogonalmente ad essa, variando lungo la striscia insonificata:

- A piccole distanze  $dy \gg dx$ , la risoluzione lungo traccia è la peggiore
- A grandi distanze  $dx \gg dy$ , la risoluzione lungo traccia è la migliore, ma accade che aree insonificate da pingo successivi possono sovrapporsi.

### ***L'equazione del sonar***

Tutti gli effetti relativi ad un sistema sonar (propagazione, rumore, reverberazione, antenne, segnali, processing, ...) possono essere sintetizzati nell'*equazione del sonar*. Questa formula può essere scritta in una forma estremamente semplificata, per dare un'idea della misura qualitativa di un sistema acustico subacqueo:

$$\text{Segnale} - \text{Rumore} + \text{Guadagno} > \text{Soglia}$$

L'equazione del sonar è una valutazione del budget di energia, usata per fare stime del SRN (rapporto segnale/rumore), facendo qualche ipotesi semplificativa, in rapporto alle performance del sistema.

- Il *segnale* considerato qui è il livello di segnale ricevuto dal sistema. Nei sistemi attivi non è altro che il livello del segnale trasmesso, corretto per la perdita di propagazione bidirezionale e per il coefficiente di backscattering del bersaglio.
- Il *rumore* che riguarda la ricezione è la somma dei contributi del rumore d'ambiente esterno e del rumore proprio del dispositivo sonar. È caratterizzato da una densità spettrale di potenza. Nei casi di sistemi attivi si considera anche la componente dovuta alla riverberazione.
- Il *guadagno* tiene conto sia dell'indice di direttività dell'array del trasduttore, sia del guadagno di processing del ricevitore.

- La *soglia*, considerata all'uscita del ricevitore, dipende dalle performance imposte dal sistema per la rivelazione di un bersaglio o per la bontà di una misura (dipende dall'applicazione dello strumento).

Quindi, il budget di potenza (tutti i termini sono in dB) di un sistema acustico attivo sottomarino possono essere descritti con maggior dettaglio:

$$SL - 2TL + TS - NL + DI + PG > RT$$

in cui  $SL$  è il livello del segnale trasmesso,  $TL$  è la perdita per trasmissione,  $TS$  è il target strength,  $NL$  è il livello del rumore,  $DI$  è l'indice di direttività,  $PG$  è il guadagno di processing e  $RT$  è la soglia in ricezione.

L'indice di direttività  $DI$  inserito nell'equazione è riferito all'antenna in ricezione. Talvolta quello relativo all'antenna in trasmissione è inserito nel parametro  $SL$ . Nel caso di sistemi sidescan sonar, si usa il solito trasduttore sia in Tx che in Rx e quindi i due guadagni sono uguali.

### ***Portata massima***

Consideriamo un sidescan sonar che lavora su bersagli estesi e che sia limitato dal rumore. Il livello dell'eco paragonato al livello di rumore è quindi:

$$EL - NL + DI > RT$$

Il livello dell'eco è dovuto allo scattering della superficie:

$$EL = SL - 2TL + TS = SL - 40 \log R - 2\alpha R + TS$$

$$TS = 10 \log \left( \phi R \frac{cT}{2 \sin \theta} \right) + BS_B(\theta)$$

in cui  $BS_B(\theta) = 10 \log(\sigma_{s,v})$  in cui  $\sigma_{s,v}$  è un'informazione relativa alla cross-section del bersaglio. Il livello dell'eco ricevuto sarà dunque:

$$EL = SL - 30 \log R - 2\alpha R + 10 \log\left(\phi c \frac{T}{2}\right) + BS_B(\theta)$$

Quindi l'equazione sonar diventa:

$$SL - 30 \log R - 2\alpha R + 10 \log\left(\phi c \frac{T}{2}\right) + BS_B - NL + PG + DI > RT$$

e la portata massima è data da:

$$30 \log R + 2\alpha R = SL + 10 \log\left(\phi c \frac{T}{2}\right) + BS_B - NL + PG + DI - RT$$

Si suppone che l'angolo di incidenza sia radente abbastanza da rendere trascurabile la proiezione di  $1/\sin \theta$  della traccia del segnale sul fondo.

### ***Copertura completa***

Un vincolo sulle performance di un sidescan sonar è la necessità di assicurare il 100% della copertura dell'area insonificata. La condizione che dev'essere completamente soddisfatta è la rimozione dei gap tra le aree insonificate da due ping successivi.

