



UNIVERSITÀ DEGLI STUDI DI FIRENZE

Facoltà di Ingegneria

Corso di Laurea in
INGEGNERIA INFORMATICA

SEGMENTAZIONE AUTOMATICA PER RICOSTRUZIONE 3D DI UN SOLIDO DI RIVOLUZIONE DA SINGOLA IMMAGINE NON CALIBRATA

Tesi di Laurea di
DARIO COMANDUCCI

Relatori:

Prof. Alberto Del Bimbo

Prof. Carlo Colombo

Correlatore:

Ing. Federico Pernici

Anno Accademico 2002/2003

Alla mia famiglia

RINGRAZIAMENTI

DESIDERO RINGRAZIARE TUTTI coloro che mi hanno seguito, supportato e sopportato durante lo sviluppo della tesi; un particolare ringraziamento va quindi ai professori Alberto Del Bimbo e Carlo Colombo, ed al loro collaboratore, Ing. Federico Pernici.

D. C.

INDICE

SOMMARIO	vii
1 COMPUTER VISION ED ORGANIZZAZIONE PERCETTIVA	1
1.1 L'organizzazione percettiva	1
1.2 La segmentazione	4
1.3 Segmentazione tramite clustering	6
1.3.1 Clustering: tecniche generali	6
1.3.2 Morfologia computazionale	10
1.4 Segmentazione per mezzo di modelli	14
1.4.1 Trattazione generale	15
1.4.2 Scelta dei residui e delle funzioni di costo	15
1.4.3 Stimatori statistici	18
1.4.4 Stimatori per la computer vision	22
2 RICONOSCERE E RICOSTRUIRE UN SOLIDO DI RIVOLU-	
ZIONE	25
2.1 Terminologia di base	25
2.2 Lavori precedenti	26
2.3 Visione d'insieme dell'algoritmo	29
2.4 Proprietà proiettive dei solidi di rivoluzione	29
2.4.1 Proprietà generali di un'omologia	29
2.4.2 Omologie nei SOR	31
2.5 Stima dell'omologia armonica e segmentazione	33
2.5.1 Inizializzazione con RANSAC	34
2.5.2 Raffinamento dell'omologia	36
2.6 Classificazione delle curve	44
2.6.1 Parametrizzazione ed estrazione delle ellissi	44
2.6.2 Raffinamento delle ellissi	50
2.6.3 Selezione e classificazione	52
2.7 Calibrazione interna e ricostruzione metrica	55
2.7.1 Calibrazione interna	56
2.7.2 Ricostruzione metrica del SOR	58
2.8 Calibrazione esterna e acquisizione della texture	62
2.8.1 Formazione dell'immagine di un SOR	62
2.8.2 Calibrazione esterna	64
2.8.3 Acquisizione della texture del SOR	68
3 ESPERIMENTI	73
3.1 Esperimenti con immagini sintetiche	73
3.1.1 Analisi dei risultati sintetici	83
3.2 Esperimenti con immagini reali	84
4 CONCLUSIONI	93
A LE COORDINATE OMOCENEED IL MODELLO "NATURAL	
CAMERA"	95
A.1 Proiezioni centrali e coordinate omogenee	95

A.1.1	Rette di un piano in coordinate omogenee	96
A.1.2	Coniche in coordinate omogenee	98
A.1.3	Trasformazioni proiettive	99
A.1.4	Punti circolari	99
A.2	Il modello “natural camera”	100
A.3	Cenni alla calibrazione	101
Bibliografia		105

SOMMARIO

L'INTERESSE SEMPRE CRESCENTE per i database di modelli tridimensionali, così come quello per la ricostruzione grafica di ambienti reali, ha rinnovato l'interesse della ricerca nel campo dell'acquisizione della forma 3D di oggetti a partire da una o più fotografie. In molti casi, queste possono essere l'unico punto di appoggio, qualora l'oggetto non sia più disponibile oppure abbia caratteristiche mal conciliabili (materiali inadatti, dimensioni eccessive) con le classiche tecniche di acquisizione 3D, come lo scanner laser. Per una ricostruzione fedele (*metrica*) dell'oggetto è necessario conoscere i parametri di fotocamera che governano la proiezione della scena sull'immagine. Tuttavia, in genere essi non sono noti; assume quindi un ruolo essenziale l'*autocalibrazione* che, a partire da vincoli geometrici noti e relativi a particolari entità rintracciabili nella foto permette la stima dei parametri di fotocamera.

I solidi di rivoluzione (indicati anche con la sigla *SOR*, da Surface Of Revolution) costituiscono una famiglia di oggetti molto diffusi negli ambienti creati dall'uomo. Le proprietà di simmetria di un solido di rivoluzione sono tali da produrre nell'immagine sufficienti vincoli geometrici per la calibrazione ed il recupero della forma del SOR. In letteratura sono documentati dei metodi per la calibrazione e ricostruzione di SOR; tuttavia, i metodi sviluppati richiedono l'intervento umano, tramite l'indicazione manuale dei tratti mostrati dal SOR nell'immagine. Sostanzialmente, questi sono costituiti sia da ellissi che derivano dalle sue sezioni trasversali (come le estremità), sia dalle due porzioni laterali della sua sagoma (quest'ultime riunite sotto il nome di *contorno apparente*). Il riconoscimento automatico dei tratti di un SOR in un'immagine permetterebbe l'automazione dell'intero processo di ricostruzione metrica. Un simile obiettivo rappresenta un problema di *segmentazione*, un campo della computer vision che cerca di organizzare in maniera automatica i dati grezzi provenienti dalle immagini. Nel caso specifico, poiché l'autocalibrazione è intrinsecamente mal condizionata, è auspicabile che la segmentazione fornisca un aiuto nel determinare i vincoli geometrici necessari alla ricostruzione del SOR.

In questa tesi sono stati studiati e sviluppati algoritmi, finalizzati alla calibrazione e ricostruzione di un SOR, per la segmentazione dei suoi tratti in una singola immagine non calibrata. In particolare, i vincoli geometrici necessari alla calibrazione sono due. Il primo vincolo riguarda la trasformazione che fa corrispondere in maniera biunivoca i punti delle due falde del contorno apparente. La necessità di conoscere, oltre al contorno (per la ricostruzione del vero profilo), anche

questo legame ha fatto da linea guida alla prima parte del lavoro. La stima della trasformazione e la determinazione dei tratti che la soddisfano viene infatti effettuata simultaneamente, inserendo, all'interno dell'algoritmo di stima, un'organizzazione in curve delle coppie di punti candidate a soddisfare la trasformazione. Punti estranei al SOR, che solo casualmente sono in relazione, agiscono infatti come *distrattori*; la loro natura casuale è rivelabile dal tipo di curva, breve e poco densa di punti, che solitamente vanno a costituire, a differenza delle coppie corrette che invece formano tratti lunghi e popolati. Un'analisi delle curve su tali caratteristiche permette l'eliminazione dei distrattori, favorendo la convergenza dei parametri verso i valori corretti. Il secondo vincolo per la calibrazione deriva da una generica coppia di ellissi che il SOR mostra nell'immagine; se nell'immagine sono visibili più di due ellissi, l'algoritmo è in grado di indicare la coppia di ellissi segmentate più indicate per rappresentare il vincolo.

La tesi è organizzata come segue:

- nel primo capitolo viene fornita una visione d'insieme sulle tecniche di segmentazione più diffuse, con un'attenzione particolare per quelle che si sono rivelate utili in questo lavoro;
- nel secondo capitolo viene affrontato il problema della segmentazione di un solido di rivoluzione e della sua ricostruzione 3D;
- nel terzo capitolo sono riportati gli esiti degli esperimenti condotti su immagini sintetiche e su fotografie;
- il quarto capitolo contiene le conclusioni sul lavoro svolto;
- la tesi è infine conclusa da una breve appendice con le nozioni basilari sulle coordinate omogenee ed il modello "natural camera", ampiamente usati nel secondo capitolo, al fine di rendere il documento *self-contained*.

COMPUTER VISION ED ORGANIZZAZIONE PERCETTIVA

LA COMPUTER VISION È UNA disciplina che si occupa di estrarre, mediante l'ausilio di un elaboratore elettronico, delle informazioni da una scena ripresa tramite dispositivi quali telecamere e macchine fotografiche digitali; il suo scopo è in definitiva quello di riconoscere e ricostruire la forma di oggetti distinti di una scena.

La natura del problema da risolvere fa della computer vision una disciplina di frontiera e come tale non inquadrabile in un'unica teoria, presentando bensì aree distinte ed in generale non organizzate tra di loro: essa ha forti connessioni con la matematica, la statistica e ovviamente l'informatica, più deboli con la fisica, la psicologia e le neuroscienze; si basa sulla comprensione del meccanismo con cui si formano le immagini per trarre semplici inferenze dal valore assunto individualmente dai pixel, separare regioni di pixel o estrarre delle forme, combinare le informazioni di più immagini riferite ad una stessa scena, riconoscere oggetti tridimensionali. Le sue applicazioni sono molteplici: esse vanno dalla navigazione dei robot all'interazione avanzata tra uomo e macchine, dalla sorveglianza di ambienti ai database di immagini, dall'analisi di immagini mediche alla computer grafica [13, 15].

1.1 L'ORGANIZZAZIONE PERCETTIVA

Il nostro sistema visivo ci permette di interpretare con successo ciò che è raffigurato nelle immagini in un'ampia gamma di condizioni. Questa abilità deriva dalla nostra capacità di organizzare i dati che gli occhi ci presentano, integrandoli con proprietà non direttamente visibili ma solo dedotte o ipotizzate.

Agli inizi del XX secolo alcuni psicologi tedeschi diedero origine alla *Psicologia della forma* o *Gestaltpsychologie*¹, il cui filone più famoso è quello elaborato, a partire dal 1911, dalla Scuola di Berlino costituita da Wertheimer, Köhler e Koffka. Le loro ricerche iniziali furono in particolare rivolte allo studio della percezione umana, per poi applicare i principi dedotti anche ad altri ambiti. Essi misero in evidenza il ruolo centrale giocato dall'organizzazione che il cervello impone sui dati grezzi ricevuti dall'esterno; il principio fondamentale alla base dei loro studi era infatti il seguente: noi percepiamo *totalità strutturate* e non ammassi di sensazioni puntuali.

¹ In tedesco, "Gestalt" significa "forma".

Pertanto, l'*organizzazione percettiva* può essere definita come l'abilità di imporre un'organizzazione strutturale sui dati rilevati, in modo da poterli riferire ad entità derivate da una causa comune. A partire dalla struttura prodotta, si possono poi effettuare ulteriori inferenze su quanto viene osservato.

Questo fenomeno è stato recepito anche nell'ambito della computer vision, in quanto esso impedisce robustezza ed efficienza computazionale agli algoritmi di visione.

In tempi recenti Kanizsa [18] ha enfatizzato il concetto di struttura organizzata come un complessivo ordinato, governato da leggi e non casuale, in contrapposizione ad un ammasso informe. Secondo Kanizsa,

- la Gestaltpsychologie non ha solo studiato il problema della percezione umana, ma ha prodotto una teoria psicologica generale per la comprensione di ogni aspetto del comportamento;
- l'esperienza percettiva dipende dal nostro interesse e dalla nostra attenzione;
- le parti che compongono un'entità complessa non sono indipendenti e non possono essere analizzate separatamente; inoltre le proprietà del entità complessa non sono date solo dalla somma di quelle delle sue parti²;
- la percezione è influenzata dalle esperienze passate: se "AB e C, ma non BC" è diventata una regola abituale, si avrà la tendenza a percepire ABC come AB e C;
- l'organizzazione è rivolta verso la *Prägnanz*, cioè la tendenza di un processo a realizzare il più regolare, ordinato, stabile stato possibile in una data situazione; la regolarità di una figura non dipende solo da una sua eventuale simmetria: essa è influenzata anche da criteri di *buona continuità* e semplicità;

Kanizsa ha inoltre suggerito che la percezione di forme può essere scomposta in due fasi: in primis si procede all'individuazione di regioni caratterizzate da regolarità spaziale o temporale; successivamente si passa al loro raggruppamento ed integrazione, all'inferenza di ciò che non è visibile. Il primo processo ha notevole importanza in quanto deve fornire i giusti ingressi al secondo stadio della percezione; se questo non avviene possiamo percepire organizzazioni prive di significato (Fig. 1(a)). Nel caso del completamento di contorni parzialmente occlusi, esso è deciso prevalentemente dal criterio di buona continuità nella direzione presso il punto di occlusione e dal criterio di minima distanza qualora il primo non dia luogo a figure chiuse; c'è inoltre una preferenza verso forme convesse rispetto a quelle simmetriche (Fig. 1(b)).

² Köhler utilizzava la formula "il tutto è *diverso* dalla somma delle sue parti".

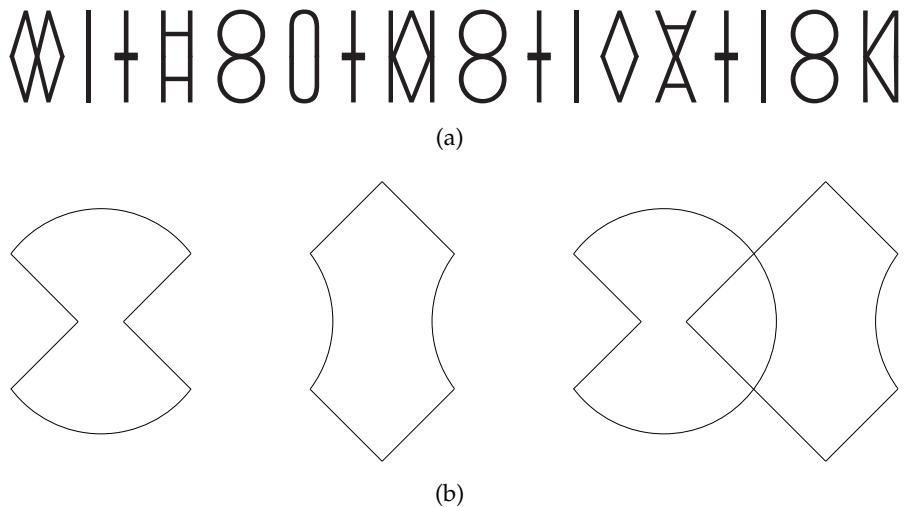


Figura 1. In (a), il raggruppamento basato su figure chiuse, simmetria e buona continuità prevale sul riconoscimento dei caratteri. In (b), la continuità delle direzioni e la convessità prevalgono sulla simmetria.

L'importanza di un'organizzazione nella percezione è stata spiegata in maniera indipendente da Lowe [21], Rock [27], Witkin e Tenenbaum [33], che hanno postulato un *principio di causa comune*: esso afferma che è altamente inverosimile che una coerenza spazio-temporale sia dovuta all'interazione fortuita di entità indipendenti, implicando quindi che se dei dati sono stati organizzati, quasi sempre ciò è dovuto alla presenza di una causa comune.

Il concetto può essere espresso anche in termini matematici: sia *Causa* l'evento per cui una serie di elementi sono parte di uno stesso oggetto e sia *Struttura* l'evento per cui è stata trovata un'organizzazione a questi elementi; la probabilità che gli elementi provengano dalla stessa causa comune, condizionata dall'aver trovato un'organizzazione per essi, è per, la legge di Bayes:

$$\Pr[Causa | Struttura] = \frac{\Pr[Struttura | Causa] \Pr[Causa]}{\Pr[Struttura]}.$$

Analizziamo i termini di questa espressione:

- $\Pr[Causa]$ esprime la probabilità a priori che un insieme di elementi provenga da una stessa causa (si può dire che cattura il significato della scena); in una situazione completamente casuale il suo valore sarebbe basso, ma nella realtà esso è invece abbastanza elevato;
- $\Pr[Struttura]$ è la probabilità a priori che una data organizzazione sia presente all'interno di un insieme di elementi e pertanto racchiude la probabilità del verificarsi casuale di una determinata struttura; il suo valore diminuisce all'aumentare della complessità dell'organizzazione;

- $\Pr\{Struttura \mid Causa\}$ è, infine, la probabilità di osservare una certa organizzazione all'interno di un insieme di elementi, condizionati dal fatto di derivare da una stessa causa (rappresenta la coerenza della materia).

In definitiva, possiamo inferire con buona sicurezza una causalità da tutte quelle organizzazioni che hanno una bassa probabilità di verificarsi casualmente ed un'alta probabilità di essere esibite dalla materia. Questi due termini non possono essere considerati in maniera isolata: ad esempio, se un'organizzazione molto articolata ($\Pr\{Struttura\}$ bassa) presenta anche una bassa probabilità $\Pr\{Struttura \mid Causa\}$ di essere osservata, il suo significato è scarso; similmente, organizzazioni che si verificano spesso negli oggetti reali ma con un alto valore per $\Pr\{Struttura\}$ (organizzazioni semplici), vedono diminuire la loro importanza.

L'insieme di organizzazioni salienti dipende dal contesto nel quale si opera: le proprietà riconosciute dalla Gestaltpsychologie nell'organizzazione percettiva di immagini reali sono *prossimità*, *continuità*, *similarità*, *chiusura* e *simmetria*; ad esse sono state aggiunte successivamente da Rock e Palmer [28] quelle di *regione comune* e di *connessione*. Le principali organizzazioni considerate nella computer vision sono *parallelismo* e *ortogonalità*, *convessità*, *continuità dei contorni* e *forme di simmetria*.

1.2 LA SEGMENTAZIONE

Nell'ambito della computer vision i concetti precedentemente esposti sono incorporati, in maniera più o meno consapevole, nei cosiddetti algoritmi di *segmentazione*, il cui scopo è quello di suddividere un'immagine in un insieme di regioni; il significato di una regione dipende dalla particolare applicazione e dal *livello dei dati*. Infatti, sebbene inizialmente la parola segmentazione indicasse solo metodi per la suddivisione di un'immagine in aree d'interesse, si sta sempre più affermando l'estensione del termine a qualsiasi procedimento che suddivida un insieme di dati estratti da un'immagine in gruppi significativi, caratterizzati dalla condivisione di qualche proprietà (ad esempio, pixel che condividono una trasformazione tra due immagini con lo stesso soggetto o all'interno della stessa foto, oppure pixel che appartengono ad una stessa curva).

Riprendendo la classificazione piramidale proposta da Sarkar e Boyer [30], possiamo individuare per i dati quattro livelli:

LIVELLO DI SEGNALE: gli elementi a cui si vuol dare un'organizzazione sono i pixel grezzi dell'immagine; ad esempio, le regioni di interesse potrebbero essere costituite da pixel con un identico livello di grigio oppure caratterizzati da un alto gradiente (punti di contorno, solitamente detti *edge*, come in Fig. 2);

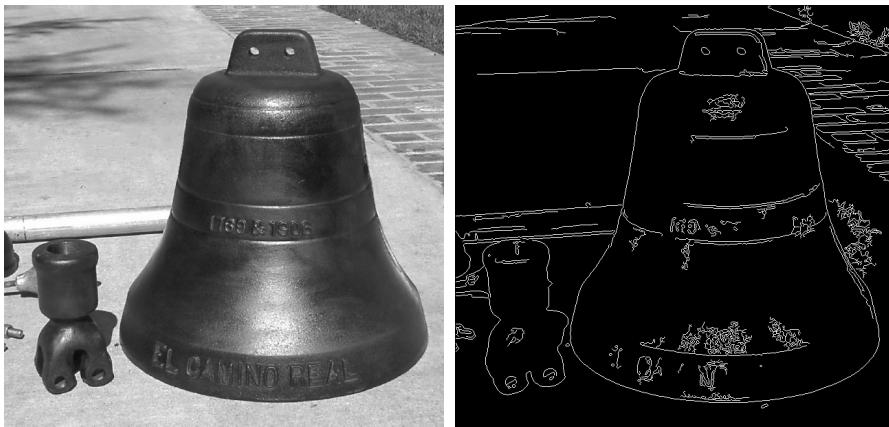


Figura 2. A destra, gli edge estratti per l'immagine di sinistra.

LIVELLO PRIMITIVO: in questo caso si cerca di organizzare le caratteristiche estratte dal precedente livello in strutture la cui dimensione è minore di quella del supporto³ a cui appartengono; nel caso di un'immagine si tratta di rette, segmenti ellittici o più in generale di curve ottenibili dai confini di regioni convesse, oppure a partire da nuvole di punti;

LIVELLO STRUTTURALE: gli elementi trovati al livello primitivo sono organizzati in entità di dimensione pari a quella del supporto; ad esempio, nel caso 2D le curve vengono organizzate in forme chiuse o simmetriche oppure in strisce parallele;

LIVELLO D'INSIEME: si cerca di assegnare un significato dipendente dal problema alle strutture trovate nel livello precedente; per esempio, nel caso di foto aeree la presenza di molti rettangoli allineati tra loro e vicini suggerisce la presenza di aree urbane.

Le tecniche computazionali adottate nella segmentazione sono varie e diversificate, anche all'interno di ciascun livello; in linea di massima si distinguono due linee, in base alla disponibilità o meno di modelli da tarare sui dati: si potrebbe voler infatti ricercare in un'immagine delle rette o delle ellissi, oppure estrarre delle curve di forma arbitraria. Nel caso che il modello non sia esplicito, la segmentazione può essere vista come un problema di *raggruppamento (clustering)*, in cui si cerca di partizionare un insieme di dati in classi (*cluster*); viceversa, se il modello è noto, il centro dell'attenzione si sposta sulla corretta stima dei suoi parametri (*model fitting*). Alcune di queste tecniche verranno discusse in dettaglio per ciascuna linea nelle prossime sezioni, analizzando più in dettaglio quelle che hanno fornito un aiuto nello svolgimento della tesi.

³ Per supporto s'intende lo *spazio* in cui la struttura viene cercata: un' immagine (dimensione 2), un filmato (2D + tempo), ...)

1.3 SEGMENTAZIONE TRAMITE CLUSTERING

Una delle caratteristiche principali della segmentazione è quella di rappresentare un'immagine in termini di classi costituite da pixel che condividono una proprietà comune: da questo punto di vista, la segmentazione rientra tra i problemi di *clustering*. In generale, dal dominio dell'immagine si passa ad un'altro insieme di dati, popolato in base all'immagine in questione, il cui dominio è un opportuno spazio di caratteristiche (*features*); una volta fatto questo, si può eseguire la segmentazione applicando un algoritmo di clustering tra i tanti esistenti in letteratura. Le caratteristiche che possono essere estratte sono molteplici: ad esempio, i pixel possono essere raggruppati in base al loro colore ed alla loro prossimità; gli edge possono essere riuniti in classi secondo le regole di vicinanza e buona continuità.

1.3.1 Clustering: tecniche generali

I metodi più semplici di clustering sono schematizzabili in due famiglie:

DIVISIVE CLUSTERING, in cui l'intero insieme di dati è visto inizialmente come un unico cluster e ricorsivamente viene diviso in cluster più piccoli;

AGGLOMERATIVE CLUSTERING, in cui i singoli elementi dell'insieme di dati costituiscono i cluster iniziali che vengono ricorsivamente fusi per formare gruppi più numerosi.

In entrambi i casi, si forma una gerarchia tra tutti i cluster, dal livello iniziale fino a quello conclusivo; per la sua rappresentazione, di solito si utilizza una struttura ad albero detta *dendrogramma* (Fig. 3).

Sorgono alcune questioni: come stabilire la partizione di un cluster o quali cluster unire? Quando interrompere la classificazione ricorsiva? Alcune delle risposte a queste domande richiedono la definizione di una distanza tra cluster: in genere, questa viene formulata a partire da una distanza definita per i singoli dati. Scelte comuni sono quelle di considerare la distanza tra due cluster come il valore minimo, oppure massimo o medio, della distanza delle coppie formate dagli elementi del primo e del secondo cluster. Nel caso di agglomerative clustering si può quindi unire di volta in volta i due cluster più vicini ed interrompere la classificazione quando la distanza tra cluster è inferiore ad una soglia o il loro numero ha raggiunto quota k (l'analisi diretta del dendrogramma è sconsigliata dalla mole dei dati solitamente in gioco). Nel caso di divisive clustering vengono impiegati dei "riepiloghi" attraverso i quali decidere la divisione di un cluster (un tipico esempio di riepilogo è l'istogramma dei colori dei pixel); se la distanza tra le due sottoclassi è sufficientemente elevata, si procede alla loro separazione.

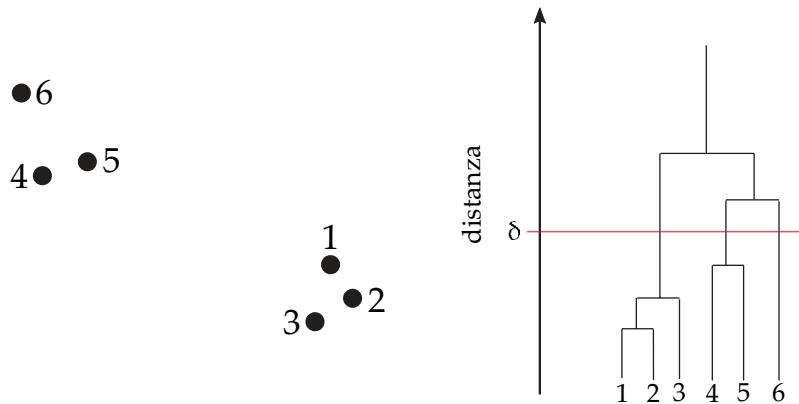


Figura 3. A sinistra, un insieme di dati; a destra, il dendrogramma ottenuto tramite agglomerative clustering. Utilizzando come distanza tra due cluster quella minima esistente tra un punto del primo ed un altro del secondo, la scelta di un particolare valore δ per essa crea una linea orizzontale che taglia il dendrogramma: i cluster sono le foglie della parte superiore del dendrogramma; nel caso illustrato, sono gli insiemi $\{1, 2, 3\}$, $\{4, 5\}$ e $\{6\}$.

K-means clustering

L'obiezione principale che si può porre alle precedenti semplici tecniche di clustering è che esse non definiscono in maniera chiara ciò che stanno cercando di ottimizzare: manca cioè loro una funzione di costo che indichi la bontà complessiva dei cluster trovati. Nell'ipotesi che il numero di classi sia noto e pari a k , una possibile funzione di costo è la seguente:

$$\Phi(\text{classi, dati}) = \sum_{i \in \text{classe}} \sum_{j \in \text{classe}(i)} (\mathbf{x}_j - \mathbf{c}_i)^\top (\mathbf{x}_j - \mathbf{c}_i),$$

dove \mathbf{x}_j è il vettore di features relativo al j -esimo dato e \mathbf{c}_i è il centro del cluster i -esimo; si ipotizza inoltre che gli elementi di un dato cluster siano vicini al suo centro. La minimizzazione di questa funzione presenta una duplice difficoltà in quanto non sono noti, in genere, né i centri dei cluster \mathbf{c}_i , né la ripartizione dei dati nelle varie classi; viceversa, se almeno una delle due informazioni fosse disponibile, la determinazione dell'altra sarebbe pressoché immediata. La soluzione al problema è fornita da un procedimento che ripete fino a convergenza i seguenti passi:

1. assumendo di conoscere la posizione di ciascun \mathbf{c}_i , si ricava la ripartizione dei dati nei vari cluster;
2. si suppone corretta la ripartizione trovata e si calcolano le coordinate dei nuovi centri.

Il tutto viene inizializzato scegliendo, in maniera casuale, le coordinate dei k \mathbf{c}_i . Il metodo esposto prende il nome di *k-means*: esso assicura la convergenza verso un minimo locale della funzione Φ , ma non viene garantito che si tratti del minimo globale.

Clustering per mezzo di grafi (analisi spettrale)

I problemi di clustering possono essere ricondotti anche al partizionamento di opportuni grafi pesati, i cui nodi sono costituiti dai dati ed i pesi degli archi indicano la *misura di affinità* degli elementi collegati: solitamente, ad archi incidenti su nodi simili corrispondono pesi elevati; viceversa, se i nodi sono molto differenti, l'affinità deve essere bassa. Nel caso di due punti \mathbf{x} e \mathbf{y} estratti da un'immagine, due possibili misure di affinità tra essi sono ad esempio le seguenti:

- *affinità per distanza*: $\text{aff}(\mathbf{x}, \mathbf{y}) = e^{((\mathbf{x}-\mathbf{y})^\top (\mathbf{x}-\mathbf{y})/\sigma^2)}$;
- *affinità per colore*: $\text{aff}(\mathbf{x}, \mathbf{y}) = e^{(\text{dist}((\mathbf{c}(\mathbf{x}), \mathbf{c}(\mathbf{y}))^2/\sigma^2)}$, dove $\mathbf{c}(\mathbf{x})$ denota il colore associato al pixel \mathbf{x} , espresso in uno spazio colore uniforme.

La costante σ deve essere elevata se si desidera raggruppare punti "distanti", piccola se i cluster devono essere composti solo da punti molto "vicini".

Un grafo può essere rappresentato tramite una matrice di adiacenza A , i cui elementi $a_{i,j}$ contengono l'affinità tra i nodi i e j ; supponendo che i cluster siano compatti e distinti, la matrice di adiacenza assume una forma simile a una diagonale a blocchi, previo un giusto ordinamento dei nodi.

Questi blocchi possono essere evidenziati attraverso lo studio degli autovalori ed autovettori di A . Una permutazione applicata simultaneamente alle righe e colonne di A (si modifica cioè l'ordine assegnato ai nodi) si ripercuote sugli autovettori con la stessa permutazione sui loro elementi, lasciando inalterati gli autovalori. L'ipotesi che sia stata trovata per A una forma simile ad una diagonale a blocchi è pertanto valida. Supponiamo che ci siano k elementi da classificare; ad ogni cluster viene inoltre assegnato un vettore di dimensione pari a k , in cui l' i -esimo componente contiene un peso relativo all' i -esimo elemento: un valore elevato denota una forte associazione tra l'elemento ed il cluster; piccolo, l'opposto. Sia \mathbf{w}_n il vettore assegnato al cluster n -esimo; definiamo una funzione obiettivo per ogni cluster:

$$\mathbf{w}_n^\top A \mathbf{w}_n.$$

Lo scopo è quello di massimizzarla al fine di estrarre i cluster più significativi; occorre però normalizzare i pesi, in quanto la loro moltiplicazione per uno scalare λ fa sì che la funzione obiettivo venga

moltiplicata per λ^2 . Imponiamo pertanto che venga soddisfatto il vincolo $\mathbf{w}_n^\top \mathbf{w}_n = 1$; utilizzando i moltiplicatori di Lagrange, dobbiamo quindi massimizzare

$$\mathbf{w}_n^\top A \mathbf{w}_n + \lambda(\mathbf{w}_n^\top \mathbf{w}_n - 1);$$

differenziando e trascurando il fattore moltiplicativo 2, infine

$$A \mathbf{w}_n = \lambda \mathbf{w}_n.$$

I vettori \mathbf{w}_n sono quindi gli autovettori di A ; il cluster più significativo è quello relativo all'autovalore più grande.

Se la forma di A è diagonale a blocchi, i suoi autovettori sono nulli tranne che in corrispondenza dei blocchi, dove compaiono gli autovettori di ciascun blocco. Se A non è perfettamente diagonale ma esistono c ($c < k$) cluster, ci aspettiamo di trovare c autovalori piuttosto elevati rispetto a tutti gli altri, con i relativi autovettori che presentano alcuni elementi con pesi significativi, indicanti l'appartenenza al cluster corrispondente, ed altri trascurabili.

La trattazione esposta presenta comunque delle debolezze: per esempio, se gli autovalori dei blocchi sono simili, potremmo avere degli autovettori che danno scarse indicazioni su come formare i cluster (il motivo di ciò risiede nel fatto che la combinazione lineare di due autovettori con lo stesso autovalore è ancora un autovettore).

Una tecnica più raffinata è quella dei *tagli normalizzati* (*normalized cuts*), in cui un grafo viene diviso in due sottografi tali che il costo del taglio sia una piccola frazione dell'affinità totale di ciascun gruppo.

Sia V l'insieme di nodi del grafo e sia $\{U, W\}$ una sua partizione ottenuta rimuovendo dal grafo gli archi che collegano i vertici in U con quelli di W . Si definisce *taglio* di U e W la somma dei pesi degli archi rimossi tra U e W :

$$\text{taglio}(U, W) = \sum_{i \in U, j \in W} a_{i,j},$$

dove $a_{i,j}$ è sempre l'elemento di coordinate (i, j) della matrice di adiacenza A . Si definisce inoltre l'*associazione* di U rispetto a V come

$$\text{assoc}(U, V) = \sum_{i \in U, j \in V} a_{i,j},$$

data dalla somma di tutti i pesi degli archi incidenti sui nodi di U .

La definizione di *taglio normalizzato* di U e W è la seguente:

$$N\text{taglio}(U, W) = \frac{\text{taglio}(U, W)}{\text{assoc}(U, V)} + \frac{\text{taglio}(U, W)}{\text{assoc}(W, V)};$$

questo valore è piccolo se la partizione divide V in due sottoinsiemi che hanno pochi archi, con pesi trascurabili, tra di loro e molti archi interni con pesi elevati. La partizione $\{U, W\}$ migliore è quindi quella che minimizza $N\text{taglio}(U, W)$.

Per minimizzare questo costo si opera una riformulazione del problema, definendo una matrice diagonale D , tale che $d_{i,i} = \sum_j a_{i,j}$ ed un vettore y a valori in $\{-1, 1\}$, la cui componente i -esima indica l'appartenenza del nodo i -esimo al cluster U ($y_i = -1$) o W ($y_i = 1$); si può dimostrare che la minimizzazione di N taglio è equivalente alla minimizzazione di

$$\frac{\mathbf{y}^\top (\mathbf{D} - \mathbf{A}) \mathbf{y}}{\mathbf{y}^\top \mathbf{D} \mathbf{y}}.$$

S tratta di un problema di programmazione intera, computazionalmente oneroso; una soluzione comunemente utilizzata è quella di rilassare il problema, permettendo ad y di assumere valori non discreti ma continui. La separazione degli elementi nei due sottoinsiemi U e W avviene confrontando i valori di y con una soglia. L'approssimazione reale di y può essere calcolata trovando una soluzione per l'equazione

$$(\mathbf{D} - \mathbf{A}) \mathbf{y} = \lambda \mathbf{D} \mathbf{y};$$

un appropriato valore per λ è il secondo autovalore generalizzato più piccolo ($\lambda = 0$ è sempre soluzione). Il procedimento viene ripetuto sui due sottoinsiemi, se necessario.

1.3.2 Estrazione di cluster mediante morfologia computazionale

I metodi discussi finora hanno un carattere generale; esiste tuttavia un'ampia letteratura che tratta il problema dell'organizzazione strutturale in maniera più prossima ai principi della Gestaltpsychologie: in particolare, ha rivestito notevole interesse il raggruppamento di punti sparsi o di edge in curve salienti. Anche in questo caso, sono stati proposti vari approcci, alcuni dei quali fanno parte di una disciplina indicata con il nome di *morfologia computazionale*, che utilizza la *geometria computazionale* per estrarre delle informazioni sulle forme. Per geometria computazionale s'intende lo studio di algoritmi volti a risolvere problemi di natura geometrica; un esempio è la determinazione della rete di lunghezza minima che connette un insieme di punti su di un piano (come sua applicazione si pensi al progetto di una rete elettrica). Questa tesi ha in parte utilizzato un approccio di questo tipo, ricorrendo ad un clustering basato sul *minimo albero ricoprente* [7, 8].

I grafi di natura geometrica, in cui i vertici sono punti in uno spazio euclideo (non necessariamente di dimensione 2), uniti da segmenti che costituiscono gli archi (i pesi sono dati dalla loro lunghezza), sono tra i descrittori di forme più diffusi; essi prendono il nome di *grafo di prossimità* e si differenziano in base alla nozione di vicinanza che s'intende esprimere. Alcune definizioni di prossimità sono le seguenti:

- due punti p e q appartenenti ad un insieme S sono i *vicini più prossimi* (*nearest neighbours*) se p è vicino a q almeno quanto ogni altro punto in S , o viceversa;
- p e q sono i *vicini più prossimi reciproci* (*reciprocal nearest neighbours*) se p è vicino a q almeno quanto ogni altro punto in S e viceversa;
- l'arco (p, q) appartiene ad un *minimo albero ricoprente* (*euclideo*) per S ;
- p e q sono *vicini relativi* (*relative neighbours*) se

$$d(p, q) \leq \min \{ \max(d(p, r), d(q, r)) : r \in S \};$$

- p e q sono *vicini secondo Grabiela* (*Gabriel neighbours*) se

$$d(p, q) \leq \min \left\{ \sqrt{d(p, r)^2 + d(q, r)^2} : r \in S \right\};$$

- p e q sono *vicini secondo Delaunay* (*Delaunay neighbours*) se i loro poligoni di Voronoi condividono un lato;
- p e q risentono di *mutua influenza* se si intersecano le sfere centrate in p e q e passanti attraverso i loro vicini più prossimi.

I grafi corrispondenti a questi criteri sono il grafo a massima prossimità (NNG), il minimo albero ricoprente (MST), il grafo di vicinanza relativa (RNG), il grafo di Gabriel (GG), la triangolazione di Delaunay (DT) e il grafo delle sfere d'influenza (SIG)⁴ (Fig. 4 ne illustra alcuni); tra di essi, alcuni formano inoltre una gerarchia:

$$\text{NNG} \subseteq \text{MST} \subseteq \text{RNG} \subseteq \text{GG} \subseteq \text{DT}.$$

Nella precedente elencazione è stato fatto riferimento ad alcuni termini che possono essere oscuri (albero ricoprente, Voronoi, triangolazione): segue una loro breve descrizione. Un *minimo albero ricoprente* di un grafo $G = \{V, E\}$ è un albero, i cui nodi sono dati da V ed i cui archi sono elementi di E , tale che la somma dei pesi degli archi che lo compongono è minima; nel caso di punti disorganizzati sul piano, consideriamo per il momento E formato da tutte le possibili coppie di nodi. I *poligoni di Voronoi* delimitano le regioni del piano che sono soluzioni del seguente problema: dato un insieme S di n punti nel piano, per ogni $p_i \in S$ qual è il luogo dei punti che sono più vicini a p_i rispetto ad ogni altro punto in S ? Per *triangolazione* s'intende infine una suddivisione del piano in regioni a forma di triangoli.

Un algoritmo cieco per la determinazione di uno MST ha complessità $\Omega(n^2)$, in quanto deve analizzare il grafo completo in cui tutte le coppie di nodi sono collegate; la natura geometrica del problema

⁴ Le abbreviazioni indicate sono gli acronimi dei nomi inglesi; ad esse viene fatto solitamente riferimento in letteratura.

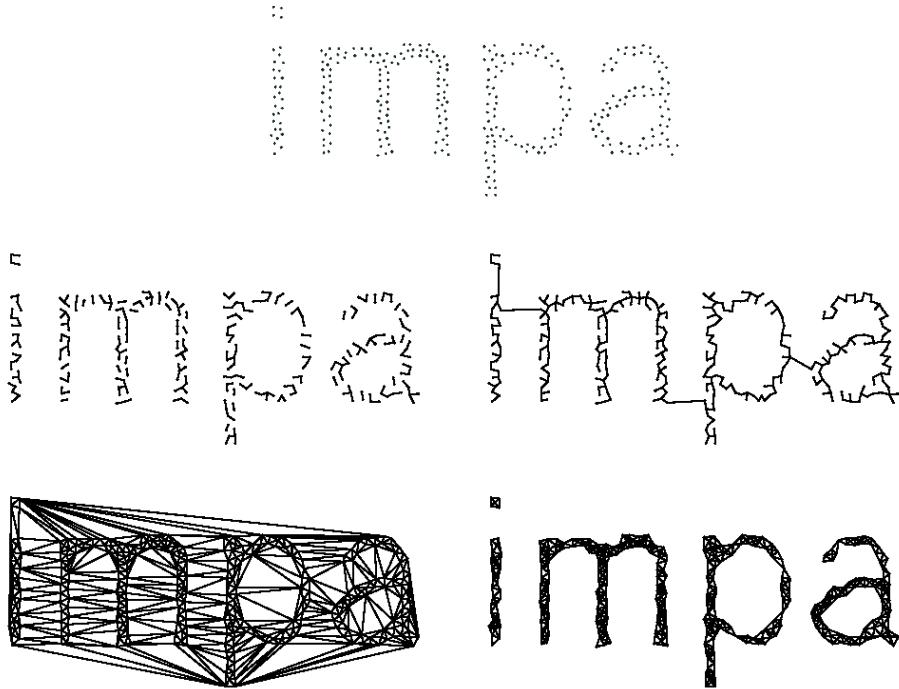


Figura 4. Sopra, l'insieme di punti disorganizzati; sotto, da sinistra verso destra e dall'alto verso il basso, NNG, MST, DT e SIG.

permette però di ridurre il numero di archi da analizzare: dalla precedente gerarchia, si ha infatti $MST \subseteq DT$ ed è possibile estrarre uno MST da DT in tempo lineare. La dimensione del quesito per i poligoni di Voronoi può essere estesa dal piano ad uno spazio di dimensioni arbitrarie; tuttavia, la complessità dell'algoritmo che determina i vicini secondo Delaunay cresce all'aumentare delle dimensioni d dello spazio in cui si trovano i punti: il numero degli archi per DT cresce almeno come $n^{\lceil d/2 \rceil}$. Comunque, nel caso bidimensionale la complessità ottimale per DT risulta essere $\Theta(n \log n)$ e pertanto anche il tempo necessario a calcolare uno MST cresce come $n \log n$ [26].

