



Regione Toscana



FAS
Fondo Aree
Sottoutilizzate
2007-2013



REPUBBLICA ITALIANA



TecnicHe per l'Esplorazione Sottomarina Archeologica mediante l'Utilizzo di Robot aUtonomi in Sciami

Report D3.3: Metodi per il riconoscimento e l'annotazione automatica di immagini e mappe

Progetto co-finanziato dalla Regione Toscana	
Livello di disseminazione	
CO	Confidenziale, ristretto ai membri dell'ATS (inclusa la Regione Toscana)

Data di consegna	30/11/2012	Data di consegna effettiva	28/02/2013
Data di inizio progetto	01/03/2011	Durata	24 mesi
Responsabile del deliverable	CNR-ISTI	Revisione	1.0
Partner coinvolti	CNR-ISTI, CVG-UNIFI		

Identificativo: Report D3.3

Nome del file: D3point_3_01.doc

Titolo: Studio e sviluppo per l'analisi e l'annotazione delle immagini e delle mappe

Indice

1	Introduzione	7
2	Condizioni sperimentali e pre-processing	8
2.1	Fonti di disturbo nei dati acquisiti	11
2.1.1	Artefatti nella visione stereoscopica subacquea	11
2.1.2	Artefatti nelle immagini side-scan sonar	11
2.2	Elaborazioni preliminari	12
2.2.1	Pre-filtering ottico	12
3	Registrazione e mosaicatura di immagini e mappe	15
3.1	Registrazione	15
3.2	Mosaicatura	16
3.2.1	Metodi per la mosaicatura di scansioni acustiche	16
3.2.1.1	Correzione della distorsione dovuta all'imbarcata	19
3.2.1.2	Correzione della distorsione dovuta al moto di beccheggio	20
3.2.2	Metodi per la mosaicatura di immagini ottiche	22
4	Algoritmi per l'interpretazione automatica di scenari subacquei	25
4.1	Interpretazione automatica mediante tecniche multi-view	25
4.1.1	Tecniche di riconoscimento 3D	27
4.2	Ricostruzione 3D da immagini sonografiche	27
5	Tecniche di screening e rilevamento automatico di oggetti da immagini	30
5.1	Riconoscimento di strutture geometriche bidimensionali	30
5.1.1	Tecniche di riconoscimento basate sulla trasformata di Hough	30
5.1.2	Tecniche di riconoscimento basate su metodi <i>parameterless</i>	31
5.2	Algoritmi di intelligenza artificiale per la classificazione di tessiture	36
5.2.1	Metodo dei filtri di Gabor	36
6	Annotazione semantica della scena e archiviazione	39
6.1	Contenuti raccolti e generati	39
6.1.1	Sensor footage	39
6.1.2	Mappe di insieme e loro metadati	39
6.1.3	Modelli 3D di Oggetti	39
6.2	Annotazione semantica dei contenuti	39
6.3	Repository per l'archiviazione dei contenuti	39
7	Integrazione di dati da più sorgenti	42
7.1	Tecniche per l'estrazione di features da mappe multi-sensoriali	42
8	Bibliografia	44

Indice delle Tabelle

Tabella 2.1: Metadati associati ad un sonogramma	8
---	----------

Indice delle Figure

Figura 1 - Mappa sidescan acquisita durante esperimento all'Isola d'Elba	9
Figura 2: Piano generale dell'AUV con evidenziati gli alloggi dei LED per l'illuminazione (a prua) e delle telecamere per l'acquisizione video (a poppa).....	10
Figura 3 - Mappatura del sonar.....	12
Figura 4 - Immagine originale a sinistra, prefiltrata a destra [13].Immagine reperibile in rete.....	13
Figura 5 - Immagini originali a sinistra, prefiltrate a destra [14], acquisite a Marciana Marina, Elba, il 18 Aprile 2012	13
Figura 6 - Immagine originale a sinistra – pre-filtrata a destra [13].....	14
Figura 7 - Visualizzazione Waterfall	16
Figura 8 - Esempi di perturbazioni rotazionali nel moto del sensore	16
Figura 9 - Esempio di acquisizioni da side scan sonar corrotte da molteplici artefatti	17
Figura 10 - Coordinate Slant e Ground	17
Figura 11 - Campionamento del side scan sonar	18
Figura 12 - Trasformazione Slant to Ground	18
Figura 13 - Mappa Sonar in coordinate Ground Range.....	18
Figura 14 - Movimento di imbardata	19
Figura 15 - Immagine originale e simulazione di un'acquisizione corrotta da moto di imbardata	20
Figura 16 - Ripristino della posizione delle linee e risultato finale dopo interpolazione	20
Figura 17 - Moto di beccheggio	20
Figura 18 – Alterazione della zona di scansione a causa del moto di beccheggio	21
Figura 19 - Apertura del fascio acustico	21
Figura 20 - Simulazione della distorsione nella mappa sonografica dovuta al moto di beccheggio	22
Figura 21 - Simulazione della correzione della distorsione dovuta a beccheggio.....	22
Figura 22 - Registrazione e Mosaicatura	23
Figura 23 - Esempio di SIFT applicato ai dati raccolti a Marciana Marina.	24
Figura 24 Schema di ricostruzione da coppia di flussi video: (a) Coppia di immagini stereo, estratte dai flussi video sincronizzati e rettificati; (b) Recupero delle coordinate tridimensionali di oggetti nella scena mediante elaborazione di depth map; (c) Ricostruzione della scena associata alla <i>depth map</i> con applicazione della texture originale.	25
Figura 25 Ricostruzione 3D da lama laser: test di laboratorio.	26
Figura 26 - Formazione dell'immagine acustica.....	28
Figura 27 - Operazione di inversione: dall'eco sonar alla superficie stimata	28
Figura 28 - Applicazione dell'algoritmo Shape From Shading Acustico su immagini simulate	29
Figura 29 - Algoritmo Hough per il riconoscimento di segmenti.....	30
Figura 30 - Algoritmo Hough per il riconoscimento di cerchi.....	31
Figura 31 - Algoritmo Hough (randomized) per il riconoscimento di ellissi	31
Figura 32 - Schema logico dell'algoritmo ELSD [ELSD]	32
Figura 33 - Riconoscimento di curve: test preliminari in vasca a Pistoia.....	33
Figura 34 - Riconoscimento di curve: fondale di Marciana Marina, piatti ed anfore	33
Figura 35 - Riconoscimento di curve: fondale di Marciana Marina, piatti ed anfore	33
Figura 36 - Schema di elaborazione del flusso ottico ai fini dell'individuazione di variazioni rilevanti nella quantità di curve regolari presenti nella scena.....	34
Figura 37 - Frame estratti dalla sequenza di Marciana Marina, con evidenziati alcuni in cui sono evidenti detection interessanti	34
Figura 38 - Grafici relativi all'algoritmo di detection applicato alla sequenza di Marciana Marina	35
Figura 39 - Riconoscimento di curve (ellissi, circonferenze, segmenti) relativo ad una sequenza di un sito archeologico molto ricco di manufatti.	35
Figura 40 - Grafici relativi all'algoritmo di detection applicato alla sequenza del sito archeologico	35
Figura 41 - Riconoscimento di curve applicato ad un sonogramma raffigurante uno pneumatico	36
Figura 42 - Filtri di Gabor con 3 orientazioni distinte	37
Figura 43 - Plot 3D di un filtro di Gabor.....	37
Figura 44 - Segmentazione applicata ad un sonogramma: evidenziata l'area dominata dai ripple.	38
Figura 45 - Architettura del repository.....	40

Identificativo: Report D3.3

Nome del file: D3point_3_01.doc

Titolo: Studio e sviluppo per l'analisi e l'annotazione delle immagini e delle mappe

Figura 46 - Schema di rappresentazione dei dati nel modello di fusione.....	42
Figura 47 - Detection rafforzata dalla concordanza di ottico e acustico	43

1 Introduzione

Il presente documento descrive i risultati dell'attività di progettazione e sviluppo software, condotta nell'ambito dell'analisi di dati multi-sensoriali relativi alle campagne di acquisizione previste nel progetto Thesaurus.

L'ambiente marino e, nel caso specifico di Thesaurus, i fondali marini definiscono uno scenario sperimentale disagiata e difficilmente controllabile. I fenomeni fisici che hanno luogo a elevate profondità marine rappresentano una dura sfida sia per gli scienziati interessati a effettuare rilevazioni e misure significative, che per gli archeologi, abituati a svolgere operazioni di scavo nel più familiare contesto della terraferma. Una missione di esplorazione in zone subacquee richiede tipicamente attrezzature costose e pianificazioni scrupolosamente dettagliate.

Con riferimento all'attuale letteratura scientifica, si possono elencare numerosi articoli riguardanti tecniche per l'esplorazione di fondali marini ([1], [2]) e la tematica dell'underwater object detection. L'approccio tipico adottato nell'ambito di queste attività di ricerca consiste nel prendere in esame una questione specifica, relativa all'esplorazione subacquea, e fornire una soluzione unica, concepita su misura per le singolari condizioni del problema: a titolo di esempio, la scelta dell'attrezzatura deve essere cautamente condotta sulla base dei requisiti imposti dal contesto sperimentale e dal paesaggio marino in cui avrà luogo la missione (acque basse o profonde, condizioni di visibilità più o meno proibitive,...). Questo carattere di molteplicità delle possibili scelte, insieme alla difficoltà di validazione dei propri risultati e di estensione e unificazione del metodo, hanno provocato il proliferare di una vasta famiglia di tecnologie ([3] [4] [5] [6] [7]) nessuna delle quali è in grado di fronteggiare pienamente una situazione sperimentale generica.

Nel contesto scientifico dell'object detection, l'idea di combinare informazioni provenienti da più dispositivi di diversa tipologia non è una novità ([8], [9]). Essa prende ispirazione dalla data fusion ([10]) che rappresenta una teoria consolidata in grado di supportare una naturale varietà di possibili applicazioni. Allo stato attuale l'analisi di immagini in ambiente subacqueo costituisce un importante ambito di ricerca, ricco di questioni aperte e problemi irrisolti.

Per le finalità del progetto Thesaurus i dispositivi, che sono stati selezionati come sensori di payload da installare a bordo del veicolo autonomo subacqueo (AUV), sono i) un sensore acustico (un Sonar a scansione laterale) e ii) un sistema di sensori ottici (un sistema di visione stereoscopico costituito da una coppia di telecamere analogiche subacquee).

Dispositivi acustici e ottici hanno un funzionamento basato su leggi fisiche diverse, restituiscono misure di tipo diverso e in generale non condividono le stesse condizioni di funzionamento ottimale. I dati grezzi prodotti dal sistema di visione sono identificati da coppie di flussi video, mentre il sonar a scansione laterale restituisce echi acustici di ritorno, cronologicamente strutturati in forma di mappa bidimensionale.

L'interpretazione dettagliata di uno scenario subacqueo, da realizzarsi mediante l'analisi dei dati raccolti dai sensori a bordo dell'AUV, ha costituito l'obiettivo primario del lavoro svolto. Tra gli scopi del progetto assume un ruolo rilevante la progettazione e lo sviluppo di un metodo per far confluire l'insieme di immagini ottiche e mappe acustiche in un'unica struttura riassuntiva, in modo da poter estrarre direttamente da quest'ultima informazioni descrittive dell'ambiente esplorato. Mediante l'elaborazione di dati ottici e acustici, e la loro successiva integrazione si offre l'opportunità di eseguire operazioni di classificazione, object detection e recognition direttamente sulla mappa di fusione, sfruttando, in ogni punto, l'eterogeneità complementare delle informazioni ottenute dai sensori. Questa complementarità irrobustisce l'interpretazione della scena, rendendola più affidabile rispetto al caso in cui siano disponibili dati prodotti da un solo tipo di sensore.

Il documento si sviluppa nella maniera seguente: nella sezione 2 si descrivono le condizioni sperimentali specifiche del progetto; nella sezione 3 si definiscono le elaborazioni preliminari con cui trattare i dati grezzi, con particolare attenzione alle differenze concettuali nel caso dei dati ottici o dei dati acustici; nelle successive sezioni 4, 5 e 6 si descrivono gli algoritmi implementati per il riconoscimento e la classificazione automatica di oggetti; nella sezione 7 si introduce il modello di fusione adottato. Il documento contiene inoltre alcuni risultati ottenuti dall'applicazione delle procedure implementate sia su immagini di simulazione sia su immagini ottenute durante alcuni test realizzati nell'ambito del progetto (dati raccolti in vasca a Pistoia e durante un esperimento condotto all'Isola d'Elba).

2 Condizioni sperimentali e pre-processing

I veicoli realizzati nel corso del progetto sono stati ideati per operare in maniera cooperativa all'interno di un gruppo, o **sciame**. In particolare, ciascun veicolo sarà equipaggiato con un solo tipo di sensore e sarà quindi specializzato nella raccolta di un particolare tipo di dato. Questa strategia introduce una prima difficoltà nell'analisi dei dati, in quanto le misure effettuate da ogni veicolo saranno caratterizzate da coordinate specifiche rapportate ad un proprio sistema di riferimento. Una corretta procedura di integrazione richiede una correzione preliminare delle misure, in modo da collocare tutti i dati all'interno di un unico sistema di coordinate.

Pertanto ogni veicolo sarà dotato di una piattaforma per rilevare la posizione, in superficie, da satellite (**GPS**); sott'acqua, dove il **GPS** non è utilizzabile, la posizione verrà stimata mediante il sistema di navigazione di bordo (piattaforma inerziale) accoppiato con un sistema di localizzazione acustica per migliorare la stima locale. Sarà nota, infine, la distanza dal fondale, misurata mediante un ecoscandaglio.

Nel seguito faremo l'ipotesi che la raccolta dei dati venga realizzata in maniera ottimale, con ogni veicolo in moto a velocità costante lungo il percorso previsto, e ad una distanza fissata rispetto al fondale.

Come è stato precedentemente ricordato, la sensoristica installabile a bordo degli AUV di Thesaurus consisterà di un dispositivo acustico (per la precisione un side-scan sonar **Tritech Seaking**) o una coppia di telecamere subacquee (**Bowtech 650D**) in configurazione stereoscopica.

Side-Scan Sonar

Il dispositivo acustico produce un'immagine bidimensionale (*sonogramma*) in cui ogni pixel rappresenta, in scala di grigi, l'intensità dell'eco di ritorno (*back-scatter echo*) rivelata dal sensore.

Il software di controllo del side-scan sonar esegue una richiesta esplicita per effettuare l'acquisizione di ogni linea. La cadenza con cui avviene questa richiesta dipende dalla scelta dei parametri di navigazione, in particolare da come viene impostata la velocità dell'AUV: maggiore è la velocità del veicolo durante la perlustrazione, maggiore dev'essere la frequenza con cui si eseguono le istanze di acquisizione.

Il software salva l'immagine in formato *jpeg*, associando al file alcuni metadati che includono i dettagli dell'acquisizione e i dati di georeferenziazione (i più importanti consistono in *latitudine*, *longitudine*, *profondità*, *angolo di rotta* e *angoli di assetto*). Attualmente i metadati vengono registrati nel commento all'immagine come coppie (*parametro*, *valore*). Una descrizione dei parametri maggiormente significativi che compaiono nei metadati è riportata in Tabella 2.1.

Tabella 2.1: Metadati associati ad un sonogramma

Parametro	Descrizione
HDCtrl	Parametro descrittivo delle varie modalità operative (scanning con motore o fisso; risoluzione a 4 o 8 bit).
RangeScale	Range massimo in <i>m</i> impostato tramite software.
TxN	Costante di trasmissione calcolata tramite la formula $TxN = (F * 2^{32}) / 32^6$. <i>F</i> è la frequenza di lavoro del sonar espressa in <i>Hz</i> .
Gain	Guadagno del ricevitore.
SlopeSettings	Pendenza iniziale della curva di guadagno.
ADSpan	Definisce lo span possibile per le ampiezze degli echi di ritorno.
ADLow	Definisce la posizione base del range dinamico degli echi di ritorno; permette di controllare la sensibilità dello strumento.
Heading Offset	Per sonar che integrano funzioni di misura dell'assetto è possibile sfruttare questo parametro per correggere dinamicamente i dati.
AD Interval	Definisce la risoluzione di campionamento per ogni bin. Si esprime in multipli di 640 <i>ns</i> e il limite inferiore possibile è 5. La massima risoluzione possibile è quindi 3.2 <i>ms</i> .
Tran. Bearing	Orientamento del trasduttore espresso in sedicesimi di <i>gradian</i> (1 <i>gradian</i> = $\pi/200$).

DBytes	Mole di dati in <i>Byte</i> generata dal sonar per ogni striscia acquisita. Dipende dal numero NBins (definito sotto) e dalla quantizzazione del segnale (4 o 8 bit).
Frequency	Frequenza del trasmettitore espressa in <i>Hz</i> , impostata da software.
Frequency (P)	Frequenza in <i>Hz</i> rilevata mediante analisi dei pacchetti inviati.
Range	Range in <i>m</i> , impostato tramite software.
RangeScale (P)	Range in <i>m</i> rilevato mediante analisi dei pacchetti inviati.
NBINS	Numero di range bins che il sonar genera per ogni ping. Il massimo valore possibile è 800.
Pulse Length (P)	Lunghezza dell'impulso trasmesso (in <i>ms</i>). Rilevato mediante analisi diretta dei pacchetti.
AD Interval (P)	Risoluzione temporale del bin, rilevata mediante analisi diretta dei pacchetti.
Scan Time	Tempo totale (in <i>s</i>) di campionamento del ping.
True Range	Range massimo in <i>m</i> rilevato mediante analisi diretta dei pacchetti.
True Resolution	Risoluzione in <i>s</i> rilevata mediante analisi diretta dei pacchetti.
Latitude	Coordinata di Latitudine espressa in <i>gradi</i> .
Longitude	Coordinata di Longitudine espressa in <i>gradi</i> .
Depth	Profondità in <i>m</i> rilevata mediante profondimetro.
Distance from Bottom	Distanza dal fondale in <i>m</i> rilevata mediante eco-scandaglio.
Heading	Angolo di rotta espresso in <i>gradi</i> .
Attitude	Angoli di assetto del veicolo in <i>gradi</i> . Sono riportati in ordine i valori di <i>yaw</i> , <i>pitch</i> e <i>roll</i> .