E' facilmente visibile che questa condizione è critica vicino alla verticale, in cui la striscia insonificata è molto sottile. Con un'apertura del lobo di direttività  $\phi$ , la larghezza della superficie interessata è  $\delta x = H\phi$ . Il sonar avanza di  $\delta x = VT_R$  tra due ping acustici, con  $V$  la velocità del natante e  $T_R$  il ritardo inter-trasmissione ottenuto dalla portata massima del sonar  $R_{max}$ , con  $T_R = 2R_{max}/c$ . Infine, si può scrivere la condizione come  $\delta x = H\phi = 2VR_{max}/c$  e, per esempio, si può imporre un limite alla velocità massima alla quale dovrà operare il natante che ospita il sidescan sonar:

$$V_{max} = \frac{cH\phi}{2R_{max}} = \frac{c\phi}{2} \cos \theta_{max}$$

in cui  $\theta_{max}$  è il massimo angolo di insonificazione. Si osserva che la condizione di copertura totale (100%) è indipendente da  $H$ . Come ordine di grandezza, un lobo di apertura  $\phi = 1^\circ$  con un angolo di tilt massimo  $\theta_{max} = 70^\circ$  limita la velocità del natante a 4.5 m/s (8.7 nodi).

### ***Utilizzo delle immagini SSS***

I sistemi SSS sono tipicamente leggeri, facilmente trasportabili da un semovente autonomo, adatti alle operazioni in acque poco profonde e capaci di coprire distanze fino a centinaia di metri. La loro dimensione ridotta, la loro semplicità di utilizzo e la qualità delle immagini prodotte li rende molto efficaci per la rivelazione di anomalie presenti sul fondale quali, ad esempio, oggetti di natura umana che possono rientrare nell'interesse dei beni archeologici. In condizioni operative particolari, come la navigazione in acque particolarmente torbide, in cui la visione dei sensori ottici sia limitata a pochi metri, l'utilizzo di un sistema acustico diventa decisivo per la mappatura del fondale.

Un primo passo verso l'identificazione di eventuali reperti può essere condotto direttamente sulle immagini grezze restituite dal dispositivo, con tecniche relativamente semplici e dal costo computazionale contenuto, ricercando *features* geometriche elementari (come segmenti e porzioni di circonferenze, ellissi o coniche in generale), la cui densità in una data immagine è senz'altro un indice della probabilità di trovarsi di fronte ad un artefatto umano immerso nell'ambiente naturale. Il vantaggio di questi metodi sta nel fatto che l'elaborazione può essere condotta online, man mano che si presenta il dato a bordo del semovente, che essendo autonomo (AUV), può compiere diverse azioni in funzione di ciò che riesce a vedere, comunicando, ad esempio ad altri dispositivi cooperanti in sciame, la possibile presenza di oggetti interessanti, fino a modificare la propria rotta per acquisire maggiori dati locali e confermare (o smentire) le decisioni precedenti.

Grazie all'*integrazione di dispositivi ausiliari* (sistema GPS, sensori inerziali per la stima del moto, sensori di controllo pressione e salinità, etc.) è possibile analizzare le immagini acquisite con un maggior grado di consapevolezza. Questo porta alla georeferenziazione del materiale memorizzato ed eventualmente ad una sua ottimizzazione grafica grazie a tecniche di correzione offline. Una volta rettificate le immagini acustiche, sulla base dei dati ausiliari acquisiti, è possibile combinarle insieme per ricreare immagini di maggiori dimensioni che restituiscono l'idea di una superficie più estesa, operazione nota come *mosaicatura*.

In condizioni operative particolari, come la navigazione in acque particolarmente torbide, in cui la visione dei sensori ottici sia limitata a pochi metri, l'utilizzo di un sistema acustico diventa decisivo per la mappatura del fondale.

Il sidescan sonar non è forse lo strumento più adatto per la ricostruzione della batimetria locale che solitamente si fa per mezzo di strumenti diversi e talvolta più sofisticati (e costosi) come i sonar multibeam, che restituiscono mappe di elevazione (*range maps*), ma che richiedono piattaforme di supporto robuste ed hanno un forte consumo energetico. Comunque, il SSS rimane (e dovrebbe rimanere ancora per qualche anno) il sistema più indicato per la produzione di immagini del fondale marino, per evidenti ragioni di risparmio economico.