Se i campioni della curva che si vuol ricostruire sono sufficientemente densi, è stato dimostrato che i minimi alberi ricoprenti riescono a strutturare correttamente tali punti ed i loro archi forniscono un'approssimazione poligonale della curva [7]. Questo è vero se la curva non presenta più componenti oppure è chiusa: un albero può modellare solo le curve più semplici e differenziabili, definendo un cammino attraverso i suoi punti; si tratta di capire quando questo cammino è di un albero di lunghezza minima. Nonostante queste limitazioni, l'utilizzo degli MST può essere esteso anche ai casi sopracitati, grazie ad alcuni accorgimenti.

In presenza di una curva chiusa, il risultato fornito da MST è comunque una catena che collega ad uno ad uno i punti campione; tutti i nodi hanno quindi grado 2, tranne gli estremi della catena (le foglie dell'albero): basta aggiungere l'arco tra le due foglie per chiudere la

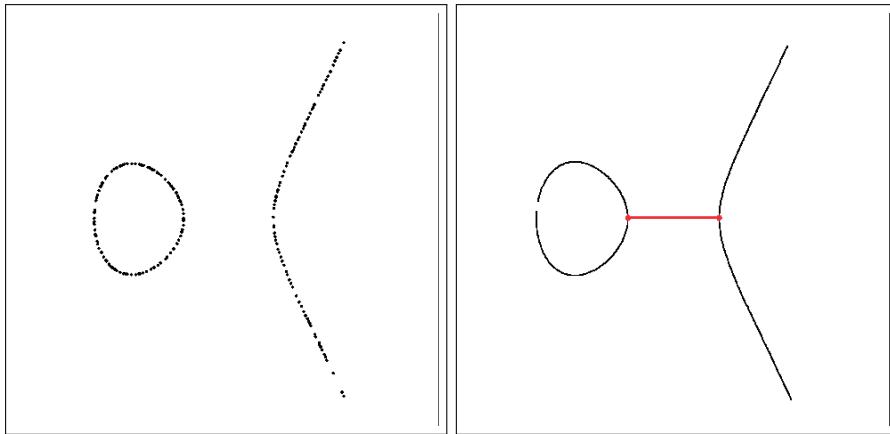


Figura 5. A sinistra, l'insieme di punti disorganizzati; a destra, MST presenta un ponte (evidenziato in rosso).

curva. Per capire se effettivamente si sta ricostruendo una curva chiusa e quindi se è necessario aggiungere un arco all'albero, si possono seguire diversi criteri:

- l'angolo definito dai segmenti incidenti sulle due foglie è circa piatto;
- la lunghezza del segmento che deve unire le due foglie è confrontabile con quella degli altri archi di MST;
- l'arco di chiusura è un arco di RNG (cfr. pag. 11).

Se la curva da ricostruire presenta più componenti separate, occorre invece rimuovere gli archi inconsistenti, o *ponti*, che le collegano tra di loro nello MST (Fig. 5): ciascuna componente deve essere infatti una curva semplice e pertanto i suoi campioni non possono essere, nel grafo che le modella, dei nodi di grado superiore a 2 (cioè ammettono al più un arco entrante ed uno uscente); i vertici con grado maggiore di 2 sono detti *punti di diramazione*.

Tuttavia tale metodo permette di ricostruire una curva a partire da un insieme di campioni sufficientemente denso, in assenza di rumore; nella pratica, questo accade raramente. In [8] si propone un metodo per l'eliminazione dei dati più rumorosi, sempre usando uno MST. Nell'ipotesi che i campioni siano comunque "sufficientemente rappresentativi" della curva, i dati più rumorosi saranno come "appesi" al cammino più lungo dell'albero, che ricostruisce la traccia della curva. Un metodo per ricavare tale cammino è quello di visitare l'albero in modalità *depth-first* per due volte: nella prima, si esplora lo MST a partire da un vertice v scelto a caso fino a trovare il nodo u a massima profondità rispetto a v ; nella seconda, si riparte da u andando a cercare il nodo w che si trova alla massima profondità rispetto ad esso. Il percorso per andare da u a w è il cammino che cerchiamo.

Altri lavori hanno visto l'utilizzo degli MST. In [19], ad esempio, si combina l'uso di MST con il metodo dei minimi quadrati mobili, al fine di assottigliare una nuvola di punti rumorosi. In [16] si fa uso dei grafi di prossimità non più per ricostruire delle curve ma delle superfici; in questo caso lo MST calcolato viene completato con altri archi che collegano i nodi u e v se almeno uno dei due appartiene all'insieme dei k punti più vicini all'altro: il grafo che si ottiene è detto *grafo di Riemann (Riemannian Graph)*. I punti che vengono collegati sono le stime del punto di tangenza tra la superficie incognita e i piani localmente tangentì, ricavati a partire dai campioni a disposizione; di questi piani è nota la direzione normale, ma non il suo verso, necessario per estrarre una distanza con segno dalla superficie di cui i piani trovati sono l'approssimazione lineare. Per avere un'orientazione dei piani consistente, si deve scegliere per uno dei punti di tangenza il verso della normale ad esso applicata e propagarla agli altri piani. L'ordine di propagazione è cruciale: è meglio non basarsi solo sulla distanza dei punti di tangenza, ma è preferibile sostituire tali valori nei pesi con una misura che esprima quanto due piani sono paralleli. Questi nuovi pesi hanno forma $1 - |\mathbf{n}_i^\top \mathbf{n}_j|$, dove \mathbf{n}_i e \mathbf{n}_j sono le normali al piano nei vertici collegati dall'arco (i, j) . La propagazione viene effettuata attraversando lo MST trovato per i nuovi pesi.

1.4 SEGMENTAZIONE PER MEZZO DI MODELLI

Oltre al clustering, l'altra grande linea in cui è schematizzabile la segmentazione consiste nella taratura di modelli, a cui devono sottostare i dati sotto esame; stavolta il modello è esplicito: si deve determinare per mezzo di uno *stimatore* il giusto valore dei suoi parametri. Le tecniche con cui vengono stimati i parametri sono, in generale, mutuate dalla statistica e si basano principalmente sull'analisi dei *residui*, cioè gli scarti tra le misure ed il modello stimato. Nella computer vision è molto importante il significato da assegnare a questi residui: i modelli della computer vision sono infatti geometrici (rette, ellissi, trasformazioni tra punti, ...), per cui è bene definire in termini geometrici anche gli scarti dal modello. Oltre ai classici stimatori statistici, i ricercatori di computer vision hanno rivolto la loro attenzione verso altri stimatori meno generali (necessitano cioè di alcune informazioni aggiuntive, di scala), al fine di affrontare dei dati estremamente disturbati da elementi estranei, che non obbediscono al modello assunto (questi dati estranei prendono il nome di *outlier*). I più importanti all'interno di questa classe di stimatori sono *Hough Transform*⁵ e *RANSAC* (RANDOM SAmple Consensus).

La prima parte della tesi ha richiesto una cooperazione tra i metodi statistici (LMedS, in particolare) e RANSAC, al fine di avere uno sti-

⁵ Hough Transform è uno stimatore nato proprio dalla Computer Vision.

matore molto robusto e generale; l'impiego di MST per la rimozione di ulteriori punti outlier ha completato il meccanismo.

1.4.1 Trattazione generale

Supposto di aver scelto il giusto modello per l'interpretazione dei dati, occorre munirsi di un criterio per mezzo del quale individuare una buona taratura.

In termini del tutto generali, sia $\mathbf{q} = [q_1, \dots, q_s]^\top$ il vettore dei parametri da stimare, e sia \mathbf{z} il vettore delle misure, interpretato come uscita del sistema che vogliamo modellare. In assenza di rumore, il sistema è descritto da una funzione per cui

$$g(\mathbf{q}, \mathbf{z}) = 0;$$

nella realtà, l'uscita \mathbf{z} è contaminata da del rumore \mathbf{n}

$$\mathbf{d} = \mathbf{z} + \mathbf{n},$$

per cui $g(\mathbf{q}, \mathbf{d}) \neq 0$. Il valore $r = g(\mathbf{q}, \mathbf{d})$, assunto da un particolare dato \mathbf{d} per un determinato vettore dei parametri \mathbf{q} , prende il nome di *residuo*.

Supponendo di avere k misure rumorose \mathbf{d}_i , $i = 1 \dots k$, dobbiamo formulare (senza perdere di generalità) una funzione di costo,

$$\mathcal{G}(\mathbf{q}, \mathbf{d}_1, \dots, \mathbf{d}_k)$$

per la quale trovare, variando \mathbf{q} , il valore minimo.

Allo scopo di misurare imparzialmente la sensibilità di uno stimatore, si definisce il suo *punto di rottura* (BP, Breakdown Point) come la più piccola frazione di outlier in un insieme di dati che può indurre lo stimatore a produrre valori per \mathbf{q} arbitrariamente cattivi [22, 29].

1.4.2 Scelta dei residui e delle funzioni di costo

La scelta della funzione $g(\mathbf{q}, \mathbf{z})$ che descrive il sistema si ripercuote nella definizione dei residui; nella computer vision questo aspetto è molto importante, data la natura geometrica dei modelli: residui che riflettano lo scostamento dal modello in termini di distanze euclidee favoriscono la corretta stima dei parametri.

I principali modelli utilizzati nella computer vision sono curve, superfici o trasformazioni tra punti. Limitando la trattazione alle immagini, ci occuperemo di curve (in forma implicita o esplicita) e di trasformazioni planari [13, 15].

Distanza da una curva

Una curva implicita nel piano è definita per mezzo di un'equazione

$$\phi(\mathbf{q}, \mathbf{d}) = 0,$$

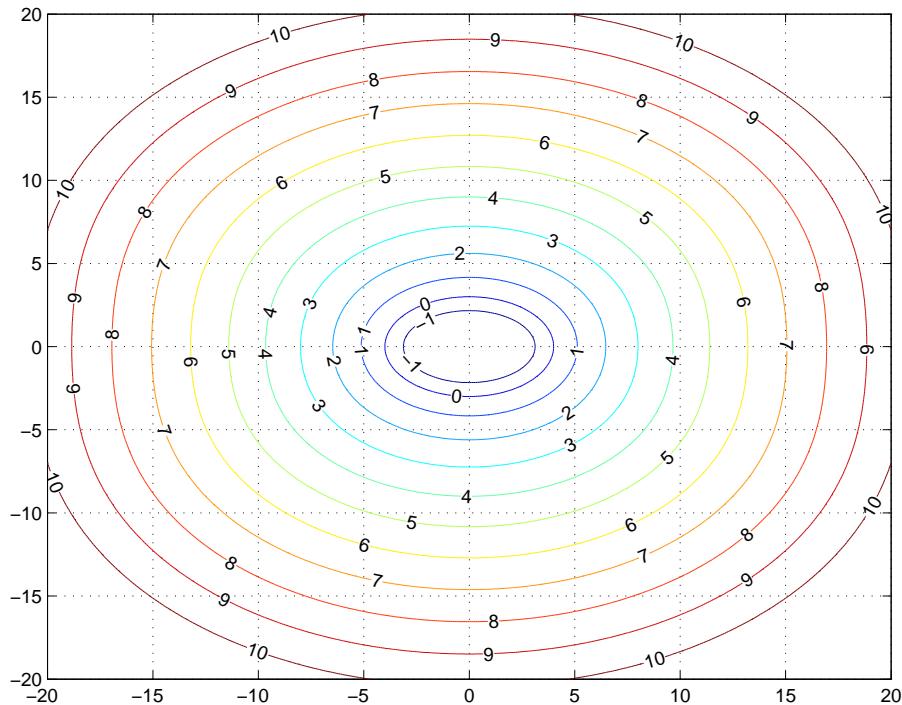


Figura 6. Approssimazione di Sampson per la distanza da un'ellisse (la curva di livello 0).

dove $\mathbf{d} = (x_d, y_d)$ è un punto del piano; se questa equazione è un polinomio, la curva è detta *algebrica* (è questo il caso più frequente). Un possibile residuo per indicare lo scostamento dal modello è quello fornito direttamente dall'equazione della curva: $r = \phi(\mathbf{q}, \mathbf{d})$; se ϕ è un polinomio, l'errore commesso prende il nome di errore algebrico. L'impiego di questo tipo di errore è generalmente sconsigliato, dato lo scarso significato geometrico; sarebbe preferibile utilizzare la vera distanza dalla curva, cioè la lunghezza del segmento congiungente il punto \mathbf{d} ed il punto sulla curva a lui più vicino. La ricerca di tale punto si basa sul fatto che il segmento che lo unisce a \mathbf{d} deve essere perpendicolare alla curva; nella pratica, tuttavia, possono insorgere dei problemi che vengono aggirati, quando possibile, con delle approssimazioni. Un esempio è fornito dalle ellissi: la distanza di un punto da esse è generalmente approssimata dall'*errore di Sampson*, che sarà introdotto nel successivo capitolo (quando verrà spiegato il procedimento adottato per la stima delle ellissi); a titolo di esempio, Fig. 6 illustra alcune curve di livello della distanza (secondo Sampson) da un'ellisse fissata.

Se la curva è in forma parametrica, essa è descritta da

$$(x(t), y(t)) = (x(t, \mathbf{q}), y(t, \mathbf{q})) \quad t \in [t_{\min}, t_{\max}].$$

In questo caso il punto su di essa che è più vicino a \mathbf{d} viene individuato dal valore del parametro t : la retta che passa per $(x(t), y(t))$ e \mathbf{d}

deve essere perpendicolare alla curva e pertanto deve occorre trovare t per cui

$$x'(t)(x_d - x(t)) + y'(t)(y_d - y(t)) = 0.$$

Rispetto alle curve implicite, la situazione non è migliore: occorre determinare lo zero della funzione definita dalla precedente equazione; spesso si deve ricorrere ad algoritmi iterativi che, se non inizializzati con un buon valore per t , possono dirigersi verso altri punti sulla curva che possono non essere i più vicini, pur rispettando la condizione di ortogonalità tra la curva e la retta che li congiunge con d .

Una volta fissata la definizione dei residui, si può procedere alla stima dei parametri minimizzando delle funzioni di costo definite su di essi, tra quelle descritte nell'esposizione di alcuni stimatori.

Distanza per le trasformazioni planari

Le trasformazioni \mathcal{H} tra due punti x e x' ($x' = \mathcal{H}(q, x)$) di maggior interesse per la computer vision sono le trasformazioni proiettive che, se espresse in coordinate omogenee⁶, sono esprimibili tramite matrici:

$$x' = \mathcal{H}(q, x) = H(q)x,$$

dove x e x' sono diventati i vettori omogenei associati ai due punti. Condurremo la trattazione delle possibili misure d'errore facendo riferimento proprio a queste trasformazioni, visto il loro impiego in questo lavoro; con abuso di notazione indicheremo la trasformazione \mathcal{H} direttamente con la sua matrice H e i parametri q diventano i coefficienti della matrice stessa o di una su parametrizzazione.

Per una trasformazione, il dato d è costituito dalla coppia $d = (x', x)$ che deve soddisfare $x' = H(q)x$. Nell'ambito di questa trattazione supporremo che le coppie siano già state fissate; in questa tesi ha invece avuto un ruolo importante la ricerca delle giuste associazioni, non note a priori, per la stima della trasformazione d'interesse (un'omologia armonica).

Anche per le trasformazioni è possibile definire, con un po' di calcoli, un errore di natura algebrica; esso presenta gli stessi problemi dell'errore algebrico relativo alle curve implicite sempre per il suo scarso significato geometrico. Preferiamo pertanto passare direttamente alla distanza geometrica tra un punto trasformato Hx e il punto associato x' come fonte d'ispirazione di alcune funzioni di costo.

La prima funzione di costo che prendiamo in esame è l'*errore di trasferimento* (*transfer error*): esso presuppone che solo uno dei due insiemi $\{x_i\}$, $\{x'_i\}$ sia costituito da misure afflitte da errori; supponiamo che l'insieme rumoroso sia il secondo, $\{x'_i\}$. Se indichiamo con $d(x, y)$ la distanza euclidea tra il punto x e y , l'errore di trasferimento per l'insieme di corrispondenze è

$$\sum_i d(x'_i, Hx_i)^2$$

⁶ Per una breve introduzione alle coordinate omogenee si veda l'appendice A.

Una possibile applicazione di questa funzione di costo si ha nella stima della trasformazione proiettiva di una scacchiera, della quale sono note le dimensioni, su di una fotografia (si tratta di una procedura di *calibrazione*).

Il motivo per cui l'errore di trasferimento è stato citato è comunque per fare da introduzione alla successiva funzione di costo, che prevede la presenza di rumore in entrambi gli insiemi delle misure. Se infatti consideriamo l'errore di trasferimento commesso dalla trasformazione diretta (H) e da quella inversa (H^{-1}) otteniamo l'*errore di trasferimento simmetrico (symmetric transfer error)*:

$$\sum_i d(x'_i, Hx_i)^2 + d(x_i, H^{-1}x'_i)^2 \quad (1)$$

Una funzione di costo più raffinata è costituita dal cosiddetto *errore di riproiezione (reprojection error)*, che coinvolge anche la correzione di ciascuna corrispondenza. In questo caso viene ricercata una trasformazione \hat{H} e delle coppie \hat{x}_i, \hat{x}'_i che si corrispondono perfettamente secondo \hat{H} , tali da minimizzare

$$\sum_i d(x'_i, \hat{x}'_i)^2 + d(x_i, \hat{x}_i)^2, \quad \text{con il vincolo } \hat{x}'_i = \hat{H}\hat{x}_i, \quad i = 1 \dots k.$$

La minimizzazione di questa funzione richiede la stima sia della matrice \hat{H} che dei suoi ingressi \hat{x}_i : lo spazio dei parametri viene così esteso di $2k$ incognite (i punti \hat{x}'_i non sono necessari, dato che sono completamente individuati dagli altri parametri); questo si ripercuote sul tempo di esecuzione della minimizzazione, che aumenta.

La funzione scelta nella tesi è stata quella data dall'errore di trasferimento simmetrico, per non appesantire troppo il carico computazionale; i risultati sono stati comunque buoni.

1.4.3 Stimatori statistici

Vengono adesso riportati alcuni degli estimatori statistici che trovano impiego nella computer vision [38].

Minimi quadrati (LS)

Una tra le più note, e storiche, funzioni di costo utilizzate come estimatore è quella *ai minimi quadrati (LS)*. Posto $\hat{\mathbf{q}}$ come stima dei parametri, sia $r_i = g(\hat{\mathbf{q}}, \mathbf{d}_i)$ il residuo del modello g sulla misura y_i ; la funzione di costo è

$$G(\mathbf{q}, \mathbf{d}_1, \dots, \mathbf{d}_k) = \sum_i r_i^2.$$

A scapito della sua semplicità, tuttavia, questo criterio è ottimo solo in presenza di disturbo gaussiano sulle misure: ha infatti il punto di rottura pari a 0: basta, cioè, un solo outlier per far fallire la stima di \mathbf{q} . Nei problemi di computer vision questo non è il suo

unico inconveniente: se applicato brutalmente alla stima di una retta ($g(\mathbf{q}, \mathbf{z}) = ax + by + c = 0$), o di una curva implicita come un'ellisse ($g(\mathbf{q}, \mathbf{z}) = ax^2 + 2bxy + cy^2 + 2dx + 2ey + f = 0$), esso è affetto da un bias dovuto allo scarso significato geometrico del residuo. Le cose migliorano se il residuo viene modificato, per rappresentare la distanza ortogonale di un punto dalla curva da modellare.

Stimatori M

Pur avendo anch'essi $\text{BP} = 0$, una classe di stimatori che solitamente opera bene nei problemi di computer vision è costituita dai cosiddetti *stimatori M* (*M estimators*). Infatti il punto di rottura è un'indice di robustezza trovato nella peggior condizione possibile; gli stimatori M riescono invece a trattare adeguatamente gli outlier e falliscono quando gli outlier sono dei *punti di leva*. Un punto di leva è un dato \mathbf{d} che si discosta dalla popolazione costituita dalle altre misure, e non è detto in generale che sia un effettivo outlier; ad esso è infatti concesso di appartenere al modello. Gli stimatori M risolvono il problema degli outlier sostituendo i quadrati dei residui con un'altra funzione $\rho(r_i)$ che, come l'elevazione al quadrato, è simmetrica, definita positiva ed ha un unico minimo in 0, differenziandosi invece nella crescita (più lenta all'aumentare di r_i). Quindi uno stimate M ricerca il vettore di parametri per cui

$$\mathcal{G}(\mathbf{q}, \mathbf{d}_1, \dots, \mathbf{d}_k) = \sum_i \rho(r_i)$$

è minima. La scelta della funzione $\rho(r)$ contraddistingue gli stimatori M tra loro. Oltre allo spazio dei parametri viene però aggiunta una nuova incognita, il *fattore di scala* σ , che regola la forma, più o meno allargata, di $\rho(r)$. Tipicamente la scala è stimata in maniera iterativa, durante la minimizzazione di \mathcal{G} (la funzione di costo è non lineare); un metodo usato è quello di porre al passo n -esimo

$$\sigma^{(n)} = 1.4826(1 + 5/(k - s))\text{med}|r_i^{(n)}(\mathbf{q}^{(n-1)})|.$$

Il "numero magico" 1.4826 è un fattore di correzione che, in presenza di rumore gaussiano, permette di avere la stessa *efficienza*⁷ di LS.

La ricerca di un minimo per \mathcal{G} comporta la determinazione di un vettore \mathbf{q} che soddisfa il sistema dato dalle m equazioni

$$\sum_i \psi(r_i) \frac{\partial r_i}{\partial q_j}, \quad j = 1, \dots, s,$$

dove $\psi(r) = d\rho(r)/dr$ viene detta *funzione d'influenza*: essa misura l'influenza di un dato sul valore della stima dei parametri; per esempio se $\rho(r) = r^2$, allora $\psi(r) = 2r$, cioè l'influenza è lineare con la

⁷ Si definisce *efficienza* il rapporto tra la più piccola varianza per i parametri stimati e l'effettiva varianza fornita dal dato metodo.

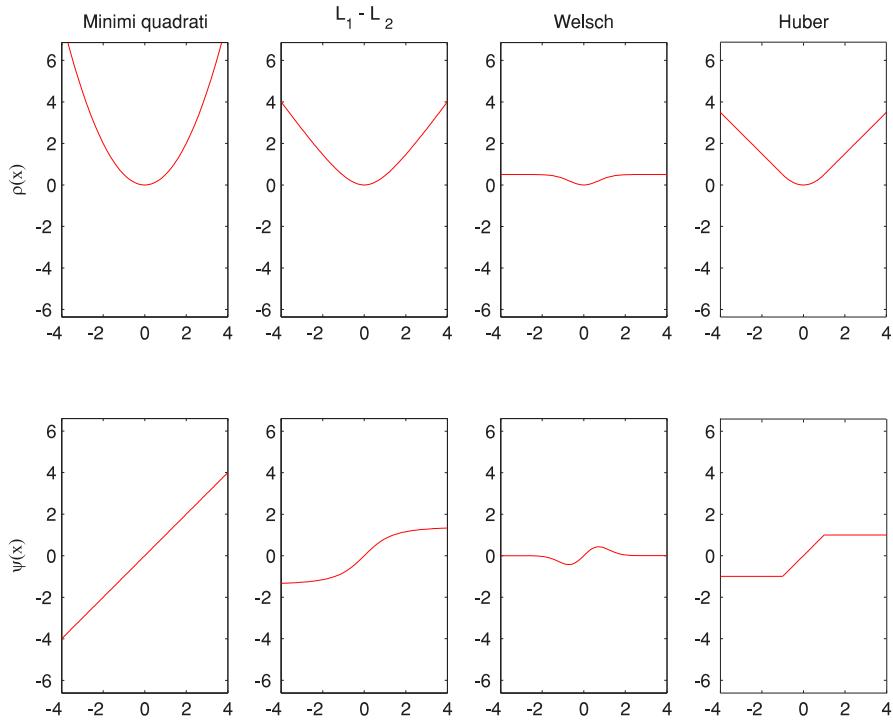


Figura 7. Alcuni estimatori M , caratterizzati secondo ρ e ψ , posti a confronto con i Minimi quadrati (LS).

dimensione dell'errore. Questa considerazione conferma la scarsa robustezza di LS, suggerendo come requisito principale di una buona $\rho(r)$ la limitatezza dell'insieme immagine della propria $\psi(r)$. Alcuni estimatori M sono quelli in Fig. 7, caratterizzati in base a ρ e ψ .

Minima mediana dei quadrati (LMedS)

LMedS [29] fa parte della classe di estimatori con il più alto BP teorico (50%) per gli estimatori statistici; esso determina i parametri del modello minimizzando la mediana della distribuzione dei quadrati dei residui:

$$\mathcal{G}(\mathbf{q}, \mathbf{d}_1, \dots, \mathbf{d}_k) = \min \text{med} r_i^2.$$

L'assenza di una formula più algebrica porta teoricamente LMedS ad effettuare una ricerca nello spazio delle possibili stime generate sui dati; essendo questo spazio troppo grande, si decide di analizzare solo un sottoinsieme casuale dei dati. L'algoritmo, utilizzato per la stima di una regressione lineare, è schematizzato nei seguenti passi.

1. Effettua sui dati una selezione casuale di N sottoinsiemi, ciascuno di cardinalità s pari al minimo numero di misure necessario per determinare un vettore di parametri \mathbf{q} .
2. Per ogni sottoinsieme \mathcal{J} stima il relativo vettore $\mathbf{q}_{\mathcal{J}}$.

3. Per ciascun \mathbf{q}_j , determina la mediana dei quadrati dei residui rispetto al modello istanziato da \mathbf{q}_j :

$$M_j = \text{medr}_i^2(\mathbf{q}_j, \mathbf{d}_i)$$

4. Seleziona il vettore \mathbf{q}_j che ha prodotto il minimo valore per M_j .

La scelta del numero N di sottoinsiemi non è stata indicata: una sua stima può essere effettuata se però si ha un'idea della massima percentuale di outlier presenti nella popolazione dei dati; sia ε questa percentuale. Equivalentemente, la probabilità di scegliere *un dato* non outlier è $w = 1 - \varepsilon$; di conseguenza, la probabilità di aver scelto *almeno un insieme* privo di outlier su N tentativi è data da $p = 1 - (1 - w^s)^N$. Risolvendo rispetto a N , si può determinare il numero di tentativi necessario per trovare, con probabilità p , almeno un sottoinsieme privo di outlier:

$$N = \frac{\log(1-p)}{\log(1-w^s)}; \quad (2)$$

soltamente, si pone $p = 0.99$.

Se enunciato in questo modo, LMedS presenta una bassa efficienza in presenza di rumore gaussiano; per compensarla, si aggiunge una minimizzazione ai minimi quadrati sui residui pesati da

$$\gamma_i = \begin{cases} 1 & \text{se } |r_i| \leq \alpha \hat{\sigma} \\ 0 & \text{altrimenti} \end{cases}, \quad (3)$$

dove solitamente $\alpha = 2.5$ e

$$\hat{\sigma} = 1.4826 \frac{1+5}{k-s} \sqrt{\min M_j}. \quad (4)$$

Nel caso monodimensionale (un solo parametro q) LMedS diventa

$$\min \text{med}(d_i - q)^2$$

e fornisce per q il valore dove la concentrazione dei punti è maggiore; si tratta, quindi, di un sorta di stimatore della moda delle misure d_i . Inoltre, il procedimento per la minimizzazione della mediana passa da casuale a deterministico, in virtù del seguente

Teorema 1.4.1. Se $\mu_\tau^2 := \text{med}(d_i - \tau)^2 = \min \text{med}(d_i - q)^2$, allora sia $\tau - \mu_\tau$ che $\tau + \mu_\tau$ sono osservazioni nei dati $\{d_1, \dots, d_k\}$.

Questo enunciato consente di stimare q individuando la *più piccola metà* delle misure: dopo aver ordinato in senso crescente l'insieme $\{d_1, \dots, d_k\}$, si deve determinare il minimo valore delle differenze

$$(d_h - d_1), (d_{h+1} - d_2), \dots, (d_k - d_{k-h+1})$$

dove $h = \lceil k/2 \rceil$. Il corrispondente sottoinsieme di h elementi è detto la metà più piccola perché di tutti i possibili sottoinsiemi con h

elementi possiede lo scarto più piccolo tra il primo e l'ultimo dei suoi elementi; il punto di mezzo di questo sottoinsieme è la stima per q .

La stima del fattore di scala $\hat{\sigma}$ (l'interpretazione di $\hat{\sigma}$ è la varianza della distribuzione priva di outlier) avviene in tre passi:

1. poni $\hat{\sigma} \leftarrow 1.4826(1 + 5/(k - 1))\sqrt{\min M_d}$;

2. calcola i pesi γ_i ;

3. $\hat{\sigma} \leftarrow \sqrt{(\sum \gamma_i r_i^2) / (\sum \gamma_i - 1)}$;

LMedS monodimensionale è stato ampiamente utilizzato in questa tesi, ogni volta che si è voluto estrarre outlier da popolazioni che potevano generare distribuzioni in una dimensione. Quando nel prossimo capitolo verrà fatto riferimento a LMedS, esso sarà sempre quello monodimensionale.

1.4.4 Stimatori per la computer vision

Nonostante l'ampia disponibilità di estimatori offerti dalla statistica, i ricercatori nel campo della computer vision hanno sentito la necessità di utilizzare degli estimatori *ad hoc*, per superare le difficoltà dovute alla presenza di outlier in numero maggiore rispetto ai dati trattati solitamente dalla statistica. Tra essi i principali sono stati gli estimatori Hough Transform e RANSAC, nati rispettivamente circa quaranta e venti anni fa. Come si vedrà, RANSAC presenta molte analogie con LMedS che, pur essendo stato proposto alcuni anni dopo rispetto a RANSAC, si è diffuso per primo nei problemi di computer vision. Solo recentemente i ricercatori hanno iniziato a sostituire LMedS con RANSAC nei loro algoritmi di visione: la necessità di impostare a priori il fattore di scala (cioè la soglia con cui individuare gli outlier) se da un lato lo rende meno generale, dall'altro gli permette di adattarsi meglio alle situazioni molto complesse (presenta infatti un BP superiore al 50%).

Hough Transform (HT)

Lo stimatore Hough Transform [14], dovuto a Hough (1959), è spesso utilizzato per estrarre linee da un'immagine binaria; la sua implementazione può comunque essere estesa alla stima di altri modelli.

Supponiamo di avere un insieme di m punti nel piano $d_i = (x_i, y_i)$, e di voler stimare la retta che meglio li descrive. Dato un punto d_i , il fascio di rette che passa per è esprimibile come $y_i = ax_i + b$ e i suoi parametri sono a e b ; dualmente, considerando lo spazio che questi parametri definiscono (detto *dominio di Hough*), ciascun punto d_i definisce in esso una retta data da $L_i = \{(a, b) : y_i = ax_i + b\}$. Se i punti d_i sono tutti allineati, tutte le rette L_i si incontrano nello stesso punto $q = (\hat{a}, \hat{b})$.

Nella pratica, HT è realizzato come segue. Innanzitutto lo spazio dei parametri viene quantizzato e ad ogni quanto corrisponde una cella in una tabella, detta *accumulatore*, in cui annotare il numero di linee L_i che intersecano il quanto; la cella con il maggior punteggio fornisce la stima della retta originaria. Se si stanno cercando più rette, si fissa una soglia e le celle che la superano individuano le rette nell'immagine.

Possiamo definire anche la funzione di costo di HT. A tale scopo, sia $B_r(\mathbf{q})$ l'intorno di raggio r centrato sul vettore dei parametri \mathbf{q} (esso serve a rappresentare il processo di quantizzazione): HT ricerca il punto \mathbf{q} per il quale è massimo il numero di linee L_i che intercettano l'intorno $B_r(\mathbf{q})$. In termini matematici, HT massimizza⁸

$$\frac{1}{m} \sum_i^m \mathbf{1}_{\{B_r(\mathbf{q}) \cap L_i \neq \emptyset\}}.$$

Si noti che $B_r(\mathbf{q}) \cap L_i \neq \emptyset$ se e solo se la distanza tra L_i ed il centro dell'intorno \mathbf{q} è minore o uguale a r . Come già detto, la dipendenza dal parametro r (la scala) rende HT meno generale rispetto agli estimatori statistici.

La generalizzazione di HT a modelli che non siano rette segue lo stesso meccanismo di quantizzazione e raccolta dei voti: per ogni dato si determina nel dominio di Hough il luogo dei parametri relativi ai modelli che descrivono il dato; si incrementa il valore di tutte le celle attraversate da ciascun luogo e si seleziona la cella con il punteggio maggiore.

Dal momento che le dimensioni dell'accumulatore crescono rapidamente all'aumentare del numero di parametri, HT non viene di solito impiegato nella stima di modelli con più di tre parametri.

RANSAC

RANSAC (RAndom SAmples Consensus) è un algoritmo proposto da Fischler e Bolles nel 1981, con lo scopo di fronteggiare un'alta percentuale di outlier. L'idea di base è simile a quella di LMedS: si selezionano casualmente dei sottoinsiemi di cardinalità s dall'insieme dei dati (s è pari al minimo numero di dati necessari ad istanziare un modello); per ciascun sottoinsieme si analizza il comportamento del relativo modello sugli altri dati. La differenza è data dall'analisi condotta: invece di osservare la mediana dei residui, si determina quanti dati hanno un residuo dal modello inferiore ad una certa soglia t . Al termine, si determina il sottoinsieme più numeroso, che diventa l'*insieme di consenso* C . Il numero di tentativi N è aggiornato durante l'esecuzione dell'algoritmo in base a Eq. 2, in cui la stima per w è data dal rapporto tra il numero complessivo di dati e la cardinalità dell'insieme di consenso corrente. Come HT, la soglia t rende RANSAC

⁸ Se cambiamo segno alla funzione di costo, ci riconduciamo ad un problema di minimizzazione.

Algoritmo 1. Stima robusta con RANSAC

- 1 Poni $N = \infty$, $k = 1$, $\mathcal{C} = \emptyset$
 - 2 Se $k > N$, stop
 - 3 Forma \mathcal{I}_k scegliendo casualmente s campioni dall'insieme di tutti i dati e istanzia il vettore dei parametri \mathbf{q}_k
 - 4 Calcola la distanza d_{\perp} di ciascun dato dal modello relativo ai parametri \mathbf{q}_k e costituisci \mathcal{C}_k con i dati per i quali $d_{\perp} < t$
 - 5 Se $\|\mathcal{C}_k\| > \|\mathcal{C}\|$, $\mathcal{C} \leftarrow \mathcal{C}_k$ e aggiorna N (con $\|\mathcal{C}\|$ si è indicata la cardinalità dell'insieme C)
 - 6 $k \leftarrow k + 1$ e vai a 2
-

meno generale rispetto a LMedS; d'altra parte gli consente, come già detto, di affrontare una maggiore quantità di outlier ($BP > 0.5$).

Lo schema di RANSAC è riassunto in Alg. 1. Poiché esso è stato utilizzato, con alcune modifiche, in questa tesi per la stima di un'omologia armonica⁹, l'esposizione che segue fa riferimento alla sua applicazione nella stima di una qualche trasformazione tra punti. La distanza d_{\perp} usata è fornita dall'errore di trasferimento simmetrico (Eq. 1): $d_{\perp}^2 = d(\mathbf{x}', H\mathbf{x})^2 + d(\mathbf{x}, H^{-1}\mathbf{x}')^2$. Si suppongono note le coppie candidate a soddisfare il modello, dalle quali devono essere eliminate quelle outlier.

Al termine del campionamento casuale viene eseguita una minimizzazione al variare di \mathbf{q} sugli elementi dell'insieme di consenso; dal momento che la funzione di costo è non lineare, il minimo viene individuato con algoritmi iterativi (Levemberg-Marquardt, solitamente) a partire dai parametri su cui l'insieme di consenso è stato determinato. Segue infine una successione di riclassificazione dei dati sui nuovi parametri e conseguente minimizzazione, fino a convergenza.

⁹ Per il momento basta sapere solo che si tratta di una trasformazione; essa verrà definita nella sezione 2.4.1.

2

RICONOSCERE E RICOSTRUIRE UN SOLIDO DI RIVOLUZIONE

ISOLIDI DI RIVOLUZIONE COSTITUISCONO una famiglia di oggetti molto diffusi negli ambienti creati dall'uomo (Fig. 8) e dal punto di vista geometrico possiedono delle interessanti proprietà che permettono la loro ricostruzione 3D a partire da una singola immagine.

2.1 TERMINOLOGIA DI BASE

Geometricamente parlando, un solido, o superficie, di rivoluzione (indicato in letteratura generalmente con l'acronimo *SOR*, dall'inglese Surface Of Revolution) si ottiene ruotando una curva planare (detta *profilo*) attorno ad un asse (Fig. 9). Facendo coincidere l'asse di rivoluzione con l'asse z di un sistema di riferimento ortogonale $Oxyz$ possiamo pertanto rappresentare il *SOR* attraverso la parametrizzazione

$$\varphi(t, \vartheta) = \begin{cases} x(t, \vartheta) = \rho(t) \cos \vartheta \\ y(t, \vartheta) = \rho(t) \sin \vartheta \\ z(t, \vartheta) = t; \end{cases} \quad (5)$$

dove $\rho(t)$ è il profilo del *SOR* e $(t, \vartheta) \in [0, 1] \times [0, 2\pi]$, come riportato in [5, 6]. Utilizzando questa parametrizzazione, possiamo definire due famiglie di curve sulla superficie del *SOR*: i *meridiani* ed i *paralleli*, costituiti rispettivamente da punti a ϑ costante e a z costante. In altre parole, i meridiani hanno tutti la forma di $\rho(t)$; i paralleli, invece, sono costituiti da cerchi il cui raggio varia secondo $\rho(t)$.

Nella silhouette assunta da un *SOR* in un'immagine sono riscontrabili due differenti tipi di curve: il *contorno apparente* ed i *bordi*: il



Figura 8. Esempi di solidi di rivoluzione.

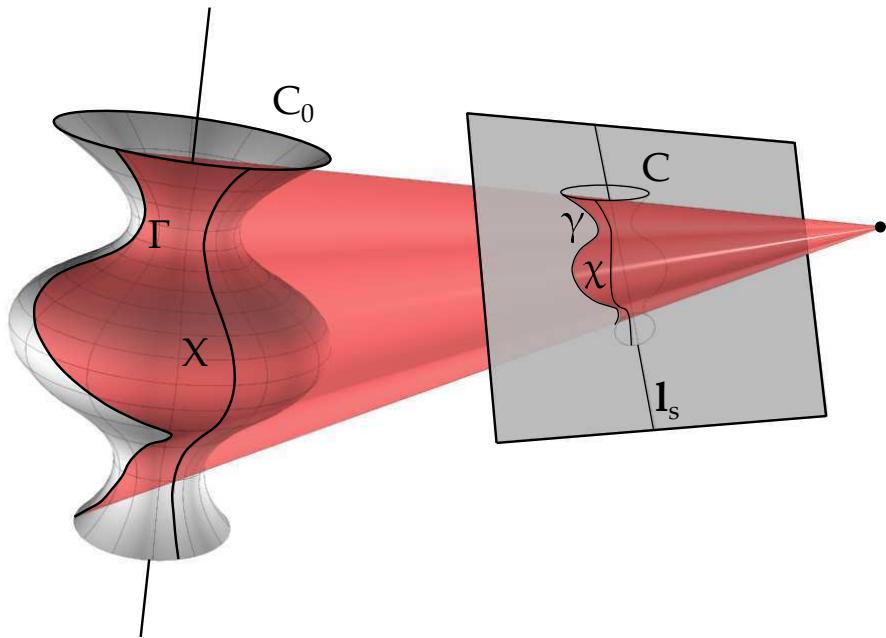


Figura 9. Geometria dell’immagine di un SOR. Γ è una parte di profilo generatore e γ è il relativo contorno apparente; X e χ sono invece un meridiano e la sua immagine, mentre l’ellisse C costituisce uno dei bordi e deriva dal parallelo C_0 . l_s rappresenta la proiezione dell’asse di rivoluzione; la superficie traslucida è l’insieme dei raggi uscenti dal centro di proiezione e tangenti al SOR lungo Γ .

primo è determinato dalla proiezione nell’immagine dei punti che costituiscono il *generatore di contorni* e cioè quelli in cui la superficie ha un andamento “dolce” ed i raggi uscenti dal centro ottico della fotocamera sono tangenti al SOR¹; gli altri derivano invece dai punti sulla superficie dove la sua normale presenta delle discontinuità.