Dimensione delle mappe sonar e stima del flusso di dati acustici

Una volta stabilito il protocollo operativo della missione, sulla base dei parametri *distanza media dal fondale*, *velocità media di ricognizione* del veicolo, e quindi il *tempo ripetizione* (T_R) che intercorre tra un impulso acustico (*ping*) e il successivo, il flusso di dati prodotto dal sensore è fissato. Supponendo nota la cadenza di campionamento (T_s) del segnale si producono $N_p = T_R/T_s$ campioni per ogni ping ricevuto. Ipotizzando una quantizzazione a 8 bit si ha un flusso di dati pari a N_p byte per ogni ping e di $(N_p * T_M/T_R)$ byte per una missione di durata pari a T_M .

La dimensione ed il numero delle immagini sonografiche dipende dalla durata della missione. Le impostazioni di acquisizione del software (risoluzione, numero massimo di linee di scansione per immagine, range massimo, ed altre ancora) possono essere modificate in dipendenza delle condizioni di navigazione. Impostando la massima risoluzione *across-track*, e 512 linee di scansione per immagine, il software produce un *jpeg* di circa 90 KB.

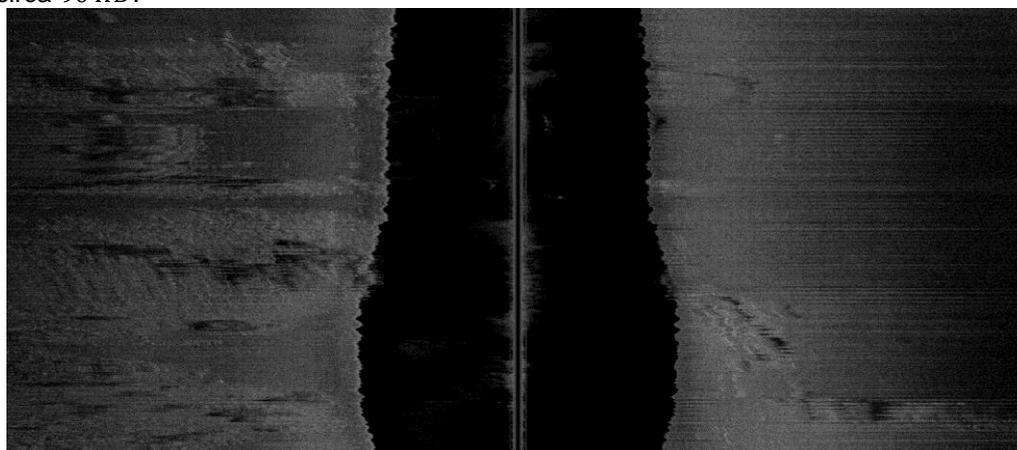


Figura 1 - Mappa sidescan acquisita durante esperimento all'Isola d'Elba

Sistema di visione stereoscopico

In condizioni sperimentali precedenti all'integrazione del sistema di visione nell'AUV in costruzione, è stato possibile produrre una coppia di flussi video a colori, in risoluzione **Full D1** e formato **PAL 720X576**, con frame rate di $25 f/s$. In tali condizioni i flussi video in formato non compresso (**RAW**) producono un throughput per ogni telecamera pari a:

$$720 * 576 * 3 \text{ byte} * 25/s \approx 30 \text{ MB/s}$$

Attualmente l'acquisizione dei video può avvenire secondo due modalità:

- La prima sfrutta il modulo *VideoCapture* della libreria *OpenCV*. Questa soluzione permette di acquisire il segnale stereo video a circa $7 - 10 f/s$ con una risoluzione di 320×240 pixel ed un ritardo sulla sincronia dovuto principalmente all'implementazione software di circa $100 ms$;
- La seconda soluzione è basata sulla libreria *libvlc*. Con questa soluzione il sistema è in grado di acquisire video a $24 f/s$ con risoluzioni possibili 640×480 o 320×240 pixel. In questo l'errore di sincronia misurato è di circa $500 - 1000 ms$. Inoltre, nel caso di risoluzione 640×480 l'acquisizione avviene in modalità interlacciata.

Per sostenere le necessità di spazio per la memorizzazione del flusso di dati stimato sopra sono stati predisposti a bordo dell'AUV due hard disk da $0,5 TB$ ciascuno; tali supporti di memoria potranno garantire missioni di durata approssimativa pari a 8 ore.

Elaborazione dei dati in tempo reale

- alert relativi al riconoscimento di indizi di reperto (linee, circonferenze, ellissi o parti di esse);
- alert relativi alla classificazione basata sulla texture.

A seconda delle possibilità del sistema potranno essere acquisiti durante la missione, le matrici di camera, il timestamp/indice di ogni frame e la geo-localizzazione del veicolo per ogni timestamp. La dimensione di tutti questi dati è comunque trascurabile rispetto a quella dei flussi video.

Filtraggio di immagini

Il segnale viene preventivamente trattato per ridurre il contributo di fonti di disturbo e per incrementare la componente di segnale utile rispetto al rumore sovrapposto. In particolare si opera un pre-filtering in modo da ottimizzare il bilanciamento del bianco e aumentare il contrasto. Si valuterà la possibilità di effettuare ulteriori step di pre-processing in dipendenza della qualità delle immagini che verranno raccolte durante le missioni.

Com'è noto le condizioni del mezzo in cui avviene la propagazione ottica (densità, temperatura, torbidità, pressione, etc.) introducono dei fenomeni di distorsione nell'immagine prodotta. Questi fenomeni vanno corretti attraverso un'operazione preliminare di calibrazione e rettifica. I file risultanti sono mappe ottiche con formato e dimensioni pari alle immagini in ingresso.

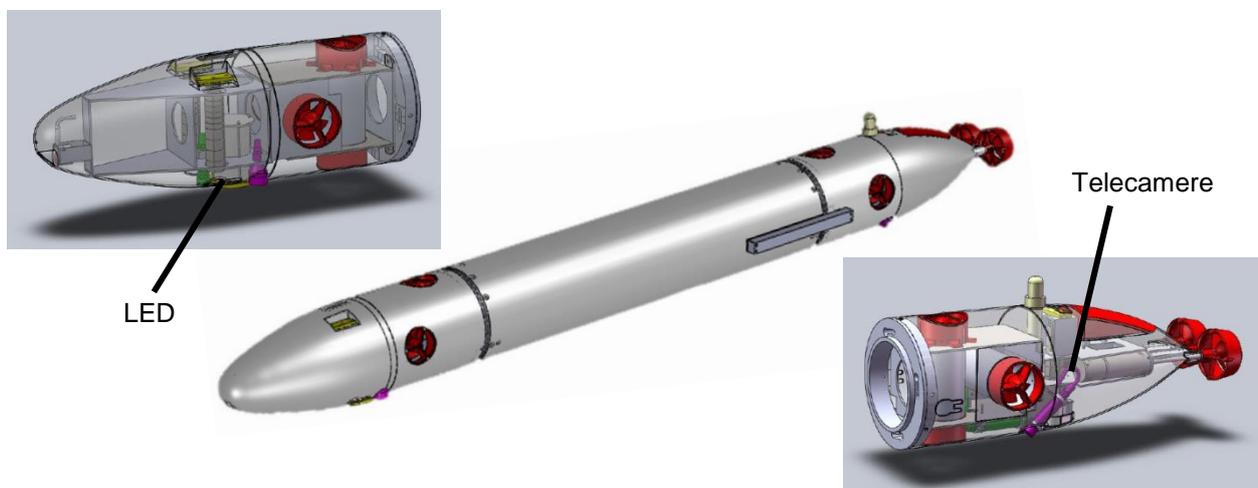


Figura 2: Piano generale dell'AUV con evidenziati gli alloggi dei LED per l'illuminazione (a prua) e delle telecamere per l'acquisizione video (a poppa).

2.1 Fonti di disturbo nei dati acquisiti

I dati raccolti durante le missioni di Thesaurus sono corrotti da una moltitudine di artefatti, i quali possono essere suddivisi in due categorie principali: la prima si riferisce a quelle sorgenti di disturbo associate alla tecnologia specifica con cui è progettato il sensore, e influenzano ogni rivelazione alterando in maniera sistematica la misura (ad esempio le distorsioni associate al passaggio della luce attraverso il sistema ottico). La seconda categoria raggruppa tutte le sorgenti esterne di rumore, dovute alle condizioni che caratterizzano l'ambiente circostante (rumore ambientale, condizioni di luce e visibilità, etc..).

Nel seguito è riportata una breve analisi delle degradazioni tipiche coinvolte nella visione stereoscopica e nella mappatura tramite side-scan sonar: la correzione di queste distorsioni è un'operazione fondamentale per garantire elaborazioni più complesse, tra cui l'implementazione di procedure di computer vision e object recognition.

2.1.1 Artefatti nella visione stereoscopica subacquea

La formazione delle immagini da dispositivi ottici è sistematicamente contaminata da distorsioni radiali e tangenziali dovute alla propagazione della luce attraverso le lenti della telecamera. Di conseguenza la geometria della scena riprodotta nell'immagine appare distorta. Questo effetto può essere corretto stimando i parametri intrinseci ed estrinseci del sistema ottico e sfruttando i valori ottenuti per effettuare la **calibrazione**. Tale operazione, che dovrà essere ripetuta all'inizio di ogni missione, consentirà la correzione dei fotogrammi acquisiti mediante una procedura di **rettifica**. Tale operazione è propedeutica alle elaborazioni off-line dedicate alla ricostruzione metrica degli ambienti esplorati.

Le immagini catturate in ambienti subacquei risentono di condizioni di visibilità tipicamente poco favorevoli. In acqua molti fattori determinano la visibilità: luminosità, indice di rifrazione, presenza di materiale in sospensione, turbolenza, diffusione, assorbimento della luce non omogeneo nello spettro del visibile. Ciò si traduce in alcune rilevanti conseguenze: perdita di contrasto, forte dipendenza della visibilità dalla distanza tra l'oggetto illuminato e le telecamere, perdita di colore, blur, illuminazione non uniforme. L'acqua si comporta in maniera analoga ad un filtro polarizzatore: la luce riflessa da oggetti sommersi si propaga attraverso l'acqua e viene raccolta dal sistema ottico dopo aver subito una riduzione del suo contenuto spettrale (effetto di sfumatura verso il blu/verde). L'utilizzo di appropriate tecniche di restauro del segnale consente il recupero parziale delle proprietà cromatiche e del contrasto della scena riprodotta.

2.1.2 Artefatti nelle immagini side-scan sonar

- L'intensità acustica degli echi di ritorno risente di perdite dovute alla propagazione attraverso l'ambiente marino. L'energia trasportata dall'onda acustica viene ceduta in parte al mezzo circostante mediante fenomeni di assorbimento; inoltre il modulo dell'ampiezza decade in modo inversamente proporzionale al quadrato della distanza, a causa della propagazione su fronti d'onda sferici (*divergenza sferica*), pertanto il rapporto segnale/rumore diminuisce drasticamente all'aumentare della distanza. Il segnale analogico in uscita dal side-scan sonar viene tipicamente pre-trattato tramite un Time Variable Gain Filter, un filtro che amplifica il segnale di eco secondo una curva di guadagno proporzionale alla distanza.

- Per una ricostruzione accurata della geometria del fondale da immagini side-scan non è sufficiente disporre le singole strisciate una sopra l'altra, seguendo l'ordine cronologico (visualizzazione *waterfall*). Difatti la collocazione di un pixel nella mappa sonografica dipende dall'istante temporale in cui l'onda acustica riflessa da un punto sul fondale raggiunge l'antenna ricevente. Questo comporta un'alterazione della geometria percepita dal sensore in quanto la distanza percorsa dall'onda, e di conseguenza l'intervallo temporale tra una rivelazione e la successiva, ha un andamento non lineare al crescere della distanza raggiunta. Il campionamento del fondale realizzato mediante la raccolta degli echi acustici risente di una compressione dei dati in prossimità della verticale (centro della mappa), in corrispondenza cioè della zona in cui si manifesta una maggiore non-linearità nella mappatura. La trasformazione delle coordinate che provoca tale effetto è rappresentato in Figura 3.

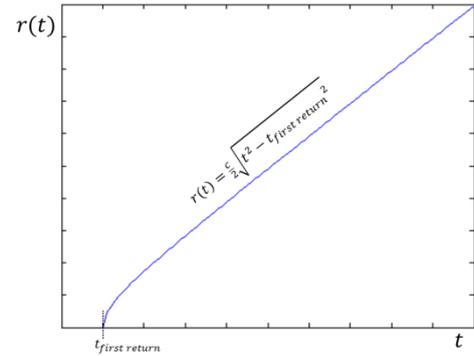


Figura 3 - Mappatura del sonar

- Allo scopo di elaborare le mappe acustiche seguendo l'approccio della computer vision è opportuno ripristinare le corrette proprietà geometriche dei dati sonar. Ciò si ottiene implementando una trasformazione da coordinate **Slant Range** a **Ground Range**. Quest'operazione comporta l'introduzione nella mappa di nuovi punti, i cui valori d'intensità dovranno essere stimati per interpolazione. Le mappe risultanti avranno dimensioni approssimativamente maggiori o uguali a quelle delle immagini in ingresso, in misura variabile a seconda del numero di interpolazioni realizzate.
- Lo stato dell'ambiente circostante (correnti marine, moto ondoso) può influenzare significativamente l'equilibrio dinamico del supporto su cui è installato il sonar, provocando un peggioramento della qualità del dato sonografico. Se il veicolo è equipaggiato con adeguati dispositivi per il controllo della navigazione e dello stato di moto è possibile contrastare questo effetto. In particolare le stime di assetto del veicolo possono essere direttamente utilizzate per riferire le scansioni ad un unico sistema di coordinate (ad esempio il sistema di coordinate proprio del **GPS**).
- Il campo acustico trasmesso dall'array di trasduttori del sonar è descritto da un caratteristico profilo di radiazione nel piano verticale. Questo comporta una copertura acustica (*insonificazione*) del fondale non uniforme. Il profilo di radiazione prodotto da una configurazione nota di trasduttori è un risultato noto in letteratura ([11]) e può essere utilizzato per correggere in maniera opportuna le misure acustiche: moltiplicando ogni valore d'intensità per un opportuno fattore di normalizzazione si compensano quei valori sfavoriti poiché dovuti a riflessioni provenienti da aree raggiunte da un campo acustico più debole.

Nei capitoli successivi si suppone che tutte le operazioni di elaborazione preliminare siano state correttamente eseguite, in modo da aver escluso quelle componenti di segnale rumoroso o di errore sistematico che contaminano il dato grezzo, isolando i segnali ottici e acustici di interesse.

2.2 Elaborazioni preliminari

In questo capitolo esplicitiamo meglio i trattamenti preliminari che abbiamo studiato al fine di scegliere quelli più performanti rispetto al nostro setting sperimentale.

2.2.1 Pre-filtering ottico

Per quanto riguarda la calibrazione del sistema visivo e la conseguente rettifica, abbiamo per ora utilizzato un toolbox di calibrazione implementato in MATLAB e ispirato al classico lavoro di Zhang ([12]). Dunque si prevede che le procedure di calibrazione vengano effettuate all'inizio di ogni missione, mediante l'acquisizione di sequenze ottiche di appositi pattern geometrici (per es. una scacchiera) *in situ*, mentre l'accurata rettifica si opera successivamente in modalità off-line, preliminarmente alla ricostruzione

Identificativo: Report D3.3

Nome del file: D3point_3_01.doc

Titolo: Studio e sviluppo per l'analisi e l'annotazione delle immagini e delle mappe

tridimensionale della scena. Tale procedimento è propedeutico ad una rappresentazione tridimensionale e metrica dei dati ottici acquisiti; ciononostante la mancata esecuzione di questo procedimento non incide il riconoscimento nella scena di forme regolari in essa presenti, ad esempio attraverso l'estrazione e la classificazione di profili 2D.



Figura 4 - Immagine originale a sinistra, prefiltrata a destra [13]. Immagine reperibile in rete

Le tecniche utilizzate per il prefiltering ottico consistono in: bilanciamento del bianco, seguendo l'approccio di Finlayson riguardo "Shades of gray and colour constancy"; miglioramento del contrasto per mezzo di un'appropriata histogram equalization (o adaptive histogram equalization), con eventuali aggiustamenti per consentire la preservazione della edge map. ([13], [14], [15]).