### ***Estensione nell'uso di immagini SSS***

Sebbene non fornisca direttamente misure di elevazione, le immagini che produce sono relative alla topografia del fondale, come anche alla sua composizione ed alla direzione d'osservazione. Di conseguenza l'analisi visuale delle immagini prodotte suggerisce che sia possibile stabilire una corrispondenza tra l'intensità del singolo pixel acquisito e la pendenza della superficie locale in un dato punto del fondale analizzato.

Questo permette la ricerca della terza dimensione (elevazione dal fondo), o meglio una sua stima, a partire da immagini acustiche che sono per loro natura bidimensionali, con la possibilità di estendere la ricerca di *features* geometriche anche a strutture nello spazio. La

struttura 3D del fondale, così ricostruita, potrà essere analizzata attraverso metodi di fitting di modelli tridimensionali di varia complessità. Un primo fitting riguarderà, come detto in precedenza, modelli geometrici semplici e generali, quali piani e quadriche.

Negli ultimi anni, sono stati fatti vari sforzi per la determinazione indiretta della topografia utilizzando immagini derivanti dal sidescan sonar. In generale l'idea fondamentale che sta dietro ai metodi di ricostruzione è la determinazione di un modello per il processo di insonificazione del fondale, ovvero per la creazione dell'immagine sonar, che sia sufficientemente semplice da poter essere invertito, per ottenere una approssimazione dei gradienti della superficie. Tali metodi sono chiamati *shape-from-shading* (letteralmente "forma dall'ombreggiatura"). E' implicito in questa descrizione che, a partire da una sola immagine bidimensionale, si possa solo fare una stima della superficie reale che genera l'immagine sonar.

Uno studio utile allo scopo è stato fatto da E. Coiras, Y. Petillot e D.M [13]. Lane i quali hanno applicato metodi di *shape from shading* ad immagini acustiche, le quali si prestano bene a queste tecniche. Le ragioni principali sono due:

- La posizione della sorgente di segnale (acustica in questo caso) è nota e coincide con la posizione del trasduttore acustico in ricezione;
- Il modello di diffusione luminosa Lambertiano è un'approssimazione molto buona per lo scattering del segnale sonar prodotto dai materiali che si trovano tipicamente sul fondo marino.

### ***Modello di diffusione acustica***

Il processo di costruzione dell'immagine acustica è rappresentato brevemente nella figura successiva. La sorgente acustica del sensore centrata nell'origine degli assi  $O$  produce un impulso che illumina (o meglio insonifica) il fondo marino.

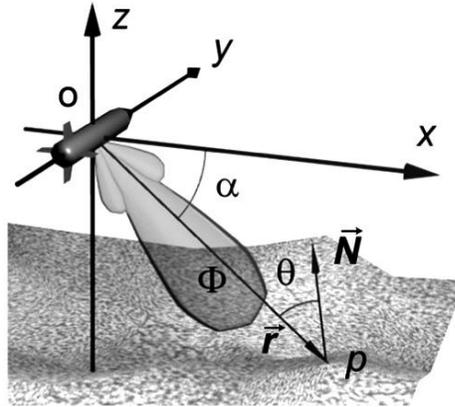


Figura 3.8. Sistema di riferimento usato nel modello di diffusione acustica

La parte di energia acustica che raggiunge il fondo nel punto  $p$ , è scatterata indietro e può essere misurata dal sensore. L'intensità del pixel corrispondente nell'immagine del sidescan sonar dipenderà dalla quantità di energia che torna indietro verso l'apparecchio dalla superficie. L'impulso non è isotropico, ma segue un particolare profilo dovuto al diagramma di direttività  $\Phi$  e dipenderà dal particolare angolo di elevazione  $\alpha$  sotteso dal vettore  $\vec{r}(p)$  da  $O$  a  $p$ . La quantità di energia scatterata dipenderà anche dalla particolare riflettività  $R$  del materiale in quel particolare punto. Statisticamente, le immagini del sidescan sonar presentano una distribuzione d'intensità di Rayleigh.

