



Università degli Studi di Parma
Facoltà di Ingegneria

Dottorato di Ricerca in Ingegneria Civile - XVIII Ciclo
Curriculum: Topografia e Cartografia (ICAR/06)

Riccardo Roncella

Sviluppo e Applicazioni di Tecniche di Automazione in Fotogrammetria dei Vicini



Dissertazione per il conseguimento del titolo di Dottore di Ricerca

Tutore: Prof. Ing. Gianfranco Forlani
Coordinatore del Dottorato: Prof. Ing. Paolo Mignosa

Parma, Gennaio 2006

Università degli Studi di Parma
Facoltà di Ingegneria

Dottorato di Ricerca in Ingegneria Civile
Curriculum: Topografia e Cartografia (ICAR/06)
(XVIII Ciclo)

Riccardo Roncella

Sviluppo e Applicazioni di
Tecniche di Automazione in
Fotogrammetria dei Vicini

Dissertazione per il conseguimento del titolo di Dottore di Ricerca

Tutore: Prof. Ing. Gianfranco Forlani
Coordinatore del Dottorato: Prof. Ing. Paolo Mignosa

Parma, Gennaio 2006

Marco Polo descrive un ponte, pietra per pietra.

-Ma qual è la pietra che sostiene il ponte?- chiede Kublai Kan.

-Il ponte non è sostenuto da questa o quella pietra,- risponde Marco -ma dalla linea dell'arco che esse formano.

Kublai Kan rimane silenzioso, riflettendo. Poi soggiunge: -Perchè mi parli delle pietre? È solo dell'arco che m'importa.

Polo risponde: -Senza pietre non c'è arco.

Italo Calvino, *Le Città Invisibili*.

Tutti i libri del mondo contengono una quantità di informazione non superiore a quella trasmessa come segnale video in una singola grande città Americana in un solo anno.

Non tutti i bit hanno la stessa importanza.

Carl Sagan

Indice

INTRODUZIONE

Presentazione degli obiettivi e sintesi del lavoro.....	1
Convenzioni e simboli.....	4

CAPITOLO 1

Structure and Motion	7
1.1. Introduzione.....	7
1.2. Operatori d'interesse.....	11
1.2.1. Operatore di Harris.....	12
1.2.2. Operatore di Föerstner.....	15
1.2.3. Operatore di Lowe (SIFT).....	16
1.2.4. Altri operatori di interesse.....	18
1.2.5. Analisi comparativa degli operatori di interesse.....	19
1.3. Feature Matching.....	24
1.3.1. Matching procustiano.....	25
1.3.2. Descrittore di Lowe.....	29
1.3.3. Ottimizzazione dell'estrazione di features.....	32
1.4. Collinearità e geometria proiettiva.....	34
1.5. Geometria epipolare.....	39
1.5.1. Matrice Fondamentale.....	39
1.5.2. Metodi di stima della matrice fondamentale.....	41
1.5.3. Stima robusta della matrice fondamentale.....	43
1.6. Geometria trifocale.....	47
1.6.1. Tensore Trifocale.....	47
1.6.2. Metodi di stima del tensore trifocale.....	51
1.6.3. Stima robusta della geometria trifocale.....	56
1.7. Ricostruzione metrica e bundle adjustment.....	58
1.8. Simulazioni, Test e Applicazioni.....	64

1.8.1. Sequenze generate artificialmente.....	64
1.8.2. S&M applicata al rilievo di pareti rocciose.....	67
1.8.3. S&M applicata al rilievo architettonico.....	69
1.8.4. S&M applicata alla navigazione di veicoli per il mobile mapping.....	71

CAPITOLO 2

Algoritmi e tecniche di matching area-based.....	79
2.1. Metodi di correlazione semplice.....	81
2.2. Least Squares Matching.....	84
2.3. Multi-Photo Geometrically Constrained Matching.....	87
2.4. Pre-elaborazione delle immagini.....	90
2.4.1. Eliminazione della distorsione.....	91
2.4.2. Ottimizzazione del contenuto informativo.....	95
2.4.3. Rettifica di immagini epipolari.....	98
2.5. Metodi di matching denso.....	104
2.5.1. Metodi paralleli.....	104
2.5.2. Metodi sequenziali.....	106
2.5.3. Metodi di rilassamento.....	109
2.6. Fotogrammetria o Laser Scanner?.....	109
2.7. Simulazioni, Test e Applicazioni.....	112
2.7.1. Precisioni di misura del least squares matching.....	112
2.7.2. Least squares matching e MPGCM.....	118
2.7.3. Generazione di nuvole di punti.....	120

CAPITOLO 3

Analisi di Point Clouds.....	125
3.1. Analisi interattiva di nuvole di punti.....	126
3.2. Segmentazione automatica di piani.....	129
3.3. Segmentazione con metodi di stima robusti.....	130
3.3.1. Generazione del DSM multi-risoluzione.....	133
3.3.2. Estrazione delle entità planari.....	135
3.3.3. Clustering gerarchico delle giaciture.....	136
3.3.4. Suddivisione della geometria.....	138
3.4. Segmentazione per mezzo della trasformata di Hough.....	139

CAPITOLO 4

Determinazione dei piani di discontinuità di pareti rocciose.....	143
4.1. Lo studio delle discontinuità negli ammassi rocciosi.....	143
4.2. Un sistema fotogrammetrico per il controllo di stabilità dei versanti in roccia.....	147

4.2.1. Progetto e ripresa delle immagini.....	148
4.2.2. Il software di elaborazione ed il flusso di informazioni.....	150
4.2.3. Versanti in roccia: Fotogrammetria o Laser Scanner?.....	150
4.3. Sito campione: Arnad.....	151
4.4. Sito campione: Longeborgne.....	158
4.5. Sito campione: Granier.....	162
Conclusioni	169
Appendice A	173
A.1. Cenni di geometria proiettiva.....	173
A.1.1. Spazi vettoriali proiettivi.....	174
A.1.2. Struttura degli spazi proiettivi.....	177
A.1.3. Coniche nello spazio proiettivo.....	179
A.1.4. Applicazioni lineari e proiettività.....	181
A.2. Metodi di stima.....	184
A.2.1. Sistemi di equazioni e decomposizione ai valori singolari.....	184
A.2.2. Il metodo di Levenberg-Marquadt.....	188
A.2.3. Line search.....	188
A.2.4. Metodi di stima robusti.....	190
A.3. Metodi di resezione.....	194
A.4. Logica fuzzy.....	198
Appendice B	201
B.1. Matrice camera.....	201
B.2. Matrice fondamentale.....	202
B.3. Tensore trifocale.....	203
Bibliografia	205
Ringraziamenti	215

Indice delle Figure

CAPITOLO 1

<i>Figura 1.1:</i> Schema a blocchi delle operazioni implementate nella Structure and Motion.....	8
<i>Figura 1.2:</i> Rappresentazione della funzione di risposta R dell'operatore di Harris.....	14
<i>Figura 1.3:</i> Immagini utilizzate per i test comparativi degli operatori di interesse.....	20
<i>Figura 1.4:</i> Test comparativi: grado di ripetibilità medio (rotazione).....	21
<i>Figura 1.5:</i> Test comparativi: grado di ripetibilità medio (variazione di scala).....	22
<i>Figura 1.6:</i> Test comparativi: grado di ripetibilità medio (trasformazione affine).....	22
<i>Figura 1.7:</i> Test comparativi: grado di ripetibilità medio (variazione del punto di vista).....	23
<i>Figura 1.8:</i> Test comparativi: grado di ripetibilità medio (variazione di luminosità).....	23
<i>Figura 1.9:</i> Finestra di ricerca nel feature matching.....	26
<i>Figura 1.10:</i> Rappresentazione di un campo di parallassi.....	26
<i>Figura 1.11:</i> Mappa delle orientazioni ed intensità (operatore di Lowe).....	29
<i>Figura 1.12:</i> Istogrammi del descrittore di Lowe.....	30
<i>Figura 1.13:</i> Ottimizzazione dell'estrazione di features.....	32
<i>Figura 1.14:</i> Localizzazione dei punti di interesse.....	33
<i>Figura 1.15:</i> Collinearità e geometria proiettiva.....	34
<i>Figura 1.16:</i> Geometria epipolare fra due fotogrammi.....	39
<i>Figura 1.17:</i> Stima della matrice fondamentale note almeno 8 corrispondenze.....	42
<i>Figura 1.18:</i> Rappresentazione della stima dell'errore algebrico e geometrico.....	43
<i>Figura 1.19:</i> Distribuzione degli errori (distanza dal modello funzionale) in un dataset.....	44
<i>Figura 1.20:</i> Stima robusta della geometria epipolare.....	46
<i>Figura 1.21:</i> Stima del tensore trifocale note 6 corrispondenze triple.....	55
<i>Figura 1.22:</i> Stima robusta della geometria trifocale.....	57
<i>Figura 1.23:</i> Ambiguità proiettiva nella ricostruzione.....	58
<i>Figura 1.24:</i> Configurazioni ottenibili dalla matrice essenziale.....	62
<i>Figura 1.25:</i> Ricostruzione metrica della structure and motion.....	63
<i>Figura 1.26:</i> Immagine artificiale per la verifica della S&M.....	64
<i>Figura 1.27:</i> Posizioni di presa e punti restituiti della sequenza 1.26 (90 immagini).....	65
<i>Figura 1.28:</i> Posizioni di presa e punti restituiti della sequenza 1.26 (18 immagini).....	65
<i>Figura 1.29:</i> Arnad: una delle pareti della Corma di Machaby.....	67
<i>Figura 1.30:</i> Posizioni di presa e punti restituiti della sequenza realizzata ad Arnad.....	67
<i>Figura 1.31:</i> Parete che sovrasta l'eremo di Longebogne.....	68
<i>Figura 1.32:</i> Posizioni di presa e punti restituiti della sequenza realizzata a Longeborgne.....	68
<i>Figura 1.33:</i> Il primo dei gruppi scultorei del Pantheon analizzati.....	69
<i>Figura 1.34:</i> Posizioni di presa e punti restituiti del gruppo scultoreo in figura 1.33.....	69
<i>Figura 1.35:</i> Il secondo gruppo scultoreo utilizzato nella prova.....	70
<i>Figura 1.36:</i> Posizioni di presa e punti restituiti del gruppo scultoreo rappresentato in figura 1.35.....	70
<i>Figura 1.37:</i> Schema del blocco elementare in applicazioni mobile mapping.....	72
<i>Figura 1.38:</i> Strategie per la structure and motion applicata al MM.....	73
<i>Figura 1.39:</i> Geometria epipolare in applicazioni mobile mapping.....	74

<i>Figura 1.40:</i> Traiettoria dal veicolo nella prova descritta.....	75
<i>Figura 1.41:</i> Differenze fra soluzione di navigazione GPS e S&M.....	77

CAPITOLO 2

<i>Figura 2.1:</i> Precisioni di posizionamento della patch, al variare delle funzioni di forma.....	85
<i>Figura 2.2:</i> Matrice disegno e matrice normale nell'MGCM.....	89
<i>Figura 2.3:</i> Schema delle operazioni di inizializzazione per l'MGCM.....	90
<i>Figura 2.4:</i> Schema delle posizioni dei punti distorti.....	93
<i>Figura 2.5:</i> Schema di scomposizione del piano immagine mediante i quad-tree.....	94
<i>Figura 2.6:</i> Trasformazione dei toni di grigio di un'immagine.....	97
<i>Figura 2.7:</i> Schema della generazione di immagini epipolari (metodo proiettivo).....	99
<i>Figura 2.8:</i> Esempio di generazione di immagini epipolari con il metodo proiettivo.....	101
<i>Figura 2.9:</i> Esempio di generazione di immagini epipolari con il metodo polare.....	103
<i>Figura 2.10:</i> Schema di traiettoria di crescita nel matching sequenziale.....	107
<i>Figura 2.11:</i> Schema delle principali operazioni nel metodo di matching denso sequenziale.....	108
<i>Figura 2.12:</i> Fotogrammetria e laser scanning: primo caso di studio.....	110
<i>Figura 2.13:</i> Fotogrammetria e laser scanning: secondo caso di studio.....	111
<i>Figura 2.14:</i> Differenze di posizione (test con slitta micrometrica).....	114
<i>Figura 2.15:</i> Errori di posizionamento e dimensioni della template (test con slitta micrometrica).....	115
<i>Figura 2.16:</i> Campo di deformazione di un provino di conglomerato bituminoso.....	116
<i>Figura 2.17:</i> Grafico della deformazione di un provino d'acciaio.....	117
<i>Figura 2.18:</i> Campo di deformazioni di un provino d'acciaio.....	118
<i>Figura 2.19:</i> Confronto con la mesh originale del DSM ottenuto tramite LSM.....	119
<i>Figura 2.20:</i> Confronto con la mesh originale del DSM ottenuto tramite MGCM.....	119
<i>Figura 2.21:</i> Particolare della cupola dell'Ospedale Vecchio di Parma.....	120
<i>Figura 2.22:</i> Ospedale Vecchio: meshing di una porzione della nuvola di punti ottenuta.....	121
<i>Figura 2.23:</i> Ospedale Vecchio: distanze geometriche dalla sfera interpolante.....	122
<i>Figura 2.24:</i> DSM della parete nord-ovest dell'ammasso di Longeborgne.....	123

CAPITOLO 3

<i>Figura 3.1:</i> Operazioni utilizzate nell'analisi di point cloud interattiva.....	127
<i>Figura 3.2:</i> Selezione interattiva di porzioni di DSM.....	128
<i>Figura 3.3:</i> Segmentazione automatica di piani con metodi di stima robusti.....	132
<i>Figura 3.4:</i> Segmentazione automatica: voxelling e successiva aggregazione.....	133
<i>Figura 3.5:</i> Distanze assolute fra livelli della piramide del DSM.....	134
<i>Figura 3.6:</i> Segmentazione automatica: clustering gerarchico.....	137
<i>Figura 3.7:</i> Segmentazione automatica: suddivisione tramite k -means.....	138
<i>Figura 3.8:</i> Trasformata di Hough: spazio euclideo e spazio dei parametri.....	140
<i>Figura 3.9:</i> Trasformata di Hough: spazio di accumulazione.....	141

CAPITOLO 4

<i>Figura 4.1:</i> Definizione degli angoli di dip e dip direction.....	144
<i>Figura 4.2:</i> Rilievo tradizionale delle superfici di discontinuità.....	145
<i>Figura 4.3:</i> Esempio di rilievo condotto per <i>scanline</i> lungo la parete.....	146
<i>Figura 4.4:</i> Schema per il rilievo, la modellazione e l'analisi di pareti rocciose.....	148
<i>Figura 4.5:</i> Inquadramento cartografico del sito di Arnad.....	152
<i>Figura 4.6:</i> Suddivisione del DSM in micro-zone e in macro-zone (<i>b.</i>).....	154
<i>Figura 4.7:</i> Suddivisione del DSM: prove di segmentazione.....	155
<i>Figura 4.8:</i> Segmentazione del DSM ottenuta per mezzo dell'algoritmo automatico.....	156
<i>Figura 4.9:</i> Arnad: stereogrammi ottenuti con le diverse modalità.....	157
<i>Figura 4.10:</i> Inquadramento cartografico del sito di Longeborgne.....	158
<i>Figura 4.11:</i> DSM della parete nord-ovest dell'ammasso di Longeborgne.....	159
<i>Figura 4.12:</i> Longeborgne: stereogrammi ottenuti con le diverse modalità.....	161
<i>Figura 4.13:</i> Inquadramento cartografico del sito del Granier.....	163
<i>Figura 4.14:</i> Posizioni di presa e punti di legame per il sito del Granier.....	164
<i>Figura 4.15:</i> DSM della parete nord del complesso del Granier.....	165
<i>Figura 4.16:</i> Granier: stereogrammi ottenuti con le diverse modalità.....	166

Appendice A

<i>Figura A.1:</i> Rappresentazione geometrica della retta proiettiva IP^1	174
<i>Figura A.2:</i> Rappresentazione della sfera unitaria.....	176
<i>Figura A.3:</i> Rappresentazione della condizione di Armijo nel line-search.....	189

Indice delle Tabelle

CAPITOLO 1

<i>Tabella 1.1:</i> Feature matching: confronto fra i due metodi di ricerca.....	28
<i>Tabella 1.2:</i> Immagini sintetiche: riepilogo dei risultati ottenuti con la structure and motion.....	66
<i>Tabella 1.3:</i> Arnad - Longeborgne: riepilogo dei risultati ottenuti con la structure and motion.....	68
<i>Tabella 1.4:</i> Pantheon: riepilogo dei risultati ottenuti con la structure and motion.....	71
<i>Tabella 1.5:</i> Mobile mapping: statistiche degli errori sui parametri di orientamento.....	76

CAPITOLO 2

<i>Tabella 2.1:</i> LSM: precisioni di misura e dimensioni della template.....	113
<i>Tabella 2.2:</i> Immagini sintetiche: differenze fra mesh originale e DSM con LSM e MGCM.....	120
<i>Tabella 2.3:</i> Ospedale Vecchio: distanze fra i 200 check point e il DSM.....	122
<i>Tabella 2.4:</i> Longeborgne: distanze fra i 200 check point e il DSM.....	123

CAPITOLO 4

<i>Tabella 4.1:</i> Arnad: risultati ottenuti dal confronto con i check point.....	153
<i>Tabella 4.2:</i> Longeborgne: risultati ottenuti dal confronto con i check point.....	160
<i>Tabella 4.3:</i> Longeborgne: distanze fra i 200 check point e il DSM.....	160

Appendice A

<i>Tabella A.1:</i> Ransac: numero di iterazioni minime assumendo la probabilità $p = 99\%$	192
<i>Tabella A.2:</i> Ransac: numero di iterazioni minime assumendo la probabilità $p = 95\%$	192

INTRODUZIONE

Presentazione degli obiettivi e sintesi del lavoro

Negli ultimi due decenni le scienze geomatiche, e fra esse la fotogrammetria, hanno assistito a notevoli cambiamenti, sia di ordine operativo sia di ordine teorico-metodologico, in virtù del crescente impatto delle nuove tecnologie in tutti gli aspetti della nostra professione. In particolare la fotogrammetria ha visto crescere una comunanza di interessi con altre discipline, prima fra tutte la visione artificiale: sebbene la prima curi soprattutto gli aspetti metrologici e la seconda le applicazioni in real-time, i presupposti operativi (specie nel close-range) e le problematiche di entrambe spesso risultano affini. La cosa non deve sorprendere se si considera che i nostri cicli operativi fanno affidamento sull'uso di calcolatori sempre più evoluti e prestanti e il divario fra costo delle risorse umane e i costi macchina tende ad aumentare in maniera esponenziale. Se quindi, da un lato, le tecniche di visione artificiale, per rispondere a richieste sempre più impegnative, devono tener conto anche degli aspetti metrologici, dall'altro la fotogrammetria non può continuare a svilupparsi se non puntando decisamente verso l'automazione delle procedure. Da ciò il sostanziale interesse che le due discipline tendono ad avere nei confronti l'una dell'altra, seppur i presupposti teorici e metodologici dai quali prendono spunto siano differenti.

L'introduzione di nuove tecnologie e una sempre maggiore integrazione di sensoristica (laser a scansione, GPS e sistemi di navigazione inerziale) hanno oggettivamente spostato l'interesse del mercato a svantaggio della fotogrammetria: lo studio e l'implementazione di nuove soluzioni, che siano in grado di fornire lo stesso livello prestazionale, soprattutto in termini di facilità d'uso, tempi di acquisizione ed esperienza richiesta all'operatore, sono essenziali per garantire lo sviluppo della disciplina e mantenerla al passo delle altre tecnologie di rilievo.

La presente dissertazione di dottorato verte proprio sullo sviluppo e l'analisi di tecniche di automazione in fotogrammetria dei vicini (close-range), affrontando (eccetto l'acquisizione) tutte le fasi del processo fotogrammetrico: orientamento, restituzione, classificazione dei dati.

Una larga parte del lavoro è stata focalizzata sul tema dell'automazione dell'orientamento di sequenze di immagini, che è stato applicato con successo sia nel Mobile Mapping, sia nel settore architettonico, sia nel rilievo di pareti rocciose.

Per quanto riguarda la restituzione, sono state implementate tecniche di correlazione a minimi quadrati multi-immagine con vincoli geometrici e strategie per il matching denso nella restituzione di superfici, applicate sia nel settore del rilievo architettonico, sia nello studio delle deformazioni dei materiali, sia nella generazione di DSM.

Particolare attenzione viene prestata al tema dell'analisi di stabilità di pareti rocciose, sul quale da alcuni anni vi è un vivo interesse del settore geotecnico in ambito internazionale: è in tale ambito che sono state sviluppate le tecniche più innovative, soprattutto nell'estrazione (interattiva o completamente automatizzata) delle discontinuità di pareti rocciose. Inoltre, la messa a punto di una procedura integrata di misurazione ed analisi dei dati, applicata a questo tipo di problemi, ha permesso di collegare tutti gli aspetti descritti in un unico processo, mostrando come l'automazione delle operazioni di rilievo sia un obiettivo perseguibile.

La struttura della dissertazione ricalca in modo sequenziale le diverse fasi di rilievo fotogrammetrico e di interpretazione dei dati.

Nel primo capitolo vengono affrontate le tematiche legate al problema della Structure and Motion, ovvero della determinazione dei parametri di orientamento e delle caratteristiche geometriche di oggetti, determinabili dallo studio di una sequenza di immagini: tematica storicamente legata al settore della visione artificiale, ha visto nell'ultimo decennio un crescente interesse da parte del mondo della fotogrammetria. Da una generica sequenza di immagini, per mezzo di operatori per l'estrazione di features, modelli matematici desunti dalla geometria proiettiva e algoritmi di stima robusta si è in grado di generare accoppiamenti fra punti individuati su immagini differenti, filtrare quelli non corretti e risolvere, senza alcuna conoscenza di parametri approssimati, le equazioni di collinearità, ottenendo, al termine della procedura, i parametri di orientamento e la posizione tridimensionale dei punti.

Dopo aver presentato e discusso tecniche di feature extraction e feature matching, si illustrano i concetti di geometria epipolare e geometria trinoculare, introducendo la matrice fondamentale ed il tensore trifocale, che sono gli elementi chiave dell'approccio proiettivo all'orientamento di una sequenza. Particolare attenzione è data inoltre al problema della robustezza di stima, essendo sempre elevata la percentuale di errori grossolani nel feature matching..

I punti di forza del metodo sono il fatto che non sia richiesta la conoscenza di alcuna informazione preliminare sulla struttura del blocco o della scena, per cui tutte le procedure sono automatiche; al tempo stesso, la sua generalità, permette che sia efficacemente applicato in molti ambiti: rilievo architettonico, navigazione, controlli industriali ma sono numerose le applicazioni anche in campo archeologico, medico, cinematografico etc. Al termine del capitolo vengono presentati i principali e più interessanti risultati ottenuti nelle applicazioni attualmente prese in considerazione.

Nel capitolo 2 vengono analizzate le tematiche legate alla restituzione densa di entità puntuali con tecniche di correlazione d'immagini, con le quali è possibile, sempre in maniera totalmente automatica sfruttando anche le informazioni ottenute durante la fase di Structure and Motion, descrivere la superficie oggetto con un insieme costituito da qualche decina di migliaia a qualche milione di punti. Dopo aver presentato i metodi di correlazione semplice ed il least squares matching, si illustra il matching multi-immagine vincolato geometricamente.

Vengono poi analizzate le principali tecniche al momento implementate per ottimizzare il contenuto informativo delle diverse immagini, che permettono di ottenere procedure di correlazione più efficienti ed accurate. Inoltre vengono illustrate le procedure e le strategie di ottimizzazione per ottenere il maggior numero possibile di punti nella nuvola.

Anche in questo caso, al termine del capitolo, vengono analizzate le effettive prestazioni dei metodi sviluppati, sia dal punto di vista delle precisioni che da quello dell'affidabilità di restituzione; vengono inoltre illustrate alcune applicazioni a nostro avviso particolarmente interessanti.

Nel terzo capitolo si vanno a presentare le tecniche sviluppate per l'analisi della "nuvola" di punti ottenuta durante la fase di matching. Tali aspetti assumono particolare importanza per lo studio del campo di discontinuità delle pareti rocciose che, come già detto, è al momento l'applicazione sviluppata che più delle altre, fa uso di tecniche di automazione: pertanto la trattazione, seppur generale, tende a mettere in luce maggiormente i problemi derivanti da questo tipo di analisi. Vengono presentati due metodi di segmentazione: uno semiautomatico, in cui l'operatore individua una porzione limitata del DSM ed il software provvede a segmentarlo, mettendo in rilievo le regioni piane presenti. Nel secondo, totalmente automatico, il calcolatore opera direttamente sulla nuvola di punti. Sebbene ancora la sperimentazione non possa dirsi conclusa, anche in virtù del fatto che la verifica dei risultati è spesso dispendiosa e complessa, la metodologia sviluppata è estremamente interessante e può essere applicata a diversi aspetti del rilievo.

Il quarto capitolo illustra i risultati ottenuti dall'applicazione congiunta di tutte le tecniche precedentemente illustrate, nel caso specifico dell'estrazione dei piani di discontinuità nello studio di analisi di stabilità di pareti rocciose. Dopo aver sintetizzato le metodiche impiegate nell'analisi di stabilità ed i limiti delle tecniche di rilievo geostrutturale attualmente adottate, si presenta un procedimento alternativo basato sull'analisi di un DSM accurato della parete, rilevato fotogrammetricamente; seguono una serie di osservazioni ed indicazioni, desunte dalle esperienze sul campo finora effettuate, legate ai problemi di progettazione e realizzazione del rilievo. Infine vengono presentati i risultati ottenuti su tre diversi siti campione che permettono una valutazione delle reali capacità dei metodi presentati.

In un capitolo conclusivo, infine, si cerca di fornire un quadro riassuntivo di tutti gli aspetti trattati nella dissertazione, con particolare riguardo alle possibilità applicative emerse, mettendo in risalto gli elementi di originalità svolti e i possibili sviluppi futuri. Molti di questi stessi aspetti verranno comunque analizzati in ciascun capitolo, per permettere un collegamento più immediato agli aspetti più particolari sia dal punto di vista teorico e metodologico che da quello applicativo.

In appendice A vengono illustrati aspetti teorici generali legati ad aspetti specifici dello sviluppo dei metodi di automazione: alcuni richiami di geometria proiettiva, metodologie di stima robusta e metodi di ottimizzazione, metodi di resezione diretta; infine verranno brevemente analizzati alcuni aspetti generali legati alla logica fuzzy che rappresenta, al momento, la direzione di sviluppo più interessante per molti degli algoritmi presentati nella tesi.

In appendice B, infine, viene presentato un breve formulario che riassume molte delle proprietà ottenibili dall'utilizzo di tecniche proiettive in problemi fotogrammetrici.

Convenzioni e simboli

Nella presente dissertazione si sono seguite le seguenti convenzioni tipografiche:

- Per quanto riguarda i riferimenti bibliografici, essi sono indicati fra parentesi quadre con il nome del primo autore e l'anno di pubblicazione: tale riferimento è rintracciabile nella bibliografia inserita al termine della presente, in cui tutte le citazioni sono elencate in ordine alfabetico e per data di pubblicazione. Qualora, per uno stesso autore, siano presenti più riferimenti pubblicati nello stesso anno, il riferimento è seguito da una lettera.
- Per quanto riguarda le equazioni inserite nel testo, quelle più significative vengono inserite in una riga isolata e numerate nella forma (x.xx), in cui (x) indica il capitolo in cui la formula è inserita e (xx) la numerazione progressiva. Gli elementi scalari vengono solitamente indicati con lettera corsiva minuscola; gli elementi vettoriali bidimensionali con lettera corsiva minuscola in grassetto, quelli tridimensionali o n -dimensionali con lettera corsiva maiuscola in grassetto; le matrici vengono sempre indicate con lettera corsiva maiuscola; gli elementi tensoriali verranno indicati con lettera corsiva minuscola o maiuscola a seconda della dimensione del corrispettivo elemento vettoriale. Gli indici vengono solitamente indicati con le lettere i, j, k in corsivo. Qualora tali convenzioni non vengano rispettate, per motivi di praticità di scrittura, il testo ne dà indicazione, illustrando la simbologia utilizzata. Con l'apice "T" si indicherà l'operazione di trasposizione di matrice, con l'apice "T" l'inversa della trasposta, mentre il punto moltiplicativo "." verrà solitamente sottinteso o indicherà il prodotto scalare fra due vettori. Il prodotto vettoriale viene indicato indifferentemente con il simbolo "×" o con il simbolo "∧"; inoltre il simbolo $[\mathbf{v}]_{\times}$, dove \mathbf{v} è un vettore tridimensionale, sta ad indicare la matrice emisimmetrica che codifica il prodotto vettoriale:

$$\mathbf{v} = (v_x, v_y, v_z)^T \Rightarrow [\mathbf{v}]_{\times} = \begin{vmatrix} 0 & -v_z & v_y \\ v_z & 0 & -v_x \\ -v_y & v_x & 0 \end{vmatrix}$$

- Per quanto riguarda le figure, esse saranno numerate nella forma (x.xx), in cui (x) indica il capitolo in cui la figura è inserita e (xx) la numerazione progressiva; l'elenco completo delle figure è inserito nell'indice corrispondente.
- Per quanto riguarda le tabelle, esse saranno numerate nella forma (x.xx), in cui (x) indica il capitolo in cui la figura è inserita e (xx) la numerazione progressiva; la numerazione delle tabelle è indipendente da quella delle figure; l'elenco completo delle tabelle è inserito nell'indice corrispondente.

Qualora, all'interno del testo, fosse necessario utilizzare simbologie differenti da quelle appena descritte per necessità di stesura, ne verrà data indicazione.

CAPITOLO 1

Structure and Motion

1.1. Introduzione

Una delle fasi più delicate del rilievo fotogrammetrico, come noto, è l'orientamento dei fotogrammi: se in fotogrammetria aerea il blocco generalmente ha una struttura ben definita e può far affidamento su tecniche e metodologie sperimentate in quasi un secolo di applicazioni, lo stesso non si può dire per le applicazioni terrestri. La fotogrammetria close-range ha subito un fortissimo sviluppo soprattutto nel corso degli ultimi anni, con la nascita di nuove applicazioni non solo nei tradizionali settori industriale ed architettonico, ma anche nel mondo della cinematografia, dell'archeologia, della medicina. Al tempo stesso la sensoristica ha fatto passi da gigante e se da un lato è possibile avere sul mercato fotocamere digitali ad alta risoluzione a prezzi contenuti, l'integrazione di tale tecnologie con altri tipi di sensori è un fenomeno in continua espansione. Risulta evidente, quindi, che ciascuna applicazione presenta i suoi aspetti peculiari sia in termini di risultati attesi, sia in termini di vincoli operativi per la progettazione e realizzazione del blocco. Da questo punto di vista l'impatto che la fase di orientamento del blocco può avere sull'intera procedura diventa cruciale, in virtù soprattutto del fatto che le caratteristiche del blocco non sono sempre uguali e che il più delle volte l'utilizzatore non ha una preparazione specifica nel settore. Anche dal punto di vista metrologico le precisioni richieste nelle diverse applicazioni possono essere assai differenti fra loro e molto spesso possono essere considerati altrettanto importanti altri aspetti connessi al rilievo: si pensi, per semplicità, allo sviluppo della computer grafica, in cui molte volte non è necessario raggiungere livelli di precisione elevatissimi, ma bisogna garantire che dal punto di vista visivo la scena o l'oggetto ricostruito siano verosimili.

All'inizio degli anni '90, con il diffondersi delle prime telecamere digitali, nel settore della visione artificiale l'attenzione si è concentrata sull'analisi di sequenze di immagini, affrontando la ricostruzione della Structure from Motion [Huang 1994], [Dhond 1989], [Mohr 1996], [Faugeras 1993] (a volte indicato anche con il nome di Structure and Motion per indicare la contemporaneità di stima che può essere richiesto

dal problema). Il problema può essere così riassunto: in quali casi ed in che modo è possibile determinare la struttura spaziale di un oggetto o più in generale di una scena avendo a disposizione una sequenza di immagini (ovviamente con punti di presa differenti) con il minor numero di informazioni a priori e parametri iniziali?

In altre parole si volevano individuare una serie di metodologie il più possibile automatizzate in grado di fornire la descrizione tridimensionale di un oggetto non disponendo di informazioni sui parametri di orientamento (interno, esterno od entrambi), eventualmente senza alcun tipo di informazione sull'oggetto, il tutto per mezzo di procedure che riducessero o eliminassero completamente l'intervento umano nel processo. La mente umana svolge quotidianamente tale operazione senza creare troppo scalpore: effettivamente per comprendere la forma e le caratteristiche spaziali di un oggetto non abbiamo quasi mai la necessità di toccarlo o di misurarlo ma ci basta poterlo guardare da alcune posizioni differenti. Il problema può quindi essere formulato come l'individuazione di tecniche e procedure in modo da "insegnare" ad un calcolatore

una metodologia per interpretare la spazialità di un insieme di oggetti che è in grado di "vedere" su una serie di immagini. Ciò dà ragione del fatto che tale problema (e le sue mille ramificazioni) sia stato inizialmente analizzato nel più vasto ambito dell'intelligenza artificiale e che tuttora trovi i maggiori spunti di sviluppo in ambienti legati alla visione artificiale. In particolare i ricercatori si sono sempre divisi sulla strada da seguire per risolvere il problema: da un lato, infatti, c'è chi sostiene che il metodo più efficiente è quello di cercare di emulare la visione "biologica" dell'essere umano sebbene, come nota O. Faugeras nella prefazione a [Hartley 2000], non si conosce probabilmente a sufficienza il funzionamento della visione umana per poter pensare

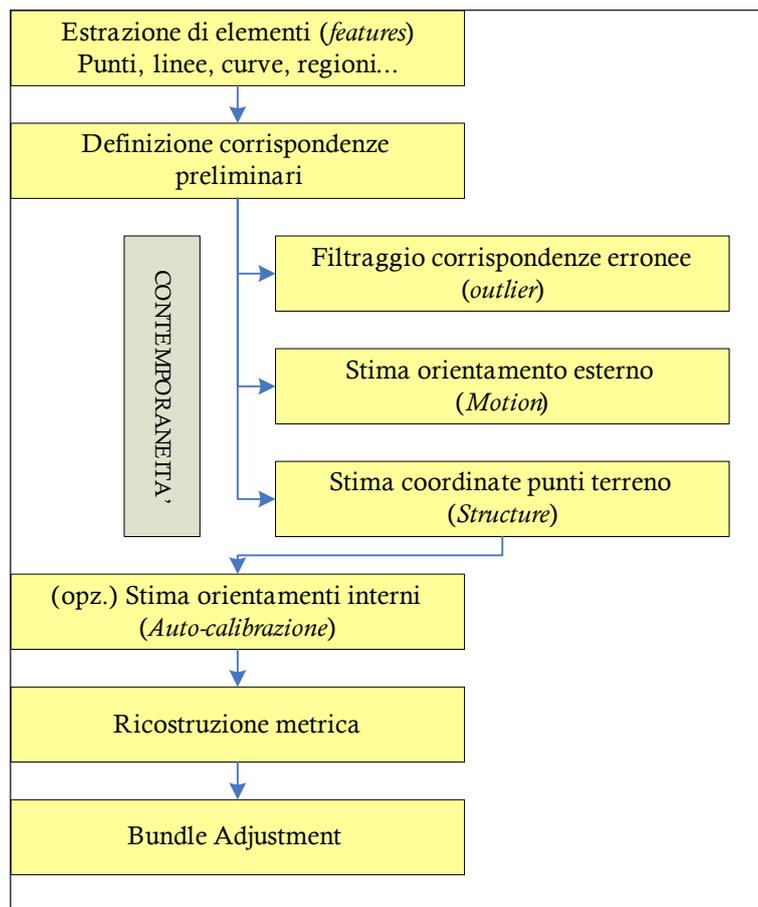


Figura 1.1: Schema a blocchi delle operazioni generalmente implementate in un problema di Structure and Motion.

di sviluppare un algoritmo o un'insieme di algoritmi che ne ricalchi il comportamento; dall'altro vi sono i ricercatori impegnati nello sviluppo di una visione "silicon-based" ovvero con regole di percezione della spazialità sostanzialmente differenti da quelle umane. Da questo punto di vista la fotogrammetria ha come disciplina l'indubbio vantaggio di avere uno sviluppo ormai quasi secolare alle spalle proprio nell'ambito dell'interpretazione matematica e algoritmica dei processi di formazione dell'immagine e della loro interpretazione; è dunque in questo senso che le differenti impostazioni possono trovare un punto di convergenza.

Lo sviluppo di una metodologia di visione e misurazione dello spazio tridimensionale non può chiaramente prescindere da una solida base teorica che fornisca il substrato metodologico per affrontare il problema in maniera efficiente da parte di un calcolatore. In questo senso diverse branche della matematica del XIX e XX secolo hanno fornito i principi teorici di base (vd. appendice A) per lo sviluppo di soluzioni mirate: ci riferiamo non soltanto alla nascita e allo sviluppo delle geometrie non euclidee ed in particolare dell'algebra proiettiva di cui si parlerà approfonditamente in seguito, ma anche di procedure di stima sempre più efficienti e robuste che permettono di risolvere i complessi sistemi di equazioni che codificano le caratteristiche del problema.

Attualmente tutti i metodi di risoluzione della structure from motion si basano su metodologie analoghe (si veda figura 1.1): inizialmente vengono estratte sulle diverse immagini una serie di elementi (solitamente puntuali, anche se non mancano implementazioni in cui si fa uso anche di elementi lineari) che, con un certo grado di probabilità siano riscontrabili su diversi fotogrammi: per mezzo di opportuni operatori matematici, si scandaglia tutta l'immagine analizzando i valori radiometrici sul dominio spaziale del fotogramma; ove tali valori presentano caratteristiche peculiari (ad esempio gradienti elevati in due direzioni mutuamente ortogonali) si suppone di aver individuato un punto maggiormente contrastato rispetto ad altri (ad esempio uno spigolo o un vertice dell'oggetto) e si ritiene che tale punto appaia ugualmente contrastato anche sulle altre immagini che compongono la sequenza.

Una volta individuato un numero sufficientemente elevato di elementi "interessanti", per mezzo di opportuni criteri geometrici e radiometrici, si individuano delle corrispondenze "putative" che formino un primo tentativo di catalogazione di corrispondenze omologhe sui diversi fotogrammi.

A questo punto è possibile stimare una geometria di presa preliminare ed eliminare le corrispondenze sbagliate. Generalmente per ottenere questa stima preliminare si fa uso della geometria epipolare fra due fotogrammi (orientamento relativo) o della geometria trifocale fra tre fotogrammi: è infatti facilmente dimostrabile che la geometria relativa fra due soli fotogrammi non è in grado di fornire un grado di robustezza sufficiente all'eliminazione sistematica di tutti gli errori di accoppiamento della fase precedente, soprattutto in quei casi in cui, a causa ad esempio di una non sufficiente caratterizzazione di tessitura, l'oggetto esaminato può condurre ad una serie numerosa di accoppiamenti ambigui; al contrario, la geometria intercorrente fra tre immagini distinte, gode di una robustezza decisamente maggiore e presenta configurazioni critiche molto raramente.

E' importante evidenziare che, fino a questo punto, solitamente, non si è ricorsi all'utilizzo di alcun parametro iniziale per vincolare la geometria di presa: in altre parole è solo la struttura matematica del fotogramma (analizzata per mezzo degli operatori di interesse), l'efficienza nell'*indovinare* gli accoppiamenti preliminari da parte degli opportuni algoritmi e dalla successiva fase di filtratura delle corrispondenze sbagliate che si arriva a definire una geometria di presa. Tale geometria non strettamente verifica le condizioni reali del blocco fotogrammetrico, nel senso che fornisce una geometria proiettivamente equivalente a quella reale. Dal momento che su tale argomento si dedicherà in seguito particolare attenzione, sia sufficiente in questa sede semplificare il problema dicendo che analiticamente è possibile ottenere, con le procedure poc'anzi illustrate, una configurazione in cui i parametri di orientamento esterno ed interno di ciascun fotogramma sia sostanzialmente differente da quello del blocco considerato, ma che in termini di residui nelle equazioni di collinearità fornisca lo stesso risultato del caso reale; così come la geometria del blocco risulta sostanzialmente deformata, anche la geometria dell'oggetto risulta compromessa pur essendo proiettivamente equivalente (secondo la definizione che verrà più avanti fornita) a quella dell'oggetto stesso.

Per ottenere infine una ricostruzione reale del blocco e della scena è sufficiente la conoscenza dei parametri di orientamento interno: la geometria interna al fotogramma infatti permette una ricostruzione conforme (ovvero equivalente a meno di una trasformazione di similitudine) del blocco e dell'oggetto considerato. Tali parametri possono essere noti a priori o, ancora una volta, possono essere stimati (seppur con un grado di approssimazione a volte non adeguato) direttamente dalle corrispondenze già determinate. In altre parole anche la ricostruzione conforme dell'oggetto e del blocco può avere luogo anche nel caso in cui siano parzialmente o completamente incogniti i parametri di calibrazione della camera utilizzata.

Per concludere, è possibile ottenere una stima ottimale (in termini di massima verosimiglianza), utilizzando i risultati ottenuti come parametri approssimati di un sistema risolvibile a minimi quadrati (*bundle adjustment*) in cui possono essere inserite informazioni aggiuntive sul blocco (punti d'appoggio, interdistanze note, orientamenti etc...). Si arriva così alla definizione completa dei parametri di orientamento esterno ed interno di ciascun fotogramma (*motion*) oltre che della geometria della scena (*structure*) per mezzo dei punti definiti dalle corrispondenze finali che risultano coerenti con la geometria stimata.

Nel seguito ciascun passaggio del metodo delineato verrà chiarito in maggior dettaglio, sviluppando i concetti matematici necessari alla sua implementazione e andando a sviluppare un'analisi critica delle tecniche attualmente disponibili. Verranno inoltre illustrati alcuni esempi reali in cui è possibile notare come l'impostazione generale del problema ne permetta l'utilizzo in una casistica estremamente ampia di applicazioni nell'ambito della fotogrammetria dei vicini.

E' importante infine evidenziare che, sebbene la struttura dell'oggetto determinata per mezzo dell'applicazione delle tecniche di *structure and motion*, non sia il più delle volte sufficiente a fornirne una descrizione esauriente (i punti vengono decisi arbitrariamente dall'operatore di interesse e possono non coincidere con i punti necessari all'operatore umano), essa può fornire i parametri iniziali per procedure di infittimento del contenuto informativo geometrico, come verrà dettagliatamente illustrato nel prossimo capitolo.

1.2. Operatori d'interesse

Il primo passo nello schema di figura 1.1. è la scelta di elementi facilmente rintracciabili nel maggior numero di fotogrammi che compongono la sequenza. Tali elementi sono generalmente di tipo puntuale in virtù del fatto che, nelle successive operazioni di filtraggio e di eliminazione di corrispondenze sbagliate, risulta più semplice e spesso più robusto lavorare con questo tipo di primitive. Non mancano però implementazioni in cui viene fatto largo uso anche di altre primitive, quali linee e curve: una vasta panoramica dei principali metodi algebrici per gestire tali problemi viene fornita in [Schmid 2000a]. Tali primitive si dimostrano particolarmente utili nello studio di scene in cui effettivamente è preponderante la presenza di elementi lineari (architettura, road modelling, alcune applicazioni industriali etc...).