Il contorno apparente non è direttamente riconducibile alla proiezione di particolari meridiani sull’immagine, dal momento che il generatore di contorni è una curva non planare; invece i bordi sono dovuti alla proiezione dei paralleli (visibili più o meno completamente) che delimitano il SOR e pertanto vengono visualizzati come segmenti ellittici, dato che l’immagine di un qualsiasi parallelo è un’ellisse; le ellissi degli altri paralleli saranno dette, con leggero abuso di terminologia, *sezioni trasversali*. Fig. 9 illustra quanto esposto.

2.2 LAVORI PRECEDENTI

L’interesse sempre crescente per i database di modelli tridimensionali, così come quello per la ricostruzione grafica di ambienti reali, ha

¹ In alternativa, al posto dei raggi, si può far riferimento ai piani contenenti il centro ottico e tangenti alla superficie del SOR [12].

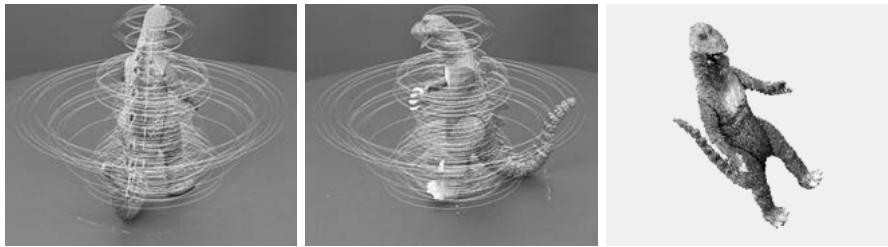


Figura 10. Esempio di SAM, Single Axis Motion; nell'ultima immagine, il modello ricostruito.

rinnovato l'interesse della ricerca nel campo dell'acquisizione della forma 3D di oggetti a partire da una o più fotografie. In molti casi, queste possono essere l'unico punto di appoggio, qualora l'oggetto non sia più disponibile oppure abbia caratteristiche mal conciliabili (materiali inadatti, dimensioni eccessive) con le classiche tecniche di acquisizione 3D, come lo scanner laser.

Per una ricostruzione fedele dell'oggetto è necessario conoscere i parametri di fotocamera che governano la proiezione della scena sull'immagine. Tuttavia, in genere essi non sono noti; assume quindi un ruolo essenziale l'*autocalibrazione* che, a partire da vincoli geometrici noti e relativi a particolari entità rintracciabili nella foto (angoli tra rette nello spazio [15], strutture che si ripetono [24], ...), permette la stima di tali parametri. In questo ambito hanno visto la luce varie ricerche per la ricostruzione di scene planari o di edifici, come in [20]. Si possono ricavare vincoli utili per la calibrazione e/o la ricostruzione anche quando la scena è sottoposta ad un moto planare noto; recenti lavori hanno riguardato la ricostruzione di oggetti a partire da una loro sequenza filmata, mentre sono fatti ruotare posti su di un tavolo ruotante (*SAM, Single Axis Motion*), come in Fig. 10 [10, 17, 23].

Il problema della calibrazione da un SOR e della sua ricostruzione è già stato ampio oggetto di studio [5, 6, 34, 35] e può essere messo in stretta relazione con SAM. Le grandezze che in SAM permettono la calibrazione sono entità che rimangono invariate lungo tutta la sequenza; d'altra parte, l'immagine di un SOR può essere pensata come una sequenza infinita in cui l'asse di rivoluzione del tavolo ruotante coincide con quello del SOR, per cui il problema della calibrazione si riconduce alla determinazione di queste entità fisse. Nella terminologia dei SOR, le principali sono:

- l'asse ed il vertice dell'omologia armonica (sezione 2.4.2) che lega i due lati del contorno apparente;
- i *punti circolari* (cfr. Appendice A) del fascio di piani ortogonali all'asse di rivoluzione.

L'autocalibrazione è un problema intrinsecamente malcondizionato, per cui la stima di questi invarianti richiede una notevole accura-

tezza; in SAM, questa è raggiunta grazie all'utilizzo della *geometria epipolare*² ed al *tracking*³ delle numerose ellissi che generano i punti "più esterni" degli oggetti sul tavolo ruotante (si vedano le prime due illustrazioni nella figura 10). Nell'immagine di un SOR questo non è possibile; inoltre, le ellissi visibili sono generalmente solo due (la base e la bocca) o poche più.

Allo stato attuale, la ricostruzione di SOR è effettuata in maniera semiautomatica: in particolare, l'intervento umano consiste nell'indicazione del contorno apparente e delle immagini di almeno un paio di paralleli [5, 6]. L'impiego di un algoritmo di segmentazione in grado di individuare queste grandezze renderebbe il processo completamente automatico; tenuto conto del malcondizionamento dell'autocalibrazione, la segmentazione deve però anche aiutare la stima delle entità fisse, al fine di raggiungere la precisione necessaria.

I precedenti lavori di segmentazione sui SOR si sono concentrati principalmente sul problema della loro localizzazione approssimata nell'immagine e del loro riconoscimento rispetto ad un archivio di modelli (non tridimensionali, ma basati su invarianti prospettici). In genere, essi sono stati trattati come sottoinsieme dei *cilindri generalizzati retti* (*SHGC, Straight Homogeneous, Generalized Cylinders*) [1, 25, 36]; i metodi proposti analizzano le relazioni di buona continuità, simmetria, chiusura, lavorando direttamente sugli edge estratti e tengono conto delle trasformazioni che legano tra di loro le immagini dei paralleli od il contorno apparente solo per localizzare il SOR (ad esempio, in [25] si fa riferimento all'omologia armonica secondo la quale si corrispondono i due lati del contorno apparente, ma viene trattata in una sua forma più semplice, quella affine). Inoltre, l'eventuale ricostruzione è *un'inferenza volumetrica*, che fornisce solo un'idea della forma del SOR; la *ricostruzione metrica*, invece, si prefigge lo scopo di recuperare le vere proporzioni tra le parti che costituiscono un'oggetto.

Il problema di una segmentazione rivolta alla ricostruzione metrica di un SOR è quindi aperto. L'algoritmo, sviluppato nel corso della tesi, che verrà illustrato in questo capitolo si differenzia da quelli precedentemente citati: esso, infatti, ricerca inizialmente dei punti candidati ad appartenere alla sagoma del SOR o alle sue ellissi visibili, per poi strutturarli in curve. Affronta anche il problema della selezione delle ellissi migliori, dato che una loro coppia permette in linea di principio la stima dei punti circolari [5, 6].

Infine, per completare il modello 3D in maniera realistica occorre riprodurre sulla superficie sintetica anche la *tessitura* (*texture*) che il SOR mostra nell'immagine (ad esempio, nel caso in cui il SOR sia un

² La geometria epipolare studia le trasformazioni proiettive che legano due immagini dello stesso scena, prese da differenti punti di vista; se pensiamo fisso il tavolo ruotante, è la telecamera che si sposta attorno ad esso.

³ Per tracking s'intende l'inseguimento, in una sequenza filmata, della posizione di oggetti o punti.

vaso artistico, la sua texture è costituita dai motivi che lo decorano). In [5, 6] si riesce ad acquisire la texture sfruttando solo i parametri di fotocamera strettamente necessari alla ricostruzione della forma del SOR. Diversamente da [5, 6], il recupero della texture svolto in questa tesi è invece conseguenza del completamento della stima del modello di fotocamera, per mezzo del metodo proposto nella sezione 2.8. Viene quindi seguito un approccio più tradizionale, in cui la conoscenza di tutti i parametri di fotocamera permette di retroproiettare i pixel dell'immagine sulla superficie sintetica.

2.3 VISIONE D'INSIEME DELL'ALGORITMO

Ciò che balza agli occhi dell'immagine di un SOR è sicuramente l'aspetto quasi simmetrico assunto dalla sua sagoma; questo è dovuto all'omologia armonica che lega i due lati del contorno apparente: la simmetria assiale ne costituisce infatti un caso particolare.

Questa forte caratterizzazione dell'immagine di un SOR, ha suggerito come primo passo da compiere quello della stima dell'omologia armonica che ne è la causa, e di tutte le curve consistenti con essa (sezione 2.5). Questa scelta ha condizionato il resto dell'algoritmo di segmentazione: lo stadio successivo deve necessariamente classificare queste curve (sezione 2.6): alcune di esse costituiscono infatti il contorno apparente ed i segmenti ellittici del SOR, mentre altre sono dovute a simmetrie casuali nella texture del SOR o nello sfondo della foto; quest'ultime rappresentano dei *distrattori* che devono essere eliminati. Sulla base dei risultati forniti dalla segmentazione, si può quindi passare al terzo stadio (sezioni 2.7 e 2.8), quello della calibrazione e ricostruzione.

Allo scopo di lavorare solo su punti che formano delle curve, come i precedenti lavori di segmentazione sui SOR, si è deciso di utilizzare un edge-detector per estrarre i contorni delle figure; in particolare è stato adoperato quello proposto da Canny [4].

2.4 PROPRIETÀ PROIETTIVE DEI SOLIDI DI RIVOLUZIONE

La forma regolare di un SOR permette alla sua immagine di esibire delle notevoli proprietà, alcune delle quali hanno rivestito un ruolo importante nello sviluppo di questo lavoro.

2.4.1 Proprietà generali di un'omologia

La maggior parte delle proprietà delle immagini di SOR possono essere espresse attraverso particolari trasformazioni proiettive nel piano, dette *omologie planari* [1, 15, 34]: nella loro forma generale, esse sono caratterizzate dall'avere una linea di punti fissi detta *asse*, più un punto chiamato *vertice* non appartenente all'asse ma fisso anch'esso.

Utilizzando le coordinate omogenee, è possibile esprimere un'omologia attraverso una matrice W di dimensioni 3×3 che trasforma i punti del piano secondo la legge $\mathbf{x}' = W\mathbf{x}$. Algebricamente, i punti fissi dell'asse e del vertice sono ricavati dagli autospazi di W (valendo infatti in coordinate omogenee l'uguaglianza $\mathbf{x} \equiv \lambda\mathbf{x}$, essi devono soddisfare la relazione $W\mathbf{x} = \lambda\mathbf{x}$); nel nostro caso due degli autovalori di W sono uguali, mentre il restante è distinto e reale: l'autospazio generato dagli autovettori corrispondenti ai due autovalori uguali costituisce l'asse; il vertice è invece rappresentato dal rimanente autovettore. Indicando l'asse ed il vertice con i rispettivi vettori omogenei \mathbf{l} e \mathbf{v} , un'omologia può essere parametrizzata secondo

$$W = I + (\mu - 1) \frac{\mathbf{v}\mathbf{l}^\top}{\mathbf{v}^\top \mathbf{l}}, \quad (6)$$

dove I è la matrice identità e μ è il rapporto tra il valore dei due autovalori uguali e quello distinto; la matrice inversa è invece

$$W^{-1} = I + \left(\frac{1}{\mu} - 1 \right) \frac{\mathbf{v}\mathbf{l}^\top}{\mathbf{v}^\top \mathbf{l}}. \quad (7)$$

Dalla parametrizzazione introdotta, si può evincere che un'omologia gode di cinque gradi di libertà: due sia per l'asse che per il vertice (\mathbf{l} e \mathbf{v} sono infatti definiti a meno di una costante moltiplicativa, in quanto vettori omogenei) ed uno dovuto a μ ; sono pertanto necessarie tre coppie di punti $(\mathbf{x}, \mathbf{x}')$ per calcolarla, dato che ogni relazione $\mathbf{x}' = W\mathbf{x}$ introduce due equazioni indipendenti. Nel caso in cui $\mu = -1$, la relativa omologia H viene detta *armonica* e per essa vale $H = H^{-1}$; riducendosi a quattro i gradi di libertà, bastano due coppie di punti per ricavarla. Altre proprietà di un'omologia sono le seguenti (Fig. 11):

- le rette che congiungono coppie di punti corrispondenti $(\mathbf{x}_i, \mathbf{x}'_i)$ si incontrano nel vertice \mathbf{v} ; rette in corrispondenza (passanti cioè una attraverso \mathbf{x}_i e \mathbf{x}_j , l'altra per i rispettivi \mathbf{x}'_i e \mathbf{x}'_j) si intersecano sull'asse \mathbf{l} (in Fig. 11, tale punto è indicato da \mathbf{p}_{12})
- dalla precedente proprietà, per curve legate da una omologia planare le rette tangenti corrispondenti si intersecano sull'asse;
- μ può essere espresso come il *birapporto* (*cross ratio*) definito dai punti $\mathbf{v}, \mathbf{x}, \mathbf{x}'$ ed \mathbf{c} (quest'ultimo dato dall'intersezione della retta \mathbf{r} congiungente $\mathbf{v}, \mathbf{x}, \mathbf{x}'$ con l'asse), avendo fissato un sistema di riferimento omogeneo su \mathbf{r} :

$$\mu = \frac{|\bar{\mathbf{v}} \bar{\mathbf{x}}||\bar{\mathbf{x}}' \bar{\mathbf{c}}|}{|\bar{\mathbf{v}} \bar{\mathbf{x}}'||\bar{\mathbf{x}} \bar{\mathbf{c}}|}$$

dove $\bar{\mathbf{q}}_k$ è il vettore omogeneo sulla retta dei punti indicati⁴ e

$$|\bar{\mathbf{q}}_i \bar{\mathbf{q}}_j| = \det \begin{bmatrix} q_{i1} & q_{j1} \\ q_{i2} & q_{j2} \end{bmatrix}$$

⁴ Se $q_{k2} = 1, k = i, j$ la quantità $|\bar{\mathbf{q}}_i \bar{\mathbf{q}}_j|$ rappresenta la distanza con segno dei punti sulla retta relativi a $\bar{\mathbf{q}}_i$ e $\bar{\mathbf{q}}_j$.

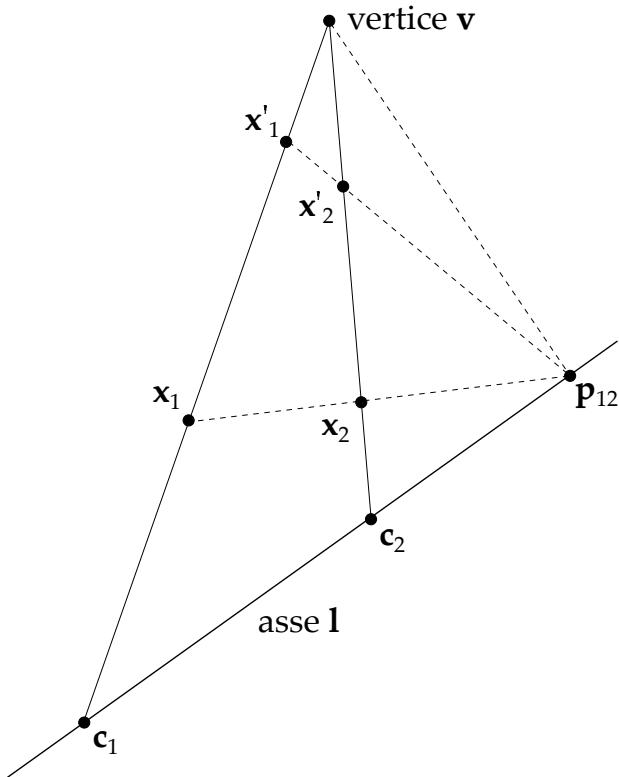


Figura 11. Rette e punti che illustrano alcune proprietà di un'omologia planare.

2.4.2 Omologie nei SOR

Nelle immagini di SOR, si manifestano sia delle omologie generali che delle omologie armoniche [1, 5, 6, 34]; più precisamente:

Proprietà 2.4.1. Ogni coppia di paralleli (i,j) nell'immagine è legata puntualmente da un'omologia W_{ij} , con il vertice v_{ij} posto sulla proiezione dell'asse di simmetria del SOR, l_s , e l'asse l dato dalla proiezione della linea di fuga dei due piani su cui giacciono i due paralleli, indicata con l_∞ ; pur condividendo tutte lo stesso asse, il vertice di ciascuna W_{ij} si trova in un punto differente di l_s ed anche μ è in funzione di i e j .

Proprietà 2.4.2. Ogni coppia di meridiani (i,j) nell'immagine è legata da un'omologia armonica H_{ij} , il cui asse è dato da l_s ed il cui vertice si trova su l_∞ ; differenti H_{ij} condividono lo stesso asse, mentre il vertice si trova dislocato diversamente su l_∞ .

Proprietà 2.4.3. Il contorno apparente viene trasformato in se stesso da un'omologia armonica H , che analogamente alle H_{ij} ha come asse l_s ed il vertice v_∞ posto su l_∞ .

Le proprietà 2.4.1 e 2.4.3 sono alla base del lavoro svolto (in particolare, sono state sfruttate riprendendo quanto esposto in [5, 6]); la proprietà 2.4.2 è stata enunciata per completezza.

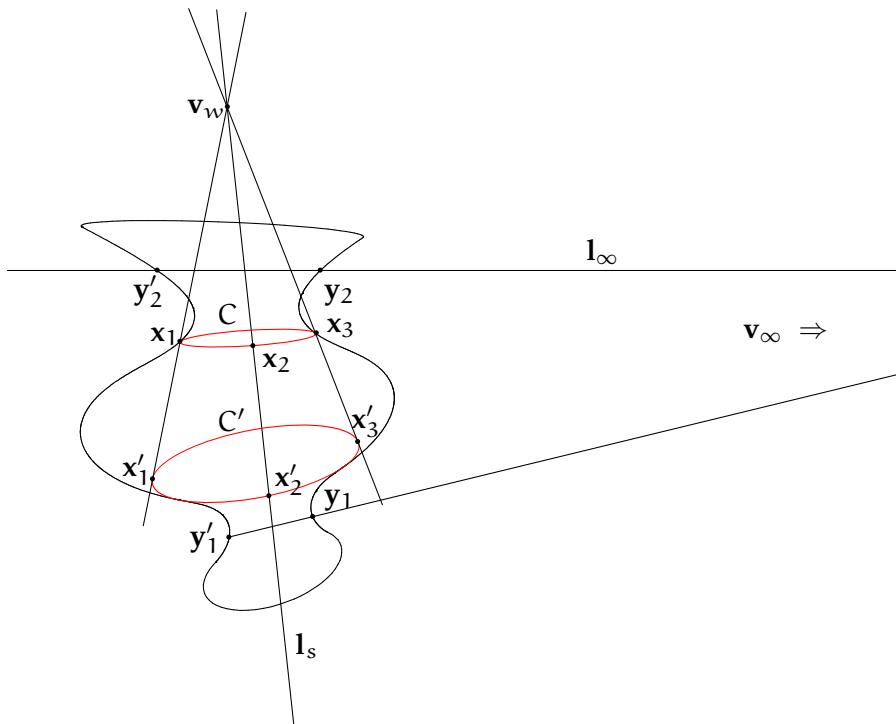


Figura 12. Rette e punti che illustrano le omologie in un SOR.

L'omologia armonica data da H generalizza il concetto di simmetria bilaterale in condizioni prospettive; questo vuol dire che anche i punti nell'immagine di un qualunque parallelo sono associati a coppie da H : in termini algebrici, indicando con C la matrice dell'ellisse relativa ad un parallelo, poiché H è una trasformazione proiettiva deve valere $C = H^{-\top}CH^{-1}$ (da cui segue $C = H^\top CH^{-1}$, essendo $H^{-1} = H^\top$). Dati pertanto due paralleli i e j nell'immagine, le rispettive ellissi descritte da C_i e C_j sono legate da W_{ij} tramite la relazione $C_i = W_{ij}^{-\top}C_jW_{ij}^{-1}$ ed entrambe si trasformano in se stesse secondo H .

Fig. 12 illustra sinteticamente le precedenti proprietà che coinvolgono le W_{ij} e H : i punti sui due paralleli x_1, x_2, x_3 e x'_1, x'_2, x'_3 si corrispondono secondo la relativa W_{ij} , il cui vertice v_W si trova su l_s ; inoltre x_1, x'_1, x_2, x'_2 vengono trasformati da H rispettivamente in x_3, x'_3, x_2, x'_2 . Anche i punti y_i, y'_i sui due lati del contorno apparente si corrispondono secondo H ; le linee $y_i \times y'_i$ che congiungono le coppie così formate si incontrano tutte in v_∞ .

Le proprietà precedenti sono completate dalle seguenti:

Proprietà 2.4.4. Le rette tangenti a due differenti paralleli (i, j) nell'immagine in punti che si corrispondono secondo W_{ij} hanno lo stesso punto di fuga posto su l_∞ .

Proprietà 2.4.5. I punti su due differenti paralleli (i, j) che si corrispondono nell'immagine sotto W_{ij} si trovano sullo stesso meridiano del SOR.

Infine, anche se il suo enunciato non coinvolge direttamente alcuna omologia, l'elenco delle proprietà viene esteso osservando che

Proprietà 2.4.6. Il contorno apparente è tangente alle ellissi dei paralleli nei punti di contatto.

Quest'ultima proprietà è alla base della tecnica adottata per riconoscere le ellissi dei paralleli visibili nell'immagine, come verrà chiarito più avanti.

2.5 STIMA DELL'OMOLOGIA ARMONICA E SEGMENTAZIONE

Durante l'esposizione della visione d'insieme dell'algoritmo di segmentazione, abbiamo anticipato che il primo passo da compiere sarebbe stato quello di stimare l'omologia armonica che lega i due lati del contorno apparente, e tutte le curve concordi con essa.

Abbiamo detto inoltre che, in condizioni particolari, l'omologia armonica si riduce ad una simmetria assiale: questo si verifica quando poniamo il vertice v_∞ all'infinito, nella direzione delle rette perpendicolari all'asse l_s ; in una foto reale di un SOR invece, v_∞ al finito, anche se molto distante da l_s (Fig. 13). Possiamo ricreare le condizioni di simmetria assiale se generiamo una versione a bassa risoluzione dell'immagine di partenza: la perdita dei dettagli si ripercuote infatti anche sulla silhouette del SOR, che assume un aspetto praticamente simmetrico. Inoltre in un'immagine a bassa risoluzione gli edge estraibili si riducono così come anche i pixel che li costituiscono.

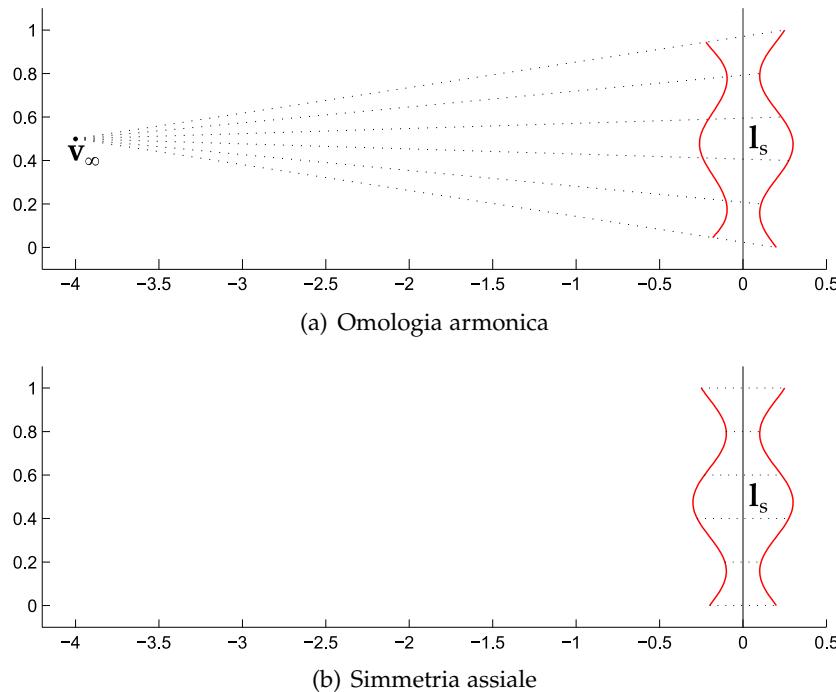


Figura 13. Omologia armonica (a) e simmetria assiale (b) a confronto.

Ricordando la parametrizzazione di Eq. 6 e ponendo $\mu = -1$, la matrice che descrive un'omologia armonica è data da

$$H(v_\infty, l_s) = I - 2 \frac{v_\infty l_s^\top}{v_\infty^\top l_s}; \quad (8)$$

i suoi quattro gradi di libertà sono dati da v_∞ e l_s , che ne possiedono due ciascuno.

Analizziamo adesso i gradi di libertà di una simmetria assiale: dai quattro di cui gode un'omologia armonica, ne dobbiamo eliminare due, perché vincoliamo la posizione di v_∞ in funzione di l_s : se $l_s = [a, b, c]^\top$, si ha una simmetria assiale solo se $v_\infty = [a, b, 0]^\top$; pertanto i gradi di libertà risiedono tutti in l_s . La diretta conseguenza è che la rappresentazione minima di una simmetria assiale è costituita da una sola coppia di punti in accordo ad essa: l'asse di simmetria si trova infatti diretto perpendicolarmente al segmento che unisce i due punti e passante per il suo punto di mezzo.

2.5.1 Inizializzazione con RANSAC

Tutte queste considerazioni (rappresentazione minima, spazio di ricerca contenuto) forniscono validi argomenti alla scelta di RANSAC come algoritmo di stima robusta; tuttavia quello che noi stiamo cercando non è una simmetria a bassa risoluzione, bensì l'omologia armonica nell'immagine originale. Occorre recuperare l'informazione persa: questo può essere fatto tramite una rappresentazione multirisoluzione della stessa foto attraverso cui propagare le stime di l_s e v_∞ ; si può pensare a questa rappresentazione come ad una piramide la cui base è costituita dall'immagine originale, mentre il vertice è dato dalla sua versione più grossolana (Fig. 14). Questa analogia non è casuale: il procedimento adottato è ripreso infatti da quanto esposto in [3], in cui l'immagine di ogni piano della piramide ha dimensioni lineari pari alla metà di quella del piano immediatamente inferiore⁵.

In definitiva, possiamo usare il campionamento casuale eseguito secondo RANSAC per ricercare attraverso esso la simmetria assiale che, nella foto a bassa risoluzione, ha maggior consenso (Fig. 15); questa costituisce il punto di partenza per la stima dell'omologia armonica a bassa risoluzione. Il suo raffinamento avviene propagandola scendendo i livelli della piramide: ad ogni livello è eseguita una minimizzazione che sfrutta la stima del livello precedente come inizializzazione. Il passaggio da simmetria assiale ad omologia armonica è eseguito

⁵ Se il vertice della piramide è al livello n , è possibile dimezzare di volta in volta le dimensioni solo se quelle originali sono multipli di 2^n ; se questo non è possibile, l'immagine di partenza viene orlata con i suoi pixel più esterni, fino al raggiungimento delle dimensioni minime necessarie. Da un punto di vista realizzativo, si è posto un limite alle dimensioni sotto cui non può scendere l'immagine al vertice della piramide: 50 pixel lungo il lato più corto.

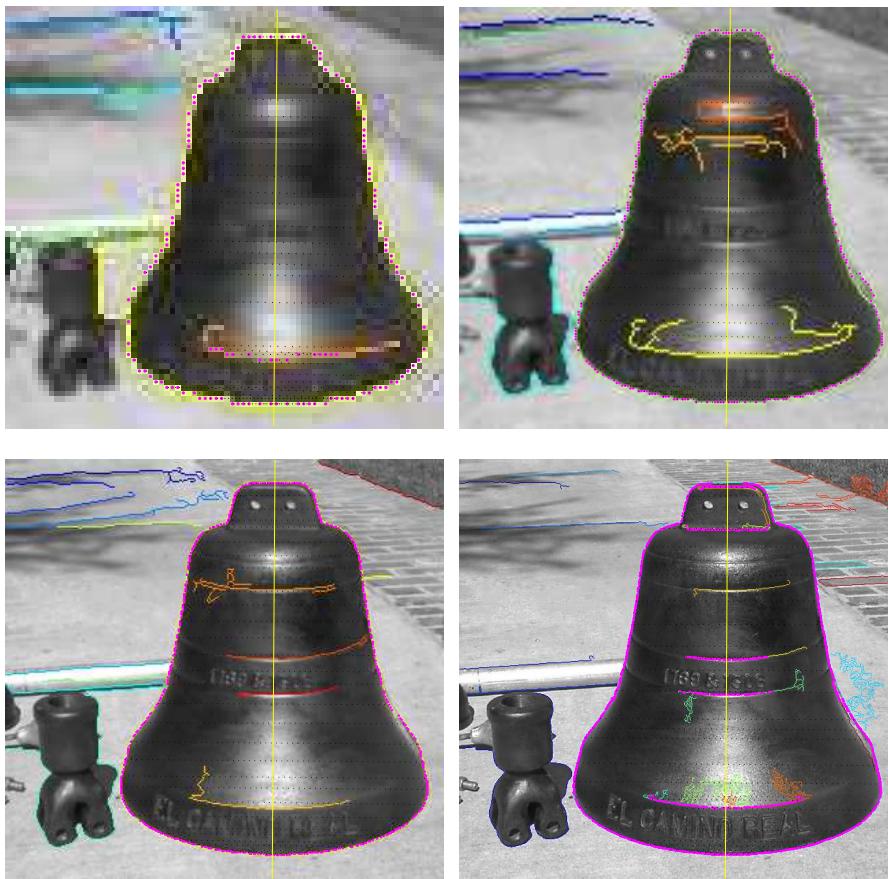


Figura 14. Propagazione della stima di H : ciascuna immagine è in realtà la metà della sua successiva; la foto in alto a sinistra è il vertice della piramide, quella in basso a destra ne è la base. Su ciascuna foto sono stati sovrapposti, a colori, gli edge estratti; i punti evidenziati sono quelli in corrispondenza secondo H .

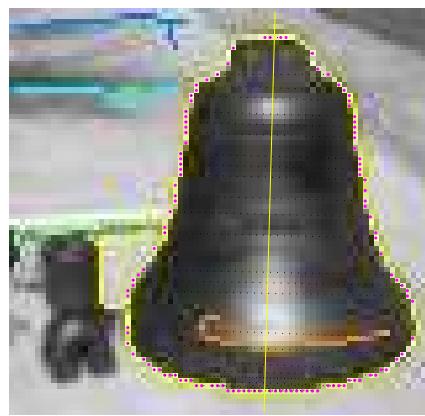


Figura 15. Simmetria a massimo consenso trovata con RANSAC; si noti il miglioramento in Fig. 14, in alto a sinistra, di H prodotto sull'immagine alla stessa risoluzione dopo la successiva fase di estensione a omologia armonica.

tramite lo stesso meccanismo di minimizzazione, che sarà esposto nella sezione 2.5.2.

La ricerca del massimo insieme di consenso avviene, come noto, per tentativi; dopo un numero sufficientemente elevato di prove, si seleziona l'insieme con la massima cardinalità. Per la determinazione del numero di tentativi, abbiamo assunto una posizione molto conservativa: in riferimento a Eq. 2

$$N = \log(1 - p) / \log(1 - w^s),$$

la percentuale w di *inlier* viene calcolata come il rapporto tra la cardinalità dell'insieme di consenso più numeroso e tutte le coppie possibili (dati n punti, esse sono $n^2 - n$); il valore per s è quindi 1 (per p si è scelto 0.99). Le coppie vengono determinate osservando, per ogni punto trasformato secondo la simmetria corrente, il punto ad esso più vicino: se la loro distanza è inferiore ad 1 pixel, la coppia costituisce un elemento dell'insieme di supporto.

Al fine di ridurre lo spazio di ricerca, viene eseguita una selezione sugli edge estratti direttamente con Canny, in base alla loro lunghezza e momento di inerzia: si tengono solo le prime n curve più lunghe e da queste si selezionano le m con il massimo momento d'inerzia più grande ($n = 20$, $m = 5$). Lo scopo è quello di eliminare edge inutili dovuti generalmente alle decorazioni presenti nella texture del SOR, se compatte e di piccole dimensioni⁶.

2.5.2 Raffinamento dell'omologia

La fase di raffinamento si differenzia per alcuni aspetti da quella del RANSAC classico: i due passi di minimizzazione e classificazione non si riferiscono più al modello utilizzato per l'insieme di massimo consenso (la simmetria assiale), ma alla sua generalizzazione (l'omologia armonica); dall'immagine a bassa risoluzione, in cui è stato trovato l'insieme di massimo consenso, si deve risalire all'immagine originale; la soglia con cui determinare le coppie *inlier* viene determinata adattivamente attraverso un'analisi statistica dei residui; infine, è stato tenuto conto del fatto che i punti conformi al modello non sono punti sparsi ma appartengono a delle curve, inglobando nella fase di classificazione un meccanismo di strutturazione di tali punti per l'estrazione di curve salienti.

Minimizzazione e propagazione

Vediamo innanzitutto le differenze minori rispetto allo schema precedente, cominciando dalla fase di minimizzazione; la classificazione è

⁶ Questa preselezione viene effettuata anche negli altri livelli della piramide (le soglie n e m vengono incrementate mano a mano che aumenta la definizione delle immagini.); lo scopo è sempre quello di eliminare dai dati una parte dei distrattori, per favorire stavolta la propagazione delle stime di v_∞ e l_s .

trattata a parte, nella sezione successiva. La funzione di costo scelta è l'errore di trasferimento simmetrico (Eq. 1); tenendo conto che per un'omologia armonica vale $H^{-1} = H$, la funzione è

$$f(\mathbf{v}_\infty, \mathbf{l}_s) = \sum_i d(\mathbf{x}_i, H(\mathbf{v}_\infty, \mathbf{l}_s)\mathbf{x}'_i)^2 + d(\mathbf{x}'_i, H(\mathbf{v}_\infty, \mathbf{l}_s)\mathbf{x}_i)^2. \quad (9)$$

La parametrizzazione non è minima: coinvolge infatti tutte e tre le componenti dei vettori \mathbf{v}_∞ e \mathbf{l}_s , senza tenere conto che in realtà ciascuno di essi è noto a meno di un fattore moltiplicativo.

Si tratta di una minimizzazione non lineare, che viene affrontata con l'algoritmo iterativo di Levenberg-Marquardt; pertanto occorre definire le stime iniziali di \mathbf{v}_∞ e \mathbf{l}_s da cui partire. La simmetria assiale con il massimo consenso fornisce il punto di partenza: poiché nella minimizzazione le componenti di \mathbf{v}_∞ sono libere, al termine dei cicli di classificazione e minimizzazione otteniamo la stima dell'omologia alla risoluzione più grezza. La propagazione dell'omologia alle risoluzioni maggiori avviene riportando in ogni livello della piramide le stime di \mathbf{v}_∞ e \mathbf{l}_s provenienti dal livello precedente: questi costituiscono il punto di partenza per una nuova fase di perfezionamento, che viene eseguita in ciascun livello.

I cicli di classificazione e minimizzazione per ogni livello l sono ripetuti fino a che, tra il passo k e $k+1$, le seguenti variazioni percentuali

$$\left\{ \begin{array}{l} \frac{|\mathbf{v}_\infty(k) - \mathbf{v}_\infty(k+1)|}{|\mathbf{v}_\infty(k)|} \\ \frac{|\mathbf{l}_s(k) - \mathbf{l}_s(k+1)|}{|\mathbf{l}_s(k)|} \end{array} \right.$$

non sono entrambe inferiori a 10^{-4} , oppure k ha raggiunto quota 100 iterazioni.

La classificazione degli inlier

Se si tiene conto del fatto che le coppie in accordo all'omologia H giacciono su delle curve, la combinazione ripetuta delle due fasi di classificazione e minimizzazione ricorda molto l'algoritmo ICP, *Iterative Closest Point*, descritto in [37]. Il problema affrontato in tale articolo riguarda la stima di una trasformazione euclidea rigida (rotazione + traslazione) tra le curve di due differenti immagini dello stesso soggetto, nell'ipotesi di piccoli spostamenti. ICP tratta curve di forma arbitraria direttamente attraverso i punti che le definiscono nell'immagine, senza l'utilizzo di particolari primitive. L'idea è quella di minimizzare, in maniera iterativa, la distanza tra le curve di una delle due immagini (diciamo la seconda) e le trasformazioni dei punti relativi alla stessa curva nell'altra immagine: la conoscenza solo puntuale delle curve obbliga l'algoritmo a ricercare, per ogni punto trasformato, il campione sulla seconda curva a lui più vicino;

da qui il nome del metodo. Nel nostro caso le due immagini sono in realtà le due porzioni di piano individuate da l_s , mentre la trasformazione euclidea è sostituita dall'omologia. Si tratta pertanto di una generalizzazione del metodo originale.

L'algoritmo in [37] presenta anche un metodo per rimuovere le coppie erronee attraverso un'analisi dei residui (la distanza tra il punto trasformato e quello associato) e degli angoli che formano le rette tangenti alle curve nei rispettivi punti di origine.

L'analisi dei residui avviene tramite il confronto con una soglia di ammissione che viene spostata adattativamente⁷. Il metodo proposto per la taratura di questa soglia è però contorto e viene nel nostro caso sostituito da LMedS che, nel caso monodimensionale, si comporta come uno stimatore (perdipiù deterministico) della moda m_d e della varianza v_r della distribuzione in esame; in particolare la soglia di confronto viene calcolata come

$$d_s = m_d + 6v_r.$$

Per quanto riguarda l'analisi sulle rette tangenti, si è proceduto come segue. Data una corrispondenza puntuale $x' = Hx$, se le rette l' e l sono le rispettive rette tangenti agli edge, esse si devono trasformare secondo $l' = H^{-T}l$: l'angolo che viene misurato è quindi, nel nostro caso, quello tra l' e $H^{-T}l$. La stima della retta tangente ad un edge in un suo punto è ottenuta dal gradiente dell'intensità dell'immagine⁸. Fig. 16 illustra tutte le coppie tutte le coppie (x, x') trovate in relazione secondo H ; l'istogramma mostra inoltre la distribuzione delle distanze tra il punto Hx e quello a più prossimo x' . In Fig. 17 è invece riportato l'effetto di sfoltimento causato dal test sulle rette tangenti agli edge nei punti precedentemente accoppiati. Per finire, Fig. 18 mostra l'azione della soglia imposta da LMedS.

La combinazione della multirisoluzione con uno metodo di regressione robusto, quale LMedS, permette di regolarizzare la superficie dell'errore di trasferimento simmetrico; tuttavia le simmetrie casuali, dovute a particolari motivi sulla texture del SOR o sullo sfondo della scena, agendo come *distrattori* possono bloccare l'intera strategia in minimi locali. Le coppie in corrispondenza devono però formare delle curve: un raggruppamento dei punti in curve, da cui scartare le meno salienti (dovute in genere alle simmetrie accidentali) permette un'ulteriore eliminazione di outlier, favorendo la convergenza verso il minimo globale. Il lavoro in [7, 8] sul minimo albero ricoprente (MST), esposto nella sezione 1.3.2, è adatto al nostro scopo: la complessità computazionale limitata ($\Theta(n \log n)$ per un numero di punti n) permette di inserire la costruzione di uno MST euclideo all'interno della fase di classificazione. La sua frammentazione in curve avviene attraverso la determinazione dei punti di diramazione, eseguendo una

⁷ Un'altra variante propone di inserire uno stimatore M nella funzione di costo [9].

⁸ La stima è grossolana, per cui si è deciso di testare gli angoli con una soglia fissa piuttosto elevata ($\pm 15^\circ$).