Per modellare il processo di scattering si fa riferimento al modello Lamberitano che permette di derivare l'intensità di ritorno dai parametri che definiscono la scena osservata. In questo semplice modello si assume che l'intensità di ritorno dipenda solamente dall'angolo di incidenza dell'impulso sonoro irradiato, ignorando la frequenza dell'impulso utilizzato. L'intensità di ritorno  $I$  da un punto  $p$  del fondo marino è rappresentata dalla seguente espressione:

$$I(p) = K\Phi(p)R(p)|\cos(\theta(p))|$$

in cui  $\Phi$  rappresenta l'intensità dell'impulso sonoro che agisce nel punto  $p$ ,  $R$  è la riflettività del fondale,  $\theta$  è l'angolo di incidenza del fronte d'onda e  $K$  è una costante di normalizzazione. Poiché molte immagini provenienti da sidescan sonar includono già gli effetti correttivi del guadagno tempo variante (TVG, time varying gain), per la

compensazione dell'attenuazione di intensità con la distanza, non è stato incluso nel modello nessun termine che tenga conto dell'attenuazione radiale. Per semplificare ulteriormente il modello, sono state raggruppate sotto il coefficiente  $\Phi$  tutte le variazioni introdotte dal profilo acustico del sensore.

La dipendenza dall'elevazione dal fondo marino è implicita nell'angolo di incidenza  $\theta(p)$ , che dipende dall'angolo di elevazione  $\alpha$  dalla sorgente acustica e dall'orientamento della normale della superficie  $\vec{N}(p)$ . Questa dipendenza si può esplicitare espandendo il coseno:

$$\cos(\theta(p)) = \frac{\vec{r}(p) \cdot \vec{N}(p)}{|\vec{r}(p)| \cdot |\vec{N}(p)|}$$

Si rappresentano poi  $\vec{r}(p)$  e  $\vec{N}(p)$  nel sistema di coordinate riferito al sensore e si esprime il punto  $p$  come  $(x, y, Z(x, y))$ , ottenendo:

$$\vec{r}(p) = (x, 0, Z(x, y))$$

$$\vec{N}(p) = \left( -\frac{\partial Z}{\partial x}(x, y), -\frac{\partial Z}{\partial y}(x, y), 1 \right)$$

in cui la coordinata  $y$  in  $\vec{r}(p)$  è 0 poiché l'impulso del sidescan sonar è sagomato in modo da irradiare solamente nei punti contenuti nel piano  $xz$ .

Combinando queste espressioni si ottiene il modello diretto per il calcolo dell'intensità  $I$  in un qualunque punto  $p$ , dati i parametri del modello  $R$ ,  $Z$  e  $\Phi$ :

$$I(x, y) = -K\Phi(x, y)R(x, y) \cdot \frac{Z(x, y) - x \cdot \frac{\partial Z}{\partial x}(x, y)}{\sqrt{x^2 + Z^2(x, y)} \cdot \sqrt{\left(\frac{\partial Z}{\partial x}(x, y)\right)^2 + \left(\frac{\partial Z}{\partial y}(x, y)\right)^2 + 1}}$$

in cui i gradienti  $\partial Z/\partial x$  e  $\partial Z/\partial y$  possono essere approssimati con le differenze finite, ottenendo un'espressione che dipende direttamente da  $Z$ .

A completamento del modello diretto rimane il calcolo della costante  $K$  che tiene conto che buona parte dei valori di intensità dell'immagine sono influenzati da un qualche tipo di scalatura dovuta ad esempio al TVG, all'attenuazione radiale, etc. L'approccio adottato dal gruppo di lavoro di E.Coiras è di calcolare il massimo ritorno possibile per una dato punto  $p$  ed utilizzarlo per normalizzare il valore del pixel calcolato. L'espressione che ne risulta è:

$$I_{max} = \frac{-Z + \frac{x^2}{-Z} \cdot \left(1 + \left(\frac{\partial Z}{\partial y}\right)^2\right)}{\sqrt{x^2 + z^2} \cdot \sqrt{1 + \left(\frac{\partial Z}{\partial y}\right)^2 + \frac{x^2}{Z} \cdot \left(1 + \left(\frac{\partial Z}{\partial y}\right)^2\right)^2}}$$

in cui è stata rimossa la dipendenza esplicita di  $Z$  da  $(x, y)$  per chiarezza. Si nota che  $I_{max}$  è sempre positiva poiché il sensore è sempre al di sopra del fondale, il che si traduce nel fatto che  $Z(x, y)$  è sempre negativa in accordo col sistema di riferimento scelto. Si nota inoltre che  $I_{max}$  è sempre compresa tra 0 e 1.