Il criterio fondamentale per valutare la reale efficienza di un operatore di interesse è rappresentato dalla sua capacità di individuare il medesimo particolare (ovvero di fornire punti omologhi) anche su fotogrammi differenti per scala, per il diverso punto di vista, per condizioni di illuminazione, ecc. In altre parole un operatore di interesse è tanto più efficiente quanto maggiore è il suo grado di invarianza rispetto a trasformazioni geometriche e radiometriche via via più generali. Un confronto particolarmente interessante fra operatori differenti si può trovare in [Schmid 2000b] e in [Mikolajczyk 2003], dove, a seconda del tipo di trasformazione che interessa successivi fotogrammi della sequenza (rotazioni, variazioni di scala, cambiamenti di illuminazione e risposta radiometrica, effetti di prospettiva), è possibile individuare l'operatore migliore ovvero quello che fornisce il più elevato grado di ripetibilità (rapporto fra riconoscimenti corretti e numero di punti individuati sul fotogramma). Altro parametro fondamentale per valutare le prestazioni di un operatore di interesse è la precisione nella localizzazione, ovvero la capacità di individuare esattamente lo stesso punto in termini di coordinate immagine. Tale precisione può essere stimata conoscendo la geometria dell'oggetto o verificando i residui delle equazioni di collinearità dopo aver eseguito un bundle adjustment al termine del riconoscimento. Nel nostro studio, dopo aver trattato i principali fondamenti teorici che contraddistinguono gli operatori attualmente implementati, ci occuperemo in particolare del primo aspetto, rimandando alle succitate opere di Schmid e Mikolajczyk per una trattazione più ampia del problema.

Quasi tutti i descrittori di interesse attualmente impiegati, a prescindere dall'implementazione scelta, analizzano la funzione $I(x, y)$ (a dominio e valori discreti) che descrive la radiometria dell'immagine; come osservato in [Moravec 1977], la funzione di auto-correlazione dei toni di grigio $f(x, y)$, calcolata nell'intorno di un dato pixel su una finestra di dimensioni assegnate, può fornire informazioni sul tipo di struttura locale dell'immagine. Posto infatti:

$$f(x, y) = \sum_{(x_k, y_k)} (I(x_k, y_k) - I(x_k + \Delta x, y_k + \Delta y))^2 \quad (1.1)$$

se la finestra è priva di tessitura, $f(x, y)$ presenta valori prossimi a zero; qualora la finestra contenga un contorno (*edge*), la correlazione si mantiene alta nella direzione del contorno e cala rapidamente in direzione ad esso ortogonale; infine, nel caso la finestra contenga uno spigolo (*corner*) la funzione di auto-correlazione presenterà variazioni significative in tutte le direzioni.

Altri tipi di operatori, di cui però non ci occuperemo in questa sede, si basano invece sull'individuazione di contorni (linee e curve) per poi andare ad analizzare la curvatura locale di tali contorni (si veda ad esempio [Asada 1986] o [Medioni 1987]).

1.2.1. Operatore di Harris

L'operatore di Harris [Harris 1988] è sostanzialmente una rivisitazione ed un superamento di alcuni problemi di ordine concettuale del metodo proposto da Moravec [Moravec 1977]. Nell'implementazione di Moravec la funzione di autocorrelazione viene calcolata dando alla finestra gli spostamenti $\{(1,0), (1,1), (0,1), (-1,1)\}$ ovvero lungo gli assi x e y e le diagonali:

$$E(x, y) = \sum_{u, v} w(u, v) \cdot f(x, y) \quad (1.2)$$

dove w indica una funzione finestra che vale 1 nella zona corrispondente all'intorno del punto (x, y) (cioè appunto la finestra considerata) e 0 altrove.

L'implementazione di Moravec ha tre controindicazioni: poichè vengono prese in considerazione solo direzioni prefissate non è possibile mettere in luce nella stessa maniera linee o spigoli diversamente inclinati; essendo i punti all'interno della finestra equipesati, il metodo risente maggiormente del rumore nei toni di grigio dell'immagine; il metodo tende a rispondere eccessivamente in corrispondenza di elementi lineari dal momento che viene preso in considerazione solo il valore di E .

La soluzione proposta da Harris permette di risolvere tali problemi. Innanzitutto la funzione di auto-correlazione deve essere calcolata in tutte le direzioni. Analiticamente ciò è possibile considerando uno sviluppo in serie di Taylor di tale funzione (per semplicità arrestato al primo ordine, considerati gli spostamenti modesti che si vogliono analizzare):

$$I(x_k + \Delta x, y_k + \Delta y) \cong I(x_k, y_k) + \left. \frac{\partial I}{\partial x} \right|_{(x_k, y_k)} \Delta x + \left. \frac{\partial I}{\partial y} \right|_{(x_k, y_k)} \Delta y \quad (1.3)$$

Sostituendo nella (1.1) otteniamo:

$$f(x, y) = \sum_{(x_k, y_k)} \left(\left. \frac{\partial I}{\partial x} \right|_{(x_k, y_k)} \Delta x + \left. \frac{\partial I}{\partial y} \right|_{(x_k, y_k)} \Delta y \right)^2 \quad (1.4)$$

Sviluppando la quantità a secondo membro otteniamo:

$$f(x, y) = |dx \ dy| \begin{vmatrix} \sum_{(x_k, y_k)} \left| \frac{\partial I}{\partial x} \right|_{(x_k, y_k)}^2 & \sum_{(x_k, y_k)} \left| \frac{\partial I}{\partial x} \right|_{(x_k, y_k)} \left| \frac{\partial I}{\partial y} \right|_{(x_k, y_k)} \\ \sum_{(x_k, y_k)} \left| \frac{\partial I}{\partial x} \right|_{(x_k, y_k)} \left| \frac{\partial I}{\partial y} \right|_{(x_k, y_k)} & \sum_{(x_k, y_k)} \left| \frac{\partial I}{\partial y} \right|_{(x_k, y_k)}^2 \end{vmatrix} \begin{vmatrix} dx \\ dy \end{vmatrix} = |dx \ dy| A(x, y) \begin{vmatrix} dx \\ dy \end{vmatrix} \quad (1.5)$$

in cui la matrice A permette di descrivere approssimativamente la funzione di auto-correlazione nell'intorno del punto (x, y) .

Inoltre è possibile utilizzare nella (1.2) come funzione finestra una funzione non costante, come ad esempio una funzione gaussiana 2D con simmetria circolare, che permetta di filtrare eventuali fenomeni di rumore:

$$w(u, v) = e^{-\frac{u^2 + v^2}{2\sigma^2}} \quad (1.6)$$

Sostituendo la (1.5) e la (1.6) nell'equazione (1.2) si ottiene infine:

$$E(x, y) = |dx \ dy| M(x, y) \begin{vmatrix} dx \\ dy \end{vmatrix} \quad (1.7)$$

con

$$M(x, y) = \begin{vmatrix} \sum_{(x_k, y_k)} \left| \frac{\partial I}{\partial x} \right|_{(x_k, y_k)}^2 w(x_k, y_k) & \sum_{(x_k, y_k)} \left| \frac{\partial I}{\partial x} \right|_{(x_k, y_k)} \left| \frac{\partial I}{\partial y} \right|_{(x_k, y_k)} w(x_k, y_k) \\ \sum_{(x_k, y_k)} \left| \frac{\partial I}{\partial x} \right|_{(x_k, y_k)} \left| \frac{\partial I}{\partial y} \right|_{(x_k, y_k)} w(x_k, y_k) & \sum_{(x_k, y_k)} \left| \frac{\partial I}{\partial y} \right|_{(x_k, y_k)}^2 w(x_k, y_k) \end{vmatrix} \quad (1.8)$$

Si noti come E sia strettamente legata alla funzione di auto-correlazione e come M possa essere utilizzata per definirne la forma locale: detti infatti α e β i due autovalori della matrice M essi rappresentano le componenti principali di curvatura della funzione di auto-correlazione. E' immediato allora notare che i valori di α e β permettono di ottenere una descrizione della forma locale della funzione di auto-correlazione invariante rispetto ad una qualsiasi rotazione.

In figura 1.2 è evidenziata la classificazione del contenuto della finestra in funzione di α e β : quando entrambi gli autovalori α e β presentano valori bassi, la funzione di auto-correlazione fornisce valori prossimi a zero, ovvero il punto considerato si trova in una zona a tessitura omogenea; se uno dei due autovalori risulta più elevato dell'altro, la funzione di auto-correlazione ha valore maggiore lungo la direzione del corrispondente autovettore: siamo quindi in presenza di un contorno; se infine entrambi gli autovalori risultano grandi, la funzione di autocorrelazione tende a variare rapidamente in tutte le direzioni: siamo in presenza di uno spigolo. Per codificare tali informazioni in un unico valore, che misuri in grado di interesse del punto considerato, Harris suggerisce di considerare un opportuno operatore:

$$R = \text{Det}(M) - k \cdot \text{Tr}(M)^2 \quad (1.9)$$

Osservando che

$$\text{Det}(M) = \alpha\beta \quad e \quad \text{Tr}(M) = \alpha + \beta \quad (1.10)$$

si ottengono valori elevati di R se lo sono entrambi gli autovalori di M ; il secondo termine della (1.9) assume il significato di fattore di forma, limitando la risposta quando la funzione di auto-correlazione risulta allungata (quando cioè uno dei due autovalori è preponderante sull'altro). Generalmente la variabile k assume valori intorno a 0.04 ± 0.06 in presenza di corner ben definiti: la determinazione di tale valore è però del tutto empirica.

Per rendere efficiente il calcolo della funzione R in ciascun punto dell'immagine si può procedere nel seguente modo: per prima cosa si calcolano le derivate locali in direzione x e y per mezzo di una convoluzione con una maschera del tipo $[-1 \ 0 \ 1]$ o $[-2 \ -1 \ 0 \ 1 \ 2]$; alcuni autori [Schmid 2000b] suggeriscono di usare al posto di una maschera per le differenze finite, come le suddette, una maschera gaussiana con $\sigma = 1$; dalla convoluzione si determinano i termini omogenei della (1.5); eseguendo una ulteriore convoluzione, stavolta con una maschera gaussiana di dimensioni prefissate, si considera l'interazione della funzione finestra e si ottengono per ogni punto gli elementi della matrice M ; è quindi possibile calcolare il valore di R in ciascun punto dell'immagine.

Per selezionare i punti di interesse si effettua una sogliatura preliminare dei valori di R , trascurando cioè i punti che presentano valori di risposta troppo bassi. Infine per mezzo di una finestra di ricerca di dimensioni prefissate si vanno a ricercare i punti di massimo locale della funzione R . Tale funzione può essere interpolata, ad esempio per mezzo di un polinomio di grado n , in modo da ottenere una stima sub-pixel della posizione di massimo della funzione che, lo ricordiamo, essendo calcolata in corrispondenza di ciascun pixel, è a dominio discreto.

In base a quanto detto, risulta evidente che l'operatore di Harris è invariante rispetto a

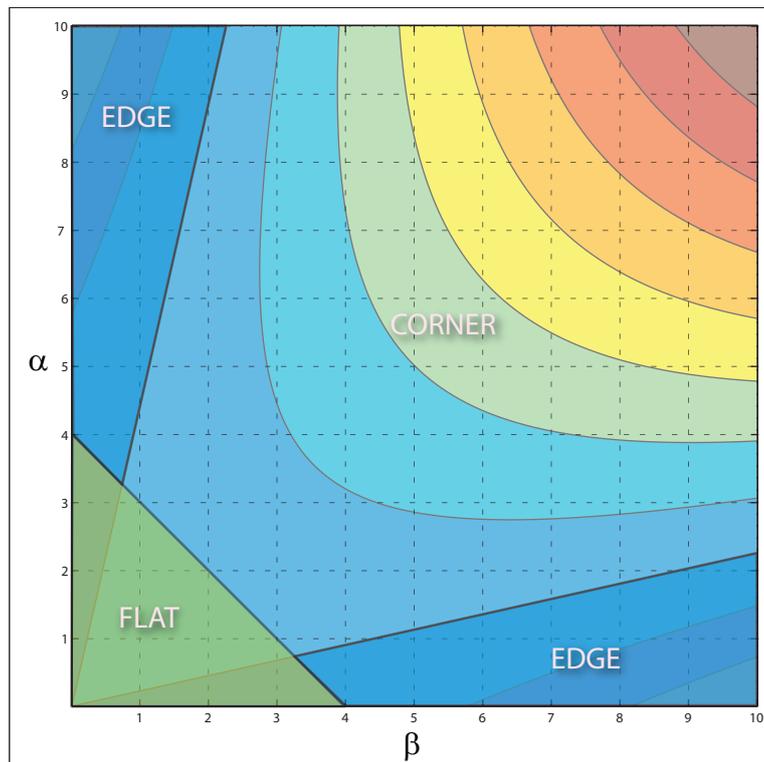


Figura 1.2: Rappresentazione per curve di livello della funzione di risposta R e della suddivisione tipologica in funzione degli autovalori α e β

rotazioni dell'immagine. Viceversa, se la trasformazione che lega localmente punti dell'oggetto in due differenti fotogrammi presenta variazioni di scala o deformazioni prospettiche, le sue prestazioni peggiorano drasticamente: mentre per rotazioni arbitrarie il grado di ripetibilità tende a mantenersi costantemente intorno a valori prossimi a 0.9 (il 90% dei punti estratti su un fotogramma coincidono con quelli estratti nell'altro), è sufficiente un fattore di scala di 1.5 per dimezzare il numero di riconoscimenti corretti. Analoghe osservazioni possono essere fatte per cambiamenti di illuminazione e variazioni prospettiche. Si rimanda all'ultima parte del paragrafo per un maggior approfondimento delle prestazioni degli operatori e della loro invarianza nei confronti di trasformazioni geometriche e radiometriche.

1.2.2. Operatore di Förstner

L'operatore di Förstner [Förstner 1987], sviluppato un anno prima di quello di Harris, pur da presupposti differenti, arriva ad un risultato assai simile. Il punto di partenza del lavoro di Förstner è infatti la correlazione a minimi quadrati o Least Squares Matching proposta precedentemente da Grün [Grün 1985]. Maggiori dettagli sul metodo verranno presentati nel paragrafo § 2.2; qui si osserva solo che, considerando la matrice dei pesi unitaria, la matrice normale del sistema a minimi quadrati è del tutto equivalente alla matrice $A(x, y)$ in (1.5):

$$N = \begin{vmatrix} \sum_{(x_i, y_i)} \left| \frac{\partial I}{\partial x} \right|_{(x_i, y_i)}^2 & \sum_{(x_i, y_i)} \left| \frac{\partial I}{\partial x} \right|_{(x_i, y_i)} \left| \frac{\partial I}{\partial y} \right|_{(x_i, y_i)} \\ \sum_{(x_i, y_i)} \left| \frac{\partial I}{\partial x} \right|_{(x_i, y_i)} \left| \frac{\partial I}{\partial y} \right|_{(x_i, y_i)} & \sum_{(x_i, y_i)} \left| \frac{\partial I}{\partial y} \right|_{(x_i, y_i)}^2 \end{vmatrix} \quad (1.11)$$

La matrice di covarianza dei due parametri spostamento (in riga e colonna) risulta quindi:

$$C_{xx} = \sigma_0^2 \cdot N^{-1} \quad (1.12)$$

Come noto, dalla matrice di covarianza si possono ricavare i parametri dell'ellisse di errore, che rappresenta in questo caso l'errore atteso nell'individuare il punto con tecniche di Least Squares Matching: maggiori le dimensioni dell'ellisse, maggiore l'incertezza nella determinazione. Förstner osservò poi che, un ellisse di errore molto allungato indica la presenza di un contorno: la posizione del punto lungo la direzione ortogonale al contorno è determinata con buona precisione, mentre in direzione parallela ad esso è mal determinata. Noti gli autovalori λ_1 e λ_2 della matrice normale (1.11), lo schiacciamento dell'ellisse è dato dal parametro *roundness*:

$$q = \frac{4 \text{Det}(N)}{\text{Tr}^2(N)} = 1 - \left(\frac{\lambda_1 - \lambda_2}{\lambda_1 + \lambda_2} \right) \quad (1.13)$$

che rappresenta il rapporto fra i due assi principali dell'ellisse, mentre l'area dell'ellisse è misurata da:

$$w = \frac{\text{Det}(N)}{\text{Tr}(N)} = \frac{\lambda_1 \lambda_2}{\lambda_1 + \lambda_2} \quad (1.14)$$

che prende il nome di valore di interesse (*interest value*).

In tal senso l'operatore di Förstner è sostanzialmente analogo a quello di Harris, non fosse che in questo caso l'autore non definisce una funzione di risposta che tenga in considerazione entrambi i parametri w e q . Valgono di conseguenza le considerazioni fatte per il comportamento dell'operatore in presenza di trasformazioni geometriche e radiometriche. A questo riguardo, però, può essere interessante osservare che, nell'impostazione di Förstner sarebbe possibile considerare un modello funzionale più complesso, andando ad esempio a considerare una trasformazione geometrica affine ed una variazione radiometrica lineare: in tal caso lo studio dell'ellissoide d'errore, seppur più complessa, porterebbe ad individuare quei punti dell'immagine che meno risentono di trasformazioni di quel tipo. Tuttavia, a nostro parere, tale percorso non porta ad un effettivo miglioramento del metodo, in quanto si presume che solo in condizioni particolari la natura del punto scelto permetta un'invarianza nella procedura di matching richiesta per determinarlo. In altre parole, è il descrittore di interesse che deve utilizzare criteri invarianti rispetto a determinate trasformazioni e non il punto a presentare una conformazione tale da rendere l'operatore invariante.

1.2.3. Operatore di Lowe (SIFT)

L'operatore di Lowe [Lowe 1999], [Lowe 2004] è attualmente uno degli algoritmi più studiati in ambito di ricerca (soprattutto in robotica e computer vision) a causa della sua estrema efficienza nel determinare features corrispondenti su più immagini. Il metodo non solo è invariante rispetto a trasformazioni di scala, rappresentando pertanto un notevole passo avanti rispetto ai metodi precedentemente analizzati, ma permette di generare per ciascuna feature dei descrittori univoci, una sorta di impronta digitale della feature, che favoriscono notevolmente le operazioni di matching.

L'invarianza rispetto alla variazione di scala è ottenuta per mezzo di un approccio piramidale in cui l'immagine sorgente viene campionata a diverse scale; al tempo stesso il suo contenuto di frequenze, a ciascun livello di scala, viene scomposto per mezzo di un filtro passa-banda basato sulla differenza di convoluzioni gaussiane. Il metodo prende spunto da studi sulla visione umana [Lindeberg 1994], dai quali risulta che l'attenzione del soggetto tende a soffermarsi su particolari differenti in funzione del livello di dettaglio percepibile (ovvero della distanza dalla quale si osserva l'oggetto) e a scale uguali si tende a concentrare l'attenzione localmente sulle bande di frequenze più significative.

Innanzitutto, dall'immagine originale si deriva il cosiddetto *Pyramid Scale Space* [Witkin 1983], ovvero una struttura dati in cui sono memorizzate immagini separate da intervalli di scala discreti; raggruppamenti di tali immagini vengono definiti da Lowe ottave. In ciascuna ottava, la differenziazione di scala non avviene per mezzo di un ridimensionamento dell'immagine, ma per mezzo di un filtro gaussiano che riduce il livello di dettaglio (*defocusing*). Tale struttura, sebbene costituita da una serie di immagini separate da intervalli di scala discreti, permette la rappresentazione continua della variazione del contenuto informativo estraibile in funzione della scala immagine. Le immagini che compongono le ottave sono le scomposizioni

in frequenza dell'immagine originale, ottenute come già anticipato per differenze fra convoluzioni con gaussiane isotrope di varianza differente: detta $I(x, y)$ la funzione a dominio e valori discreti che descrive l'immagine ad una determinata scala, risulta:

$$L(x, y, \sigma) = G(x, y, \sigma) * I(x, y) \quad (1.15)$$

dove si è indicato

$$G(x, y, \sigma) = \frac{1}{2\pi\sigma^2} e^{-\frac{(x^2 + y^2)}{2\sigma^2}} \quad (1.16)$$

Le immagini che compongono l'ottava ad una determinata scala sono date dalla differenza fra due funzioni derivate dalla (1.15) utilizzando differenti valori di σ :

$$D(x, y, \sigma) = (G(x, y, k\sigma) - G(x, y, \sigma)) * I(x, y) = L(x, y, k\sigma) - L(x, y, \sigma) \quad (1.17)$$

Come osservato in [Lindeberg 1994] e prima ancora in [Koenderink 1984] per ottenere una corretta rappresentazione dello spazio delle scale, in cui ad ogni livello sia garantito lo stesso grado di contenuto informativo, la funzione di ricampionamento deve essere una funzione Laplaciana: in [Lowe 2004] viene effettivamente dimostrato che la differenza di due gaussiane è proporzionale ad una funzione Laplaciana ed è quindi corretta per ottenere l'invarianza di scala.

Si determina un intervallo incrementale della varianza in (1.16) in modo tale che in ciascuna ottava l'ultima immagine abbia valore di σ uguale al doppio di quella iniziale. Il numero di ottave da considerare dipende dall'intervallo di scale che si vogliono analizzare: in alternativa, si può stabilire una dimensione minima di immagine in cui non ha più significato cercare di estrarre informazioni. Viceversa il numero di immagini all'interno di un'ottava può essere fissato in base al livello di discretizzazione che si vuole operare per lo scale-space: in [Lowe 2004] viene mostrato empiricamente che un valore pari a 3÷4 immagini è solitamente sufficiente.

Per determinare i punti di interesse si procede per mezzo di una ricerca dei massimi e minimi locali della funzione $D(x, y, \sigma)$ in ciascuna ottava: per mezzo di un volume di ricerca cubico (ad esempio di dimensione 3x3x3) si va a verificare quali punti risultano massimi locali in tale volume (in letteratura viene consigliato di scartare punti con valori di D minori a 0.03). Una volta individuato tale punto, per rendere più precisa l'individuazione del massimo o minimo locale, è possibile procedere come segue [Brown 2002]: si considera una funzione quadratica tridimensionale che interpoli i valori della funzione nello scale-space $D(x, y, \sigma)$ nel punto estremo trovato; tale funzione per mezzo di uno sviluppo in serie di Taylor arrestato al secondo ordine risulta:

$$D(x, y, \sigma) = D(\mathbf{x}) = D(\mathbf{x}_0) + \frac{\partial D^T}{\partial \mathbf{x}} \mathbf{x} + \frac{1}{2} \mathbf{x}^T \frac{\partial^2 D}{\partial \mathbf{x}^2} \mathbf{x} \quad (1.18)$$

Derivando la (1.18) e risolvendo per \mathbf{x} si ottiene la posizione dell'estremo:

$$\frac{\partial D}{\partial \mathbf{x}} = \frac{\partial D^T}{\partial \mathbf{x}} + \frac{\partial^2 D}{\partial \mathbf{x}^2} \mathbf{x} = 0 \quad \Rightarrow \quad \hat{\mathbf{x}} = - \frac{\partial D}{\partial \mathbf{x}} \frac{\partial^2 D^{-1}}{\partial \mathbf{x}^2} \quad (1.19)$$

Andando a sostituire la posizione trovata per mezzo della (1.19) nella (1.18) è possibile ricavare il valore corrispondente della funzione D :

$$D(\hat{\mathbf{x}}) = D(\mathbf{x}_0) + \frac{1}{2} \frac{\partial D^T}{\partial \mathbf{x}} \hat{\mathbf{x}} \quad (1.20)$$

Le derivate possono essere stimate localmente, ad esempio per mezzo di un approccio alle differenze finite.

Se non si considera la curvatura della funzione D non vi è garanzia che il punto determinato non giaccia su un bordo e quindi sia facilmente caratterizzabile in una direzione ma non in quella ortogonale. Seguendo un procedimento analogo a quello sviluppato per l'operatore di Harris si può allora utilizzare l'Hessiana di D (calcolata unicamente rispetto alle due variabili spaziali x e y):

$$H = \begin{vmatrix} \frac{\partial^2 D}{\partial x^2} & \frac{\partial^2 D}{\partial x \partial y} \\ \frac{\partial^2 D}{\partial y \partial x} & \frac{\partial^2 D}{\partial y^2} \end{vmatrix} \quad (1.21)$$

Gli autovalori della (1.21) permettono di determinare la forma locale (o meglio la curvatura nel piano xy) della funzione D e forniscono una valutazione sulla ripetibilità di riconoscimento del punto trovato.

In [Lowe 2004] viene inoltre proposto un metodo per classificare ciascuna feature trovata, costruendo un descrittore matematico che, pur non univoco, equivale ad una sorta di impronta digitale della feature, che sia esplicitamente confrontabile con le feature trovate in altre immagini. Al tempo stesso il metodo impiegato per il matching è estremamente veloce dal punto di vista computazionale. Di tale descrittore si parlerà più approfonditamente nel prossimo paragrafo in cui si analizzeranno gli algoritmi sviluppati per il feature matching.

1.2.4. Altri operatori di interesse

Oltre ai tre operatori sopra descritti, che sono stati adottati nella nostra implementazione della S&M, ci sembra opportuno presentarne brevemente altri, largamente utilizzati sia in ambito fotogrammetrico sia nell'ambito della visione artificiale, che vorremmo sperimentare prossimamente. Prestazioni comparabili a quella degli operatori di Harris e di Foerstner possono essere ottenuti ad esempio con l'operatore di Heitger ([Heitger 1992]) o di Reisfeld ([Reisfeld 1995]). Il primo è basato su un approccio ispirato dalla visione biologica dell'uomo e consiste nell'analizzare una serie di convoluzioni monodimensionali dell'immagine per mezzo di un filtro appropriato; il secondo invece considera la norma e la direzione dei gradienti locali

dei toni di grigio dell'immagine e ricerca i punti di interesse in quelle zone in cui sono soddisfatte particolari condizioni di simmetria.

La ricerca, negli ultimi anni, si concentra su operatori sempre meno suscettibili ai cambiamenti di scala e di punto di vista: gli studi iniziali di Crowley [Crowley 1981] nel dominio della scala immagine (si veda § 1.2.3.) hanno aperto la strada per i successivi sviluppi di Lindeberg, già citato per l'operatore di Lowe. Approcci differenti sono stati proposti da Kadir e Brady [Kadir 2001] in cui viene introdotta una quantità, da loro definita entropia, che generalmente risulta massima in corrispondenza di punti invarianti alle variazioni di scala.

Il campo forse di maggior interesse al momento è rappresentato dagli operatori invarianti rispetto alle trasformazioni affini: in questo caso, estendendo i concetti delineati nel precedente paragrafo, si analizzano livelli di scala differenti lungo due direzioni mutuamente ortogonali; in tale frangente infatti, l'operatore di Lowe, così come gli altri operatori basati sugli stessi concetti, hanno percentuali di ripetibilità notevolmente inferiori. In [Deriche 1993] e in seguito in [Alvarez 1997] si applica un'analisi multi-scala utilizzando una funzione di forma affine e si generano delle catene di punti di interesse riconosciuti a scale differenti, permettendo di localizzare i più stabili. Poiché il metodo ipotizza che gli angoli (*corner*) presenti sull'immagine derivino da elementi che si intersecano perpendicolarmente, si presta maggiormente ad applicazioni in ambiente *indoor*.

Di grande interesse sono poi i miglioramenti introdotti negli ultimi anni nell'operatore di Harris: in [Mikolajczyk 2001] esso viene esteso al caso tridimensionale considerando anche lo spazio delle scale, acquistando proprietà analoghe a quelle dell'operatore di Lowe. Gli studi di Baumberg [Baumberg 2000] hanno poi permesso di estendere il metodo anche al caso dell'invarianza affine ([Schaffalitzky 2001], [Schaffalitzky 2002]).

Partendo da presupposti differenti, infine, altri ricercatori sono arrivati ad ottenere operatori e descrittori invarianti rispetto a trasformazioni affini: ricordiamo in questa sede, per brevità, solamente il lavoro compiuto da Tuytelaars [Tuytelaars 1999], [Tuytelaars 2000] che probabilmente rappresenta una delle migliori implementazioni attualmente sviluppate.

1.2.5. Analisi comparativa degli operatori di interesse

Pur senza la completezza e il grado di approfondimento presente in [Schmid 2000b] e in [Mikolajczyk 2003] si è voluto analizzare le effettive prestazioni degli algoritmi implementati in relazione ad una serie di casi differenti e la corretta implementazione.

Sono state considerate 5 diverse immagini (reali o realizzate artificialmente con un software di modellazione) che possano fornire un quadro di situazioni sufficientemente ampio. Per ciascuna di tali immagini si è testato ciascun algoritmo andando a imporre trasformazioni note (rotazione, variazione di scala, trasformazione affine, variazioni di luminosità e contrasto) e nel caso delle immagini sintetiche si è anche

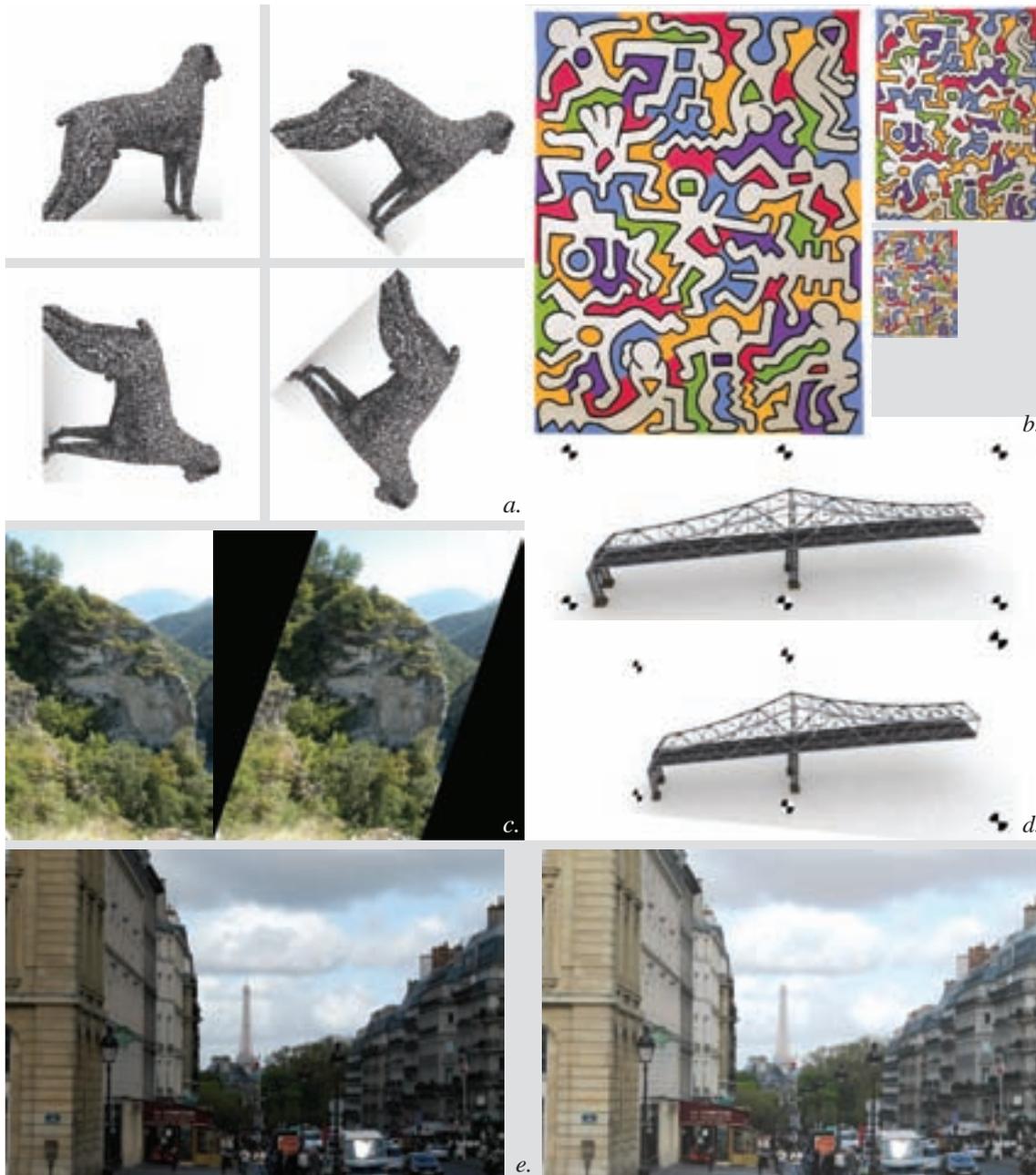


Figura 1.3: Immagini utilizzate per i test comparativi degli operatori di interesse implementati. Ognuna viene utilizzata per mostrare una delle trasformazioni applicate nella prova: *a.* trasformazione di rotazione; *b.* trasformazione di scala; *c.* trasformazione affine; *d.* cambiamento del punto di vista (omografia) utilizzata solo per i due casi generati artificialmente (*a.* e *d.*); *e.* trasformazione complessa di luminosità, contrasto e correzione gamma.

provveduto a considerare cambiamenti nel punto di vista. Si sono così ottenuti 22 differenti test in cui si sono analizzate le prestazioni di ciascuno dei 3 operatori al momento implementati, variando i parametri di trasformazione.

Di seguito vengono presentati i grafici riepilogativi dei risultati emersi: si è ritenuto corretto considerare come parametro più significativo per valutare l'efficienza degli operatori il grado di ripetibilità che, lo ricordiamo, è definito come la percentuale di features corrispondenti estratte dall'operatore su entrambi i fotogrammi. Tale valore dipende strettamente dall'errore di localizzazione, definito come la distanza di accettazione massima a cui devono trovarsi le features dalla loro posizione corretta valutata a partire dalle posizioni sul fotogramma di riferimento e dai parametri della trasformazione imposta: l'errore di localizzazione, in tutte le prove, è stato posto a 1.5 pixel. È importante osservare a questo riguardo che in molti casi pratici non è richiesta in operazioni di feature matching una precisione troppo elevata, in quanto, una volta identificati i punti corrispondenti, per mezzo di algoritmi di matching *Area Based* è possibile migliorare il posizionamento.

Si cominci con l'analizzare il test relativo alla trasformazione di rotazione: come possibile vedere in figura 1.3.a, in cui sono mostrate alcune delle immagini trasformate utilizzate relative ad una delle immagini utilizzate per la prova, si è realizzata una cornice monocromatica attorno al fotogramma originale in modo da garantire che l'area analizzata dall'operatore fosse sempre uguale per tutte le rotazioni imposte. I risultati sono presentati nel grafico di figura 1.4, in cui, come in tutti gli altri casi presi in considerazione, vengono visualizzati solamente i valori medi di ripetibilità ottenuti su tutte le immagini: è possibile notare che tutti gli operatori, visti i criteri sui quali sono basati, si comportano abbastanza bene di fronte a questo tipo di trasformazione; fa in parte eccezione l'operatore di Foerstner che presenta valori di ripetibilità più bassi del 10% ÷ 15% rispetto agli altri; tale comportamento è però identico in tutte le altre prove e non sembra quindi essere relativo solamente al caso della rotazione. È poi possibile osservare che l'andamento della curva tende a decrescere per i primi 45° per poi aumentare fino a 90°, diminuire nuovamente fino a 135° e infine aumentare di nuovo fino a 180°; tale comportamento può essere dovuto principalmente al fatto che nel caso delle rotazioni per 90° e 180° il ricampionamento dell'immagine è esatto, mentre per tutte le altre l'interpolazione (bilineare) utilizzata per calcolare la trasformazione dei toni di grigio può indurre cambiamenti in termini di contrasto in una certa percentuale di punti estratti con l'operatore.

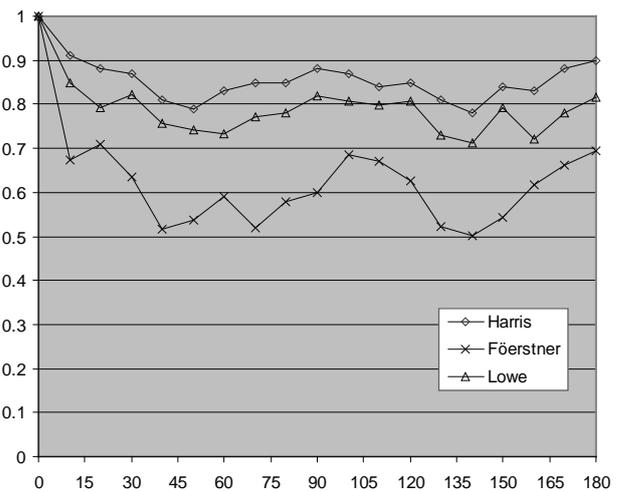


Figura 1.4: Grado di ripetibilità medio per le 5 immagini relativamente alla rotazione del piano immagine.

Nel caso delle variazioni di scala (figura 1.3.b) è immediato verificare, osservando i risultati del grafico di figura 1.5, che solo l'operatore di Lowe permette di ottenere gradi di ripetibilità elevati quando la variazione di scala supera il 25%: con solo il 50% di variazione l'operatore di Harris presenta coefficienti inferiori a 0.5, attestandosi rapidamente, per fattori più alti, su valori prossimi a 0.2; viceversa l'operatore di Lowe per riduzioni fino ad un terzo della dimensione originale presenta valori di ripetibilità approssimativamente pari a $0.7 \div 0.6$, valori solitamente accettabili per le procedure di feature matching di cui si parlerà nel prossimo paragrafo.

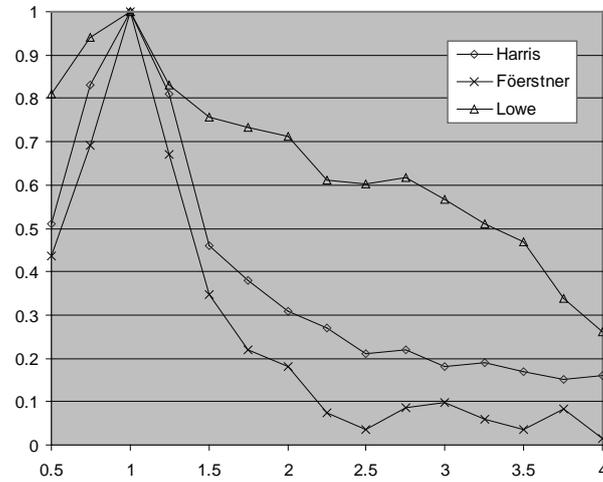


Figura 1.5: Grado di ripetibilità medio per le 5 immagini relativamente alla variazione di scala del piano immagine.

Per la trasformazione affine (figura 1.3.c) valgono sostanzialmente le stesse considerazioni per l'operatore di Lowe; al contrario gli altri operatori, si veda figura 1.6, presentano valori di ripetibilità lievemente superiori al caso precedente: la nostra interpretazione del fenomeno è che il tipo di trasformazione tenda ad sfavorire il riconoscimento di alcuni punti che presentano gradienti principali in direzione della variazione di scala, mentre influenza in grado minore quelli ortogonali.

Per considerare la variazione del punto di vista si è proceduto come segue: per mezzo di un programma di modellazione si sono generati due differenti immagini sintetiche (rappresentate in figura 1.3.c e 1.3.d); tali immagini sono state applicate ad una superficie piana alla quale sono stati solidarizzati sei target circolari: la scelta di considerare un oggetto piano deriva dalla necessità di poter stimare facilmente una trasformazione fra i punti dei vari fotogrammi per riportarli in un sistema di riferimento comune; al tempo stesso con tale geometria non si ha il problema di poter riscontrare occlusioni che possono drasticamente ridurre il grado di ripeti-

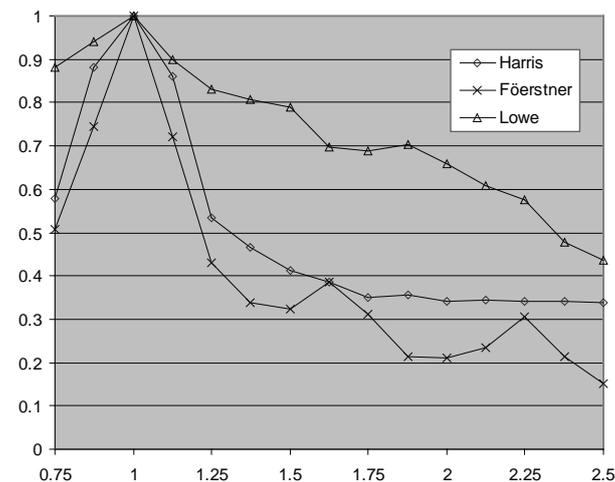


Figura 1.6: Grado di ripetibilità medio per le 5 immagini relativamente ad una trasformazione affine del piano immagine.

bilità, in quanto l'operatore va ad estrarre punti in zone differenti dell'oggetto. Si è poi provveduto a variare il punto di vista muovendo una camera virtuale su una circonferenza, variando l'angolo d'incidenza dell'asse principale con la normale al piano da 0° a 70° con passi costanti di 5° ; una volta estratti i punti con i vari operatori si sono determinate con LSM le posizioni dei centri dei vari target e si è stimata la trasformazione omografica con il piano immagine iniziale (cioè quello corrispondente a 0°); infine si sono trasformati, grazie all'omografia stimata, le posizioni corrispondenti di tutti i punti estratti. Per tenere in conto i possibili errori di centrimento dei target e della conseguente stima della trasformazione si è stimato mediamente di dover incrementare di 0.2 pixel la soglia di accettazione per il calcolo del grado di ripetibilità.

Come è possibile vedere in figura 1.7, e come del resto era prevedibile, tutti gli operatori non forniscono prestazioni accettabili superati i $35^\circ \div 45^\circ$; sia l'operatore di Harris che quello di Foerstner consentono di ottenere una buona ripetibilità fino a $15^\circ \div 20^\circ$, mentre quello di Lowe, che come di consueto si dimostra più robusto quando intervengono variazioni di scala, si può pensare di ottenere risultati accettabili fino a 45° . Ciononostante, per quanto osservato precedentemente sulle occlusioni e sulle altre possibili cause di disturbo bisogna considerare che, in casi reali non ottimali, il grado di ripetibilità può essere nettamente inferiore.

Infine si è considerato il caso di variazioni complesse delle caratteristiche illuminometriche del fotogramma: sono state considerate 15 repliche dei fotogrammi originali, per ognuna delle quali è stata applicata una trasformazione casuale di luminosità e contrasto oltre ad una correzione gamma (vd. § 2.4.2) sempre casuale; come si vede in figura 1.3.e tali variazioni sono comunque contenute in modo da non avere effetti di *clipping* dei toni di grigio. Il grafico

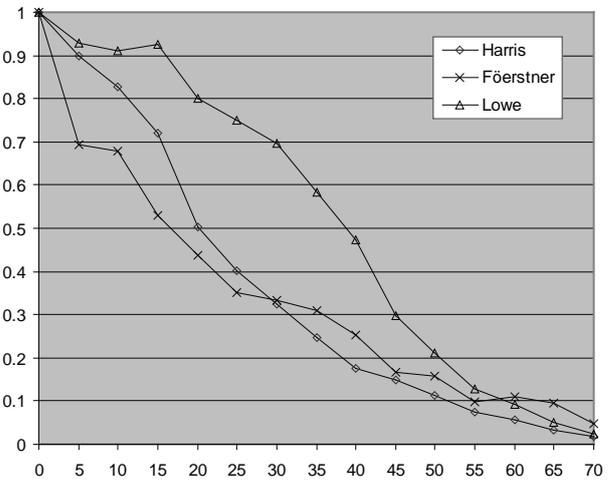


Figura 1.7: Grado di ripetibilità medio per le 2 immagini sintetiche relativamente alla variazione del punto di vista da cui è generata l'immagine.