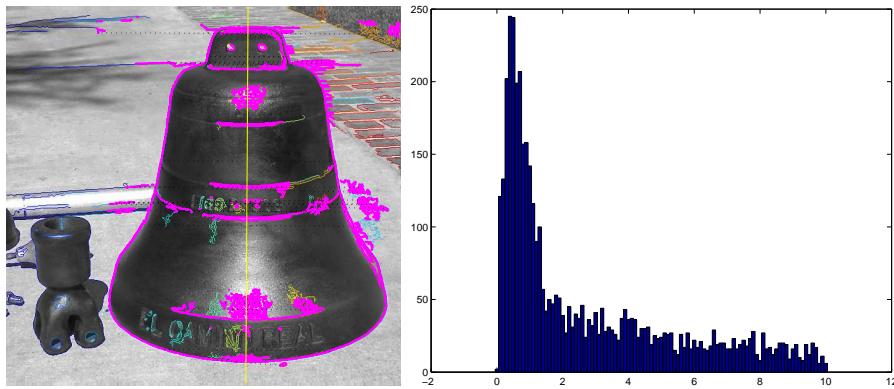


Figura 16. A sinistra, tutte le coppie (x, x') per le quali x' è il punto più vicino a Hx , in un suo intorno di raggio 10 pixel. A destra, l'istogramma mostra la distribuzione delle distanze, valutata per ogni coppia, tra x' e Hx .

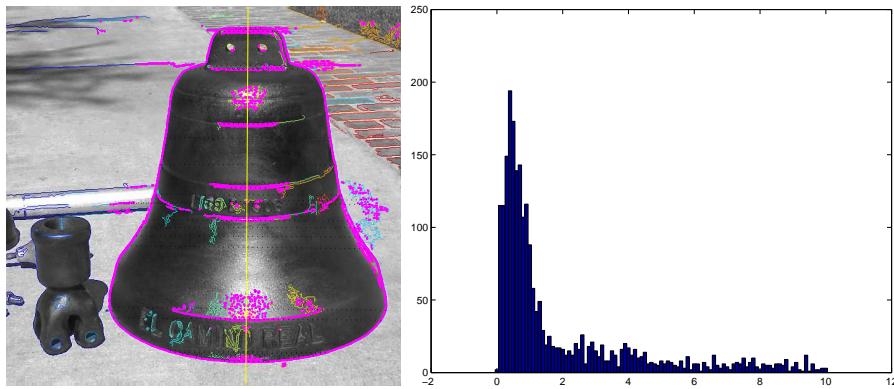


Figura 17. A sinistra, le coppie (x, x') rimaste dopo il test sulle rette tangenti. A destra, la distribuzione delle distanze all'interno di ciascuna coppia.

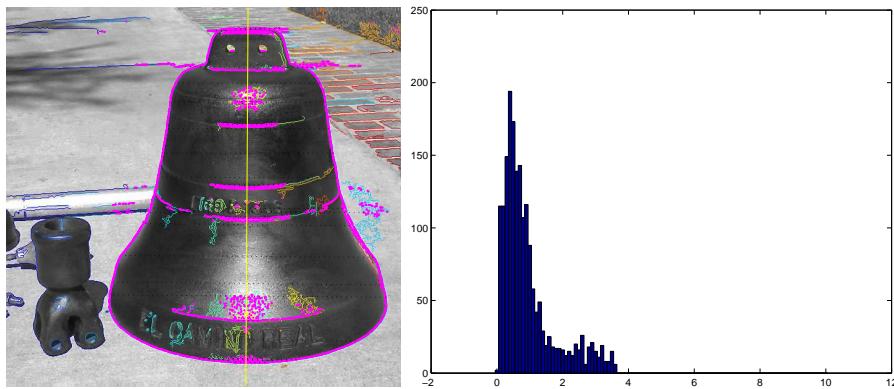


Figura 18. A sinistra, le coppie (x, x') rimaste dopo il taglio eseguito da LMedS. A destra, la distribuzione delle distanze all'interno di ciascuna coppia.

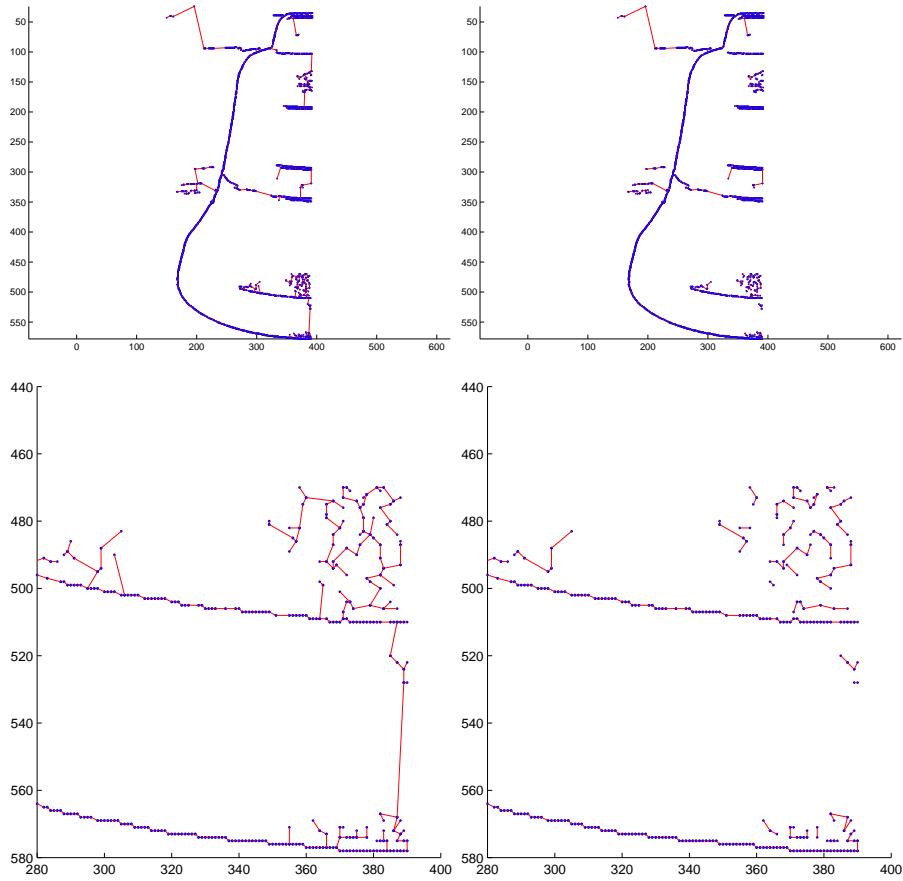


Figura 19. In alto a sinistra lo MST; a destra il grafo sconnesso dopo l'eliminazione dei ponti. Particolari di entrambi in basso.

visita depth-first dell’albero; per ogni punto di diramazione vengono rimossi tutti i ponti, lasciando solo i due archi con peso minimo.

Allo scopo di assicurarsi l’estrazione di curve simmetriche, oltre che per ridurre il carico computazionale, MST è calcolato solo su uno dei due semipiani (quello che contiene v_∞); una volta selezionate le curve più salienti, vengono ricreate le corrispondenze con l’altro semipiano. Inoltre, viene operato un campionamento casuale delle coppie in accordo all’omologia corrente: se inquadriamo l’estrazione delle curve salienti nella strategia globale, oltre a ridurre ulteriormente il numero di punti n su cui calcolare MST, ciò fornisce una via d’uscita dai minimi locali. Fig. 19 illustra sia lo MST istanziato sui punti relativi a Fig. 17, sia la sua frammentazione in tratti separati.

In alternativa alla distanza euclidea tra i nodi, è stato provato anche un altro sistema di pesi, per cercare di tener conto (seppur in maniera naïf) anche del criterio di buona continuità, oltre a quello di prossimità. Consideriamo la possibilità che un arco tra due punti sia giusto: si può ipotizzare che più i due punti sono vicini, più tale valore aumenta; inoltre, più le rette tangenti agli edge sono allineate, più l’arco è allineato con esse e maggiore è la probabilità che i due

nodi vadano collegati. Pertanto, un peso dell'arco (i, j) che ne indichi la "bontà" può essere

$$bnt(i, j) = \frac{1}{\text{dist}(i, j)} |\mathbf{t}_i^\top \mathbf{t}_j|,$$

dove \mathbf{t}_i e \mathbf{t}_j sono i versori tangenti nei nodi i e j ai rispettivi edge di appartenenza. Entrambi i termini sono compresi tra 0 e 1 ($\text{dist}(i, j) \geq 1$, in quanto i e j hanno coordinate intere; $\mathbf{t}_i^\top \mathbf{t}_j$ è invece un prodotto scalare tra versori): possiamo associare a bnt un significato probalistico e ricercare un albero ricoprente che massimizzi

$$\prod bnt(i, j).$$

Applicando il logaritmo a questa nuova funzione di costo e cambiando segno, ci possiamo però ricondurre alla determinazione di un minimo albero ricoprente standard con pesi $p(i, j)$ sugli archi dati da

$$p(i, j) = -\log(bnt(i, j)). \quad (10)$$

Da un punto di vista implementativo, la variante allo MST euclideo viene determinata sul grafo di Delaunay con i nuovi pesi; non sono stati analizzati altri grafi di prossimità.

Una volta frammentato l'albero in curve semplici $\{C_k\}$, a ciascuna di esse viene assegnato un valore di salienza: le curve che stiamo cercando sono in genere lunghe e dense, mentre le curve create dalla connessione di punti che soddisfano casualmente l'omologia sono di solito brevi e con pochi punti; per queste considerazioni, la salienza è valutata come

$$\text{Sal}(C_k) = \frac{s_k}{\frac{1}{s_k-1} \sum_i^{s_k-1} \text{dist}(x_i, x_{i+1})}, \quad (11)$$

dove s_k denota il numero di punti di MST che sono stati associati alla curva C_k e $\text{dist}(x_i, x_{i+1})$ è la distanza tra due punti consecutivi. Questa misura assegna valori bassi alle curve con pochi elementi e frammentate, molto alti a quelle lunghe e dense. Nell'ipotesi che quest'ultime siano poche rispetto alle prime, possiamo applicare nuovamente LMedS sulla distribuzione di $\text{Sal}(C_k)$, andando stavolta a scartare tutte le curve con punteggio inferiore alla soglia. L'estrazione delle curve più salienti è illustrata in Fig. 20.

Le performance del clustering basato sullo MST euclideo e di quello con MST modificato sono simili: quello modificato funziona, forse, leggermente meglio, anche se entrambi non sono esenti da errori.

Riassumendo, il procedimento per la stima dell'omologia ed il raggruppamento in curve è descritto in Alg. 2.

Classificazione finale

Al termine del procedimento di stima si esegue un'ultima classificazione su tutti gli edge estratti con Canny, reintegrando anche quelli

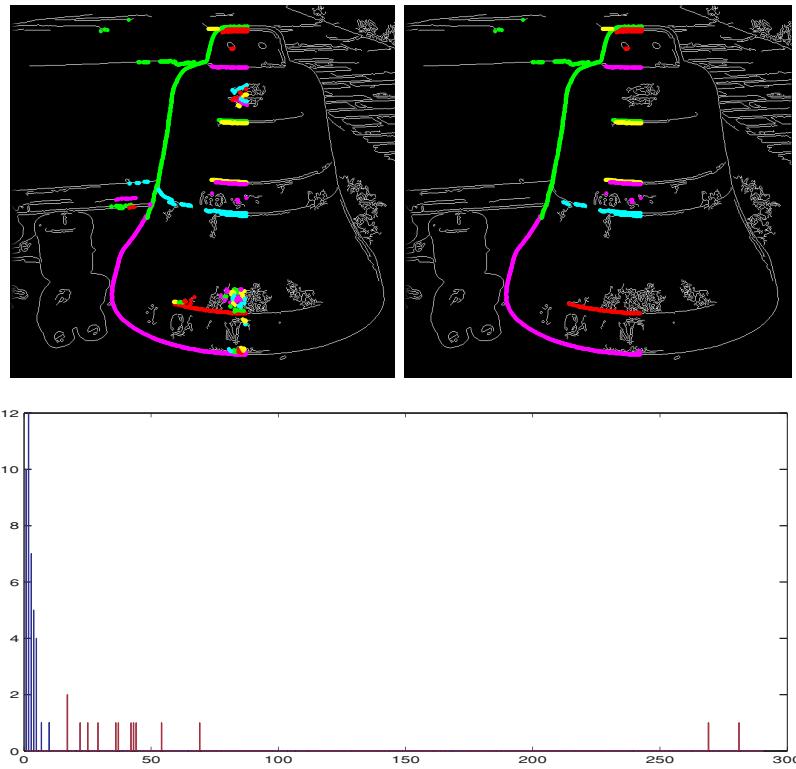


Figura 20. In alto, l'insieme di tratti prima e dopo il taglio sulla salienza. In basso, l'istogramma della salienza: in rosso le salienze selezionate.

Algoritmo 2. Stima dell'omologia armonica

- 1 Crea la piramide di immagini $\{P_l\}$, $l = 0 \dots L$ ed estrai gli edge ad ogni livello.
 - 2 Per $l=0$ (risoluzione minima) calcola $H^{(0)}$ su P_0 con RANSAC.
 - 3 Per l'immagine P_l , ripeti fino a convergenza:
 - trasforma ogni punto x_i degli edge di P_l con $H^{(l)}$ e forma la coppia (x_i, x'_i) con il punto x'_i più vicino a $H^{(l)}x_i$.
 - Scarta le coppie le cui tangenti non soddisfano $(H^{(l)})^{-T}$
 - Trova le coppie inlier (LMedS su istogramma delle distanze)
 - Calcola MST sui punti del semipiano che contiene v_∞ .
 - Rimuovi i ponti da MST per creare l'insieme di curve $\{C_k\}$
 - Estrai le curve più significative dall'istogramma di salienza, cioè quelle con valori superiori alla soglia data da LMedS.
 - Aggiorna l'omologia secondo $H^{(l)} \leftarrow \arg \min_{H^{(l)}} f(v_\infty, I_s)$ tramite ottimizzazione non lineare (Levemberg-Marquardt).
 - 4 Se $l < L$, ponì $l \leftarrow l + 1$, ricava $H^{(l)}$ da $H^{(l-1)}$ e vai a 3.
-

rimossi in base alla loro lunghezza e momento d'inerzia, nel caso fossero stati eliminati eventuali frammenti di edge appartenenti al contorno apparente o alle ellissi del SOR.

Infine, le curve estratte $\{C_k\}$ (e le loro controparti $\{C'_k\}$ secondo H) sono interpolate con delle smoothing spline $(x(t), y(t))$ [31], necessarie per i due stadi successivi. Gli “istanti” t_i , relativi ai punti x_i di ciascuna curva, sono stabiliti scegliendo un verso di percorrenza della catena di punti, da un suo estremo all’altro: a quello di partenza viene assegnato $t = 0$; gli altri istanti sono generati con incrementi Δt pari alla distanza che separa ciascun punto dal suo successivo. I punti x_i non vengono accantonati bensì costituiscono i *punti campione* della curva (Fig. 21).

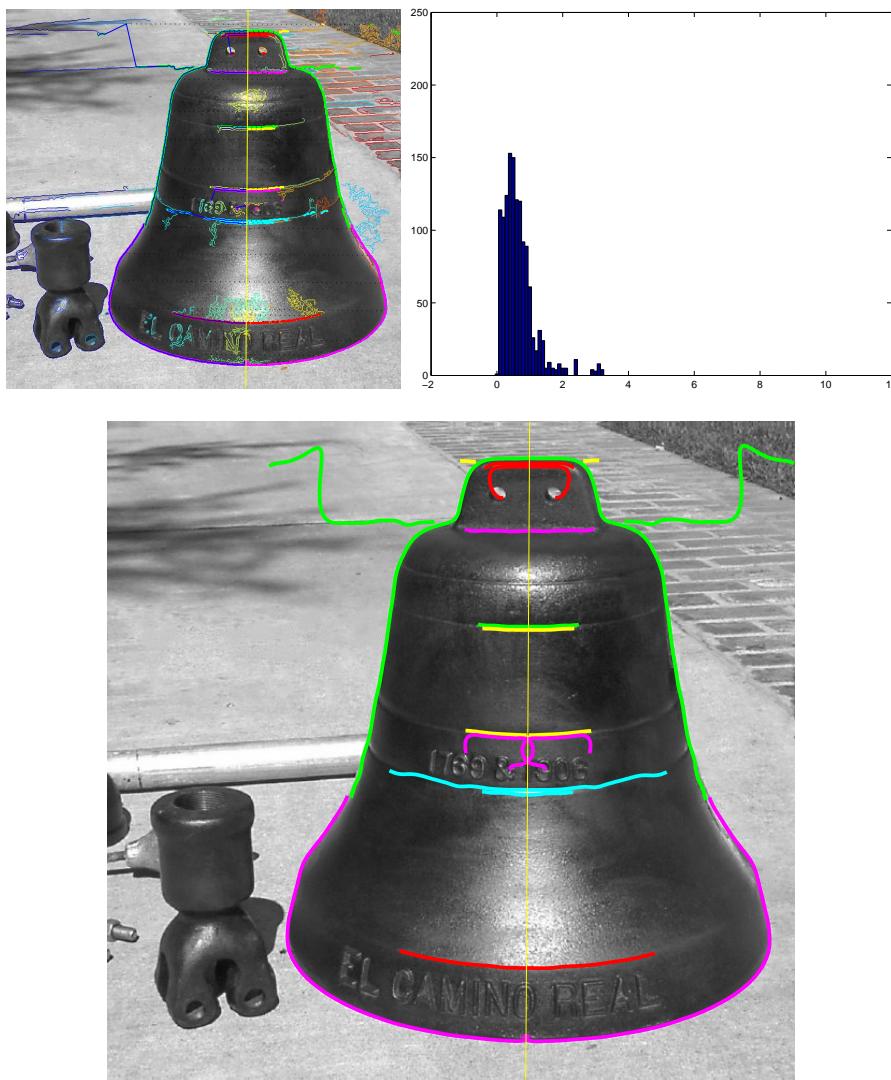


Figura 21. In alto a sinistra, le curve sono ancora sotto forma di catene; a destra l’istogramma finale delle distanze. Sotto, le curve interpolano i tratti con smoothing spline; i punti delle catene diventano i *punti campione* delle curve.

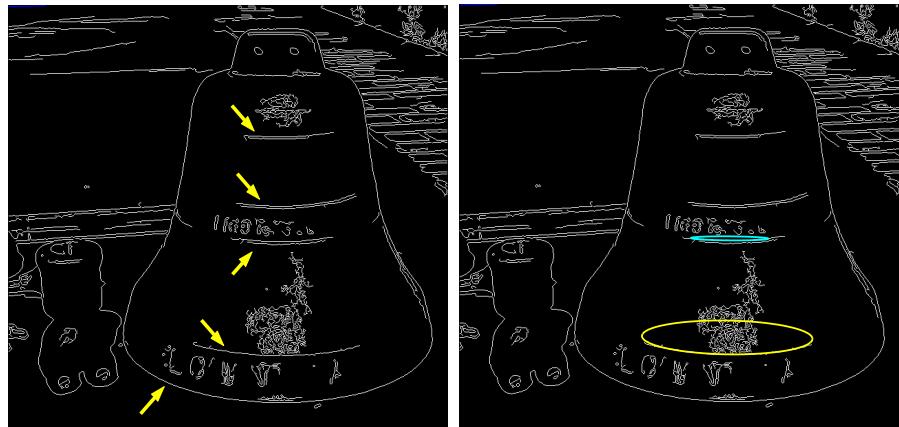


Figura 22. La stima delle ellissi a partire dai loro segmenti visibili crea delle difficoltà ai classici stimatori algebrici [11] (nella figura di destra, due ellissi mal stimate).

2.6 CLASSIFICAZIONE DELLE CURVE

Le ellissi dovute ai paralleli di un SOR giocano un ruolo essenziale per la sua ricostruzione metrica: è necessario che la loro stima sia la migliore possibile. Se osserviamo i segmenti ellittici sulla campana in Fig. 22, vediamo come essi siano dei tratti con poca curvatura: in queste situazioni le misure a disposizione sono di scarsa utilità per i classici metodi per la stima delle ellissi [11, 38]. Nel nostro caso, tuttavia, le ellissi derivano da un modello ben preciso (il SOR) e devono soddisfare precise proprietà: questa informazione addizionale guida sia l'estrazione delle ellissi dalle curve fornite dal precedente stadio di segmentazione, sia la classificazione delle altre curve in contorno apparente e sporco.

La classificazione delle curve si svolge in tre livelli: il primo (sezione 2.6.1) prevede l'estrazione delle ellissi con una stima grossolana dei loro coefficienti; il secondo (sezione 2.6.2) consiste nel perfezionamento di tali equazioni; il terzo (sezione 2.6.3) infine, opera una selezione su tutte le ellissi trovate e classifica le curve tangenti ad esse come parti di contorno apparente.

2.6.1 Parametrizzazione ed estrazione delle ellissi

Come già anticipato, la proprietà alla base dell'estrazione delle ellissi riguarda la condizione di tangenza tra l'immagine di un qualsiasi parallelo ed i due lati del contorno apparente (proprietà 2.4.6).

La chiave di volta per descrivere questa proprietà in maniera compatta e mediante grandezze misurabili sull'immagine, è la parametrizzazione del fascio di ellissi passanti per due punti e tangentì in ognuno di essi ad una retta data [2]; in coordinate omogenee ciò è

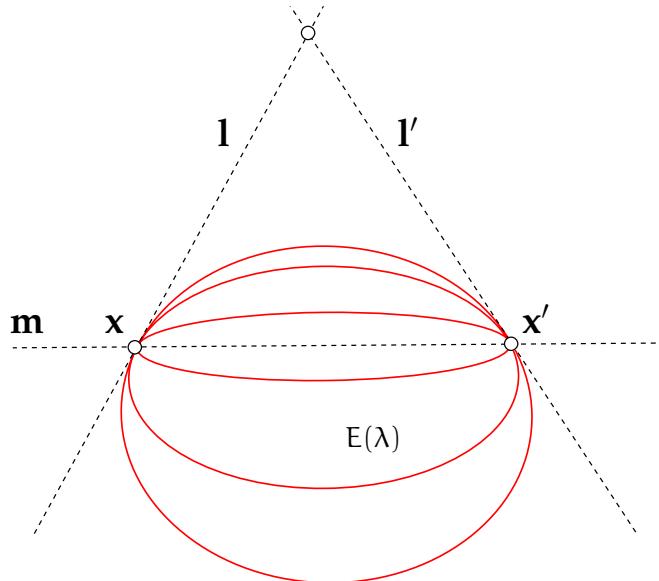


Figura 23. Fascio di ellissi $E(\lambda)$ generato da x, x', l, l' .

esprimibile come:

Proprietà 2.6.1. Dati due punti x e x' e due rette l e l' passanti per essi, tutte le ellissi tangenti a l , l' e passanti per x , x' sono generate dal fascio

$$E(\lambda) = \mathbf{m}\mathbf{m}^T + \lambda(l'l'^T + l'l^T),$$

dove $\mathbf{m} = x \times x'$ è la retta che congiunge x e x' .

$E(\lambda)$ è la matrice omogenea che descrive l'ellisse: per $\lambda = 0$ l'ellisse degenera nella retta \mathbf{m} ; se si ha la cura di orientare l e l' nello stesso verso, le ellissi sono istanziate solo per λ con valori positivi, mentre per $\lambda < 0$ vengono generate delle iperboli. All'aumentare di λ , le ellissi crescono di dimensioni fino a quando, per un dato λ_0 si forma una parabola; per $\lambda > \lambda_0$ si ottengono nuovamente delle iperboli. Il valore di λ_0 deriva da una condizione che deve essere soddisfatta dai coefficienti di E , affinché la conica da essa rappresentata sia un'ellisse: se infatti da $E(\lambda)$ ricaviamo l'equazione algebrica di una conica

$$ax^2 + 2bxy + cy^2 + 2dx + 2ey + f = 0,$$

per avere un'ellisse deve essere $ac - b^2 > 0$.

Fig. 23 illustra alcune ellissi del fascio all'aumentare di λ , per le grandezze x, x', l, l' in essa istanziate.

Questa parametrizzazione ha permesso lo sviluppo di due procedure per l'estrazione delle ellissi corrette. Le procedure sono due perché due sono i tipi di ellissi visibili nell'immagine di un SOR: nella terminologia adottata nella sezione 2.1 esse si dividono in bordi (i cui tratti costituiscono parte della sagoma del SOR) e sezioni trasversali (visibili solo all'interno della silhouette). La prima ad essere analizzata è la procedura per le sezioni trasversali; quella per i bordi, mostrata successivamente ne è, per certi versi, una specializzazione.

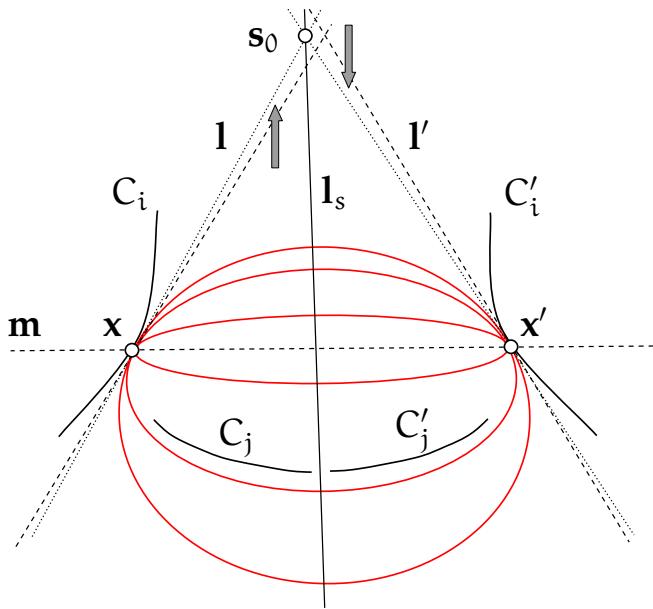


Figura 24. Correzione delle tangenti: le rette tangenti a C_i e C'_i vengono fatte passare per s_0 . Si ricerca l'ellisse che descrive la curva C_j e la sua controparte C'_j .

Estrazione delle sezioni trasversali

L'estrazione delle sezioni trasversali avviene per mezzo di una procedura a voto, simile alla Hough Transform. Se una curva C_k e la sua controparte secondo H , C'_k , costituiscono una parte di contorno apparente, allora, dati i punti x e x' in relazione su di esse, la sezione trasversale che contiene i due punti x e x' è un'ellisse tra quelle in $E(\lambda)$; ovviamente l e l' devono essere le rette tangenti a C_k , C'_k rispettivamente in x e x' . Per ogni ellisse, le rette l e l' sono ottenute derivando nei punti di contatto x e x' le spline dei due tratti di contorno: le loro equazioni però non sono legate solo dalla corrispondenza tra i punti di contatto, ma deve valere anche $l' = H^{-T}l$. Nella pratica questa relazione è soddisfatta solo in maniera approssimata, avendo derivato le due curve separatamente; per soddisfare questa relazione, si preferisce allentare il vincolo di tangenza tra l'ellisse e la spline, correggendo il coefficiente angolare delle due tangenti. Se la corrispondenza tra l e l' fosse esatta, esse intersecherebbero l_s nello stesso punto: modifichiamo quindi l'angolazione delle due rette imponendo il passaggio di entrambe per il punto sull'asse che si trova a metà strada tra le due intersezioni (Fig. 24).

L'individuazione di una sezione trasversale avviene quindi attraverso la determinazione di una curva C_i con opportuni x e λ (x' è funzione di x , $x' = Hx$) che trovino un adeguato consenso sui punti di un'altra curva C_j : la prima verrà accreditata come contorno apparente, l'altra come sezione trasversale. Il consenso sulla curva C_j viene verificato considerando entrambi i lati, a destra e a sinistra dell'asse

I_s , che la compongono: nel seguito, quando sarà fatto riferimento ad una curva C_j nell'accezione di sezione trasversale, la notazione C_j sarà da interdersi estesa anche alla controparte C'_j .

Singole istanze di un dato fascio $E(\lambda)$ sono generabili imponendo il passaggio per un punto \mathbf{p} , diverso da \mathbf{x} e \mathbf{x}' ; basta porre

$$\lambda_{\mathbf{p}} = -\frac{\mathbf{p}^\top (\mathbf{m}\mathbf{m}^\top)\mathbf{p}}{\mathbf{p}^\top (\mathbf{l}\mathbf{l}'^\top + \mathbf{l}'\mathbf{l}^\top)\mathbf{p}}, \quad (12)$$

verificando se $\lambda_{\mathbf{p}} \in (0, \lambda_0)$.

La procedura consiste in una ricerca esaustiva in cui una curva $C_i = (x_i(t), y_i(t))$ viene ritenuta un'ipotetica parte di contorno apparente; essa viene percorsa in tutta la sua lunghezza a piccoli passi⁹ (inducendo un movimento analogo anche lungo la controparte C'_i) e per ogni sosta $(x_i(t_k), y_i(t_k))$, si ricerca sulle altre curve C_j il valore per $\lambda_{i,j}(t_k)$ che permetta di descrivere il maggior numero dei punti campione di ogni C_j con un'ellisse E_{ij} tangente a C_i e a C'_i in $x(t_k)$ e $x'(t_k)$. Questo viene fatto con una selezione casuale dei punti campione $\{\mathbf{p}\}$ della curva C_j , candidata a sezione trasversale: grazie a Eq. 12, per ogni \mathbf{p} si ha una sola (o nessuna) possibile ellisse, per la quale vengono contati gli altri punti campione di C_j la cui distanza da essa è inferiore ad una soglia δ costante, posta sperimentalmente pari a $\sqrt{2}$. Le distanze sono calcolate con l'approssimazione di Sampson:

$$d_{Samp}(\mathbf{p}, E_{ij}) = \frac{|\mathbf{p}^\top E_{ij} \mathbf{p}|}{2\sqrt{(E_{ij} \mathbf{p})_1^2 + (E_{ij} \mathbf{p})_2^2}}, \quad (13)$$

dove $(E_{ij} \mathbf{p})_n$ indica l' n -esima componente del vettore $(E_{ij} \mathbf{p})$.

Anche se adesso per C_j è disponibile una versione analitica, si è preferito l'impiego dei suoi punti campione per mantenere un legame con le misure originali dell'immagine e tarare l'ellisse su di esse. Per ogni coppia contorno-sezione (C_i, C_j) candidata, si tiene traccia della massima frazione (*di contenimento*) $w(i, j)$ di punti campione di C_j per i quali è stata trovata un'ellisse, istanziata da determinati t_k e $\lambda_{i,j}(t_k)$ per C_i , che li contiene nel suo intorno di ampiezza δ . Al termine di tutti i test incrociati, le curve C_j con $\max_i w(i, j)$ superiore al 70% vengono considerate come ellissi: la loro equazione è fornita dalla matrice per cui hanno conseguito la massima percentuale di contenimento. I punti campione all'interno della loro fascia vengono ad esse associati come *inlier*.

Allo scopo di ridurre il tempo sprecato in ricerche inutili, il procedimento esposto viene aiutato da alcune euristiche, isirate a criteri fisici, numerici e statistici.

Se una curva non è parte del contorno apparente, il suo impiego in questo ruolo fa solo sprecare del tempo; per limitare questi fenomeni,

⁹ Per come sono stati scelti i t_i dei campioni, $\Delta t = 1$ permette già di compiere degli spostamenti di dimensioni paragonabili a quelle di un pixel, arrivando ad una risoluzione sub-pixel nei tratti con elevata curvatura.

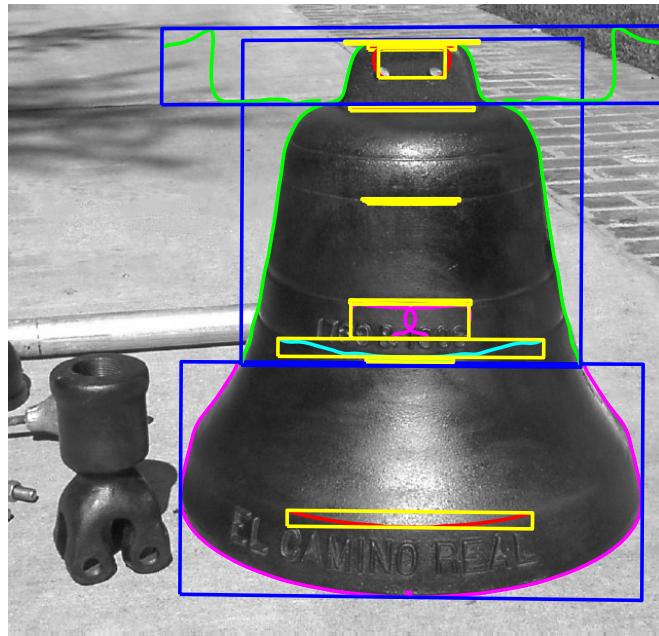


Figura 25. Le curve con bounding box gialla non possono essere parte del contorno apparente. I limiti delle bounding box sono individuati dalle rette tracciate parallelamente a l_s e da quelle del fascio per v_∞ che sottendono tutta la curva.

si costruisce per ogni curva, e relativa controparte, la bounding box che le contiene entrambe: tutte le curve con bounding box contenuta in quella di un'altra curva non partecipano come possibili parti del contorno apparente (Fig. 25). Un'altra euristica interviene nella selezione delle curve candidate a sezione trasversale, in funzione della posizione dei punti x e x' : la zona dell'immagine in cui può trovarsi una sezione trasversale deve essere circoscritta; è inutile cercare ellissi al di fuori di essa. Questa regione è sicuramente delimitata dalle rette l e l' ; possiamo comunque abbozzare una sua ulteriore delimitazione intersecandola con una fascia centrata sulla linea m , con semiampiezza pari alla distanza tra x e x' : solo le curve le cui bounding box hanno almeno un vertice all'interno di questa zona trapezoidale intervengono nel ruolo di sezione.

Altre considerazioni di tipo fisico riguardano la forma dell'ellisse. Essa infatti deve avere l'asse minore diretto grossomodo come l'asse del SOR (se l'angolo tra i due supera i 5° l'ellisse viene scartata subito, prima di saggiarla sugli altri punti campione). Inoltre, dato che i segmenti ellittici possono essere quasi rettilineo, si deve evitare che essi siano descritti da ellissi quasi degeneri, con l'asse minore pressoché nullo rispetto a quello maggiore: è stato quindi imposto che il rapporto tra le lunghezze dei due assi sia superiore a 0.05.

Per quanto riguarda gli aspetti numerici, occorre tener presente di come viene determinato $x'(t_k)$; l'omologia infatti non viene usata direttamente tramite la matrice H perché $Hx(t_k)$ viene a trovarsi vicino

Algoritmo 3. Estrazione delle sezioni trasversali

- 1 Forma gli insiemi delle curve candidate ad essere contorno apparente o ellissi.
 - 2 Per tutte le coppie (C_i, C_j) contorno-sezione possibili, ripeti i passi da 3 a 5.
 - 3 Campiona il dominio della spline che interpola la curva C_i ($\Delta t = 1$); poni $w(i, j) = 0$.
 - 4 Per ogni t_k scegli $N_{i,j}$ punti campione di C_j e per ognuno di essi istanzia, con Eq. 12, l'ellisse del fascio relativo a t_k . Per ogni istanza, determina la percentuale di contenimento $w_p(i, j, k)$.
 - 5 Se $\max_p\{w_p(i, j, k)\} \geq w(i, j)$, allora $w(i, j) \leftarrow \max_p\{w_p(i, j, k)\}$ e aggiorna anche $N_{i,j}$.
 - 6 Seleziona come ellissi le curve C_j con $\max_i\{w(i, j)\} \geq 0.7$
-

alla curva C' , ma non su di essa. L'individuazione del punto su C' per il quale $x'(t_k) \simeq Hx(t_k)$ avviene quindi intersecando C' con la retta passante per v_∞ e $x(t_k)$. Il calcolo di questa intersezione avviene cercando lo zero della funzione (sul parametro t) che si ottiene mettendo a confronto $C' = (x'(t), y'(t))$ con la retta relativa al vettore omogeneo $r = v_\infty \times x(t_k)$; la soluzione è trovata da un algoritmo iterativo a partire da $t = t_k$. Quando r e C' sono però vicini alla tangenza, la convergenza verso il giusto zero può presentare dei problemi: per evitare situazioni di questo tipo, viene controllato l'angolo formato da 1 e $1'$; si cerca di istanziare un'ellisse solo quando tale angolo supera i 5° (tra l'altro, quando 1 e $1'$ sono quasi allineate, l'informazione che forniscono è praticamente nulla).

Un altro aspetto numerico tiene conto del criterio adottato per la selezione delle ellissi: l'uso della percentuale sui punti campione fa sì che curve C_j con pochi punti abbiano però alti valori per $\max_i\{w(i, j)\}$ e possano essere selezionate anche se non sono delle ellissi. Per evitare questa situazione, prima di iniziare i test incrociati, vengono scartate le curve con un numero di punti campione inferiore a 50.

Infine, l'euristica a carattere statistico: per ogni t_k è infatti inutile istanziare tante ellissi quanti sono i punti campione della curva candidata a sezione trasversale; ne bastano solo alcuni scelti a caso. Dobbiamo quindi tarare, per ogni coppia contorno-sezione (C_i, C_j) in esame, il loro numero $N_{i,j}$ in base alla percentuale massima di contenimento $w(i, j)$ trovata fino a t_k (Eq. 2, con $s = 1$ e $p = 0.99$). Quando in teoria $N_{i,j}$ raggiunge quota 0, esso viene forzato a valere 3, per permettere di continuare a saggiare, con pochi tentativi, C_j : può darsi che il vero punto di contatto non sia il primo trovato, ma che sia situato nelle vicinanze.

Per un quadro riassuntivo, l'estrazione delle ellissi (candidate a essere classificate come sezioni circolari) è schematizzata in Alg. 3.

Estrazione dei bordi

La procedura per l'estrazione dei bordi da un eventuale tratto di contorno apparente avviene in un modo simile al precedente; stavolta però la curva in tutta la sua lunghezza interpreta il contorno, mentre la parte compresa tra il punto dato dall'istante t_k corrente e l'estremo da cui si è iniziato a percorrerla assume il ruolo di candidato a bordo. Le differenze principali con il caso precedente sono tre: una curva può essere percorsa in entrambi i sensi, qualora abbia dei punti vicini all'asse di simmetria (le ellissi devono necessariamente intersecarlo) sia nella sua parte inferiore che superiore¹⁰; il punto di contatto tra bordo e contorno apparente è dato dall'istante t_k per il quale è massima la differenza tra il numero M_p dei punti campione contenuti dall'ellisse (istanziata assieme a $\lambda(t_k)$) e il numero W di quelli esterni ad esso (questo accorgimento serve ad evitare casi come in figura); infine, la percentuale di contenimento w interviene solo nel calcolo del numero di tentativi N ed è data dal rapporto tra il massimo numero di campioni contenuti su tutti quelli che costituiscono la parte corrente di curva candidata a bordo.

Lo schema riassuntivo è quello in Alg. 4 (con l'espressione $\|C_j\|$ si è indicata la cardinalità dei punti campione della curva in esame).

2.6.2 Raffinamento delle ellissi estratte

Chiamiamo \mathcal{E} l'insieme di tutte le ellissi estratte, sia i possibili bordi che le possibili sezioni; sia inoltre E_i la matrice relativa all'ellisse i -esima \mathcal{E}_i di \mathcal{E} . Con la parametrizzazione adottata, le ellissi in \mathcal{E} dipendono essenzialmente solo da due valori: il punto di contatto t_k ed il coefficiente λ_p , determinato da un unico punto p . In un certo senso si tratta di una istanziazione minima, stile RANSAC, che va arricchita dell'informazione apportata da tutti i punti inlier che sono stati associati ad ognuna di esse: occorre cioè minimizzare, per ogni ellisse, la somma delle distanze (sempre con Sampson, Eq. 13) dei propri inlier:

$$\sum_{p \in \text{inlier}(E_i)} d_{S_{mp}}^2(p, E_i). \quad (14)$$

Come in RANSAC, la minimizzazione non viene effettuata una volta sola, ma integrata con una ulteriore classificazione dei punti campione della curva da cui l'ellisse è stata tratta; i due passi vengono ripetuti fino a quando l'insieme degli inlier non varia più.

Terminata questa fase, si cerca di raggruppare eventuali ellissi trovate su tratti differenti, ma originati dallo stesso parallelo, in un'unica entità. Un'altra possibile causa di ellissi "cloni" è data inoltre dalla duplice estrazione di un'ellisse con entrambi i metodi, qualora non

¹⁰ La soglia di vicinanza è stata posta pari a 10 pixel.