Il coefficiente di normalizzazione sarà dunque:

$$K(x, y) = \frac{1}{I_{max}(x, y)}$$

### ***Inversione Sonar***

Il problema inverso, ovvero il calcolo dei parametri del modello a partire dalle intensità osservate, è indeterminato, utilizzando una sola osservazione di  $I$  per ogni punto da cui calcolare i parametri. Per risolvere questo problema, il gruppo di E.Coiras propone di utilizzare il metodo dei minimi quadrati (LMS) tentando di minimizzare la somma  $E$  del quadrato delle differenze tra i punti dell'immagine osservata  $I(x, y)$  e quelli del modello  $\hat{I}(x, y)$ :

$$E = \sum_{x,y} E(x, y) = \sum_{x,y} (I(x, y) - \hat{I}(x, y))^2$$

Perciò, dovrà essere affrontato il seguente problema di ottimizzazione:

$$(Z, R, \Phi) = \operatorname{argmin}(E)$$

Esistono molti modi per approssciare un problema simile, ed il gruppo suddetto ha scelto di usare l'algoritmo del gradiente discendente, così da convergere su un minimo locale. Per assicurare la convergenza su un minimo "realistico" deve essere fatta particolare attenzione al punto di partenza con cui inizializzare l'algoritmo, introducendo delle conoscenze a priori sulla scena. L'algoritmo proposto è iterativo e può essere sintetizzato con il seguente diagramma di flusso:

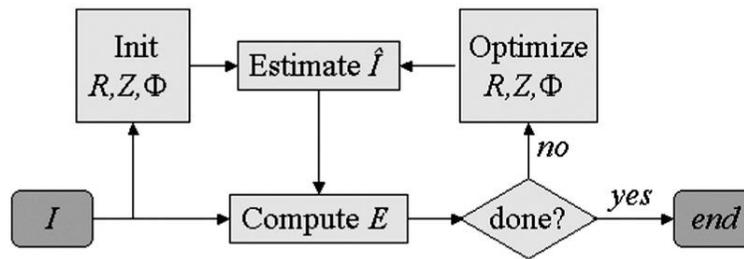


Figura 3.9. Diagramma di flusso della procedura di ottimizzazione

Nel primo stadio, le stime per i parametri del modello sono utilizzate per calcolare una stima della mappa acustica delle intensità  $\hat{I}(x, y)$ . Per far questo si sostituiscono nel modello di diffusione acustica i parametri  $R(x, y)$ ,  $\Phi(x, y)$  e  $Z(x, y)$  stimati nell'iterazione precedente.

Nello stadio successivo (ottimizzazione) si usa direttamente l'approccio del gradiente discendente per minimizzare  $E$  aggiornando i parametri del modello come segue:

$$R(x, y) \leftarrow R(x, y) - \lambda \cdot \frac{\partial E}{\partial R}(x, y)$$

$$\Phi(x, y) \leftarrow \Phi(x, y) - \lambda \cdot \frac{\partial E}{\partial \Phi}(x, y)$$

$$Z(x, y) \leftarrow Z(x, y) - \lambda \cdot \frac{\partial E}{\partial Z}(x, y)$$

dove  $\lambda$  è una piccola costante usata per controllare la velocità di aggiornamento dell'algoritmo. Le espressioni sono reiterate finchè l'errore  $E$  scende sotto una soglia voluta. Un valore suggerito come compromesso tra errore finale e velocità di calcolo è  $\lambda = 0.25$ .

Esplicitando i calcoli si ottengono le seguenti espressioni per l'aggiornamento dei parametri:

$$R(x, y) \leftarrow R(x, y) + 2\lambda \cdot \frac{\hat{I}(x, y)}{R(x, y)} (I(x, y) - \hat{I}(x, y))$$

$$\Phi(x, y) \leftarrow \Phi(x, y) + 2\lambda \cdot \frac{\hat{I}(x, y)}{\Phi(x, y)} (I(x, y) - \hat{I}(x, y))$$

$$Z \leftarrow Z - 2\lambda \hat{I}(I - \hat{I}) \cdot \left( \frac{-\frac{\partial Z}{\partial y} - \frac{\partial Z}{\partial x}}{1 + \left(\frac{\partial Z}{\partial y}\right)^2 + \left(\frac{\partial Z}{\partial x}\right)^2} + \frac{1 + x}{x \left(\frac{\partial Z}{\partial y}\right) - Z} + \frac{Z}{x^2 + Z^2} \right)$$

in cui, nell'ultima equazione, è stata rimossa la dipendenza esplicita dei parametri da  $(x, y)$  per semplificare la leggibilità.