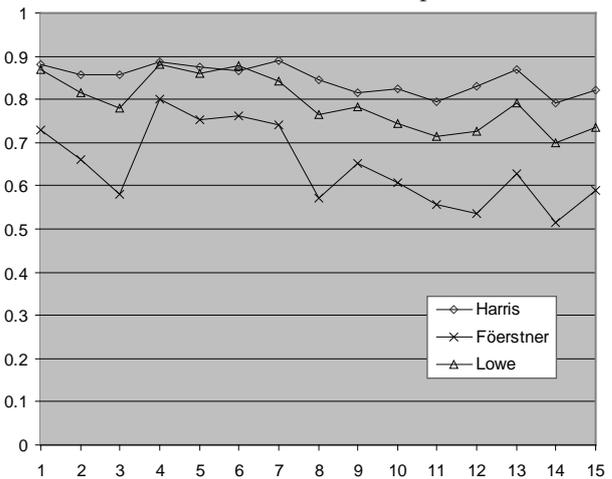


Figura 1.8: Grado di ripetibilità medio per le 5 immagini relativamente ad una variazione complessa delle caratteristiche di luminosità dell'immagine.

di figura 1.8 riporta i risultati di ripetibilità ottenuti nei vari casi: si può osservare come il comportamento dei vari operatori sia approssimativamente costante, ad eccezione dell'operatore di Foerstner che presenta oscillazioni un po' più pronunciate. Il grado di ripetibilità medio complessivo per l'operatore di Harris è stato dell'84.7 %, per Foerstner del 64.5 % e del 79.2 % per l'operatore di Lowe.

Cercando di riassumere i risultati emersi, possiamo dire che solitamente, quando intervengono variazioni fra le immagini limitate, è preferibile utilizzare l'operatore di Harris, in quanto è quello che fornisce (seppur lievemente) i risultati migliori e garantisce soprattutto uno sforzo computazionale minore; in parte esso, come del resto l'operatore di Foerstner è in grado di fornire risultati accettabili, se non ottimali, anche per variazioni del punto di vista, purché relativamente contenute ($15^\circ \div 20^\circ$); ciononostante in tali frangenti è sempre preferibile utilizzare l'operatore di Lowe, o comunque altri approcci invarianti rispetto alle trasformazioni di scala, in quanto forniscono prestazioni decisamente maggiori.

1.3. Feature Matching

Il *feature based matching* riguarda la determinazione di accoppiamenti fra punti omologhi su due o più fotogrammi, una volta individuate su tali fotogrammi, per mezzo di operatori di interesse, delle primitive geometriche significative. Pur facendo uso di algoritmi di stima robusti in grado di gestire percentuali significative di accoppiamenti sbagliati, la limitazione di tali errori grossolani non solo permette di ottenere con maggior probabilità una soluzione corretta, ma limitano grandemente il numero di operazioni richieste dall'elaboratore e dunque il carico computazionale dell'intero processo.

Si rimanda al precedente paragrafo per una descrizione dettagliata dei possibili operatori di interesse (di tipo puntuale) da utilizzare ed all'analisi delle loro capacità di determinare punti omologhi sui diversi fotogrammi a fronte di variazioni geometriche e radiometriche di diversa significatività. Nel presente paragrafo si andranno al contrario ad analizzare le tecniche attualmente sviluppate per determinare tali accoppiamenti.

E' importante evidenziare che proprio questa fase è di primaria importanza per garantire il successo delle successive elaborazioni: gli accoppiamenti determinati per mezzo delle procedure di feature matching vengono mantenute sostanzialmente per tutta la durata del processo; sebbene, una volta stimata una geometria di presa preliminare, si possa cercare di determinare ulteriori accoppiamenti, le coppie omologhe individuate in questa fase vengono mantenute per tutto il processo (a meno di non venire eliminate se catalogate come erronee). In altre parole se il processo condotto non ha permesso di individuare un numero sufficiente di accoppiamenti corretti, non solo i passaggi successivi possono portare a risultati non soddisfacenti, ma è anche probabile che la structure and motion dell'intera sequenza non venga determinata. Inoltre mentre la stima della geometria di presa e il filtraggio degli errori grossolani sia una fase sostanzialmente basata su tecniche analitiche largamente sperimentate e collaudate, il matching di features risente maggiormente

della vasta casistica di situazioni che si possono presentare (tipologie del blocco, variazioni di geometria, radiometria etc...) da richiedere una maggior possibilità di adattamento alle particolari condizioni del problema.

1.3.1. Matching procustiano

In molte differenti implementazioni di algoritmi per la structure and motion descritte in letteratura la fase di generazione di coppie omologhe preliminari avviene nel seguente modo: per ciascun punto estratto nella prima immagine vengono determinati i punti dell'altra con valore di parallasse (intesa come differenza lungo le due direzioni x e y fra le coordinate nei due fotogrammi) minore di una certa soglia; tra questi viene scelto il punto che presenta il massimo valore di similitudine dei toni di grigio con quelli del punto originale (ad esempio con metodi di correlazione semplice, si veda il prossimo capitolo). Sebbene il metodo fornisca risultati il più delle volte accettabili, il confronto effettuato con questa tecnica è strutturalmente locale, poiché non vengono presi in considerazione contemporaneamente tutti gli accoppiamenti possibili e le relazioni geometriche e topologiche fra punti vicini. In termini più rigorosi possiamo asserire che nel metodo descritto si determina l'accoppiamento migliore di r punti a partire da due insiemi di m ed n punti cercando di minimizzare una funzione obiettivo globale (rappresentata ad esempio dalle differenze in termini di coordinate e di toni di grigio negli intorni dei punti) analizzando separatamente una serie di sottoinsiemi, non necessariamente disgiunti, del suo dominio è che non è possibile in tal modo garantire una convergenza al minimo globale. Si è ottenuto un deciso miglioramento in termini di affidabilità sviluppando un algoritmo basato sull'analisi procustiana, che cerca la miglior coerenza degli accoppiamenti analizzando contemporaneamente tutte le possibili combinazioni.

Il metodo, dovuto inizialmente a Scott e Longuet-Higgins [Scott 1991] e ulteriormente sviluppato da Pilu [Pilu 1997], si basa sulla decomposizione ai valori singolari (SVD) di una particolare matrice di similitudine. Si costruiscono anzitutto due matrici D ed S di dimensione $m \times n$, una contenente l'informazione sulle parallasse, l'altra sui valori di correlazione dei toni di grigio, pesate entrambe tramite una funzione gaussiana. Scorrendo la riga i -esima di una delle due matrici si legge cioè una funzione delle parallasse o delle similitudini dei g.v. tra il punto i della prima immagine e tutti gli n punti estratti nella seconda immagine. Le matrici vengono poi moltiplicate elemento per elemento, in modo da ottenere una matrice Q che tenga in conto contemporaneamente i due aspetti. Infine la matrice Q viene decomposta con la SVD e, dopo aver posto a 1 tutti i valori singolari nella matrice diagonale, le due matrici ortogonali destra e sinistra vengono moltiplicate ottenendo la matrice M . Si esaminano poi i valori non diagonali di M : l'accoppiamento tra i punti i e j è (con elevata probabilità) corretto se l'elemento $M(i, j)$ è il massimo sia per riga sia per colonna. Si tratta di un metodo certamente costoso dal punto di vista computazionale, ma comunque del tutto affrontabile con le risorse di calcolo odierne, anche con un numero di punti per immagine superiori al migliaio.

Nelle due implementazioni di Scott e di Pilu, come già detto, viene utilizzata una funzione gaussiana a simmetria circolare per pesare i valori di parallasse e di correlazione, i cui parametri vengono fissati a priori e mantenuti costanti durante il processo.

L'uso della funzione gaussiana può essere interpretato come l'assegnazione di probabilità agli accoppiamenti, in funzione della loro distanza dall'origine della distribuzione (ovvero dalla posizione del punto omologo, ipotizzata a priori). Con una simmetria circolare della gaussiana (ovvero con autovalori coincidenti e covarianza nulla) non ci si adatta tuttavia alle condizioni reali della sequenza.

Si è pertanto proposto in [Roncella 2005a] di sfruttare i gradi di libertà per far assumere alla funzione peso una forma ellittica (vedi figura 1.9). Il campo di parallasse fra due fotogrammi successivi in una sequenza presenta una componente dipendente dalla geometria dell'oggetto, che per i nostri scopi può essere

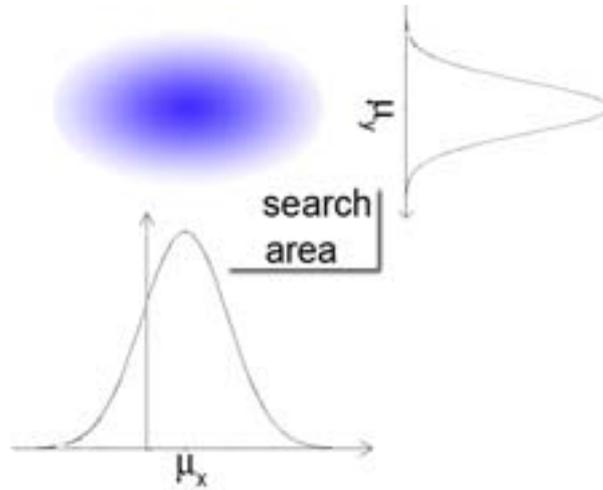


Figura 1.9: Rappresentazione della finestra di ricerca definita dai parametri di media e di dispersione delle parallasse in direzione x e y .

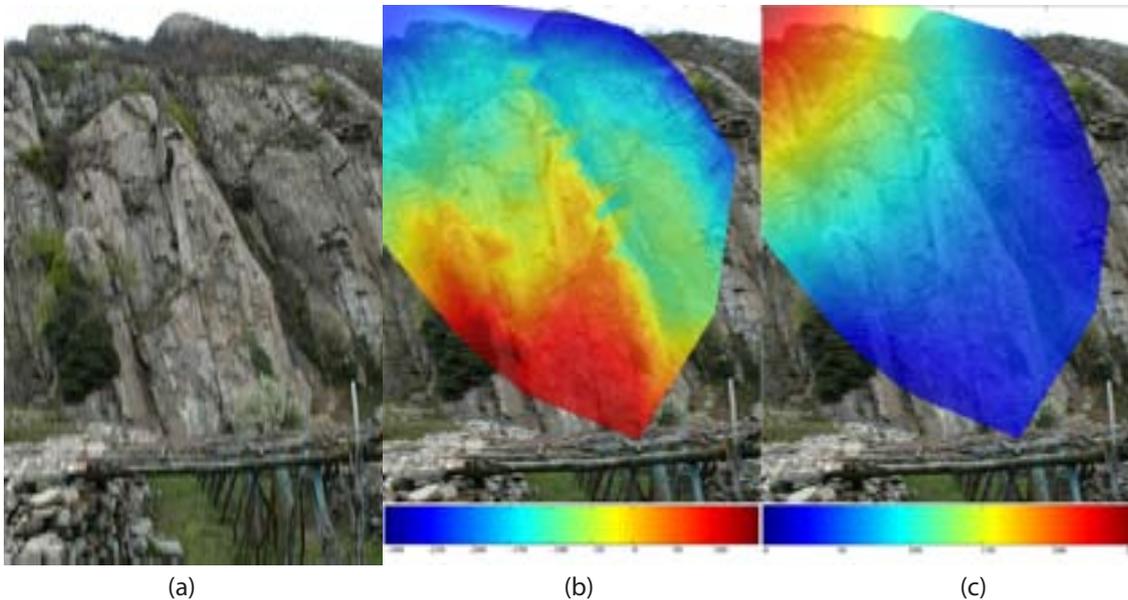


Figura 1.10: Rappresentazione mediante mappe di colore di un campo di parallasse in una sequenza analizzata per mezzo della structure and motion. (a) Una delle immagini della sequenza. (b) Campo di parallasse in direzione x . (c) Campo di parallasse in direzione y . A posteriori si sono determinati i valori medi e di dispersione: $\mu_x = -62.1$ pixel; $\sigma_x = 94.0$ pixel; $\mu_y = 80.7$ pixel; $\sigma_y = 55.3$ pixel.

considerata di tipo casuale, ed una componente sistematica, strettamente dipendente dal moto della camera (si veda figura 1.10); la funzione gaussiana descrive quindi, seppur in maniera approssimativa, la distribuzione della componente casuale: è evidente che, dipendendo dalla geometria dell'oggetto (in questa fase incognita) tale componente non risulta praticamente mai affine ad una distribuzione normale: ciononostante la funzione gaussiana permette di ottenere risultati accettabili.

Pertanto, per quanto riguarda la componente sistematica, è possibile considerare una traslazione della variabile casuale (anche variabile da zona a zona nel fotogramma) in modo da limitare l'influenza del moto della camera.

Analiticamente otteniamo, per quanto riguarda la valutazione della matrice di parallasse:

$$D_{ij} = e^{-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})^T C^{-1}(\mathbf{x} - \boldsymbol{\mu})} \cong e^{-\frac{(x_i + y_j)^2}{2}} \quad (1.22)$$

dove il vettore $\mathbf{x} = x_i - x_j$ rappresenta il vettore parallasse fra il punto i -esimo sulla prima immagine e il punto j -esimo sulla seconda, il vettore $\boldsymbol{\mu}$ rappresenta il valore atteso di parallasse e la matrice C la dispersione attesa del campo di parallassi e dove, infine, nel secondo membro (approssimato) si è posto:

$$X_{ij} = \frac{(x_i - x_j - \mu_x)^2}{\sigma_x^2} \quad e \quad Y_{ij} = \frac{(y_i - y_j - \mu_y)^2}{\sigma_y^2} \quad (1.23)$$

Si noti che, sebbene sarebbe più corretto considerare nell'espressione (1.22) anche il termine misto, ovvero la covarianza fra x ed y , solitamente fornisce risultati eccellenti anche il solo utilizzo di una finestra ellittica orientata secondo i due assi principali di riferimento.

Per quanto riguarda la matrice di correlazione dei toni di grigio definito:

$$C_{ij} = \frac{\sum_k \sum_l (A_{kj} - \bar{A}) (B_{kl} - \bar{B})}{\sqrt{\sum_k \sum_l (A_{kj} - \bar{A})^2 \sum_k \sum_l (B_{kl} - \bar{B})^2}} \quad (1.24)$$

il valore di correlazione dei toni di grigio fra una finestra A centrata sull' i -esimo elemento del primo fotogramma (con media \bar{A}) e la finestra B centrata sul j -esimo elemento della seconda (con media \bar{B}), si ottiene:

$$S_{ij} = e^{-\frac{(C_{ij} - 1)^2}{2\gamma^2}} \quad (1.25)$$

E' dunque possibile definire la matrice di similitudine globale calcolando il prodotto di Hadamard (ovvero il prodotto elemento per elemento) delle due matrici (1.22) e della (1.24):

$$M_{ij} = S_{ij} \cdot D_{ij} \quad (1.26)$$

in cui il parametro γ della (1.25) permette di variare il peso della correlazione in rapporto alla misura di

parallasse.

Nell'implementazione realizzata, dunque, i valori di parallasse vengono stimati in maniera più coerente alla geometria di presa della sequenza. Ad esempio, in una sequenza di rilievi stereo nel Mobile Mapping (si veda §1.8.4) i parametri dell'area di ricerca, ovvero i due valori medi μ_x e μ_y e le due dispersioni σ_x e σ_y , possono essere definiti a priori note la distanza fra le due camere e la velocità di scatto dei fotogrammi, o possono essere interattivamente stimati durante il processo, utilizzando per la previsione i valori di parallasse evidenziati dagli accoppiamenti corretti dei fotogrammi precedenti, ottenendo un'ulteriore diminuzione degli accoppiamenti sbagliati (si veda tabella 1.1).

	Corrispondenze estratte	Corrispondenze filtrate (epipolare)	Corrispondenze dopo matching guidato
Metodo circolare (Scott, Pilu)	56	15	54
Metodo ellittico sviluppato	291	99	288
Incremento percentuale	519%	660%	533%

Tabella 1.1: Confronto fra i risultati ottenuti con l'applicazione del metodo proposto da [Pilu 1997] e il metodo sviluppato in cui si considera una finestra di ricerca ellittica, in diverse fasi della ricerca di omologhi. Si nota che, con il metodo proposto, non solo si ottengono molte più corrispondenze, ma anche il grado di correttezza è maggiore.

E' importante evidenziare che, la necessità di modellare correttamente il campo di parallassi in termini di componenti di traslazione è di fondamentale importanza in quanto, l'approccio numerico descritto non coincide esattamente con un metodo procustiano generalizzato [Crosilla 1983], nel qual caso, tale componente sistematica non avrebbe effetti nella stima di similitudine. Considerare nella (1.26) anche la componente legata alla similitudine dei toni di grigio comporta che il modulo del vettore (sistematico) di traslazione abbia influenza su quale dei due aspetti (correlazione dei g.v. e distribuzione dei punti) debba avere maggiore importanza.

Se si hanno informazioni a priori sul moto della camera, questo può aiutare a predire sulle immagini successive la forma del campo di parallassi. Ad esempio, sempre per una applicazione di S&M nel Mobile Mapping, si sta sviluppando una procedura che, tramite tecniche quali il flusso ottico, individua nella sequenza situazioni in cui il campo di parallasse subisce una brusca modifica, come accade ad esempio nell'ingresso in curva del veicolo, correggendo opportunamente i quattro parametri in quelle circostanze.

Per quanto concerne la stima del valore di correlazione un ulteriore miglioramento delle prestazioni potrebbe essere rappresentato dalla sostituzione, nella valutazione della similitudine dei toni di grigio, dell'algoritmo di correlazione semplice con il Least Squares Matching (si veda in proposito il capitolo seguente), che permetta una più corretta modellazione delle variazioni prospettiche dei punti nelle diverse immagini. Tale opzione, al momento, è stata accantonata per il drastico aumento della mole di calcolo (ricordiamo che ciascun punto estratto nella prima immagine viene confrontato con ciascun punto nella seconda) sebbene

possa fornire un deciso aumento delle prestazioni in quelle circostanze in cui il blocco fotogrammetrico presenti basi di presa particolarmente ampie e la correlazione semplice non sia in grado di fornire una stima corretta del grado di similitudine.

1.3.2. Descrittore di Lowe

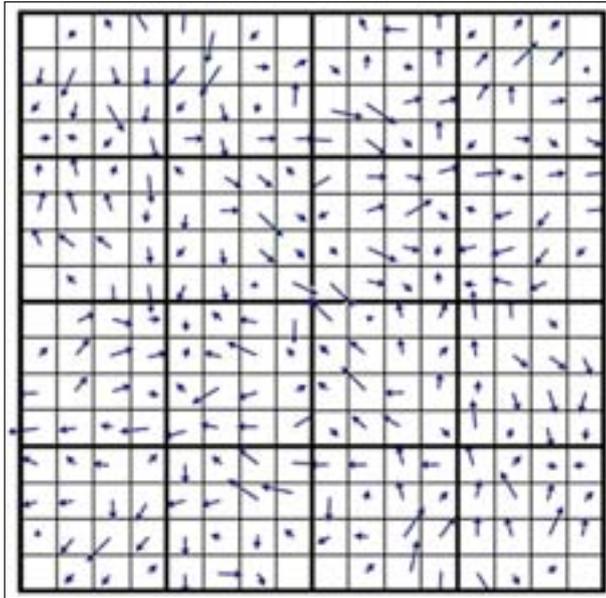


Figura 1.11: Mappa delle orientazioni ed intensità nell'intorno di un punto estratto per mezzo dell'operatore di Lowe.

Si è visto nel precedente paragrafo come l'operatore di Lowe, a causa della sua sostanziale invarianza in presenza di variazioni di scala, sia estremamente efficiente laddove queste non possano essere considerate trascurabili fra fotogrammi successivi della sequenza. In [Lowe 2004] viene definito un descrittore, univoco per ciascuna feature estratta, per facilitare eventuali operazioni di matching.

Innanzitutto per ciascuna feature viene determinata nel piano xy del fotogramma l'orientazione del gradiente dei toni di grigio per mezzo della (1.21). La direzione calcolata, alla quale sono riferiti i valori del descrittore, fa sì che tale descrittore risulti invariante rispetto a trasformazioni di rotazione dell'immagine.

Per determinare tale orientazione viene scelta l'immagine con il fattore di scala prossimo al valore di σ calcolato per il punto considerato. Su tale immagine viene determinata la norma del gradiente per mezzo di:

$$m(x, y) = \sqrt{(L(x+1, y) - L(x-1, y))^2 + (L(x, y+1) - L(x, y-1))^2} \quad (1.27)$$

e la sua direzione per mezzo di:

$$\theta(x, y) = \tan^{-1}((L(x, y+1) - L(x, y-1)) / (L(x+1, y) - L(x-1, y))) \quad (1.28)$$

in cui $L(x, y)$ rappresenta l'informazione contenuta nell'immagine differenza di gaussiane alla scala scelta. Per ogni feature viene quindi calcolato, prendendo in considerazione le informazioni in una regione locale

attorno al keypoint, un istogramma degli orientamenti che, codificando le caratteristiche locali attorno al punto determinato, ne fornisce una descrizione univoca: l'intervallo $[0-2\pi]$ viene suddiviso in n diverse orientazioni (nella nostra implementazione $n = 36$). Per ciascun punto nell'intorno della feature inserita nell'istogramma vengono calcolate le due grandezze (1.27) e (1.28) come mostrato in figura 1.11.

Inoltre, considerato che risultano più significative le informazioni estratte dai punti nelle immediate vicinanze della feature i valori di intensità vengono pesati per mezzo di una funzione circolare gaussiana (come quella in (1.16)). Per semplificare le successive operazioni di riconoscimento, infine, le direzioni locali di ciascun gradiente vengono riferite alla direzione calcolata precedentemente.

A questo punto, resta da definire analiticamente un descrittore che sintetizzi le informazioni estratte nel punto di interesse e nel suo intorno. Lo studio della visione umana [Edelman 1997] mostra che i complessi neuronali nella corteccia visiva primaria rispondono a gradienti aventi una particolare orientazione e frequenza, mentre la posizione all'interno della retina non deve essere necessariamente predeterminata, ma può presentare delle variazioni limitate di posizione in cui trovarsi. In altre parole i ricettori della retina, agendo all'unisono, forniscono al cervello valori mediati delle informazioni riguardanti gli oggetti: per il riconoscimento piccole variazioni di posizione non modificano drasticamente tali valori. Sono stati presentati risultati in cui un approccio simile, nel campo della visione artificiale, permette di triplicare i riconoscimenti corretti rispetto a tecniche di matching tradizionali (come ad esempio la correlazione semplice) quando, a causa della tridimensionalità dell'oggetto rappresentato, intervengono deformazioni prospettiche non trascurabili; al tempo stesso, il considerare l'andamento del gradiente, fornisce risultati molto più stabili ed affidabili.

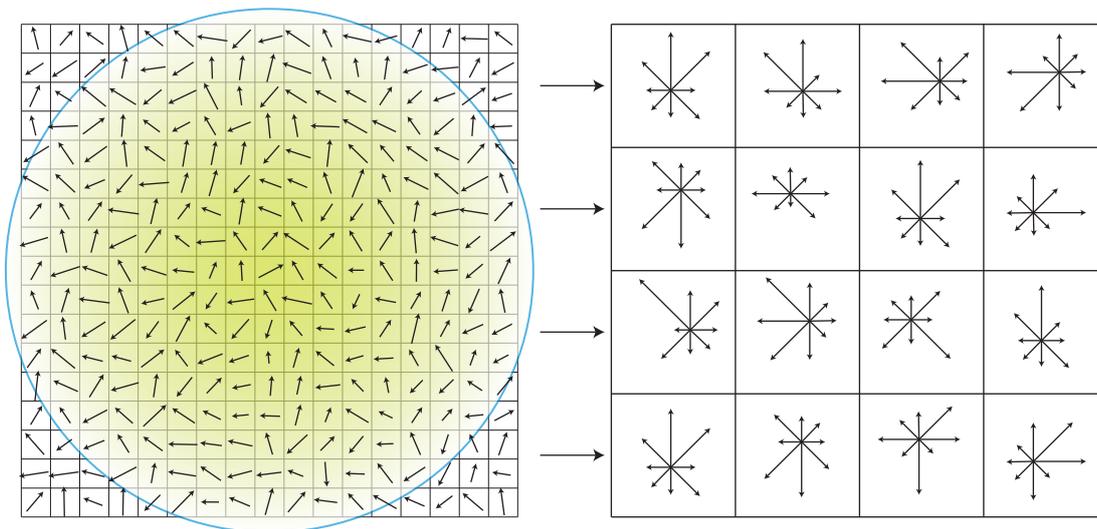


Figura 1.12: Dai gradienti nell'intorno del punto (a sinistra) si determinano, pesando i valori con una funzione gaussiana circolare, gli istogrammi relativi a ciascuna macro-area (a destra), che andranno a fornire i dati del descrittore.

La soluzione adottata definisce, nell'intorno del punto scelto, una serie di regioni (nel nostro caso 4x4 regioni quadrate) per ciascuna delle quali si calcola l'istogramma per i punti appartenenti interni alla regione stessa, come mostrato in figura 1.12. Stabilito un intervallo di discretizzazione per tali istogrammi (ad esempio 8 intervalli) si calcola l'intensità dei gradienti in ciascun intervallo.

La struttura dati in cui viene memorizzato il descrittore è un semplice vettore che contiene tutte le informazioni su ciascun istogramma, la lunghezza dei vettori rappresentati in figura 1.12: il descrittore è quindi rappresentato da un punto in uno spazio n -dimensionale (con riferimento a figura 1.6, $n = 128$). La scelta del numero di dimensioni di tale spazio dipende da un lato dall'onere computazionale che si è disposti a compiere, dall'altro dal grado di discriminazione (ovvero al grado di univocità) che si vuole raggiungere per descrivere il punto. La scelta dei parametri illustrata finora rappresenta, anche in riferimento a quanto dedotto dalla letteratura, il miglior compromesso fra questi due aspetti.

Per determinare gli accoppiamenti preliminari fra feature estratte su fotogrammi differenti ciascun punto sulla prima immagine viene confrontato per mezzo del suo descrittore con i punti sull'altra. Il confronto può avvenire nello spazio n -dimensionale del descrittore andando a calcolare l'angolo formato dai due vettori individuati dal punto i -esimo nella prima immagine e il punto j -esimo nella seconda:

$$\theta = \cos^{-1}(\mathbf{v}_i \cdot \mathbf{v}_j) \quad (1.29)$$

Qualora tale angolo risulti minore di una certa soglia, o più semplicemente il minore relativamente al punto i -esimo a fronte di tutti i possibili accoppiamenti con i punti j -esimi, tale accoppiamento, con buona probabilità si dimostra corretto.

Anche in presenza di un numero molto elevato di features estratte su ogni singolo fotogramma, il metodo è, dal punto di vista computazionale, estremamente efficiente. Inoltre, facendo affidamento sul fatto che l'operatore di Lowe (così come il suo descrittore) si dimostra invariante rispetto a variazioni di scala, il numero di corrispondenze corrette determinabili è molto maggiore rispetto al metodo descritto nel precedente paragrafo, soprattutto in quelle circostanze dove, proprio a causa di basi di presa non trascurabili, la correlazione semplice non fornisce buoni risultati. Il matching procustiano si dimostra più robusto sotto altri punti di vista, in quanto considera gli accoppiamenti in termini globali, mantenendo i rapporti geometrici e topologici fra i punti distribuiti sui diversi fotogrammi mentre il metodo proposto da Lowe è intrinsecamente seriale: ogni feature sul primo fotogramma viene associata alla feature migliore individuata sul secondo, indipendentemente dai risultati emersi dagli altri punti. Per risolvere le ambiguità è però sufficiente (ed al momento in fase di sperimentazione) definire una matrice di similitudine fra punti analoga alla (1.25) in funzione dell'angolo di somiglianza θ di (1.29) e inserire quindi il descrittore di Lowe nelle procedure di matching procustiano. Si pesano così i risultati ottenuti per mezzo del descrittore in funzione della reciproca distribuzione dei punti sui diversi fotogrammi, ottenendo un criterio globale di accoppiamento.

1.3.3. Ottimizzazione dell'estrazione di features

Per ottenere la massima efficienza nella fase di feature matching spesso non è sufficiente avere una buona implementazione degli algoritmi di accoppiamento dei punti: è altrettanto importante limitare le ambiguità nella fase di ricerca e filtraggio degli abbinamenti preliminari. Soprattutto nel caso di scene con scarsa tessitura la correlazione perde capacità di discriminazione: in tali casi è bene impostare il parametro λ in (1.25) ad un valore abbastanza alto, per dare maggior risalto alle informazioni relative alla parallassi fra i punti. Se, però, i punti estratti dall'operatore sono molto vicini il comportamento dell'algoritmo può portare a risultati inattesi: ad esempio, con un moto della camera nel senso della profondità, la scena inquadrata presenta forti variazioni in termini di parallassi su una porzione o su tutto il fotogramma; in tali casi i parametri relativi alle parallassi presentano una dispersione maggiore e la conoscenza del campo può non essere sufficientemente corretta per l'individuazione delle coppie omologhe. Punti moderatamente vicini possono quindi creare situazioni di ambiguità e rendere la successiva fase di filtraggio e stima della geometria di presa instabile.

Per ovviare a queste difficoltà si impongono criteri geometrici per evitare l'addensamento, in fase di estrazione, di punti di interesse: si va inizialmente a ricompilare la lista di punti estratti in funzione del valore di risposta in ordine decrescente (tale valore può essere quello di (1.9) per Harris, quello di (1.13) per Foerstner e quello derivato dalla (1.21) per Lowe); si va poi ad analizzare la lista ottenuta, scartando quei punti che risultano troppo vicini ad altri che li precedono: essendo infatti la lista ordinata in modo decrescente, punti successivi sono caratterizzati da risposte più basse e risultano riconoscibili con minor probabilità sui fotogrammi successivi.

Bisogna evidenziare che tale criterio, in alcuni casi, si dimostra controproducente ai fini della stabilità dell'intera procedura: se molti punti vicini presentano valori di risposta simili può infatti accadere che di volta in volta, sui diversi fotogrammi, il loro ordine sia differente; vengono allora eliminate un gran numero di corrispondenze che con buona probabilità sarebbero state correttamente indivi-

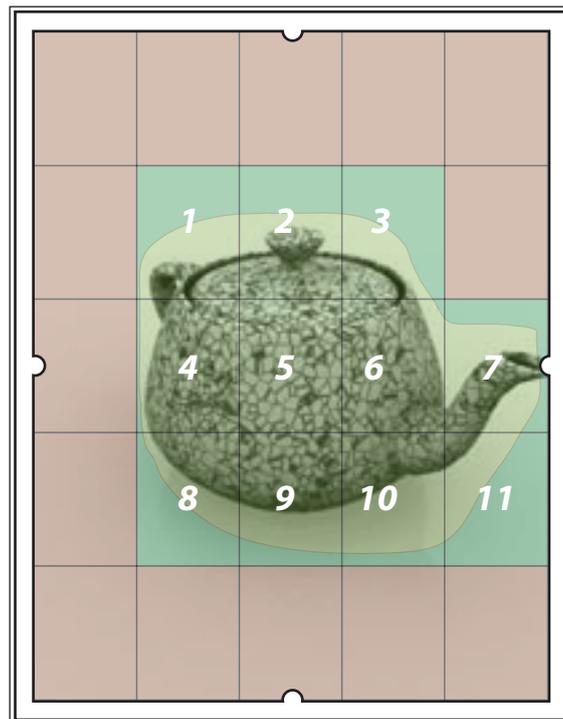


Figura 1.13: Per garantire una uniforme distribuzione di feature vengono individuate zone rettangolari in cui la densità di feature deve risultare approssimativamente uguale; per mezzo di una maschera si possono selezionare quali rettangoli utilizzare (quelli numerate) escludendo quelli in cui non ci si aspetta che vengano determinate features o che rappresentano porzioni di scena non interessanti.

duate dagli algoritmi di feature matching. Ciononostante, in molte altre circostanze il metodo si dimostra necessario; con oggetti dalle forme irregolari (come ad esempio le foglie di un albero) su uno sfondo ben contrastato, l'operatore di interesse etichetta la maggior parte dei punti sul contorno di tali forme come significativi, generando una densità di feature molto elevata; in queste condizioni gli algoritmi di feature matching producono spesso risultati errati.

Oltre alla caratterizzazione locale dei punti, per favorire una stima più robusta della geometria di presa nelle successive fasi di elaborazione si cerca di ottenere una copertura del fotogramma estesa ed uniforme. Questo non solo migliora stima della geometria di presa, ma garantisce che tutte le parti dell'oggetto ripreso vi contribuiscano. Soprattutto nell'orientamento relativo accade che, con una copertura locale dell'oggetto, i parametri stimati danno luogo a residui in tolleranza per i punti appartenenti a tale zona, ma inaccettabili al di fuori.

In termini più analitici, una disposizione concentrata di punti usati per la stima dei parametri può portare ad avere forti componenti di correlazione fra tali parametri: in tal caso non si è certi che la soluzione trovata, seppur verifichi test statistici sui residui, sia effettivamente la soluzione corretta.

Per superare il problema si suddivide il fotogramma in zone rettangolari di dimensioni prefissate (vedi

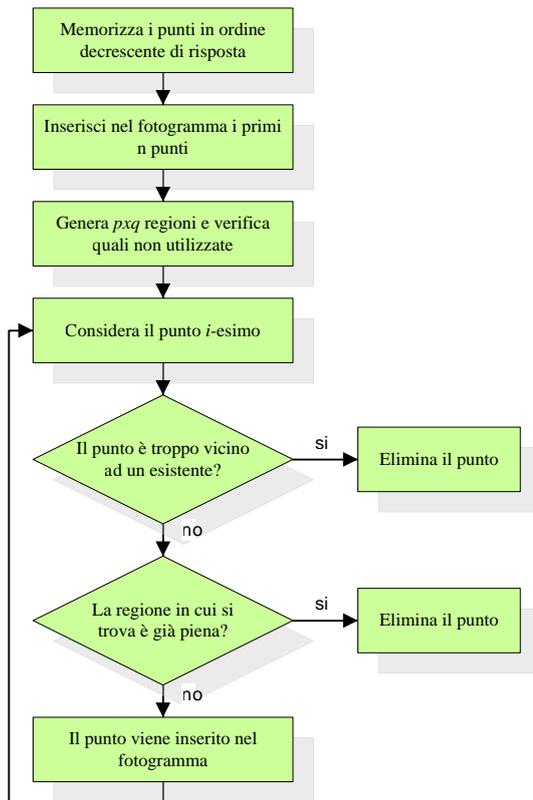


Figura 1.14: Diagramma di flusso per l'ottimizzazione della localizzazione dei punti di interesse.

figura 1.13). Ordinati come in precedenza i punti in senso decrescente di risposta dell'operatore, si inseriscono in modo permanente i primi n punti della lista indipendentemente dalla loro posizione sul formato; si prescrive infine che in ogni zona vi sia, con una certa tolleranza, lo stesso numero di features. Se tale condizione non è soddisfatta (ad esempio perché in alcune zone si ha una risposta molto maggiore che in altre) semplicemente non si vanno più ad aggiungere punti in quelle regioni che appaiono "sovraffollate" rispetto alle altre. Come nel caso precedente, bisogna porre attenzione a non utilizzare soglie di tolleranza troppo ristrette perché ciò può condurre ad avere una distribuzione uniforme di punti poco significativi in quanto la maggior parte dei punti a risposta più elevata sono stati scartati per favorire il riempimento delle zone con minor contenuto informativo.

A questo scopo è sufficiente mascherare (vedi figura 1.13) il fotogramma, considerando solo le zone che, scelte da un operatore o da un opportuno filtro che vada ad esempio ad analizzare le deviazioni locali dei

toni di grigio, si dimostrino utili per la stima della geometria di presa. L'intero processo è riassunto dallo schema di figura 1.14: esso va inserito subito dopo il calcolo, per mezzo dell'operatore di interesse, dei valori di risposta dei singoli punti. Per rendere il procedimento ulteriormente automatizzato e garantire una scelta coerente dei parametri di tolleranza sono allo studio alcuni algoritmi basati sulla logica fuzzy (vd. appendice A) che, in funzione del campo di risposta, delle informazioni sulla varianza locale dei toni di grigio e di altri parametri legati alla distribuzione dei punti di interesse trovati, sia in grado di fornire una distribuzione ottimale. Al momento tuttavia la ricerca è ancora in una fase iniziale e non si è in grado di dire se tale percorso può fornire un reale miglioramento delle prestazioni dell'algoritmo.

1.4. Collinearità e geometria proiettiva

Qualunque siano le procedure di feature matching impiegate, parte degli accoppiamenti preliminari sono errati e devono essere messi in evidenza imponendo che soddisfino un modello matematico che descrive la formazione dell'immagine (la camera) e la geometria di presa.

A questo scopo, la formulazione tradizionale delle equazioni di collinearità e di complanarità non è particolarmente indicata, in quanto, lo si ricorda, non si dispone di parametri iniziali e la natura non lineare di tali equazioni rende difficoltosa la soluzione del problema. Facendo invece ricorso alla geometria proiettiva (per alcuni cenni introduttivi su tale argomento si rimanda all'appendice A) tali problemi possono essere riformulati con approcci lineari. Consideriamo inizialmente un sistema di riferimento terreno con origine nel punto principale della camera e abbia l'asse z coincidente con l'asse ottico. Da semplici considerazioni di similitudine (vd. figura 1.15) si ha:

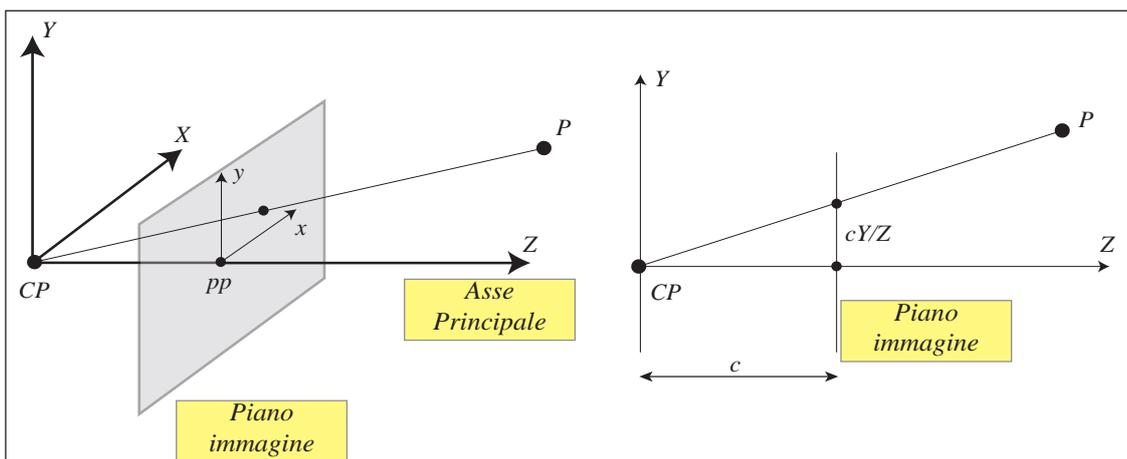


Figura 1.15: Sistema di riferimento terreno (XYZ) e immagine (xy) (a sinistra); rappresentazione della proiezione sul piano immagine del punto P (a destra).

$$\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} c \frac{X}{Z} \\ c \frac{Y}{Z} \end{bmatrix} \quad (1.30)$$

ovvero in coordinate omogenee:

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} cX \\ cY \\ Z \end{bmatrix} = \begin{bmatrix} c & 0 & 0 \\ 0 & c & 0 \\ 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} \quad (1.31)$$

Tenendo conto che il punto principale può non coincidere con l'origine del sistema di riferimento immagine

$$\begin{aligned} (X \ Y \ Z)^T &\mapsto (cX/Z + p_x \ cY/Z + p_y) \\ \Rightarrow \begin{bmatrix} cX + Zp_x \\ cY + Zp_y \\ Z \end{bmatrix} &= \begin{bmatrix} c & 0 & p_x \\ 0 & c & p_y \\ 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} \end{aligned} \quad (1.32)$$

si vede che la matrice

$$K = \begin{bmatrix} c & 0 & p_x \\ 0 & c & p_y \\ 0 & 0 & 1 \end{bmatrix} \quad (1.33)$$

codifica tutti i parametri di orientamento interno e prende pertanto il nome di matrice di calibrazione. Tale matrice può assumere forma più complessa (pur mantenendosi triangolare superiore) in modo da considerare trasformazioni più generiche dello spazio immagine:

$$K = \begin{bmatrix} c & s & p_x \\ 0 & r \cdot c & p_y \\ 0 & 0 & 1 \end{bmatrix} \quad (1.34)$$

In tal caso il parametro r (*aspect ratio*) permette di utilizzare scale dimensionali differenti per la coordinata x e y , ovvero casi in cui la forma del pixel è rettangolare, mentre s (coefficiente di taglio o *skew*) permette di considerare un sistema di riferimento con assi coordinati non ortogonali: è facile derivare che

$$s = \tan \theta \quad (1.35)$$

in cui θ è l'angolo formato dai due assi coordinati.

Considerando inoltre che sistema immagine e terreno solitamente non coincidono:

$$\tilde{X}_{cam} = R (\tilde{X} - \tilde{C}) \quad (1.36)$$

dove il simbolo \sim rappresenta il fatto che le coordinate sono espresse nel sistema di riferimento terreno in forma non omogenea; si ottiene quindi:

$$\mathbf{X}_{cam} = \begin{bmatrix} R & -R\tilde{C} \\ 0 & 1 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} = \begin{bmatrix} R & -R\tilde{C} \\ 0 & 1 \end{bmatrix} \mathbf{X} \quad (1.37)$$

da cui

$$\mathbf{x} = KR \begin{bmatrix} I & -\tilde{C} \end{bmatrix} \mathbf{X} \quad (1.38)$$

E' allora possibile andare a definire una matrice $P_{3 \times 4}$ che permette la proiezione del punto terreno \mathbf{X} sul punto immagine \mathbf{x} :

$$P = KR \begin{bmatrix} I & -\tilde{C} \end{bmatrix} = K \begin{bmatrix} R & -R\tilde{C} \end{bmatrix} \quad (1.39)$$

$$\mathbf{x} = P\mathbf{X} \quad (1.40)$$

E' facile, sviluppando i termini che compongono P e ricavando le coordinate non omogenee nello spazio immagine, ottenere una formulazione in tutto identica alle ben note equazioni di collinearità.

Al tempo stesso la matrice camera P permette, in maniera estremamente sintetica, di codificare tutti gli elementi necessari alla proiezione di punti (coordinate del centro di presa, angoli di assetto del fotogramma, parametri di orientamento interno). Tali valori possono essere estratti dalla matrice con semplici operazioni algebriche (si veda in proposito il relativo paragrafo in appendice).