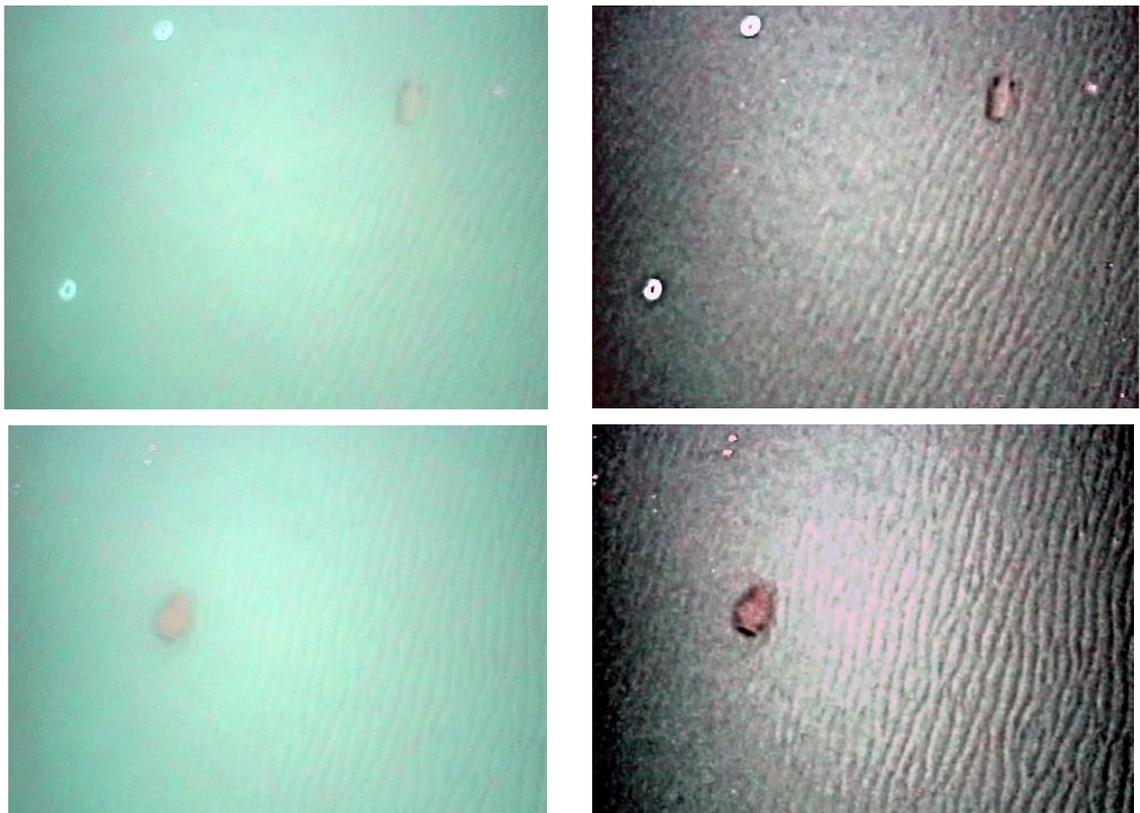


Figura 5 - Immagini originali a sinistra, prefiltrate a destra [14], acquisite a Marciana Marina, Elba, il 18 Aprile 2012

Identificativo: Report D3.3

Nome del file: D3point_3_01.doc

Titolo: Studio e sviluppo per l'analisi e l'annotazione delle immagini e delle mappe

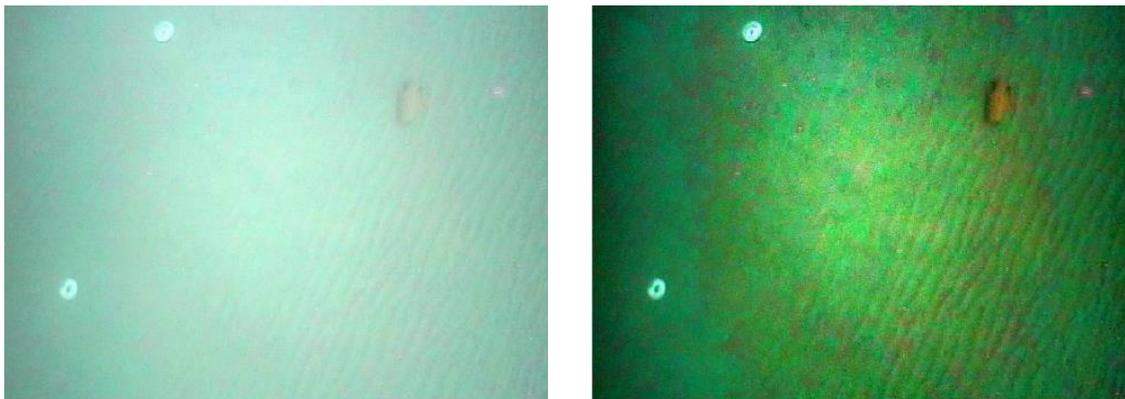


Figura 6 - Immagine originale a sinistra – pre-filtrata a destra [13]

3 Registrazione e mosaicatura di immagini e mappe

In questa sezione si descrivono in dettaglio alcune tecniche di analisi delle immagini che permettono di ottenere mappe su larga scala della zona esplorata. Nel corso dell'attività di progetto sono state implementate procedure che realizzano la fusione di singole immagini di una stessa zona, e restituiscono una mappa d'insieme, risultato dell'unione delle informazioni contenute in ogni singola ripresa. Quest'operazione può essere condotta, seguendo procedimenti distinti, sia per dati ottici che per dati acustici.

3.1 Registrazione

Si ricorda che le misure raccolte dal sonar sono scansioni lineari strutturate in una mappa bidimensionale. Per questo tipo di dato l'operazione di registrazione si traduce nell'assegnare ad ogni scansione un insieme di coordinate relative ad un unico sistema di riferimento, ad esempio quello del **GPS**. Questo è possibile se la velocità di aggiornamento del dato geografico è sincrona o almeno comparabile con la cadenza di acquisizione delle linee sonografiche. In tal caso ogni strisciata può essere geo-referenziata e correttamente riportata su una mappa.

Per quanto riguarda la fusione di dati ottici catturati da molteplici punti e angolazioni una prima essenziale elaborazione consiste nel ricondurre tutte le mappe ad un unico sistema di coordinate: per fare ciò si stima la trasformazione geometrica (*omografia*) che esegue la mappatura di un pixel di una qualsiasi delle immagini, con coordinate riferite a quella immagine, in un pixel con coordinate riferite ad un sistema globale, opportunamente fissato. Il risultato dell'applicazione di questa trasformazione ad ogni immagine costituisce un insieme di mappe bidimensionali geometricamente coerenti tra loro, in quanto relative ad un unico sistema di riferimento.

In letteratura esiste un'ampia attività di sviluppo di algoritmi in grado di svolgere il compito suddetto. Un metodo solidamente accettato da parte della comunità scientifica si basa sull'utilizzo della *Scale Invariant Feature Transform (SIFT)*, [16] per individuare e caratterizzare i punti salienti (*keypoints*) in ciascuna immagine. In particolare, questa tecnica consiste nella ricerca, in intorno dei pixel di un'immagine, di punti di massimo e minimo di una funzione di costo propriamente definita. Il requisito fondamentale su cui si basa questa procedura è che i suddetti punti estremali risultino invarianti per scala, orientazione e distorsione affine e parzialmente invarianti a cambi di illuminazione, (e cioè consistenti sotto l'applicazione di filtri definiti dalla differenza tra nuclei gaussiani). In più, questo metodo risulta particolarmente robusto rispetto al rumore dell'immagine, condizione abbastanza significativa per le immagini sottomarine. L'identificazione dei punti salienti avviene reiterando la procedura e incrementando, ogni volta di un'ottava, la risoluzione dell'immagine in ingresso.

Una volta individuato un set preliminare di corrispondenze tra i punti salienti (*match*) in diversi fotogrammi si procede alla stima della matrice della trasformazione omografica H mediante l'algoritmo di *Random Sampling Consensus (RANSAC)*, [17]), e sulla base dei risultati, si eliminano i *match* spuri. In conclusione, dati due pixel di coordinate x_a e x_b , che identificano uno stesso punto nello spazio ma catturato in due distinti fotogrammi, e la matrice $H_{a \rightarrow b}$ che definisce la trasformazione geometrica da un sistema all'altro, la procedura di registrazione si formalizza come:

$$x_b = H_{a \rightarrow b} x_a$$

Una volta identificati i punti corrispondenti in più immagini e dopo aver effettuato la registrazione, le singole mappe vengono "incollate" seguendo il metodo descritto nella sezione 3.2. Il risultato è la creazione di una mappa completa della regione esplorata. L'immagine prodotta avrà dimensioni maggiori dei singoli file di ingresso, in proporzione alle dimensioni dell'intera zona esplorata.

Nell'attuale implementazione del software le mappe di insieme sono corredate da una serie di metadati, in modo da fornire un'accurata descrizione della scena.

3.2 Mosaicatura

In questa sezione viene descritto il procedimento di creazione di mappe di insieme, noto come *mosaicatura*, con riferimento sia al caso acustico che ottico.

3.2.1 Metodi per la mosaicatura di scansioni acustiche

L'acquisizione dei dati sonografici può essere trattata in modo da produrre mappe del fondale marino, utilizzabili per l'analisi dettagliata dell'ambiente esplorato. Il procedimento di creazione di queste mappe comporta che le singole scansioni sonografiche vengano collocate in posizioni adiacenti, in modo da formare un'unica mappa estesa. Questa procedura, realizzata in passato ritagliando e incollando direttamente le scansioni stampate su una mappa cartacea, oggi è implementata all'interno di appropriati software commerciali. Per permettere un posizionamento corretto delle scansioni è necessario che i dati sonografici vengano combinati con i dati di navigazione. L'operazione essenziale della mosaicatura di sonogrammi consiste nel visualizzare ogni pixel sonar tenendo conto della traiettoria del supporto del sensore ovvero del veicolo subacqueo. La giustapposizione diretta delle linee di scansione (visualizzazione *waterfall*, esempio in Figura 7) induce a pensare che la traiettoria del veicolo sia perfettamente lineare e che i profili dei vari echi di ritorno forniscano una riproduzione reale della morfologia del fondale.

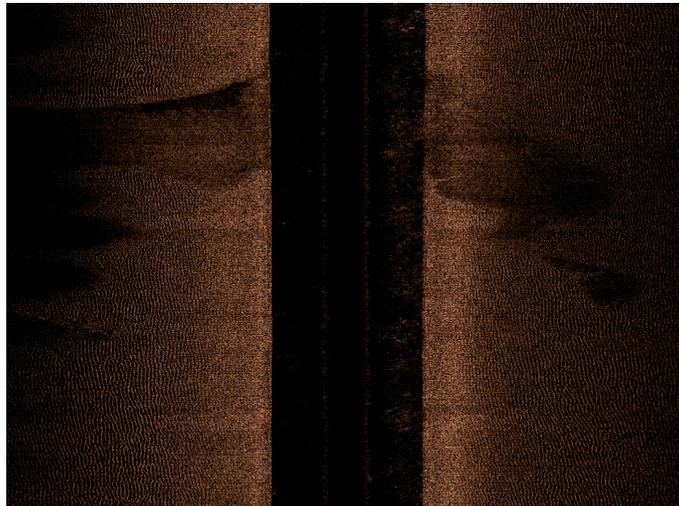


Figura 7 - Visualizzazione Waterfall

In realtà il percorso del veicolo coincide raramente con la linea retta che caratterizza la semplice visione *waterfall* del sonogramma.

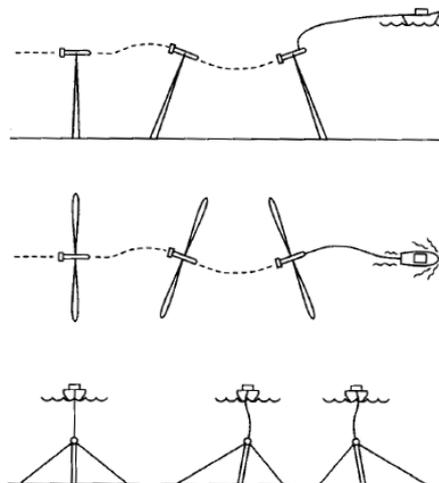


Figura 8 - Esempi di perturbazioni rotazionali nel moto del sensore (dall'alto rispettivamente movimento di pitch, yaw e roll)

Fattori che influiscono sullo stato di assetto del veicolo e sull'instabilità del moto sono in particolare rappresentati dall'azione delle correnti marine. Le possibili perturbazioni nell'assetto si suddividono in due tipi: *traslazionali* e *rotazionali*. Le prime sono dovute al moto di spostamento laterale o verticale del veicolo rispetto alla direzione rettilinea desiderata. Questo tipo di distorsioni è correggibile nel caso in cui sia nota la posizione del sensore nello spazio, la quale può essere ricavata mediante sensori di profondità installati a bordo veicolo. Integrando tutte le misure di assetto si ottiene una stima periodicamente aggiornata della posizione dei veicoli nello spazio.

Le perturbazioni del secondo tipo sono illustrate in Figura 8. I movimenti di *pitch* (*beccheggio*) e *yaw* (*imbardata*) introducono distorsioni geometriche nei dati acquisiti, mentre variazioni di *roll* (*rollio*) provocano distorsioni di intensità dell'eco di ritorno in quanto l'orientamento dei lobi acustici viene temporaneamente alterato rispetto alla sua configurazione di equilibrio. In caso di concomitanza delle perturbazioni sopra descritte (esempio in Figura 9) l'acquisizione dei dati può risultare significativamente corrotta e, conseguentemente, la successiva elaborazione ed analisi delle immagini diventa complessa e facilmente soggetta ad interpretazioni scorrette.

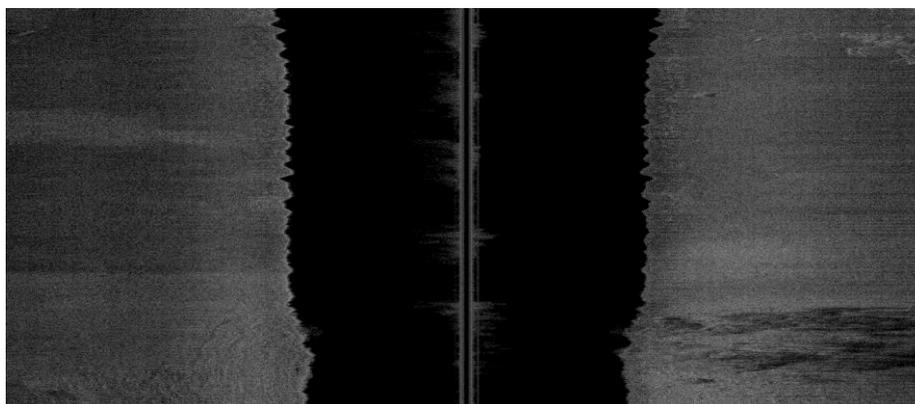


Figura 9 - Esempio di acquisizioni da side scan sonar corrotte da molteplici artefatti

A tutti gli effetti è necessario effettuare la geo-referenziazione per ogni strisciata sonar, posizionandola in maniera corretta all'interno di un opportuno sistema di coordinate. Ne consegue che la qualità della mappa finale dipenderà fortemente dalle modalità e dall'accuratezza con cui sono state effettuate le acquisizioni. Operativamente è di fondamentale importanza garantire una navigazione fluida, evitando brusche o repentine variazioni nella traiettoria e nell'assetto del veicolo.

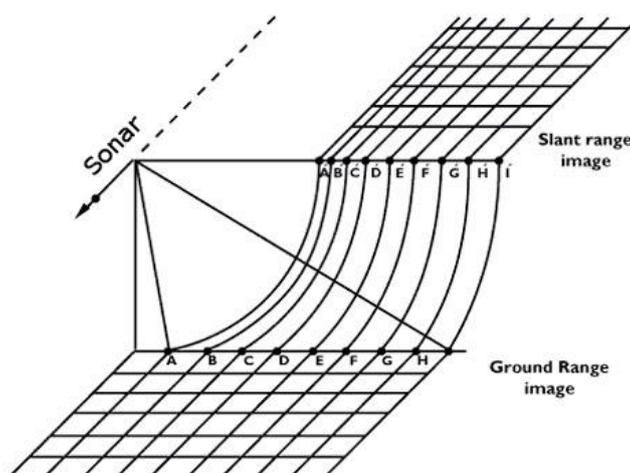


Figura 10 - Coordinate Slant e Ground

La procedura che conduce alla mappa finale prevede alcuni passi operativi: in primo luogo il sonogramma dev'essere corretto per rimuovere l'informazione inutilizzata relativa al passaggio della perturbazione acustica attraverso la colonna d'acqua. Questo si ottiene implementando la trasformazione da coordinate Slant Range, che definiscono una visualizzazione in modalità *waterfall*, a Ground Range (vedi Figura 10).

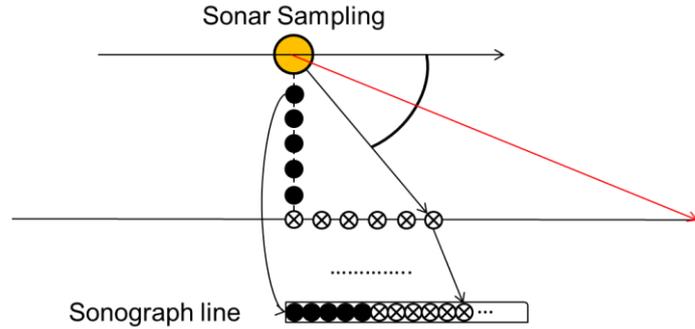


Figura 11 - Campionamento del side scan sonar

La geometria che caratterizza il campionamento del sidescan sonar permette di intuire la procedura di elaborazione necessaria per correggere le distorsioni geometriche dovute alla "slanted view" del sonar. In Figura 12 è rappresentata la trasformazione che permette di recuperare la posizione effettiva dei campioni rispetto alla verticale.

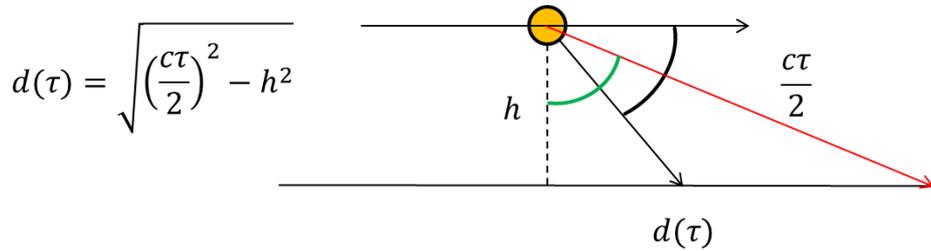


Figura 12 - Trasformazione Slant to Ground

Per poter applicare questa trasformazione è necessario conoscere la distanza del sonar dal fondale per ogni scansione effettuata. In tal modo è possibile calcolare il tempo di attraversamento della perturbazione acustica attraverso la colonna ed eliminare da ogni scansione l'insieme di campioni superflui.

È stato sviluppato un codice in ambiente MATLAB per implementare la procedura suddetta; a titolo di esempio, il suo funzionamento è stato verificato sull'immagine in Figura 7. La mappa utilizzata non contiene informazioni di assetto del veicolo né di posizionamento geografico pertanto i valori di quota sono stati stimati mediante elaborazione diretta dell'immagine. Il risultato è riportato in Figura 13.