Poiché questo metodo si basa sull'analisi di ogni singolo pixel è utile uno schema di regolarizzazione per mantenere coerenza tra i punti vicini delle superfici, che si compie alla fine di ogni ciclo iterativo filtrando la riflettività stimata e le mappe del pattern di direttività. Infatti è data come ipotesi che i valori di  $\Phi$  per tutti i punti sottesi dallo stesso angolo  $\alpha$  dal sensore siano gli stessi perché la struttura del sensore rimane costante, e sono fissati al loro valore mediano.

Come già accennato, è necessario studiare le condizioni che inizializzano l'algoritmo, ovvero i valori di partenza di  $R(x, y)$ ,  $\Phi(x, y)$  e  $Z(x, y)$ . La mappa di riflettività  $R$  è posta ad un valore costante, tipicamente 0.9 (i valori possibili variano tra 0 e 1, ovvero assenza di scattering e riflessione totale). Il valore di  $\Phi$  è calcolato partendo dall'immagine originale e regolarizzato come descritto in precedenza. Infine, la mappa delle elevazioni  $Z$  è inizializzata in modo che ogni punto  $(x, y)$  sia fissato alla quota stimata del primo ritorno  $(0, y)$ , grazie a sensori ausiliari che restituiscono elevazione rispetto al fondo e quota operativa (a mezzo del controllo della pressione).

## ***Limiti della ricostruzione***

Il primo limite sta nelle ipotesi fatte per l'assunzione del modello e per la sua implementazione. Comunque il secondo e più importante deriva dalle performance del sensore stesso.

Poiché il sidescan sonar integra tutti i ritorni ricevuti dalla superficie ad un tempo fissato, diverse configurazioni della superficie possono produrre esattamente la stessa immagine sonar. Un esempio è mostrato nella figura successiva, in cui si presentano due casi equivalenti.

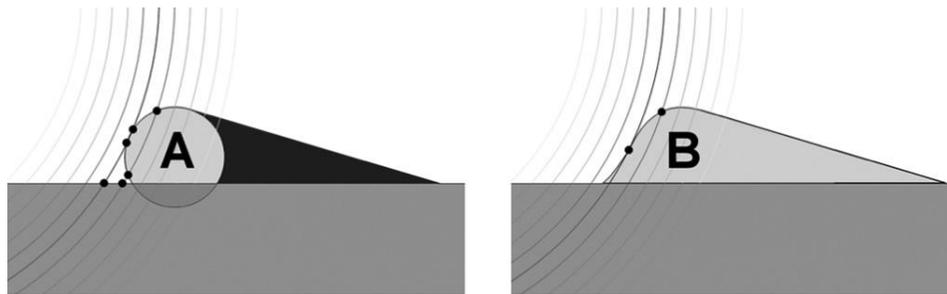


Figura 3.10. Due configurazioni del fondale marino che generano la stessa immagine sonar

Nella caso A è graficata una conduttura adagiata sul fondale piatto; ogni fronte d'onda acustica interseca sia la conduttura che il fondo, in uno o più punti, e questi contributi vengono integrati nel valore di ritorno misurato dal sensore. Inoltre tutti i punti della superficie dietro l'ombra acustica causata dalla conduttura (colorata in nero) sono effettivamente nascosti al sensore e non possono essere ricostruiti. Il caso B rappresenta una superficie ricostruita in modo che sia intercettata dal fronte d'onda acustico solamente un punto alla volta e restituisca la stessa energia del caso A. I "bersagli" A e B generano la stessa immagine sidescan sonar. La seconda configurazione è quella ricostruibile dal metodo proposto poiché si assume implicitamente che  $Z(r)$  sia una funzione a valori singoli per ogni raggio  $r$ .

## Capitolo 4 - Conclusioni

Se pur facenti parte di due mondi diversi, i metodi per l'indagine ottica e quelli per l'indagine acustica hanno lo stesso scopo di mappatura dei fondali e di ricerca di siti d'interesse, distinguendo tra le caratteristiche naturali dell'habitat marino e quelle artificiali dovute alla presenza umana. L'identificazione di tali luoghi, a mezzo di veicoli autonomi sottomarini (AUV, autonomous underwater vehicle), può essere condotta in maniera sistematica su larga scala in quanto le procedure di ricerca possono beneficiare di natanti dotati di un'autonomia sufficiente a coprire vaste aree di territorio. A tale scopo, l'intervento umano diretto, tramite operatori specializzati per missioni subacquee in acque anche relativamente profonde, si riduce al minimo, solamente nel caso in cui vengano rilevate condizioni particolari, reputate interessanti, una volta analizzati i dati disponibili.