Algoritmo 4. Estrazione dei bordi

- 1 Per ogni curva C_i nell'insieme dei possibili tratti di contorno apparente ripeti i passi da 2 a 6.
 - 2 Campiona il dominio della spline che interpola la curva C_i ($\Delta t = 1$); poni $M = 0$, $w = 0$. Sia t_{\max} l'istante associato all'estremo della curva C_i , opposto a quello con $t = 0$.
 - 3 Se C_i presenta nella sua parte inferiore ($t < t_{\max}/3$) dei punti campione vicini a l_s , scorri C_i a partire da $t = 0$ fino a t_{\max} :
 - per ogni t_k sia $C_j = \{\text{punti campione di } C_i \text{ situati tra } 0 \text{ e } t_k\}$
 - scegli N punti campione $\{p\}$ di C_j per istanziare le ellissi con Eq. 12; per ciascuna valuta M_p e W_p
 - se $M \leq \max_p\{M_p - W_p\}$, $M \leftarrow \max_p\{M_p - W_p\}$
 - $w \leftarrow M/\|C_j\|$ e aggiorna N .
 - 4 Estrai l'ellisse relativa al punto t_k per il quale M è stato aggiornato l'ultima volta. Azzera M .
 - 5 Se C_i presenta nella sua parte superiore ($t > 2/3 t_{\max}$) dei punti campione vicini a l_s , scorri C_i a partire da $t = t_{\max}$ fino a 0:
 - per ogni t_k sia $C_j = \{\text{punti campione di } C_i \text{ situati tra } 0 \text{ e } t_k\}$
 - scegli N punti campione $\{p\}$ di C_j per istanziare le ellissi con Eq. 12; per ciascuna valuta M_p e W_p
 - se $M \leq \max_p\{M_p - W_p\}$, $M \leftarrow \max_p\{M_p - W_p\}$
 - $w \leftarrow M/\|C_j\|$ e aggiorna N .
 - 6 Estrai l'ellisse relativa al punto t_k per il quale M è stato aggiornato l'ultima volta.
-

sia stata trovata una chiara inclusione del tratto che la contiene in una bounding box più esterna.

Per stabilire se effettuare delle fusioni, per ogni coppia non ordinata del prodotto cartesiano di \mathcal{E} con se stesso vengono determinate la percentuale degli inlier che un'ellisse riesce a descrivere dell'altra, oltre alla varianza dei loro residui; a ciascuna coppia viene assegnata la massima tra le due percentuali incrociate e relativa varianza. A partire dalla coppia con maggior percentuale¹¹ (che deve essere comunque superiore al 70%), tra le due ellissi si sacrifica quella che descrive peggio gli inlier dell'altra; si minimizza quindi il costo di Eq. 14 relativo alla restante, sull'unione dei suoi inlier con i campioni che è riuscita a prendere dall'altra. Di nuovo, il procedimento di minimizzazione e

¹¹ a parità di percentuale, si sceglie la coppia a varianza minima

classificazione viene ripetuto sugli inlier dell'ellisse sacrificata fino a convergenza. Effettuata la fusione di una coppia, si procede ad una nuova serie di test incrociati fino ad esaurimento delle fusioni. Queste tre fasi corrispondono rispettivamente a Fig. 26(a)–26(c); le due coppie di ellissi estremamente ravvicinate correttamente non vengono fuse: infatti, sono state trovate su segmenti ellittici paralleli e non dovuti (due a due) allo stesso parallelo.

Nella minimizzazione la parametrizzazione dell'ellisse viene estesa, con lo scopo di migliorare anche l'orientamento delle rette tangenti nei punti di contatto. Al termine del primo livello, infatti, le due rette l e l' sono state corrette imponendo il passaggio per il punto di mezzo tra le loro intersezioni con l'asse l_s : spostando tale punto, si modificano contemporaneamente i coefficienti angolari di entrambe. Sia pertanto $\mathbf{q} = \mathbf{m} \times l_s$ l'intersezione tra l'asse di simmetria e la retta \mathbf{m} congiungente x e x' ; sia inoltre s_0 il punto su l_s da cui sono state fatte passare l e l' : avendo cura di normalizzare \mathbf{q} e s_0 rispetto alla loro terza componente, un qualsiasi punto su l_s si può sempre scrivere come

$$\mathbf{s}_\mu = \mu \mathbf{q} + \mathbf{s}_0.$$

In definitiva, la minimizzazione del costo di Eq. 14 avviene, per ogni ellisse, su tre parametri:

- il punto di contatto t ,
- il coefficiente del fascio λ ,
- il coefficiente di orientamento delle tangenti μ .

2.6.3 Selezione delle ellissi e classificazione delle curve

Rimane da eseguire l'ultimo passo, quello che completa la segmentazione. Giunti a questo punto, dobbiamo operare una selezione delle ellissi estratte e classificare come tratti di contorno apparente le curve su cui sono stati costruiti i loro vincoli di tangenza. Nel caso in cui un'ellisse sia stata inizialmente estratta come bordo, la parte della curva che va a formare il contorno apparente è quella successiva al punto di contatto, nel verso in cui la curva è stata percorsa.

La selezione finale ha lo scopo di rimuovere ellissi inesistenti o male stimate. Per prima cosa, si eliminano tutte le ellissi troppo piccole¹² (Fig. 26(d)): non è detto infatti che la selezione sul numero minimo di punti campione, fatta all'inizio del secondo livello, ci abbia messo al riparo dalla falsa manifestazione di piccole ellissi; inoltre, anche se appartenessero effettivamente al SOR, qualora venissero scelte per la calibrazione il loro contributo sarebbe esiguo. Questa selezione

¹² Sperimentalmente è stata fissata una soglia minima per l'area di un'ellisse, pari a 700 pixel² in un'immagine 640 × 480.

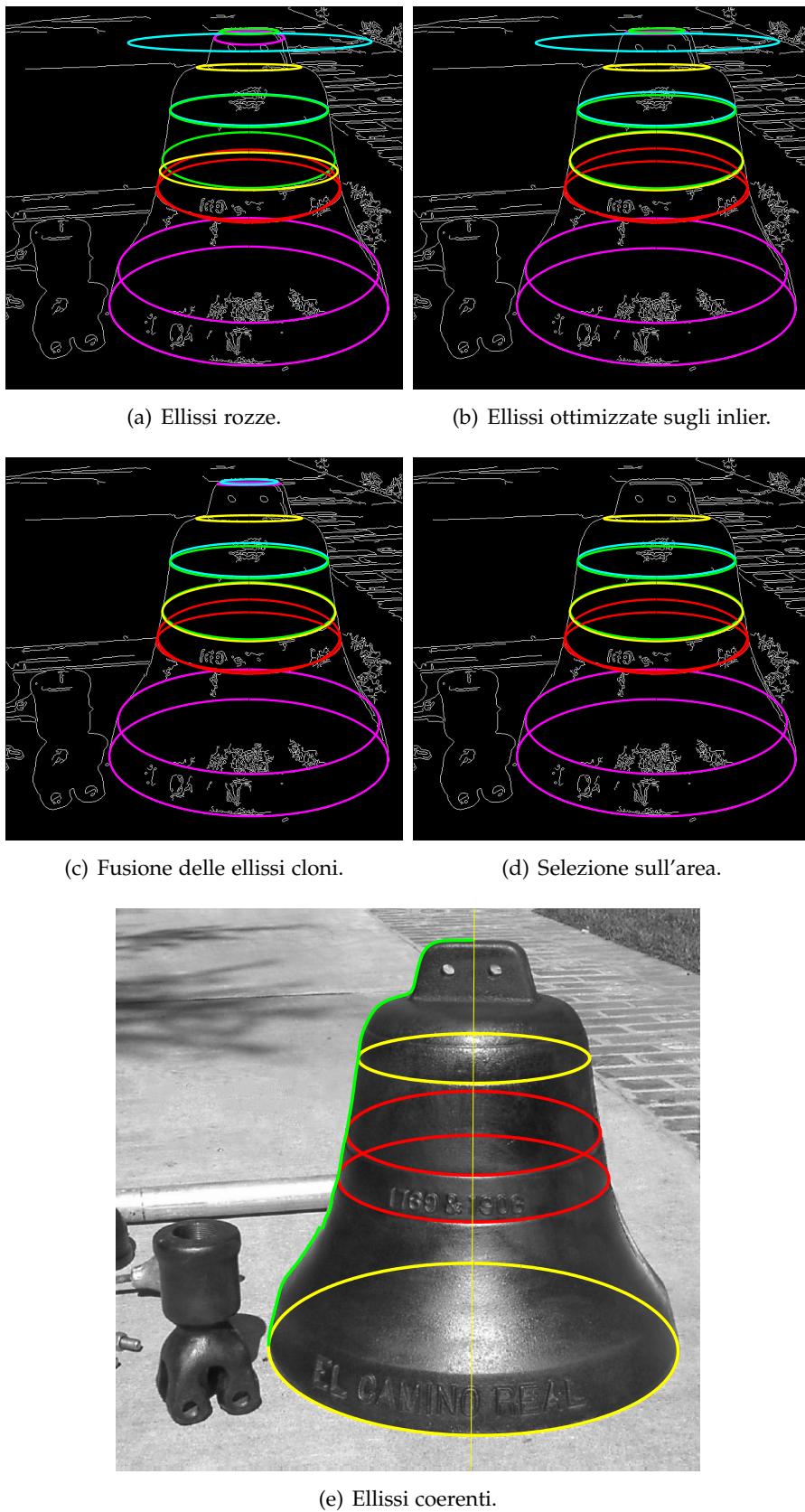


Figura 26. Da (a) ad (e), passi seguiti nell'estrazione delle ellissi.

può tuttavia creare problemi nella ricostruzione di SOR con estremità piccole, quali le bottiglie.

La selezione più importante è però quella rivolta a determinare il gruppo con il maggior numero di ellissi in accordo allo stesso modello (Fig. 26(e)): per fare ciò si ricorre al fatto che \mathbf{v}_∞ è ricavabile anche tramite una coppia di ellissi corrette [6].

Se infatti calcoliamo le loro intersezioni, dalle equazioni poste a sistema si ottengono quattro soluzioni, in generale complesse coniugate, di cui al più solo due sono reali (quando le due ellissi si intersecano davvero); per ellissi dovute ai paralleli di un SOR, quattro intersezioni reali non si possono verificare. Calcolando l'equazione delle "rette" passanti per ciascuna coppia complessa coniugata, l'intersezione di quest'ultime è proprio \mathbf{v}_∞ .

Possiamo pertanto considerare tutte le possibili coppie di ellissi e per ognuna di esse determinare la distanza del \mathbf{v}_∞ trovato con RANSAC (che fa da riferimento) da quello stimato con esse, $\hat{\mathbf{v}}_\infty$. Qualora due ellissi presentino quattro intersezioni reali, la distanza viene per definizione posta a infinito (come già detto, per i SOR, è un caso fisicamente impossibile). A questo punto, ci possiamo aspettare che coppie di ellissi ben stimate determinino un $\hat{\mathbf{v}}_\infty$ vicino a quello di riferimento, mentre le intersezioni che coinvolgono almeno un'ellisse errata forniranno un $\hat{\mathbf{v}}_\infty$ ben più distante.

Una selezione in base alla distanza delle stime su \mathbf{v}_∞ sembra un criterio plausibile, ma va perfezionata. Infatti, fissata una soglia τ per la distanza, all'interno del cerchio centrato sul \mathbf{v}_∞ di riferimento e con raggio pari a τ possono cadere anche coppie (errate) che solo casualmente hanno fornito un $\hat{\mathbf{v}}_\infty$ sufficientemente vicino; addirittura, la loro stima potrebbe essere la più vicina di tutte al \mathbf{v}_∞ di riferimento.

Se però iniziamo a controllare anche le altre intersezioni che coinvolgono una data coppia con un $\hat{\mathbf{v}}_\infty$ all'interno del cerchio, possiamo sperare che, se le due ellissi sono giuste, trovino accordo anche con altre ellissi corrette; viceversa una buona stima, ma casuale, di $\hat{\mathbf{v}}_\infty$ verrebbe rivelata dalla mancanza di ulteriori conferme.

In definitiva, fissare una distanza dal \mathbf{v}_∞ di riferimento implica la creazione di una relazione binaria di *coerenza* tra le ellissi; la determinazione dell'insieme più numeroso (e a parità di elementi, con somma delle distanze minima) in cui tutte le ellissi sono concordi tra di loro può fornire la risposta al problema.

La formulazione del problema tramite un grafo pesato, lo riconduce alla determinazione della *cricca* (*clique*) che, tra tutte quelle massime, ha la somma dei pesi degli archi più piccola. Si definisce *cricca* un sottoinsieme massimale di nodi che siano a due a due collegati tra loro; una *cricca massima* è una cricca con il più grande numero possibile di nodi. Il grafo viene così costruito: ad ogni ellisse corrisponde un nodo; se l'intersezione tra due ellissi fornisce un $\hat{\mathbf{v}}_\infty$ distante da quello di riferimento meno della soglia τ , i loro due no-

di vengono collegati da un arco con peso pari alla distanza tra $\hat{\mathbf{v}}_\infty$ e \mathbf{v}_∞ . I problemi che coinvolgono le cricche sono computazionalmente difficili; tuttavia, in questo caso il numero di nodi è molto limitato (le ellissi visibili di un SOR sono infatti solitamente poche), consentendo un'analisi pressoché esaustiva di tutte le cricche.

Quale valore usare per τ ? Purtroppo stavolta non possiamo usare LMedS, visto l'esiguo numero di dati a disposizione; una soglia fissa d'altra parte è difficile da stabilire (\mathbf{v}_∞ è infatti molto distante dalla regione di piano circoscritta dalla foto e ha poco senso parlare di intorni con raggio 100 piuttosto che 200, o un altro valore). La soluzione che abbiamo adottato è un compromesso: τ viene tarato prendendo il valore massimo tra la mediana delle distanze e quello dato da 1.5 volte la distanza più piccola.

Oltre al problema della soglia τ , la selezione delle ellissi corrette necessita purtroppo di ulteriori approfondimenti: dalle prove condotte su foto reali, infatti, il criterio della distanza da \mathbf{v}_∞ non ha fornito sempre le giuste indicazioni (in particolare, due ellissi stimate sufficientemente bene ma vicine nella foto sono in scarso accordo tra di loro, per cui solitamente una delle due viene scartata).

2.7 CALIBRAZIONE INTERNA E RICOSTRUZIONE METRICA

Nella sezione 2.2, sui lavori presenti in letteratura, è stato fatto riferimento alle analogie che ci sono tra la geometria di un SOR e la ricostruzione in SAM (Single Axis Motion); in particolare era stato anticipato che per la calibrazione (e ricostruzione) è necessaria la conoscenza di alcune grandezze che in SAM restano costanti lungo tutta la sequenza filmata. Avevamo detto anche che nella terminologia dei SOR, le principali tra queste grandezze sono l'asse ed il vertice dell'omologia armonica tra i due lati del contorno apparente, oltre ai i punti circolari del fascio di piani ortogonali all'asse di rivoluzione. Il primo stadio dell'algoritmo (sezione 2.5) ha permesso una stima robusta dei parametri dell'omologia armonica, per cui restano da determinare solo i punti circolari.

Per mezzo di queste grandezze è possibile calibrare i parametri interni di fotocamera, se utilizziamo come modello per essa la "natural camera"¹³, riassunta dalla matrice di proiezione in Eq. 52:

$$\mathbf{P} = \mathbf{KR}[\mathbf{I}_{3 \times 3} \mid -\mathbf{c}],$$

con

$$\mathbf{K} = \begin{bmatrix} f & 0 & p_x \\ 0 & f & p_y \\ 0 & 0 & 1 \end{bmatrix}$$

I parametri interni sono dati dai coefficienti della matrice \mathbf{K} e grazie ad essi è possibile la ricostruzione metrica del SOR.

¹³ Per una sua breve trattazione, cfr. Appendice A.

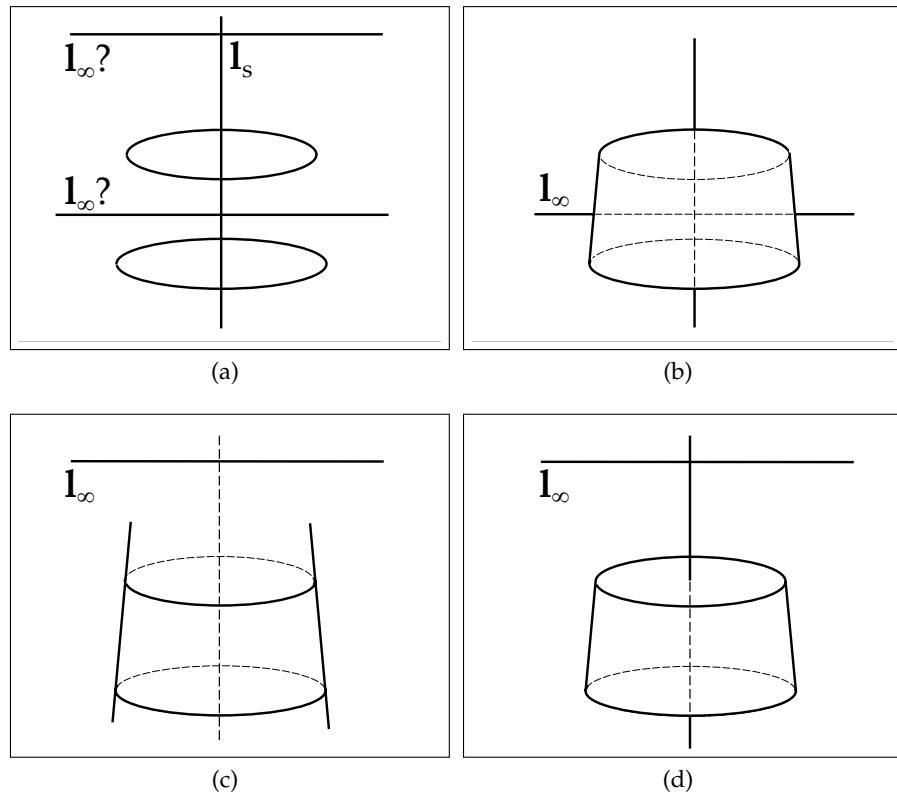


Figura 27. Indeterminazione della linea dell'orizzonte.

2.7.1 Calibrazione interna

Riprendiamo l'intersezione di due ellissi del SOR e supponiamo per il momento che ci siano due radici reali: le altre due radici complesse costituiscono necessariamente i punti circolari **i** e **j** che stiamo cercando. Questa affermazione deriva dal fatto che entrambe le ellissi li condividono sulla linea dell'orizzonte l_∞ generata dai piani ortogonali all'asse di rivoluzione; quindi queste due grandezze complesse coniugate devono soddisfare il sistema formato dalle due coniche. Essendo due delle radici reali, le altre due devono essere necessariamente i punti circolari.

Quando invece siamo in presenza di quattro radici complesse coniugate, sorge un'ambiguità che può essere risolta dall'intervento di un'altra ellisse, oppure tramite una semplice euristica [6]. L'origine dell'ambiguità è evidente in Fig. 27: la conoscenza di due ellissi che non si intersecano nell'immagine non permette di capire se il punto di vista giace a mezz'altezza tra i due paralleli o è esterno (nel caso riportato in figura, i due paralleli potrebbero essere osservati dall'alto). Le rette passanti per ciascuna coppia coniugata costituiscono le due possibili linee dell'orizzonte. Se osserviamo Fig. 27(b), si nota che quando il punto di vista è compreso tra i due paralleli, l_∞ è compreso tra le due ellissi e le loro porzioni visibili sono opposte: una mostra in

pratica solo il suo lato superiore, l'altra solo quello inferiore. Questa situazione non si verifica negli altri casi, sia che il punto di vista stia sopra i due paralleli, sia che stia sotto. Questa è la semplice euristica che ci viene in aiuto: quando le due ellissi presentano la stragrande maggioranza dei propri inlier solo sui due lati opposti, la linea dell'orizzonte corretta è quella che passa tra le due ellissi; viceversa, va scelta l'altra retta.

Noti \mathbf{i} e \mathbf{j} , dall'analogia con SAM possiamo impostare il seguente sistema di equazioni che deve soddisfare l'immagine della conica assoluta (IAC) ω :

$$\begin{cases} \mathbf{i}^\top \omega \mathbf{i} = 0 \\ \mathbf{j}^\top \omega \mathbf{j} = 0 \\ \mathbf{l}_s \times \omega \mathbf{v}_\infty = \mathbf{0} \end{cases} \quad (15)$$

Tale sistema possiede tre vincoli lineari indipendenti, sufficienti per la stima dei parametri interni del modello "natural camera". Riscrivendo infatti la terza equazione come $\mathbf{l}_s \times \omega \mathbf{v}_\infty = \mathbf{0}$, otteniamo un sistema lineare risolvibile mediante fattorizzazione SVD. Nota la IAC, poiché $\omega = \mathbf{K}\mathbf{K}^\top$, si ottiene \mathbf{K} dalla sua fattorizzazione di Choleski. Purtroppo, nel caso in cui \mathbf{v}_∞ è un punto improprio, il punto principale (la terza colonna di \mathbf{K}) viene a trovarsi in una posizione indeterminata su \mathbf{l}_s e non si può portare a termine la stima dei parametri interni; se l'immagine non è stata ritagliata, una soluzione comune può comunque essere quella di indicare come punto principale il punto su \mathbf{l}_s più vicino al centro della foto. Gli effetti di questa indeterminazione si avvertono ognqualvolta il punto principale si colloca vicino a \mathbf{l}_s , peggiorando la stima della matrice \mathbf{K} ; la soluzione migliore per evitare tutto ciò è fare delle foto in cui il SOR si trovi leggermente defilato rispetto al centro della scena inquadrata.

Fin qui, la teoria; nella pratica, le ellissi riconosciute come appartenenti al SOR possono essere più di due (quali scegliere?) e, inoltre, \mathbf{v}_∞ difficilmente (praticamente mai) viene a trovarsi sulla linea dell'orizzonte calcolata dalle ellissi.

Per quanto riguarda la scelta delle ellissi da cui ricavare la linea dell'orizzonte, il criterio di selezione ha tenuto conto sia della distanza tra i due centri delle due ellissi (ellissi lontane evitano i problemi di malcondizionamento che hanno invece quelle vicine; inoltre, descrivono una porzione ampia del SOR), sia della vicinanza del loro $\hat{\mathbf{v}}_\infty$ da quello di riferimento (si riduce l'errore su $\mathbf{l}_\infty^\top \mathbf{v}_\infty = 0$): in definitiva, vengono utilizzate le due ellissi per le quali è massimo

$$\frac{\text{dist}(\mathbf{e}_i, \mathbf{e}_j)}{\text{dist}(\mathbf{v}_\infty, \hat{\mathbf{v}}_\infty)},$$

essendo \mathbf{e}_k il centro della k -esima ellisse. In Fig. 26(e) le ellissi scelte per determinare la linea dell'orizzonte sono evidenziate in giallo.

Successivamente, si vincola \mathbf{l}_∞ a passare per \mathbf{v}_∞ secondo un modello in cui i punti circolari sono mobili, minimizzando la loro distanza

(approssimazione di Sampson) dalle due ellissi. Nell'ipotesi che \mathbf{v}_∞ non sia un punto improprio, sia $\mathbf{v}_\infty = [v_1, v_2, 1]^\top$; esprimiamo inoltre la linea dell'orizzonte come $\mathbf{l}_\infty = [1, m, l_3]^\top$ (si perde la linea parallela all'asse y dell'immagine): dal vincolo $\mathbf{l}_\infty^\top \mathbf{v}_\infty = 0$ si ottiene $l_3 = -(v_1 + mv_2)$. Sia adesso $\mathbf{i} = [\alpha + i\beta, \gamma + i\delta, 1]^\top$ uno dei due punti circolari; dalla relazione $\mathbf{l}_\infty^\top \mathbf{i} = 0$ otteniamo l'equazione

$$(\alpha + i\beta) + m(\gamma + i\delta) - (v_1 + mv_2) = 0,$$

da cui, uguagliando a 0 sia la parte reale che quella immaginaria, si ha:

$$\begin{cases} \delta = -\beta/m \\ \gamma = -(\alpha - (v_1 + mv_2))/m. \end{cases}$$

In definitiva, \mathbf{i} può essere parametrizzato come

$$\mathbf{i}(m, \alpha, \beta) = \begin{bmatrix} \alpha + i\beta \\ -(\alpha - (v_1 + mv_2) + i\beta)/m \\ 1 \end{bmatrix}$$

e la stima di \mathbf{l}_∞ avviene minimizzando su m, α, β con Levenberg-Marquardt la somma delle distanze (secondo Sampson) di $\mathbf{i}(m, \alpha, \beta)$ dalle due ellissi, usando come seme i valori inizialmente trovati per \mathbf{i} ; l'impiego anche del complesso coniugato non porta ulteriore informazione.

2.7.2 Ricostruzione metrica del SOR

La ricostruzione metrica del SOR avviene in due fasi [5, 6]: la prima comporta la ricostruzione di un meridiano χ sull'immagine, seguita dalla determinazione della linea di fuga, \mathbf{m}_∞ , del piano sul quale giace; la seconda consiste nella rettificazione di χ grazie all'impiego di ω e \mathbf{m}_∞ .

I singoli tratti che sono stati classificati come parti del contorno apparente vengono campionati ($\Delta t = 1$), per aggregare poi insieme tali punti tracciando il cammino più lungo [8] sullo MST euclideo dato da essi (Fig. 28): così facendo, si cerca di eliminare eventuali punti estranei, dovuti a punti campione erroneamente associati alle curve in questione. Sui punti selezionati si adatta una spline che, oltre ad essere necessaria nella ricostruzione del meridiano (per il calcolo delle tangenti al contorno apparente), nell'ipotesi di un profilo dolce riempie eventuali lacune tra i singoli tratti. Un approccio più conservativo avrebbe potuto comunque essere quello di ricostruire separatamente i singoli tratti, lasciando indeterminate le zone mancanti.

Ricostruzione del meridiano χ

Come esposto in [5, 6], per la costruzione di un meridiano χ vengono sfruttate le ultime proprietà proiettive dei SOR enunciate nella sezione 2.4 (in dettaglio, si tratta delle proprietà 2.4.4, 2.4.5 e 2.4.6).

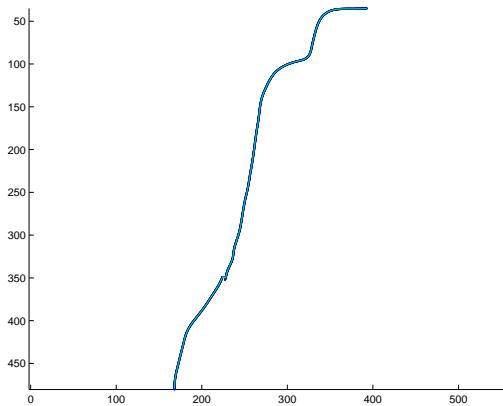


Figura 28. Ricomposizione del contorno apparente con il cammino più lungo sullo MST.

Come mostrato in Fig. 29, dato il contorno apparente γ , esiste un'unico parallelo nell'immagine C' che include il generico punto $x'_\gamma \in \gamma$; inoltre, nota l_∞ , esiste un'unica omologia planare W che porta un'ellisse di riferimento C su C' (proprietà 2.4.1). Al variare di x'_γ su γ , cambiano sia il vertice v dell'omologia, che si muove su l_s , sia il coefficiente μ : questa dipendenza è individuabile con l'approccio costruttivo di [5, 6], per cui la costruzione del meridiano avviene fissando un punto x_γ sull'ellisse di riferimento e propagandolo lungo il SOR con le W date dai campioni presi sul contorno apparente.

L'ellisse che viene scelta come riferimento è quella, tra le due utilizzate per l_∞ , con area maggiore. Nel caso della campana, il meridiano ricostruito è quello in Fig. 30.

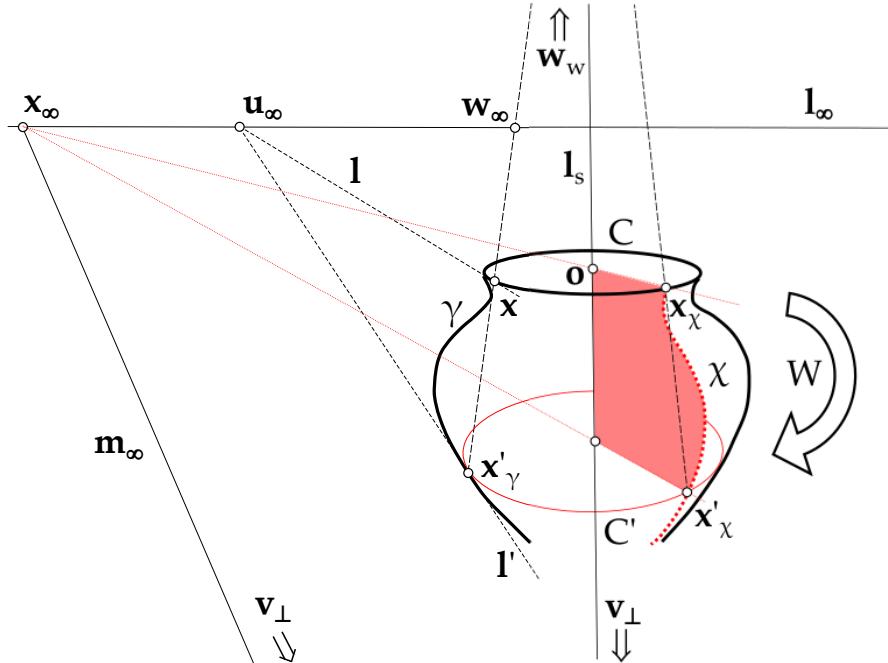


Figura 29. Ricostruzione dell'immagine di un meridiano.

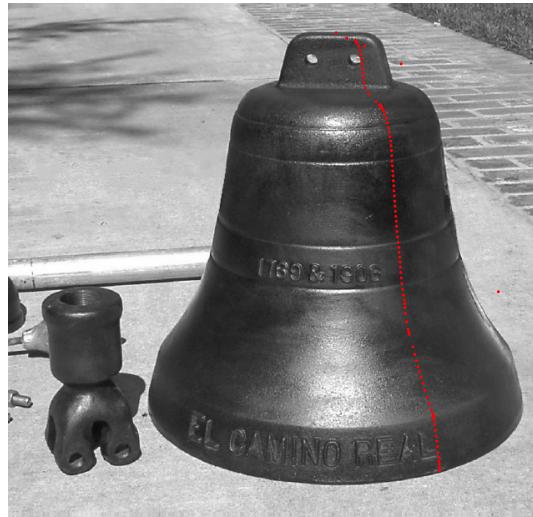


Figura 30. Ricostruzione di un meridiano nell'immagine.

Rettificazione del meridiano χ

La rettificazione del meridiano così ricostruito richiede la rimozione della distorsione prospettica sulla sua immagine, cioè occorre applicare (ai punti che lo individuano nella foto) l'*omografia* che lega il piano dell'immagine al piano in cui giace.

Per determinare un'*omografia* rettificante del meridiano in questione occorre la conoscenza di \mathbf{m}_∞ , cioè la linea di fuga del piano su cui giace nello spazio. Su tale piano si trova sicuramente la retta che congiunge \mathbf{x}_γ con l'immagine del centro del parallelo su cui si trova l'origine di \mathbf{x}_γ (si veda anche Fig. 29); essa è sempre in relazione polo-polare con \mathbf{l}_∞ . Pertanto, l'intersezione di questa retta con \mathbf{l}_∞ fornisce un punto della linea di fuga \mathbf{m}_∞ ; in formule, detto \mathbf{x}_∞ questo punto:

$$\mathbf{x}_\infty = (\mathbf{x}_\gamma \times C^{-1}\mathbf{l}_\infty) \times \mathbf{l}_\infty.$$

Per tracciare \mathbf{m}_∞ basta conoscere un suo altro punto; questo è fornito dal punto di fuga di tutte le rette sul piano di χ che sono perpendicolari all'asse di rivoluzione, ottenibile dalla relazione polo-polare con ω e \mathbf{l}_∞ . L'intersezione di \mathbf{m}_∞ con ω permette infine di ricavare le coordinate dei punti circolari $\mathbf{i}_\chi, \mathbf{j}_\chi$ del piano su cui giace χ .

Concludendo, l'*omografia* rettificante M_r che, se applicata ai punti di χ , restituisce la forma del meridiano (cioè il profilo $\rho(t)$) è quella che riporta $\mathbf{i}_\chi, \mathbf{j}_\chi$ alle loro forme canoniche:

$$M_r = \begin{bmatrix} \beta_\chi^{-1} & -\alpha_\chi \beta_\chi^{-1} & 0 \\ 0 & 1 & 0 \\ m_1 & m_2 & 1 \end{bmatrix} \quad (16)$$

essendo

$$\mathbf{m} = \begin{bmatrix} m_1 \\ m_2 \\ 1 \end{bmatrix}, \quad \mathbf{i}_\chi = \text{conj}(\mathbf{j}_\chi) = \begin{bmatrix} \alpha_\chi - i\beta_\chi \\ 1 \\ -m_2 - m_1 \alpha_\chi + im_1 \beta_\chi \end{bmatrix}.$$

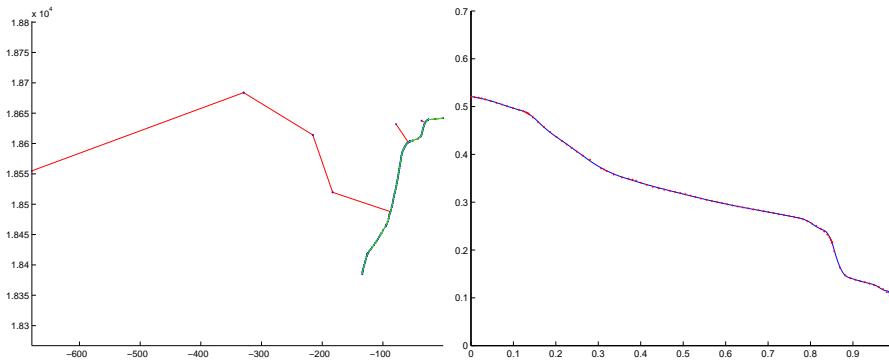


Figura 31. A sinistra, i punti scelti come campioni del profilo, una volta rettificato il meridiano; a destra la loro interpolazione mediante spline, ad altezza unitaria del SOR.

Quando l'omografia è applicata su punti che in realtà non appartengono al contorno apparente, bensì a dei bordi che non sono stati rilevati come tali, i punti trasformati si discostano molto dal profilo del SOR; di nuovo, il cammino più lungo sullo MST dei punti rettificati fornisce un metodo per l'eliminazione di questi elementi in realtà estranei, unito alla rimozione dei punti, tra quelli selezionati, che sono connessi da archi troppo lunghi (Fig. 31). Ricostruito il profilo, la sua rotazione attorno all'asse z genera il modello 3D del SOR.

In Fig. 32 si può vedere il modello della campana che ha fatto da esempio per tutta la trattazione; purtroppo la foto è stata eseguita frontalmente facendo apparire simmetrica anche la sagoma del pezzo di aggancio, per cui la ricostruzione 3D ha un "cappello" di troppo.

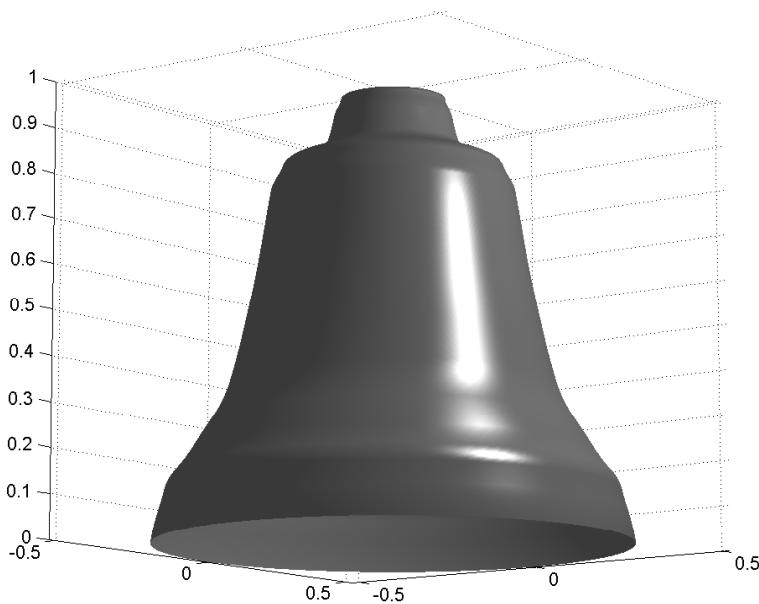


Figura 32. Modello tridimensionale della campana.

2.8 CALIBRAZIONE ESTERNA E ACQUISIZIONE DELLA TEXTURE

Nella matrice di proiezione $P = KR[I_{3 \times 3} | -c]$, oltre alla matrice K, intervengono altri due termini che servono a legare un sistema di riferimento che si trova nella scena e quello solidale con la fotocamera. Essi costituiscono i *parametri esterni* di fotocamera: in particolare, il vettore c esprime le coordinate che il centro di proiezione della fotocamera possiede nella scena, mentre R è la matrice di rotazione ortogonale che orienta il sistema di riferimento posto sul centro di proiezione.

2.8.1 Formazione dell'immagine di un SOR

La completa conoscenza dei parametri della fotocamera permette di descrivere in termini algebrici le curve che un SOR crea in un'immagine: le ellissi dovute ai paralleli ed il contorno apparente.

Formazione dell'immagine dei paralleli

Riprendiamo la parametrizzazione di un SOR esposta in Eq. 5:

$$\varphi(t, \vartheta) = \begin{cases} x(t, \vartheta) = \rho(t) \cos \vartheta \\ y(t, \vartheta) = \rho(t) \sin \vartheta \\ z(t, \vartheta) = t; \end{cases}$$

essa definisce l'equazione parametrica della circonferenza con raggio $\rho(t)$ e giacente sul piano π , parallelo a quello formato dagli assi x e y relativi alla parametrizzazione del SOR, e posto ad altezza $z = t$.

Attraverso la matrice P è possibile definire la trasformazione che proietta i punti di tale piano su quello immagine: indicando con X il vettore omogeneo associato ad un generico punto posto ad altezza t , il relativo punto sull'immagine è dato da

$$\begin{aligned} x = P\mathbf{X} &= P \begin{bmatrix} x \\ y \\ t \\ 1 \end{bmatrix} = \\ &= x\mathbf{p}_1 + y\mathbf{p}_2 + (t\mathbf{p}_3 + \mathbf{p}_4) = \\ &= [\mathbf{p}_1 \quad \mathbf{p}_2 \quad (t\mathbf{p}_3 + \mathbf{p}_4)] \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \\ &= H_t \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \end{aligned} \tag{17}$$

La matrice $3 \times 3 H_t$ ottenuta dalle colonne di P (la notazione \mathbf{p}_j indica la j -esima colonna della matrice P) rappresenta la trasformazione

tra tutti i punti del piano ad altezza t ed il piano immagine; il parallelo del SOR giacente sul piano π è espresso in coordinate omogenee dalla matrice

$$C = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -\rho(t)^2 \end{bmatrix};$$

ne consegue che esso si trasforma sul piano immagine nella conica data dalla matrice

$$C' = H_t^{-T} C H_t^{-1}.$$

Formazione del contorno apparente

Il contorno apparente è dato dalla proiezione sul piano immagine dei punti che costituiscono il generatore di contorni e cioè tali che il piano passante per essi e tangente al SOR comprenda anche il centro di proiezione della fotocamera. Se descriviamo il generatore di contorni attraverso una curva 3D parametrica $\psi(t) : [0, 1] \rightarrow \mathbb{R}^3$, il contorno apparente è ricavabile applicando la trasformazione proiettiva P della fotocamera a $\psi(t)$.