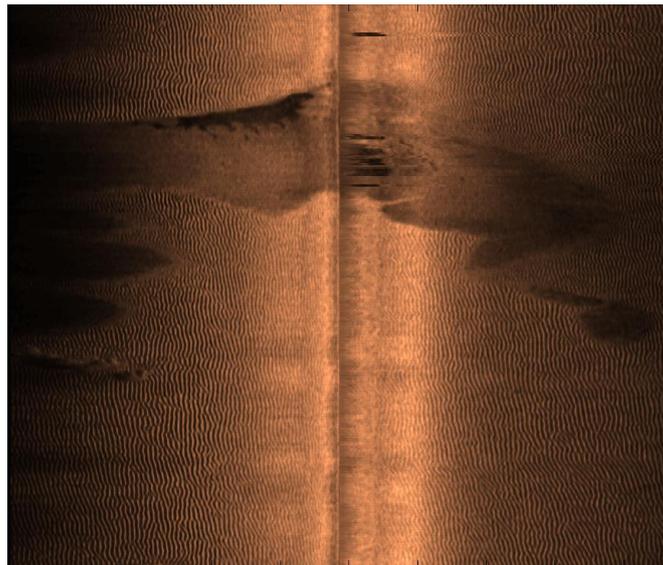


Figura 13 - Mappa Sonar in coordinate Ground Range

Una volta ottenuta una mappa in coordinate relative al piano del fondale marino i dati possono essere ulteriormente elaborati per correggere distorsioni geometriche di tipo rotazionale. Di particolare importanza è la rimozione di artefatti associati a movimenti di beccheggio e imbardata, il cui effetto tipico è quello di

replicare alcune scansioni nella mappa, alterando la percezione ed eventualmente la stima delle reali dimensioni degli oggetti. Una perturbazione di beccheggio e imbardata ad elevata intensità può far sì che il fascio acustico ricopra più volte una stessa regione del fondale. Gli oggetti che si trovano posizionati in una regione soggetta a scansioni molteplici appaiono triplicati nel sonogramma: un'immagine dovuta alla prima scansione, una seconda causata dalla perturbazione ed un'ultima nel moto inverso durante il quale il veicolo si riporta alle condizioni di equilibrio.

3.2.1.1 Correzione della distorsione dovuta all'imbardata

Come anticipato nella sezione precedente la perturbazione d'imbardata può essere corretta applicando una trasformazione che ripristini la posizione effettiva delle linee scansionate in un opportuno sistema di coordinate. In Figura 14 è rappresentata la perturbazione di imbardata e la conseguente rotazione del fascio acustico su una regione diversa da quella attesa.

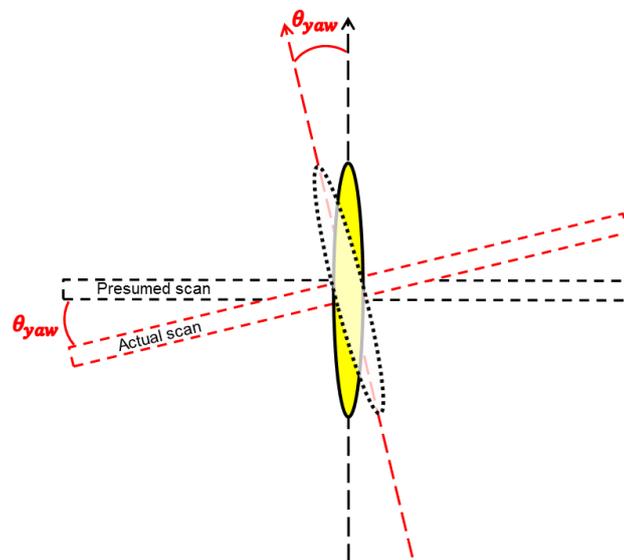


Figura 14 - Movimento di imbardata

In condizioni di elevate perturbazioni di imbardata l'immagine ottenuta dalla sovrapposizione diretta delle linee di scansione può risultare molto corrotta. Come già accennato se si hanno a disposizione informazioni di assetto del veicolo, misurate ad esempio mediante una bussola giroscopica, si possono correggere le distorsioni introdotte nell'immagine. Per simulare un'acquisizione di dati contaminati si è considerato l'immagine in Figura 15. Supponendo che ogni riga sia il risultato di una scansione acustica, si può pensare di effettuare nuovamente la scansione ruotando ogni riga di una quantità casuale e memorizzando ogni valore angolare in un opportuno vettore. Come si vede nella parte destra della Figura 15 la mappa risulta visibilmente corrotta e in alcune zone dell'immagine risulta difficile interpretare correttamente la scena e riconoscere gli oggetti che vi sono rappresentati.



Figura 15 - Immagine originale e simulazione di un'acquisizione corrotta da moto di imbardata

In Figura 16 si può apprezzare il risultato della trasformazione inversa che permette il recupero del reale orientamento delle linee di scansione (precedentemente memorizzato). Si può notare che parte dell'informazione è inevitabilmente andata perduta in quanto una mappatura acustica non uniforme contiene punti in cui l'informazione è assente (gap all'interno della mappa). Mediante un procedimento di interpolazione che permetta di stimare i valori dei pixel andati perduti si ottiene il risultato visibile nella parte destra di Figura 16.



Figura 16 - Ripristino della posizione delle linee e risultato finale dopo interpolazione

3.2.1.2 Correzione della distorsione dovuta al moto di beccheggio

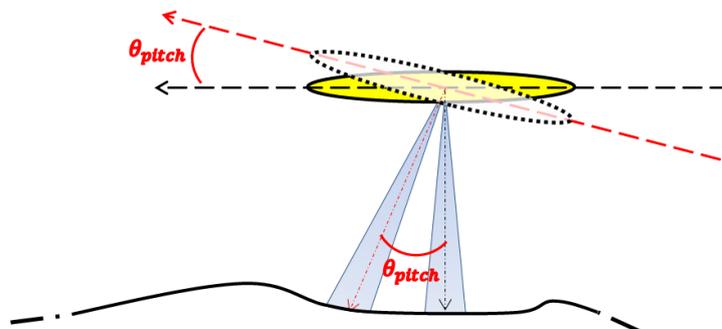


Figura 17 - Moto di beccheggio

A causa dei movimenti di beccheggio indotti dalle correnti circostanti l'area definita dall'intersezione tra fascio acustico e superficie del fondale si sposta rispetto alla verticale. A causa della maggiore distanza

percorsa dalla perturbazione acustica la proiezione del fascio sul fondale evidenzia una striscia di lunghezza più estesa rispetto al caso di beccheggio nullo. La conseguenza più significativa consiste nella effettiva scansione di una porzione di fondale traslata rispetto al nadir. In Figura 18 è schematizzato l'effetto descritto.

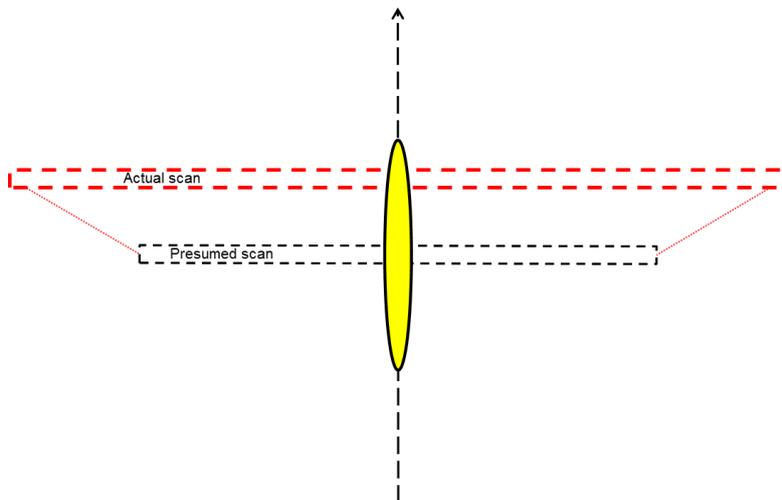


Figura 18 – Alterazione della zona di scansione a causa del moto di beccheggio

Per correggere l'effetto si ricorre nuovamente all'uso delle informazioni di assetto misurate mediante dispositivi opportuni installati sull'AUV.

Supponiamo che il side scan sonar sia stato installato in modo da garantire che, per ogni canale, il fascio acustico ricopra una regione angolare, nel piano *across-track*, di ampiezza θ_{beam} con origine a partire dalla verticale (Figura 19).

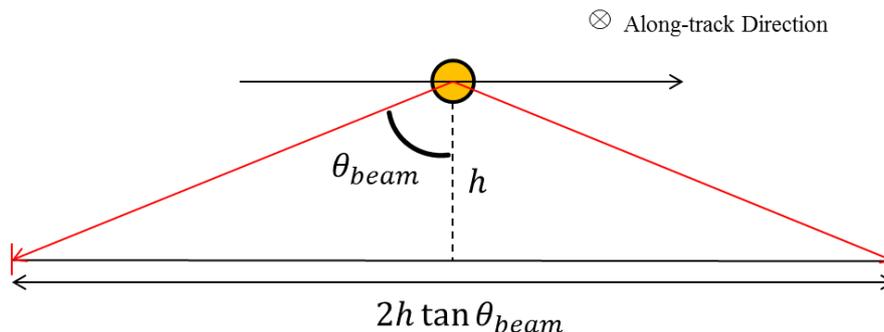


Figura 19 - Apertura del fascio acustico

Se il veicolo ruota di un angolo θ_{pitch} rispetto all'orizzontale il fascio acustico proietterà una striscia sul fondale il cui punto più vicino al veicolo, corrispondente al primo eco di ritorno nel sonogramma, è situato ad una distanza $h / \cos \theta_{pitch}$. La lunghezza della proiezione del fascio ruotato diventa $2h \tan \theta_{beam} / \cos \theta_{pitch}$ e la distanza sul fondale tra il punto di primo eco di ritorno e l'intersezione del nadir è pari a $h \tan \theta_{pitch}$. Supponendo un'apertura del fascio nel piano *along-track* di valore φ si ha che la relazione tra il valore di riga del sonogramma i e il valore i' corrispondente alla posizione effettiva della striscia di scansione è stabilito da:

$$|i - i'| = \frac{\tan \theta_{pitch}}{\varphi} \quad (1)$$

Conoscendo θ_{pitch} dalle misure della bussola giroscopica è possibile invertire l'equazione (1) per recuperare il valore i' .

Nell'ipotesi che gli angoli di beccheggio si mantengano tanto piccoli da poter approssimare $\cos \theta_{pitch} \sim 1$ si può supporre che la lunghezza della nuova strisciata si mantenga inalterata sotto l'effetto della rotazione di beccheggio. Sotto tale ipotesi la correzione della distorsione di beccheggio si traduce nel riposizionamento delle righe secondo la regola stabilita dall'equazione (1).

Per simulare un'acquisizione di dati contaminati si è nuovamente considerata l'immagine in Figura 15. L'introduzione nell'immagine di una perturbazione di beccheggio può essere simulata ridisponendo le righe secondo la regola definita in (1). I valori di θ_{pitch} vengono generati casualmente e memorizzati in un

opportuno vettore. Il risultato della simulazione è apprezzabile nella parte a sinistra della Figura 20, dove si nota come i dati siano stati significativamente corrotti.



Figura 20 - Simulazione della distorsione nella mappa sonografica dovuta al moto di beccheggio

Nella parte destra della Figura 20 si può apprezzare il risultato della trasformazione inversa che permette il recupero del reale posizionamento delle linee di scansione. Così come per la correzione della distorsione di imbardata si nota che parte dell'informazione è inevitabilmente andata perduta a causa della disomogeneità della mappatura acustica. Sono infatti presenti numerose regioni in cui l'informazione associata al dato acustico risulta dispersa o inesistente. Mediante un procedimento di interpolazione che permetta di stimare i valori dei pixel andati perduti si ottiene il risultato in Figura 21.



Figura 21 - Simulazione della correzione della distorsione dovuta a beccheggio

3.2.2 Metodi per la mosaicatura di immagini ottiche

La mosaicatura di immagini ottiche è da lungo tempo un importante argomento di studio per la comunità scientifica e le numerose implementazioni esistenti vengono impiegate per molteplici usi ed applicazioni. Le possibilità sono varie e spesso contengono anche l'integrazione dei dati ottici con quelli di posizione ottenuti dal sistema di navigazione dell'AUV. Quest'ultima opzione risulta irrinunciabile nel caso in cui il sistema di controllo della navigazione preveda l'integrazione con i dati ottici (ad esempio qualora si decida di usare parte dei dati SLAM come correttivi per quelli di navigazione e viceversa). Questo non è il caso dell'AUV prodotto per il progetto THESAURUS (*Tifone*), in cui i dati ottici vengono elaborati ed integrati in real-time al fine di ottenere un indice di probabilità di scoperta di oggetti di rilievo archeologico (ed avviare eventuali ripianificazioni della missione originaria), ed off-line al fine di ricostruire accuratamente ciò che è stato ritenuto interessante.

Rimane dunque di fondamentale importanza un'operazione di mosaicatura della regione esplorata che consenta la creazione di una mappa di insieme e la ricostruzione della scena in cui allocare gli oggetti di interesse a cui si prevede di associare un dettagliato modello 3D.

In questa sezione si è descritta la procedura adottata in Thesaurus per la creazione di un'unica mappa su larga scala, sia a partire da un insieme di rivelazioni sonografiche che da un insieme di singole immagini cronologicamente ordinate (flusso video).

Nel caso ottico il procedimento di elaborazione prevede:

- i) la stima delle matrici di omografia per ogni coppia di immagini consecutive tramite SIFT
- ii) l'applicazione della trasformazione che mappa ogni immagine all'interno di un unico sistema di riferimento
- iii) la giustapposizione delle mappe sfruttando le aree di sovrapposizione e le relative corrispondenze.

In considerazione dell'onere computazionale associato a questo tipo di algoritmi le operazioni di registrazione e mosaicatura dei dati verranno eseguite unicamente in modalità off-line.

A questo scopo abbiamo circoscritto la ricerca di descrittori significativi per i punti salienti a due: Harris ([18]) e SIFT ([16]), scegliendo infine il secondo, in quanto si dimostra più robusto rispetto al rumore.

Il SIFT (di cui si è già accennato nella sezione 3.1) è un metodo usato per determinare i punti salienti dell'immagine invarianti rispetto a cambi di scala, prospettiva, illuminazione, e particolarmente robusti rispetto al rumore nell'immagine. L'algoritmo estrae i punti di interesse nell'immagine grazie ad una serie di procedure, come la procedura multi-risoluzione nota come *pyramid of images*, differenze di Gaussiane (come il filtro passa-banda), e la ricerca di punti estremali di un'opportuna funzione costo. Ognuno di tali punti, con la direzione del campo gradiente nel punto stesso, si dice *Keypoint*. Per ogni Keypoint si definisce così un descrittore, dato da un vettore di 128 componenti che rappresenta l'orientazione del gradiente nel punto stesso; la norma del gradiente si definisce considerando una finestra 4x4 attorno al punto saliente. Il SIFT consente di produrre un insieme denso di punti salienti.

Creazione di una mappa su larga scala

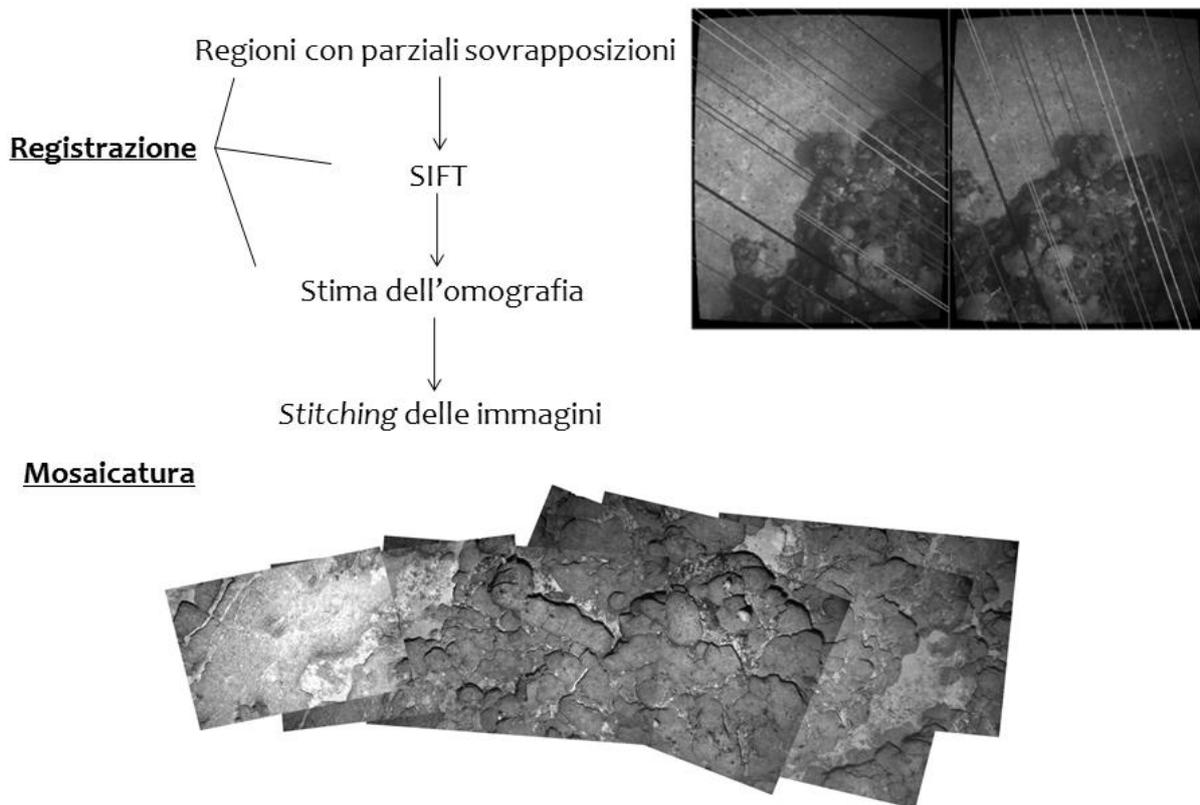


Figura 22 - Registrazione e Mosaicatura

Identificativo: Report D3.3

Nome del file: D3point_3_01.doc

Titolo: Studio e sviluppo per l'analisi e l'annotazione delle immagini e delle mappe

In questo modo è possibile produrre il mosaico desiderato. Notiamo in particolare che l'integrazione dei dati ottici in un orto-mosaico (mosaico composto da immagini rettificate e fuse opportunamente) è il passo base per la fusione con gli altri tipi di dati che sono acquisiti durante la missione, come ad esempio la classificazione delle aree di fondale in base alla tessitura ottica e acustica, la referenziazione geografica degli alert relativi alla presenza di strutture regolari, la probabile presenza di oggetti di interesse.