I sensori per l'acquisizione di tali dati possono nascere come dispositivi specifici oppure essere *general purpose*, termine con cui si intendono dispositivi che non siano dedicati ad un solo possibile utilizzo, ma dispositivi versatili che di solito prevedono componenti software che sono invece soluzioni specifiche a una particolare esigenza. La seconda soluzione fornisce un più ampio ventaglio di possibilità operative dettate sia dalle necessità previste, sia quelle che si presentano *in itinere*, fornendo scalabilità al sistema, nella sua interezza. Componente non

trascurabile è il costo relativamente contenuto di questi apparati, come semplici telecamere stagne e sidescan sonar, che sono già da tempo utilizzati per varie applicazioni e su diverse piattaforme. Le dimensioni ridotte permettono la loro integrazione sullo stesso veicolo e possono essere alloggiati in spazi dedicati al payload scientifico.

### ***Integrazione ottico/acustica***

L'aspetto che desta interesse è l'utilizzo congiunto di dati per loro natura diversi, provenienti da sensori che operano in domini diversi (ottico, acustico, magnetico, etc.) che possono essere integrati fra loro in vari modi. Grazie all'utilizzo di grandezze ausiliare registrate contestualmente all'acquisizione dai sensori di indagine, è possibile georeferenziare i dati, organizzandone una grande mole in maniera da ottimizzarne la consultazione. Questo è un passo essenziale all'utilizzo congiunto e si estende nell'operazione di *registration* (tradotta talvolta col termine "registratura") ovvero nel processo che permette la trasformazione di differenti insiemi di dati (ognuno dei quali adotta un proprio sistema di riferimento) in un sistema dove ogni coordinata spaziale corrisponde, rendendo possibile ogni possibile correlazione o integrazione. Grazie alla referenziazione congiunta dei dati è possibile valutare anche l'evoluzione temporale (oltre che spaziale) qualora vengano fatte diverse acquisizioni, per stimare come evolvano le condizioni che afferiscono ad una stessa area geografica.

L'integrazione vera e propria delle informazioni provenienti dai diversi sensori può essere suddivisa in categorie per meglio comprenderne gli obiettivi:

- Integrazione "a cascata": ci poniamo nella situazione in cui le condizioni di acquisizione non siano ottimali per ogni tipo di sensore alloggiato sul natante; le informazioni registrate non avranno tutte la stessa qualità o addirittura alcune saranno infruibili. In questo caso l'itegrazione è "operativa", ovvero i dati buoni che possano rivelare la presenza di siti interessanti serviranno per pianificare una missione successiva, per colmare le lacune della precedente. Ad esempio, in un ipotetico scenario di missione,

attraverso un primo passaggio in acque torbide si rivelano punti interessanti grazie al sensore acustico (che è in qualche modo immune alla torbidità delle acque) e successivamente, in tempi diversi, si torna sul luogo per confermare (o smentire) la presenza di qualche anomalia d'interesse. Oppure si immagina che il natante stia navigando a distanze troppo elevate dal fondo marino e che gli illuminatori di bordo (led o laser) non riescano a penetrare fino al fondale; anche in questo caso le informazioni derivanti dal sidescan sonar sono essenziali alla missione. Le immagini acustiche serviranno per classificare la scena osservata in base alla *features* estratte. Tale classificazione verrà sfruttata per una eventuale ri-pianificazione distribuita della missione di esplorazione.

- Fusione di basso livello: nell'ipotesi che esista una quantità di dati fruibili, e dopo aver eseguito la procedura di *registration*, è possibile gestire la diversa natura delle informazioni stratificandole in *layer*. Questa operazione permette di correlare grandezze, per loro natura diverse, in modo da facilitarne la consultazione sia da parte di un operatore, sia in maniera automatizzata. Nella stessa scena sarà possibile sovrapporre informazioni di tipo magnetico per valutare la presenza di anomalie rispetto al fondo naturale, di tipo acustico (ad esempio nelle stime di riflettività del fondo, indice del tipo di materiale che riflette l'onda acustica) e di tipo ottico da cui è possibile estrarre molte delle *features* che caratterizzano l'osservato (peculiarità geometriche, texture che identificano un particolare tipo di fondale, etc.).
- Fusione di alto livello: l'esplorazione su larga scala dell'ambiente marino restituisce una mole di informazioni difficilmente consultabili da uno o più operatori, nella loro interezza; il tempo necessario alla loro analisi sarebbe troppo elevato. E' utile quindi la ricerca automatizzata dei punti di interesse, qualsiasi sia la loro natura, per mezzo di algoritmi discriminanti che suggeriscano anche una "misura" di quanto buona sia la loro rivelazione.