Inoltre la linearità nello spazio omogeneo della (1.40) permette un più diretto approccio a tutte le problematiche legate alla stima dei parametri incogniti della camera: si consideri, ad esempio, la stima dei parametri di orientamento interno ed esterno, noti alcuni punti nel sistema terreno e le corrispondenti immagini sul fotogramma; come ben noto tale problema di resezione è non lineare: per effettuare una stima a minimi quadrati è necessario avere parametri approssimati delle incognite. Sono stati proposti numerosi algoritmi di soluzione diretta, alcuni con l'uso dei quaternioni (si veda appendice A.3). L'esperienza mostra che non sempre essi garantiscono buoni risultati e, se non si ha una buona conoscenza dei parametri di orientamento interno, possono portare a risultati inattesi.

Al contrario, mediante l'impostazione proiettiva, detti a_{ij} gli elementi incogniti della matrice camera, P^i i vettori riga che la compongono, \mathbf{X}_i le coordinate omogenee nel sistema terreno dell' i -esimo punto, $\mathbf{x}_i = (x_i, y_i, w_i)$ le corrispondenti coordinate nel sistema fotogramma è possibile scrivere la (1.40) come:

$$\mathbf{x} = P\mathbf{X} \Rightarrow \mathbf{x} \wedge P\mathbf{X} = 0 \quad (1.41)$$

da cui

$$\mathbf{x}_i \wedge P\mathbf{X}_i = \mathbf{x}_i \wedge \begin{bmatrix} P^{1T} \mathbf{X}_i \\ P^{2T} \mathbf{X}_i \\ P^{3T} \mathbf{X}_i \end{bmatrix} = \begin{bmatrix} y_i P^{3T} \mathbf{X}_i - w_i P^{2T} \mathbf{X}_i \\ w_i P^{1T} \mathbf{X}_i - x_i P^{3T} \mathbf{X}_i \\ x_i P^{2T} \mathbf{X}_i - y_i P^{1T} \mathbf{X}_i \end{bmatrix} = 0 \quad (1.42)$$

Ricavando dunque gli elementi incogniti della matrice camera nella (1.42) si giunge a

$$\begin{vmatrix} 0^T & -w_i X_i^T & y_i X_i^T \\ w_i X_i^T & 0^T & -x_i X_i^T \\ -y_i X_i^T & x_i X_i^T & 0^T \end{vmatrix} \begin{vmatrix} P^1 \\ P^2 \\ P^3 \end{vmatrix} \quad (1.43)$$

ovvero ad un sistema omogeneo facilmente risolvibile in maniera diretta; per contro non è possibile inserire nel processo di stima informazioni a priori su alcuni dei parametri. Ad esempio, se è noto l'orientamento interno occorrerebbe esprimere i termini della matrice camera in funzione di essi; si torna però ad avere equazioni risolventi non lineari. Il metodo, sostanzialmente equivalente alla Direct Linear Transformation (DLT, [Abdel-Aziz 1971]), è dunque vantaggioso solo quando i parametri di orientamento interno non sono noti (camera non calibrata) o per ottenere una stima approssimata, da impiegare per la linearizzazione del problema a minimi quadrati.

In maniera analoga si giunge ad una formulazione lineare per il calcolo dell'intersezione (triangolazione) fra due o più raggi omologhi. Considerando per semplicità il caso stereoscopico, detti \mathbf{x} ed \mathbf{x}' due punti omologhi si ha:

$$\mathbf{x} = P\mathbf{X}, \quad \mathbf{x}' = P'\mathbf{X} \quad (1.44)$$

Utilizzando nuovamente il prodotto vettoriale come in (1.42) e isolando la componente vettoriale \mathbf{X} si ottiene:

$$\begin{vmatrix} xP^{3T} - P^{1T} \\ yP^{3T} - P^{2T} \\ x'P^{3T} - P^{1T} \\ y'P^{3T} - P^{2T} \end{vmatrix} \mathbf{X} = 0 \quad (1.45)$$

in cui si è tralasciato, per ciascuna immagine, la terza equazione, linearmente dipendente dalle altre due. L'estensione ad un numero di immagini superiore a due è immediato.

Dalle verifiche numeriche effettuate è emerso che il metodo risulta sufficientemente stabile, sebbene, minimizzando una grandezza algebrica anziché geometrica, porta a coordinate oggetto non identiche a quelle ottenute dalla soluzione a minimi quadrati delle equazioni di collinearità: il metodo dunque è preferibilmente da utilizzare per una stima dei valori approssimati da assegnare alle coordinate terreno dei punti.

Inoltre, in condizioni di geometria di presa sfavorevole (basi piccole, angoli di intersezione ridotti etc.) [Stewénius 2005] è consigliabile seguire le indicazioni fornite in [Hartley 1997a]: in tali condizioni il sistema (1.45) può essere numericamente instabile a causa del differente peso che ciascuna equazione ha nella stima; tali pesi non derivano da una scelta dell'utente o da considerazioni statistiche, ma dal fatto che i valori presenti nella matrice di (1.45) possono essere, a seconda della loro grandezza, più o meno influenzati da variazioni del vettore \mathbf{X} .

Per far fronte a tale problema, che in certi casi può rendere impossibile la risoluzione della S&M, si

sono implementate due soluzioni. La prima introduce un metodo iterativo che, al termine delle iterazioni, porta a residui di riproiezione sostanzialmente identici su ciascun fotogramma: seppur tale criterio non sia geometricamente corretto quando il numero di immagini è superiore a due o quando si desiderano pesare differentemente le osservazioni, i risultati sono accettabili per fornire coordinate approssimate affidabili per la stima a minimi quadrati. Si risolve il sistema (1.45) una prima volta per ottenere un valore iniziale di \mathbf{X} ; per mezzo della (1.40) si riproietta tale punto sui vari fotogrammi calcolando i residui con i punti originari; qualora non fossero simili (con geometrie instabili si possono avere residui anche dell'ordine di $10 \div 20$ pixel su un fotogramma e sostanzialmente nulli sull'altro) si risolve nuovamente il sistema (1.45) moltiplicando le equazioni relative a ciascun fotogramma per il valore

$$w^i = \frac{1}{r} \quad \text{con } r = \|\mathbf{x}^i - P_i \mathbf{X}\| \quad (1.46)$$

Tale coefficiente ha proprio il compito di ridurre l'importanza delle equazioni per le quali il sistema, all'iterazione precedente, ha minimizzato i residui di riproiezione. Nella maggior parte dei casi il metodo illustrato è in grado di andare a convergenza e di fornire dunque un risultato sufficientemente affidabile. Bisogna inoltre considerare che, pur utilizzando un metodo iterativo, l'algoritmo è in grado di fornire ottime prestazioni in quanto le due operazioni di risoluzione della (1.45) e riproiezione per mezzo della (1.40) hanno oneri computazionali minimi.

In alternativa si può utilizzare il metodo sviluppato in [Hartley 1997a], testo al quale si rimanda per una descrizione approfondita in quanto nella nostra implementazione si sono seguite esattamente le stesse linee di sviluppo. Come osservato in [Beardsley 1996] i cosiddetti metodi di triangolazione con il punto mediano, ovvero quei metodi che determinano l'intersezione fra due raggi omologhi come il punto medio del segmento di minima distanza fra di essi (si veda [Beardsley 1996] o [Cooper 1996]) non forniscono criteri di ottimalità in spazi proiettivi generici: in tali spazi la metrica euclidea non ha lo stesso significato che in uno spazio reale, così come del resto il concetto di perpendicolarità; si arriva quindi alla definizione di un punto di intersezione che, una volta trasformato in uno spazio reale euclideo, non coincide con il punto di mezzo che rappresenterebbe la soluzione corretta del problema. Il metodo proposto da [Hartley 1997a] utilizza la matrice fondamentale (che codifica la geometria relativa fra due fotogrammi) per ottenere una stima parametrica del residuo di riproiezione di un generico punto. La funzione obiettivo che descrive il residuo totale, grazie al vincolo epipolare, può essere scritta in funzione di un solo parametro t ; determinando i minimi locali di tale funzione e scegliendo quello minore, si perviene ad una soluzione geometricamente corretta.

Nel caso in cui il numero di fotogrammi, ovvero di punti immagine, utilizzato per la stima sia superiore a due, sorge il problema di poter estendere il legame epipolare ad un numero di immagini superiore; dal momento che tale estensione non è banale e conduce ad una complessità numerica indesiderabile, è preferibile applicare il metodo a solo due dei fotogrammi presenti, scegliendoli secondo criteri geometrici (ad esempio

quelli che presentano una base di presa maggiore o danno luogo ad angoli di intersezione più prossimi a 90°); una volta ottenuta una posizione approssimata per il punto oggetto, come nel caso precedente, si può considerare il contributo di tutti i fotogrammi per mezzo di una stima a minimi quadrati.

1.5. Geometria epipolare

Di seguito vengono illustrate le tecniche proiettive per stimare la geometria relativa fra due fotogrammi: anche in questo caso si perviene ad una formulazione lineare, con metodi di risoluzione più semplici e stabili. Gli studi su tale argomento risalgono, in fotogrammetria a [VonSanden 1908] sebbene, per avere un approccio sistematico al problema, si debba attendere gli anni ottanta in cui, in computer vision si comincia a studiare le proprietà della matrice essenziale [Longuet-Higgins 1981], [Huang 1989], [Horn 1990] e della matrice fondamentale [Faugeras 1992a], [Hartley 1992]. La prospettiva con cui gli argomenti verranno presentati è sempre quella della S&M, ovvero ammettendo che tra le osservazioni impiegate per la stima vi sia una percentuale anche elevata di errori grossolani.

1.5.1. Matrice Fondamentale

Si considerino i due fotogrammi in figura 1.16. Sia π un piano generico dello spazio che contiene il punto terreno X visibile su ambo i fotogrammi in posizione rispettivamente x ed x' , con l'unica condizione che esso non contenga la base di presa b ; le coordinate del punto x e quelle del punto X in un sistema piano solidale a π sono legate da una trasformazione di collinearità; analoga osservazione vale per il punto x' . Componendo le due trasformazioni è possibile ricavare l'omografia H che lega i punti immagine di π sul primo fotogramma con quelli sul secondo.

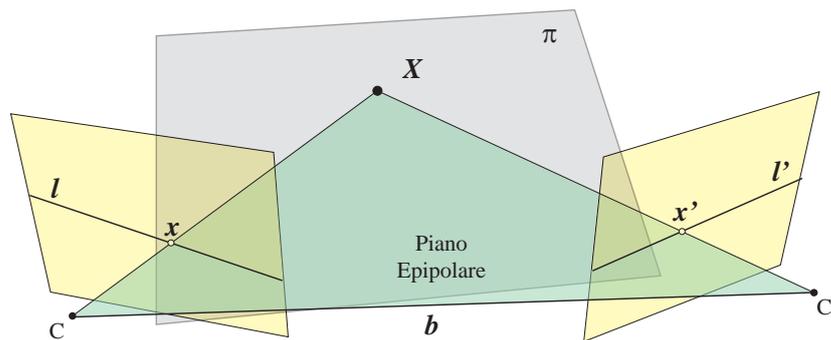


Figura 1.16: Rappresentazione della geometria epipolare fra due fotogrammi (in giallo); è possibile notare i punti immagine e il corrispondente punto oggetto, le corrispondenti linee epipolari, il piano epipolare che le genera, e il piano di trasferimento π .

$$\mathbf{x}' = H\mathbf{x} \quad \mathbf{x} = H^{-1}\mathbf{x}' \quad (1.47)$$

Ulteriore condizione che lega i due punti immagine è il fatto che \mathbf{x}' deve giacere sulla linea epipolare generata da \mathbf{x} e viceversa; siano dunque \mathbf{e} ed \mathbf{e}' i due epipoli relativi al primo e secondo fotogramma, individuati dall'intersezione della retta contenente la base di presa con i due piani immagine (ovvero per la condizione di collinearità dall'immagine del centro di presa di un fotogramma sul piano immagine dell'altro).

E' possibile determinare la linea epipolare corrispondente al punto \mathbf{x} sul secondo fotogramma come la retta passante per l'epipolo \mathbf{e}' ed il punto \mathbf{x}' (si veda Appendice A.1.2):

$$\mathbf{l}' = \mathbf{e}' \wedge \mathbf{x}' = [\mathbf{e}']_{\times} \mathbf{x}' \quad (1.48)$$

da cui, sostituendo la (1.47) si ottiene

$$\mathbf{l}' = [\mathbf{e}']_{\times} H\mathbf{x} = F\mathbf{x} \quad (1.49)$$

in cui la matrice

$$F = [\mathbf{e}']_{\times} H_{\pi} \quad (1.50)$$

è detta matrice fondamentale ed è in grado di definire il legame esistente fra i punti omologhi nei due fotogrammi; è facile osservare che essendo la matrice $[\mathbf{e}']_{\times}$ di rango due anche la matrice fondamentale ha il medesimo rango; tale matrice rappresenta una trasformazione dallo spazio \mathbb{P}^2 dei punti del primo fotogramma all'insieme delle linee epipolari (ovvero del fascio di rette con centro in \mathbf{e}') sul secondo fotogramma; essendo dunque una trasformazione da uno spazio di dimensione 2 ad uno spazio di dimensione 1 è necessario che il rango della matrice che descrive la trasformazione sia pari a due. Tale osservazione risulta di particolare importanza in fase di stima degli elementi della matrice fondamentale: se tale matrice risulta di rango pieno il codominio della trasformazione ha anch'esso dimensione 2; ciò equivale a dire che l'epipolo non è univoco o, in termini più corretti, che non tutte le rette (1.49) passano per uno stesso punto coincidente con l'epipolo del fotogramma. Si noti inoltre che, legando grandezze omogenee, la matrice F è definita a meno di un fattore di proporzionalità.

Nel ricavare la (1.50) è necessario introdurre un piano virtuale π per permettere il trasferimento omografico fra i due fotogrammi; esso è stato preso arbitrariamente e non è necessario per l'esistenza di F : il punto \mathbf{X} non deve necessariamente essere considerato giacente sull'oggetto ma, per dar luogo alla stessa retta epipolare, deve solo giacere sul raggio di proiezione per \mathbf{x} ; è semplice allora scegliere un piano π che dia luogo ad una intersezione \mathbf{X} con qualsiasi raggio del primo fotogramma.

Per verificare se due punti sono omologhi è sufficiente controllare se uno dei due giace sulla linea epipolare corrispondente all'altro: questa condizione può essere espressa per mezzo della (1.49) in maniera estremamente sintetica ricordando che, in geometria proiettiva, un punto giace su una retta se e solo se il loro prodotto scalare è uguale a zero:

$$\mathbf{x}' \in \mathbf{l}' \iff \mathbf{x}' \cdot \mathbf{l}' = \mathbf{x}' F \mathbf{x} = 0 \quad (1.51)$$

Facendo riferimento al formulario in appendice B è facile accorgersi che la conoscenza della matrice fondamentale è equivalente a quella dell'orientamento relativo fra due fotogrammi: da essa si possono ricavare infatti non solo le linee epipolari e la condizione (1.51) ma anche la posizione dei due epipoli e, sotto determinate condizioni (si veda il paragrafo 1.7), anche i valori delle due matrici camera relative ai fotogrammi, dalle quali come noto si può risalire ai parametri di orientamento interno ed esterno.

1.5.2. Metodi di stima della matrice fondamentale

L'utilizzo della matrice fondamentale in problemi connessi alla geometria relativa di presa di due fotogrammi è estremamente semplice dal punto di vista dell'implementazione ed efficiente dal punto di vista computazionale. Nell'ambito della structure and motion il problema principale è tuttavia quello di stimare la geometria di presa (e quindi la matrice fondamentale) da una serie di corrispondenze omologhe.

Sviluppando l'equazione (1.51) indicando con (x, y, w) le coordinate del punto \mathbf{x} , con (x', y', w') le coordinate del punto \mathbf{x}' ed infine con f_{ij} gli elementi della matrice fondamentale, si ottiene:

$$x'x f_{11} + x'y f_{12} + x'w f_{13} + y'x f_{21} + y'y f_{22} + y'w f_{23} + w'x f_{31} + w'y f_{32} + w'w f_{33} = 0 \quad (1.52)$$

Date n corrispondenze si può ottenere un sistema di equazioni omogenee nelle variabili $\mathbf{f} = (f_{11} \dots f_{33})^T$:

$$\begin{vmatrix} x'_1 x_1 & x'_1 y_1 & x'_1 w_1 & y'_1 x_1 & y'_1 y_1 & y'_1 w_1 & w'_1 x_1 & w'_1 y_1 & w'_1 w_1 \\ \vdots & \vdots \\ x'_n x_n & x'_n y_n & x'_n w_n & y'_n x_n & y'_n y_n & y'_n w_n & w'_n x_n & w'_n y_n & w'_n w_n \end{vmatrix} \mathbf{f} = A \mathbf{f} = 0 \quad (1.53)$$

Gli elementi di F sono definiti a meno di una costante di proporzionalità e il rango di F deve risultare pari a due. Tali condizioni possono essere imposte nella stima del vettore \mathbf{f} in (1.53) come

$$\|\mathbf{f}\| = 1; \quad \det(F) = 0 \quad (1.54)$$

Note almeno 7 corrispondenze fra i due fotogrammi, il sistema (1.53), tenendo conto della prima condizione in (1.54), può essere risolto per mezzo della SVD (vedi A.2.3.); detta $A = UDV^T$ la scomposizione per mezzo della SVD, si ottiene una famiglia di soluzioni $\alpha F_1 + (1 - \alpha) F_2$ in cui F_1 ed F_2 sono le matrici corrispondenti ai due generatori dello spazio nullo di A . Parametrizzando tale soluzione con il termine α ed andando a calcolare il determinante è possibile imporre la seconda delle condizioni (1.54), che si traduce in un'equazione di terzo grado nella variabile α ; dando luogo a tre soluzioni (una reale e due complesse o tre reali), la scelta della soluzione corretta può essere fatta (nel caso in cui siano tutte reali) utilizzando un ottavo punto e verificando quale delle tre fornisce risultati corretti.

Al tempo stesso, se si dispone di più di 7 corrispondenze, la soluzione della (1.53) per mezzo della SVD è unicamente determinata: in tal caso, tuttavia, non è garantito che tale soluzione verifichi anche il rango della matrice; come già anticipato, se tale condizione non è verificata, non si è in grado di garantire che tutte le linee epipolari passino per l'epipolo. Per tale motivo si stima ([Longuet-Higgins 1981]) una matrice F^* tale che $\det(F^*) = 0$, imponendo che $\|F - F^*\|$ sia minima. Tale problema può, ancora una volta essere risolto per mezzo della SVD: scomponendo la matrice F in $F = UDV^T$, riordinando i termini a secondo membro in modo che in D i valori singolari siano disposti in ordine decrescente dall'alto verso il basso, è possibile dimostrare con un approccio identico a quello utilizzato in appendice A.2.3 che la matrice cercata è ottenibile dal prodotto

$$F^* = UD^*V^T \quad (1.55)$$

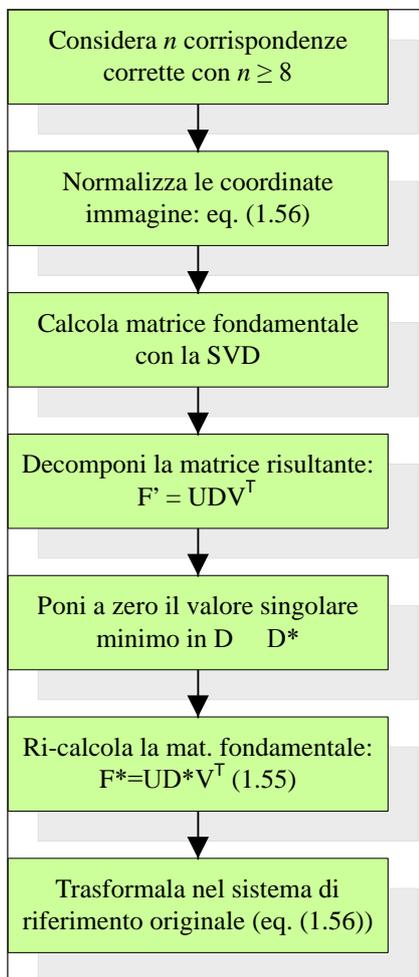


Figura 1.17: Principali operazioni per la stima della matrice fondamentale note almeno 8 corrispondenze.

in cui D^* è la matrice ottenuta da D ponendo a zero il minore elemento singolare.

Come osservato in [Hartley 1997] è preferibile normalizzare le coordinate dei punti prima della stima in modo da ottenere una soluzione più stabile, soprattutto se le osservazioni sono affette da errore: supponendo infatti che i valori w e w' vengano posti uguali ad 1 nella trasformazione da coordinate non omogenee ad omogenee ci si accorge facilmente che le quantità che compongono la matrice A in (1.53), considerate in colonna, possono presentare ordini di grandezza molto differenti; in analogia a quanto accade per la triangolazione può dunque succedere che nella stima di f si prediligano alcuni elementi rispetto ad altri, introducendo un bias inaccettabile nella soluzione del sistema. Le coordinate vengono trasformate su ciascun fotogramma per mezzo delle matrici T e T' in modo che il loro baricentro coincida con l'origine del sistema di riferimento e che il loro momento polare, rispetto a tale punto, risulti uguale a 2; una volta ottenuta la stima di F^* la matrice fondamentale viene ritrasformata nei sistemi di riferimento originari:

$$\begin{aligned} \bar{x}'_i &= T'x'_i & \bar{x}_i &= Tx_i \\ F^* &= T'^T \bar{F}^T T \end{aligned} \quad (1.56)$$

L'insieme delle operazioni descritte è riassunto, per maggior chiarezza, nello schema di figura 1.17.

Infine, qualora sia richiesto il massimo grado di precisione, è

possibile stimare gli elementi della matrice fondamentale per mezzo di un approccio a minimi quadrati. In tal caso, fissato uno degli elementi non nulli della matrice F a 1, si calcola la soluzione cercando di rendere minima la quantità

$$\sum_{i=1}^n d(\mathbf{x}'_i, F\mathbf{x}_i) + d(\mathbf{x}_i, F\mathbf{x}'_i) \quad (1.57)$$

Gli algoritmi risolutivi visti in precedenza possono essere utilizzati per fornire i valori iniziali per la stima.

1.5.3. Stima robusta della matrice fondamentale

Gli accoppiamenti derivanti dal feature matching, come già detto, presentano in percentuale variabile abbinamenti errati. Utilizzare tali corrispondenze nel sistema (1.53) può portare ad errori di stima non trascurabili. Per risolvere il problema si è adottato un approccio basato su Ransac (vd. Appendice A.2.4.).

Come visto nel paragrafo precedente il numero minimo di corrispondenze necessarie per stimare la matrice fondamentale è pari a 7: sebbene con tale numero di corrispondenze sia necessario risolvere una soluzione di terzo grado, è preferibile utilizzare meno corrispondenze per limitare il numero di iterazioni massime che devono essere eseguite in Ransac (si veda a riguardo le tabelle A.1 e A.2). Una volta ottenuta la soluzione del sistema di terzo grado si escludono le eventuali soluzioni complesse; se al contrario le soluzioni sono tutte reali, dal momento che la presenza di outlier non consente di impiegare un'altra corrispondenza per determinare quella corretta, si fa l'ipotesi che quella che fornisce il massimo consenso sia quella da tenere in considerazione. Al termine della procedura, come di consueto, si considera corretta (nel senso che il corrispondente campione non conteneva outlier), la soluzione con il massimo consenso. Per stimare

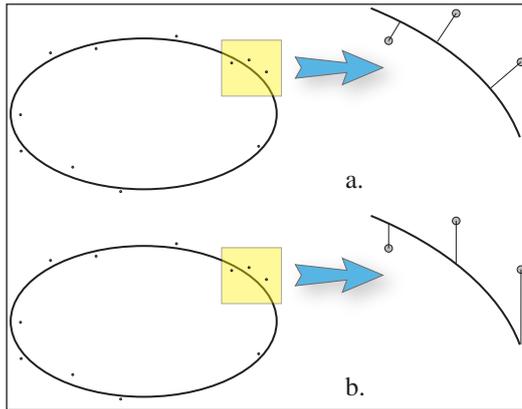


Figura 1.18: Rappresentazione della stima dell'errore (a.) con criteri geometrici come formulato in (1.58) e (b.) con criteri algebrici come formulato in (1.1)

l'errore di ciascun punto, da confrontare con la soglia di accettazione, non si utilizza l'equazione (1.51). Come osservato in [Torr 1997] e in [Torr 2002] essa fornisce una stima geometricamente non corretta della distanza di un punto dalla linea epipolare. La stima della matrice fondamentale può essere pensata come la stima di una quadrica degenera: con riferimento alla figura 1.18 in cui è rappresentata un'ellissi per semplicità di rappresentazione, la quantità rappresentata dal primo membro della (1.51) non rappresenta la distanza geometrica dalla conica, calcolata ortogonalmente alla tangente, ma una distanza algebrica; è allora evidente che tale grandezza risulta fortemente correlata al gradiente della curva nel punto considerato e può fornire una stima non corretta

della geometria epipolare. Si calcola quindi la distanza del punto dall'epipolare individuata dal suo omologo sull'altro fotogramma, ovvero:

$$\varepsilon = \frac{|\mathbf{x}'F\mathbf{x}|}{\sqrt{a^2 + b^2}}, \quad \text{con } F\mathbf{x} = (a, b, c)^T \quad (1.58)$$

Al termine della procedura Ransac, la matrice fondamentale viene ristimata, utilizzando tutti gli inlier del set di massimo consenso, con l'algoritmo a 8 punti: può accadere allora che la nuova geometria individui un set di consenso differente da quello ottenuto con Ransac; si procede quindi iterativamente al calcolo dei nuovi inlier ed alla ristima della matrice, finché non si raggiunge la convergenza, ovvero il set di consenso non cambia. Solitamente, se la distribuzione delle distanze dalle epipolari presenta due classi nettamente separate (vd. figura 1.19.a) tale procedura non è richiesta, in quanto durante la stima robusta si riescono ad isolare chiaramente inlier ed outlier; al contrario (figura 1.19.b), se tale distribuzione non presenta classi distinte di punti, si avranno punti che possono essere sicuramente considerati outlier, punti che possono essere considerati inlier ed una serie di punti che presentano valori di accettazione prossimi alla soglia imposta che,

per piccoli cambiamenti nella matrice fondamentale, transitano dall'una all'altra classe. In questo secondo caso una stima più ridondante assicura una maggior correttezza geometrica ed una più chiara distinzione fra inlier ed outlier: nella maggior parte dei casi il numero di corrispondenze corrette tende ad aumentare in maniera considerevole; a volte si assiste, nelle prime iterazioni del metodo, ad una eliminazione di corrispondenze che permette di convergere ad una nuova soluzione che porta ad un consenso molto più elevato rispetto a quello ottenuto con Ransac; infine, molto di rado ed in corrispondenza di geometrie di presa molto instabili (basi di presa piccole e percentuale molto bassa di inlier), il sistema tende a divergere numericamente, nel senso che nel corso delle iterazioni vengono continuamente scartate corrispondenze, ottenendo un consenso inferiore a quello di partenza. Circostanze simili si presentano anche nel caso della geometria trifocale di cui si parlerà nel seguente paragrafo. Ritenendo comunque migliore la soluzione di massimo consenso, per ogni iterazione se ne registra il valore scegliendo al termine la soluzione corrispondente al massimo. In analogia a quanto suggerito in [Torr 1998] il valore di consenso viene calcolato come

$$C = \sum_i \rho(\varepsilon_i^2) \quad (1.59)$$

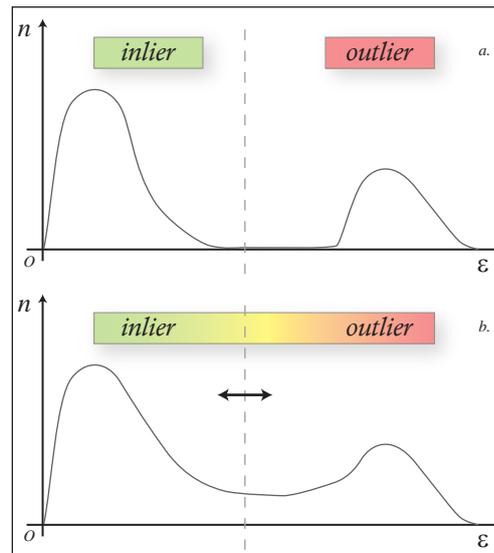


Figura 1.19: Rappresentazione schematica della possibile distribuzione degli errori (distanza dal modello funzionale) in un dataset.

in cui

$$\rho(\epsilon^2) = \begin{cases} \epsilon^2 & \text{se } \epsilon^2 < T^2 \\ T^2 & \text{se } \epsilon^2 \geq T^2 \end{cases} \quad (1.60)$$

La soluzione viene utilizzata per inizializzare una stima a minimi quadrati calcolata con tutti gli inlier corrispondenti.

Infine, per aumentare il numero di corrispondenze, viene svolta un'operazione di matching guidato in cui tutti i punti presenti sul fotogramma, che non sono risultati appartenere a coppie omologhe, vengono rivalutati tramite tecniche di feature matching in cui il vincolo geometrico è rappresentato dal legame epipolare stimato: vengono presi in considerazione tutti i punti (non accoppiati) estratti sul primo fotogramma; per ciascuno di essi si calcola, tramite la (1.49), la retta epipolare sul secondo fotogramma; si determinano i punti su tale immagine che si trovano ad una distanza inferiore alla soglia di accettazione precedentemente stabilita e si calcola per ciascun abbinamento un valore di similitudine dei toni di grigio (per mezzo di tecniche di correlazione semplice o, dal momento che in questo caso le combinazioni di punti sono solitamente poco numerose, per mezzo di tecniche più sofisticate come il LSM); viene quindi memorizzato l'abbinamento che presenta il massimo di tali valori a patto che esso sia superiore ad una certa soglia minima; infine, per garantire che tali accoppiamenti siano bilaterali [Ullman 1979] si procede allo stesso modo utilizzando i punti sul secondo fotogramma: al termine di tale operazione si considerano validi solo gli accoppiamenti che risultano uguali in entrambi i casi; può infatti accadere che un punto sul primo fotogramma si accoppi in maniera ottimale con un punto sul secondo che, a sua volta, risulta legarsi meglio ad un altro punto sul primo fotogramma.

Al termine, qualora sia richiesto dall'applicazione, si può ulteriormente affinare la stima della geometria epipolare con un sistema a minimi quadrati in cui vengono utilizzati tutti gli accoppiamenti trovati, condurre test di significatività e di presenza di eventuali errori grossolani sfuggiti nel corso dell'intera procedura.

Tale stima, molte volte, non si dimostra necessaria, in quanto non è in questa fase che si desidera individuare la geometria di presa del blocco: quello che conta maggiormente è l'individuazione di corrispondenze sui vari fotogrammi che siano funzionali ad ulteriori procedure di filtraggio ed analisi; in tal senso è dunque di scarso interesse stimare la geometria epipolare finale e la presenza di qualche errore grossolano non pregiudica i passaggi seguenti in cui gli accoppiamenti vengono ulteriormente filtrati. Lo schema dell'intero procedimento è illustrato in figura 1.20.

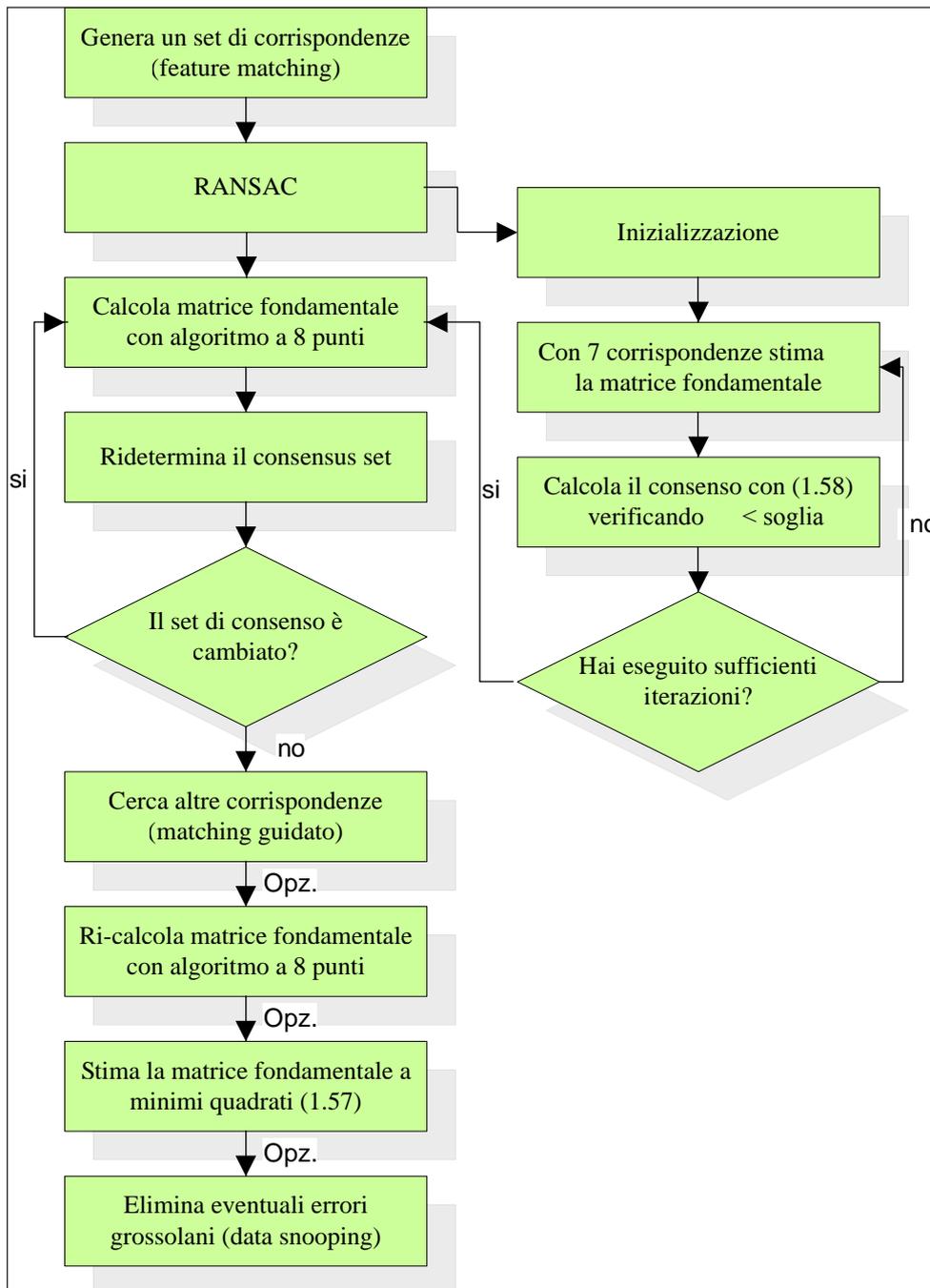


Figura 1.20: Schema delle operazioni implementate per la stima robusta della geometria epipolare.

1.6. Geometria trifocale

Il vincolo epipolare è, come noto, poco robusto: tutti i punti oggetto che giacciono sul piano epipolare sono accettabili; la scarsa caratterizzazione di tessitura o la sua ripetitività può portare a condizioni di ambiguità negli accoppiamenti. Le prove effettuate con diversi tipi di scena, geometria di blocco e qualità delle immagini indicano che, al termine della stima della matrice fondamentale una percentuale variabile fra il 10% e il 40% degli accoppiamenti accettati è in effetti non corretto. Stimare la geometria di presa con una presenza così consistente di errori grossolani non è possibile, se si vuole garantire una sufficiente affidabilità. Risulta allora necessario filtrare ulteriormente l'insieme delle corrispondenze con un vincolo geometrico più forte, come quello rappresentato dalla geometria trifocale, ovvero quella che si instaura fra tre immagini distinte. In questo caso, verrà dimostrato nel seguito, le configurazioni critiche sono praticamente inesistenti e la geometria risulta molto più robusta: il vincolo trifocale infatti non implica soltanto che due geometrie epipolari siano contemporaneamente soddisfatte (ovvero che un punto sulla terza immagine si trovi sull'intersezione delle due linee epipolari corrispondenti ai punti punto sul primo e sul secondo fotogramma). Nel caso in cui tali linee fossero pressoché coincidenti il doppio vincolo epipolare non si dimostrerebbe efficace, al contrario di quello imposto dal tensore trifocale.

Il tensore trifocale, che rappresenta un'estensione della matrice fondamentale, mette in relazione i parametri di orientamento assoluto delle tre camere (in un sistema arbitrario) e le coordinate oggetto del punto: pertanto, la configurazione critica poc'anzi illustrata non è problematica in quanto definendo la posizione spaziale del punto non vi è possibilità di ambiguità sulla linea epipolare. Imponendo con soglie opportune il soddisfacimento del vincolo trifocale, la nostra esperienza ci insegna, vengono eliminati tutti gli errori grossolani.

Il tensore trifocale viene inizialmente introdotto in [Weng 1988] e [Spetsakis 1991], anche se è dovuto ad [Hartley 1994a] l'estensione al caso in cui non si conoscano i parametri di orientamento interno. Le prime applicazioni del tensore al problema della structure and motion sono invece dovuti a [Torr 1995] e [Beardsley 1996].

1.6.1. Tensore Trifocale

Si consideri inizialmente una corrispondenza fra linee su tre distinti fotogrammi: sia L una retta nello spazio oggetto ed l, l' ed l'' l'immagine di L sui tre fotogrammi; si consideri poi un sistema di riferimento terreno tale che la prima camera risulti in forma canonica, ovvero con centro di proiezione nell'origine del sistema ed assi coincidenti con gli assi coordinati: $P = [I | \mathbf{0}]$; le altre matrici camera avranno forma $P' = [A | \mathbf{a}_1]$ e $P'' = [B | \mathbf{b}_1]$.

È facile verificare che i due vettori \mathbf{a}_1 e \mathbf{b}_1 coincidono con gli epipoli corrispondenti al primo fotogramma, rispettivamente sulla seconda e terza immagine: è infatti

$$\begin{aligned} \mathbf{e}' &= P'C = P' |0 \ 0 \ 0 \ 1|^T = \mathbf{a}_i \\ \mathbf{e}'' &= P''C = P'' |0 \ 0 \ 0 \ 1|^T = \mathbf{b}_i \end{aligned} \quad (1.61)$$

Ciascuna linea sui vari fotogrammi individua un piano nello spazio oggetto in funzione della matrice camera:

$$\boldsymbol{\pi} = P^T \mathbf{l} = \begin{vmatrix} \mathbf{l} \\ 0 \end{vmatrix} \quad \boldsymbol{\pi}' = P'^T \mathbf{l}' = \begin{vmatrix} A^T \mathbf{l}' \\ \mathbf{a}_i^T \mathbf{l}' \end{vmatrix} \quad \boldsymbol{\pi}'' = P''^T \mathbf{l}'' = \begin{vmatrix} B^T \mathbf{l}'' \\ \mathbf{b}_i^T \mathbf{l}'' \end{vmatrix} \quad (1.62)$$

Infatti un punto \mathbf{x} giace su \mathbf{l} se e solo se $\mathbf{x}^T \mathbf{l} = 0$ (vd. equazione A.4); al tempo stesso un generico punto oggetto \mathbf{X} viene proiettato sul piano immagine in $P\mathbf{X}$: tale punto giace sulla retta se e solo se $\mathbf{X}^T P^T \mathbf{l} = 0$; è allora immediato verificare che il punto \mathbf{X} giace sul piano $P^T \mathbf{l}$.

Dal momento che i tre piani in (1.62) hanno in comune la retta L , i loro parametri devono essere linearmente dipendenti: in particolare detti \mathbf{m}_1 , \mathbf{m}_2 ed \mathbf{m}_3 i tre vettori dei parametri dei piani in (1.62) deve risultare

$$\mathbf{m}_1 = \alpha \mathbf{m}_2 + \beta \mathbf{m}_3 \quad (1.63)$$

è immediato allora ricavare, sostituendo le grandezze vettoriali ottenute in (1.62), che

$$\alpha = k (\mathbf{b}_i^T \mathbf{l}''), \quad \beta = -k (\mathbf{a}_i^T \mathbf{l}') \quad (1.64)$$

in cui k è un valore reale diverso da 0. Ponendo tale valore pari a 1 dal momento che si stanno utilizzando grandezze omogenee definite a meno di una costante di proporzionalità, si ottiene sostituendo nella (1.63)

$$\mathbf{l} = (\mathbf{b}_i^T \mathbf{l}'') A^T \mathbf{l}' - (\mathbf{a}_i^T \mathbf{l}') B^T \mathbf{l}'' = (\mathbf{l}''^T \mathbf{b}_i) A^T \mathbf{l}' - (\mathbf{l}'^T \mathbf{a}_i) B^T \mathbf{l}'' \quad (1.65)$$

la componente i -esima della retta \mathbf{l} risulta dunque

$$l_i = \mathbf{l}''^T (\mathbf{b}_i \mathbf{a}_i^T) \mathbf{l}' - \mathbf{l}'^T (\mathbf{a}_i \mathbf{b}_i^T) \mathbf{l}'' \quad (1.66)$$

in cui \mathbf{a}_i e \mathbf{b}_i rappresentano le colonne i -esime delle due matrici camera P' e P'' . Posto

$$T_i = \mathbf{a}_i \mathbf{b}_i^T - \mathbf{a}_i \mathbf{b}_i^T \quad (1.67)$$

si ottiene allora la relazione di incidenza che lega tre rette corrispondenti sui tre fotogrammi:

$$l_i = \mathbf{l}''^T T_i \mathbf{l}' \quad (1.68)$$

L'insieme delle tre matrici $T_i = [T_1, T_2, T_3]$ rappresenta il tensore trifocale espresso in notazione matricia-

le. Generalmente alla notazione matriciale si preferisce quella tensoriale, in quanto più immediata: il tensore trifocale corrisponde infatti ad una matrice tridimensionale $3 \times 3 \times 3$ che può essere indicata come T_i^{jk} . In notazione tensoriale la (1.68) si trasforma in

$$l_i = l'_j l''_k T_i^{jk} \quad (1.69)$$

in cui il tensore trifocale è definito, in analogia alla (1.67) come

$$T_i^{jk} = a_i^j b_4^k - a_4^j b_i^k \quad (1.70)$$

Si ricorda che la ripetizione degli indici in notazione tensoriale corrisponde a considerare la sommatoria nell'indice ripetuto: la (1.69) equivale dunque a

$$l_i = \sum_j \sum_k l'_j l''_k T_i^{jk} \quad (1.71)$$

E' facile osservare come la (1.71) sia del tutto equivalente alla (1.68).

Dalla (1.68) è immediato ottenere la condizione di corrispondenza di tre rette omologhe

$$(\mathbf{l}^T [T_1, T_2, T_3] \mathbf{l}'') [\mathbf{l}]_{\times} = \mathbf{0}^T \quad (1.72)$$

che in forma tensoriale diventa

$$(l_i \varepsilon^{ris}) l'_j l''_k T_i^{jk} = 0^s \quad (1.73)$$

in cui il tensore ε_{ris} è definito per $r, s, t = 1, \dots, 3$ come

$$\varepsilon_{ris} = \begin{cases} 0 & \text{se } r = s \mid r = t \mid s = t \\ +1 & \text{se } rst \text{ e' una permutazione diretta di } 123 \\ -1 & \text{se } rst \text{ e' una permutazione inversa di } 123 \end{cases} \quad (1.74)$$

e permette di svolgere un'operazione equivalente al prodotto vettoriale in (1.72).