Figura 23 - Esempio di SIFT applicato ai dati raccolti a Marciana Marina.

4 Algoritmi per l'interpretazione automatica di scenari subacquei

In questo capitolo esponiamo una lista di algoritmi per l'interpretazione automatica di scenari subacquei. In particolare, nei successivi paragrafi verranno messi in evidenza, tra quelli studiati, gli algoritmi scelti per Thesaurus.

4.1 Interpretazione automatica mediante tecniche multi-view

L'area di ricerca della ricostruzione 3D affonda le radici nella fotogrammetria classica; grazie alla crescente richiesta di nuove tecnologie, negli ultimi decenni molte di queste sono state dedicate proprio al complesso settore dei beni culturali. Le tecniche a cui ci riferiamo per la ricostruzione 3D nel caso dei beni archeologici subacquei sono:

Ricostruzione 3D da Immagini Stereo

Ogni acquisizione video, tramite il sistema di telecamere in configurazione stereoscopica, fornisce una ripresa della stessa scena da due punti di vista diversi. L'elaborazione di questi dati mediante algoritmi di tipo *SFM* e *SLAM* consente una stima del profilo 3D del fondale. La creazione delle mappe avviene stimando le coordinate spaziali di un insieme ristretto di punti ed interpolando per estendere l'informazione al dominio considerato. Il formato di uscita è un file contenente le coordinate spaziali dei punti appartenenti alla scena.

Nelle immagini qui di seguito è possibile osservare lo schema di ricostruzione utilizzato nel caso di flussi video sincronizzati e rettificati acquisiti durante la missione in Marciana Marina. Nel caso specifico del test, è stato possibile ricostruire un modello della scena piuttosto grezzo, principalmente a causa della bassa risoluzione, in rapporto alla distanza intercorrente tra le telecamere e l'anfora posta sul fondale (circa 8 metri).

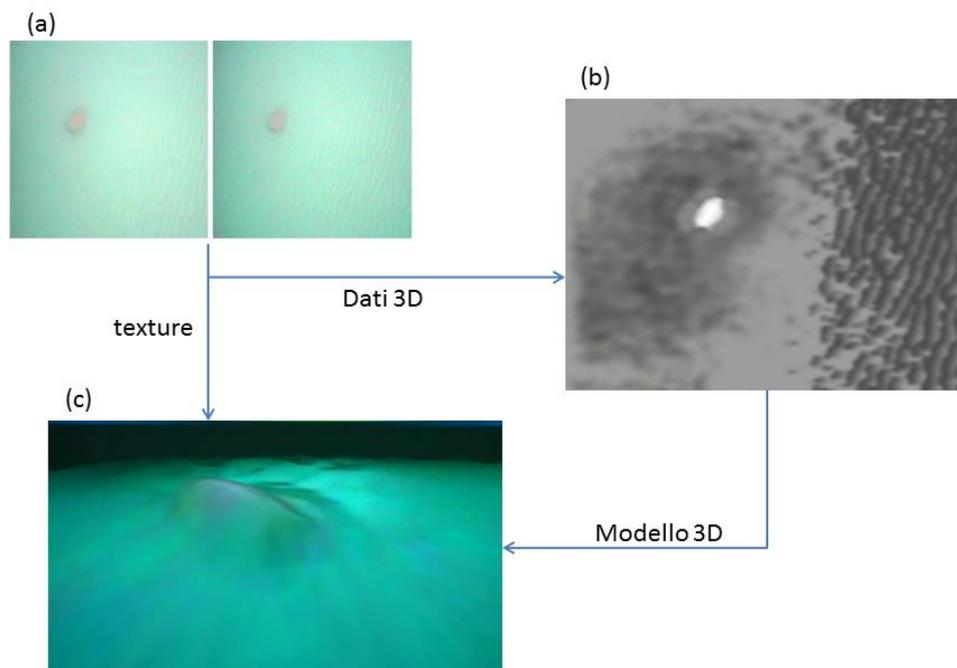


Figura 24 Schema di ricostruzione da coppia di flussi video: (a) Coppia di immagini stereo, estratte dai flussi video sincronizzati e rettificati; (b) Recupero delle coordinate tridimensionali di oggetti nella scena mediante elaborazione di depth map; (c) Ricostruzione della scena associata alla *depth map* con applicazione della texture originale.

Identificativo: Report D3.3

Nome del file: D3point_3_01.doc

Titolo: Studio e sviluppo per l'analisi e l'annotazione delle immagini e delle mappe

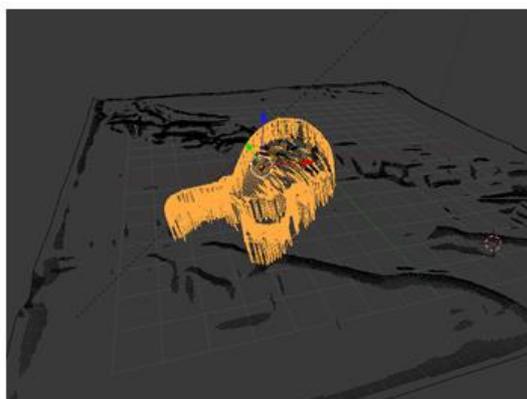
La dimensione dei file di output dipende dal tipo di mesh scelta per costruire la superficie che viene generata dalla nuvola di punti, dalla risoluzione desiderata per la mappa finale e dal passo di interpolazione scelto. Per esempio, se si usa una mesh di triangoli, un'interpolazione comune è quella di aggiungere il centro di ogni triangolo, moltiplicando così per 3 il numero di facce che rappresentano il modello. Nel caso di un file di dimensione 1.8 MByte quest'operazione comporta l'incremento di un fattore 16 della dimensione del file. Questa tipologia di elaborazione verrà eseguita solo in modalità off-line.

Ricostruzione 3D da Lama Laser

La proiezione sul fondale esplorato di un fascio di luce laser a profilo laminare, fornisce informazioni sulle coordinate dei punti nella scena osservata. Algoritmi di ricostruzione applicati a immagini stereo in cui sia visibile la proiezione del fascio laser producono mappe 3D dense. Il formato dell'output è un file contenente le coordinate dei punti nello spazio tridimensionale. Le osservazioni fatte al punto precedente (*Ricostruzione 3D da immagini stereo*) relativamente alla stima della dimensione dei file di output valgono anche per i modelli 3D ottenuti tramite lama laser. Questa tipologia di elaborazione verrà eseguita solo in modalità off-line. In questo caso sono stati condotti dei test in laboratorio per validare lo schema di ricostruzione 3D con l'utilizzo della lama laser. Sono state montate su uno stesso supporto due telecamere ed un illuminatore capace di proiettare una lama laser, e fatte muovere parallelamente al piano d'appoggio dell'oggetto (anfora). In questo modo è stato possibile acquisire due flussi video relativi all'oggetto, ed operare una ricostruzione tramite un algoritmo sviluppato in ambiente MATLAB. Successivamente si è operato alla ricostruzione del modello tridimensionale (in MeshLab, per poter fare alcune operazioni di ottimizzazione della mesh di triangolazione della superficie), ed alla sua importazione in un programma di grafica 3D (Blender), per consentire l'applicazione della texture originale alla superficie ricostruita, come si vede nella figura sottostante.



1. Frame estratto dalla sequenza video



2. Costruzione della mesh



3. Ricostruzione in ambiente virtuale con applicazione della texture originale

Figura 25 Ricostruzione 3D da lama laser: test di laboratorio.

4.1.1 Tecniche di riconoscimento 3D

Un output atteso del progetto THESAURUS è la realizzazione di un modello tridimensionale del fondale marino per la rappresentazione e la descrizione della scena esplorata durante una o più missioni. Questo, oltre a fornire una visualizzazione intuitiva dei dati raccolti dagli AUV, permetterà di effettuare ricerche nel 3D per l'identificazione di manufatti.

La ricerca di oggetti 3D permette di analizzare la scena senza il limite imposto dall'osservazione da un unico punto di vista – come invece accade per tecniche di riconoscimento che operano nel 2D. Inoltre l'impiego di caratteristiche geometriche per il riconoscimento è necessario in quanto tecniche *texture-based* sono di scarsa utilità poiché gli oggetti di interesse possono risultare sedimentati e mimetizzati con l'ambiente sottomarino.

Per questo motivo sono state studiate tecniche per il riconoscimento di oggetti tridimensionali e siamo in attesa di dati – che saranno disponibili dopo la prima missione in mare del prototipo di AUV realizzato – sui cui validare tali ipotesi.

Al momento è possibile delineare una pipeline di tre step principali:

1. identificazione del *ground-plane*: il piano principale della scena potrebbe essere identificato andando a fissare sulla nuvola di punti 3D densa un modello planare sia localmente sia globalmente;
2. identificazione di strutture regolari: escludendo i punti appartenenti al *ground-plane* andiamo a valutare la regolarità delle strutture che si staccano dal piano principale. Eseguendo fitting di elementi geometrici elementari (piani e quadriche) si vuole identificare punti 3D che con buona probabilità appartengono a oggetti non naturali (i.e. realizzati dall'uomo);
3. comparazione con modelli 3D di artefatti pre-acquisiti: nella terza fase, le strutture 3D rimanenti sono confrontate con modelli 3D di artefatti come vasellame, anfore, ancore, ecc. collezionati in precedenza e raccolti in uno specifico database.

Mentre per eseguire i primi due passi dovrebbe essere sufficiente eseguire classici metodi di fitting (ad esempio RANSAC vincolato dall'equazione del piano), il terzo step richiede l'utilizzo di tecniche più avanzate.

Gli approcci più interessanti sono quelli basati su *3D detector* [19] e *descriptor* [20], [21] robusti a trasformazioni rigide come traslazioni e rotazioni. In modo analogo a come avviene l'estrazione e la descrizione di punti salienti nel 2D, anche nel 3D è possibile andare a identificare punti di particolare rilevanza e quindi costruire descrittori che permettano l'identificazione, per quanto possibile univoca, del punto e conseguentemente il match fra i modelli 3D.

4.2 Ricostruzione 3D da immagini sonografiche

La mappa sonografica permette di ottenere informazioni sulla batimetria della regione esplorata sfruttando le informazioni sui gradienti di luminosità della mappa stessa. L'esplorazione dettagliata del fondale marino e la sua ricostruzione hanno spinto negli ultimi anni la ricerca in due direzioni fondamentali: da un lato un avanzamento delle tecnologie, con il miglioramento della risoluzione degli strumenti, e con l'impiego di altri sonar come ad esempio quelli ad apertura sintetica, i multibeam e sub-bottom profiler; dall'altro l'elaborazione di nuovi metodi per l'interpretazione dei dati acustici.

Nel caso di Thesaurus, lo strumento acustico in dotazione è un side scan ad alta risoluzione (Tritech Seaking). La maggior parte degli studi esistenti riguardo l'utilizzo di side scan sonar per la ricostruzione del fondale sono orientati a task di individuazione di ostacoli, oppure di classificazione tessiturale o riconoscimento di specifici oggetti-target. Seguendo ricerche più recenti ([22]) ci siamo orientati alla ricostruzione 3D del fondale da immagini sonografiche attraverso algoritmi di *Shape From Shading*: ipotizzando un opportuno modello fisico di diffusione acustica (modello Lambertiano) e applicando tali algoritmi è possibile ottenere una stima della mappa di elevazione. Più in dettaglio, la base teorica si compone di due concetti fondamentali:

- la formazione dell'immagine, seguendo il modello fisico di diffusione di Lambert;
- l'inversione, come da *shape from shading* ottico.

Secondo il modello fisico di insonificazione assunto, l'intensità dell'eco di ritorno relativa al punto p di coordinate $(x,y,Z(x,y))$ dipende da tre parametri:

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

Osserviamo che il passo di inversione, come spesso accade, consiste nell'ottimizzazione di un funzionale energia col metodo dei minimi quadrati, e dunque non assicura (nel caso non convesso) l'esistenza di un minimo globale; nel nostro caso abbiamo un insieme di minimi locali come possibili soluzioni del procedimento di ottimizzazione.

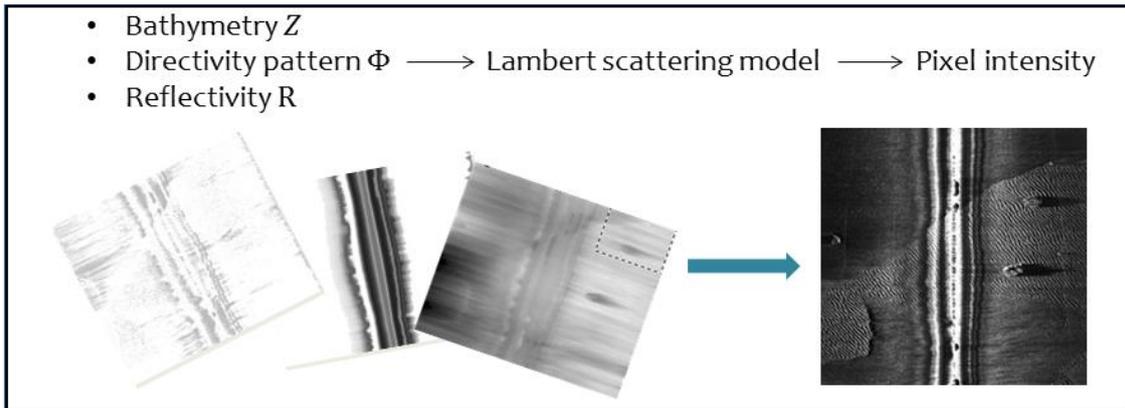


Figura 26 - Formazione dell'immagine acustica

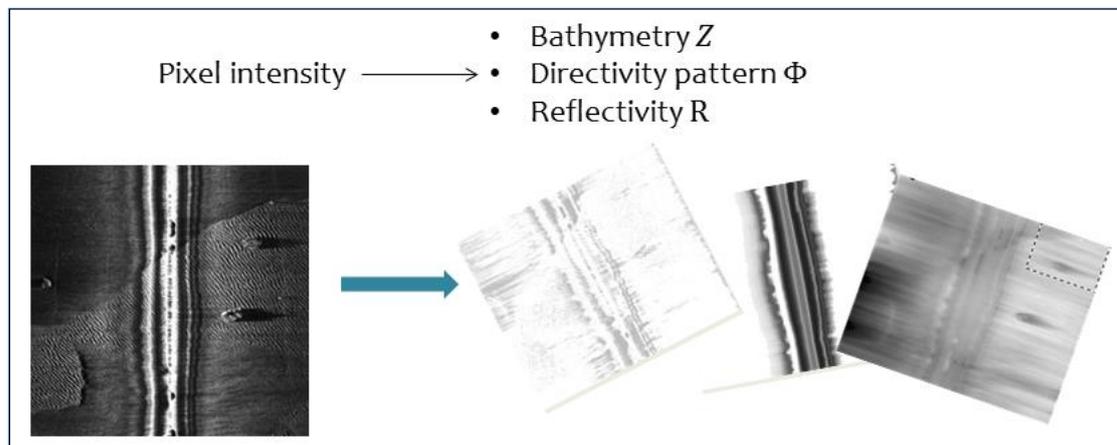
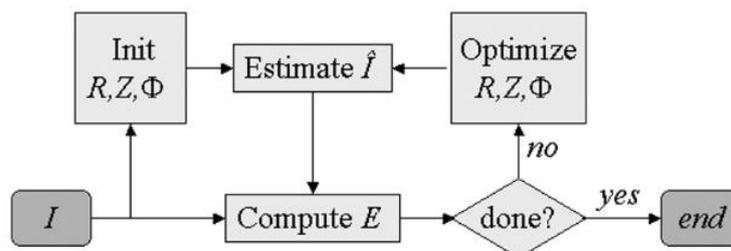


Figura 27 - Operazione di inversione: dall'eco sonar alla superficie stimata

Nella seguente immagine il procedimento di stima dei parametri è visualizzato in uno schema a blocchi.



Fino ad oggi non è stato possibile acquisire dati sonografici in condizioni qualitative sufficientemente soddisfacenti da poter verificare l'attendibilità del procedimento (ad esempio in una regione di cui sia nota la

batimetria del fondale). L'algoritmo di *Shape from Shading* è stato implementato in ambiente MATLAB ed è stato testato in simulazione. In figura un esempio di simulazione.

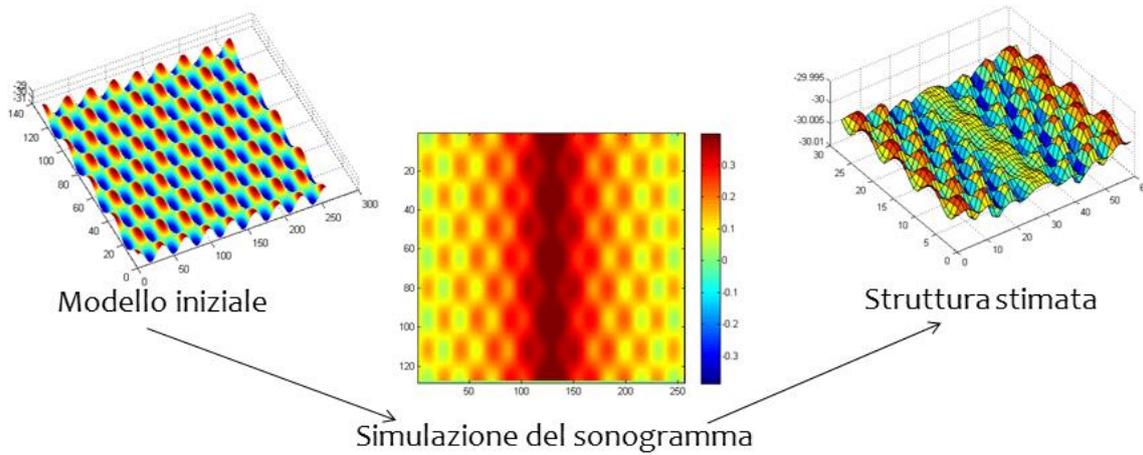


Figura 28 - Applicazione dell'algoritmo Shape From Shading Acustico su immagini simulate

Il formato dei file di output di questi algoritmi contiene le coordinate dei punti nello spazio tridimensionale. Questa tipologia di elaborazione verrà eseguita solo in modalità off-line.