Per le telecamere e le immagini acustiche si possono applicare metodi per l'identificazione di *features* geometriche di tipo bidimensionale, quali rilevanti porzioni di rette e di coniche (curve elementari), o di tipo tridimensionale rilevate mediante i

metodi di ricostruzione 3D descritti. Dai sensori magnetici si possono misurare le variazioni di campo e associarne le intensità alla presenza di strutture sommerse nel fondo (sabbioso, fangoso, o sotto della vegetazione) e quindi invisibili agli altri sensori.

Anche se ognuno di questi algoritmi di ricerca deve agire necessariamente nel proprio dominio di appartenenza, questo non significa che i risultati che essi producono debbano rimanere disgiunti. Infatti può essere fatta un'integrazione dei risultati e delle stime provenienti dal singolo sensore per affinare delle "mappe di probabilità" create per dare un senso di misura (in termini di probabilità di buona rilevazione, appunto) alla presenza degli oggetti ricercati. La densità con cui vengano rilevate *features* di rilievo in una data immagine o in una mappa acustica o magnetica è senz'altro un indice della probabilità di trovarsi di fronte ad un artefatto. La ricerca di corrispondenze nei diversi domini può sia rafforzare la rivelazione che smentirla, affinando di fatto la decisione automatica.

# Bibliografia

- [1] Hartley - Zisserman, *Multiple View Geometry in Computer Vision*, Cambridge University Press, 2004
- [2] Jean-Yves Bouguet, *Visual methods for three-dimensional modeling*, California Institute of Technology, Pasadena, California, 1999
- [3] Luigi Di Stefano, *Principi ed Applicazioni della Visione Stereo*, University of Bologna, Computer Vision Laboratory (CVLAB), <http://www.vision.deis.unibo.it/>
- [4] Stefano Mattocchia, *Stereo Vision: Algorithms and Applications*, Università di Bologna, Dipartimento di Elettronica Informatica e Sistemistica (DEIS), <http://www.vision.deis.unibo.it/smatt/>
- [5] Stefano Mattocchia, *Introduzione alla Visione Stereo*, Università di Bologna, Dipartimento di Elettronica Informatica e Sistemistica (DEIS), <http://www.vision.deis.unibo.it/smatt/>
- [6] Yoav Y. Schechner and Nir Karpel, *Clear Underwater Vision*, Dept. of Electrical Engineering Technion - Israel Inst. Technology, Haifa 32000, Israel, 2004
- [7] Tritech International Limited, *Tritech SeaStripe - ROV / AUV Laser Line Generator*, <http://www.tritech.co.uk/products/products-seastripe.htm>
- [8] Jean-Yves Bouguet, *Camera Calibration Toolbox for Matlab*, [http://www.vision.caltech.edu/bouguetj/calib\\_doc/](http://www.vision.caltech.edu/bouguetj/calib_doc/)
- [9] D.K.Naidu, R.B.Fisher, *A Comparative Analysis of Algorithms for Determining the Peak Position of a Stripe to Sub-pixel Accuracy*, Department of Artificial Intelligence, University of Edinburgh, Edinburgh, 1991
- [10] Visual Computing Lab (ISTI - CNR), *MeshLab*, Open Source Software, <http://meshlab.sourceforge.net/>
- [11] Xavier Lurton, *An Introduction to Underwater Acoustics, Principles and Applications*, Springer-Praxis, 2002
- [12] Tritech International Limited, *SeaKing ROV/AUV Sidescan Sonar System*, [http://www.tritech.co.uk/products/products-sidescan\\_sonar.htm](http://www.tritech.co.uk/products/products-sidescan_sonar.htm)
- [13] "Multiresolution 3-D Reconstruction From Side-Scan Sonar Images", Enrique Coiras, Yvan Petillot, David M. Lane, IEEE Transactions On Image Processing, Vol. 16, No. 2, Feb 2007