Dalla (1.68) è possibile ricavare un'ulteriore condizione di incidenza: date due rette corrispondenti \mathbf{l}' ed \mathbf{l}'' sul secondo e terzo fotogramma, un punto \mathbf{x} sul primo corrisponde ad un punto \mathbf{X} sulla retta L nello spazio terreno se e solo se $\mathbf{x}^T \mathbf{l} = 0$, in cui \mathbf{l} è la retta determinata per mezzo della (1.68). Ne deriva

$$\mathbf{x}^T (\mathbf{l}^T [T_1, T_2, T_3] \mathbf{l}'') = \mathbf{l}^T \left(\sum_i \mathbf{x}^i T_i \right) \mathbf{l}'' = 0 \quad (1.75)$$

$$\mathbf{x}^i l'_j l''_k T_i^{jk} = 0 \quad (1.76)$$

Si consideri a questo punto l'omografia che si viene a generare ad esempio fra il primo ed il terzo fotogramma per mezzo di un trasferimento per il piano π' individuato da una retta l' sul secondo fotogramma: per un punto sul terzo fotogramma vale $\mathbf{x}'' = H\mathbf{x}$ e per una retta sul primo fotogramma risulta $l = H^T l'$; sempre in relazione alla (1.68) deriva

$$H = [\mathbf{h}_1 \quad \mathbf{h}_2 \quad \mathbf{h}_3] \quad \text{con} \quad \mathbf{h}_i = T_i^T l' \quad (1.77)$$

che può essere più efficacemente scritta come

$$H_{13}(l') = [T_1^T \quad T_2^T \quad T_3^T] l' \quad (1.78)$$

dalla quale si ricava la relazione di incidenza che lega due punti (ad esempio sul primo e terzo fotogramma) ed una linea (sul secondo fotogramma)

$$\mathbf{x}'' = H_{13}(l') \mathbf{x} = [T_1^T l' \quad T_2^T l' \quad T_3^T l'] \mathbf{x} \quad (1.79)$$

$$l'^T \left(\sum_i x^i T_i \right) [\mathbf{x}'']_{\times} = \mathbf{0}^T \quad (1.80)$$

che in notazione tensoriale diventa

$$x^i l'_j (x''^k \varepsilon_{kqs}) T_i^{jq} = 0_s \quad (1.81)$$

Infine si consideri la retta l' passante per un punto \mathbf{x}' sul secondo fotogramma: tale retta può essere individuata per mezzo di un altro punto \mathbf{y}' come $l' = \mathbf{x}' \wedge \mathbf{y}' = [\mathbf{x}']_{\times} \mathbf{y}'$; sostituendo nella (1.80) si ottiene

$$\mathbf{y}'^T [\mathbf{x}']_{\times} \left(\sum_i x^i T_i \right) [\mathbf{x}'']_{\times} = \mathbf{0}^T \quad (1.82)$$

Si osservi a questo punto che, per considerare la relazione di incidenza fra i tre punti \mathbf{x} , \mathbf{x}' e \mathbf{x}'' , la relazione (1.82) deve valere per ciascuna retta l' , e di conseguenza per ciascun \mathbf{y}' : ne deriva che

$$[\mathbf{x}']_{\times} \left(\sum_i x^i T_i \right) [\mathbf{x}'']_{\times} = \mathbf{0}_{3 \times 3} \quad (1.83)$$

$$x^i (x'^j \varepsilon_{jpr}) (x''^k \varepsilon_{kqs}) T_i^{pq} = 0_{rs} \quad (1.84)$$

sono le relazioni di incidenza, rispettivamente in forma matriciale e tensoriale che lega tre punti omologhi su altrettanti fotogrammi.

Le relazioni di incidenza finora descritte sono riassunte nel formulario in appendice B in cui vengono

anche illustrate le relazioni che permettono di ricavare tutte le grandezze di interesse dal tensore trifocale: come nel caso della matrice fondamentale, il tensore incapsula tutte le informazioni relative alla geometria di presa ed è quindi possibile ottenere direttamente da esso le tre matrici fondamentali che legano a coppie i vari fotogrammi, gli epipoli, le matrici camera e così via.

1.6.2. Metodi di stima del tensore trifocale

Per la stima del tensore trifocale è necessaria la conoscenza dei 27 elementi che lo compongono (definiti, come al solito, a meno di una costante di proporzionalità); il numero di gradi di libertà del tensore è tuttavia più basso: come nel caso della matrice fondamentale esistono infatti dei vincoli geometrici interni. Per determinarne il numero si può ragionare come segue: il tensore deriva da tre matrici camera distinte espresse in un sistema proiettivo generico: in tale sistema, verrà dimostrato nel prossimo paragrafo, vengono implicitamente fissati gli elementi di una omografia tridimensionale (definita da 15 parametri indipendenti) che lega lo spazio reale a quello arbitrario; per fissare le matrici camera sono richiesti 11 parametri di orientamento per ciascuna (6 di orientamento esterno e 5 di orientamento interno, in virtù del fatto che la matrice camera tiene in conto anche l'aspect ratio ed il coefficiente di taglio); ne deriva che per definire correttamente il tensore trifocale sono richieste 18 condizioni indipendenti. Dai risultati emersi nel precedente paragrafo, considerando che si utilizzano grandezze omogenee e pertanto, per ciascuna grandezza vettoriale solo due delle tre componenti sono linearmente indipendenti, deriva che condizioni di incidenza di tre punti corrispondono a 4 condizioni linearmente indipendenti, mentre condizioni fra tre linee corrispondono a 2 condizioni; in applicazioni di structure and motion sono questi due tipi di relazioni ad essere prevalentemente utilizzati (nella nostra implementazione vengono usate solamente condizioni fra punti): le altre condizioni non vengono di fatto considerate.

Il caso risolutivo minimo richiede la conoscenza di sole 6 corrispondenze: in questo caso infatti il tensore viene calcolato direttamente dalle tre matrici camera, in modo da garantire che i vincoli interni siano soddisfatti, ovvero che il tensore corrisponda ad una situazione reale di presa. L'algoritmo, piuttosto complesso, viene qui brevemente riassunto, rimandando ad [Hartley 2000] per maggiori dettagli. Tale algoritmo è basato sul teorema di dualità di Carlsson-Weinshall ([Carlsson 1995], [Weinshall 1995], [Carlsson 1998]) che permette di invertire il ruolo fra punti omologhi su diversi fotogrammi e i centri di proiezione dei fotogrammi stessi; si considerino i quattro punti che formano la base canonica dello spazio proiettivo \mathbb{IP}^3 : $E_1 = (1, 0, 0, 0)^T$, $E_2 = (0, 1, 0, 0)^T$, $E_3 = (0, 0, 1, 0)^T$ e $E_4 = (0, 0, 0, 1)^T$; si considerino, inoltre i punti che formano la base canonica di \mathbb{IP}^2 $e_1 = (1, 0, 0)^T$, $e_2 = (0, 1, 0)^T$, $e_3 = (0, 0, 1)^T$ oltre al punto $e_4 = (1, 1, 1)^T$; si consideri infine una matrice camera P che non giaccia su nessun asse coordinato (in modo che i punti PE_i non siano allineati). Si può allora determinare una omografia H che porti le immagini tramite P della base di \mathbb{IP}^3 nei corrispondenti punti di \mathbb{IP}^2 .

$$e_i = HPE_i \quad (1.85)$$

La matrice che si ottiene dal prodotto $P_{red} = HP$ prende il nome di matrice camera ridotta; per verificare la (1.85) essa deve avere la forma

$$P_{red} = \begin{vmatrix} a & 0 & 0 & d \\ 0 & b & 0 & d \\ 0 & 0 & c & d \end{vmatrix} \quad (1.86)$$

che rappresenta una camera con centro nel punto $(a^{-1}, b^{-1}, c^{-1}, -d^{-1})$: per qualsiasi punto oggetto $\mathbf{X} = (X, Y, Z, W)^T$ vale allora la seguente uguaglianza

$$P_{red}\mathbf{X} = \begin{vmatrix} a & 0 & 0 & d \\ 0 & b & 0 & d \\ 0 & 0 & c & d \end{vmatrix} \begin{vmatrix} X \\ Y \\ Z \\ W \end{vmatrix} = \begin{vmatrix} X & 0 & 0 & W \\ 0 & Y & 0 & W \\ 0 & 0 & Z & W \end{vmatrix} \begin{vmatrix} a \\ b \\ c \\ d \end{vmatrix} = \begin{vmatrix} aX + dT \\ bY + dT \\ cZ + dT \end{vmatrix} \quad (1.87)$$

in cui si sono effettivamente invertiti i ruoli fra camera e punti. Sebbene tale dualità si dimostri poco intuitiva, e possa far pensare ad un semplice artificio numerico, essa si dimostra di grande utilità per la stima di grandezze legate alla geometria di presa. In particolare si può dimostrare che, grazie al teorema di dualità di Carlsson è possibile, da un qualsiasi algoritmo che riguardi n fotogrammi e $m+4$ punti, ottenerne uno affine che lega $n+4$ punti e m immagini: nel nostro caso, considerata la prima camera che forma la tripletta fissa in posizione canonica ($P = [I \ \mathbf{0}]$), si vuole ottenere la ricostruzione dei restanti 2 fotogrammi utilizzando 6 punti.

Si scelgono innanzitutto quattro punti non allineati a tre a tre per stimare una trasformazione di collinearità che porti tali punti a coincidere con gli \mathbf{e}_i ; come noto un'omografia rappresenta la trasformazione di collinearità più generale e richiede, per la sua stima almeno quattro punti; tale trasformazione viene applicata ai restanti due punti del set; lo stesso procedimento viene applicato agli altri fotogrammi, ottenendo infine 3 coppie di punti.

In analogia a quanto fatto per la matrice camera si definisce matrice fondamentale ridotta la matrice F_{red} tale che $\mathbf{e}_i^T F_{red} \mathbf{e}_i = 0$, $i = 1, \dots, 4$, con $i = 1, \dots, 4$; è immediato verificare che essa deve avere elementi nulli lungo la diagonale principale e che la somma degli elementi deve risultare uguale a zero; in altre parole la sua forma generica deve risultare

$$F_{red} = \begin{vmatrix} 0 & p & q \\ r & 0 & s \\ t & -(p + q + r + s + t) & 0 \end{vmatrix} \quad (1.88)$$

Riferendoci ai risultati ottenuti nel paragrafo sulla stima della matrice fondamentale è facile osservare che, utilizzate quattro corrispondenze per stimare la trasformazione di riduzione (cioè l'omografia poc' anzi

descritta) da altre tre corrispondenze è possibile ottenere una famiglia di soluzioni che, una volta introdotto il vincolo $\det(F_{red}) = 0$, permette di risolvere compiutamente la matrice, in tutta analogia all'algoritmo a 7 punti. Le tre coppie di punti, ottenute singolarmente su ciascun fotogramma, vengono dunque utilizzate per stimare tale matrice. Dalla fondamentale ridotta è particolarmente semplice risalire ad una coppia di matrici camera ridotte che la generino: in questo caso non è possibile fissare la prima camera in posizione canonica, in quanto il centro di proiezione in questo caso coinciderebbe con il punto E_4 ; il centro di proiezione della prima camera può allora essere fissato in corrispondenza del punto $(1, 1, 1, 1)^T$:

$$P = \begin{vmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 \end{vmatrix} \quad P' = \begin{vmatrix} a & 0 & 0 & d \\ 0 & b & 0 & d \\ 0 & 0 & c & d \end{vmatrix} \quad (1.89)$$

Ricavando la matrice fondamentale dalle due matrici camera (vd. formulario appendice B) si ottiene che la fondamentale ridotta risulta

$$F_{red} = \begin{vmatrix} 0 & b(d-c) & -c(d-b) \\ -a(d-c) & 0 & c(d-a) \\ a(d-b) & -b(d-a) & 0 \end{vmatrix} \quad (1.90)$$

la quale è immediato verificare, soddisfa i vincoli della (1.88).

Ottenuta dunque la F_{red} numericamente, dai suoi elementi si ricavano i parametri della seconda matrice camera ridotta risolvendo un sistema in forma lineare: il rapporto a:b:c può essere ottenuto risolvendo l'equazione omogenea

$$\begin{vmatrix} f_{12} & f_{21} & 0 \\ f_{13} & 0 & f_{31} \\ 0 & f_{23} & f_{32} \end{vmatrix} \begin{vmatrix} a \\ b \\ c \end{vmatrix} = 0 \quad (1.91)$$

in cui f_{ij} è l'elemento corrispondente alla i-esima riga e j-esima colonna della F_{red} calcolata. Da tale rapporto è possibile ottenere due equazioni linearmente indipendenti in (a, b, c, d) .

Analogamente il rapporto $(d-a):(d-b):(d-c)$ può essere ottenuto risolvendo

$$|d-a \quad d-b \quad d-c| F_{red} = 0 \quad (1.92)$$

che fornisce altre due equazioni linearmente indipendenti. Unendo le equazioni ottenute con la (1.91) e quelle con la (1.92) si ottiene un sistema lineare in quattro equazioni nelle quattro incognite (a, b, c, d) , la cui conoscenza permette di risalire alla seconda matrice camera ridotta di (1.89). A questo punto, invocando nuovamente il principio di dualità è possibile trasformare le due matrici camera di (1.89) nei punti oggetto $X_1 = (1, 1, 1, 1)^T$ e $X_2 = (a, b, c, d)^T$; inoltre sono note le coordinate terreno corrispondenti agli altri punti in

quanto, per la trasformazione duale $X_{2+k} = E_k$. A questo punto, dunque, si è in grado di ottenere le matrici camera per mezzo di procedure di resezione, la più semplice delle quali, in virtù del fatto che il sistema è monodeterminato, è la DLT (equazione (1.43)). In tale resezione conviene utilizzare direttamente le coordinate immagine originarie, in modo da non dover trasformare successivamente le matrici camera da spazio ridotto a spazio iniziale, cosa che, come osservato in [Hartley 2000] può portare a soluzioni non ottimali per l'influenza del noise di misura. L'intera procedura è riassunta schematicamente in figura 1.21.

Un metodo di risoluzione ridondante, se si dispone di numerose corrispondenze è rappresentato dall'algoritmo lineare: esso, in analogia all'algoritmo ad otto punti per la matrice fondamentale, parte con una stima lineare del tensore per mezzo di equazioni di osservazione derivate dalle equazioni di incidenza illustrate precedentemente (equazioni da (1.72) a (1.84)) determinando tutti gli elementi del tensore, mentre i vincoli geometrici interni vengono imposti in un secondo momento. Anche in questo caso è conveniente pre-normalizzare le coordinate immagine per rendere numericamente stabile la stima: dette H, H' e H'' le matrici di trasformazione come di consueto i punti immagine divengono $\bar{x} = Hx$ e le rette $\bar{l} = H^{-T}l$; si ottiene un sistema risolvente del tipo

$$At = 0 \quad (1.93)$$

in cui t è il vettore delle 27 componenti incognite del tensore trifocale e A è la matrice disegno costituita da 27 colonne ed n righe, con $n = 4n_{ppp} + 2n_{ppl} + n_{pll} + n_{lll}$, in cui n_{ppp} rappresenta il numero di corrispondenze fra punti omologhi utilizzate nella stima, n_{ppl} sono le corrispondenze punto-punto-linea, n_{pll} quelle punto-linea-linea e n_{lll} le corrispondenze fra tre rette omologhe. Una volta risolto il sistema, ad esempio con una SVD, detto T_r^{st} il tensore risultante, si ottiene quello corrispondente agli spazi immagine originari per mezzo di:

$$T_i^{jk} = H_i^r (H'^{-1})_s^j (H''^{-1})_t^k T_r^{st} \quad (1.94)$$

E' ovvio che il tensore (1.94) non soddisfa, in generale, le condizioni geometriche interne del tensore; per ottenerne uno geometricamente coerente si procede come segue: si ricavano da T_i^{jk} i due epipoli e' ed e'' sul primo fotogramma; in virtù della (1.70) un tensore, geometricamente corretto, che dia origine agli stessi epipoli sul primo fotogramma è dato da

$$T_i^{jk} = \alpha_i^j e'^k - e''^j b_i^k \quad (1.95)$$

si genera la matrice E nelle componenti incognite di α_i^j e b_i^k (raggruppate ordinatamente nel vettore a) in modo che $t = Ea$; si risolve, ad esempio con il metodo descritto in appendice A.2.1., per mezzo della SVD, il problema

$$\|AEa\| = \min, \quad \text{con } \|Ea\| = 1 \quad (1.96)$$

In tal modo il tensore trifocale è geometricamente corretto; tuttavia, avendo utilizzato gli epipoli de-

rivanti dalla soluzione geometricamente incoerente è possibile che la soluzione non risulti ottimale: per ovviare al problema si calcola il vettore errore

$$\boldsymbol{\varepsilon} = A\mathbf{E}\mathbf{a} \quad (1.97)$$

e si considera la funzione da \mathbb{R}^6 a \mathbb{R}^{27} che porta le componenti dei due epipoli all'errore algebrico relativo al tensore; si può quindi cercare di minimizzare la funzione obiettivo (1.97) con metodi iterativi, ripetendo la stima ottenuta per mezzo delle (1.95) e (1.96), ad esempio con l'algoritmo derivato dal metodo di Levenberg-Marquadt illustrato in appendice A.2.2. Infine, anche nel caso del tensore trifocale, si può decidere di terminare la stima con un metodo a minimi quadrati cercando di minimizzare quantità geometriche più significative, come ad esempio l'errore di riproiezione

$$\sum_i d(\mathbf{x}_i, \bar{\mathbf{x}}_i) + d(\mathbf{x}'_i, \bar{\mathbf{x}}'_i) + d(\mathbf{x}''_i, \bar{\mathbf{x}}''_i) \quad (1.98)$$

Tale ultimo passaggio può anche essere sostituito da un bundle block adjustment (con equazioni di osservazione per eliminare la deficienza di rango) in cui vengono tenuti incogniti contemporaneamente alcuni dei parametri di orientamento e le coordinate terreno dei punti. Il sistema può essere risolto nello spazio proiettivo arbitrario come illustrato in [Triggs 1999] o, a seguito di una ricostruzione metrica, imponendo i parametri di orientamento interno e seguendo i metodi tradizionali. Solitamente al termine della stima del tensore trifocale, in problemi di structure and motion, si fa sempre seguire una ricostruzione metrica della scena ed un bundle adjustment finale, e pertanto, il più delle volte, questo ultimo passaggio viene tralasciato.

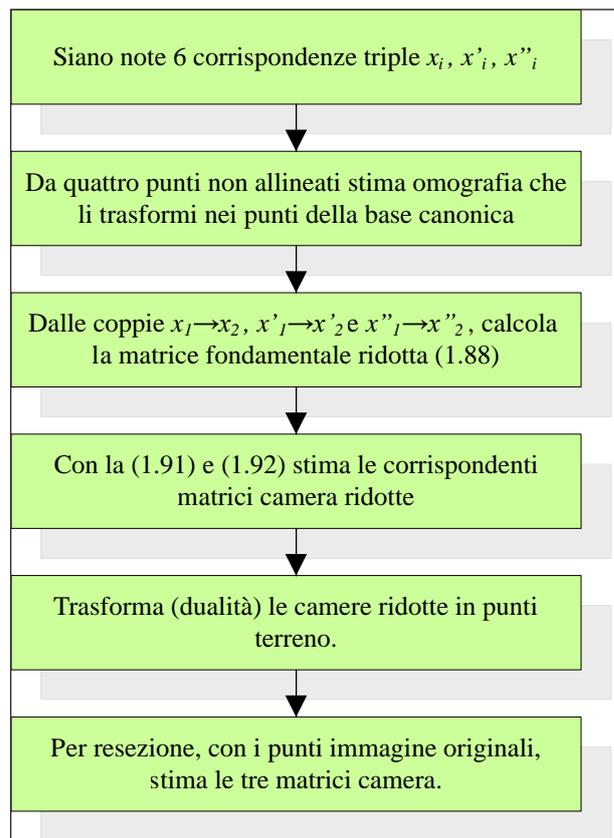


Figura 1.21: Schema delle principali operazioni da eseguire per la stima del tensore trifocale note sei corrispondenze triple.

1.6.3. Stima robusta della geometria trifocale

Gli accoppiamenti filtrati con la matrice fondamentale vengono analizzati per ottenere corrispondenze triple sui fotogrammi; le coppie che non trovano riscontro sul terzo fotogramma vengono escluse in quanto non corrette: in realtà un punto doppio che verifichi la geometria di presa pur essendo sbagliato non causa errori nella stima dei parametri di orientamento esterno, ma solo nella ricostruzione dell'oggetto; dal momento che questa è fondamentale nei passaggi successivi alla structure and motion, come ad esempio il matching denso (vd. capitolo seguente), tali corrispondenze vengono eliminate.

Le linee di sviluppo della stima robusta del tensore trifocale seguono sostanzialmente quelle utilizzate per la stima della matrice fondamentale: per mezzo dell'algoritmo a sei punti precedentemente descritto si determina un tensore trifocale che permetta di definire il set di consenso; come visto si determinano le tre matrici camera relative ai tre fotogrammi, si calcolano le coordinate terreno di tutti i punti corrispondenti alle triplette disponibili per mezzo della (1.45) o del metodo sviluppato in [Hartley 1997]; in questo secondo caso, come descritto nel paragrafo 1.4., la triangolazione è basata sulla matrice fondamentale fra due fotogrammi: si scelgono il primo e il terzo fotogramma, supponendo che siano quelli con base di presa maggiore, e si calcola la posizione del punto terreno; il metodo quindi non prende in considerazione le informazioni sul fotogramma di mezzo e richiede uno sforzo computazionale maggiore rispetto all'altro; ciononostante il risultato è più affidabile dal momento che le matrici camera vengono ottenute in uno spazio proiettivo generico.

Al termine della procedura Ransac, il tensore trifocale viene ristimato, utilizzando tutti gli inlier del set di massimo consenso, per mezzo dell'algoritmo descritto nel paragrafo precedente, in modo da ottenere una stima più ridondante: anche in questo caso può accadere che la nuova geometria individui un set di consenso differente da quello ottenuto con Ransac; si procede quindi iterativamente al calcolo dei nuovi inlier ed alla ristima del tensore, fino al raggiungimento della convergenza. Il calcolo del consenso viene di nuovo eseguito seguendo le (1.59) e (1.60).

Infine si può decidere, con la conoscenza delle tre matrici camera, di eseguire un feature matching guidato, per aumentare il numero di corrispondenze: in questo caso non è più necessario valutare la similitudine dei toni di grigio fra i vari accoppiamenti in quanto è già stato eseguito nei passaggi precedenti. In questo caso, però, tale matching è solitamente inutile e ridondante, in quanto si presume che gli accoppiamenti possibili siano già stati analizzati compiutamente all'inizio della procedura ed il dataset non possa essere ulteriormente ampliato.

Inoltre, normalmente, non è richiesta una stima a minimi quadrati della geometria di presa, in quanto essa viene definitivamente stimata durante la ricostruzione metrica e il conseguente bundle block adjustment. La soluzione viene allora utilizzata per inizializzare una stima a minimi quadrati in cui, come osservazioni, vengono utilizzati tutti gli inlier corrispondenti.

Lo schema dell'intero procedimento è illustrato in figura 1.22.

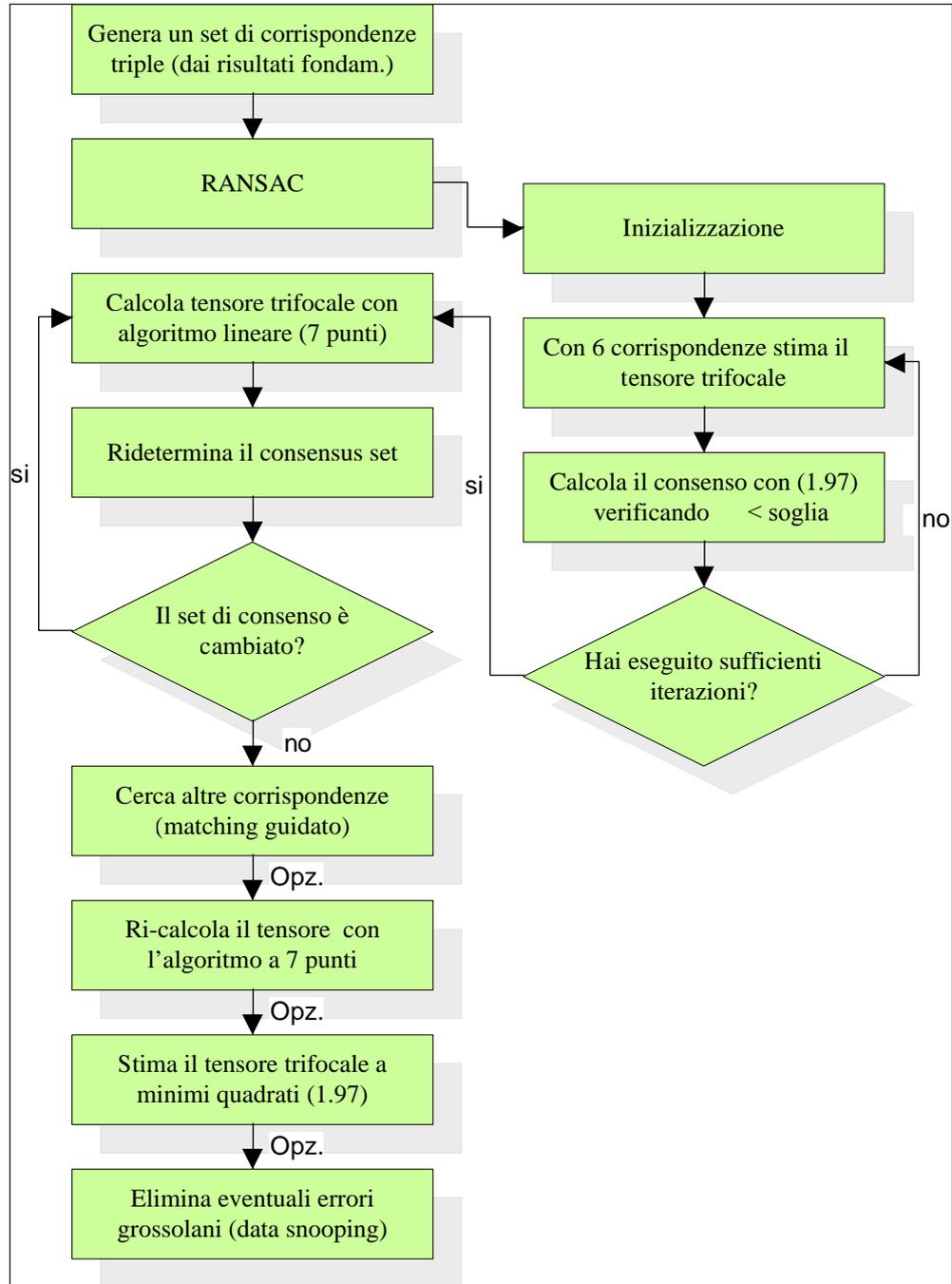


Figura 1.22: Schema delle operazioni implementate per la stima robusta della geometria trifocale.

1.7. Ricostruzione metrica e bundle adjustment

Al termine della stima del tensore trifocale il blocco fotogrammetrico è esente, di fatto, da errori grossolani. Prima di procedere alla ricostruzione tridimensionale dell'oggetto va osservato che le matrici camera estratte con il tensore trifocale, seppur corrette da un punto di vista proiettivo, non coincidono con quelle reali: la forma dell'oggetto ed i parametri delle camere sono stati calcolati a partire dalle sole coordinate immagine, senza sfruttare eventuali caratteristiche delle camere utilizzate (parametri di orientamento interno), o prenoscenze sulla geometria dell'oggetto (punti d'appoggio). Si possono allora ottenere configurazioni proiettivamente equivalenti, nel senso che forniscono residui nelle equazioni di collinearità identici, in cui sia la geometria dell'oggetto che i parametri di orientamento delle camere risultano modificati rispetto a quelle effettive. Si consideri infatti il generico punto oggetto X e la camera rappresentata dalla matrice P . Moltiplicando opportunamente per una matrice di trasformazione H (omografia tridimensionale) entrambe le quantità a secondo membro della (1.40) si ottiene

$$\mathbf{x} = (PH^{-1})(HX) \quad (1.99)$$

in cui è facile osservare che una matrice camera PH^{-1} fornisce le stesse coordinate immagine nel proiettare un punto oggetto HX sul fotogramma (figura 1.23). L'omografia H è dunque la trasformazione proiettiva arbitraria che lega lo spazio euclideo reale allo spazio proiettivo in cui si sono fin'ora ottenuti i risultati della structure and motion. L'omografia H agisce su tutti gli elementi della matrice camera e di conseguenza modifica sia i parametri di orientamento esterno sia quelli di orientamento interno. Ciò spiega anche perché, nella determinazione del tensore trifocale, non sia possibile partire dalla conoscenza delle matrici fondamentali che legano, a coppie, i fotogrammi interessati: ciascuna matrice è infatti definita nel suo proprio riferimento proiettivo e non può dunque essere messa in rapporto con le altre; in altri termini, qualora si disponga, ad esempio, di due matrici fondamentali, generate da un fotogramma co-

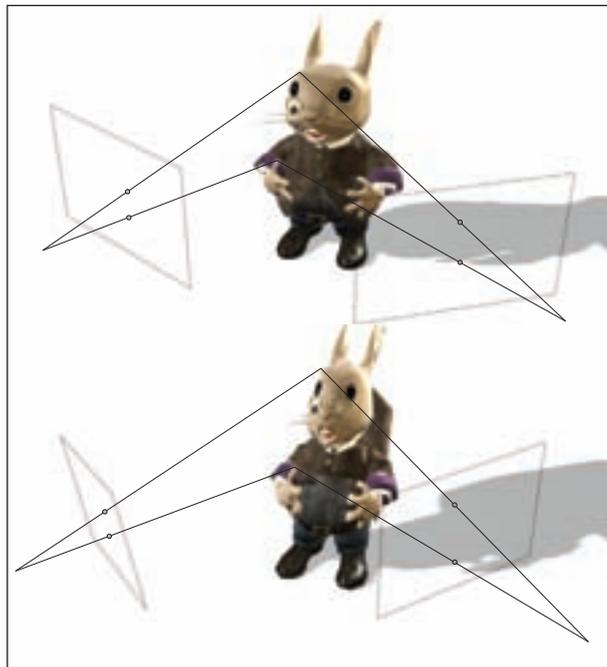


Figura 1.23: I risultati ottenuti con la structure and motion sono definiti a meno di una trasformazione proiettiva: configurazioni geometricamente differenti possono condurre a risultati, in termini di riproiezione, identici.

mune, la matrice camera corrispondente (e dunque i parametri di orientamento interno ed esterno) non è la stessa nelle due fondamentali. E' possibile tuttavia, disponendo di un numero di fotogrammi sufficienti, stimare la trasformazione proiettiva che riporta l'intero blocco in uno spazio euclideo in cui i parametri di orientamento interno siano gli stessi per tutti i fotogrammi e coincidano con quelli reali.

Algoritmi di auto-calibrazione che permettono di svolgere tale procedura partendo ancora una volta solamente dalle corrispondenze ottenute sui diversi fotogrammi, sono stati sviluppati per la prima volta in [Faugeras 1992b] per mezzo delle equazioni di Kruppa. Altri riferimenti importanti sono dovuti a [Hartley 1994b], [Mohr 1993] e [Pollefeys 1996].

Qualora si disponga dei parametri di orientamento interno, la ricostruzione metrica è molto più semplice, potendosi utilizzare la matrice essenziale [Longuet-Higgins 1981], che rappresenta l'analogo calibrato della matrice fondamentale. Si consideri una generica matrice camera del tipo:

$$P = K [R \mid R\tilde{C}] = K [R \mid \mathbf{t}] \quad t.c. \quad \mathbf{x} = PX \quad (1.100)$$

Si moltiplichino ambo i membri della (1.38) a sinistra per K^{-1}

$$K^{-1}\mathbf{x} = K^{-1}K [R \mid \mathbf{t}]X = [R \mid \mathbf{t}]X \quad (1.101)$$

è immediato osservare che, se si considerano le quantità derivate dalle coordinate immagine (coordinate immagine normalizzate)

$$\bar{\mathbf{x}} = K^{-1}\mathbf{x} \quad (1.102)$$

la matrice camera equivalente (matrice camera normalizzata) ha come matrice di calibrazione la matrice identità ed, in altre parole, è indipendente dai parametri di orientamento interno. Considerando dunque due matrici camera normalizzate, la prima in forma canonica $P = [I \mid \mathbf{0}]$ (normalizzata implicitamente), e la seconda $P' = [R \mid \mathbf{t}]$, è possibile definire una matrice concettualmente analoga alla matrice fondamentale per mezzo della (1.50)

$$E = [\mathbf{t}]_{\times} R \quad (1.103)$$

Tale matrice, che prende il nome di matrice essenziale, fornisce risultati analoghi alla fondamentale con l'unica differenza che opera su quantità (coordinate immagine) normalizzate ed è generata da matrici camera dipendenti unicamente dai parametri di orientamento interno; in particolare

$$\bar{\mathbf{x}}^T E \bar{\mathbf{x}} = 0 \quad (1.104)$$

in analogia alla (1.51) da cui è possibile ricavare il legame fra matrice essenziale e fondamentale per una stessa coppia di fotogrammi:

$$F = K'^T E K^{-1} \Rightarrow E = K'' F K \quad (1.105)$$

Per la matrice essenziale valgono quindi tutte le considerazioni ed i risultati ottenuti in § 1.5.2 e in § 1.5.3.

E' possibile dimostrare che una matrice, per essere compatibile algebricamente con una matrice essenziale, deve avere nuovamente rango 2 (come la fondamentale) e i suoi due valori singolari non nulli devono essere uguali: considerata infatti la matrice $[\mathbf{t}]_{\times}$ in (1.102) essa risulta, per costruzione, anti-simmetrica; si scomponga tale matrice per mezzo di

$$S = [\mathbf{t}]_{\times} = k \cdot UZU^T \quad \text{con } Z = \begin{vmatrix} 0 & 1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 0 \end{vmatrix} \quad (1.106)$$

in cui k è una costante di proporzionalità e la matrice U , in virtù della decomposizione agli autovalori ([Golub 1989]), è una matrice reale ortogonale; tale decomposizione può essere efficacemente affrontata per mezzo del metodo di Jacobi [Press 1988]. Si consideri poi la matrice

$$W = \begin{vmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{vmatrix} = -\text{diag}(1, 1, 0) Z \quad (1.107)$$

Sostituendo la (1.107) nella (1.106) è immediato verificare che, a meno di un fattore di scala

$$S = U\text{diag}(1, 1, 0)WU^T \quad (1.108)$$

da cui deriva, sostituendo nella (1.103) e ricordando che $S = [\mathbf{t}]_{\times}$

$$E = SR = k \cdot U\text{diag}(1, 1, 0)WU^T R = U (k \cdot \text{diag}(1, 1, 0))(R^T U W^T)^T = UDV^T \quad (1.109)$$

E' facile allora osservare che la quantità ad ultimo membro equivale ad una decomposizione ai valori singolari in cui i due valori singolari non nulli sono uguali a k .

La dimostrazione appena sviluppata ci permette di mettere in luce una proprietà importante della matrice essenziale: a differenza della matrice fondamentale che è caratterizzata dall'aver 7 gradi di libertà, la matrice essenziale, ha solo 5 gradi di libertà, ovvero quelli dell'orientamento relativo, in cui sono noti i parametri di orientamento interno delle camere. A questo punto, per ricostruire metricamente la scena, si determina una coppia di fotogrammi della sequenza che abbia un numero sufficiente di punti di legame (almeno una ventina) e che siano sufficientemente distanziate (abbiano cioè una base di presa sufficientemente ampia) lungo la sequenza; per mezzo dei punti di legame si può stimare, con uno dei metodi di § 1.5.2., la matrice fondamentale per poi determinare la matrice essenziale per mezzo della (1.105) o, alternativamente, si può stimare direttamente tale matrice normalizzando i punti (sempre con i metodi di § 1.5.2.). A questo punto si decompone tramite SVD la matrice essenziale; solitamente, a causa degli errori di misura sui punti

di legame, la matrice ottenuta non presenta esattamente due valori singolari uguali; per ovviare al problema si può semplicemente calcolare la media dei valori singolari ottenuti per mezzo della SVD e sostituirla alle quantità non corrette, senza introdurre particolari errori; per mezzo della (1.106) si scompone la matrice essenziale come $E = SR$: tale decomposizione porta a due possibili matrici di rotazione:

$$S = k \cdot UZU^T \rightarrow R = UWV^T \text{ oppure } R = UW^T V^T \quad (1.110)$$

in cui U e V sono le due matrici ortogonali ottenute dalla SVD (vd. equazione (1.109)) e Z e W sono le matrici introdotte nella (1.106) e (1.107).

E' immediato verificare, svolgendo il prodotto $E = SR$ che entrambe le (1.110) verificano l'uguaglianza; al tempo stesso si può dimostrare che esse sono le uniche fattorizzazioni possibili: la matrice di rotazione R può infatti essere scritta come $R = UXV^T$ in cui X è una generica matrice di rotazione; dalla (1.109) deriva immediatamente che

$$U \text{diag}(1, 1, 0) V^T = SR = (UZU^T)(UXV^T) = U (ZX) V^T \quad (1.111)$$

Ne consegue banalmente che, essendo X una matrice di rotazione e dovendo essere $ZX = \text{diag}(1, 1, 0)$, la matrice Z non può che essere uguale o a W o a W^T .

Per quanto riguarda il vettore \mathbf{t} dalla (1.106) abbiamo che $S = k \cdot UZU^T = [\mathbf{t}]_{\times}$; se $k = 1$ risulta $\|\mathbf{t}\| = 1$ che corrisponde a considerare una base di presa di lunghezza unitaria; dal momento che il sistema di riferimento nel quale viene eseguita la ricostruzione è arbitrario e, se non si dispone di punti d'appoggio o altre informazioni tale ricostruzione è definita a meno di un fattore di scala e una rotazione (cioè una trasformazione conforme), si può senz'altro porre $S = UZU^T$. Dal momento che $S\mathbf{t} = [\mathbf{t}]_{\times}\mathbf{t} = 0$ ne deriva che

$$\mathbf{t} = U (0, 0, 1)^T = \mathbf{u}_3 \quad \text{oppure} \quad \mathbf{t} = -U (0, 0, 1)^T = -\mathbf{u}_3 \quad (1.112)$$

In conclusione dalla matrice essenziale è possibile ricavare quattro matrici camera, e conseguentemente quattro differenti scelte di parametri di orientamento esterno, che corrispondono a

$$P_1 = [UWV^T \mid +\mathbf{u}_3], P_2 = [UWV^T \mid -\mathbf{u}_3], P_3 = [UW^T V^T \mid +\mathbf{u}_3], P_4 = [UW^T V^T \mid -\mathbf{u}_3] \quad (1.113)$$

Di tali configurazioni solo una risulta geometricamente valida in quanto, come mostrato in figura 1.24, le altre corrispondono a casi in cui uno qualsiasi dei punti di legame utilizzati, in coordinate oggetto, si trova dalla parte opposta del centro di proiezione rispetto al piano immagine come mostrato in [Hartley 1993].

A questo punto è possibile determinare per triangolazione diretta (essendo in uno spazio euclideo, si può usare anche un metodo basato sul punto medio) le coordinate terreno dei punti di legame. La posizio-

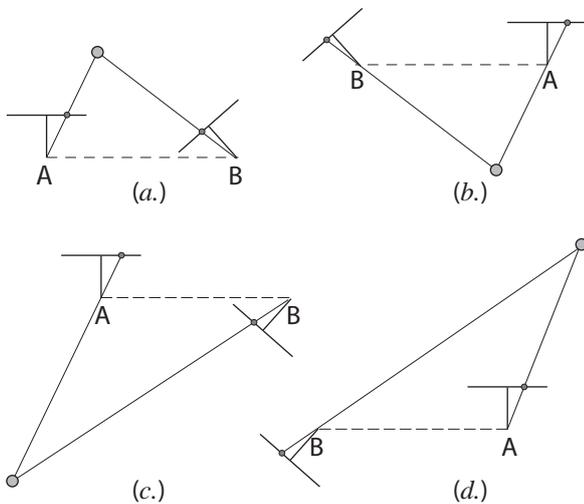


Figura 1.24: Le quattro diverse configurazioni derivanti dalla (1.112); è facile osservare che solo in una di esse (a.) il punto terreno si trova nel semi-spazio opposto del centro di proiezione rispetto alla positiva dell'immagine.

ne dei punti e dei restanti fotogrammi si ottiene combinando resezione (si veda appendice A.3) e triangolazione. Qualora il blocco sia costituito da un numero elevato di fotogrammi o si abbiano più strisciate è possibile applicare il metodo finora illustrato per ciascun sotto-blocco; al termine, grazie ai punti di legame fra i vari sottoblocchi si possono stimare le trasformazioni per portare ciascun modello multi-immagine (ovvero sotto-blocco) in un sistema di riferimento comune.

Al termine i risultati della procedura vengono utilizzati come parametri iniziali di un bundle block adjustment imponendo vincoli o informazioni a priori (punti d'appoggio, inter-distanze note etc.). Nel nostro caso il sistema viene risolto tramite CALGE ([Forlani 1986]).

La letteratura è comunque ricca di lavori riguardanti il soggetto (si veda ad esempio [Triggs 2000]). Tutti i passaggi procedurali illustrati nel presente paragrafo sono riassunti schematicamente in figura 1.25.

Nel paragrafo conclusivo verranno presentati alcuni dei risultati ottenuti per mezzo della nostra implementazione e i principali campi di applicazione finora affrontati.

Attualmente l'implementazione dei metodi è stata effettuata in un linguaggio non compilato (Matlab) che, come noto, è un tipico linguaggio di sviluppo, con prestazioni in termini di velocità di elaborazione molto limitate.

Si è comunque notato che i tempi di elaborazione cambiano notevolmente da caso a caso: dal momento che il numero di iterazioni richiesto da Ransac dipende strettamente dal numero di inlier presenti, il metodo è più efficiente, a parità di scena esaminata, su sequenze con basi di presa modeste, pur essendo in tal caso il numero di fotogrammi maggiore. La ragione è una maggior affidabilità del feature matching avvantaggiato da migliori condizioni di correlazione fra le diverse immagini, per cui il numero di outlier risulta estremamente contenuto; quando al contrario la base di presa assume valori elevati non solo si ha maggiore difficoltà ad individuare zone corrispondenti sui diversi fotogrammi, ma la maggior incertezza sul campo di parallassi porta ad avere un numero molto maggiore di accoppiamenti sbagliati.