5 Tecniche di screening e rilevamento automatico di oggetti da immagini

Object Detection

La presenza di strutture e pattern geometrici regolari (ad es. circonferenze, linee, etc.) viene quantificata mediante algoritmi di object detection.

Il risultato viene memorizzato in un file in cui si specificano le proprietà geometriche della struttura rilevata (ad es. posizione del centro della circonferenza, raggio, etc.). Le dimensioni tipiche di questi dati sono significativamente inferiori rispetto alle dimensioni dei dati di ingresso.

5.1 Riconoscimento di strutture geometriche bidimensionali

L'alta concentrazione di strutture geometriche bidimensionali sembra essere essenziale nel rilevamento automatico della possibile presenza di manufatti sul fondale marino. Due sono i fattori ambientali che possiamo prevedere ostacolano la performance degli algoritmi implementati: uno è la fisiologica presenza di forme geometriche abbastanza regolari, come pesci e conchiglie, l'altro si riferisce al deterioramento delle superfici degli oggetti di interesse, spesso dovute a concrezioni e depositi di vario genere. Teoricamente il primo produce dei falsi positivi, mentre il secondo ostacola il corretto riconoscimento dei manufatti. Utilizzando alcuni flussi video gentilmente concessi da terzi per il testing degli algoritmi di object detection, abbiamo potuto operare un fine tuning dei parametri di run; purtroppo entrambe le questioni su citate sono non di poco conto, e necessitano di più sperimentazioni sul campo, per verificare l'entità dei falsi prodotti, e studiare dei moduli correttivi ad hoc.

5.1.1 Tecniche di riconoscimento basate sulla trasformata di Hough

Tra le tecniche di riconoscimento di strutture geometriche bidimensionali più classiche annoveriamo l'*Hough Transform*. Purtroppo, sebbene questa tecnica risulti abbastanza efficace e performante nel caso di segmenti, così non è per curve parametriche più complesse. Riportiamo, per completezza, nelle figure seguenti alcuni dei risultati ottenuti applicando ai dati in nostro possesso gli algoritmi derivati dalla trasformata di Hough (vedere ad esempio [23]), ma non riteniamo poterli utilizzare in real-time per supportare l'AUV nel task di object detection.

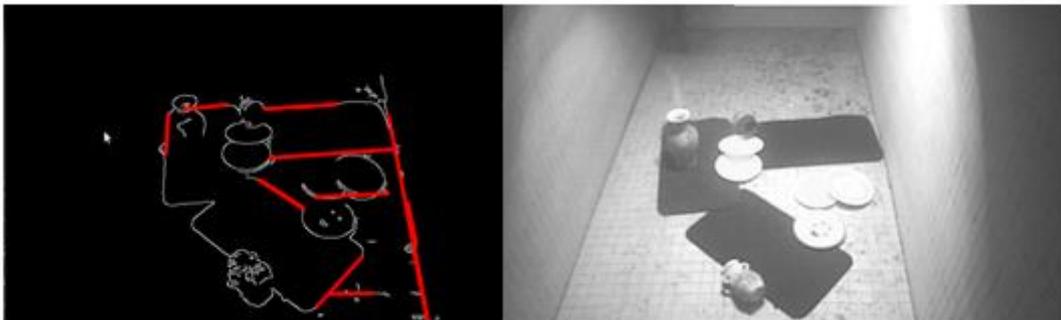


Figura 29 - Algoritmo Hough per il riconoscimento di segmenti



Figura 30 - Algoritmo Hough per il riconoscimento di cerchi

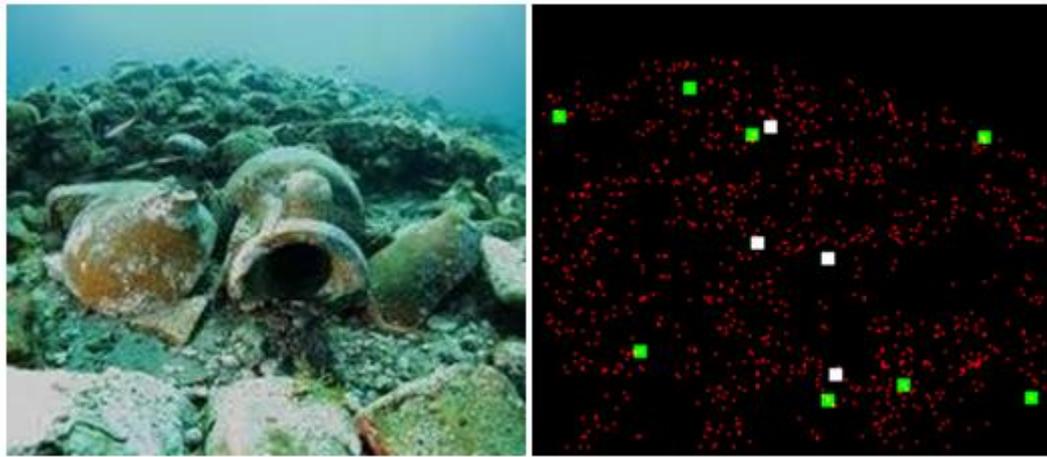


Figura 31 - Algoritmo Hough (randomized) per il riconoscimento di ellissi

5.1.2 Tecniche di riconoscimento basate su metodi *parameterless*

I metodi di fit sono tra i più sfruttati per il riconoscimento e la classificazione di curve regolari o porzioni di esse come segmenti e archi di circonferenze o ellissi. Purtroppo molti non sono abbastanza performanti oppure richiedono il settaggio di molti parametri. L'algoritmo che è stato implementato ad hoc per l'elaborazione di dati subacquei (ELSD) in nostro possesso è descritto in [24]. Tale algoritmo, che si può sinteticamente schematizzare come nella figura sottostante, affonda le radici nella teoria della *Gestalt* applicata alla Computer Vision ([25]). E' interessante evidenziare le caratteristiche che lo rendono un promettente punto di partenza per un sistema completo per la object detection a cui miriamo:

- Riconoscimento di segmenti con accuratezza sub-pixel
- Tempo di esecuzione lineare
- Assenza di parametri da regolare
- Controllo interno del numero dei falsi positivi
- Robustezza rispetto al noise Gaussiano
- Fitting non iterativo per le ellissi.

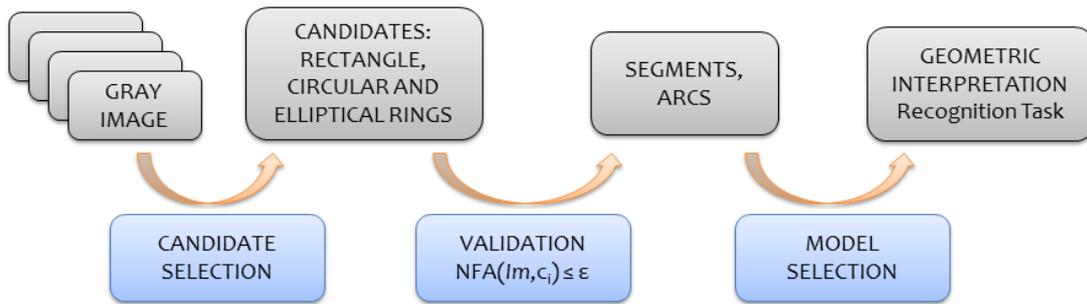


Figura 32 - Schema logico dell'algoritmo ELSD [ELSD]

Qui di seguito lo pseudo codice relativo alle operazioni del primo dei tre blocchi di cui è composto ELSD.

Input: Immagine x , **parametri:** no

Output: L_f – lista delle features valide (segmenti, archi circolari ed ellittici).

$grad \leftarrow$ calcolo di gradiente(x);

foreach pixel p_i in x **do**

$R \leftarrow$ region grow (p_i , $grad$);

$C \leftarrow$ curve grow (R , $grad$);

$line \leftarrow$ fit rettangolo (R);

$circle \leftarrow$ fit corona circolare (C);

$ellipse \leftarrow$ fit corona ellittica (C);

$(NFA_{line}, NFA_{circle}, NFA_{ellipse}) = NFA(line, circle, ellipse)$;

$NFA_{min} \leftarrow \min(NFA_{line}, NFA_{circle}, NFA_{ellipse})$;

if $NFA_{min} \leq 1$ **then**

add feature corrispondente a NFA_{min} in L_f ;

end

Dopo aver applicato l'algoritmo così com'è a sequenze di immagini subacquee relative ai test preliminari che abbiamo condotto, con i risultati di cui è possibile osservare un esempio nelle immagini sottostanti, abbiamo deciso di integrare l'algoritmo in un processo dedicato alla produzione degli alert relativi all'alta probabilità di scoperta di manufatti sul fondale.

Identificativo: Report D3.3

Nome del file: D3point_3_01.doc

Titolo: Studio e sviluppo per l'analisi e l'annotazione delle immagini e delle mappe



Figura 33 - Riconoscimento di curve: test preliminari in vasca a Pistoia

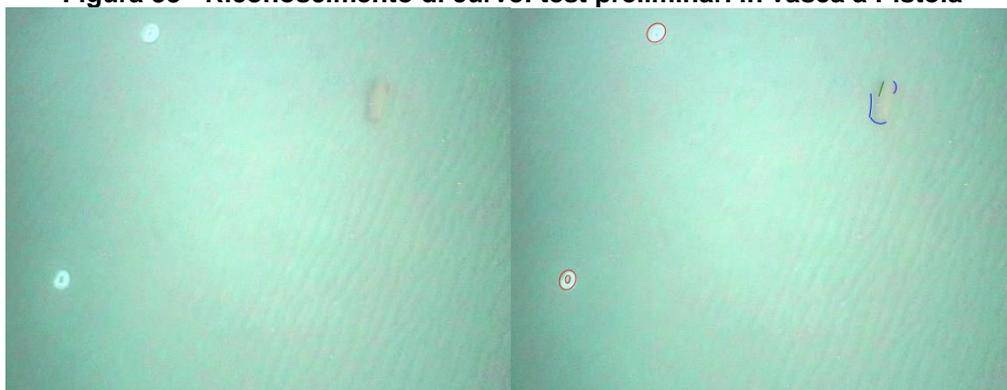


Figura 34 - Riconoscimento di curve: fondale di Marciana Marina, piatti ed anfore

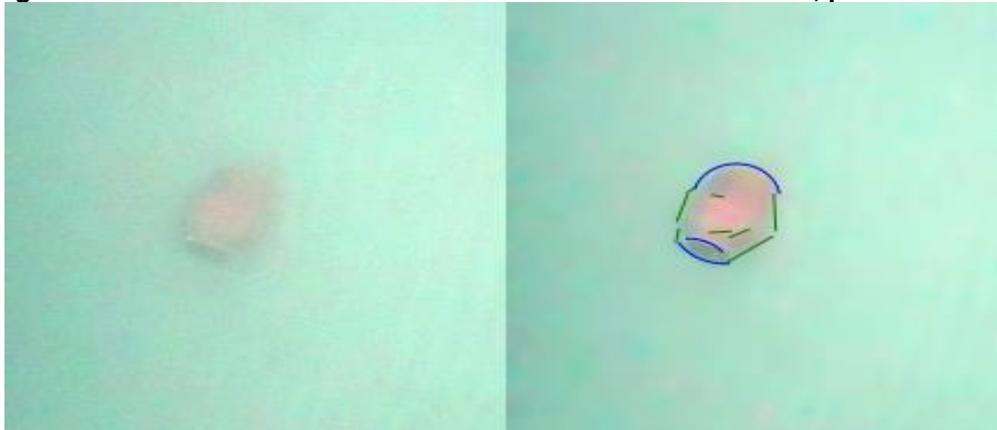


Figura 35 - Riconoscimento di curve: fondale di Marciana Marina, piatti ed anfore

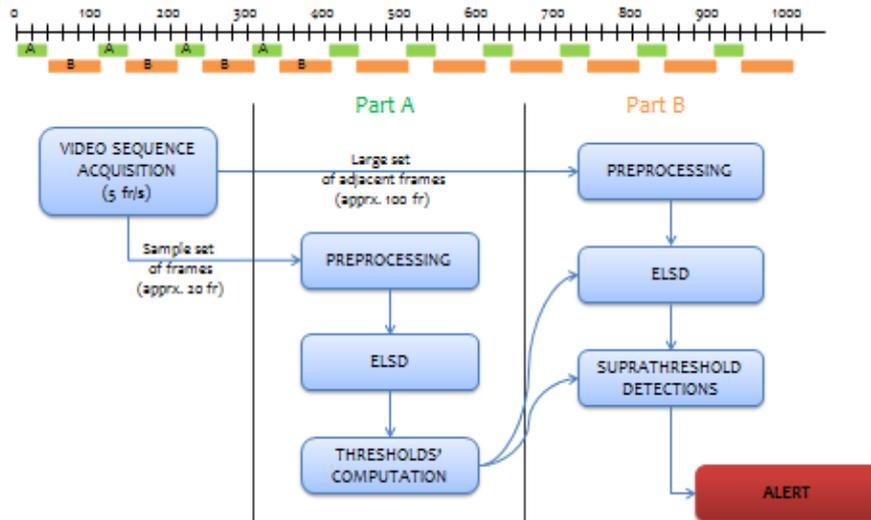


Figura 36 - Schema di elaborazione del flusso ottico ai fini dell'individuazione di variazioni rilevanti nella quantità di curve regolari presenti nella scena

Più in dettaglio, in [26] si può trovare una trattazione più dettagliata sull'applicazione di questo algoritmo per finalità di object detection. L'idea di base è che vada costruita una feature corrispondente alla "ricchezza di curve 2D regolari", considerata un buon indicatore per la presenza di manufatti di interesse. Per estrarre questa feature ci affidiamo attualmente ad una media pesata calcolata sul numero di segmenti ed archi ellittici o circolari. Tale media pesata non possiede una soglia assoluta oltre la quale far scattare l'alert relativo al riconoscimento di manufatti. L'articolo di cui sopra si dedica appunto a questo, proponendo un metodo che possa scandire l'analisi della sequenza video in cicli A-B, in modo rendere la soglia di scoperta relativa ad ogni ciclo rappresentativa di una variazione sostanziale del numero di curve regolari presenti in un dato ambiente. La Figura 36 riporta uno schema riassuntivo del suddetto metodo.

Tale metodo è già stato applicato sia a video acquisiti da noi, sia ad una sequenza video (cortesia di un collaboratore), ed i grafici relativi alla variazione della soglia di scoperta in rapporto alla scoperta (valutata dall'operatore) è riportato qui di seguito.

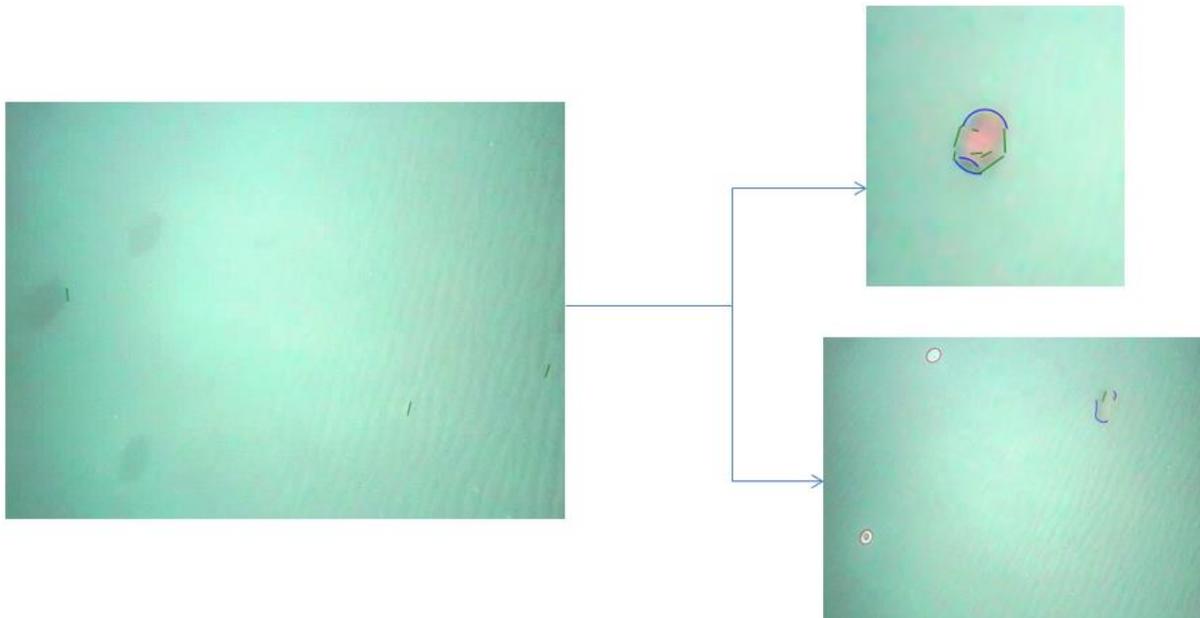


Figura 37 - Frame estratti dalla sequenza di Marciana Marina, con evidenziati alcuni in cui sono evidenti detection interessanti