Parametrizzazione del generatore di contorni di un SOR

La formula del piano tangente ad una superficie parametrica $\varphi(u, v)$ in un suo punto $\varphi(u_0, v_0)$ è

$$\begin{vmatrix} x - x(u_0, v_0) & y - y(u_0, v_0) & z - z(u_0, v_0) \\ \frac{\partial x}{\partial u}(u_0, v_0) & \frac{\partial y}{\partial u}(u_0, v_0) & \frac{\partial z}{\partial u}(u_0, v_0) \\ \frac{\partial x}{\partial v}(u_0, v_0) & \frac{\partial y}{\partial v}(u_0, v_0) & \frac{\partial z}{\partial v}(u_0, v_0) \end{vmatrix} = 0. \quad (18)$$

Applicando Eq. 18 a Eq. 5 e imponendo il passaggio del piano per il centro di proiezione della fotocamera, otteniamo

$$\begin{vmatrix} c_x - \rho(t) \cos \vartheta & c_y - \rho(t) \sin \vartheta & c_z - t \\ \rho'(t) \cos \vartheta & \rho'(t) \sin \vartheta & 1 \\ -\rho(t) \sin \vartheta & \rho(t) \cos \vartheta & 0 \end{vmatrix} = 0$$

sse

$$\rho(t)(\rho(t) + \rho'(t)(c_z - t) - c_x \cos \vartheta - c_y \sin \vartheta) = 0. \quad (19)$$

Pertanto, i punti del generatore di contorni sono individuati dalle coppie (t, ϑ) che soddisfano Eq. 19; equivalentemente, essendo

$$\begin{cases} \cos \vartheta = \frac{1 - \tau^2}{1 + \tau^2} \\ \sin \vartheta = \frac{2\tau}{1 + \tau^2} \\ \tau = \tan \frac{\vartheta}{2} \end{cases} \quad (20)$$

si devono determinare le coppie (t, τ) che rendono valida

$$\rho(t) \left(\rho(t) + \rho'(t)(c_z - t) - c_x \frac{1 - \tau^2}{1 + \tau^2} - c_y \frac{2\tau}{1 + \tau^2} \right) = 0. \quad (21)$$

Trascurando la soluzione banale $\rho(t) = 0$, si risolve Eq. 21 rispetto a τ :

$$(1 + \tau^2)(\rho(t) + \rho'(t)(c_z - t)) - c_x(1 - \tau^2) - 2c_y\tau = 0,$$

da cui

$$\tau_{\pm}(t) = \frac{c_y \pm \sqrt{c_y^2 + c_x^2 - (\rho(t) + \rho'(t)(c_z - t))^2}}{c_x + \rho(t) + \rho'(t)(c_z - t)}. \quad (22)$$

La duplice soluzione riflette la natura fisica del generatore dei contorni, in quanto la sua immagine deve generare i due lati del contorno apparente. Per finire, applicando Eq. 20 in Eq. 5 e sostituendo a τ i valori forniti da Eq. 22 otteniamo l'equazione parametrica del generatore dei contorni:

$$\psi_{\pm}(t) = \begin{cases} x(t) = \rho(t) \frac{1 - \tau_{\pm}(t)^2}{1 + \tau_{\pm}(t)^2} \\ y(t) = \rho(t) \frac{2\tau_{\pm}(t)}{1 + \tau_{\pm}(t)^2} \\ z(t) = t \end{cases}, \quad (23)$$

sempre con $t \in [0, 1]$ e tale da non rendere negativo l'argomento della radice presente in Eq. 22.

Resta un'ultima osservazione da fare: il metodo esposto tratta il vaso come se fosse trasparente, senza tenere conto delle possibili occlusioni di alcuni paralleli. Questo fenomeno è illustrato nella sezione dedicata all'acquisizione della texture.

2.8.2 Calibrazione esterna

Una volta trovati i parametri interni della "natural camera" è possibile sfruttare le proprietà proiettive di un SOR per recuperare anche i parametri esterni: più precisamente la matrice di rotazione R ed il centro di proiezione, ottenibile a meno di un fattore di scala, sono ricavabili dall'asse di simmetria l_s e dell'orizzonte l_{∞} , assieme all'equazione di un'ellisse immagine di un qualche parallelo.

Per fare tutto ciò, una volta fissato un sistema di riferimento (per convenzione *destrorso*) Oxyz nella scena possiamo supporre, senza perdere di generalità, che il SOR sia appoggiato sul piano dato dagli assi x e y , con l'asse di simmetria coincidente con l'asse z ; il centro di proiezione viene invece ritenuto giacere sul semipiano formato dagli assi x e z , con x positiva.

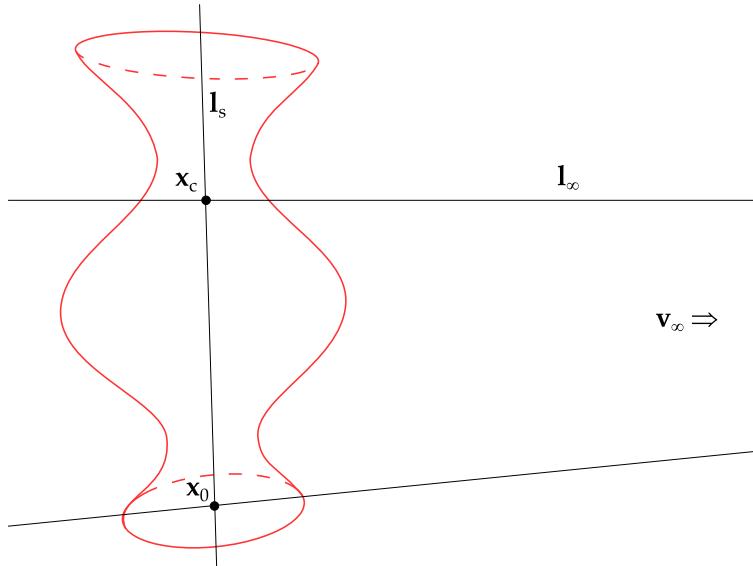


Figura 33. Rette e punti per la stima della matrice di rotazione R .

Stima della matrice di rotazione R

Con le ipotesi fatte, per stimare la matrice R occorre determinare i versori \mathbf{n}_y e \mathbf{n}_z , espressi nel sistema di fotocamera, degli assi y e z del sistema $Oxyz$, avendo traslato l'origine di ques'ultimo sul centro di proiezione: il restante versore \mathbf{n}_x è dato dal prodotto vettoriale dei primi due.

Se ad un punto dell'immagine viene applicata la matrice K^{-1} , quello che si ottiene è un vettore (non omogeneo) che fornisce la direzione, espressa nel sistema di fotocamera, lungo la quale si propaga il raggio che passa per il centro di proiezione ed il punto in questione [15].

Pertanto la scelta di due punti sulla linea l_s permette di determinare due vettori nello spazio giacenti sul piano (x, z) ; il loro prodotto vettoriale a sua volta fornisce un vettore perpendicolare a tale piano. Applicando il ragionamento anche a due punti sulla linea dell'orizzonte l_∞ , otteniamo un vettore perpendicolare al piano (x, y) . Scegliendo opportunamente l'ordine degli operandi nei prodotti vettoriali, il verso dei vettori restituiti coincide con quello degli assi del sistema $Oxyz$; per avere i versori cercati non resta che normalizzare i vettori così ottenuti.

Vediamo come procedere attraverso un esempio (Fig. 33). I punti ritenuti più idonei per calcolare questi versori sono tre:

- \mathbf{v}_∞ , vertice dell'omologia armonica che lega i due lati del contorno apparente e quindi appartenente a l_∞ ;
- x_0 , la proiezione del centro del parallelo assunto come base e pertanto posto su l_s (questo particolare parallelo è stato scelto per semplicità; basterebbe scegliere un punto qualsiasi su l_s);
- x_c , intersezione di l_s con l_∞ .

Si ritiene inoltre noto dalla rettificazione del meridiano anche il verso dell'asse z , cioè se l_s sia diretto da x_0 verso x_c o viceversa.

In Fig. 33 l_s è orientato da x_0 verso x_c . La retta che collega x_0 con v_∞ è l'immagine dell'asse y del sistema Oxyz; avendo posto la fotocamera sul semipiano formato dagli assi x e z , con x positiva, tale retta deve necessariamente essere orientata da x_0 verso v_∞ affinché il sistema di riferimento Oxyz sia destrorso; questo orientamento si ripercuote anche sulla linea dell'orizzonte che viene orientata da x_c verso v_∞ . In virtù di questa situazione, per calcolare il vettore \mathbf{m}_y , perpendicolare al piano (x, z) e diretto come l'asse y , ed il vettore \mathbf{m}_z , perpendicolare al piano (x, y) e diretto come l'asse z , l'ordine degli operandi nei prodotti vettoriali è il seguente:

$$\mathbf{m}_y = (K^{-1}x_0) \times (K^{-1}x_c) = K^\top(x_0 \times x_c) \quad (24a)$$

$$\mathbf{m}_z = (K^{-1}v_\infty) \times (K^{-1}x_c) = K^\top(v_\infty \times x_c) \quad (24b)$$

Da questi vettori si ricava infine la terna cartesiana

$$\left\{ \begin{array}{l} \mathbf{n}_y = \frac{\mathbf{m}_y}{\|\mathbf{m}_y\|} \\ \mathbf{n}_z = \frac{\mathbf{m}_z}{\|\mathbf{m}_z\|} \\ \mathbf{n}_x = \mathbf{n}_y \times \mathbf{n}_z \end{array} \right. \quad (25)$$

e pertanto

$$R = [\mathbf{n}_x \ \mathbf{n}_y \ \mathbf{n}_z] \quad (26)$$

Fig. 33 illustra un particolare caso delle configurazioni che possono assumere i tre punti x_0 , x_c e v_∞ in relazione al verso dato ad l_s : di volta in volta si deve analizzare l'ordine degli operandi nei prodotti vettoriali presenti in Eq. 24 in modo da soddisfare le ipotesi fatte.

Stima del centro di proiezione

Effettuato il calcolo della matrice R , resta da stimare la posizione del centro di proiezione; per fare questo abbiamo bisogno solo di conoscere i coefficienti di un'ellisse immagine di un parallelo ad altezza nota, oltre ad l_s .

Sia quindi $\rho(t)$ il raggio del parallelo ad altezza $z = t$; riprendendo la trasformazione proiettiva tra il piano su cui giace tale circonferenza ed il piano immagine

$$H_t = [\mathbf{p}_1 \ \mathbf{p}_2 \ (t \ \mathbf{p}_3 + \mathbf{p}_4)],$$

introdotta in Eq. 17, abbiamo che il centro del cerchio in questione si proietta sul piano immagine nel punto di coordinate (x_t, y_t) , il cui vettore omogeneo è (si ricorda che \mathbf{p}_j è la j -esima colonna di P):

$$\sigma \mathbf{x}_t = \sigma \begin{bmatrix} x_t \\ y_t \\ 1 \end{bmatrix} = H_t \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} = t \mathbf{p}_3 + \mathbf{p}_4. \quad (27)$$

Un punto qualsiasi sul parallelo considerato si trasforma sull'immagine nel punto

$$\begin{aligned} \mathbf{x}(\vartheta) &= H_t \begin{bmatrix} \rho(t) \cos \vartheta \\ \rho(t) \sin \vartheta \\ 1 \end{bmatrix} = \\ &= \rho(t) \cos \vartheta \mathbf{p}_1 + \rho(t) \sin \vartheta \mathbf{p}_2 + (t \mathbf{p}_3 + \mathbf{p}_4) = \\ &= \rho(t) \cos \vartheta \mathbf{p}_1 + \rho(t) \sin \vartheta \mathbf{p}_2 + \sigma \mathbf{x}_t \end{aligned} \quad (28)$$

e quindi le sue coordinate cartesiane sono

$$\begin{cases} x_\vartheta = \frac{\mathbf{x}_1(\vartheta)}{\mathbf{x}_3(\vartheta)} \\ y_\vartheta = \frac{\mathbf{x}_2(\vartheta)}{\mathbf{x}_3(\vartheta)} \end{cases},$$

da cui¹⁴

$$\frac{y_\vartheta}{x_\vartheta} = \frac{\mathbf{x}_2(\vartheta)}{\mathbf{x}_1(\vartheta)} = \frac{\rho(t) \cos \vartheta p_{2,1} + \rho(t) \sin \vartheta p_{2,2} + \sigma y_t}{\rho(t) \cos \vartheta p_{1,1} + \rho(t) \sin \vartheta p_{1,2} + \sigma x_t}$$

Risolvendo rispetto a σ abbiamo infine

$$\sigma = \frac{(p_{2,1}x_\vartheta - p_{1,1}y_\vartheta) \cos \vartheta + (p_{2,2}x_\vartheta - p_{1,2}y_\vartheta) \sin \vartheta}{y_\vartheta x_t - x_\vartheta y_t} \rho(t). \quad (29)$$

Osservando la forma della matrice P , si nota che il centro di proiezione c interviene solo nella quarta colonna,

$$\mathbf{p}_4 = -KRc; \quad (30)$$

la sostituzione di Eq. 30 in Eq. 27 porta a

$$\begin{aligned} c &= (KR)^{-1}(-\sigma \mathbf{x}_t + t \mathbf{p}_3) = \\ &= -\sigma R^\top K^{-1} \mathbf{x}_t + (0, 0, t)^\top, \end{aligned} \quad (31)$$

dato che \mathbf{p}_3 è anche la terza colonna della matrice (KR) .

Le grandezze in Eq. 29, Eq. 31 sono quindi tutte determinabili:

- con l'algoritmo di ricostruzione del profilo si risale all'altezza t del parallelo relativo ad una delle ellissi trovate dall'algoritmo di segmentazione o ricostruita tramite l'opportuna omologia;
- la proiezione x_t del centro del parallelo in questione è ricavabile dalla relazione *polo-polare* dell'ellisse dato con la linea dell'orizzonte;
- preso infine un punto qualsiasi sull'ellisse considerato, si risale all'angolo ϑ utilizzando la formula di Laguerre [6] (in particolare, il punto *più vicino* tra i due di intersezione tra \mathbf{l}_s e l'ellisse assume $\vartheta = 0$; il segno di ϑ è stabilito in base a come la linea dell'orizzonte è stata orientata nel calcolo della matrice R).

¹⁴ La notazione $p_{i,j}$ indica l'elemento (i, j) della matrice P .

2.8.3 Acquisizione della texture del SOR

La completa conoscenza dei parametri interni ed esterni, congiuntamente al profilo ricostruito, permette di determinare in quale punto sull'immagine viene proiettato un generico punto sul SOR. Ovviamente tutto ciò ha senso per i punti che sono effettivamente visibili; le occlusioni che la vista del SOR può presentare hanno duplice natura.

Facendo riferimento alla parametrizzazione di Eq. 5 in (t, ϑ) del SOR, fissato un parallelo infatti sono visibili solo alcune delle sue intersezioni con i meridiani, quelli *di fronte*, (*occlusioni in ϑ o angolari*); similmente, fissato un meridiano non è detto che siano visibili le sue intersezioni con tutti i paralleli (*occlusioni in ϑ o in altezza*) [6].

Le occlusioni in altezza possono essere rilevate individuando nel contorno apparente i punti in cui la curva non è derivabile (cioè la retta ad essa tangente non è univoca)¹⁵; le occlusioni angolari possono essere invece determinate calcolando per ogni parallelo non occluso in altezza l'angolo sotteso dalla fotocamera. Tale angolo può essere determinato prendendo i punti sul contorno apparente che appartiene all'ellisse immagine del parallelo considerato e misurando sempre con la formula di Laguerre l'angolo che essi formano; un metodo più immediato può essere invece quello di utilizzare le coordinate del centro di proiezione per calcolare, tramite Eq. 22, i due punti che quel parallelo fornisce al generatore di contorni.

Qualora sia stata fatta una buona stima dei parametri di fotocamera, la proiezione del modello sull'immagine permette di determinare velocemente il colore posseduto da ogni punto visibile del SOR in alternativa al metodo proposto in [6], comunque meno sensibile agli errori di calibrazione, in quanto utilizza solo i parametri di fotocamera interni. L'acquisizione della texture su di una superficie piana è effettuata secondo un approccio di *inverse texture mapping*, proiettando i punti del SOR ortogonalmente ad un cilindro coassiale con il SOR e srotolandolo poi tale cilindro (una classica proiezione cartografica): pertanto all'immagine piana della texture, che riporta in ascissa l'angolo ϑ ed in ordinata lo scalare t della parametrizzazione di Eq. 5 del SOR, viene applicata una griglia regolare e per le coppie visibili (t, ϑ) così campionate, si procede alla determinazione delle coordinate immagine e del colore assunto dal pixel relativo.

Le figure seguenti mostrano la texture acquisita ed adattata sul modello 3D attraverso la proiezione di quest'ultimo sull'immagine usando la matrice P calcolata con le tecniche esposte nella sezione 2.8.2: per il calcolo del centro di proiezione è stata usata l'ellisse della base del vaso ($t = 0$) e come punto campione x_ϑ è stato scelto quello con $\vartheta = 0$. Oltre a mostrare la campana usata finora (Fig. 34), che ha una texture poco significativa, è stato deciso di usare anche un vaso riccamente decorato (Fig. 35).

¹⁵ Potrebbe essere sviluppato in tal senso un opportuno algoritmo di segmentazione.



(a) Modello 3D con texture sovrapposta.



(b) Foto originale della campana.

Figura 34. Modello 3D della campana con texture sovrapposta (a) confrontato con la fotografia di partenza (b).



(a) Immagine originale, con le curve segmentate automaticamente (ellissi e contorno apparente).

(b) Modello 3D del vaso, osservato dal punto di vista mostrato in Fig. 36.

Figura 35. Altro esempio di segmentazione (a) e ricostruzione 3D con texture (b).

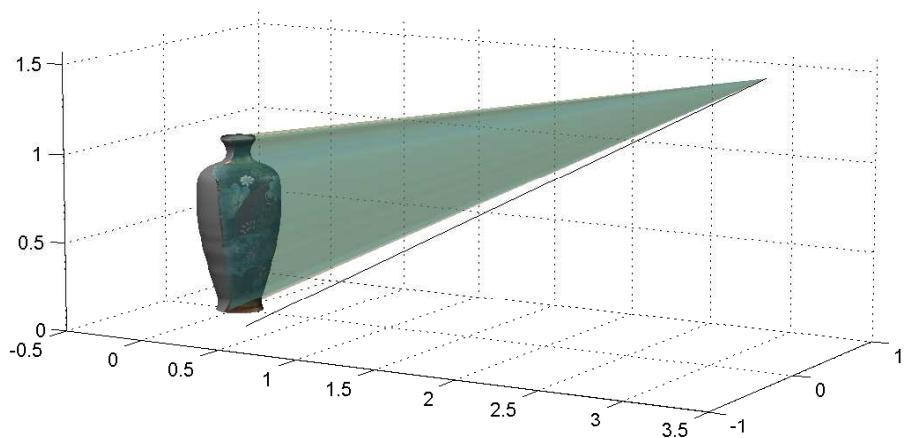


Figura 36. Ricostruzione del punto di vista (centro di proiezione e direzione dell'asse ottico).



Figura 37. Proiezione cartografica in coordinate cilindriche della texture del vaso: le righe dell'immagine corrispondono ai paralleli del SOR, mentre le colonne ai meridiani.

Fig. 36 illustra la ricostruzione del punto di vista da dove la fotografia in Fig. 35(a) è stata fatta, seguendo il procedimento della sezione 2.8.2. Attraverso la parametrizzazione Di Eq. 5 (qui riportata per semplicità)

$$\varphi(t, \vartheta) = \begin{cases} x(t, \vartheta) = \rho(t) \cos \vartheta \\ y(t, \vartheta) = \rho(t) \sin \vartheta \\ z(t, \vartheta) = t; \end{cases}$$

è infine possibile mappare in coordinate cilindriche sul piano (ϑ, t) la texture del vaso, come mostrato in Fig. 37.

3

ESPERIMENTI

QUESTO CAPITOLO RIPORTA il comportamento dell'algoritmo sia su immagini sintetiche (per confronti con dei dati di ground-truth), sia su immagini reali, per le quali è possibile fornire solo un giudizio qualitativo. L'algoritmo è stato implementato in MATLAB® 6.5, ed è stato provato su un computer dotato di un processore a 1.66 GHz; il tempo di esecuzione sugli esempi riportati varia di caso in caso, in base alla dimensione delle immagini, ricchezza di dettagli presenti ed al numero di curve estratte (indicativamente, cinque minuti su immagini di media risoluzione, come 800×600 o 640×480).

3.1 ESPERIMENTI CON IMMAGINI SINTETICHE

Sono stati condotti una serie di esperimenti su delle sagome ottenute proiettando un SOR, con profilo $\rho(t)$ noto (Fig. 38), su un piano tramite una matrice $P = KR[I_{3 \times 3} | -c]$ (descrivente quindi il modello “natural camera”) fissata a priori; la funzione per il profilo è:

$$\rho(t) = \frac{\cos\left(\frac{\pi}{2}\left(\frac{19}{3}t + 1\right)\right) + 2}{10}.$$

Gli esperimenti hanno riguardato la ricostruzione del SOR sia in condizioni vicine al caso degenere (il punto principale viene a trovarsi sull'immagine dell'asse di rivoluzione l_s), sia meno critiche; sarà fatto riferimento ad essi come *caso degenere* e *non degenere*. I due casi sono illustrati in Fig. 39; sono riportati anche l_s , la linea dell'orizzonte creata dal piano su cui è appoggiato il SOR, ed il punto principale.

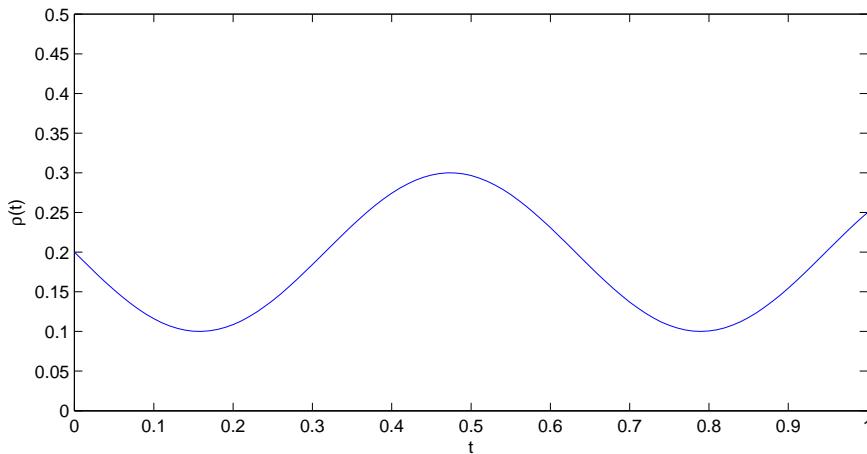


Figura 38. Il profilo usato per la generazione del SOR sintetico.

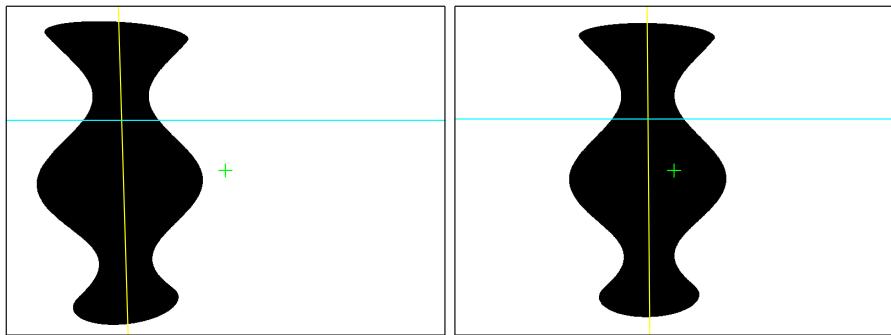


Figura 39. Sagome utilizzate per gli esperimenti sintetici: a destra, il caso degenere; a sinistra, la disposizione non degenera. La retta orizzontale è la linea dell'orizzonte creata dal piano su cui è appoggiato il SOR; l'altra linea è l'asse di rivoluzione 1_s ; la croce + indica il punto principale.

Le due immagini, con risoluzione di 800×600 pixel, sono state ottenute utilizzando per entrambe gli stessi valori per \mathbf{c} e K ; ciò che sposta la posizione del SOR è la differente matrice R impiegata. Più precisamente, i parametri per P sono stati:

$$\begin{aligned}\mathbf{c} &= [1.6, 0, 0.7]^\top \\ K &= \begin{bmatrix} 750 & 0 & 400 \\ 0 & 750 & 300 \\ 0 & 0 & 1 \end{bmatrix} \\ R_{\text{deg}} &= \begin{bmatrix} 0.062378 & 0.998050 & 0 \\ 0.123560 & -0.007722 & -0.99231 \\ -0.990380 & 0.061898 & -0.12380 \end{bmatrix} \\ R_{\text{ndeg}} &= \begin{bmatrix} 0.242540 & 0.970140 & 0 \\ 0.116790 & -0.029198 & -0.992730 \\ -0.963090 & 0.240770 & -0.120390 \end{bmatrix}\end{aligned}$$

Poiché la matrice K per la “natural camera” ha struttura

$$K = \begin{bmatrix} f & p_x \\ f & p_y \\ 1 \end{bmatrix}$$

si evince che per la focale vale $f = 750$ e che le coordinate del punto principale sono $(400, 300)$. Le coordinate del centro di proiezione \mathbf{c} sono espresse nel sistema di riferimento solidale con il SOR, con l'asse z coassiale ad esso, e con il piano (x, y) come suo piano d'appoggio. Le matrici di rotazione R_{deg} e R_{ndeg} sono state ottenute indicando un punto nello spazio verso il quale rivolgere l'asse ottico

(l'asse z del sistema di fotocamera); gli assi x e y di questo sistema sono stati calcolati, senza perdere di generalità, in maniera tale che la linea dell'orizzonte nell'immagine fosse orizzontale. I punti utilizzati per dirigere l'asse ottico sono stati $[0 \ 0.4 \ 0.5]^\top$ e $[0 \ 0.1 \ 0.5]^\top$, rispettivamente nel caso degenero e non. La seconda coordinata di \mathbf{c} è stata posta a 0 per soddisfare le ipotesi della sezione 2.8.2 e confrontare così direttamente i dati ricavati con il metodo esposto in essa.

Gli esperimenti sono stati svolti con un approccio *Montecarlo*: alle sagome di Fig. 39 è stato aggiunto del rumore Gaussiano a media nulla, diretto secondo la normale al contorno apparente, e si è analizzato il comportamento della stima di P e $\rho(t)$ al crescere della varianza σ^2 del rumore. Per ogni valore di σ fissato, sono state generate 50 istanze di rumore e per ognuna di esse è stato lanciato l'algoritmo¹; per ogni gruppo di 50 esperimenti, i dati ritenuti d'interesse sono stati:

- l'angolo formato da \mathbf{l}_s e dalla sua stima, $\hat{\mathbf{l}}_s$;
- \mathbf{v}_∞ ;
- la focale;
- il punto principale;
- l'angolo formato dai versori individuati dalle colonne di R e da quelli stimati;
- il centro di proiezione \mathbf{c} , ricavato con due differenti valori (0° e 90°) per l'angolo ϑ in Eq. 29;
- le massime escursioni (positive e negative) tra $\rho(t)$ e la sua stima, $\hat{\rho}(t)$;
- l'errore complessivo nella ricostruzione del profilo.

Per la valutazione di quest'ultimo, sono stati usati un valor medio ed una deviazione standard, definiti rispettivamente come

$$ric_{med} = \int_0^1 |\hat{\rho}(t) - \rho(t)| dt, \quad (33a)$$

$$ric_{std} = \sqrt{\int_0^1 (\hat{\rho}(t) - \rho(t))^2 dt}. \quad (33b)$$

Dal momento che ciascuna istanza del rumore aggiunto è casuale, tutte le grandezze del precedente elenco diventano delle variabili aleatorie e pertanto descrivibili per mezzo del loro valor medio e deviazione standard, stimati sulle loro 50 manifestazioni. I valori scelti per "l'intensità del rumore" σ sono stati 0, 0.1, 0.2, 0.4, 0.8 e 1.6; anche nel caso di assenza di disturbo ($\sigma = 0$) ha senso di parlare di media

¹ Un numero superiore di prove avrebbe richiesto tempi molto lunghi: l'analisi di ciascun gruppo di 50 istanze ha richiesto infatti più di 4 ore.

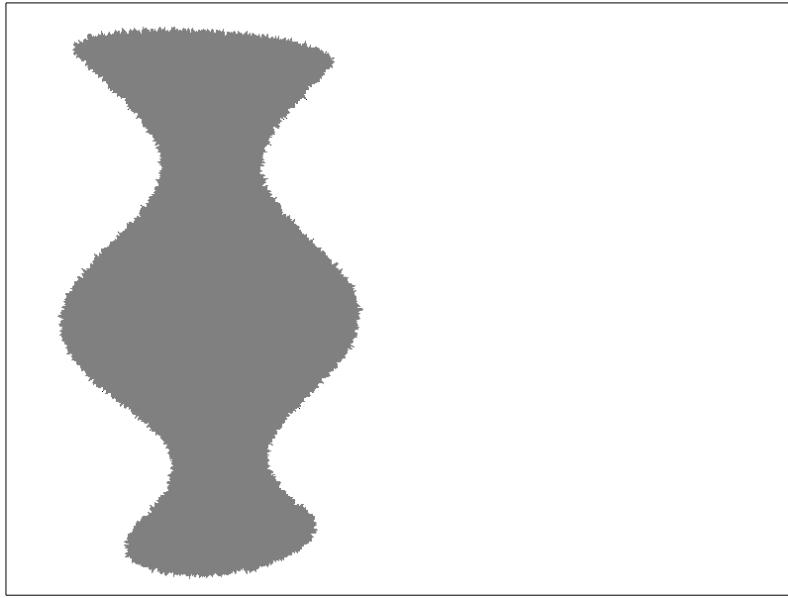


Figura 40. Istanza di disturbo aggiunto alla sagoma del caso non degenere, per $\sigma = 1.6$.

e deviazione standard, in quanto il meccanismo casuale alla base dell'algoritmo fa sì che due esecuzioni sulla stessa immagine producano stime leggermente differenti. Fig. 40 mostra la sagoma del SOR nel caso non degenere, alterata da rumore con $\sigma = 1.6$; tra i valori scelti per σ , questo è quello che distorce maggiormente la silhouette.

Al fine di automatizzare la fase dei test, l'algoritmo è stato diviso in tre parti; ciascuna parte viene eseguita su tutte le istanze che riceve, filtrate da due procedure di controllo per l'individuazione dei casi in cui l'algoritmo è “fuori strada”:

- la prima procedura, subito dopo la parte di stima dell'omologia armonica, analizza con LMedS la distribuzione dell'angolo formato da l_s e \hat{l}_s (ritenendo erronei i casi distanti dalla moda più di 6 volte la deviazione standard di tale distribuzione);
- la seconda, al termine della fase di segmentazione, scarta i casi in cui o non sono state individuate entrambe le due ellissi, o non è possibile fattorizzare la IAC con Cholesky (è necessario che la matrice sia definita positiva) per la stima di K .

I risultati sono stati raccolti in tabelle, contenenti i dati sia per il caso degenere che per quello non degenere. Le prime tre tabelle (Tab. 1–Tab. 3) riportano l'analisi condotta dalle procedure di controllo, a cui seguono i risultati per la stima di v_∞ (Tab. 4 e Fig. 41).

Si passa quindi ai dati relativi alla calibrazione interna: Tab. 5 e Fig. 42 sono relative alla stima del punto principale, mentre Tab. 6 riguarda la stima della lunghezza focale.

Per quanto concerne la calibrazione esterna, la stima della matrice di rotazione R viene riportata in tre tabelle (Tab. 7–Tab. 9, una per

ogni colonna di R); la stima del centro di proiezione² è riassunta da Tab. 10 e Tab. 11, a seconda che l'angolo ϑ usato nel procedimento di stima (in Eq. 28) sia 0° o 90° , ed illustrata da Fig. 43.

Gli esperimenti sintetici sono chiusi da Tab. 12 dove sono riportati i risultati per l'errore di ricostruzione ric_{med} (Eq. 33a), così come in Tab. 13 per ric_{std} (Eq. 33b). Infine, Tab. 14 riassume i dati sulle escursioni tra il profilo ricostruito e quello di ground truth.

Tabella 1. Analisi della distribuzione dell'angolo (misurato in gradi) tra \mathbf{l}_s e $\hat{\mathbf{l}}_s$ con LMedS.

σ	moda $\angle(\mathbf{l}_s, \hat{\mathbf{l}}_s)$		varianza $\angle(\mathbf{l}_s, \hat{\mathbf{l}}_s)$	
	deg.	non deg.	deg.	non deg.
0	-0.0003825	-0.0128	0.008216	0.00500
0.1	-0.0170299	-0.0095	0.011640	0.00812
0.2	-0.0158114	-0.0074	0.013167	0.01487
0.4	-0.0168158	-0.0107	0.019467	0.01184
0.8	-0.0221651	-0.0057	0.023335	0.02195
1.6	0.0199828	-0.0325	0.042331	0.06720

Tabella 2. Analisi degli outlier secondo LMedS nella distribuzione dell'angolo tra \mathbf{l}_s e $\hat{\mathbf{l}}_s$.

σ	# outlier		min $\angle(\mathbf{l}_s, \hat{\mathbf{l}}_s)$		max $\angle(\mathbf{l}_s, \hat{\mathbf{l}}_s)$	
	deg.	non deg.	deg.	non deg.	deg.	non deg.
0	2	2	0.0976	0.0347	0.0992	0.0375
0.1	5	1	0.0930	89.5910	0.1021	89.5910
0.2	1	1	0.1066	1.8443	0.1066	1.8443
0.4	2	1	0.1045	89.6640	0.1061	89.6640
0.8	1	0	0.1489	—	0.1489	—
1.6	0	16	—	89.4410	—	89.778

Tabella 3. Casi in cui le ellissi estratte non permettono il proseguimento della ricostruzione.

σ	# casi	
	deg.	non deg.
0	0	0
0.1	0	0
0.2	0	0
0.4	2	0
0.8	2	0
1.6	4	5

² In conseguenza delle ipotesi fatte nella sezione 2.8.2, la coordinata y del centro di proiezione deve avere valore nullo; nella realtà ciò non sempre succede a causa di piccoli errori (forse di rappresentazione finita dei numeri) nella matrice R . Poiché gli scarti sono trascurabili (dell'ordine di 10^{-16}) si è deciso di ometterne l'esposizione.

Tabella 4. Stima di \mathbf{v}_∞ . Ground truth: (12493.0, 206.4) per il caso degenere (3422.0, 209.0) nell’altro.

σ	media(x)		std(x)		media(y)		std(y)	
	deg.	non deg.	deg.	non deg.	deg.	non deg.	deg.	non deg.
0	12879.9	3426.8	222.6	6.4	203.0	206.8	2.0	0.4
0.1	12703.3	3432.5	260.5	9.9	203.2	207.1	1.6	0.4
0.2	12624.2	3426.3	219.0	11.9	203.6	208.1	2.5	0.7
0.4	12586.9	3423.8	280.0	17.3	203.8	208.1	3.5	1.2
0.8	12702.7	3436.6	462.0	36.3	203.5	207.6	5.6	1.6
1.6	12494.3	3449.5	842.3	56.9	206.1	209.0	15.3	3.1

Tabella 5. Stima del punto principale. Ground truth: (400, 300).

σ	media(x)		std(x)		media(y)		std(y)	
	deg.	non deg.	deg.	non deg.	deg.	non deg.	deg.	non deg.
0	397.5	403.6	0.8	4.4	300.2	298.5	2.6	0.4
0.1	397.9	400.8	1.0	3.9	300.6	299.1	2.4	0.6
0.2	398.1	399.8	0.9	5.6	300.1	300.1	3.2	1.0
0.4	398.2	399.5	1.1	6.9	299.0	299.9	4.9	1.2
0.8	398.0	406.8	1.8	22.9	300.3	299.4	7.3	2.1
1.6	427.0	428.6	82.8	132.4	297.7	299.9	18.1	4.9

Tabella 6. Stima della focale f. Ground truth 750.

σ	media(f)		std(f)		min(f)		max(f)	
	deg.	non deg.	deg.	non deg.	deg.	non deg.	deg.	non deg.
0	743.2	757.8	1.5	7.7	738.1	742.7	745.8	773.9
0.1	741.4	753.2	1.9	6.6	735.5	740.6	744.5	769.3
0.2	740.5	750.5	3.0	9.2	735.9	738.2	750.9	788.0
0.4	739.7	749.7	4.8	11.7	731.8	739.9	753.5	803.8
0.8	741.8	763.5	8.1	35.2	723.3	729.3	763.3	943.0
1.6	864.0	783.5	363.6	7.1	646.4	723.1	2050.0	1448.4

Tabella 7. Angolo (misurato in gradi) tra la prima colonna di R, \mathbf{n}_x , e la sua stima, $\hat{\mathbf{n}}_x$.

σ	media($\angle(\mathbf{n}_x, \hat{\mathbf{n}}_x)$)		std($\angle(\mathbf{n}_x, \hat{\mathbf{n}}_x)$)		min($\angle(\mathbf{n}_x, \hat{\mathbf{n}}_x)$)		max($\angle(\mathbf{n}_x, \hat{\mathbf{n}}_x)$)	
	deg.	non deg.	deg.	non deg.	deg.	non deg.	deg.	non deg.
0	0.403	0.104	0.149	0.031	0.234	0.070	0.515	0.128
0.1	0.128	0.125	0.062	0.029	0.056	0.097	0.167	0.155
0.2	0.082	0.144	0.017	0.039	0.066	0.102	0.100	0.178
0.4	0.353	0.236	0.264	0.046	0.048	0.195	0.507	0.286
0.8	0.231	0.445	0.089	0.170	0.130	0.252	0.295	0.570
1.6	0.567	0.539	0.131	0.390	0.445	0.089	0.706	0.764

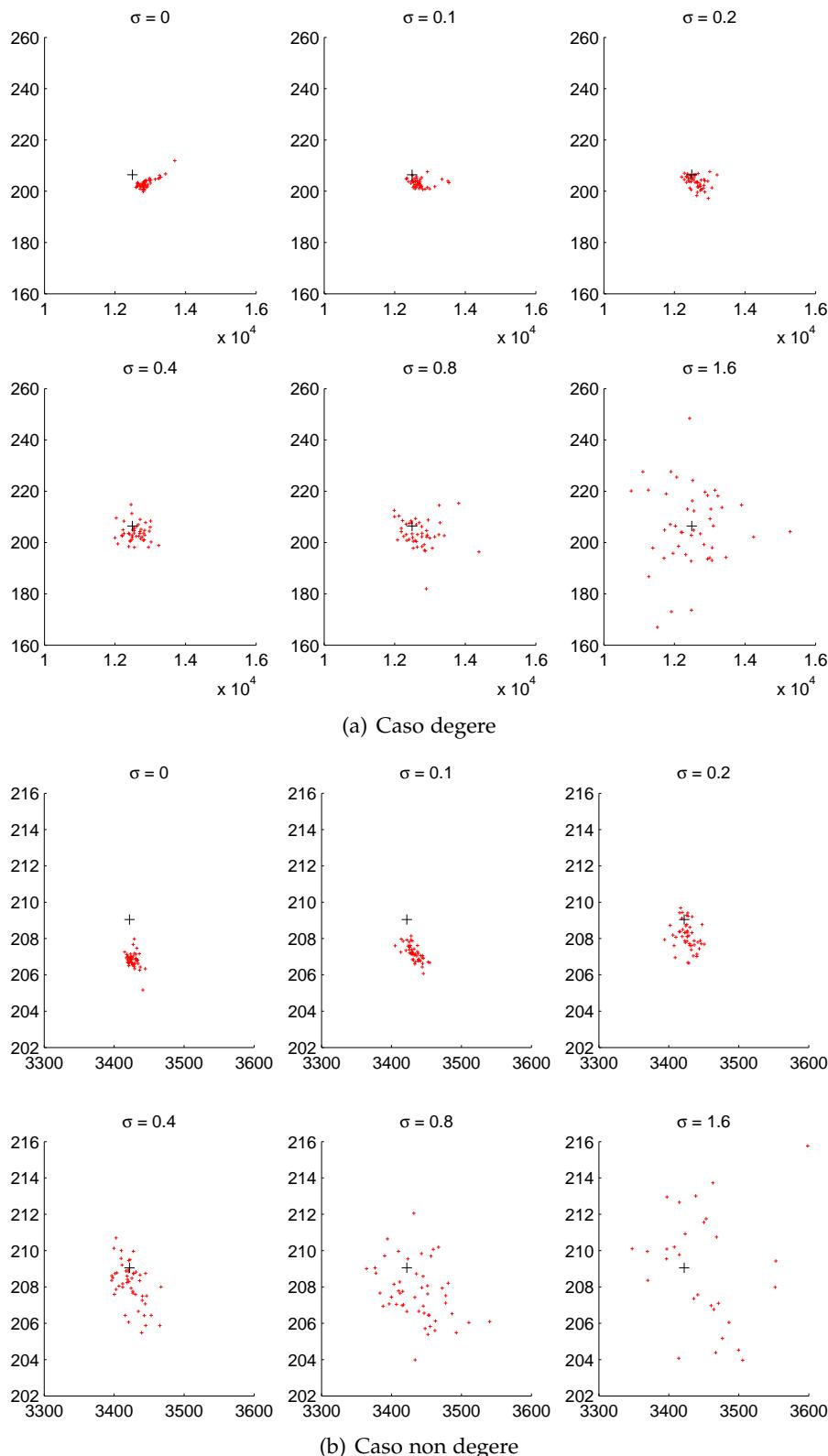


Figura 41. Stime di v_∞ per ogni gruppo di esperimenti; la croce maggiore indica il vero v_∞ .