L'altro parametro, che condiziona pesantemente i tempi di elaborazione, è il numero di features estratte su ciascun fotogramma, che può essere specificata inizialmente dall'operatore. In situazioni normali esso è dell'ordine di alcune centinaia di punti; in altri casi, dove ad esempio si hanno molti elementi potenzialmente estraibili, è necessario specificare un numero di estrazioni molto più elevato (dell'ordine di alcune

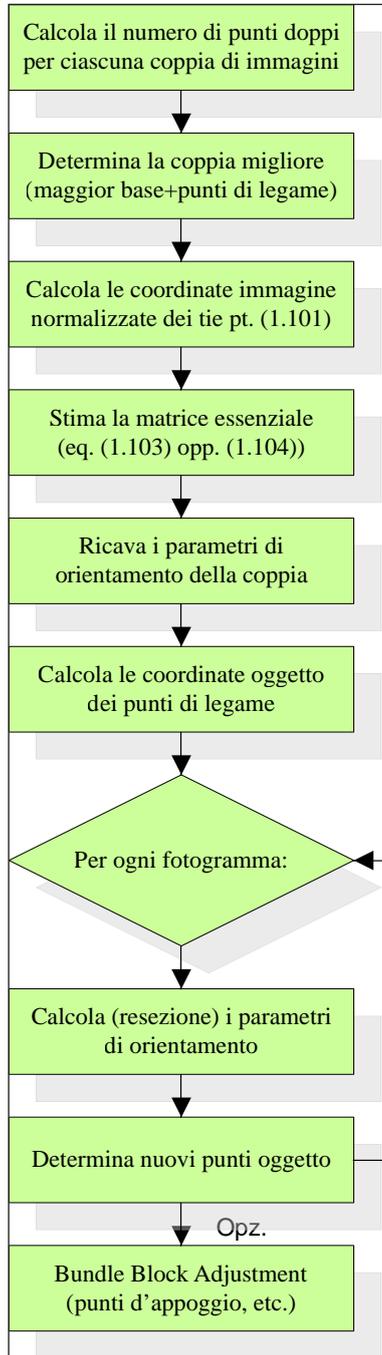


Figura 1.25: Schema delle operazioni implementate per la ricostruzione metrica della structure and motion.

migliaia) per garantire che su tutti i fotogrammi vengano estratti quasi sempre gli stessi elementi. Come termine di paragone, si consideri che l'esempio riportato in figura 1.26 nel prossimo paragrafo ha richiesto all'incirca mezz'ora di elaborazione per 90 fotogrammi e circa 15 minuti per elaborarne solo 18; in letteratura, per casi analoghi con implementazioni ottimizzate, sono citati tempi di qualche minuto.

Dal momento che, per le applicazioni attualmente sviluppate, non si ha la necessità di ottenere risultati in real-time o in tempi stretti, e che, come sperimentato, il passaggio da codice interpretato a codice compilato può portare a tempi di elaborazione anche di 10÷100 volte inferiori, non consideriamo come limitante l'attuale livello prestazionale del software sviluppato.

1.8. Simulazioni, Test e Applicazioni

Di seguito vengono presentate alcune delle simulazioni utilizzate per verificare la correttezza nell'implementazione e alcune delle, a nostro avviso, più interessanti applicazioni sviluppate con l'utilizzo della structure and motion: oltre ad essere utilizzato per orientare i blocchi utilizzati per il rilievo delle discontinuità di pareti rocciose, di cui si parlerà approfonditamente nel capitolo 4, le tecniche finora trattate sono anche state applicate in condizioni impegnative (scarsa caratterizzazione di tessitura degli oggetti) per alcuni gruppi scultorei del Pantheon di Parigi; infine si presenta una applicazione per la determinazione dei parametri e delle traiettorie di navigazione in ambito di rilievi ad alta produttività per mezzo di veicoli per il mobile mapping, che al momento rappresenta il caso di maggior complessità (scarsa tessitura, geometria di presa complessa, etc.) finora affrontato per mezzo della structure and motion.

1.8.1. Sequenze generate artificialmente

Per verificare la correttezza dell'implementazione delle tecniche precedentemente esposte, e analizzare le capacità del metodo in diverse condizioni operative, si sono innanzitutto realizzate una serie di simulazioni con immagini sintetiche (ovvero realizzate artificialmente) per avere maggior controllo sui dati in ingresso: le immagini (per un esempio vd. figura 1.26) sono state generate con un apposito software di modellazione tridimensionale. Oltre a poter controllare i parametri ambientali e le tecniche di generazione delle immagini, la geometria della scena non affetta da errore, i parametri di orientamento interno ed esterno sono noti, non si hanno effetti di distorsione ottica e la geometria degli oggetti viene ottenuta per mezzo di una mesh poligonale dalla quale è facile sia determinare punti d'appoggio sia valutare gli errori di restituzione una volta terminata la procedura.

Per uniformità si sono utilizzati in tutti i test gli stessi parametri di stima: in particolare si è considerata una soglia di accettazione per la geometria epipolare una distanza pari ad un pixel, in linea con il rendimento ottimale degli operatori di interesse utilizzati (vd. §1.2.5), mentre per la stima del tensore trifocale, considerando che in questo caso si vogliono eliminare solo le corrispondenze palesemente errate, si è utilizzata una soglia lievemente maggiore, pari a 1.5 pixel.



Figura 1.26: Immagine generata mediante l'utilizzo di un apposito software di modellazione per la verifica dei risultati ottenuti con i moduli sviluppati per la S&M.

La sequenza di figura 1.27, ha permesso di analizzare le precisioni di restituzione e di stima dei parametri di orientamento esterno in funzione di vari parametri, fra cui la lunghezza della base di presa e il numero di features inizialmente estratte: essa riprende un oggetto di grandezza approssimativamente pari ad un metro muovendosi su una circonferenza di raggio 125 cm centrata nel baricentro dell'oggetto, coprendo un settore di 90° . La fase di rendering delle immagini è stata impostata generando 90 immagini distanziate sulla circonferenza di un arco di 1° utilizzando una camera virtuale con distanza principale di 35 mm su un formato $24 \times 36\text{ mm}$, generando immagini con risoluzione di $1600 \times 1200\text{ pixel}$. Per ricostruire la scena al termine della procedura, in modo da determinare la trasformazione conforme per riportarsi al caso reale, si sono fissati i parametri di orientamento del primo, dell'ultimo e del fotogramma di mezzo della sequenza.

Con circa 750 punti per fotogramma, utilizzando tutte le immagini della sequenza (figura 1.27) si sono ottenuti i risultati illustrati in tabella 1.2 (*sim 01*); le precisioni di restituzione non sono ottimali, soprattutto se confrontate con quelle ottenute impiegando solamente $1/5$ dei fotogrammi (figura 1.28). Ciò è dovuto principalmente al fatto che nel primo caso la maggior parte dei punti non viene tracciata per più di $4 \div 5$ fotogrammi: essendo estremamente modesta la base di presa fra fotogrammi adiacenti (dell'ordine di $2 \div 3\text{ cm}$) gli angoli di intersezione sono decisamente piccoli e pertanto le precisioni di stima non sono elevate. Inoltre, il numero di fotogrammi collegati a quelli fissi solo dai punti di legame, è pari a 43, per cui si può avere una deriva della soluzione fotogrammetrica in corrispondenza dei fotogrammi mediani. Nel secondo caso (*sim 02*) è interessante osservare che il numero di punti ottenuti è solo lievemente inferiore all'altro, ma le preci-

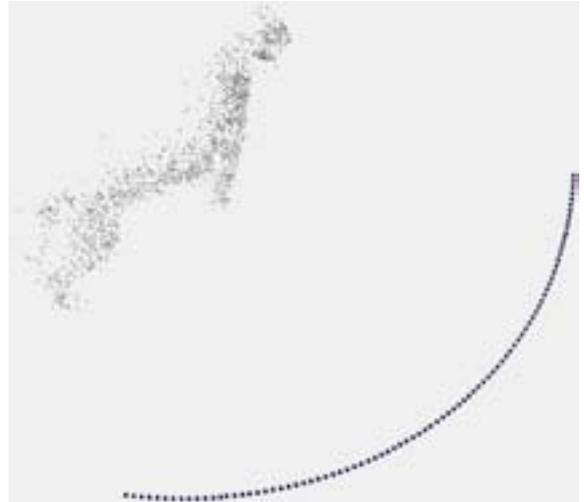


Figura 1.27: Posizioni di presa e punti restituiti utilizzando le 90 immagini della sequenza 1.26.

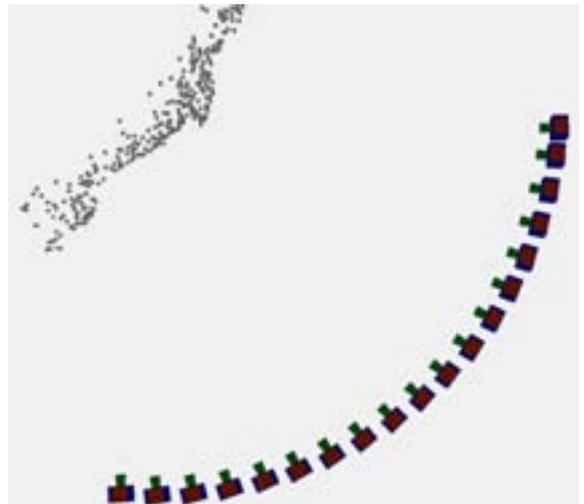


Figura 1.28: Posizioni di presa e punti restituiti utilizzando le 18 immagini della sequenza 1.26.

sioni sono decisamente più soddisfacenti: anche qui le basi di presa non sono particolarmente grandi (circa 10 cm) il che facilita il riconoscimento di punti omologhi con tecniche di correlazione semplice.

Blocco fotogrammetrico			Structure and motion		
	<i>sim 01</i>	<i>sim 02</i>		<i>sim 01</i>	<i>sim 02</i>
Caratteristiche blocco			Feature extraction		
Immagini	90	18	Pt. estratti (med)	689	723
Punti oggetto	1796	335	Feature Matching		
punti x imm. (min)	35	26	matching (med)	124.3	156.7
punti x imm. (max)	102	81	matching (min)	106	112
punti x imm. (med)	73	61	matching (max)	143	184
Molteplicità pt. (med)	4	4	Fondamentale (Ransac)		
Molteplicità pt. (max)	19	9	n.inlier (med)	106.4	103.2
Residui eq. di collinearità - per punti (pixel)			% inlier (med)	85.6%	65.9%
RMS (med)	0.32	0.3	Fondamentale (dopo matching guidato)		
RMS min delle medie	0.03	0.04	n. inlier (med)	114.8	115.1
RMS max delle medie	2.4	1.13	n. inlier (min)	96	88
RMS max	4.78	1.6	n. inlier (max)	146	145
Residui eq. di collinearità - per immagine (pixel)			% inlier (med)	92.4%	73.5%
RMS (med)	0.36	0.35	Trifocale (Ransac)		
RMS min delle medie	0.26	0.27	n.inlier (med)	43.0	41.2
RMS max delle medie	0.53	0.46	% inlier (med)	84.3%	67.2%
Errori di restituzione punti (cm)			Trifocale (dopo matching guidato)		
RMSE	1.13	0.312	n. inlier (med)	44.2	43.4
errore max	3.89	0.791	n. inlier (min)	27	20
errore min	0.01	0.007	n. inlier (max)	51	48
dev. Standard	1.07	0.308	% inlier (med)	86.7%	70.8%

Tabella 1.2: Riepilogo dei risultati ottenuti per mezzo dell'orientamento automatico nei due casi in cui si sono utilizzate immagini generate artificialmente.

Assumendo che camere adiacenti siano assimilabili a prese normali e supponendo una precisione nella collimazione σ_c pari a 0.5 pixel (ragionevole se confrontata con i risultati di tabella 1.2), si ottiene [Krauss 1992]

$$\sigma_z = m_b \frac{Z}{B} \sigma_{p_e} = \frac{Z^2}{cB} \sigma_{p_e} \quad (1.114)$$

con $\sigma_{p_e} = \sqrt{2} \sigma_c \cong 0.7 \text{ pixel} = 15.75 \mu\text{m}$

che corrisponde, considerando i fotogrammi centrali della sequenza, punti nel centro del formato in modo

da poter trascurare in prima approssimazione le componenti di errore in x e y , e ipotizzando una base di presa di 8 cm (corrispondente a quattro fotogrammi) nel primo caso a $\sigma_z = 9\text{ mm}$, mentre nel secondo, considerando una base media di circa 30 cm , a $\sigma_z = 2.3\text{ mm}$.

1.8.2. S&M applicata al rilievo di pareti rocciose

Come anticipato nell'introduzione, una delle applicazioni della structure and motion cui si è dedicato maggior spazio nel nostro lavoro è il rilievo fotogrammetrico di pareti rocciose per l'analisi di stabilità, allo scopo di permettere un facile orientamento delle prese anche a personale non particolarmente esperto nelle tecniche fotogrammetriche. Di seguito vengono brevemente illustrati due casi in cui si è applicato il metodo; in entrambi si è utilizzata una camera digitale Nikon D100 (risoluzione 3008×2000 , dimensione del pixel $7.8\ \mu\text{m}$) con ottica Nikkor da 18 mm , appositamente calibrata.

Nel primo [Roncella 2005c] il rilievo interessa un settore della Corma di Machaby (Valle d'Aosta) su un fronte di 150 m ed un'altezza di 90 m (per maggiori dettagli si rimanda al § 4.2).

Il blocco fotogrammetrico si sviluppa (vd. figura 1.30) approssimativamente in direzione parallela alla parete ad una distanza di circa 60 m ed è costituito da 17 immagini scattate ad una distanza variabile fra 3 m e 5 m , per un'estensione complessiva di 65 m . Essendo il primo caso pratico in cui si è impiegata la structure and motion per l'orientamento dei fotogrammi, si è preferito adottare basi di presa modeste per facilitare le fasi di correlazione di immagine: nella restituzione finale si sono invece utilizzate basi di presa più ampie, per una migliore intersezione dei raggi omologhi. Sono state estratte circa 800 features per fotogramma:



Figura 1.29: Una delle pareti della Corma di Machaby (Valle d'Aosta) utilizzate per analizzare le tecniche di structure and motion.



Figura 1.30: Posizioni di presa e punti restituiti utilizzando le 17 immagini della sequenza realizzata nel sito di Arnad.

complessivamente sono stati determinati 650 punti oggetto corrispondenti a circa 3500 punti immagine (mediamente ogni punto è visto quasi su 6 immagini).

Per il controllo e l'appoggio sono stati misurati sulla parete 37 punti per mezzo di un teodolite reflectorless; il rilievo è stato georeferenziato con due stazioni GPS.

Per l'orientamento assoluto del blocco fotogrammetrico sono stati utilizzati solamente 3 punti d'appoggio. I restanti sono stati collimati manualmente sui fotogrammi, una volta fissati i parametri di orientamento esterno, e sono stati usati come check point per verificare la precisione di restituzione. I risultati, riassunti in tabella 1.3, sono stati giudicati in linea con le precisioni attese.



Figura 1.31: Una delle pareti dell'ammasso roccioso che sovrasta l'eremo di Longebogne (Svizzera Vallese) utilizzata per analizzare le tecniche di structure and motion.

	Arnad (23 check point)			Longebogne (18 check point)		
	X (cm)	Y (cm)	Z (cm)	X (cm)	Y (cm)	Z (cm)
Media	3.1	-0.3	3.1	0.4	0.2	-0.1
RMSE	6	8.2	9.1	1.1	1.6	1.2
Dev. Std.	5.7	9.2	9.5	1.1	1.6	1.2

Tabella 1.3: Media, valore quadratico medio e deviazione standard delle differenze misurate dai check point nelle tre direzioni, nei due casi illustrati.

Il secondo esempio riguarda un'estensione di circa 30×25 m della parete che sovrasta l'eremo di Longebogne a Bramois (Svizzera Vallese): per maggiori dettagli si rimanda al § 4.3. L'esiguità degli spazi idonei ad effettuare prese ha dato vita ad un blocco (figura 1.32) costituito da tre strisciate parallele di 4, 5 e 6 fotogrammi ad una distanza media dalla parete rispettivamente di 48 m, 53 m e 57 m. La realizzazione di più strisciate ha principalmente lo scopo di aumentare la molteplicità dei punti ed evitare eventuali problemi di occlusione nella successiva generazione del DSM, e non tanto di irrobustire la geometria di presa, vincolata



Figura 1.32: Posizioni di presa e punti restituiti utilizzando le 16 immagini della sequenza realizzata nel sito di Longebogne.

al poco spazio a disposizione per la presa. Dopo la ricostruzione della structure and motion, sono stati rimisurati automaticamente tutti i punti con LSM vincolato geometricamente (MGCM); complessivamente sulle 15 immagini sono stati identificati 400 punti di legame (6000 punti immagine). Dei 27 punti d'appoggio collimati sulla parete tre sono stati utilizzati per la georeferenziazione del blocco, i restanti come check point. Le informazioni e i risultati relativi al blocco sono riassunti in tabella 1.4.

1.8.3. S&M applicata al rilievo architettonico

L'utilizzo della structure and motion può essere estremamente interessante in ambito architettonico o di conservazione dei beni culturali (monumenti, statue, siti archeologici, manufatti etc.): per saggiarne le capacità in alcune condizioni tipiche di questo ambito applicativo, ovvero la possibilità che porzioni assai rilevanti del fotogramma non presentino buone caratteristiche di tessitura, sono stati scelti due esempi.

Entrambi riguardano gruppi scultorei ripresi con la medesima camera Nikon D100 all'interno del Pantheon di Parigi. Purtroppo al momento non si dispone di un rilievo d'appoggio specifico per i due gruppi e pertanto la valutazione dei risultati può avvenire solamente in termini di errori di riproiezione sui fotogrammi e capacità del metodo di estrarre un sufficiente numero di elementi puntuali affidabili.

Nella prima sequenza (figura 1.33) il blocco è costituito da 12 fotogrammi ripresi ad una distanza dal gruppo scultoreo di circa 10 m muovendosi approssimativamente su un settore circolare di ampiezza di circa 90° (vd. figura 1.34), con basi di presa mediamente da 1÷2 m. Al termine della procedura si sono ottenuti 276 punti oggetto corrispondenti a 1049 punti immagine. I risultati ottenuti sono riassunti in tabella 1.4.

Il secondo caso risulta, in termini di tessitura, maggiormente difficoltoso rispetto al primo, come si può



Figura 1.33: Il primo dei due gruppi scultorei del Pantheon di Parigi utilizzato per le prove in ambito architettonico.

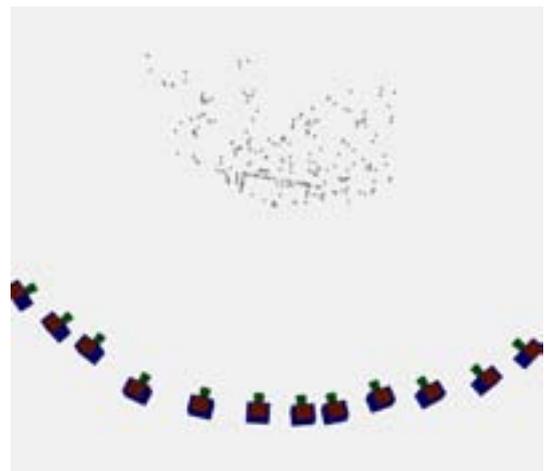


Figura 1.34: Posizioni di presa e punti restituiti utilizzando le 12 immagini della sequenza del gruppo scultoreo in figura 1.33.

vedere in figura 1.35. La conformazione del blocco (figura 1.36) è sostanzialmente analoga al caso precedente, (9 fotogrammi con basi di presa di circa 2 m). Utilizzando le immagini originali si sono ottenuti 184 punti oggetto corrispondenti a 648 punti immagine. Per cercare di aumentare il numero di corrispondenze si è riapplicata la procedura su immagini ricampionate per mezzo del filtro di Wallis (si veda § 2.4.2) in modo da aumentare il contrasto sul fotogramma, mettere in luce un numero maggiore di features e rendere più stabile il loro matching. Il filtro è stato applicato sulle immagini a colori e si è proceduto in un secondo momento alla loro trasformazione in toni di grigio, in modo da mettere in evidenza anche quelle componenti di contrasto visibili solamente in un canale radiometrico. Si è ottenuto quindi un nuovo blocco in cui sono presenti 258 punti immagine (con un incremento del 40%) e a 899 punti immagine (incremento pari al 39%). Grazie all'applicazione del filtro di Wallis, è decisamente aumentato il numero di corrispondenze estraibili ma non la loro molteplicità (ovvero la tracciabilità lungo la sequenza), che rimane infatti immutata. Il confronto fra le posizioni dei centri di presa stimati nei due casi evidenzia differenze medie di circa 1.5 cm, valore non del tutto trascurabile vista la scala fotogramma utilizzata; non avendo ancora a disposizione check point, non si può, al momento, stabilire il grado di precisione nei due casi, e quindi se l'uso del filtro di Wallis possa migliorare la ricostruzione.

Le informazioni e i risultati relativi ai due esempi sono riassunti in tabella 1.4 a pagina seguente.



Figura 1.35: Il secondo gruppo scultoreo utilizzato nella prova.

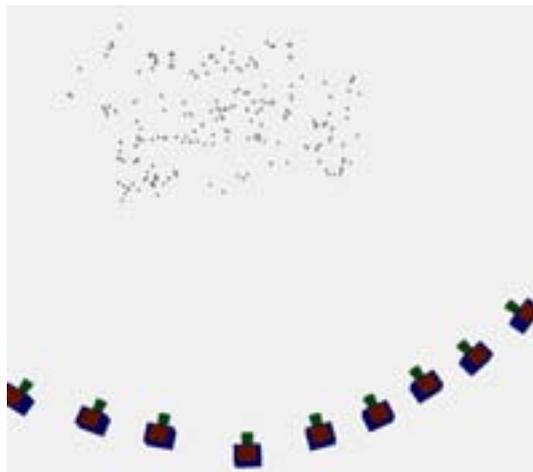


Figura 1.36: Posizioni di presa e punti restituiti utilizzando le 9 immagini della sequenza del gruppo scultoreo rappresentato in figura 1.35.

Blocco fotogrammetrico			Structure and motion		
	<i>statua 01</i>	<i>statua 02</i>		<i>statua 01</i>	<i>statua 02</i>
Caratteristiche blocco			Feature extraction		
Immagini	12	9	Pt. estratti (med)	759	765
Punti oggetto	276	258	Feature Matching		
punti x imm. (min)	33	51	matching (med)	348.3	301.3
punti x imm. (max)	146	131	matching (min)	318	204
punti x imm. (med)	87	97	matching (max)	367	352
Molteplicità pt. (med)	5	4	Fondamentale (Ransac)		
Molteplicità pt. (max)	9	8	n.inlier (med)	127.3	121
Residui eq. di collinearità - per punti (pixel)			% inlier (med)	36.5%	40.2%
RMS (med)	0.52	0.67	Fondamentale (dopo matching guidato)		
RMS min delle medie	0.06	0.16	n. inlier (med)	201.3	171
RMS max delle medie	1.28	1.54	n. inlier (min)	136	70
RMS max	1.6	2.29	n. inlier (max)	241	211
Residui eq. di collinearità - per immagine (pixel)			% inlier (med)	57.8%	56.8%
RMS (med)	0.59	0.74	Trifocale (Ransac)		
RMS min delle medie	0.51	0.64	n.inlier (med)	47.6	35.8
RMS max delle medie	0.65	1.01	% inlier (med)	40.8%	42.3%
Errori di restituzione punti			Trifocale (dopo matching guidato)		
RMSE	<i>nd</i>	<i>nd</i>	n. inlier (med)	62.8	41.3
errore max	<i>nd</i>	<i>nd</i>	n. inlier (min)	44	28
errore min	<i>nd</i>	<i>nd</i>	n. inlier (max)	83	46
dev. Standard	<i>nd</i>	<i>nd</i>	% inlier (med)	53.8%	48.8%

Tabella 1.4: Riepilogo dei risultati ottenuti per mezzo dell'orientamento automatico nei due casi considerati.

1.8.4. S&M applicata alla navigazione di veicoli per il mobile mapping

Uno dei principali problemi nelle applicazioni cinematiche GPS di precisione è la perdita del segnale in prossimità di alberi o edifici, nell'attraversamento di sottopassi, etc. Per eliminare questa vulnerabilità e per una miglior determinazione di *pitch* e *roll* e *yaw*, nel Mobile Mapping si affianca al GPS un sistema di navigazione inerziale. Spesso le interruzioni sono molto brevi (meno di 100-200 m): in tali casi, una alternativa al sistema inerziale è orientare le immagini con una strategia puramente fotogrammetrica. In questo caso, ovviamente, è impensabile utilizzare approcci manuali per l'orientamento del blocco fotogrammetrico, in quanto richiederebbe tempi di lavoro inaccettabili. Inoltre la disponibilità di sequenze orientate automaticamente con la S&M consente, ai veicoli che dispongono già di un sistema di navigazione, di migliorare la

stima dei parametri di orientamento ricavati dai dati GPS, se applicata in modalità ISO (Integrated Sensor Orientation).

Il metodo si articola in due fasi ([Roncella 2005a], [Forlani 2005]): una fase di orientamento relativo automatico “per concatenamento” della sequenza di immagini, per mezzo della structure and motion, ed un orientamento assoluto finale “a sbalzo” della sequenza, partendo dai dati calcolati con GPS agli estremi. Le principali difficoltà (che rappresentano al contempo l’aspetto di maggior interesse dal punto di vista teorico) sono dovute alle condizioni di presa sfavorevoli, alla tessitura delle immagini, alla propagazione degli errori di stima.

L’integrazione di dati fotogrammetrici e osservazioni GPS per l’orientamento di sequenze in applicazioni di Mobile Mapping è stata proposta da diversi autori negli ultimi dieci anni: in [Da Silva 1998] e [Da Silva 2003] è presentato lo sviluppo di un veicolo in cui l’uso della fotogrammetria è fondamentale per il posizionamento e l’orientamento dei fotogrammi. In [Crosilla 1998] sono proposti a più riprese metodi per la ricerca di punti omologhi e di punti d’appoggio, riproiettando sulle immagini punti di coordinate note da Cartografia Numerica e GIS. Il gruppo di ricerca della scuola di Calgary si è occupato dell’utilizzo congiunto di dati IMU/GPS e fotogrammetrici per l’orientamento automatico delle sequenze, in particolare durante le perdite di segnale da parte del sistema di posizionamento (si veda a riguardo [Chaplin 1998], [Tao 1999]). Sebbene a tutt’oggi l’argomento riceva considerevole attenzione ([Kim 2004], [Chaplin 2001]), nessuna delle soluzioni proposte sembra essere divenuta realmente operativa su un veicolo rilevatore.

Una sequenza di MM ha caratteristiche che rendono difficile la ricostruzione della S&M: i fotogrammi presentano notevoli variazioni di scala immagine; anche con un intervallo spaziale di acquisizione delle immagini piuttosto ridotto (2-3 m), i punti omologhi nelle immediate vicinanze del veicolo presentano parallassi e deformazioni prospettiche elevate; la presenza di oggetti in movimento sul corpo stradale genera false corrispondenze su immagini asincrone nonché numerose occlusioni della scena, rendendo difficoltoso il tracciamento dei punti nella sequenza. Anche la geometria di presa è tutt’altro che ottimale (vd. figura 1.37): le basi di presa risultano estremamente piccole se confrontate con la profondità della scena (nei dati da noi elaborati, la base è di circa 1.70 m mentre alcuni dei punti tracciati nella S&M si trovano anche a 200 m dal veicolo); gli angoli di convergenza dei raggi omologhi risultano pertanto generalmente bassi e portano ad una scarsa accuratezza nella stima degli orienta-

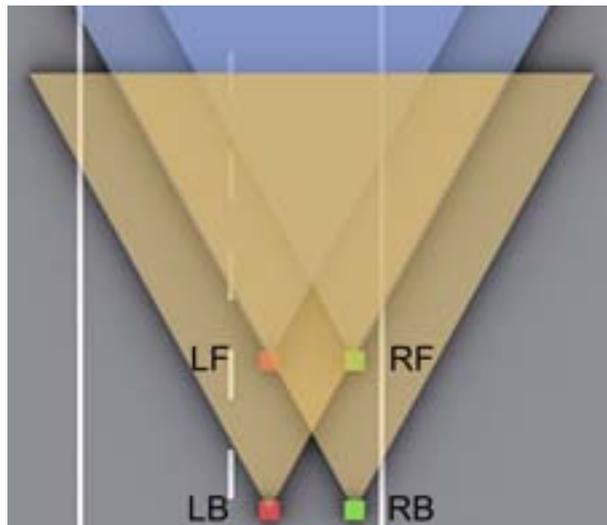


Figura 1.37: Schema del blocco elementare in applicazioni mobile mapping.

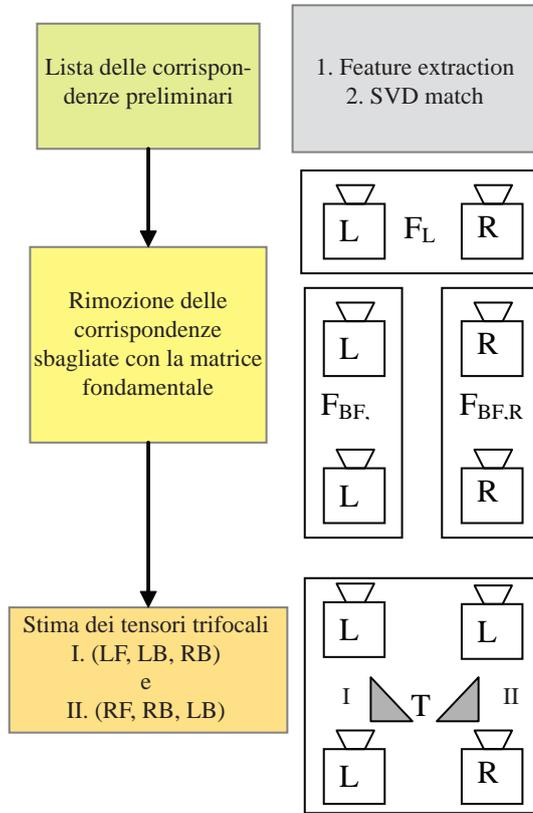


Figura 1.38: Strategie per la structure and motion applicata al MM.

menti e delle coordinate oggetto dei punti.

In un veicolo con camere frontali stereo, una sequenza da orientare può essere schematizzata come una doppia strisciata con sovrapposizione longitudinale di circa il 60-70% (variabile in funzione della velocità del veicolo) e con sovrapposizione trasversale di circa l'80-90%. Gli orientamenti esterni della prima e dell'ultima coppia stereoscopica della sequenza sono noti dal sistema GPS, essendo il tratto di sequenza costruito proprio dalla porzione non orientata e da due coppie di immagini ancora orientate via GPS (le ultime prima dell'interruzione e le prime orientate con successo appena riacquisito il segnale). Indichiamo con L=left e R=right le camere destra e sinistra, con B=backward e F=forward immagini successive lungo la sequenza (figura 1.38). Considerando due coppie successive di immagini, si nota (figura 1.39 a pagina seguente) che per coppie BF le linee epipolari tendono a coincidere con le linee di fuga principali della scena (si pensi per semplicità alle strisce stradali) mentre quelle fra coppie LR si dispongono parallelamente all'orizzontale. In

quest'ultimo caso non è infrequente che punti scelti ad esempio sulla segnaletica orizzontale su un lato della strada vengano impropriamente accoppiati con punti della striscia sull'altro lato, perché caratterizzati da valori del coefficiente di correlazione lineare elevato; in ogni caso, la notevole differenza prospettica fra le due prese e i valori di parallasse elevati rendono il processo di matching preliminare più difficoltoso rispetto all'altra configurazione. Ciononostante, le coppie LR permettono una stima più stabile e controllata della geometria epipolare e forniscono, in virtù della maggior sovrapposizione dei fotogrammi anche nella parte bassa del formato, una miglior distribuzione dei punti di legame.

Le caratteristiche delle due configurazioni di presa LF e BF risultano pertanto complementari: l'utilizzo del tensore trifocale permette l'unione delle due in una geometria ottimale, entro i limiti imposti dal problema: nelle coppie BF vengono determinati un maggior numero di accoppiamenti corretti, mentre la geometria LR permette l'eliminazione di accoppiamenti ambigui e stabilizza la ricostruzione almeno nei punti relativamente vicini.

Per ottenere una configurazione simmetrica vengono prese in considerazione due triplette distinte per ciascun istante della sequenza: viene dapprima stimata la geometria trifocale fra i fotogrammi (facendo



Figura 1.39: Geometria epipolare (a.) fra coppie di fotogrammi asincroni (coppie BF) e (b.) fra coppie sincrone (coppie LR).

referimento alla notazione in figura 1.38) LB-RB-LF e in seguito la geometria fra LB-RB-RF. Vengono considerati attendibili solo gli accoppiamenti ritenuti corretti in entrambe le triplette. In altre parole, per la successiva ricostruzione della S&M vengono utilizzati solamente punti quadrupli: sebbene il vincolo imposto tenda a ridurre considerevolmente il numero complessivo di matching (la zona di sovrapposizione dei quattro fotogrammi risulta assai ridotta) la loro affidabilità risulta estremamente elevata. La sequenza viene perciò suddivisa in “blocchetti” elementari, costituiti da due coppie stereoscopiche consecutive: ciascun blocchetto risulta legato al successivo condividendo due dei quattro fotogrammi che lo compongono. La dimensione del blocchetto risulta piuttosto regolare in quanto un lato coincide con la distanza fra le due camere (circa 1.70 m), mentre l’altro, dipendente dal frame rate, si attesta su una distanza variabile da 3 a 5 metri.

Per ottenere una maggior stabilità di stima è possibile, come illustrato in [Forlani 2005], utilizzare nel bundle adjustment finale il vincolo aggiuntivo rappresentato dalla geometria relativa fra fotogrammi sincrni, la quale può essere stimata inizialmente con un apposito processo di calibrazione.

Di seguito vengono presentati i risultati relativi alla sequenza più impegnativa finora elaborata: per mettere alla prova l'intero sistema si è deciso di analizzare una sequenza ripresa in ambito extraurbano, in cui sono minori le possibilità di individuare punti con buona tessitura e in cui la velocità tenuta dal veicolo, e di conseguenza le differenze prospettiche fra i fotogrammi, sono più elevate. La sequenza è costituita da 91 coppie di fotogrammi sincroni, ripresi con due camere digitali Basler Af101 (risoluzione 1300x1030, dimensione pixel $6.7 \mu\text{m}$) dotate di un ottica con distanza principale di circa 8 mm . Essa si sviluppa per circa 285 m lungo un tratto di strada con curva e rettifili (si veda a riguardo figura 1.40). Le due camere montate frontalmente sul tetto del veicolo

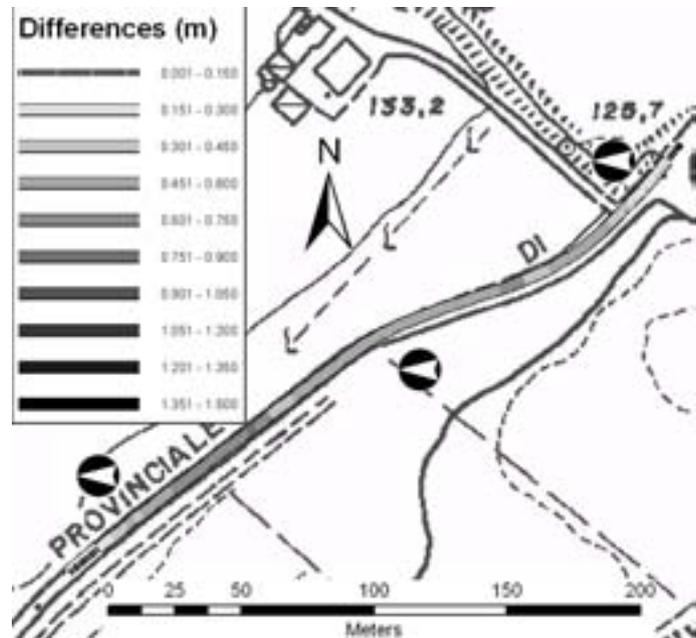


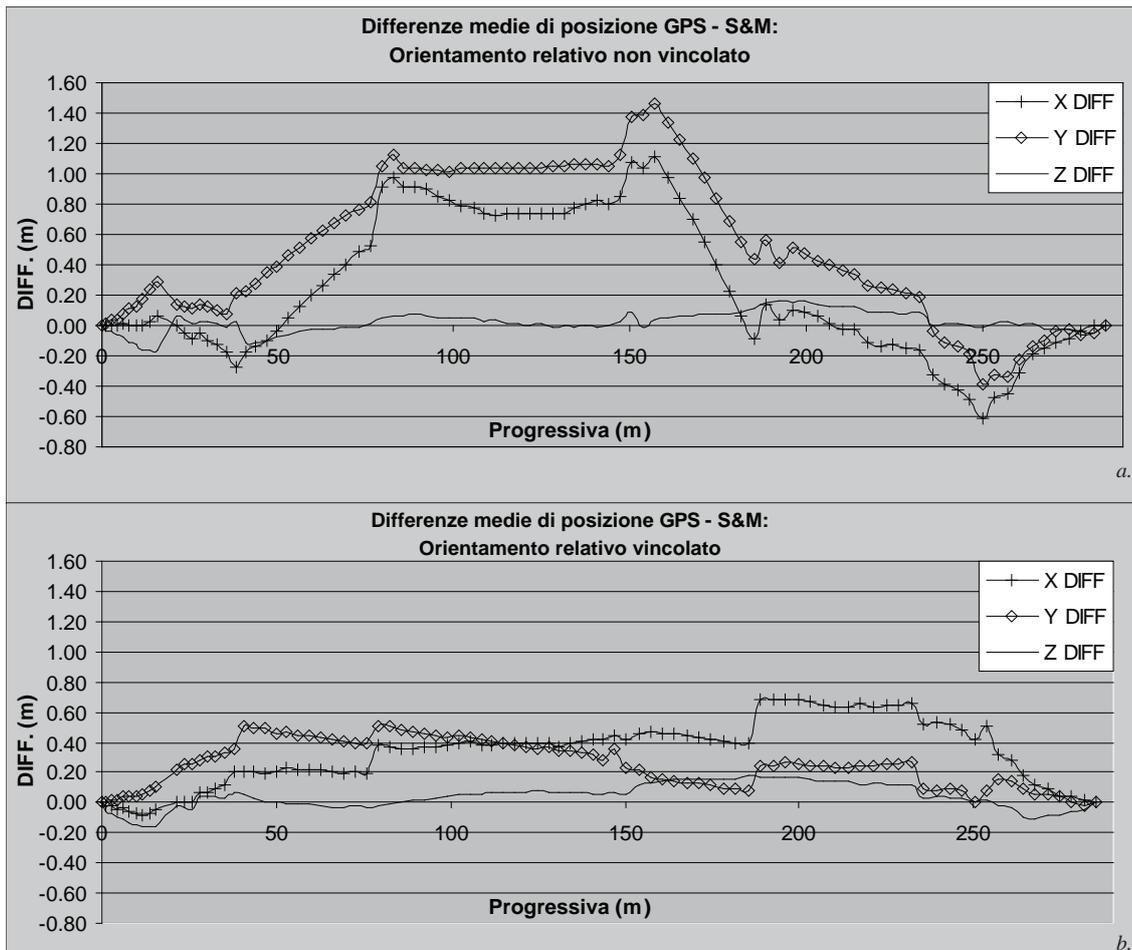
Figura 1.40: Percorso seguito dal veicolo nella prova descritta: in scale di grigio vengono rappresentate le differenze riscontrate fra la soluzione di navigazione GPS e quella ottenuta con la structure and motion vincolando la geometria relativa.

sono ad una distanza di 1.70 m ; la distanza media fra coppie asincrone è di circa 3.15 m . Complessivamente sono stati tracciati 1933 punti terreno (di cui 300 sono stati esclusi dall'algoritmo perchè generanti intersezioni troppo deboli) corrispondenti all'incirca a 7400 punti immagine; mediamente in ciascuna immagine sono presenti 34 punti immagine per una copertura del formato pari al 56%.

Particolare interesse assume il confronto fra i risultati ottenuti con e senza l'imposizione del vincolo relativo fra immagini sincrone, riassunti in tabella 1.5 e nei grafici di figura 1.41: è facile innanzitutto osservare che in entrambi i casi, essendo l'intero blocco appoggiato solamente sulla prima e ultima coppia di fotogrammi, gli errori di posizionamento risultano maggiori nella parte centrale del blocco in cui la soluzione è più sensibile ad effetti di deriva; ciononostante nel caso della soluzione non vincolata (a.) tali errori, oltre ad assumere valori circa tripli (relativamente al posizionamento planimetrico) rispetto al caso vincolato (b.), tendono ad aumentare in maniera vistosa per un numero maggiore di fotogrammi: in altre parole gli effetti di deriva continuano ad esercitare la loro influenza per tutta la durata della sequenza; nel caso opposto, seppur come ovvio l'errore tenda ad aumentare in direzione del centro del blocco, la crescita di tale errore è più evidente nei primi/ultimi $40 \div 50$ fotogrammi, mantenendosi in seguito all'incirca costante. Il vincolo geometrico sull'orientamento relativo è quindi efficace nel limitare gli effetti di deriva.

	Senza Vincolo Geometria Relativa			Con Vincolo Geometria Relativa		
	ΔX (m)	ΔY (m)	ΔZ (m)	ΔX (m)	ΔY (m)	ΔZ (m)
Media	0.24	0.50	0.02	-0.32	-0.25	-0.03
Dev. Std.	0.45	0.48	0.07	0.22	0.16	0.08
RMSE	0.51	0.69	0.07	0.39	0.30	0.08
Max	1.15	1.49	0.21	0.70	0.51	0.20
	$\Delta\omega$ (deg)	$\Delta\phi$ (deg)	$\Delta\kappa$ (deg)	$\Delta\omega$ (deg)	$\Delta\phi$ (deg)	$\Delta\kappa$ (deg)
Media	-0.43	-0.71	-0.18	0.10	-0.12	-1.68
Dev. Std.	0.62	0.57	1.67	0.51	0.14	1.57
RMSE	0.75	0.91	1.67	0.52	0.19	2.30
Max	1.71	1.63	3.64	1.36	0.51	4.73

Tabella 1.5: Media, deviazione standard, errore quadratico medio e valore massimo degli errori sui parametri di orientamento esterno nel caso di bundle block adjustment finale senza vincolo sugli orientamenti relativi (a sinistra) e con vincolo sugli orientamenti relativi (a destra).