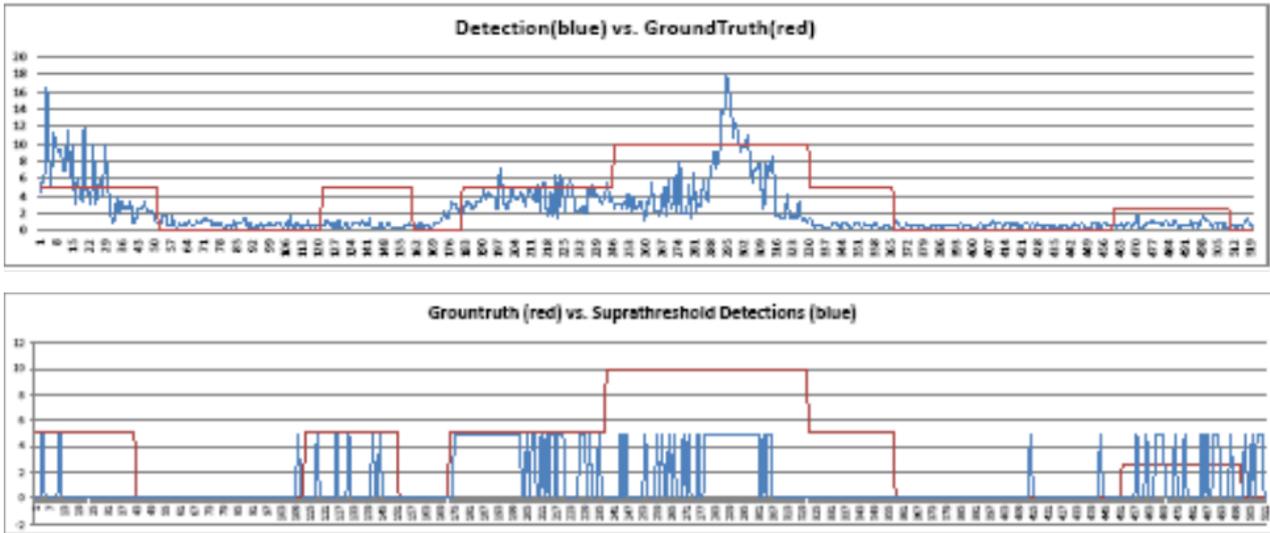


Figura 38 - Grafici relativi all'algorithmo di detection applicato alla sequenza di Marciana Marina

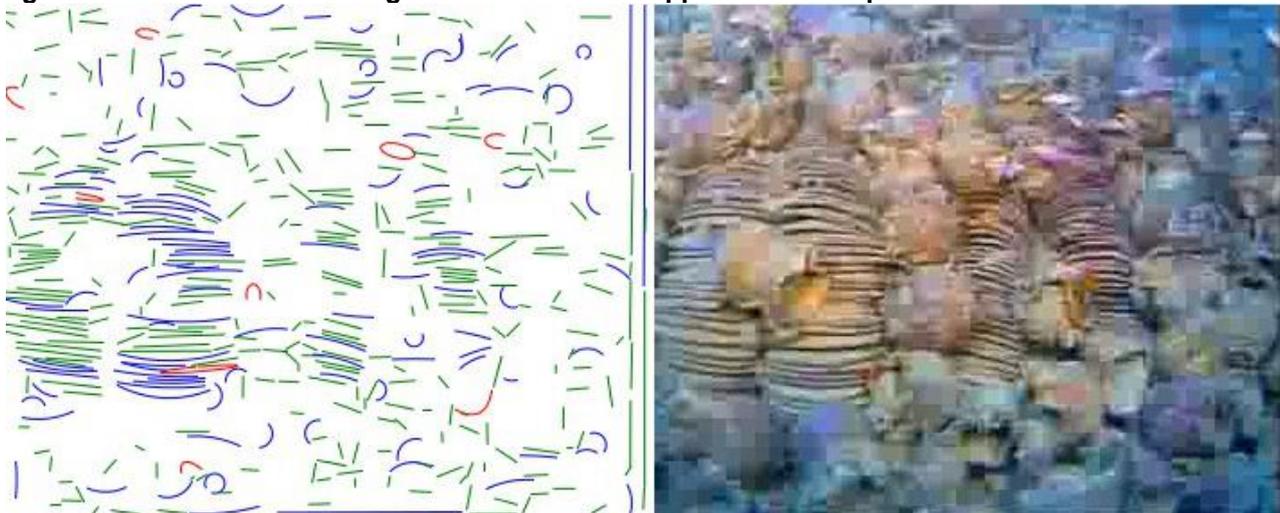


Figura 39 - Riconoscimento di curve (ellissi, circonferenze, segmenti) relativo ad una sequenza di un sito archeologico molto ricco di manufatti.

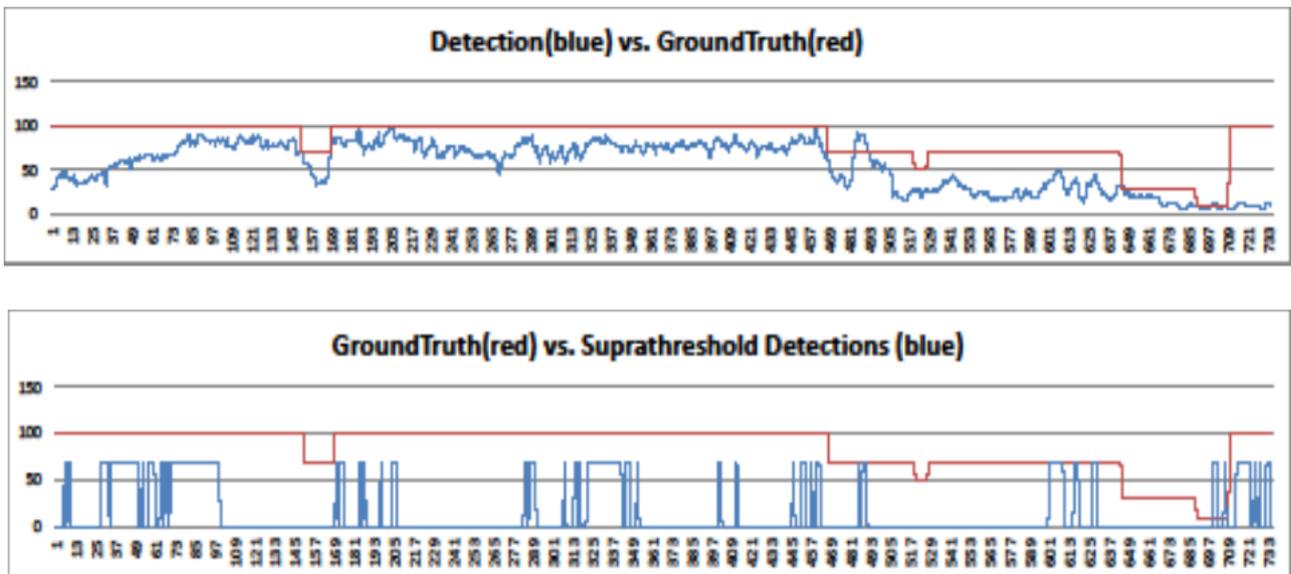


Figura 40 - Grafici relativi all'algorithmo di detection applicato alla sequenza del sito archeologico

Reputiamo che una versione dell'algoritmo ELSD possa produrre buoni risultati anche nell'applicazione ai dati acustici. Ovviamente sarà possibile elaborare le modifiche da apportare (si pensa ad un diverso setting dei parametri interni e ad un filtering basato sulla lunghezza delle curve riconosciute) solo dopo aver raccolto una serie di dati acustici (da side scan sonar) che siano rappresentativi rispetto ad un setting di scoperta.

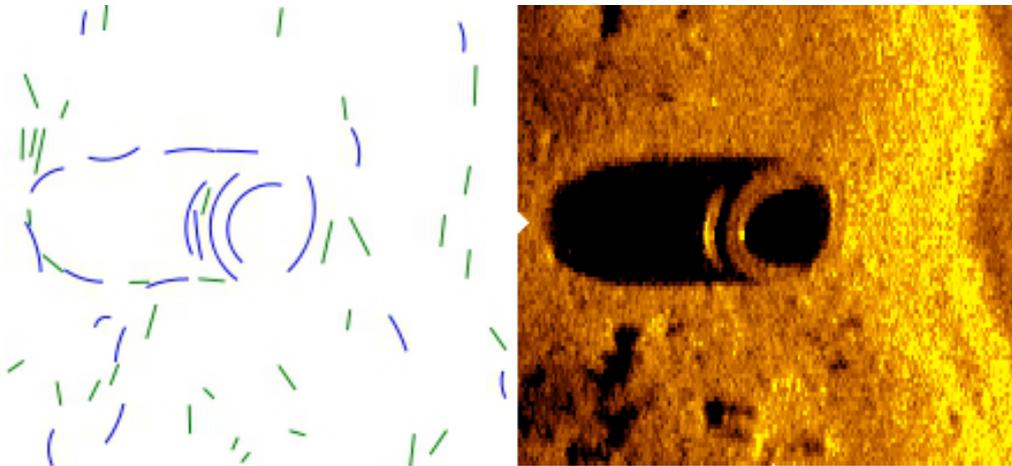


Figura 41 - Riconoscimento di curve applicato ad un sonogramma raffigurante uno pneumatico

5.2 Algoritmi di intelligenza artificiale per la classificazione di tessiture

Il concetto di tessitura si riferisce a pattern visuali che hanno particolari proprietà di omogeneità derivanti dall'alternarsi regolare di più colori o d'intensità all'interno di un'immagine. La tessitura cattura importanti informazioni sulla struttura delle superfici e sulla loro relazione con l'ambiente circostante. L'analisi quantitativa della tessitura può essere utilizzata come guida nel processo di segmentazione e di classificazione delle immagini.

A questo fine sono stati introdotti svariati descrittori della tessitura, ossia degli insiemi di metriche calcolate mediante strumenti di elaborazione delle immagini per quantificare la percezione della tessitura in un'immagine o in una regione selezionata della stessa. In particolare, un descrittore della tessitura deve fornire informazioni sulla disposizione spaziale di colore o intensità.

In base ai valori calcolati di un insieme di descrittori di tessitura è possibile discriminare regioni distinte delle immagini. Il risultato della discriminazione può essere utile ai fini della segmentazione in regioni omogenee per tessitura oppure al rilevamento e alla classificazione automatica delle regioni in cui si riscontrano dei pattern di particolare interesse.

Nell'ambito dell'esplorazione sottomarina, l'analisi tessiturale dei fondali è utile per catalogare i fondali in varie categorie (sabbia, fango, roccia, praterie di posidonia, ...) e a valutare la presenza di zone con tessitura anomala di potenziale interesse. Questo tipo di analisi può essere applicato ai flussi video ottici, ai sonogrammi o a entrambe le tipologie di dati congiuntamente per il loro screening automatico.

5.2.1 Metodo dei filtri di Gabor

Tra i vari descrittori di tessitura disponibili in letteratura, sono stati testati metodi basati sull'utilizzo dei filtri di Gabor. Tali filtri hanno ricevuto una notevole attenzione nel campo della visione artificiale poiché è stato dimostrato che mimano il comportamento di alcune cellule della corteccia visiva dei mammiferi. In questo senso, i filtri di Gabor rappresentano un metodo ispirato dalla visione biologica. Inoltre i filtri di Gabor hanno dimostrato di possedere buone proprietà di localizzazione sia nel dominio spaziale sia in quello delle frequenze. Per questa ragione sono adatti a problemi di segmentazione di strutture. Tra le varie applicazioni dei filtri di Gabor citiamo la segmentazione di strutture complesse, il rilevamento di target, l'analisi OCR dei documenti, il rilevamento dei bordi e l'identificazione della retina.

Informalmente un filtro Gabor può essere visto come un'onda sinusoidale di frequenza e orientamento specifici, modulata da una Gaussiana (vedi Figura 42 e Figura 43). Convolvere il filtro con l'immagine da analizzare si ottiene la risposta del filtro in cui risultano enfatizzati i luoghi dell'immagine aventi a livello locale componenti con quella frequenza e orientamento specifici. Convolvere con un insieme di filtri aventi diverse frequenze e orientamenti, si ottiene un insieme di risposte che consentono di classificare complessivamente la tessitura punto per punto.

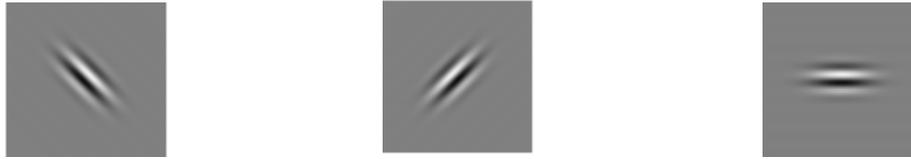


Figura 42 - Filtri di Gabor con 3 orientazioni distinte

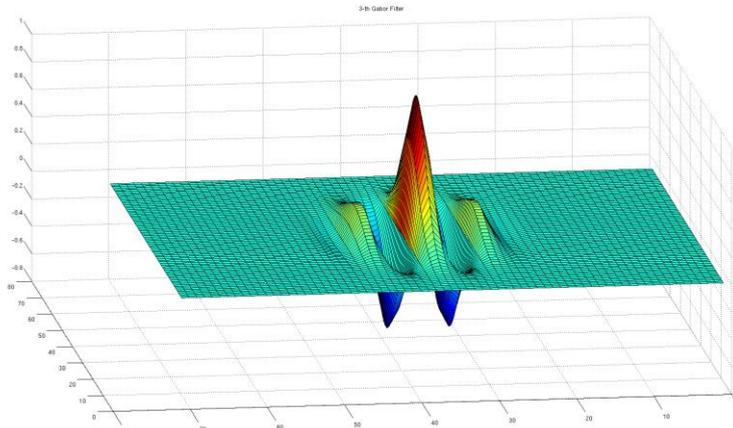


Figura 43 - Plot 3D di un filtro di Gabor

Riportiamo una porzione di pseudocodice per il calcolo della risposta di un'immagine e la segmentazione in k regioni con caratteristiche omogenee:

Input: Immagine x ,

Parametri: Numero di regioni da estrarre K ; Orientazione $\alpha_1, \dots, \alpha_p$;

Frequenze: η_1, \dots, η_q ; Deviazione standard del kernel Gaussiano σ

Output: L — immagine indicizzata con K valori (stessa dimensione di x)

foreach orientation α_i , frequency η_j **do**

$K \leftarrow \text{computeGaborFilter}(\alpha_i, \eta_j, \sigma)$;

$x(i,j) \leftarrow \text{convolve}(K, x)$; //This is the response of the filter

end

$\text{clusterCentroids} = \text{clusterize}(\{x(i,j)\}_{i,j}, K)$ // here we clusterize the $(s \cdot q)$ -dimensional vector of pixel
// responses into K clusters and we store their centroids.

foreach pix in x **do**

$\text{label} \leftarrow \text{labelPixelWithCluster}(\text{pix}, \text{clusterCentroids})$; //each pixel belongs to a cluster

$L(\text{pix}) \leftarrow \text{label}$

end

In Figura 44 si mostra l'applicazione dell'algoritmo descritto nello pseudocodice alla segmentazione di un sonogramma per evidenziare l'area dominata dai ripple.

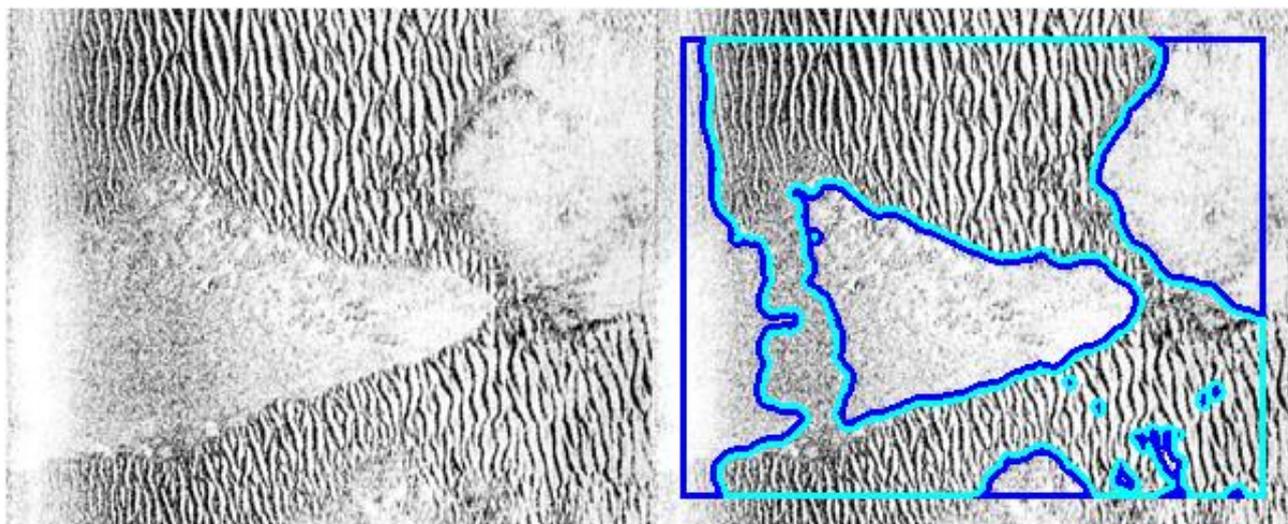


Figura 44 - Segmentazione applicata ad un sonogramma: evidenziata l'area dominata dai ripple.

6 Annotazione semantica della scena e archiviazione

La piattaforma di acquisizione dati di Thesaurus consente la raccolta di una mole di dati eterogenei. D'altro canto i metodi descritti nei capitoli precedenti generano dati secondari (quali modelli tridimensionali degli oggetti sul fondale) e metadati contenenti informazioni sulla scena osservata. Entrambe le tipologie di dati devono essere opportunamente organizzate e archiviate in un repository facendo uso dei relativi metadati per consentire il recupero e la fruizione ottimale di tutte le informazioni contenute.

6.1 *Contenuti raccolti e generati*

In Thesaurus, sono coinvolti i) i dati grezzi raccolti dai sensori durante la navigazione (sensor footage), i dati elaborati consistenti in ii) mappe di insieme e iii) modelli tridimensionali e i relativi metadati associati.

6.1.1 *Sensor footage*

I dati acquisiti durante le missioni sono direttamente archiviati in un disco presente all'interno di ogni AUV. Una volta terminata la missione, i relativi dati sono resi disponibili al sistema informativo tramite l'utilizzo di servizi dedicati. I servizi si occupano, quindi, della ricezione del dato, ove necessario della sua interpretazione e del suo inserimento in un'apposita struttura atta alla gestione ed alla fruizione del dato stesso.

6.1.2 *Mappe di insieme e loro metadati*

Elaborando il sensor footage ed estraendo i dati di navigazione sono create mappe di insieme ottenute fondendo e georeferenziando i dati raccolti durante una missione sottomarina, mediante script ad hoc basati sulle tecniche descritte nella Sezione 3.