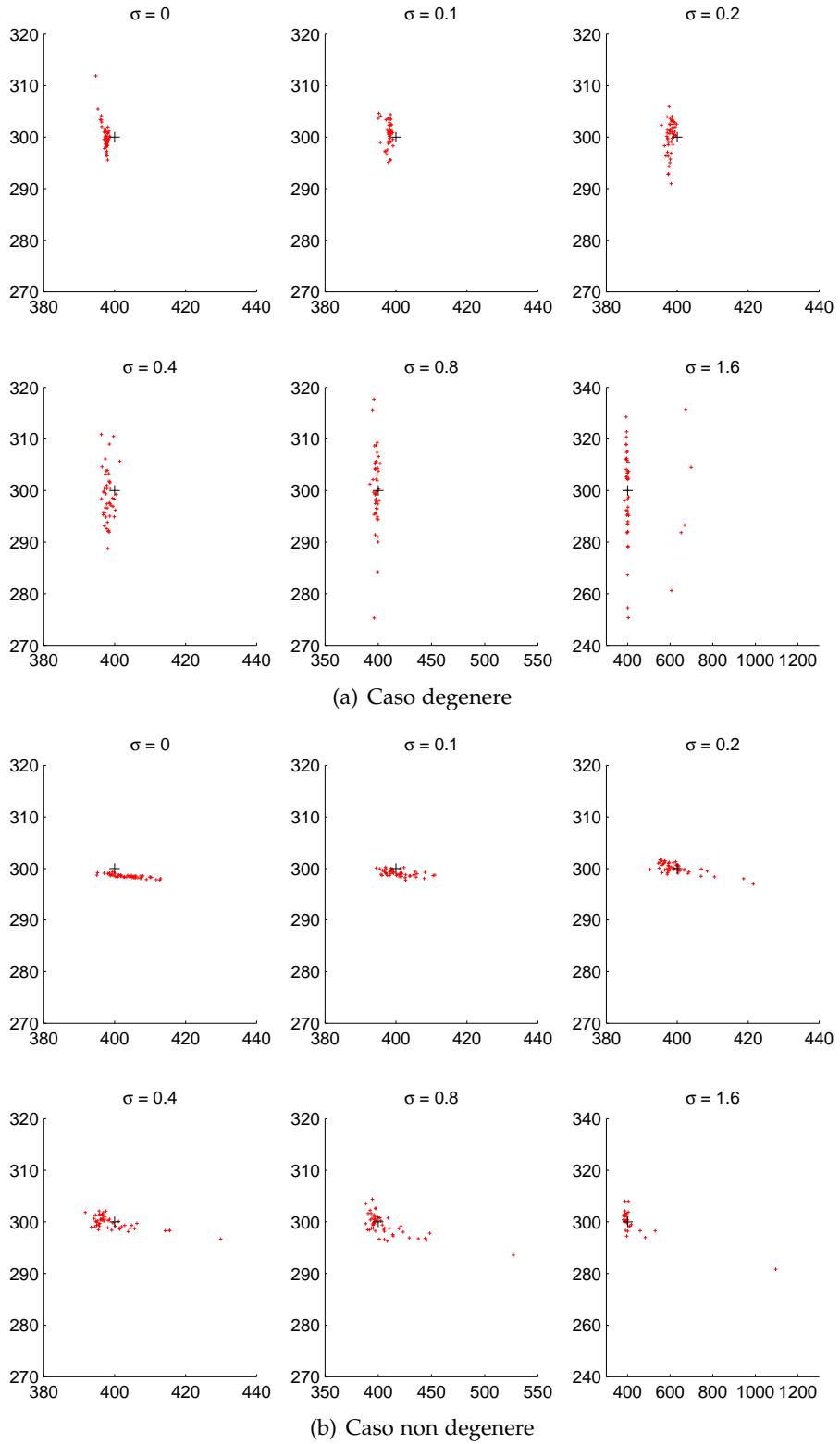


Figura 42. Stime del punto principale per ogni gruppo di esperimenti; la croce maggiore indica il vero punto principale.

Tabella 8. Angolo (misurato in gradi) tra la seconda colonna di R , \mathbf{n}_y , e la sua stima, $\hat{\mathbf{n}}_y$.

σ	media($\angle(\mathbf{n}_y, \hat{\mathbf{n}}_y)$)		std($\angle(\mathbf{n}_y, \hat{\mathbf{n}}_y)$)		min($\angle(\mathbf{n}_y, \hat{\mathbf{n}}_y)$)		max($\angle(\mathbf{n}_y, \hat{\mathbf{n}}_y)$)	
	deg.	non deg.	deg.	non deg.	deg.	non deg.	deg.	non deg.
0	0.104	0.243	0.022	0.119	0.084	0.106	0.128	0.321
0.1	0.166	0.059	0.093	0.011	0.059	0.046	0.224	0.068
0.2	0.371	0.156	0.230	0.063	0.105	0.083	0.509	0.201
0.4	0.294	0.101	0.194	0.057	0.070	0.036	0.409	0.135
0.8	0.246	0.428	0.172	0.265	0.047	0.122	0.347	0.587
1.6	1.401	3.624	1.191	0.769	0.026	3.038	2.089	4.494

Tabella 9. Angolo (misurato in gradi) tra la terza colonna di R , \mathbf{n}_z , e la sua stima, $\hat{\mathbf{n}}_z$.

σ	media($\angle(\mathbf{n}_z, \hat{\mathbf{n}}_z)$)		std($\angle(\mathbf{n}_z, \hat{\mathbf{n}}_z)$)		min($\angle(\mathbf{n}_z, \hat{\mathbf{n}}_z)$)		max($\angle(\mathbf{n}_z, \hat{\mathbf{n}}_z)$)	
	deg.	non deg.	deg.	non deg.	deg.	non deg.	deg.	non deg.
0	0.141	0.822	0.032	0.282	0.113	0.505	0.176	1.044
0.1	0.089	0.098	0.033	0.054	0.051	0.036	0.112	0.132
0.2	0.257	0.251	0.181	0.059	0.049	0.190	0.362	0.308
0.4	0.278	0.310	0.139	0.169	0.118	0.114	0.366	0.409
0.8	0.232	0.210	0.165	0.102	0.041	0.092	0.328	0.270
1.6	0.958	0.428	0.749	0.147	0.094	0.261	1.392	0.538

Tabella 10. Stima delle coordinate del centro di proiezione per $\vartheta = 0^\circ$.
Ground truth: $x = 1.6$, $z = 0.7$.

σ	media(x)		std(x)		media(z)		std(z)	
	deg.	non deg.	deg.	non deg.	deg.	non deg.	deg.	non deg.
0	1.587	1.615	0.003	0.016	0.702	0.699	0.002	0.005
0.1	1.585	1.604	0.003	0.013	0.701	0.703	0.002	0.004
0.2	1.582	1.600	0.006	0.019	0.700	0.700	0.002	0.005
0.4	1.581	1.600	0.009	0.024	0.700	0.696	0.003	0.006
0.8	1.582	1.625	0.015	0.072	0.702	0.694	0.005	0.018
1.6	1.799	1.664	0.659	0.283	0.813	0.724	0.324	0.192

Tabella 11. Stima delle coordinate del centro di proiezione per $\vartheta = 90^\circ$. Ground truth: $x = 1.6$, $z = 0.7$.

σ	media(x)		std(x)		media(z)		std(z)	
	deg.	non deg.	deg.	non deg.	deg.	non deg.	deg.	non deg.
0	1.587	1.615	0.003	0.016	0.702	0.700	0.002	0.005
0.1	1.585	1.604	0.003	0.013	0.701	0.703	0.002	0.004
0.2	1.583	1.600	0.006	0.019	0.700	0.700	0.002	0.005
0.4	1.581	1.600	0.009	0.024	0.700	0.696	0.003	0.006
0.8	1.582	1.626	0.015	0.073	0.702	0.694	0.005	0.018
1.6	1.799	1.665	0.659	0.283	0.813	0.724	0.324	0.192

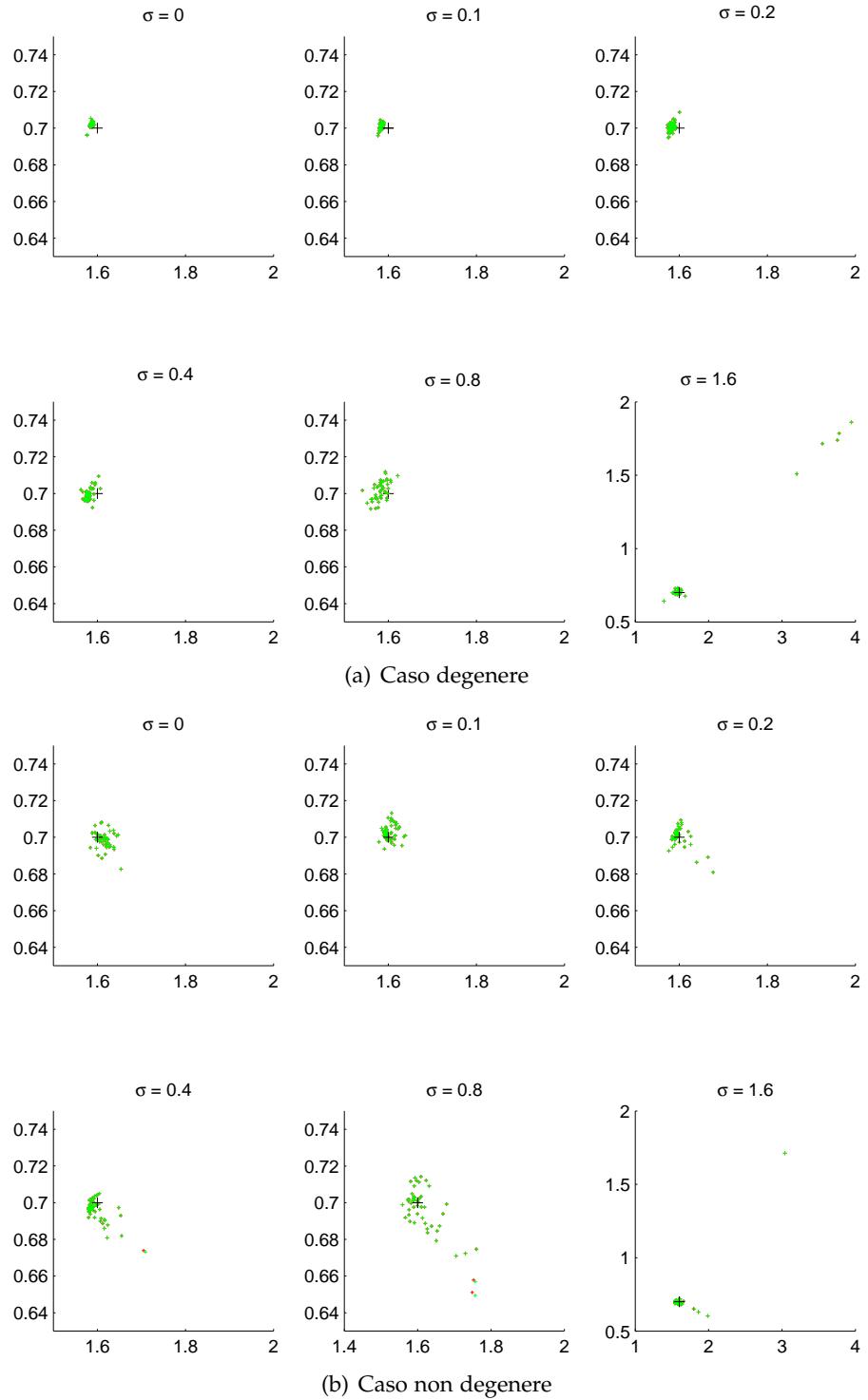


Figura 43. Stime del centro di proiezione nel piano (x, z) per ogni gruppo di esperimenti; la croce maggiore indica il vero centro, i punti in verde corrispondono a $\vartheta = 90^\circ$, quelli in rosso a $\vartheta = 0^\circ$.

Tabella 12. Errore ric_{med} per $\hat{\rho}(t)$.

σ	media(ric_{med})		std(ric_{med})		min(ric_{med})		max(ric_{med})	
	deg.	non deg.	deg.	non deg.	deg.	non deg.	deg.	non deg.
0	0.001	0.001	0.0002	0.0002	0.001	0.001	0.002	0.002
0.1	0.002	0.001	0.0002	0.0002	0.001	0.001	0.002	0.001
0.2	0.002	0.001	0.0002	0.0003	0.001	0.001	0.002	0.002
0.4	0.002	0.001	0.0002	0.0003	0.001	0.001	0.002	0.003
0.8	0.002	0.002	0.0004	0.0015	0.001	0.001	0.003	0.011
1.6	0.004	0.004	0.0029	0.0061	0.002	0.001	0.012	0.033

Tabella 13. Errore $ric_{std} \cdot 10^{-3}$ per $\hat{\rho}(t)$.

σ	media(ric_{med})		std(ric_{med})		min(ric_{med})		max(ric_{med})	
	deg.	non deg.	deg.	non deg.	deg.	non deg.	deg.	non deg.
0	0.002	0.001	0.001	0.001	0.001	0.001	0.004	0.003
0.1	0.003	0.002	0.002	0.002	0.002	0.001	0.016	0.012
0.2	0.003	0.003	0.001	0.002	0.001	0.001	0.007	0.012
0.4	0.003	0.003	0.001	0.002	0.002	0.001	0.006	0.014
0.8	0.004	0.011	0.002	0.033	0.002	0.002	0.012	0.232
1.6	0.029	0.065	0.052	0.242	0.005	0.002	0.196	1.284

Tabella 14. Massime escursioni di $\hat{\rho}(t)$ rispetto a $\rho(t)$.

σ	min($\rho(t) - \hat{\rho}(t)$)		max($\rho(t) - \hat{\rho}(t)$)	
	deg.	non deg.	deg.	non deg.
0	-0.003733	-0.004804	0.002446	0.003163
0.1	-0.004483	-0.003925	0.001914	0.004080
0.2	-0.004752	-0.008090	0.002147	0.008013
0.4	-0.005686	-0.010955	0.003031	0.012367
0.8	-0.006754	-0.027293	0.004406	0.047915
1.6	-0.048916	-0.030763	0.030719	0.058598

3.1.1 Analisi dei risultati sintetici

La prima osservazione che è possibile compiere riguarda il deterioramento delle performance, al crescere del disturbo: questo è osservabile, ad esempio, attraverso le deviazioni standard di tutte le stime, che tendono ad aumentare.

Ai fini della ricostruzione del profilo $\rho(t)$, i risultati sono stati piuttosto buoni: ad esempio, per l'errore complessivo ric_{med} , (tranne il massimo riscontrato pari a 0.033, per il caso non degenere con $\sigma = 1.6$) gli errori commessi sono piccoli, anche tenendo conto che l'integrale di $\rho(t)$ sullo stesso dominio vale circa 0.181 (il confronto è lecito: ric_{med} può essere infatti interpretato come l'area della regione di piano compresa tra $\hat{\rho}(t)$ e $\rho(t)$).

Per quanto riguarda la calibrazione esterna, e più in particolare la

stima del centro di proiezione, non si notano particolari differenze tra l'uso di $\vartheta = 0^\circ$ e $\vartheta = 90^\circ$; il confronto tra il caso degenere e non, sembra indicare la presenza di un bias nella stima della coordinata x per il caso degenere, meno evidente nel secondo. L'analisi delle varianze però è favorevole al primo che mostra dei valori più bassi. Meno chiara è la valutazione sulla matrice R , anche se gli angoli tra i versori (dati dalle sue colonne) veri e quelli stimati sono frazioni di grado tranne che per $\sigma = 1.6$.

La calibrazione interna presenta un comportamento curioso: confrontando le deviazioni standard sia per la focale che per il punto principale, notiamo come esse siano più ridotte per il caso degenere. Inoltre, andando ad osservare la distribuzione del punto principale nel piano (x, y) (Fig. 42) si notano dei cluster allungati secondo direzioni diverse nei due casi: nel caso degenere l'incertezza è maggiore lungo l'asse y , nell'altro lungo l'asse x . Infine, alle intensità di rumore più elevate, il caso non degenere presenta cluster più compatti.

Per terminare, un commento sulla stima di v_∞ : in entrambi i casi, degenere e non, si nota un bias nella sua stima (Fig. 41) in assenza di rumore, che poi tende a scomparire quando il rumore aumenta. Una possibile spiegazione è che esso sia dovuta all'edge-detector: quest'ultimo infatti, produce degli edge che, orlando la sagoma, sono leggermente spostati rispetto al vero contorno apparente; il bias si attenua quando la sagoma viene zigrinata dal rumore perché in tal caso gli edge oscillano intorno alla vera silhouette.

3.2 ESPERIMENTI CON IMMAGINI REALI

In questa sezione sono riportate alcuni esempi di ricostruzione di SOR utilizzando delle fotografie, trovate perlopiù su internet, che ritraggono per lo più dei comuni vasi; l'assenza di dati di ground truth lascia spazio solo ad una valutazione qualitativa del risultato finale.

L'unico tipo di analisi che è stato possibile condurre riguarda il caso in cui sono state disponibili più foto dello stesso SOR, scattate da differenti punti di vista: in tale situazione è stato possibile eseguire la proiezione del modello, ricavato da una delle due foto, sull'altra ed osservare la differenza tra l'immagine vera ed il profilo del modello.

Una buona sovrapposizione indica che le due ricostruzioni si sostengono a vicenda, come mostrato in Fig. 44; in Fig. 45 si mettono inoltre a confronto i modelli 3D ricavati dalle due foto. Infine, Fig. 46 mostra la rettificazione della banconota visibile in Fig. 44(d), ottenuta riportando i punti circolari delle due ellissi del SOR nella loro posizione canonica: la banconota infatti giace su un piano parallelo alle due ellissi, per cui i punti circolari sono in comune.

Segue una carrellata di immagini segmentate (Fig. 47–Fig. 54), con le ricostruzioni dei SOR in esse ritratti; le didascalie riportano anche eventuali commenti sul comportamento dell'algoritmo nei vari casi.

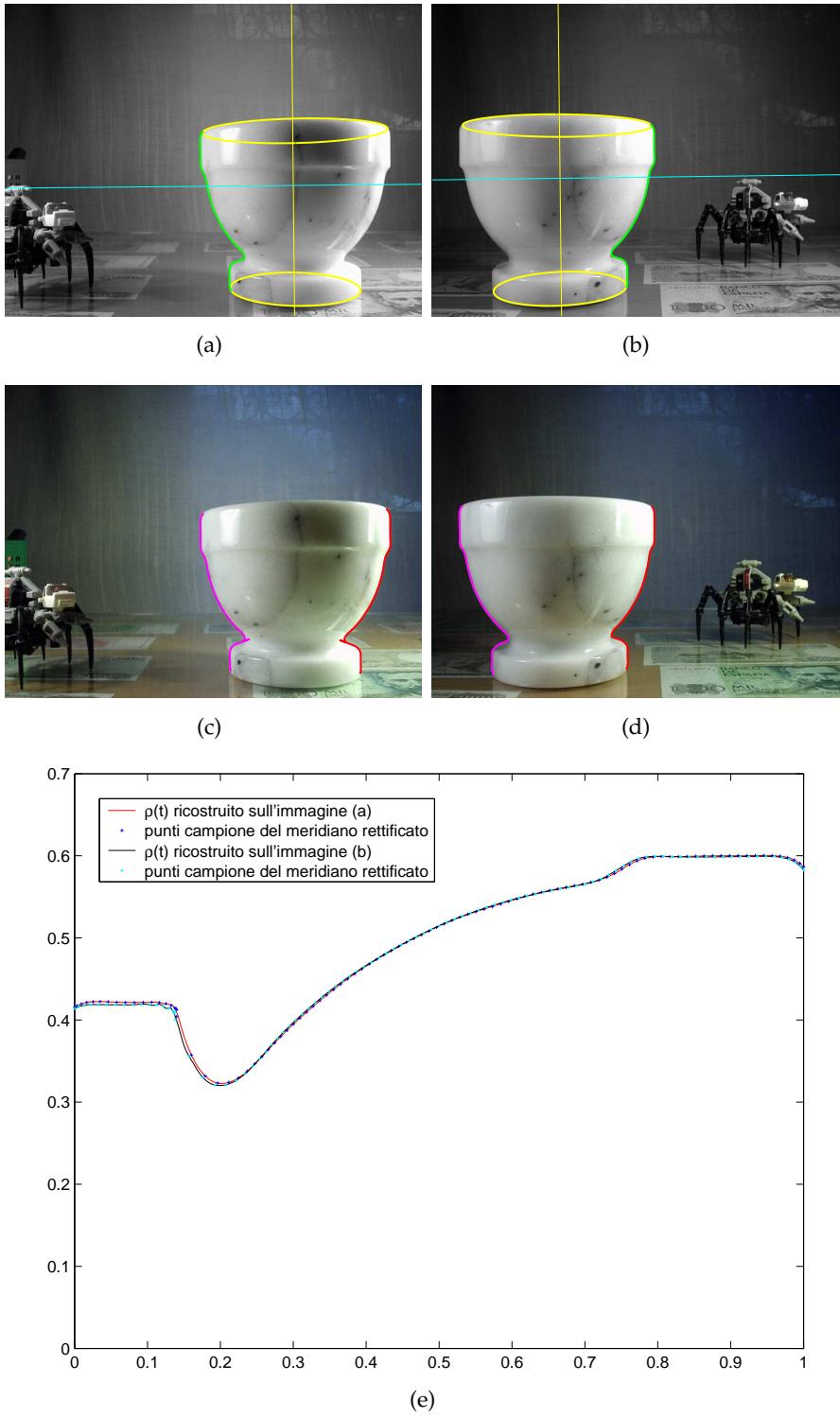


Figura 44. In alto, (a) e (b), due diverse viste dello stesso SOR. Al centro, (c) ed (d), la proiezione incrociata del profilo ricostruito da (b) su (a) e da (a) su (b). In basso, i profili ricostruiti da (a) e (b) a confronto.

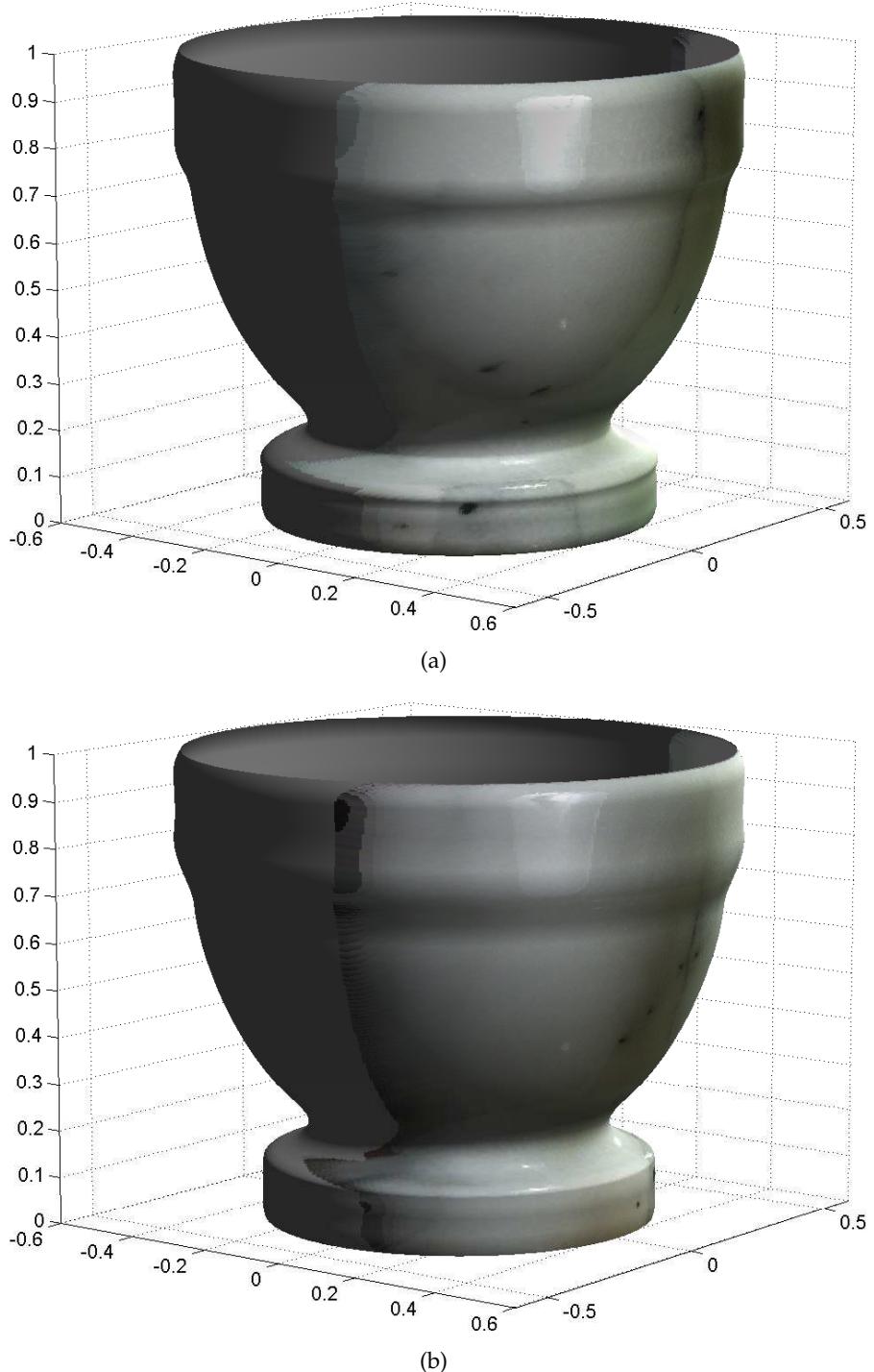


Figura 45. I modelli ottenuti da Fig. 44(a) e Fig. 44(b).



Figura 46. Rettificazione della banconota visibile in Fig. 44(d).

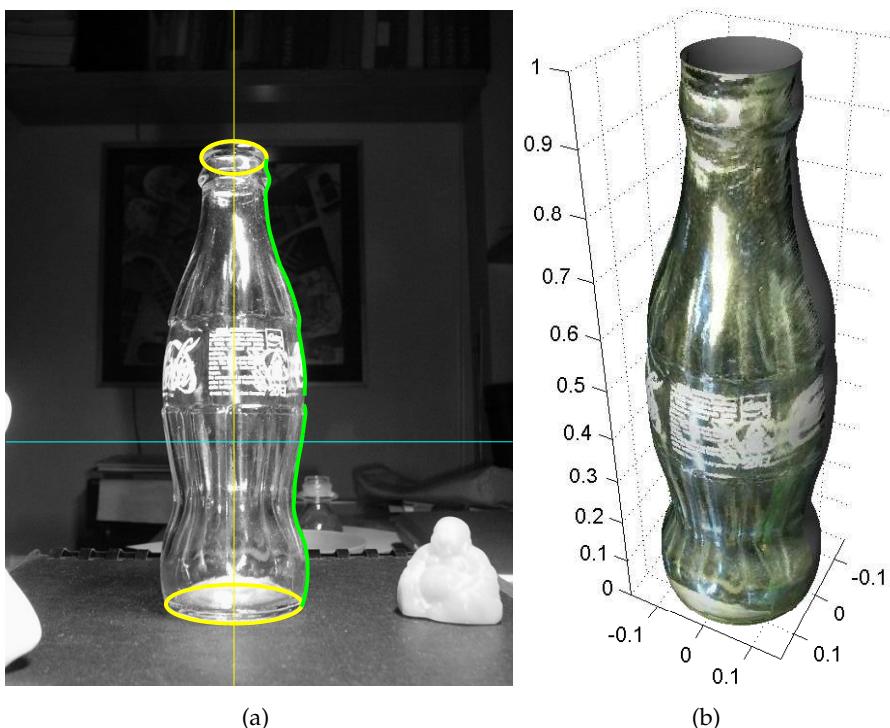
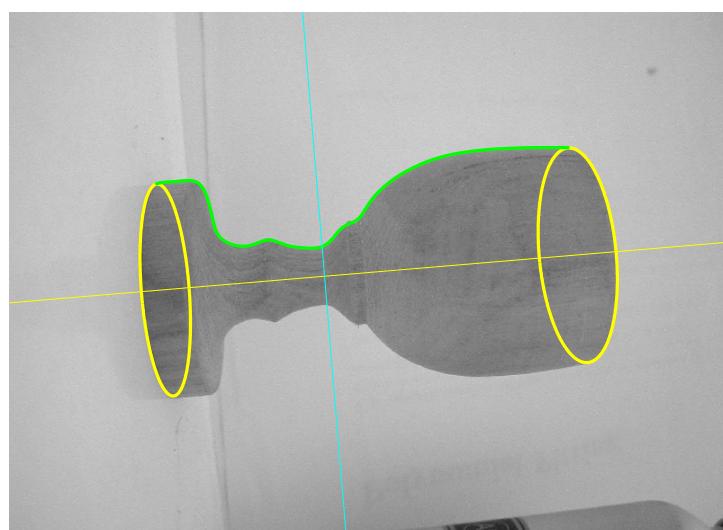
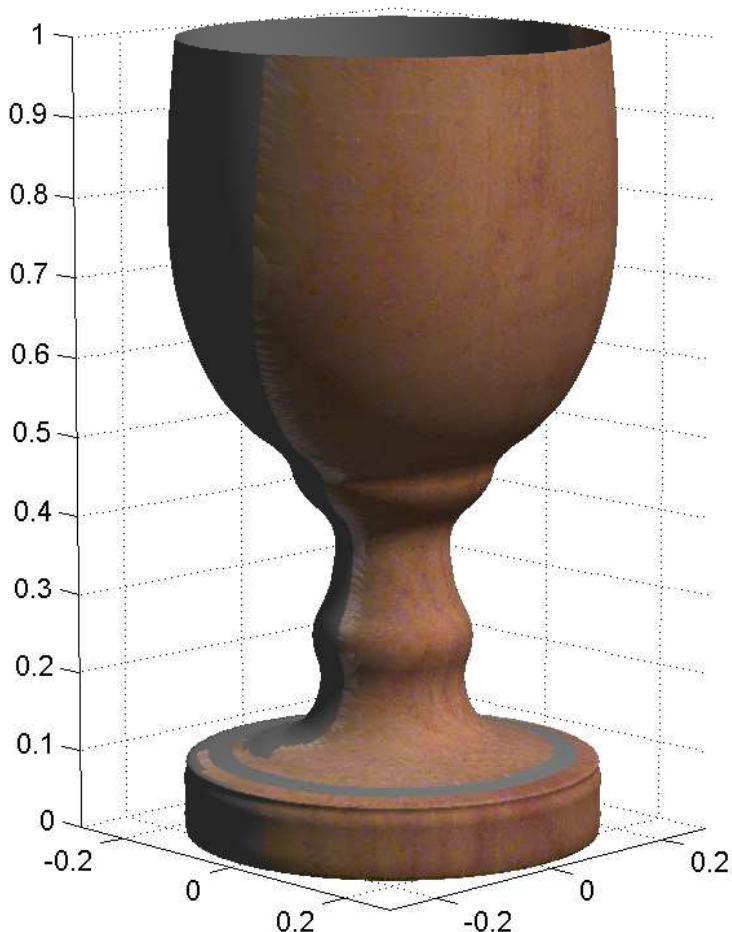


Figura 47. A sinistra, l'immagine con le curve segmentate; a destra il modello 3D. La bottiglia è di vetro, un materiale inadatto per la ricostruzione tramite scanner laser. In questo particolare caso, la bottiglia si discosta leggermente dal modello del SOR, poiché presenta delle incisioni longitudinali, che il modello 3D non può riprodurre.



(a)



(b)

Figura 48. In alto, l'immagine con le curve segmentate; in basso il modello 3D. La fascia di texture mancante nella parte superiore del piedistallo è dovuta alla difficoltà di derivare, in quella zona, la funzione del profilo.

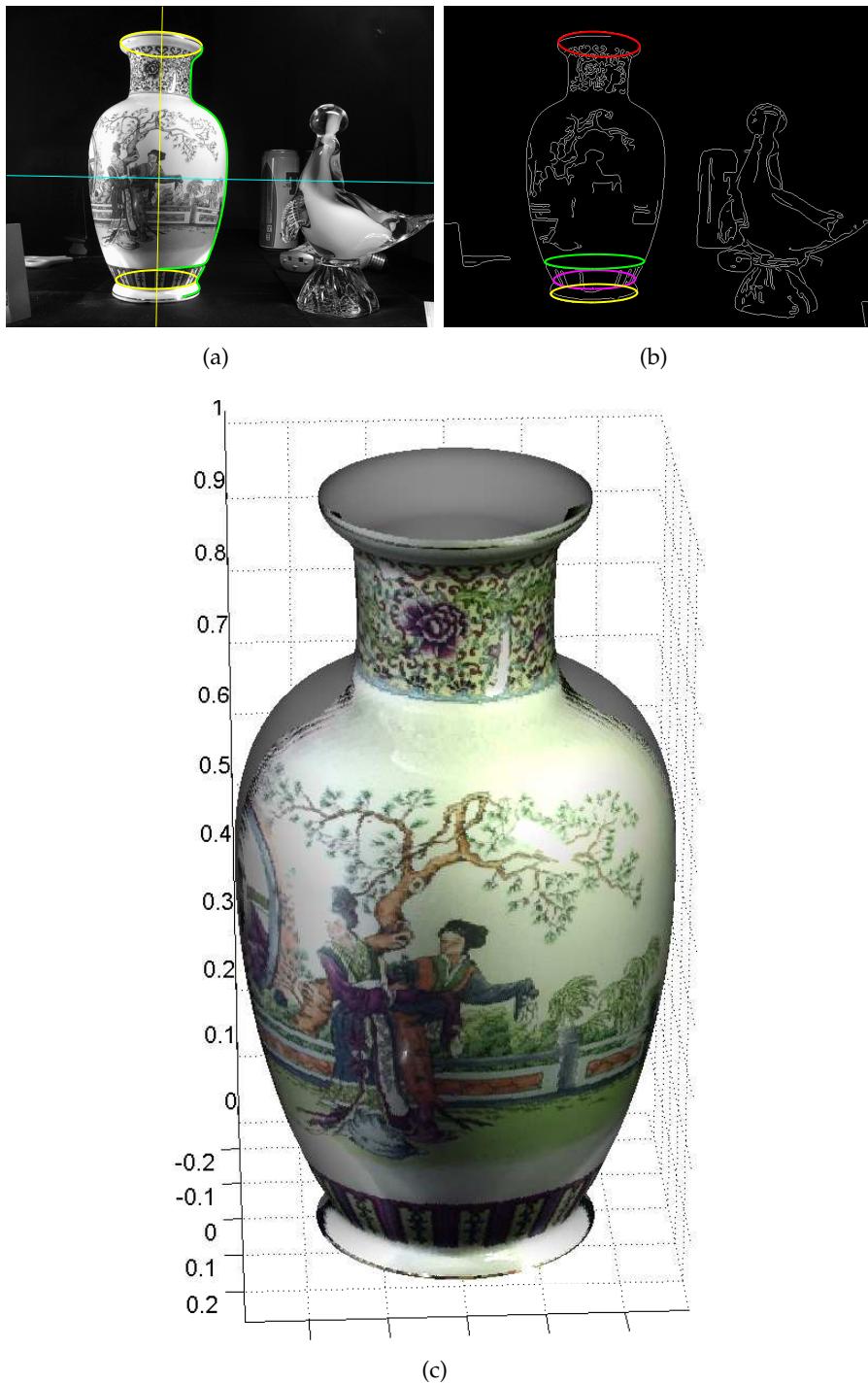


Figura 49. La selezione delle ellissi (a) ha scartato delle ellissi corrette (b), a causa di una soglia troppo bassa durante la creazione del grafo per l'estrazione della cricca massima; tuttavia, il cammino di massima lunghezza applicato sull'unione dei due tratti del contorno apparente ha permesso la rimozione del segmento ellittico non riconosciuto. Lo scarso dell'ellisse alla base del SOR ha impedito comunque la ricostruzione della prima parte del profilo.

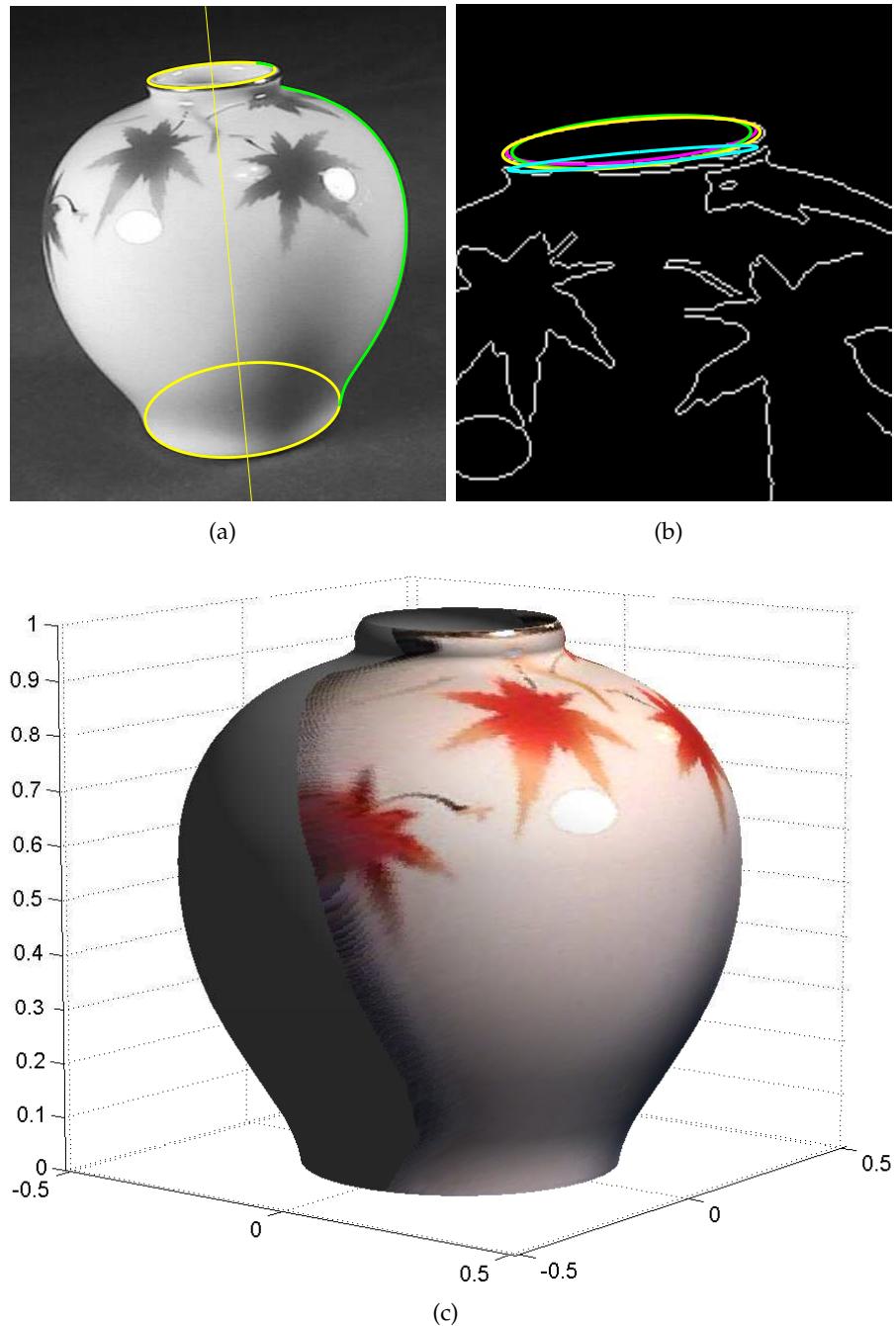


Figura 50. L'ellisse superiore in (a) è stata ottenuta dalla fusione delle ellissi cloni che erano state trovate in (b) sulla sezione superiore del vaso.