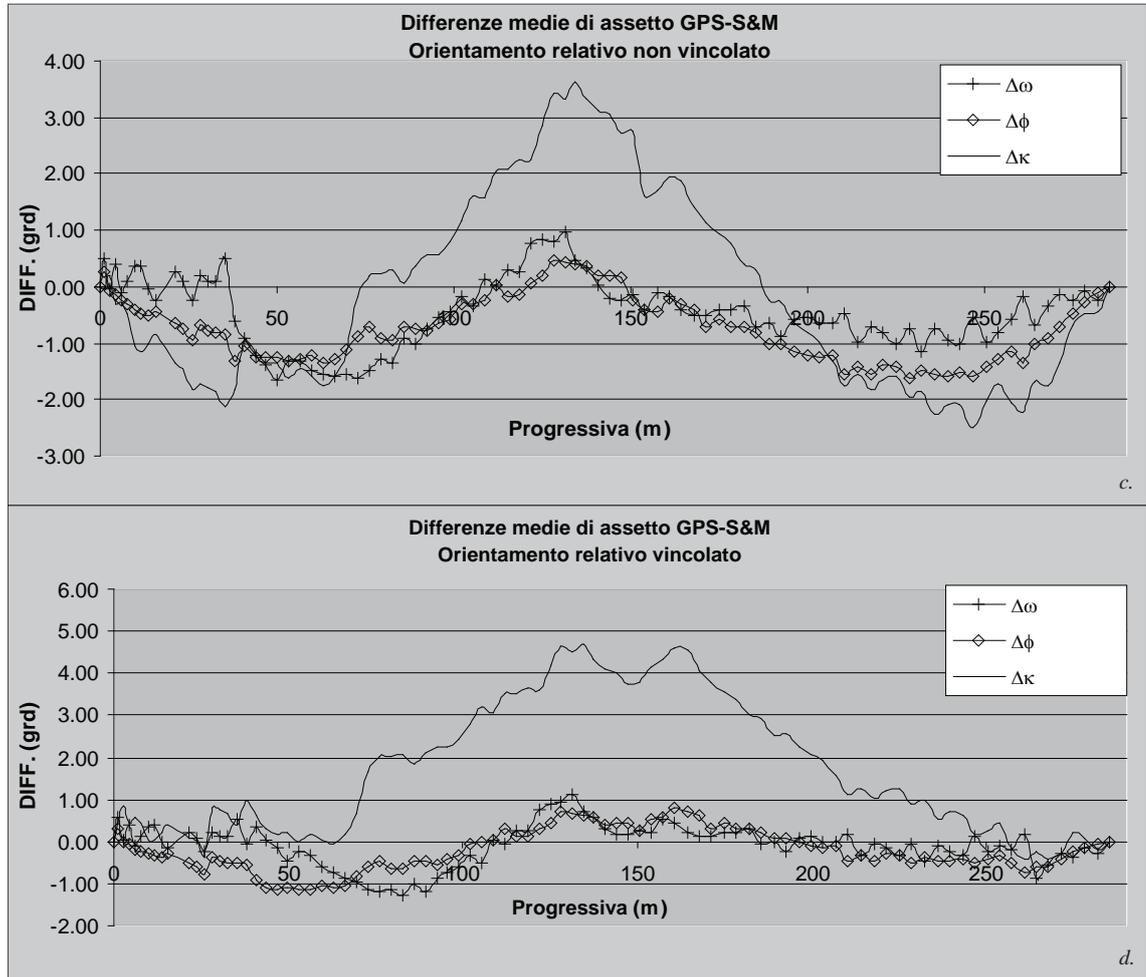


Figura 1.41: Differenze fra soluzione di navigazione GPS e quella ottenuta tramite structure and motion: (a.) differenze medie di posizione in X, Y e Z per il caso non vincolato; (b.) differenze medie di posizione in X, Y e Z per il caso vincolato; (c.) differenze medie di assetto in ω , ϕ e κ per il caso non vincolato; (d.) differenze medie di assetto in ω , ϕ e κ per il caso vincolato.

Infine in [Roncella 2005a] viene mostrato come l'utilizzo di operatori meno suscettibili agli improvvisi cambiamenti del punto di vista (operatore di Lowe), permetta di risolvere porzioni di traiettoria in cui il veicolo affronti ad esempio curve con raggi di curvatura ristretti: in tale frangente infatti gli altri operatori non conseguono una ripetibilità sufficiente, impedendo di individuare punti di legame in numero e disposizione sufficiente ad una buona stima dei parametri di orientamento.

CAPITOLO 2

Algoritmi e tecniche di matching area-based

In prospettiva, l'elevato grado di automazione oggi implementabile in diverse fasi del processo fotogrammetrico, e più in generale nelle scienze geomatiche, non può prescindere dallo sviluppo di tecniche di riconoscimento ed interpretazione dei dati altrettanto automatizzate, ovvero con minimo intervento di un operatore. E' quindi essenziale garantire la robustezza delle procedure di stima, l'efficienza computazionale delle operazioni e il grado di intelligenza interna al processo nel riconoscere e classificare i diversi elementi presenti sulle immagini. L'impiego sempre più diffuso dei database richiede che le informazioni siano interpretate ma la quantità di dati da inserire o aggiornare preclude spesso un approccio esclusivamente manuale alla restituzione. Si pensi all'evoluzione del laser a scansione negli ultimi anni: seppur ancora poco diffusa come tecnologia a causa dei costi di ammortamento della strumentazione e degli applicativi software necessari per processare i dati, l'interesse da parte degli specialisti sia in close-range sia in ambito territoriale e il notevole impiego di risorse da parte dei centri di ricerca dimostra che la tecnologia è ormai matura. Se solo qualche anno fa si manifestavano alcune riserve, principalmente per la difficoltà di gestire e rielaborare l'elevatissima mole di dati che lo strumento è in grado di fornire, oggi tali problemi sono, se non superati, almeno molto ridotti: sono presenti sul mercato diversi applicativi software in grado di gestire progetti costituiti da milioni o decine di milioni di punti e sono sempre più efficienti gli algoritmi per la semplificazione e segmentazione di tali dati. Esiste, a nostro avviso, la tendenza da parte di una certa classe di utilizzatori, forse affascinata dalle potenzialità del metodo, a sottovalutare le problematiche della precisione di misura, della semplicità di utilizzo del prodotto finale, della progettazione del rilievo, del controllo e dell'affidabilità dei risultati. In altre parole ci sembra, soprattutto in campo close-range, che tali tecniche vengano scelte a volte in maniera non sufficientemente ragionata, affrontando costi e complessità delle operazioni non commisurati ai reali benefici. Non si è forse ancora sviluppata una sensibilità nel valutare criticamente i risultati e non è ancora presente una casistica sufficientemente ampia per poter definire chiaramente metodologie di controllo e di collaudo. Ciononostante è palese lo sforzo da parte del settore produttivo, soprattutto del mondo della ricerca, di arrivare ad una definizione chiara e completa di procedure, metodologie e criteri di standardizzazione. In tale panorama la fotogrammetria classica, soprattutto

in ambiente close-range, viene da alcuni ritenuta una tecnologia superata o quantomeno non competitiva col laser a scansione; altri, con maggior equilibrio, vedono le due tecnologie come complementari e prefigurano lo sviluppo di nuovi prodotti. Dal punto di vista hardware, l'integrazione con un sistema laser di fotocamere digitali aumenta marginalmente i costi e non complica particolarmente le operazioni di rilievo; analogamente l'efficienza in campo di restituzione può sfruttare efficacemente i punti di forza di entrambe le tecnologie portando ad un flusso operativo più snello ed efficiente.

Nonostante ciò, a nostro avviso, la fotogrammetria è oggi in grado di fornire, da sola, in molti ambiti applicativi differenti, soluzioni paragonabili a quelle ottenibili con strumentazione più sofisticata (e costosa): visti i progressi di tipo tecnologico e metodologico in campo fotogrammetrico, esistono già da quasi un decennio, le possibilità per ottenere una descrizione geometrica dell'oggetto analizzato costituita da una moltitudine di punti (point cloud), in tutta analogia a quanto si è in grado di ottenere da un laser a scansione. Ovviamente una restituzione tridimensionale per mezzo della fotogrammetria richiede sempre un approccio stereoscopico o multi-immagine: con questo vogliamo sottolineare che, la scelta di utilizzo della fotogrammetria, è sempre più difficoltosa, in termini operativi, rispetto ad altre tecnologie più dirette. Tuttavia, come si vedrà fra breve, la fotogrammetria ha anche degli indubbi vantaggi rispetto agli altri metodi, non soltanto in termini di costi.

Il presente capitolo affronta le tecniche attualmente analizzate e sviluppate per ottenere una descrizione fitta della geometria di un oggetto, determinando una elevata mole di punti sulla sua superficie. Tali tecniche sono basate sostanzialmente su criteri di correlazioni (*matching*) di immagine, ovvero su metodi per determinare zone corrispondenti su fotogrammi differenti. Generalmente si determinano, su una o più immagini di riferimento (*master image*), che raffigurino l'oggetto da analizzare, una serie di punti di cui si vogliono conoscere le coordinate spaziali; nell'intorno di ciascuno di tali punti viene poi considerata una finestra di dimensioni prefissate (*template*) che viene messa in correlazione con una analoga finestra (*patch*) sulle restanti immagini (*slave image*), cercando di ottenere una stima della posizione che presenta il massimo grado di similitudine.

Le tecniche per effettuare tale operazione sono diverse, e nella presente si vanno ad analizzare con particolare attenzione quelle attualmente implementate; tratto comune di tutte le diverse metodologie è comunque il fatto che, per garantire buoni risultati in termini di precisione ed affidabilità, le immagini sulle quali vanno ad operare devono presentare un buon grado di contrasto: in altre parole le zone che si desidera mettere in correlazione devono presentare gradienti dei toni radiometrici piuttosto marcati, per favorire una più semplice localizzazione. A tal riguardo si andranno ad analizzare le tecniche di pre-processamento delle immagini che consentono di ottenere fotogrammi ricampionati da cui sia più semplice ottenere informazioni coerenti e corrette.

Inoltre si andranno ad analizzare le metodologie finora sviluppate per applicare ricorsivamente tali algoritmi di matching, in modo da ottenere la descrizione densa (in termini di distanza fra punti adiacenti) della geometria dell'oggetto che si vuole analizzare.

Anche in questo caso, infatti, è possibile ottenere procedure altamente automatizzate, in cui l'intervento di un operatore umano è di fatto eliminato: seppur in maniera meno diretta, dunque, e richiedendo in alcuni casi un maggior controllo, le tecniche fotogrammetriche che si andranno ad analizzare permettono di ottenere risultati equivalenti a quelli ottenibili mediante strumenti più sofisticati (laser scanner).

Si vedrà inoltre come, in alcuni casi, si possa utilizzare nelle procedure di matching, informazioni sulla geometria di presa dei fotogrammi, per rendere il processo più stabile e preciso: in tal caso è possibile utilizzare i dati ottenuti dall'orientamento automatico del blocco visto nel precedente capitolo, in modo da avere una soluzione integrata in cui, ancora una volta, l'intervento umano richiesto sia ridotto ai minimi termini.

Al termine del capitolo vengono analizzate alcune simulazioni numeriche ed una serie di test utilizzati per controllare la correttezza dei metodi implementati e verificare le effettive capacità degli algoritmi. Al tempo stesso si andranno ad illustrare alcune applicazioni in cui si fa largamente uso di queste tecniche di correlazione, ciò a dimostrazione di come esse siano da considerarsi un tema di grande interesse e di grandissima attualità. Si vedrà infatti come le metodologie fotogrammetriche, seppur meno costose, non siano da meno a molte altre tecniche, e permettano, grazie anche alle risorse di calcolo oggi disponibili, di sviluppare applicazioni fino a pochi decenni fa assolutamente impensabili.

Si vedrà infine, nel prossimo capitolo, come i risultati ottenuti per mezzo della correlazione di immagini possano costituire la base per una serie di ulteriori analisi, che permettono lo sviluppo di soluzioni complesse, assolutamente integrate, completamente automatizzabili, e decisamente competitive.

2.1. Metodi di correlazione semplice

I metodi di matching *area based* più semplici rientrano nella categoria della correlazione semplice: spostando con passi regolari sull'immagine di lavoro (*slave*) una finestra di dimensioni prefissate, si cerca il massimo di una misura di somiglianza dei toni di grigio dei due fotogrammi.

E' noto che, salvo condizioni particolari, la precisione di questi metodi è limitata. Tuttavia essi rimangono attuali perchè estremamente veloci nel determinare corrispondenze su due immagini: il numero di operazioni richieste per il calcolo dei descrittori statistici è infatti molto minore rispetto ad altre tecniche di matching. Pertanto, quando l'area di ricerca sull'immagine di lavoro è particolarmente estesa e sono quindi necessari numerosi tentativi prima di ottenere la corrispondenza corretta o nelle applicazioni in real time essi sono largamente utilizzati.

Le misure di correlazione più comunemente usate in fotogrammetria e in computer vision sono:

Somma delle differenze quadratiche (Sum of Squared Differences o SSD)

$$SSD(i, j) = \sum_{m=0}^M \sum_{n=0}^N (A(i + m, j + n) - S(m, n))^2 \quad (2.1)$$

Correlazione normalizzata (Normalized Cross-Correlation o NCC)

$$NCC(i, j) = \frac{\sum_{m=0}^M \sum_{n=0}^N (A(i+m, i+n) B(m, n))}{\sqrt{\sum_{m=0}^M \sum_{n=0}^N A(i+m, i+n)^2 \sum_{m=0}^M \sum_{n=0}^N B(m, n)^2}} \quad (2.2)$$

Coefficiente di correlazione (Zero Mean Normalized Cross Correlation o ZNCC)

$$ZNCC(i, j) = \rho(i, j) = \frac{\sum_{m=0}^M \sum_{n=0}^N ((A(i+m, i+n) - \bar{A})(B(m, n) - \bar{B}))}{\sqrt{\sum_{m=0}^M \sum_{n=0}^N (A(i+m, i+n) - \bar{A})^2 \sum_{m=0}^M \sum_{n=0}^N (B(m, n) - \bar{B})^2}} \quad (2.3)$$

in cui A e B rappresentano i valori dei toni di grigio rispettivamente della patch e della template e \bar{A} e \bar{B} i valori medi di tali grandezze.

E' importante osservare che solo la (2.3) risulta invariante nei confronti di variazioni di luminosità e contrasto fra template e patch, mentre le prime due grandezze possono essere affette in maniera minore (la (2.2)) o maggiore (la (2.1)) da cambiamenti di illuminazione. Il maggiore carico computazionale richiesto dalla (2.3) non è fra l'altro eccessivo ed è per questo motivo che nel seguito verrà considerato solo questa tipologia di correlazione semplice. Esistono poi altre grandezze che possono essere utilizzate per valutare il grado di similitudine fra due porzioni di immagini (si veda a riguardo [Moravec 1977], [Faugeras 1990] o [Aschwanden 1992]). Al fine di ottenere l'invarianza, anche solo parziale, rispetto a trasformazioni geometriche quali rotazioni o variazioni di scala, diversi autori suggeriscono di utilizzare nell'applicazioni degli algoritmi non l'immagine originale ma quella ottenuta per mezzo di appropriate convoluzioni [Martin 1995]: si può ad esempio dimostrare facilmente che l'immagine derivata dalla norma del gradiente in ciascun punto si dimostra invariante rispetto alla rotazione. Quando fra la template e la patch intervengono variazioni geometriche di una certa entità (è sufficiente che il punto di vista cambi in maniera non trascurabile) i valori di similitudine misurati dalla correlazione semplice sono bassi, anche quando il punto omologo sull'immagine slave viene individuato correttamente. Applicato in via preliminare, tuttavia, esso fornisce posizioni iniziali sufficientemente accurate per inizializzare altri metodi di matching più raffinati. In altre parole i metodi di correlazione possono essere utilizzati, vista la loro rapidità di esecuzione, per analizzare ampie regioni di immagine alla ricerca di posizioni omologhe; una volta individuate approssimativamente tali zone è un altro algoritmo a individuare con precisione la posizione del punto.

Il numero di operazioni per il calcolo del coefficiente di correlazione cresce con il quadrato della dimensione della template; inoltre il numero di tentativi da effettuare nella ricerca è direttamente proporzionale al numero di pixel che compongono la regione analizzata. Diversi studi sono allora stati compiuti per ottimizzare tali operazioni di ricerca. Di seguito presentiamo l'approccio di Lewis [Lewis 1995], implementato

nel nostro software.

Si consideri l'equazione (2.3) e si indichi rispettivamente con $A'(m, n) = A(m, n) - \bar{A}$ e con $B'(m, n) = B(m, n) - \bar{B}$ le funzioni che descrivono patch e template rimuovendo la media; il numeratore della (2.3) può allora essere espresso come:

$$\rho^{num}(i, j) = \sum_{m=0}^M \sum_{n=0}^N A'(i + m, i + n) B'(m, n) \quad (2.4)$$

Per il suo calcolo, considerando una dimensione della template pari a M^2 su una finestra di ricerca di dimensione N^2 , sono richieste approssimativamente $N^2(M-N+1)^2$ operazioni di somma e $N^2(M-N+1)^2$ moltiplicazioni. Al tempo stesso la (2.4) può essere vista come una convoluzione della template per mezzo della maschera $A'(m, n)$ e può essere calcolata per mezzo di una trasformata di Fourier. Per calcolare la trasformata i moderni algoritmi di FFT richiedono approssimativamente $18M^2 \log_2 M$ somme e $12M^2 \log_2 M$ moltiplicazioni. E' facile allora notare che, nel caso in cui N non sia trascurabile rispetto ad M il secondo metodo richiede un numero di operazioni decisamente inferiore.

Si considerino di nuovo l'equazione (2.3) e la (2.4): osservando che nel corso della ricerca la template non cambia possiamo scrivere:

$$\rho^{num}(i, j) = \sum A'(i + m, j + n) B(m, n) - \bar{B} \sum A'(i + m, j + n) \quad (2.5)$$

Dal momento che, per costruzione, A' ha media zero, l'ultimo termine a secondo membro della (2.5) può essere eliminato. Anche in questo caso il calcolo può essere condotto per mezzo di una trasformata di Fourier.

Delle due quantità sotto radice a denominatore nella (2.3) la seconda, essendo relativa alla template, rimane fissa durante tutta la procedura di ricerca e può quindi essere calcolata una sola volta. La prima, al contrario, deve essere ristimata ad ogni spostamento della patch e comporta l'onere computazionale principale ma può essere efficientemente calcolata per mezzo di due tabelle contenenti una gli integrali dei toni di grigio dell'immagine e l'altra il quadrato di tali valori. Tali tabelle possono essere calcolate per mezzo delle cosiddette somme cumulate (*running sum*):

$$s(u, v) = f(u, v) + s(u - 1, v) + s(u, v - 1) - s(u - 1, v - 1) \quad (2.6)$$

$$s^2(u, v) = f^2(u, v) + s^2(u - 1, v) + s^2(u, v - 1) - s^2(u - 1, v - 1) \quad (2.7)$$

La costruzione di tali tabelle richiede approssimativamente $3M^2$ operazioni ed è quindi sostanzialmente trascurabile se paragonata alle $3N^2(M-N+1)^2$ richieste per il calcolo del denominatore per ciascun pixel della finestra di ricerca.

Infine il calcolo del denominatore può essere ottenuto in maniera estremamente rapida in quanto richie-

de solamente poche operazioni sulle somme e le somme dei quadrati dei toni di grigio.

Solitamente, grazie all'algoritmo descritto, si è in grado con finestre di ricerca di dimensioni notevoli, di ottenere tempi di elaborazione anche dieci volte inferiori a quelli dell'approccio tradizionale.

2.2. Least Squares Matching

Come sottolineato, i metodi di correlazione non sono robusti di fronte a trasformazioni geometriche diverse dalla traslazione; il metodo dunque può essere applicato singolarmente solo in determinate circostanze, come ad esempio nell'analisi di sequenze in cui fotogrammi successivi presentano basi di presa molto piccole: in tali circostanze, tuttavia, la geometria del problema non garantisce buone affidabilità e precisioni ed essendo modesta la tracciabilità dei punti lungo la sequenza non si riesce a migliorare la base. Per superare il problema sono stati sviluppati negli ultimi vent'anni diversi algoritmi che permettono di considerare differenze geometriche più generali. Di questi, la tecnica più conosciuta in fotogrammetria, sviluppata per il presente studio, è la correlazione di immagini a minimi quadrati [Grün 1985] o Least Squares Matching (LSM).

Prese due porzioni di immagine, una estratta sull'immagine di riferimento (*master image*) chiamata template e la seconda sull'immagine in cui si vuole trovare il punto omologo (*slave image*) chiamata picture o patch, lo scopo dell'algoritmo è quello di determinare punti corrispondenti sulle due immagini, stimando una opportuna trasformazione che renda la patch localmente identica (a meno degli inevitabili residui) alla template. I toni di grigio di ciascun pixel della template $f(i, j)$ vengono confrontati con quelli ritenuti omologhi sulla patch $g(u, v)$: il grado di somiglianza viene misurato dalla somma dei quadrati delle differenze dei valori radiometrici dei pixel corrispondenti delle due finestre.

Applicando alla funzione discreta $g(u, v)$ trasformazioni parametriche sia di tipo geometrico che di tipo radiometrico si ottiene un'immagine $g'(u, v)$ più simile alla template. In termini analitici risulta:

$$f(x, y) - g'(x, y) = e(x, y) \quad (2.8)$$

in cui $e(x, y)$ è il residuo per il punto di coordinate (x, y) nel riferimento della master image. Supponendo che tale valore, nel dominio della template, si comporti come una variabile casuale a media zero e varianza σ , è possibile ottenere una stima di massima verosimiglianza dei parametri che determinano la trasformazione $g(u, v) \mapsto g'(x, y)$ minimizzando la norma L2 del vettore dei residui e . Per modellare la trasformazione geometrica della patch, ovvero la sua forma sulla slave image, si impiega una trasformazione affine:

$$\begin{cases} x = a_1u + a_2v + a_3 \\ y = b_1u + b_2v + b_3 \end{cases} \quad (2.9)$$

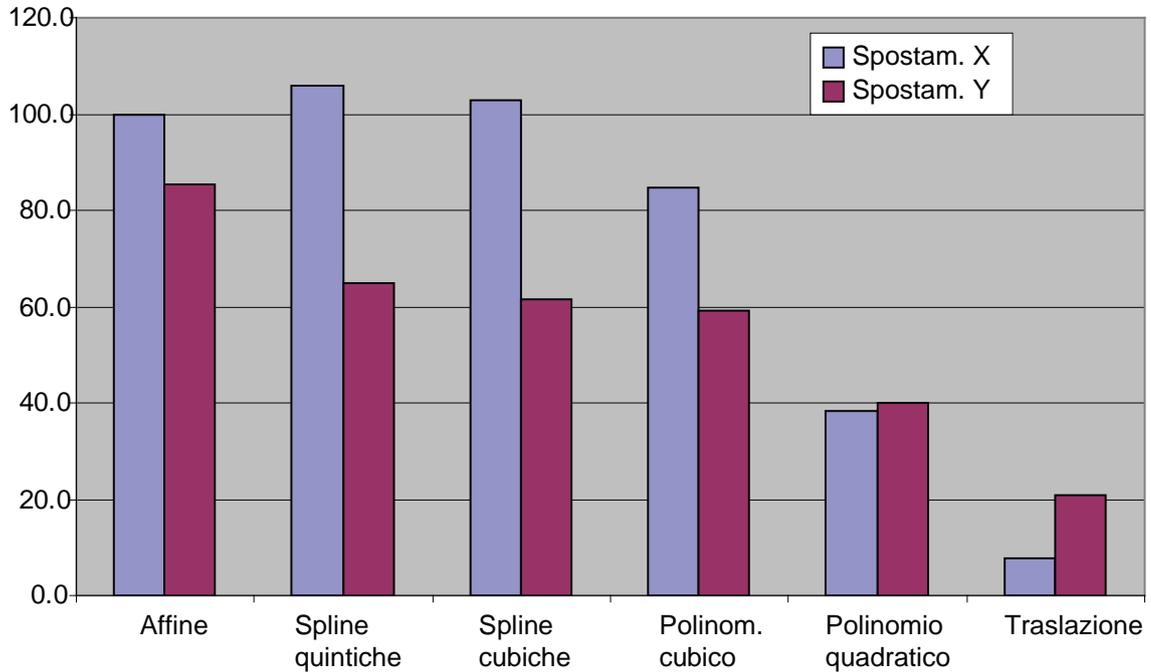


Figura 2.1: Precisioni ($1/\text{RMSE}$) nel posizionamento della patch, al variare delle funzioni di forma utilizzate per la stima a minimi quadrati. L'immagine di riferimento era stata riscalata del 10%.

Considerato che la template ha dimensioni normalmente inferiori a 15×15 pixel i relativi raggi proiettanti sono pressoché paralleli; assunto inoltre che la porzione di oggetto inquadrata sia piana, tale trasformazione assorbe gran parte delle deformazioni prospettiche legate alla variazione del punto di presa. Poiché in alcune applicazioni si utilizzano dimensioni della template piuttosto elevate sono state sperimentate anche altre funzioni di forma come suggerito da [Sutton 1988], [Bruck 1989], [Lu 2000]. I risultati emersi, brevemente illustrati in figura 2.1, mostrano come funzioni di forma più semplici (sole traslazioni) si dimostrano migliori (con un incremento di precisione tuttavia trascurabile) solo in assenza di deformazioni di tipo prospettico mentre forniscono risultati inaccettabili non appena intervengono trasformazioni differenti. Funzioni di forma più complesse (funzioni quadratiche polinomiali) non mostrano sostanziali differenze rispetto alla (2.9) se non con template molto grandi (caso in cui l'ipotesi di planeità dell'oggetto è meno valida).

Oltre ad una trasformazione geometrica viene anche considerata una trasformazione radiometrica dei toni di grigio, in modo da considerare eventuali cambiamenti di illuminazione e contrasto fra template e patch, generalmente espressa da una affine monodimensionale:

$$g'(x, y) = r_0 + (1 + r_1)g(x, y) \quad (2.10)$$

Si noti che, se i parametri di illuminazione sono molto simili nella patch e nella template, e quest'ultima

non è ben caratterizzata in termini di tessitura (non sono presenti gradienti dei toni di grigio significativi), i parametri r_0 ed r_1 risultano fortemente correlati e possono in qualche modo compromettere l'operazione di stima. Per irrobustire il processo durante la stima a minimi quadrati si verificano le correlazioni fra i vari parametri, escludendo dalla stima quelli fortemente correlati.

La funzione discreta che descrive i toni di grigio nella patch è una funzione generica, pertanto l'equazione (2.8) non risulta lineare e deve essere approssimata con uno sviluppo in serie di Taylor e la stima a minimi quadrati avviene per via iterativa:

$$f(x, y) - e(x, y) = g_0(x, y, p_0) + \frac{\partial g(x, y, p_0)}{\partial x} dx + \frac{\partial g(x, y, p_0)}{\partial y} dy \quad (2.11)$$

Facendo riferimento alle trasformazioni espresse dalla (2.9) i differenziali a secondo membro risultano:

$$\begin{aligned} \frac{\partial g(x, y, p_0)}{\partial x} dx &= \frac{\partial g(x, y, p_0)}{\partial x} \frac{\partial x}{\partial a_1} \Big|_0 da_1 + \frac{\partial g(x, y, p_0)}{\partial x} \frac{\partial x}{\partial a_2} \Big|_0 da_2 + \frac{\partial g(x, y, p_0)}{\partial x} \frac{\partial x}{\partial a_3} \Big|_0 da_3 \\ \frac{\partial g(x, y, p_0)}{\partial y} dy &= \frac{\partial g(x, y, p_0)}{\partial y} \frac{\partial y}{\partial b_1} \Big|_0 db_1 + \frac{\partial g(x, y, p_0)}{\partial y} \frac{\partial y}{\partial b_2} \Big|_0 db_2 + \frac{\partial g(x, y, p_0)}{\partial y} \frac{\partial y}{\partial b_3} \Big|_0 db_3 \end{aligned} \quad (2.12)$$

Per rendere la stima più stabile, anche in presenza di parametri approssimati non ottimali o di configurazioni geometriche o radiometriche instabili, la soluzione del sistema a minimi quadrati è ottenuta con l'algoritmo di line-search come suggerito in [Börlin 2002]; per una descrizione approfondita della tecnica si rimanda al relativo paragrafo in appendice A.

Il modello a minimi quadrati, ad ogni iterazione ed al termine della stima permette il calcolo delle seguenti quantità:

$$\text{sigma zero:} \quad \sigma_0^2 = \frac{1}{n - m} \mathbf{v}^T P \mathbf{v} \quad (2.13)$$

$$\text{matrice di covarianza:} \quad Q = \sigma_0^2 (J^T P J)^{-1} \quad (2.14)$$

$$\text{varianza dei singoli parametri:} \quad \sigma_i^2 = Q_{ii} \quad (2.15)$$

$$\text{covarianza degli scarti:} \quad C_{\mathbf{v}\mathbf{v}} = \sigma_0^2 \left[P^{-1} - J (J^T P J)^{-1} A^T \right] \quad (2.16)$$

in cui n è il numero di osservazioni nel sistema, ovvero il numero di pixel che compongono la template, m è il numero di gradi di libertà del sistema, ovvero il numero di parametri stimati, P è la matrice dei pesi ed infine J è la matrice jacobiana stimata per mezzo della (2.12) all'iterazione considerata. La matrice dei pesi P è di difficile determinazione in quanto non è possibile conoscere a priori se alcuni pixel risultano più significativi di altri, così come non è facile stabilire un valore a priori del σ_0 (come osservato in [Baltasvias

1991]) e pertanto viene posta uguale all'unità. Per questo motivo la grandezza di (2.13) e le singole varianze dei parametri in (2.15) sono di difficile interpretazione. Come parametro di qualità del matching si è soliti, pertanto, considerare il coefficiente di correlazione (2.3) fra template e patch ricampionata, il cui calcolo è computazionalmente trascurabile se confrontato con quello richiesto per la stima a minimi quadrati. Le precisioni di misura, come già detto, non possono essere facilmente desunte dalla (2.14); pertanto il rapporto v/σ_v , per mettere in luce eventuali errori grossolani deve essere opportunamente riscaloato per effettuare il data snooping. Tali errori possono essere causati da una molteplicità di fattori (effetti di noise del sensore, riflessi dell'oggetto o dell'obbiettivo della camera, occlusioni, etc...). Va tuttavia sottolineato che, essendo il matching una tecnica di misura dinamica, eliminare osservazioni in via definitiva nelle fasi iniziali può ostacolare la convergenza alla soluzione corretta, se la posizione di partenza è poco approssimata.

Nel paragrafo conclusivo del presente capitolo verranno illustrati alcuni test utilizzati per stimare la precisione raggiungibile per mezzo del LSM, dell'ordine del centesimo di pixel in condizioni ottimali.

2.3. Multi-Photo Geometrically Constrained Matching

Il LSM si basa unicamente sulla similitudine tra i toni di grigio e non è quindi in grado di garantire che le corrispondenze estratte siano geometricamente compatibili: ad esempio non è detto che il punto estratto su un'immagine si trovi sulla linea epipolare individuata dal punto corrispondente alla template. A maggior ragione quindi, se il matching viene eseguito su più di due immagini, i punti estratti su ogni slave sono indipendenti gli uni dagli altri.

Per superare tali limitazioni una tecnica estremamente più versatile è offerta dal Multi-Photo Geometrically Constrained Matching (MPGCM o MGCM) introdotto da [Grün 1988] e [Baltasvias 1991]: si tratta di un'estensione del LSM in cui la soluzione è vincolata a soddisfare le condizioni di collinearità e di congruenza nello spazio oggetto dei raggi omologhi. Il metodo, grazie alle condizioni di collinearità, mette in relazione le osservazioni su più fotogrammi traendo beneficio dalla maggior ridondanza ottenibile con l'uso congiunto di tutte le immagini, sia in termini di affidabilità sia di precisione di misura.

I vincoli geometrici vengono imposti trattando le equazioni di collinearità come osservazioni aggiuntive nel sistema a minimi quadrati. In linea di principio, oltre alle coordinate terreno del punto, anche i parametri di orientamento interno ed esterno possono essere stimati; tuttavia la dinamicità della misura e la potenziale instabilità che nasce per le forti correlazioni tra parametri sconsiglia di adottare questa soluzione se la geometria non è molto robusta e ridondante. Le dimensioni del sistema risolvibile divengono poi rapidamente molto elevate e l'esperienza mostra che, se non si dispone di parametri approssimati molto buoni, esso può risultare piuttosto instabile, portando a risultati del tutto inattesi. Per questo motivo nel seguito si considererà solamente il caso in cui i parametri di orientamento esterno ed interno sono noti a priori e vengono mantenuti fissi durante la stima, limitandosi a considerare solamente le coordinate oggetto del punto.

Si considerino allora le equazioni di collinearità:

$$\begin{aligned} x - x_0 &= c \frac{r_{11}(X - X_0) + r_{21}(Y - Y_0) + r_{31}(Z - Z_0)}{r_{13}(X - X_0) + r_{23}(Y - Y_0) + r_{33}(Z - Z_0)} \\ y - y_0 &= c \frac{r_{12}(X - X_0) + r_{22}(Y - Y_0) + r_{32}(Z - Z_0)}{r_{13}(X - X_0) + r_{23}(Y - Y_0) + r_{33}(Z - Z_0)} \end{aligned} \quad (2.17)$$

in cui c è la distanza principale, (x_0, y_0) le coordinate del centro principale, (X_0, Y_0, Z_0) le coordinate del centro di presa nel sistema oggetto ed r_{ii} le diverse componenti della matrice di rotazione; una volta osservato che risultano non lineari nelle variabili indipendenti X, Y e Z è possibile approssimarle per mezzo di uno sviluppo in serie di Taylor:

$$\begin{aligned} dx &= x_0 + \left. \frac{\partial x}{\partial X} \right|_0 dX + \left. \frac{\partial x}{\partial Y} \right|_0 dY + \left. \frac{\partial x}{\partial Z} \right|_0 dZ \\ dy &= y_0 + \left. \frac{\partial y}{\partial X} \right|_0 dX + \left. \frac{\partial y}{\partial Y} \right|_0 dY + \left. \frac{\partial y}{\partial Z} \right|_0 dZ \end{aligned} \quad (2.18)$$

Si noti che, scrivendo le (2.18) nel sistema pixel, dx e dy sono proprio quelle in (2.12): esse costituiscono quindi il legame tra le osservazioni radiometriche a quelle geometriche; va però osservato che mentre le quantità (x, y) in (2.12) sono prese direttamente sul fotogramma, e sono dunque affette generalmente da distorsione, le medesime quantità in (2.18) devono soddisfare la condizione di collinearità e devono quindi essere prive di effetti di distorsione. Il valore molto piccolo che assumono le correzioni dx e dy ad ogni iterazione fa sì che, seppur a rigori esse andrebbero corrette secondo la distorsione locale, possano senz'altro essere considerate identiche. In termini matematici:

$$\begin{aligned} x &= \bar{x} + \Delta x, & y &= \bar{y} + \Delta y \\ dx &\cong d\bar{x}, & dy &\cong d\bar{y} \end{aligned} \quad (2.19)$$

in cui Δx e Δy rappresentano le correzioni da apportare alle coordinate pixel \bar{x} e \bar{y} per ottenere le coordinate prive di distorsione x e y .

Consideriamo allora lo schema che assume il sistema risolvete e la matrice normale (figura 2.2): è facile accorgersi che la matrice disegno è decisamente sparsa e presenta una chiara struttura a blocchi: per motivi computazionali è quindi preferibile calcolare la matrice normale direttamente sfruttando le proprietà delle matrici a blocchi.

La soluzione del sistema è data, con l'impostazione di Gauss-Markov, da:

$$\hat{\mathbf{x}} = (A^T P A)^{-1} A^T P (Y - \mathbf{y}_0) \quad (2.20)$$

in cui A è la matrice disegno, P è la matrice dei pesi, Y sono le osservazioni e \mathbf{y}_0 il vettore contenente i termini noti. Come nel LSM resta il problema di stabilire che peso dare alle equazioni radiometriche; la situazione è ora ulteriormente complicata dalla presenza di osservazioni differenti ed è pertanto necessario

almeno fissare i rapporti di peso tra le due. Tale problema risulta di fondamentale importanza per la corretta applicazione del metodo e viene indicato come uno dei maggiori punti di debolezza già nello stesso [Baltsavias 1991]: se viene dato poco peso alle equazioni di collinearità si ottiene sostanzialmente lo stesso risultato che si avrebbe avuto applicando il solo LSM con le posizioni delle patch indipendenti le une dalle altre; se viceversa è troppo elevato il peso delle equazioni di collinearità, si rischia di bloccare la soluzione nella sua posizione iniziale (verrà più avanti spiegato che inizialmente i punti vengono scelti in modo da non generare residui nelle (2.17)) impedendo alla parte radiometrica di migliorare il grado di somiglianza

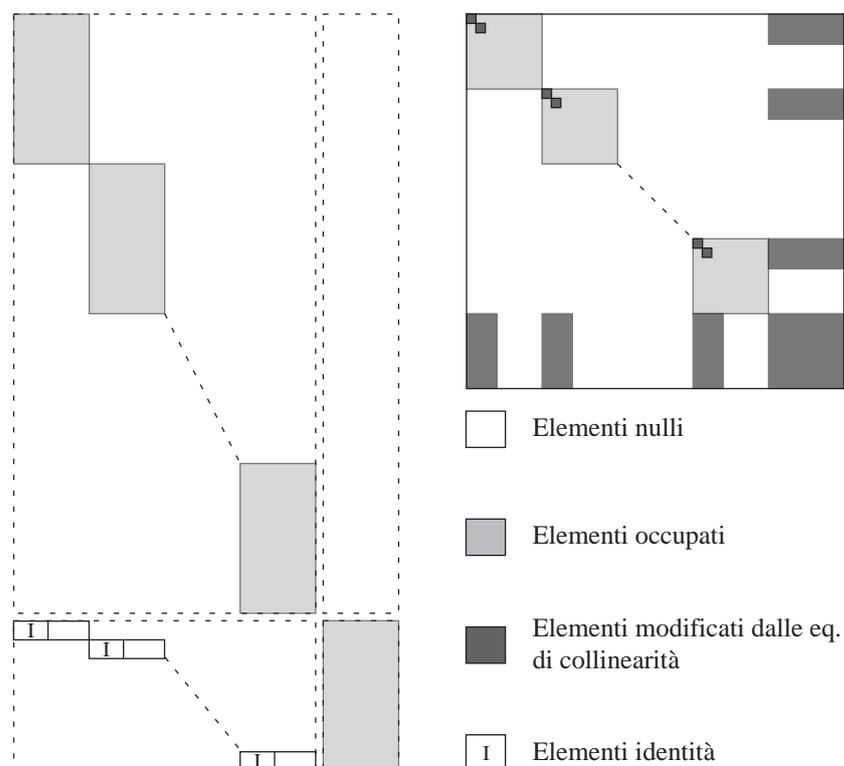


Figura 2.2: Schema della struttura della matrice disegno (a sinistra) e della matrice normale (a destra) per MFCM in cui si considerano variabili solo le coordinate oggetto del punto.

delle patch con la template.

L'implementazione utilizza un approccio adattivo al problema, così come suggerito anche da Baltsavias. Ad ogni iterazione vengono analizzati i residui delle equazioni di collinearità su ciascun fotogramma: se tali residui risultano contenuti (inferiori al decimo di pixel) si diminuisce il peso delle equazioni di collinearità in modo da garantire una maggior libertà alla parte radiometrica; se al contrario i residui della collinearità divengono non trascurabili il peso delle equazioni viene opportunamente aumentato secondo una legge esponenziale dipendente dalla norma di tali residui: in tal modo si costringe il sistema, nell'ite-

razione successiva, a riassumere una configurazione coerente con la geometria del problema. Detto quindi σ_i il valore della norma dei residui di collinearità relativi al fotogramma i -esimo per il punto considerato, il peso da dare alle equazioni geometriche (sempre relative al fotogramma i -esimo) risulta:

$$\frac{1}{\sigma_i^2} = \frac{1}{\sigma_0^2} \cdot 10^A \quad (2.21)$$

in cui il valore di σ_0 viene impostato dall'utente in funzione principalmente della qualità dell'orientamento dei fotogrammi. La scelta di una funzione esponenziale è di tipo empirico; si è visto, nel corso della sperimentazione, che tale legge rispetto ad altre rende più stabile il sistema: una legge lineare o più in generale polinomiale, ad esempio, con scarsa caratterizzazione di tessitura o presenza di occlusioni su uno o più fotogrammi, conduce ad un continuo alternarsi di pesi alti e bassi per le equazioni di collinearità che non portano a convergenza il sistema.

Nella fase di inizializzazione si calcola la posizione iniziale delle patch in modo da partire con una configurazione geometrica corretta (ovvero senza residui nella collinearità): ciò permette di impostare un peso iniziale molto basso, per poi eventualmente correggerlo al termine della prima iterazione per mezzo della (2.21).

Le operazioni di inizializzazione sono riassunte in figura 2.3: per prima cosa, date le posizioni iniziali sui diversi fotogrammi, si stima a minimi quadrati l'intersezione dei raggi omologhi, ovvero le coordinate del punto oggetto; volendo mantenere inalterata la posizione della template sulla master image, tale punto viene proiettato sul raggio di proiezione passante per il suo centro; il punto così ottenuto viene riproiettato su tutti gli altri fotogrammi per mezzo delle equazioni di collinearità, il che garantisce l'annullarsi dei residui; le posizioni sul fotogramma, vengono poi corrette tenendo conto di eventuali distorsioni. Le posizioni iniziali sulle diverse immagini possono essere ottenute con correlazione semplice, anche se questo può introdurre errori grossolani nelle corrispondenze.

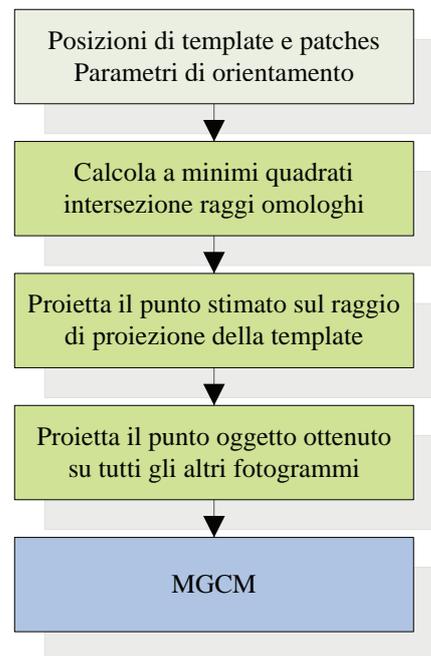


Figura 2.3: Schema delle operazioni di inizializzazione per l'MGCM.

2.4. Pre-elaborazione delle immagini

Per rendere più affidabile ed efficiente il riconoscimento di punti omologhi su due o più fotogrammi nella fase di Area Based Matching (ABM) si può decidere di far precedere tale fase da una serie di procedure

di pre-processamento: tali procedure permettono di ottimizzare la distribuzione dei toni di grigio, ottenendo valori di luminosità e contrasto migliori rispetto a quelli presenti nell'immagine originale, si possono ottenere trasformazioni dei fotogrammi in modo da migliorare la similitudine fra aree corrispondenti e ridurre o eliminare i fenomeni di distorsione ottica. Non sempre necessarie, queste pre-elaborazioni si dimostrano in molti casi essenziali per avere risultati soddisfacenti nelle successive fasi di matching. Inoltre, pur rappresentando un maggior onere computazionale all'inizio della procedura, portano a prestazioni più elevate in fase di correlazione, con un risparmio di risorse nelle successive fasi di elaborazione. La gestione dell'onere computazionale è infatti un tema centrale in questo tipo di applicazioni, adottata per la ricostruzione densa della superficie dell'oggetto (DSM), in cui sono coinvolti da alcune decine di migliaia ad alcuni milioni di punti per ciascun fotogramma.

Di seguito tali procedure verranno brevemente illustrate indicando i metodi e gli algoritmi attualmente implementati.