Le immagini ottiche e sonografiche georeferenziate e i relativi mosaici possono essere così inseriti in apposite strutture. In particolare, per ciascuna mappa di insieme sono registrate le coordinate della zona coperta, il tempo di acquisizione, gli identificativi (URI) delle immagini coinvolte e la loro georeferenziazione.

6.1.3 *Modelli 3D di Oggetti*

Il sensor footage unito ai dati di navigazione è inoltre ulteriormente processato per generare modelli 3D a scale e risoluzioni diverse. Anche in questo caso si utilizzano script ad hoc, contenenti chiamate agli algoritmi di visione artificiale descritti nella sezione 4, per generare:

- Modelli 3D da dati sonar, mediante tecniche di shape form shading;
- Modelli 3D di elevazione del fondale mediante tecniche di structure from motion da dati ottici;
- Modelli 3D di oggetti di interesse singoli, mediante successioni di immagini con lama laser.

6.2 *Annotazione semantica dei contenuti*

I metadati utilizzati per la descrizione delle immagini, delle mappe e dei modelli precedentemente descritti possono essere integrati da ulteriori metadati contenenti informazioni sulla semantica della scena osservata. In particolare, dopo la prima fase di creazione, le mappe di insieme sono arricchite con opportuni tag, sempre georeferenziati, indicanti le aree in cui sono stati rilevati oggetti di interesse mediante gli algoritmi descritti nella sezione 5. Tali tag possono essere inoltre confermati, modificati e arricchiti tramite procedure assistite. Anche in questo caso sono utilizzati script ad hoc, basati su chiamate agli algoritmi di visione, per accedere alle immagini, generare i tag e modificare i metadati delle mappe di insieme.

6.3 *Repository per l'archiviazione dei contenuti*

Per la gestione dei dati in Thesaurus si è scelto di sviluppare un repository in grado di i) trattare i dati indipendentemente dal loro contenuto e dalla loro forma, permettendo quindi una grande libertà nell'organizzazione e archiviazione di dati eterogenei, ii) fornire un ampio supporto per i metadati e iii) gestire informazioni geografiche nel rispetto degli standard dell'Open Geospatial Consortium per i sistemi informativi

territoriali. Coerentemente con le scelte effettuate nell'implementazione del database storico-archivistico, si è quindi scelto di realizzare un repository ibrido basato sul database relazionale postGIS (l'estensione di PostgreSQL ai dati spaziali) e sul database XML nativo eXist. Il primo viene sfruttato per gestire i sistemi di riferimento e i dati spaziali relativi alla geolocalizzazione dei dati multimediali, mentre le potenzialità del secondo vengono utilizzate per il mantenimento dei metadati e la gestione delle annotazioni semantiche, effettuate con la medesima terminologia utilizzata nel database storico-archivistico.

Complessivamente il repository, come mostrato in **Error! Reference source not found.**, mette in relazione fra loro la parte di database (relazionale +XML) ed il file system su cui i documenti (immagini, modelli 3d,...) sono salvati fisicamente.

Infatti, durante la fase di archiviazione di un documento (immagine, flusso video, modello 3D,...) viene creato un metadato XML contenente informazioni inerenti al documento. Il metadato viene archiviato nel database XML mentre il documento in sé è inserito nel file system in una apposita cartella. Inoltre il metadato XML contiene uno o più riferimenti a informazioni geografiche inserite in postGIS.

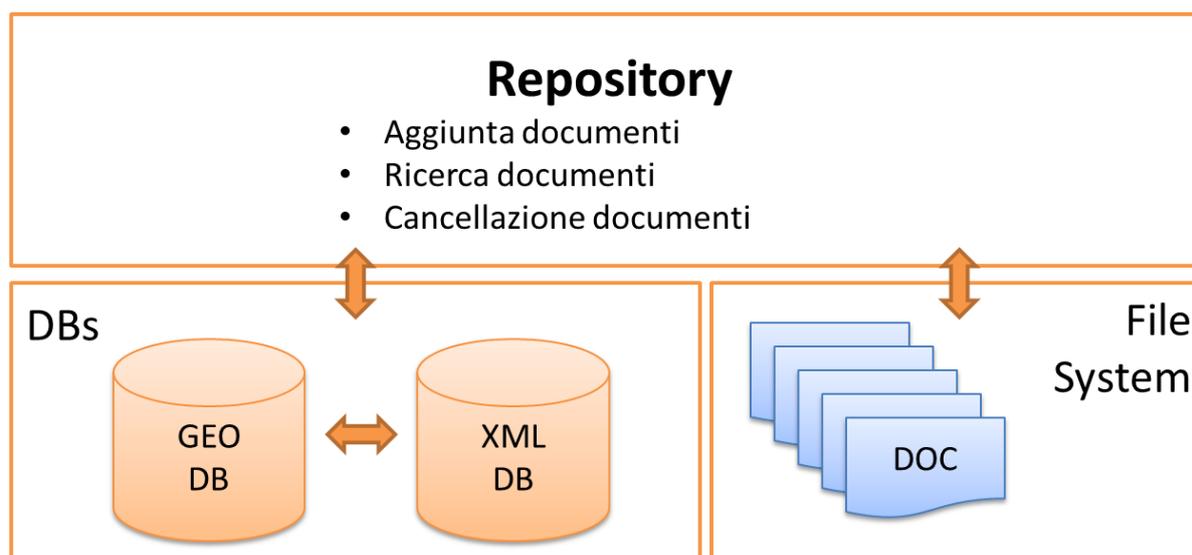


Figura 45 - Architettura del repository

Il repository sfrutta la collaborazione fra postGIS, il database XML, il file system e l'unità di controllo, scritta in linguaggio Java, che si occupa di:

- gestire le comunicazioni con i database;
- gestire il file system (sia in scrittura per l'upload che in lettura per la ricerca);
- creare dinamicamente le query;
- creare i documenti XML relativi ai dati inseriti nel sistema e i record nelle opportune tabelle postGIS.

Sfruttando questa architettura è possibile gestire genericamente qualsiasi tipo di dato indifferentemente dal suo contenuto e dalla sua struttura. All'upload di un nuovo documento, infatti, viene associato ad esso un metadato, contenente le informazioni immesse dall'utente compilando i vari campi durante la fase di upload, che verrà usato per ritrovare il documento in un secondo momento. Le ricerche all'interno del repository infatti non coinvolgeranno i documenti inviati ma solo i loro metadati; il risultato di ogni ricerca sarà la lista dei path per recuperare i documenti trovati.

Durante le fasi di screening dei documenti archiviati, gli algoritmi di visione artificiale genereranno ulteriori metadati contenenti informazioni sulla semantica della scena (tipo fondale, coordinate di oggetti di interesse, etc). Tali nuovi metadati andranno a complementare i metadati già presenti e diventeranno immediatamente sfruttabili per query al repository. Un esempio di metadato associato ad un sonogramma è il seguente:

```
<sonogramma>
  <path>/Repository/Sonogramma/AUV_rawdata_20130126_20130126.jpg</path>
h>
  <filename>AUV1_rawdata_20130126_20130126.jpg</filename>
  <originalname>sonogramma.jpg</originalname>
  <utente>cnr</utente>
  <sorgente>AUV1</sorgente>
  <datatype>rawdata</datatype>
  <date>
```

Identificativo: Report D3.3

Nome del file: D3point_3_01.doc

Titolo: Studio e sviluppo per l'analisi e l'annotazione delle immagini e delle mappe

```
<inizio>20130126</inizio>
<fine>20130126</fine>
</date>
<coordinate>
  <geo_id>8</geo_id>
</coordinate>
<step>2m</step>
<zone_di_interesse>
  <zona_di_interesse>
    <annotazione>relitto</annotazione>
    <zona_geo_id>15</zona_geo_id>
  </zona_di_interesse>
</zone_di_interesse>
</sonogramma>
```

Il documento XML è sufficientemente human-readable. Le parti in color aqua contengono i metadati inseriti al momento dell'upload del documento e includono la sorgente del dato e informazioni per la sua geolocalizzazione in un sistema informativo territoriale. In particolare non vengono fornite direttamente le coordinate geografiche in formato testo, ma all'interno del tag `geo_id` viene riportata una chiave per il recupero dell'informazione spaziale da postGIS. Inoltre è riportato il path al file fisico contenente il sonogramma in formato jpeg. È stata evidenziata col colore orange la porzione delle file contenente un'informazione ottenuta mediante annotazione assistita dell'immagine. In questo caso il tag semantico "relitto" è associato al luogo geografico in cui è stato riscontrato. Anche in questo caso l'informazione geografica è solamente riferita mediante una chiave ad un record di postGIS.

7 Integrazione di dati da più sorgenti

In generale per fusione si intende un metodo di ottimizzazione di informazioni eterogenee o omogenee derivante da dati complementari o ridondanti. Esistono vari livelli di fusione possibili [27]: dato, feature, decisionale; a questi corrispondono diversi livelli di rappresentazione: segnale/pixel, feature, simbolo. La fusione a livello di segnale/pixel è la combinazione dei dati raw acquisiti da sorgenti diverse in una sola immagine, mentre a livello più alto occorre prima effettuare il passaggio di estrazione delle diverse feature, o registrare l'output dei processi decisionali, relativi al dato di ogni singola sorgente, e poi integrarli opportunamente.

7.1 Tecniche per l'estrazione di features da mappe multi-sensoriali

I metodi di fusione più comunemente utilizzati si differenziano nella collocazione temporale attribuita al processo di fusione (e cioè prima o dopo l'aver effettuato l'estrazione di feature rilevanti). A causa del fatto che l'ambiente sottomarino consente un'estrazione poco robusta delle feature relative ai dati ottici (che sono tra i più significativi, rispetto ai nostri obiettivi), ci è sembrato opportuno progettare un modello di fusione che fosse in grado di contenere tutta la complessità dei dati raccolti, ma non solo. Ciò significa effettuare lo step di fusione a tutti i livelli, mettendo i dati ottici e acustici (che si suppongono registrati correttamente) insieme con tutte le informazioni estratte tramite l'analisi di texture, regolarità di pattern geometrici e regolarità di curve e superfici, opportunamente georeferenziate. Il senso generale di questo modello non è comprensibile se non si specifica operativamente come si effettua la fusione tra tutti i livelli. Operativamente assegniamo ad ogni pixel p dell'ortomosaico della regione esplorata un vettore $v(p)$ i cui valori sono raggruppati (e per ora limitati) come segue:

- INTENSITA': valori RGB, e intensità dell'eco acustica;
- ALTITUDINE: elevazione (coordinata z) come stimata tramite lo *shape from shading* acustico, e tramite la *depth map* relativa al sistema di visione stereoscopico;
- SUPERFICIE: classificazione delle texture, sia nell'ottico, sia nell'acustico;
- GEOMETRIA: appartenenza (eventualmente contata con molteplicità) a feature o pattern di tipo geometrico.

Una volta che i dati relativi alla scena esplorata vengono così rappresentati è possibile operare specifici algoritmi per trarre il massimo dell'informazione utile. A titolo di esempio:

- i) Comparazione tra i livelli ottico ed acustico per riconoscere elementi in comune;
- ii) Selezione di alcune componenti di uno specifico $v(p)$ al fine di individuare nella mappa tutti gli altri punti che mostrano un alto grado di similarità con esso.



Figura 46 - Schema di rappresentazione dei dati nel modello di fusione

Identificativo: Report D3.3

Nome del file: D3point_3_01.doc

Titolo: Studio e sviluppo per l'analisi e l'annotazione delle immagini e delle mappe

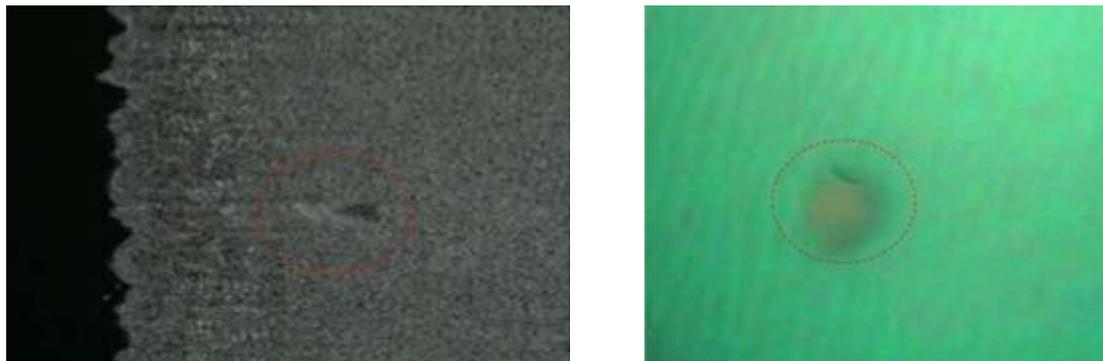


Figura 47 - Detection rafforzata dalla concordanza di ottico e acustico

8 Bibliografia

- [1] R. E. a. H. S. O. Pizarro, «Large area 3D reconstructions from underwater optical surveys,» *IEEE Journal of Oceanic Engineering*, vol. 34(2), pp. 150-169, 2009.
- [2] A. .. B. a. G. O. F. Bonin, «Imaging systems for advanced underwater vehicles,» *Journal of maritime research*, vol. VII(1), pp. 65-86, 2011.
- [3] C. R. L. W. a. D. Y. H. Singh, «Advances in fusion of high resolution underwater optical and acoustic data,» in *Proc. IEEE Int. Symp. Underwater Technology*, 2000.
- [4] S. W. a. I. Mahon, «Simultaneous localization and mapping on the great barrier reef,» in *IEEE Proceedings Int. Conf. on Robotics and Automation*, 2004.
- [5] A. F. a. V. Murino, «Augmented scene modeling and visualization by optical and acoustic sensor integration,» *IEEE Transactions on Visualization and computer Graphics*, vol. 10(6), pp. 625-635, 2004.
- [6] H. S. a. H. P. S. Negahdaripour, «Optiacoustic stereo imaging, system calibration and 3D reconstruction,» in *Proc. IEEE Beyond Multi-View Geometry*, 2007.
- [7] Q. Z. a. R. Pless, «Extrinsic calibration of a camera and laser range finder,» in *IEEE/RSJ Proceedings Int. Conf. on Intelligent Robots and Systems*, 2004.
- [8] H. D. L., *Mathematicl Techniques in Multisensor Data Fusion*, Artech House, 1992.
- [9] B. Dasarthy, «Sensor fusion potential exploitation - innovative architectures and illustrative applications,» in *Proceddings of the IEEE*, 1997.
- [10] P. Varshney, «Multisensor data fusion,» in *Lecture Notes in Computer Science*, Springer Berlin / Heidelberg, 2000, pp. 103-186.
- [11] A. B. E. C. I. T. R. C. S. a. Y. P. C.G. Capus, «Data correction for visualization and classification of sidescan SONAR imagery,» *IET Radar, Sonar and Navigation*, vol. 2(3), pp. 155-169, 2008.
- [12] Q. Zhang, «Flexible Camera Calibration by Viewing a Plane from Unknown Orientations,» in *The Proceedings of the Seventh IEEE International Conference on Computer Vision*, 1999.
- [13] J. S. a. X. T. Kaiming He, «Single image haze removal using dark channel prior,» *IEEE Transaction on Pattern Analysis and Machine Intelligence*, vol. 33 (12), 2011.
- [14] I. Q. L. J. a. J. M. S. Bazeille, «Automatic underwater image pre-processing,» in *Proceedings of the SEA TECH WEEK, Caracterisation du milieu marin*, 2006.
- [15] R. Fattal, «Single image dehazing,» in *ACM Transactions on Graphics*, 2008.
- [16] D. Lowe, «Distinctive image features from scale-invariant keypoints,» *Internation journal of Computer Vision*, vol. 60(2), pp. 91-100, 2004.
- [17] M. F. a. R. Bolles, «Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography,» *Communications of ACM*, vol. 24(6), pp. 381-395, 1981.
- [18] C. H. a. M. Stephens, «A combined corner and edge detection,» in *Proc. of the Fourth Alvey Vision Conference*, 1988.
- [19] S. S., T. F. and D. S. L., "A Performance Evaluation of 3D Keypoint Detectors," in *Proceedings of 3DIMPVT '11, International Conference on 3D Imaging, Modeling, Processing, Visualization and Transmission*, Washignton DC, 2011.
- [20] A. Johnson, "Spin Images: A Representation for 3-D Surface Matching," Robotics Institute, Carnegie Mellon University, Pittsburgh, 1997.
- [21] D. H. R. K. T. B. a. J. M. A. Frome, «Recognizing Objects in Range Data Using Regional Point Descriptors,» in *European Conference on Computer Vision*, Prague, 2004.
- [22] Y. R. P. a. D. M. L. E. Coiras, «Multiresolution 3D reconstruction from side-scan sonar images,» *IEEE Transactions on Image Processing*, vol. 16 (2), pp. 382-390, 2007.
- [23] M. T. a. R. B. C. Basca, «Randomized Hough Transform for ellipse detection with result clustering,» in *Proceedings of the EUROCON 2005*, 2005.
- [24] P. G. a. R. G. v. G. V. Patraucean, «A parameterless line segment and elliptical arc detector with enhanced ellipse fitting,» in *ECCV 2012 - Lecture notes in computer science*, 2012.
- [25] L. M. a. J. M. M. A. Desolneux, *From Gestalt Theory to Image Analysis: A Probabilistic Approach*, Springer, 2008.

Identificativo: Report D3.3

Nome del file: D3point_3_01.doc

Titolo: Studio e sviluppo per l'analisi e l'annotazione delle immagini e delle mappe

- [26] M. A. P. M. R. a. O. S. D. Moroni, «Underwater scene understanding by optical and acoustic data integration,» *Proceedings Of Meetings on Acoustics*, vol. 17, 2013.
- [27] D. Ballard, «Generalizing the Hough Transform to detect arbitrary shapes,» *Pattern Recognition*, vol. 13(2), pp. 111-122, 1981.
- [28] T. N. a. B. Kerautret, «Ellipse detection through decomposition of circular arcs and line segments,» in *Internal Conference ICIAP '11*, Ravenna, 2011.
- [29] A. S. R. W. P. H. a. P. B.-C. J. Esteban, «A review of data fusion models and architectures: towards engineering guidelines,» *Journal of Neural Computing and Applications*, vol. 14 (4), pp. 273-281, 2005.