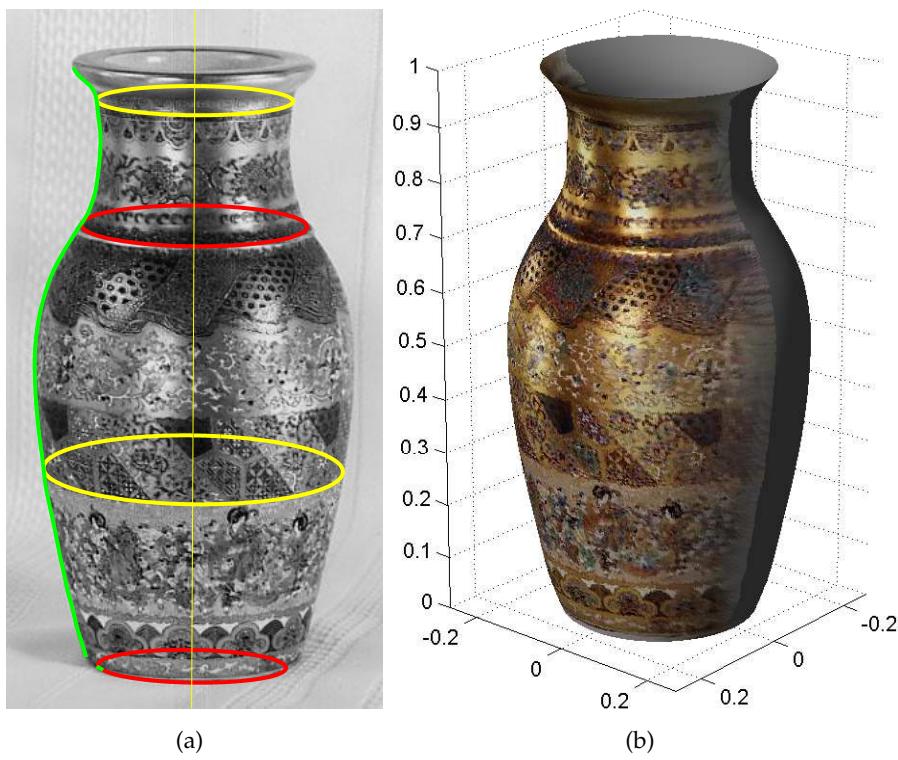


Figura 51. L'ellisse della base, pur essendo mal stimata, è stata selezionata come coerente con le altre; comunque, le ellissi usate per la calibrazione sono solo quelle in giallo: la forma ricostruita del vaso non risente dell'ellisse sbagliata.

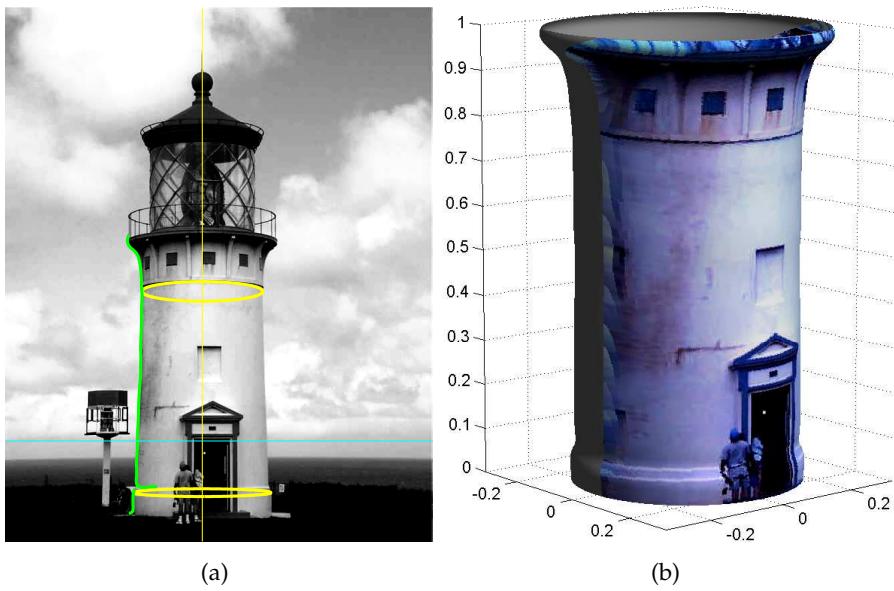


Figura 52. Segmentazione parziale del faro; la linea dell'orizzonte è più alta di quella del mare perché le ellissi trovate non sono stimate con la dovuta accuratezza.



Figura 53. La linea inferiore del computer ha agito come distrattore, mostrando a bassa risoluzione una simmetria maggiore di quella del bicchiere.

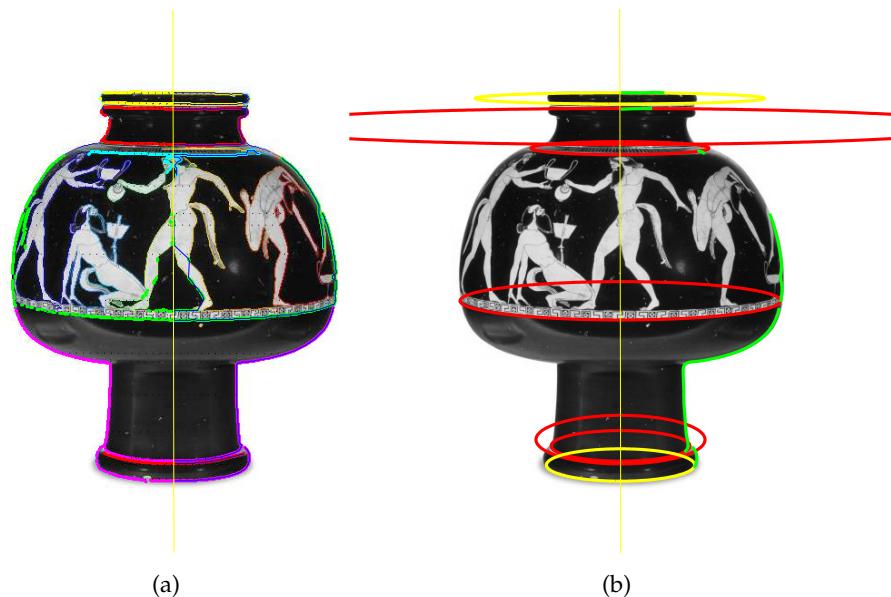


Figura 54. Segmentazione non riuscita: i punti in accordo all'omologia (a) non sono stati organizzati bene dallo MST ed inoltre le ellissi selezionate contengono degli outlier evidenti (b).

4

CONCLUSIONI

LA TESI SI È OCCUPATA dell'identificazione dei tratti che un solido di rivoluzione mostra in una sua immagine (il contorno apparente, le sezioni trasversali ed i bordi) ed ha definito degli algoritmi per la stima delle entità che individuano i vincoli geometrici necessari alla calibrazione e ricostruzione metrica del SOR.

I risultati ottenuti rappresentano un contributo sia alla ricerca sulla ricostruzione metrica di oggetti (permettendo la ricostruzione non supervisionata di SOR) sia per la stessa segmentazione (in virtù dell'alta precisione ottenibile con le procedure proposte). Tutto ciò è stato possibile grazie all'impiego mirato di varie tecniche, ciascuna delle quali ha permesso di superare le difficoltà che si sono presentate durante lo sviluppo del lavoro.

Le parti della tesi che hanno richiesto maggior attenzione sono state due: la stima della trasformazione (l'omologia armonica) che lega le due falde del contorno apparente, simultaneamente al raggruppamento in curve dei punti in accordo ad essa, e la stima delle ellissi.

L'impiego di stimatori robusti come RANSAC e LMedS nell'algoritmo di registrazione ICP, inserendo al suo interno anche un raggruppamento in curve dei punti in accordo all'omologia armonica, ha permesso la stima simultanea dei tratti candidati ad appartenere al SOR e dei parametri della trasformazione in questione, necessari per la calibrazione. L'utilizzo di una stratificazione a differenti risoluzioni dell'immagine, attraverso la quale propagare la stima dei parametri, ha inoltre contribuito all'addolcimento della funzione di costo da minimizzare rispetto all'omologia armonica.

Per quanto riguarda invece la stima delle ellissi, si è ricorsi ad una parametrizzazione che tiene conto del modello fisico dal quale esse derivano, riscuotendo risultati superiori a quelli ottenibili con stimatori generici di ellissi.

Inoltre, il lavoro è stato arricchito da un semplice procedimento per il recupero dei parametri di calibrazione esterni, fornendo come corollario un metodo per l'acquisizione della tessitura presente sulla superficie del SOR.

Gli aspetti che richiedono un ulteriore approfondimento sono tre (la figura 54 ci fornisce a questo proposito alcuni spunti). Il primo riguarda il processo di estrazione di contorni: l'edge detector di Canny lavora solo su immagini monocromatiche, trascurando completamente i colori, e possiede alcune soglie da tarare. In questo lavoro queste soglie sono state inserite a mano e lasciate costanti per tutti

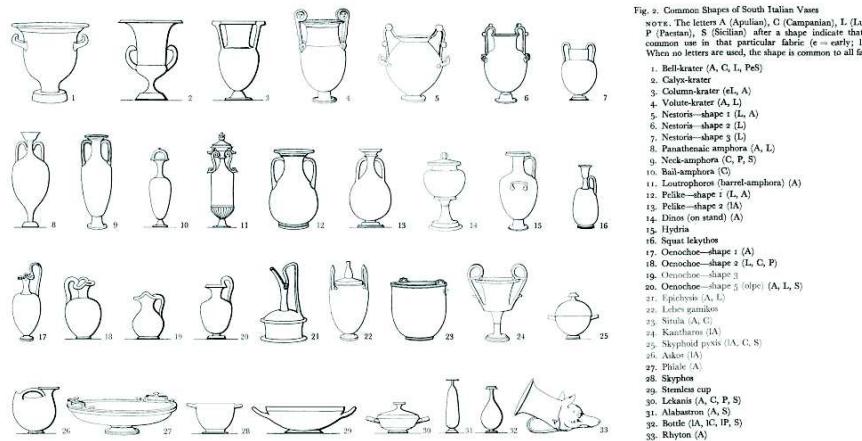


Figura 55. Classificazione di antichi vasi greci [32].

i livelli della multirisoluzione: potrebbero essere condotti studi su come tarare questi parametri, e su quanto scendere di risoluzione nell'immagine su cui lanciare RANSAC.

Il secondo riguarda l'organizzazione da imporre alla nuvola di punti che, al termine di ICP, sono in accordo all'omologia armonica trovata; l'uso del minimo albero ricoprente potrebbe essere infatti limitato all'interno di ICP, per poi usare criteri più complessi. Un esempio potrebbe essere l'impiego del colore: i punti che si trovano all'interno della sagoma del vaso presentano solitamente una tessitura differente da quelli che compongono il resto della scena. Anche la selezione delle curve più salienti potrebbe impiegare metriche che vadano oltre la lunghezza e la densità di punti, per comprendere anche un'analisi della curvatura presentata da ciascun tratto.

Per terminare, anche la selezione delle ellissi migliori necessita di ulteriori ricerche: il criterio qui adottato (un'analisi della relazione di coerenza definita sulle stime del vertice dell'omologia armonica, fornite dalle ellissi prese due a due, rispetto al vertice trovato tramite la segmentazione) è forse la parte che, più delle altre, è migliorabile.

Volgendo invece lo sguardo verso nuove frontiere, il naturale proseguimento del lavoro potrebbe coinvolgere lo studio di metriche per il riconoscimento delle forme ricostruite. Si potrebbe così creare un motore di ricerca sui database popolati da SOR: i modelli di vasi antichi potrebbero essere accessibili via web, favorendo le ricerche degli studiosi di archeologia (Fig. 55).

Un'ulteriore estensione del lavoro potrebbe riguardare l'impiego di più fotografie dello stesso solido: i modelli estratti da ognuna verrebbero fusi in un modello unico, permettendo il recupero di alcune delle parti che in una singola immagine risultano occluse.

A

LE COORDINATE OMOGENEE ED IL MODELLO “NATURAL CAMERA”

NELLA GEOMETRIA PROIETTIVA è conveniente usare un sistema di coordinate diverso da quello più familiare, cartesiano: le coordinate omogenee. Come si vedrà, attraverso esse le trasformazioni proiettive ed i modelli di fotocamera sono esprimibili in maniera lineare, tramite delle matrici.

Nelle prossime pagine verranno esposte le nozioni di base per comprendere il formalismo delle equazioni che impiegano le coordinate omogenee, terminando il capitolo con la trattazione del modello di fotocamera “natural camera”, forti delle notazioni acquisite. Per una trattazione approfondita, si rimanda a [15].

A.1 PROIEZIONI CENTRALI E COORDINATE OMOGENEE

Come già detto, il modello di fotocamera considerato è quello “natural camera”, basato sulla proiezione centrale di punti nello spazio 3D su di un piano: pur descrivendo solo il comportamento dei primi modelli di camera oscura, senza lenti, ai nostri fini costituisce una buona approssimazione delle moderne fotocamere.

In una camera oscura, il centro di proiezione (detto anche *centro di fotocamera* o *centro ottico*) è individuato dal foro praticato su di una delle sue pareti, mentre il piano su cui vengono proiettati i punti (noto come *piano immagine* o *piano focale*) è costituito dalla parete opposta al foro; la proiezione avviene grazie ai raggi di luce che, provenienti dall'esterno, passano attraverso la piccola apertura del centro ottico e colpiscono il piano focale.

Senza perdere di generalità possiamo comunque considerare il piano focale posto tra il centro di proiezione e l'esterno della fotocamera (Fig. 56): poniamo quindi nel centro di proiezione un sistema di rife-

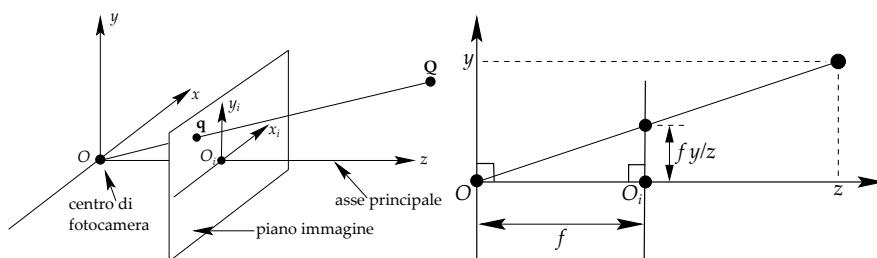


Figura 56. Schematizzazione di una proiezione centrale.

rimento euclideo $Oxyz$, considerando come piano immagine quello individuato dall’equazione $z = f$. Un punto Q nello spazio con coordinate $[x, y, z]^\top$ viene proiettato sul piano immagine dal raggio che lo congiunge al centro di proiezione O ; sfruttando la similitudine dei triangoli (Fig. 56, a destra), osserviamo che le coordinate del punto q , mappato sul piano immagine, sono $[fx/z, fy/z, f]^\top$. Trascurando la terza coordinata (quella del piano) otteniamo che tra i punti nello spazio e quelli sul piano immagine esiste la seguente relazione:

$$[x, y, z]^\top \mapsto [fx/z, fy/z]^\top \quad (34)$$

La semiretta che dal centro ottico è diretta perpendicolarmente al piano immagine è detta *asse principale*, mentre il punto individuato dalla loro intersezione prende il nome di *punto principale*; il valore assunto da f è indicato come *distanza focale*. In Eq. 34 abbiamo in pratica individuato un sistema euclideo $O_i x_i y_i$ di coordinate sul piano focale, centrato sul punto principale e con gli assi ottenuti proiettando ortogonalmente quelli x e y del sistema $Oxyz$.

Da quanto esposto, si evince che ad un punto q sul piano immagine con coordinate $[x_i, y_i]^\top$ possono corrispondere infiniti punti dello spazio 3D, tutti giacenti sul raggio che si propaga dal centro ottico e passa attraverso q ; pertanto q rappresenta una classe di equivalenza: tralasciando la distanza focale, possiamo associare in maniera biunivoca a q una classe di punti $x = \{[x, y, z]^\top : x_i = x/z, y_i = y/z\}$. Questa classe di punti equivalenti prende il nome di *vettore omogeneo* e costituisce la base di partenza per le coordinate omogenee: lo stesso principio è valido non solo per i punti in due dimensioni, ma anche nel caso di una dimensione (retta)¹ o tre (spazio 3D). In generale, un punto verrà indicato attraverso il relativo vettore omogeneo.

A.1.1 Rette di un piano in coordinate omogenee

Continuiamo la trattazione delle coordinate omogenee nel caso bidimensionale.

Per prima cosa vediamo come rappresentare una retta su di un piano: in coordinate cartesiane, l’equazione di una retta è data da

$$ax + by + c = 0; \quad (35)$$

associando ad un generico punto $[x, y]^\top$ che giace su di essa il suo vettore omogeneo $[x, y, 1]^\top$, possiamo esprimere Eq. 35 come

$$[x, y, 1] [a, b, c]^\top = 0. \quad (36)$$

¹ In particolare, nello spazio monodimensionale, viene fissato un punto di riferimento O ed un senso di percorrenza sulla retta; la classe rappresentativa di un punto su di essa è quindi data dall’vettore $\bar{w} = [x, 1]^\top$ e da tutti i suoi multipli e sottomultipli, essendo x la distanza con segno del punto da O .

Questa notazione ci induce a voler rappresentare la retta in questione attraverso il vettore $[a, b, c]^\top$; la scelta è in effetti giusta e porta a notevoli risultati, come esposto nel proseguo. Notiamo che anche in questo caso ci troviamo di fronte ad una classe di equivalenza: moltiplicare il vettore $[a, b, c]^\top$ per uno scalare $k \neq 0$ non altera il luogo dei punti espresso da Eq. 35 o da Eq. 36. Nel seguito, quando verrà fatto riferimento ad una retta, essa sarà rappresentata dal vettore omogeneo associato.

Le rappresentazioni omogenee di punti e rette sono duali, in quanto da qualsiasi enunciato che li riguarda se ne può derivare un altro scambiando i ruoli dei punti e delle rette tra di loro (*principio di dualità*). Un esempio di tale principio ci viene fornito da Eq. 36: la relazione è infatti simmetrica, per cui retta e punto possono essere invertiti nella formula. Tra l'altro tale equazione ci permette di enunciare un primo risultato:

Teorema A.1.1. Il punto x giace sulla retta r sse $x^\top r = 0$.

Altri enunciati di interesse sono i seguenti:

Teorema A.1.2. L'intersezione di due rette r ed r' è data dal punto $x = r \times r'$ (l'operatore \times denota il prodotto cartesiano tra vettori).

Teorema A.1.3. La retta passante per i punti x ed x' è data dalla formula $r = x \times x'$.

Anche in questo caso i teoremi A.1.2 e A.1.3 sono l'uno duale dell'altro. Il teorema A.1.2 ci consente di mostrare come le coordinate omogenee permettano di sostituire il concetto di rette parallele (che nella geometria euclidea non s'incontrano mai) con quello di *rette incidenti in un punto all'infinito* fornendo quindi un ampliamento del piano. Siano infatti $r_1 = [a, b, c_1]^\top$ e $r_2 = [a, b, c_2]^\top$ due rette parallele; la loro intersezione è data dal punto

$$x = [x_1, x_2, x_3]^\top = r_1 \times r_2 \equiv [b, -a, 0]^\top$$

che non è associabile a nessun punto del piano (in quanto $x_3 = 0$): si tratta del *punto all'infinito* (o *punto improprio*) con il quale sono state estese le due rette (per $x_3 \rightarrow 0$, $x_1/x_3 \rightarrow \infty$ e $x_2/x_3 \rightarrow \infty$). Tale punto rappresenta la direzione della retta considerata, dato che in coordinate cartesiane $[b, -a]^\top$ è il vettore tangente alla retta in questione. Il luogo dei punti all'infinito è costituito da una retta, $l_\infty = [0, 0, 1]^\top$ (in conformità al teorema (A.1.1), $[x_1, x_2, 0] l_\infty = 0$); essa prende il nome di *retta all'infinito*.

In generale ogni vettore $[\alpha, \beta, \gamma]^\top$, sia esso rappresentativo di una classe di equivalenza relativa ad un punto o ad una retta, costituisce un elemento dello spazio proiettivo $\mathbb{P}^2 = \mathbb{R}^3 - [0, 0, 0]^\top$ (va escluso il vettore 0 perché non corrisponde ad alcun punto o retta del piano).

A.1.2 Coniche in coordinate omogenee

In coordinate cartesiane una conica (ellisse, iperbole o parabola) nel piano è descritta da un'equazione di secondo grado della forma

$$ax^2 + 2bxy + cy^2 + 2dx + 2ey + f = 0; \quad (37)$$

utilizzando per i punti $[x, y]^\top$ la notazione omogenea $\mathbf{x} = [x_1, x_2, x_3]^\top$, dove $x \mapsto x_1/x_3$, $y \mapsto x_2/x_3$, Eq. 37 diventa

$$ax_1^2 + 2bx_1x_2 + cx_2^2 + 2dx_1x_3 + 2ex_2x_3 + fx_3^2 = 0$$

che può essere espressa nella forma matriciale

$$\mathbf{x}^\top \mathbf{C} \mathbf{x} = 0, \quad (38)$$

essendo

$$\mathbf{C} = \begin{bmatrix} a & b & d \\ b & c & e \\ d & e & f \end{bmatrix} \quad (39)$$

La matrice \mathbf{C} è definita a meno di un fattore di proporzionalità (pertanto possiede 5 gradi di libertà) e costituisce la rappresentazione omogenea di una conica.

Nel caso in cui tale matrice non abbia rango massimo, la conica è detta *degenera*, riducendosi a due rette \mathbf{l} ed \mathbf{m} nel caso di rango 2 ($\mathbf{C} = \mathbf{ml}^\top + \mathbf{lm}^\top$), o ad una retta \mathbf{l} ripetuta nell'eventualità di rango 1 ($\mathbf{C} = \mathbf{ll}^\top$).

Il principio di dualità permette di definire una conica non solo per mezzo di punti (Eq. 37 è infatti relativa ad una conica *puntuale*), ma anche tramite le rette ad essa tangenti: data una conica puntuale \mathbf{C} , si può dimostrare che una retta \mathbf{l} tangente a \mathbf{C} soddisfa $\mathbf{l}^\top \mathbf{C}^* \mathbf{l} = 0$, dove \mathbf{C}^* è la matrice aggiunta di \mathbf{C} (per una matrice simmetrica e non singolare \mathbf{C} , $\mathbf{C}^* = \mathbf{C}^{-1}$ a meno di un fattore di scala); \mathbf{C}^* rappresenta pertanto la conica *duale* di \mathbf{C} e costituisce una conica *lineare*. Una conica lineare ha anch'essa cinque gradi di libertà e quindi, come cinque punti servono per definire una conica puntuale, cinque rette sono necessarie per individuare una conica duale. Analogamente alle coniche puntuali, anche le coniche lineari possono essere degeneri: se hanno rango 1, si tratta di un punto ripetuto; se il rango è pari a 2, la conica può essere scritta come $\mathbf{C}^* = \mathbf{xy}^\top + \mathbf{yx}^\top$ e consiste di tutte le rette passanti per i punti \mathbf{x} o \mathbf{y} .

Attraverso una conica è possibile definire anche una particolare relazione tra punti e rette: un punto \mathbf{x} ed una retta \mathbf{r} si dicono in *relazione polo-polare* se $\mathbf{r} = \mathbf{Cx}$; la retta \mathbf{r} prende il nome di *polare* del punto \mathbf{x} rispetto alla conica \mathbf{C} , mentre \mathbf{x} è il *polo* di \mathbf{r} rispetto a \mathbf{C} . Il significato geometrico di tale legame è il seguente: la polare di \mathbf{x} interseca la conica \mathbf{C} in due punti; le tangenti a \mathbf{C} in questi due punti si incontrano in \mathbf{x} . Come conseguenza di ciò abbiamo che se \mathbf{x} appartiene alla conica, $\mathbf{r} = \mathbf{Cx}$ è la retta tangente alla conica in \mathbf{x} .

A.1.3 Trasformazioni proiettive

Introduciamo adesso il concetto di trasformazioni proiettive nel piano e di come vengono trasformate attraverso di esse rette e coniche.

Una *trasformazione proiettiva* od *omografia* su di un piano è una funzione invertibile $h : \mathbb{P}^2 \rightarrow \mathbb{P}^2$ che trasforma rette in rette: cioè, se x_1, x_2 e x_3 appartengono ad una stessa retta, allora anche $h(x_1), h(x_2), h(x_3)$ giacciono su di una retta comune. La definizione data astrae dal sistema di coordinate; tuttavia esiste un teorema il quale ci assicura che, utilizzando le coordinate omogenee, l'omografia è esprimibile attraverso una matrice 3×3 , non singolare: indicando con H tale matrice, abbiamo pertanto che $x' = h(x)$ è scrivibile come $x' = Hx$.

Se una trasformazione proiettiva H viene applicata ai punti di un piano, indicando con $H^{-\top}$ la matrice inversa e trasposta di H , le rette si trasformano secondo la relazione

$$\mathbf{r}' = H^{-\top} \mathbf{r} \quad (40)$$

mentre le coniche ubbidiscono a

$$C' = H^{-\top} CH^{-1} \quad (41)$$

nel caso di coniche puntuali oppure

$$C^{*\prime} = HC^*H^\top \quad (42)$$

se si tratta di coniche lineari.

In generale, l'applicazione di una trasformazione proiettiva sui punti del piano comporta che i punti appartenenti a \mathbf{l}_∞ diventino propri e che rette ortogonali tra loro non restino tali.

A.1.4 Punti circolari

L'equazione di una circonferenza è un caso particolare di Eq. 38, in cui $a = c$ e $b = 0$ (senza perdere di generalità, possiamo porre anche $a = 1$); imponendo l'intersezione con \mathbf{l}_∞ ($x_3 = 0$), otteniamo

$$x_1^2 + x_2^2 = 0,$$

la cui soluzione è data dai punti $\mathbf{I} = [1, 1, 0]^\top$ e $\mathbf{J} = [1, -1, 0]^\top$, che prendono il nome di punti circolari. Una caratteristica importante di questi punti è che sono fissi rispetto ad un solo tipo di trasformazioni proiettive, le *similarità*, che preservano la forma delle curve; se viene applicata una generica trasformazione proiettiva a delle curve su di un piano, è possibile ricostruire la loro forma originaria (a meno di un fattore di scala), una volta identificati i punti circolari ed averli riportati nella loro posizione canonica.

La conica

$$C_\infty^* = \mathbf{IJ}^\top + \mathbf{JI}^\top \quad (43)$$

è duale ai punti circolari; si tratta di una conica lineare di rango 2 che è appunto formata da tutte le rette per i punti circolari e, in un sistema euclideo, assume la forma

$$C_{\infty}^* = \begin{bmatrix} 1 \\ \imath \\ 0 \end{bmatrix} [1 \ \imath \ 0] + \begin{bmatrix} 1 \\ -\imath \\ 0 \end{bmatrix} [1 \ -\imath \ 0] = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} \quad (44)$$

La conica C_{∞}^* permette di ricavare il vero angolo θ formato da due rette dopo che esse sono state sottoposte ad una trasformazione proiettiva; si può infatti dimostrare che

Teorema A.1.4. la seguente espressione è invariante sotto qualsiasi trasformazione proiettiva H :

$$\cos(\theta) = \frac{\mathbf{l}^T C_{\infty}^* \mathbf{m}}{\sqrt{(\mathbf{l}^T C_{\infty}^* \mathbf{l})(\mathbf{m}^T C_{\infty}^* \mathbf{m})}}. \quad (45)$$

Nel caso che C_{∞}^* abbia la forma riportata in Eq. 44, è facile ricavare l'usuale formula per il coseno dell'angolo formato da due rette nel piano cartesiano.

A.2 IL MODELLO "NATURAL CAMERA"

La formula di Eq. 34, che descrive la proiezione centrale dei punti $\mathbf{Q} = [x, y, z]^T$ nello spazio su quelli $\mathbf{q} = [x_i, y_i]^T$ del piano immagine, può essere scritta tramite una matrice se per i punti \mathbf{Q} e \mathbf{q} si utilizzano le coordinate omogenee, con i rispettivi vettori \mathbf{X} e \mathbf{x} (il vettore omogeneo \mathbf{X} associato a \mathbf{Q} è $\mathbf{X} = [x_1, x_2, x_3, x_4]^T$, con $x_1/x_4 = x$, $x_2/x_4 = y$ e $x_3/x_4 = z$)²:

$$\begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix} \longmapsto \begin{bmatrix} fx \\ fy \\ z \end{bmatrix} = \begin{bmatrix} f & & 0 \\ & f & 0 \\ & & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix}, \quad (46)$$

ossia in maniera più compatta

$$\mathbf{x} = \mathbf{P}\mathbf{X}, \quad (47)$$

dove $\mathbf{P} = \text{diag}(f, f, 1)[I_{3 \times 3} \mid 0]$.

La formula di Eq. 34 era stata ricavata centrando nel punto principale il sistema di riferimento del piano immagine; in generale questo non è detto, per cui assumeremo che il punto principale abbia coordinate $[p_x, p_y]^T$: mantenendo comunque inalterate le direzioni degli assi sul piano focale, la relazione che adesso lega le coordinate di quest'ultimo a quelle del sistema Oxyz è

$$\mathbf{x} = \mathbf{K}[I_{3 \times 3} \mid 0]\mathbf{X}_{\text{cam}}, \quad (48)$$

² Anche se al momento l'utilizzo del vettore omogeneo per \mathbf{Q} è una forzatura, il motivo del suo impiego sarà evidente alla fine.

avendo posto

$$K = \begin{bmatrix} f & p_x \\ f & p_y \\ 1 & \end{bmatrix}. \quad (49)$$

La notazione \mathbf{X}_{cam} serve ad enfatizzare che le coordinate del punto nello spazio sono prese rispetto al sistema di fotocamera Oxyz.

A questo punto, possiamo fare un'ulteriore generalizzazione, in quanto le coordinate dei punti nello spazio possono non essere prese rispetto al sistema di fotocamera, bensì avendo fissato un sistema di coordinate *nel mondo* $O'x'y'z'$. I due riferimenti sono legati da una traslazione che porta O' in O e da rotazioni per sovrapporre gli assi: siano quindi $\mathbf{Q}' = [x', y', z']^\top$ e $\mathbf{Q}_{\text{cam}} = [x, y, z]^\top$ le coordinate cartesiane di uno stesso punto, rispettivamente secondo $O'x'y'z'$ e Oxyz; quanto detto è sintetizzato dalla formula $\mathbf{Q}_{\text{cam}} = R(\mathbf{Q}' - \mathbf{c})$, dove R è la matrice responsabile della rotazione degli assi e \mathbf{c} rappresenta le coordinate del centro di fotocamera espresse nel sistema del mondo³. Passando alle coordinate omogenee, (indichiamo pertanto con \mathbf{X}' il vettore omogeneo dei punti espressi secondo il riferimento $O'x'y'z'$), abbiamo

$$\mathbf{X}_{\text{cam}} = \left[\begin{array}{ccc|c} R & & & -R\mathbf{c} \\ 0 & 0 & 0 & 1 \end{array} \right] \mathbf{X}'; \quad (50)$$

la composizione di Eq. 50 con Eq. 48 porta alla relazione

$$\mathbf{x} = KR[\mathbf{I}_{3 \times 3} \mid -\mathbf{c}] \mathbf{X}'. \quad (51)$$

Questo è il legame del modello “natural camera” tra le coordinate nel mondo e quelle dell’immagine; la matrice che lo descrive è

$$P = KR[\mathbf{I}_{3 \times 3} \mid -\mathbf{c}] \quad (52)$$

e possiede 9 gradi di libertà: 3 contenuti in K (f, p_x, p_y , i cosiddetti *parametri interni di fotocamera*), 3 in R (le rotazioni dei tre assi) e 3 in \mathbf{c} (le coordinate del centro di fotocamera); questi restanti 6 costituiscono invece i *parametri esterni*.

A.3 CENNI ALLA CALIBRAZIONE

Ogni punto \mathbf{x} in un’immagine individua un raggio nello spazio passante per esso e per il centro di proiezione. La direzione di questo raggio è esprimibile (nel sistema di riferimento solidale con la fotocamera) grazie alla matrice K , in virtù di come è stata introdotta. Più precisamente, vale il seguente

³ Mediante i *coseni direttori* si può dimostrare che le righe di R sono costituite dai versori delle direzioni del sistema di fotocamera espressi nel sistema nel mondo; quindi possiamo assumere R come matrice ortogonale.

Teorema A.3.1. il vettore *non omogeneo* \mathbf{d} che, rispetto al sistema solidale con la fotocamera, indica nello spazio la direzione del raggio passante per il centro di proiezione ed un generico punto \mathbf{x} sul piano immagine è dato da:

$$\mathbf{d} = \mathbf{K}^{-1} \mathbf{x}. \quad (53)$$

Come corollario a questo risultato, se \mathbf{x}_1 e \mathbf{x}_2 sono due punti nell'immagine, possiamo ricavare tramite \mathbf{K} l'angolo formato dai due raggi \mathbf{d}_1 e \mathbf{d}_2 passanti per essi ed il centro di proiezione; in virtù della familiare formula per il coseno dell'angolo ϑ compreso da due vettori, abbiamo:

$$\cos(\vartheta) = \frac{\mathbf{d}_1^\top \mathbf{d}_2}{\sqrt{\mathbf{d}_1^\top \mathbf{d}_1} \sqrt{\mathbf{d}_2^\top \mathbf{d}_2}} = \frac{\mathbf{x}_1^\top \boldsymbol{\omega} \mathbf{x}_2}{\sqrt{\mathbf{x}_1^\top \boldsymbol{\omega} \mathbf{x}_1} \sqrt{\mathbf{x}_2^\top \boldsymbol{\omega} \mathbf{x}_2}}, \quad (54)$$

avendo posto $\boldsymbol{\omega} = (\mathbf{K} \mathbf{K}^\top)^{-1}$.

La matrice $\boldsymbol{\omega}$ ha un significato teorico estremamente importante: si può dimostrare infatti che essa rappresenta la proiezione della *conica assoluta* Ω_∞ nell'immagine; per questa ragione viene solitamente indicata con la sigla IAC. La *conica assoluta* sta allo spazio, come i punti circolari stanno al piano: la sua principale caratteristica è infatti quella di essere invariante a trasformazioni di similarità. Inoltre, l'intersezione di un piano con Ω_∞ individua i punti circolari dei cerchi che giacciono su quel piano. Si può dimostrare che la IAC è essa stessa una conica immaginaria (planare); inoltre un punto \mathbf{x} ed una retta \mathbf{l} nel piano immagine, in relazione polo-polare con la IAC, si proiettano nello spazio rispettivamente in una retta ed un piano ortogonali tra loro e passanti entrambi per il centro di proiezione della fotocamera.

In definitiva, una fotocamera per la quale è nota la matrice \mathbf{K} costituisce un *sensore di direzione*, capace di misurare le direzioni dei raggi uscenti da essa. Una fotocamera per cui la matrice \mathbf{K} è nota viene detta *calibrata*; \mathbf{K} prende il nome di *matrice di calibrazione*.

BIBLIOGRAFIA

- [1] S. M. Abdallah. *Object Recognition via Invariance*. Tesi di Dottorato di Ricerca, The University of Sydney, Sidney, Australia, 2000.
- [2] N. Bennett, R. Burridge, e N. Saito. A method to detect and characterize ellipses using the hough transform. *IEEE Transaction On Pattern Analysis and Machine Intelligence*, 21(7):652–657, 1999.
- [3] P. J. Burt e E. H. Adelson. The laplacian pyramid as a compact image code. *IEEE Transactions on Communications*, 31(4):532–540, 1983.
- [4] J. Canny. A computational approach to edge detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 8(6):679–698, 1986.
- [5] C. Colombo, A. Del Bimbo, e F. Pernici. Uncalibrated 3D metric reconstruction and flattened texture acquisition from a single view of a surface of revolution. In *First International Symposium on 3D Data Processing Visualization and Transmission*, 2002.
- [6] C. Colombo, A. Del Bimbo, e F. Pernici. Metric 3d reconstruction and texture acquisition of surfaces of revolution from a single uncalibrated view. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, submitted 2003.
- [7] L. H. de Figueiredo. *Computational Morphology of Implicit Curves*. Tesi di Dottorato di Ricerca, IMPA - Instituto de Matematica Pura e Aplicata, Rio de Janeiro, Brazil, 1992.
- [8] L. H. de Figueiredo e J. Gomes. Computational morphology of curves. *The Visual Computer*, 11(2):105–112, 1995.
- [9] A. W. Fitzgibbon. Robust registration of 2D and 3D point sets. In *British Machine Vision Conference*, pp. 662–670, 2001.
- [10] A. W. Fitzgibbon, G. Cross, e A. Zisserman. Automatic 3D model construction for turn-table sequences. In *3D Structure from Multiple Images of Large-Scale Environments*, 1998.
- [11] A. W. Fitzgibbon, M. Pilu, e R. B. Fisher. Direct least square fitting of ellipses. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 21(5):476–480, 1999.
- [12] D. A. Forsyth, J. L. Mundy, A. Zisserman, e C. A. Rothwell. Recognising rotationally symmetric surfaces from their outlines. In *European Conference on Computer Vision*, 1992.

- [13] D. A. Forsyth e J. Ponce. *Computer Vision: A Modern Approach*. Prentice Hall, 2002.
- [14] A. Goldenshluger e A. Zeevi. The Hough transform estimator. *The Annals of Statistics*, 32(5):1908–1932, 2004.
- [15] R. I. Hartley e A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, 2000.
- [16] H. Hoppe. *Surface reconstruction from unorganized points*. Tesi di Dottorato di Ricerca, Department of Computer Science and Engineering, University of Washington, Seattle, WA, USA, 1994.
- [17] G. Jiang, H. T. Tsui, L. Quan, e A. Zisserman. Single axis geometry by fitting conics. In *European Conference on Computer Vision*, 2002.
- [18] G. Kanizsa. *Organization in Vision*. Praeger, 1979.
- [19] I.-K. Lee. Curve reconstruction from unorganized points. *Computer Aided Geometric Design*, 17(2):161–177, 2000.
- [20] D. Liebowitz, A. Criminisi, e A. Zisserman. Creating architectural models from images. In *Annual Conference of the European Association for Computer Graphics (Eurographics)*, volume 18, pp. 39–50, 1999.
- [21] D.G. Lowe. *Perceptual Organization and Visual Recognition*. Kluwer, 1985.
- [22] P. Meer, C.V. Sewart, e D.E. Tyler. Robust computer vision: An interdisciplinary challenge. *Computer Vision and Image Understanding*, 78(1):1–7, 2000.
- [23] P.R.S. Mendonça, K.-Y.K. Wong, e R. Cipolla. Epipolar geometry from profiles under circular motion. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(6):604–616, 2001.
- [24] J. L. Mundy e A. Zisserman. Repeated structures: Image correspondence constraints and ambiguity of 3D reconstruction In *Applications of Invariance in Computer Vision*. A cura di J. L. Mundy, A. Zisserman, e D. Forsyth, pp. 89–106. Springer-Verlag, 1994.
- [25] N.H.E. Pillow. *Recognition of Generalized Cylinders Using Geometric Invariance*. Tesi di Dottorato di Ricerca, Oriel College, University of Oxford, Oxford, UK, 1996.
- [26] F.P. Preparata e M.I. Shamos. *Computational Geometry: An Introduction*. Springer-Verlag, 1985.
- [27] I. Rock. *The Logic of Perception*. MIT Press, 1983.

- [28] I. Rock e S. Palmer. The legacy of gestalt psychology. *Scientific American*, 263(6):84–91, 1990.
- [29] P. J. Rousseeuw e A. M. Leroy. *Robust Regression and Outlier Detection*. John Wiley and Sons, 1987.
- [30] S. Sarkar e K. L. Boyer. Perceptual organization in computer vision: A review and a proposal for classificatory structure. *IEEE Transaction on systems, man and cybernetics*, 23(2):382–399, 1993.
- [31] B. Shahraray e D. J. Anderson. Optimal estimation of contour properties by cross-validated regularization. *IEEE Transaction On Pattern Analysis and Machine Intelligence*, 2(6):600–610, 1989.
- [32] A. D. Trendall. *South Italian vase-painting*. The British Museum, 1966.
- [33] A. Witkin e J. Tenenbaum. On the role of structure in vision. In *Human and Machine Vision*, 1983.
- [34] K.-Y. K Wong, P. R. S. Mendonça, e R. Cipolla. Camera calibration from surfaces of revolution. *IEEE Transaction On Pattern Analysis and Machine Intelligence*, 25(2):147–161, 2003.
- [35] K.-Y. K. Wong, P. R. sS. Mendonça, e R. Cipolla. Reconstruction of surfaces of revolution from single uncalibrated views. In *British Machine Vision Conference*, 2002.
- [36] M. Zerroug e R. Nevatia. Volumetric descriptions from a single intensity image. *International Journal of Computer Vision*, 20(1):11–42, 1996.
- [37] Z. Zhang. Iterative point matching for registration of free-form curves. Relazione Tecnica RR-1658, INRIA, 1992.
- [38] Z. Zhang. Parameter estimation techniques: A tutorial with application to conic fitting. Relazione Tecnica RR-2676, INRIA, 1995.