2.4.1. Eliminazione della distorsione

Modelli matematici per la correzione della distorsione e delle deformazioni dei fotogrammi sono stati lungamente studiati dagli anni '70 a partire da uno studio fondamentale di [Brown 1971] che ha posto le premesse per le tecniche di auto-calibrazione. Per maggiori dettagli su tali aspetti e sulle metodologie richieste per la stima dei parametri di calibrazione si rimanda a testi quali [Beyer 1992] [Cooper 1996] o [Fraser 2001]; ci si limita qui a ricordare, per sottolineare l'importanza del problema, che i valori di distorsione nelle camere digitali impiegate correntemente in fotogrammetria possono superare di parecchio il centinaio di μm e, ad esempio, in obiettivi per telecamere impiegate per Mobile Mapping o applicazioni dinamiche, arrivare a 500-600 μm ed oltre.

Per riportare le coordinate di punti estratti dal fotogramma alle loro corrispettive coordinate "indistorte" si applicano quindi termini correttivi che modellano le diverse cause di distorsione presenti. Normalmente il termine prevalente è quello dovuto alla distorsione radiale:

$$\begin{aligned}\Delta x_r &= \frac{x}{r} \Delta r \\ \Delta y_r &= \frac{y}{r} \Delta r\end{aligned}\tag{2.22}$$

dove

$$\bar{x} = x - x_0, \bar{y} = y - y_0, r^2 = \bar{x}^2 + \bar{y}^2\tag{2.23}$$

rappresentano la distanza dal punto considerato dal centro del fotogramma individuato dal punto principale e la grandezza Δr è ottenuta per mezzo di una opportuna funzione polinomiale:

$$\Delta r = K_1 r^3 + K_2 r^5 + K_3 r^7 + \dots \quad (2.24)$$

Si possono inoltre considerare effetti dovuti al non perfetto centramento delle diverse lenti che compongono l'obbiettivo, che generano effetti di distorsione di tipo tangenziale:

$$\begin{aligned} \Delta x_d &= P_1 (r^2 + 2\bar{x}^2) + 2P_2 \bar{x} \bar{y} \\ \Delta y_d &= P_2 (r^2 + 2\bar{y}^2) + 2P_1 \bar{x} \bar{y} \end{aligned} \quad (2.25)$$

Seguendo quanto indicato in [Fraser 2001] si possono poi considerare la non perfetta planeità del sensore fotosensibile, che può raggiungere valori elevati in quelli di maggiori dimensioni. La modellazione delle distorsioni indotte è tuttavia complessa a fronte di risultati modesti, per cui questo errore non viene considerato nel seguito. Il sistema di riferimento individuato dal sensore CCD può avere fattori di scala differenti nelle due direzioni principali (ovvero il pixel può non essere quadrato) e può presentare assi coordinati non perfettamente ortogonali (effetto di taglio). Di tali effetti è però possibile tenere direttamente conto, con riferimento al paragrafo 1.4, per mezzo della matrice camera.

In definitiva si passa dalle coordinate (x, y) di un punto misurate sul fotogramma alle corrispondenti coordinate immagine indistorte con la seguente:

$$\begin{aligned} x &= x_0 + \Delta x = x_0 + \Delta x_r + \Delta x_d \\ y &= y_0 + \Delta y = y_0 + \Delta y_r + \Delta y_d \end{aligned} \quad (2.26)$$

in cui le correzioni che compaiono a secondo membro sono rappresentate dalle espressioni (2.22) e (2.25).

La (2.26) è non lineare nelle variabili x e y e non è invertibile in forma chiusa. In altre parole, con la (2.26) è possibile calcolare direttamente la posizione indistorta di un punto nota la sua posizione distorta; non è invece possibile determinare la posizione distorta nota quella calcolata per proiezione analitica, cioè dalle equazioni di collinearità (ovvero dalle coordinate indistorte).

Tale problema diventa particolarmente importante quando si debbano contemporaneamente utilizzare entrambe le informazioni contemporaneamente, come nel caso dell'MGCM o in quei casi in cui si debba ricampionare l'immagine secondo parametri legati alle sue caratteristiche di orientazione (foto-raddrizzamenti, immagini epipolari etc...).

Alcuni autori (ad esempio [D'Apuzzo 2003]) hanno proposto una procedura iterativa per determinare approssimativamente le contro-correzioni per mezzo delle (2.22) e (2.25), sostituendo le coordinate del punto distorto con quelle del punto indistorto e supponendo, pertanto, che localmente la funzione di distorsione possa considerarsi sostanzialmente costante. Determinando i residui di una ri-trasformazione a coordinate indistorte e correggendo il punto di partenza di una tale quantità, dopo un certo numero di iterazioni, sotto certe condizioni, il sistema va a convergenza fornendo una buona approssimazione (dell'ordine di centesimi o millesimi di pixel) della corretta posizione indistorta. Quando la curva di distorsione ha pendenze molto elevate, tuttavia, il metodo richiede un numero molto elevato di iterazioni e può non arrivare

a convergenza.

Per tale motivo è stato realizzato un approccio più consistente, collaudato su ottiche con distorsioni anche molto elevate. Esso si fonda sull'uso di una struttura dati che fornisca contemporaneamente le posizioni deformate del fotogramma (ovvero un insieme finito delle possibili coordinate pixel) e le corrispondenti posizioni ottenute eliminando la distorsione ottica per mezzo delle (2.26). In tal modo la corrispondenza fra coordinate deformate e indeformate è sempre esatta (a meno delle precisioni di calcolo dell'elaboratore) indipendentemente dall'influenza dei parametri di distorsione. Quando viene richiesta la posizione deformata in funzione di una posizione indeformata si va ad interrogare questa seconda lista e si va a ricavare la corrispondente posizione affetta da distorsione nel primo elenco. Per ottenere precisioni elevate, dal momento che difficilmente la posizione che interessa l'utente coinciderà esattamente con una di quelle memorizzate nella struttura, si può decidere di considerare i punti che circondano tale posizione nel campo indistorto e ricavare per interpolazione la corrispondente posizione distorta corretta.

Al momento è stato implementato un metodo di interpolazione bilineare: si considerano i quattro punti (indistorti) che circondano il punto P richiesto (vd. Figura 2.4); determinato il punto NW della griglia più vicino, gli altri punti sono facilmente rintracciabili nella struttura dati a griglia regolare rettangolare; le coordinate distorte (x_d, y_d) corrispondenti vengono determinate, dette (x_u, y_u) le corrispondenti coordinate indistorte, considerando come funzione approssimante un paraboloide iperbolico:

$$\begin{aligned} x_d &= a_1 x_u + a_2 y_u + a_3 x_u y_u + a_4 \\ y_d &= b_1 x_u + b_2 y_u + b_3 x_u y_u + b_4 \end{aligned} \quad (2.27)$$

i coefficienti a_i e b_i di ogni maglia si calcolano risolvendo il sistema :

$$\begin{bmatrix} x_{u, nw} & y_{u, nw} & x_{u, nw} y_{u, nw} & 1 \\ x_{u, ne} & y_{u, ne} & x_{u, ne} y_{u, ne} & 1 \\ x_{u, sw} & y_{u, sw} & x_{u, sw} y_{u, sw} & 1 \\ x_{u, se} & y_{u, se} & x_{u, se} y_{u, se} & 1 \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ a_3 \\ a_4 \end{bmatrix} = \begin{bmatrix} x_{d, nw} \\ x_{d, ne} \\ x_{d, sw} \\ x_{d, se} \end{bmatrix} \quad (2.28)$$

in cui le osservazioni sono le coordinate distorte dei vertici della maglia (per il calcolo dei parametri b_i è sufficiente cambiare il vettore delle osservazioni).

Fissato il passo della griglia, l'insieme di coordinate distorte funge da descrittore del campo di distorsione sul fotogramma: se gli effetti di distorsione sono limitati è possibile scegliere un passo di griglia relativamente grande in modo da ottenere liste non eccessivamente lunghe e facilitando l'elaborazione del calcolatore; se, al contrario, i parametri di distorsione sono elevati, si può decidere di utilizzare passi inferiori ed una descrizione molto più fitta del campo.

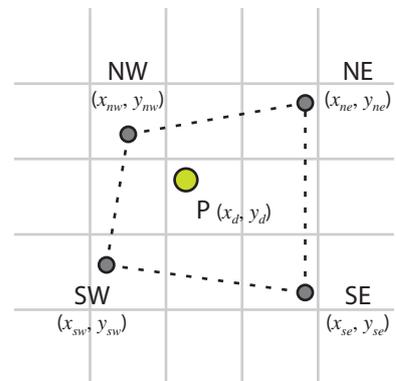


Figura 2.4: Schema delle posizioni dei punti distorti.

Si noti che, non essendo le posizioni indistorte disposte secondo una maglia regolare, la ricerca del punto più vicino (ovvero del punto NW di figura 2.4) richiede di analizzare tutta la lista di corrispondenze; indipendentemente dalla curvatura locale del campo di distorsione, tale lista è costituita da un numero molto elevato di abbinamenti e la ricerca può richiedere uno sforzo computazionale non trascurabile. Si è perciò implementato un sistema di ricerca basato sui *quad-tree* (decomposizione ad albero quaternario) in cui l'intera lista viene suddivisa in sottogruppi di dimensione massima prefissata (generalmente alcune decine di punti); ciascun punto della griglia iniziale

viene di volta in volta inserito in un gruppo in base alla sua posizione sul fotogramma: inizialmente tale fotogramma viene suddiviso in quattro rettangoli uguali (vd. Figura 2.5); nel momento in cui, in uno di tali sottogruppi si raggiunge una situazione di overflow, ovvero viene inserito un numero maggiore di punti rispetto a quello massimo prefissato, si suddivide ulteriormente tale regione in quattro quadranti più piccoli; la procedura viene iterata fintanto che tutti i punti della lista non sono stati assegnati. Nel momento in cui l'utente richiede la posizione deformata di un punto fornendo quella indistorta, l'algoritmo determina a quale sottogruppo (foglia) corrisponde la posizione richiesta, e determina, per le sole posizioni presenti in tale sottogruppo, quale sia quella di minima distanza. In tal modo, a fronte inizialmente di un onere computazionale appena maggiore (la costruzione della struttura ad albero richiede uno sforzo superiore di pochi punti percentuali), si ha una modalità di interrogazione della lista molto più efficiente.

Infine, qualora siano richieste numerosissime interrogazioni si può decidere, seguendo lo schema riportato, di calcolare le trasformazioni fra posizioni indistorte e distorte disponendo le prime su una griglia rettangolare regolare: memorizzando tali informazioni in una struttura dati adeguata (ad esempio una look-up table) risulta molto più semplice il calcolo dell'interpolazione bilineare e la trasformazione avviene in tempi addirittura inferiori a quelli richiesti per passare da coordinate distorte ad indistorte utilizzando la (2.26). E' anche possibile ricampionare direttamente l'immagine originale ottenendone una versione indistorta ed evitando così la trasformazione fra un sistema e l'altro nelle procedure seguenti. Ottenendo precisioni, pressoché con qualsiasi tipo di ottica, dell'ordine del decimillesimo di pixel, è facile notare come il metodo sia più che adeguato per qualsiasi genere di applicazione fotogrammetrica anche di altissima precisione. Per quanto riguarda i tempi di elaborazione bisogna considerare i tre possibili utilizzi del metodo:

1. per un utilizzo interattivo si genera l'albero quadramentale e si vanno poi ad interrogare singolarmente

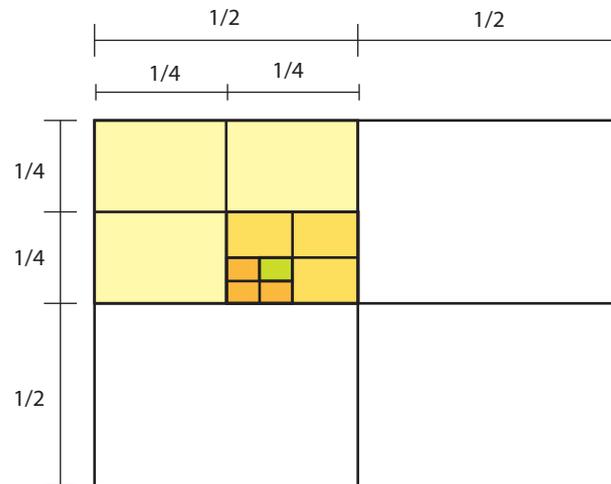


Figura 2.5: Schema di scomposizione del piano immagine mediante i quad-tree.

le singole posizioni; l'onere di calcolo maggiore viene affrontato inizialmente proprio per la generazione dell'albero, che richiede solitamente alcuni secondi per liste costituite da qualche milione di punti; le successive interrogazioni, grazie all'efficienza del metodo di ricerca, hanno tempi dell'ordine del millesimo di secondo;

2. se le interrogazioni sono inserite in procedure automatiche e la loro numerosità risulta estremamente elevata (MGCM) risulta più efficace la seconda metodologia proposta, in cui i punti indistorti vengono opportunamente grigliati; la struttura dati complessiva, per la cui codifica nella look-up table può essere richiesta una decina di minuti, può essere salvata in un file ed essere riutilizzata a piacimento. In tal caso, come già detto, il calcolo delle posizioni è pressoché istantaneo.
3. qualora si decida di ricampionare direttamente l'immagine, una volta formata la look-up table, bastano pochi decimi di secondo per immagini di medie dimensioni (da 3 a 6 Mpixel).

2.4.2. Ottimizzazione del contenuto informativo

Spesso le immagini che compongono la sequenza presentano zone più o meno estese in cui il contenuto informativo non è ottimale per le operazioni di matching: in alcuni casi tale problema è dovuto alla scarsa caratterizzazione di tessitura che contraddistingue l'oggetto, situazione che rende il matching instabile e poco affidabile; in altri il problema può essere rappresentato da sostanziali differenze di illuminazione fra l'immagine di riferimento (*master image*) e quelle di lavoro (*slave image*). In tali casi è possibile intervenire sull'immagine con trasformazioni dei toni di grigio.

La correzione del fattore gamma permette di tenere in considerazione la risposta non lineare del sensore in funzione dei cambiamenti di illuminazione: essa assume analiticamente la forma:

$$\bar{A}(x, y) = A(x, y)^\gamma \quad (2.29)$$

in cui A ed \bar{A} rappresentano le funzioni a valori discreti che descrivono i toni di grigio dell'immagine rispettivamente originale e ricampionata e in cui γ è il fattore correttivo. E' immediato verificare che il termine λ in (2.29) tende a schiarire l'immagine quando presenta valori maggiori di 1 ed a scurirla per valori inferiori. Nei casi in cui la template presenti forti variazioni di luminosità, i parametri radiometrici del LSM (2.10) non sono sufficienti a modellare le variazioni, per cui i residui sono piuttosto elevati anche qualora i fattori di forma vengano determinati correttamente

Quando le immagini presentano una caratterizzazione di tessitura non molto elevata si può aumentare il contrasto ri-scalando i toni di grigio dell'immagine in modo da utilizzare una gamma dinamica più ampia; ipotizzando che l'immagine in questione sia costituita da 8 bit per canale, e quindi i toni di grigio vadano da 0 a 255, si scalano linearmente tali valori facendo coincidere un determinato valore minimo con 0 ed un determinato valore massimo con 255. Il metodo permette quindi di aumentare i gradienti fra i toni di grigio (il fattore di scala è sempre maggiore di 1), e di rendere pertanto più riconoscibili eventuali sfumature

nella tessitura. Bisogna però osservare che, nel caso in cui il range dei toni di grigio ricopre tutta la scala, e di conseguenza i valori massimi e minimi coincidono con gli estremi del dominio, la rielaborazione non conduce a sostanziali miglioramenti.

Per far fronte a quest'ultimo problema, ovvero ai casi in cui l'istogramma dei toni di grigio non risulti compresso in un determinato intervallo ma si estenda su tutto il dominio di luminosità, si possono utilizzare tecniche di equalizzazione: esse cercano di ottenere una ridistribuzione dei toni di grigio, utilizzando una funzione che può essere lineare o non lineare, in modo da ottenere una distribuzione all'incirca uniforme nell'intero istogramma. Così non solo si raggiunge l'obiettivo di utilizzare l'intera gamma cromatica disponibile ma si bilanciano anche zone chiare e zone scure. La funzione da applicare nel caso lineare risulta:

$$g_i = \frac{(M-1)}{n_i} \sum_{j=0}^i n_j \quad (2.30)$$

in cui n_i è il numero totale di punti che costituiscono l'immagine, n_i è il numero di punti che hanno scala di grigio pari ad i ed M è il numero totale di scale di grigio (solitamente pari a 256).

Anche in questo caso la tecnica può funzionare in maniera autonoma ed automatica sulle immagini in ingresso.

Si può infine impiegare il filtro di Wallis [Wallis 1976] per riportare i valori di luminosità e contrasto di una regione di dimensioni predefinite attorno a ciascun punto dell'immagine a valori determinati. La sua efficienza nel fornire immagini con contenuti informativi più leggibili dagli algoritmi di matching [Seiz 2003] è testimoniata dalle numerose implementazioni del filtro rintracciabili in letteratura ([Grin 2003], [Baltavias 1996], etc...). I valori radiometrici dell'immagine ricampionata si ottengono da quella originale nel modo seguente: l'immagine viene suddivisa in blocchi rettangolari di dimensione prefissata e a distanza nota gli uni dagli altri; vengono poi determinati per ciascun blocco i parametri r_1 ed r_2 definiti come:

$$\begin{aligned} r_1 &= c \frac{s_f}{cs_g + (1-c)s_f} \\ r_2 &= bm_f + (1-b-r_1)m_g \end{aligned} \quad (2.31)$$

in cui m_g ed s_g rappresentano la media e la deviazione standard del blocco, m_f e s_f i valori obiettivo per le stesse grandezze e b e c sono due costanti che permettono di forzare il raggiungimento dei valori obiettivo: in particolare se b è prossimo a 0 l'immagine presenta medie prossime a quelle originali, se b è prossimo a 1 si ottengono valori più prossimi a quelli obiettivo; la costante c permette di modificare il contrasto finale in relazione a quello obiettivo: per c minore di 1 il contrasto risulta minore rispetto al valore fissato, viceversa se c è maggiore di 1.

Si calcolano inoltre i valori di r_0 ed r_1 per ciascun punto dell'immagine per mezzo di un processo di interpolazione (solitamente si utilizza un'interpolazione bilineare) basandosi sui valori determinati precedentemente. Infine si calcolano i nuovi valori di luminosità per ciascun punto mediante la seguente equa-

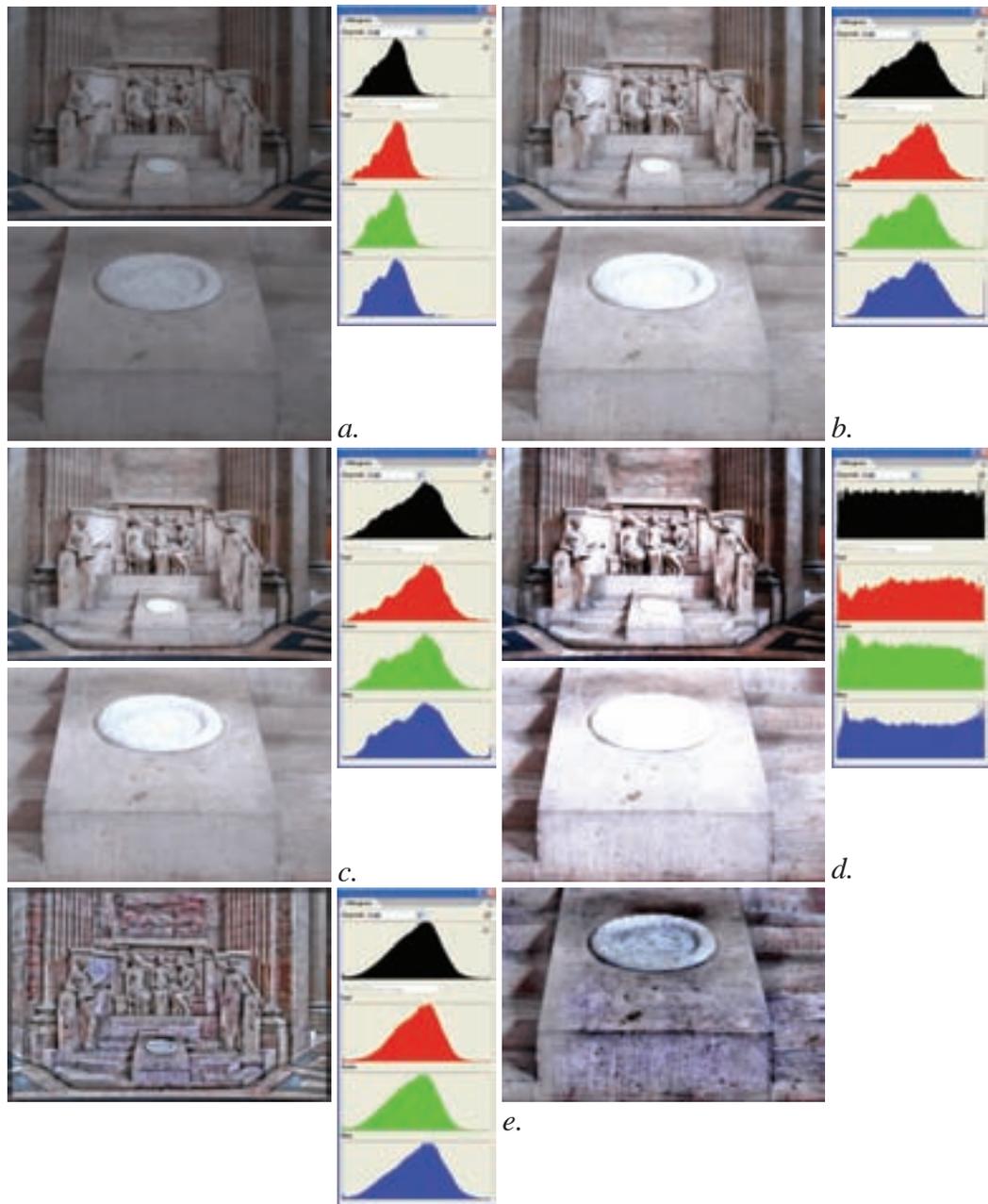


Figura 2.6: Trasformazione dei toni di grigio per migliorare le prestazioni del matching. Vengono presentati: il fotogramma completo, l'istogramma dei toni radiometrici per ciascuna canale RGB ed un particolare del fotogramma. (a.) Immagine originale; (b.) Correzione gamma; (c.) Correzione del contrasto; (d.) Equalizzazione; (e.) Filtro di Wallis. In questo caso il filtro di Wallis ottiene i migliori risultati, evitando problemi di clipping (aree sovra- o sotto-saturate) che invece, in misura differente affliggono gli altri metodi.

zione:

$$\bar{A}(x, y) = r_1(x, y) \cdot A(x, y) + r_0(x, y) \quad (2.32)$$

Anche in questo caso la procedura può essere utilizzata a scatola chiusa ed è quindi pienamente automatizzabile nel processo.

Nella pagina precedente sono riportati alcuni esempi delle tecniche appena illustrate (figura 2.6) in cui è possibile anche notare come le diverse procedure operino trasformazioni differenti negli istogrammi dei toni di grigio.

2.4.3. Rettifica di immagini epipolari

Molto spesso, nonostante le potenzialità dei metodi di matching precedentemente descritti, le procedure di correlazione non sortiscono buon esito a causa dell'eccessiva differenza fra le immagini utilizzate. Il problema è particolarmente importante nel cosiddetto matching denso, cioè nella ricostruzione di superficie a fotogrammi già orientati: mentre la S&M seleziona i punti migliori, qui occorre individuare omologhi su una maglia densa, indipendentemente dalla qualità del segnale radiometrico. Senza una buona conoscenza di tutti i parametri approssimati impiegati nel LSM (non solo le traslazioni, ma anche le rotazioni e le scale) il metodo difficilmente ottiene buoni risultati. Tali parametri, come si vedrà nel seguito, possono essere stimati dalla disposizione di punti già ottenuti dalla S&M, purché sufficientemente ben distribuiti, in relazione alla complessità dell'oggetto. Quando le posizioni approssimate non risultano molto precise, o quando l'oggetto ripreso non presenta una buona caratterizzazione di tessitura, la scelta dei parametri geometrici del LSM deve evitare l'instabilità del sistema risolvibile. L'utilizzo delle equazioni di collinearità può parzialmente porre rimedio a questo aspetto, vincolando le possibili direzioni che la soluzione può prendere; tuttavia, come visto, bisogna introdurre nel problema incognite aggiuntive (le coordinate terreno del punto) e non si può derivare da tali equazioni le variazioni di forma presenti sui fotogrammi.

Un metodo che permette di risolvere, seppur parzialmente, i problemi descritti è la rettifica epipolare. L'idea è quella di determinare una trasformazione proiettiva che permetta di ricampionare entrambi i fotogrammi in modo che punti omologhi appartengano alla medesima retta orizzontale: lo spazio di ricerca dei punti corrispondenti diviene monodimensionale e si ottiene una rappresentazione dell'oggetto nei due fotogrammi più simile. Il vantaggio per il matching è duplice: da un lato il sistema a minimi quadrati è più stabile: non sono richieste incognite aggiuntive ma al contrario si fissa una variabile; dall'altro la maggior similitudine fra le immagini permette di ottenere più facilmente accoppiamenti corretti; vengono inoltre eliminati effetti di rotazione del formato: non occorre quindi stimare i parametri approssimati di forma della patch e si può far precedere la fase di LSM da una ricerca più vasta della corrispondenza iniziale con la correlazione semplice.

Al momento sono stati inseriti nella procedura due differenti metodi di generazione delle immagini

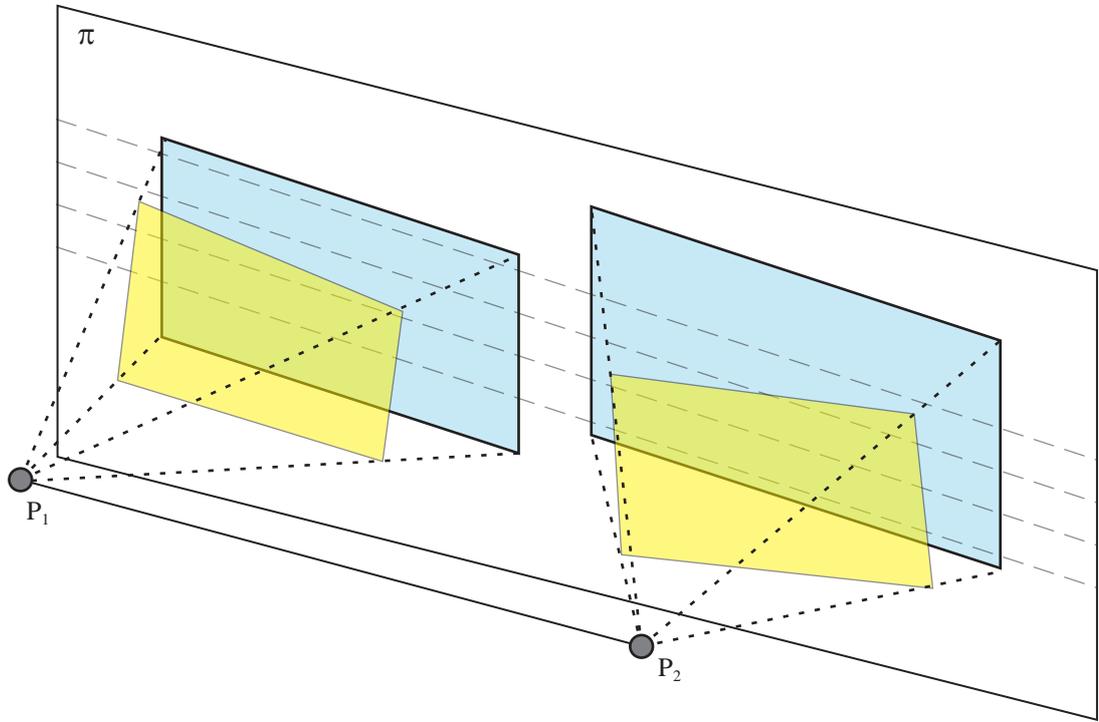


Figura 2.7: Schema della generazione di immagini epipolari: si stima l'omografia che lega i due piani immagine originali (in giallo) ad un nuovo piano π parallelo alla base di presa P_1P_2 ; punti corrispondenti sulle nuove immagini (in azzurro) vengono, per costruzione, a trovarsi sulla stessa linea orizzontale.

epipolari: un primo metodo, descritto in [Hartley 2000], è basato sulla stima di un piano proiettivo comune ad entrambi i fotogrammi e legato a questi da due particolari omografie. La rettifica di immagini epipolari richiede infatti che i punti corrispondenti agli epipoli vengano “mappati”, per mezzo di una trasformazione appropriata, all'infinito. Tale trasformazione può essere effettivamente assimilata ad una collinearità, nel qual caso l'intero processo assume il significato di riproiezione su un piano immagine “virtuale” che sia parallelo alla base di presa dei due fotogrammi (vedi figura 2.7). È facile osservare che i piani che verificano tale condizione costituiscono ∞^1 fasci impropri di rette con generatrici rette parallele alla base di presa. Analiticamente ciò può essere dimostrato, riferendoci per semplicità ad un solo fotogramma, dal fatto che una volta determinata la corrispondenza fra l'epipolo (di coordinate (x_p, x_2, x_3)) e il punto all'infinito di coordinate $(1,0,0)$ restano ancora 4 gradi di libertà (1 rotazione, 2 di traslazione e 1 di scala del piano π) per fissare l'omografia evitando di dar luogo a trasformazioni proiettive indesiderate.

Una prima condizione da imporre è quindi limitare al massimo le deformazioni prospettiche nell'intorno di un punto del fotogramma (ad esempio il centro del fotogramma o il baricentro dell'area che si vuole analizzare nel matching). Si supponga, per semplicità che l'epipolo del fotogramma considerato abbia coordinate $e = (f, 0, 1)$. Si consideri allora la trasformazione omografa individuata dalla matrice:

$$G = \begin{vmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -1/f & 0 & 1 \end{vmatrix} \quad (2.33)$$

E' facile dimostrare che se il punto di interesse è prossimo all'origine (0, 0, 1) la deformazione nell'intorno del punto è approssimativamente nulla: il punto di coordinate generiche (x, y, 1), a seguito dell'applicazione della collinearità risulta trasformato in

$$\mathbf{x}' = G\mathbf{x} = (x, y, 1 - x/f)^T = \left(\frac{x}{1 - x/f}, \frac{y}{1 - x/f}, 1 \right)^T \quad (2.34)$$

Approssimando tali coordinate per mezzo di uno sviluppo in serie di Taylor supponendo che il rapporto $|x/f| \ll 1$, ovvero che l'epipolo si trovi sufficientemente distante dal formato del fotogramma, si ottiene

$$\mathbf{x}' \cong (x(1 + x/f + \dots), y(1 + x/f + \dots), 1) \quad (2.35)$$

Lo jacobiano della trasformazione, in altre parole, risulta:

$$J = \begin{vmatrix} 1 + 2x/f & 0 \\ y/f & 1 + x/f \end{vmatrix} \quad (2.36)$$

che, in virtù delle ipotesi fatte, risulta assimilabile alla matrice identità. Se l'epipolo ha coordinate generiche ed il punto di interesse non si trova nell'origine, detta R la matrice di rotazione richiesta per portare l'epipolo in (f, 0, 1) e T la matrice di trasformazione che rappresenta la traslazione per portare il punto di interesse x_0 nell'origine, l'omografia complessiva può essere ottenuta come

$$H = GRT \quad (2.37)$$

Il problema, a questo punto, consiste nel determinare un'omografia da applicare al secondo fotogramma in modo che due condizioni vengano soddisfatte: innanzitutto anche in questo caso l'epipolo deve essere portato all'infinito, in modo da portare punti omologhi sulla stessa retta orizzontale; inoltre, visto che si può dimostrare che la precedente condizione non è in grado di fissare tutti i gradi di libertà, è necessario stabilire un criterio per fissare i rimanenti parametri liberi: tale condizione può ad esempio essere quella di scegliere la trasformazione che rende minime le parallassi fra le due immagini:

$$\sum_i d(Hx_i, H'x'_i)^2 = \min \quad (2.38)$$

E' possibile dimostrare ([Hartley 2000]) che la prima condizione è verificata se e solo se essa risulta

$$H' = (I + Hea^T)HM \quad \text{con } F = [e]_x M \quad (2.39)$$

in cui \mathbf{a} è un vettore che contiene gli elementi incogniti dell'omografia. Dal momento che, a seguito della trasformazione (2.37) l'epipolo e viene a trovarsi in $(1, 0, 0)$, la quantità tra parentesi nella (2.39) risulta:

$$H_A = (I + H\mathbf{e}\mathbf{e}^T) = I + \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \cdot \begin{bmatrix} \mathbf{a} & \mathbf{b} & \mathbf{c} \end{bmatrix} = \begin{bmatrix} \mathbf{a} & \mathbf{b} & \mathbf{c} \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad (2.40)$$

Detti allora $\hat{\mathbf{x}}_i = H\mathbf{x}_i$, $H_0 = HM$ e $\hat{\mathbf{x}}'_i = H_0\mathbf{x}'_i$ è possibile riscrivere la condizione (2.38) come

$$\sum_i d(\hat{\mathbf{x}}_i, H_A \hat{\mathbf{x}}'_i)^2 = \min \quad (2.41)$$



Figura 2.8: Generazione di immagini epipolari con il metodo proiettivo.

che riscritta in funzione delle componenti (a, b, c) di tale vettore porta a:

$$\sum_i (a\hat{x}_i + b\hat{y}_i + c - \hat{x}'_i)^2 + (\hat{y}_i - \hat{y}'_i)^2 \quad (2.42)$$

Osservando che la seconda quantità $(\hat{y}_i - \hat{y}'_i)$, per le trasformazioni determinate, risulta costante, la quantità da minimizzare risulta:

$$\sum_i (a\hat{x}_i + b\hat{y}_i + c - \hat{x}'_i)^2 \quad (2.43)$$

che può essere risolto semplicemente per mezzo di una stima a minimi quadrati.

In figura 2.8 a pagina precedente viene mostrato l'effetto della trasformazione su una coppia di immagini. Al termine del matching si calcolano le coordinate immagine sui due fotogrammi originali applicando le trasformazioni inverse H^{-1} e $(H')^{-1}$ alle coordinate ottenute.

E' inoltre possibile restituire i punti determinati sulle immagini epipolari in funzione dei parametri di orientamento esterno (si veda [Roncella 2005b] per maggiori dettagli).

L'ipotesi che l'epipolo si trovi distante dal punto di interesse x_0 , e quindi il rapporto x/f sia prossimo a zero, non è sempre verificata, soprattutto in applicazioni close-range. Osservando lo jacobiano della trasformazione (2.36) si comprende che se l'epipolo viene portato all'infinito, con l'approccio descritto, si introducono trasformazioni (deformazioni) del fotogramma troppo elevate per permettere un matching affidabile. Risulta allora più efficiente il metodo introdotto da [Pollefeys 1999], nel quale le immagini vengono ricampionate per mezzo di una trasformazione polare e non più una proiezione. Dalla conoscenza della matrice fondamentale F si ricava la relazione che lega linee epipolari corrispondenti sulle due immagini [Luong 1996]; l'idea è piuttosto intuitiva se ci si riferisce alla figura 1.16 e si considerano le rette corrispondenti determinate dalle intersezioni fra i due piani immagine e un fascio di rette con generatrice la base di presa. L'omografia che mappa una linea epipolare sull'altra risulta:

$$H = [e']_x F + e' a^T \quad (2.44)$$

in cui, ancora una volta, il vettore a è un generico vettore tale che $\det(H) \neq 0$. Il passaggio può anche essere ottenuto considerando un punto qualsiasi su una delle due rette epipolari e calcolando la corrispondente linea sull'altra immagine.

Si considera dunque una trasformazione, per la prima immagine, fra lo spazio cartesiano immagine ed uno spazio in coordinate polari con origine sull'epipolo; una nuova immagine, rappresentante lo spazio $O\rho\theta$ viene ottenuta per ricampionamento; il passo di campionamento in ρ può essere posto pari ad un pixel, mentre per θ si può stabilire un passo fisso o variabile come mostrato in [Pollefeys 1999]. In questo secondo caso per non avere perdite nel contenuto informativo, per ciascuna successiva linea epipolare considerata corrisponde un'ampiezza di θ differente. Le trasformazioni da spazio immagine a spazio ricampionato e viceversa risultano semplicemente

$$\rho = \sqrt{(x - e_x)^2 + (y - e_y)^2}$$

$$\theta = \arctan\left(\frac{(y - e_y)}{(x - e_x)}\right)$$
(2.45)

$$x = \rho \cos \theta$$

$$y = \rho \sin \theta$$
(2.46)

I valori di θ , se a passo variabile, possono essere memorizzati in una tabella che permette di applicare la (2.46) immediatamente note le coordinate sull'immagine epipolare.

Un esempio dell'applicazione della tecnica è presentata in figura 2.9.

In analogia con l'altro metodo, per ottenere la minimizzazione delle parallassi, si può stimare a minimi quadrati un'ulteriore trasformazione affine in modo da verificare la condizione (2.38) fra le due immagini ricampionate:

$$x' = ax + by + c$$

$$y' = y$$
(2.47)

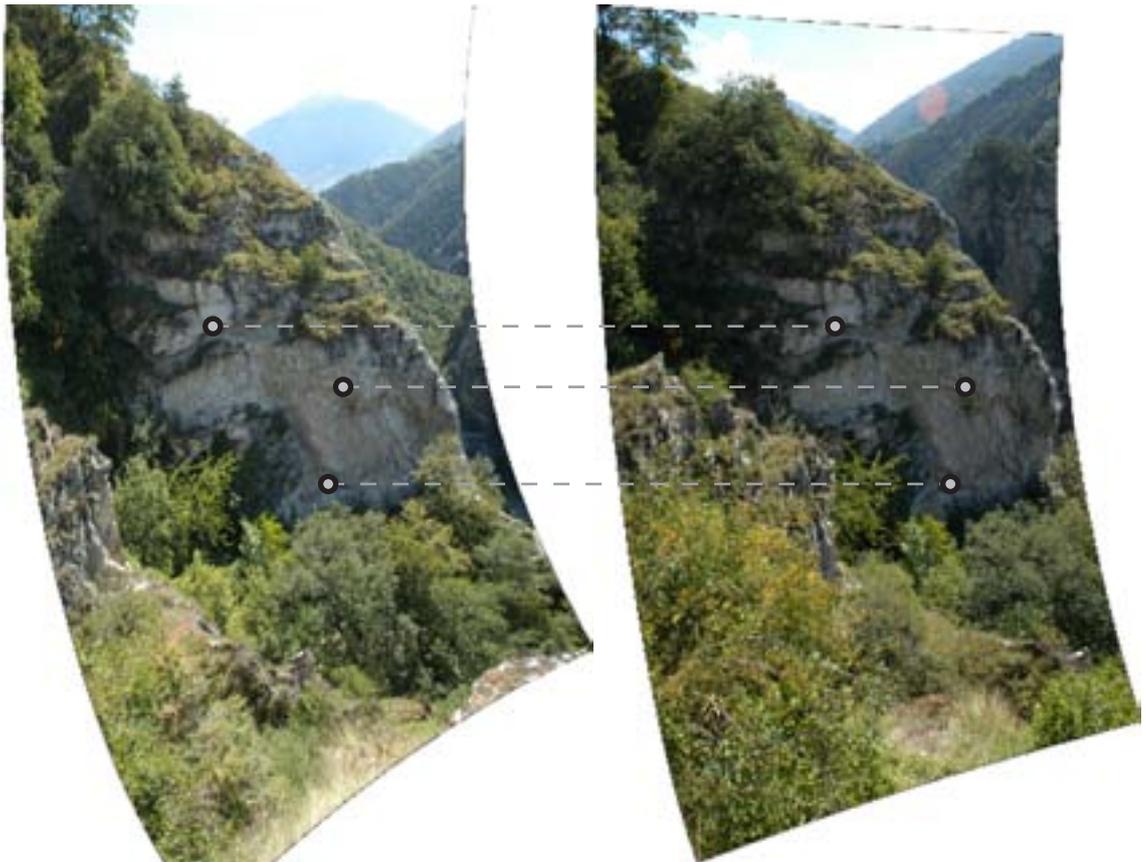


Figura 2.9: Generazione di immagini epipolari con il metodo polare.

Si osservi che la (2.47) è del tutto equivalente alla (2.43).

Sebbene ciò rappresenti un ulteriore passaggio, ed una maggior complicazione per trasformare le coordinate estratte dal matching nelle corrispondenti coordinate riferite allo spazio immagine originale, in alcuni casi può permettere un notevole miglioramento delle prestazioni degli algoritmi di correlazione.

2.5. Metodi di matching denso

Come già anticipato nell'introduzione uno degli scopi principali nello sviluppo di tecniche di correlazione di immagine sempre più efficienti è strettamente legata alla necessità di ottenere, dall'analisi di due o più immagini, la restituzione di un numero di punti molto elevati. Dal momento che tali punti, nell'ottica di una metodologia automatica, non possono essere determinati in posizioni ritenute significative, come potrebbe fare un operatore umano, l'elevato numero di punti estratti ha proprio lo scopo di dare una descrizione generalizzata dell'oggetto scelto. In analogia a quanto viene solitamente fatto per mezzo di strumenti a scansione, una volta ottenuta una nuvola di punti sufficientemente densa e ricostruita la superficie dell'oggetto, le operazioni di misurazione possono avvenire direttamente sul modello digitale ottenuto. Dal momento che il numero di punti richiesto per tale descrizione può anche raggiungere il milione di unità, i metodi utilizzati devono innanzitutto essere molto veloci. Non a caso, nei precedenti paragrafi, si sono più volte illustrate le possibili tecniche per ottimizzare lo sforzo computazionale richiesto dai diversi algoritmi. Inoltre particolare importanza deve essere posta sull'affidabilità offerta dai diversi metodi in quanto, vista la numerosità delle informazioni estratte, risulta impossibile pensare ad un controllo manuale dei risultati.

I metodi di restituzione automatica basati su tecniche di correlazione possono essere suddivise in tre grandi categorie [Rosenfeld 1976], [Zhang 1992]: metodi paralleli, metodi sequenziali, metodi di rilassamento. Nel sistema sviluppato vengono utilizzate tecniche appartenenti alle prime due categorie mentre metodi di rilassamento sono attualmente in fase di sviluppo. Di seguito vengono brevemente descritte le loro differenze.

2.5.1. Metodi paralleli

Si basano sull'applicazione iterata di un dato algoritmo di matching considerando indipendentemente ciascun punto nello spazio immagine: in altre parole il risultato ottenuto nella correlazione di un punto non influenza in alcun modo la stima dei punti che si trovano nelle sue immediate vicinanze.

I parametri iniziali per il matching vengono calcolati inizialmente per ciascun punto da correlare. A tale scopo, le informazioni a disposizione sono le coordinate immagine e terreno dei punti di legame ottenuti in fase di orientamento del blocco. Il loro numero può variare da alcune decine (soprattutto se l'orientamento avviene in modo manuale) ad alcune centinaia uniformemente distribuiti sul formato (nel caso di orientamento automatico). La variazione spaziale dei parametri sul formato, soprattutto se si sta analizzando un

oggetto complesso, può essere estremamente elevata e non è dunque semplice garantire una buona stima dei valori iniziali.

Per superare tale difficoltà si è adottato un approccio multi-risoluzione: generati i diversi livelli della piramide, vengono ricalcolate le coordinate immagine di ciascun punto di legame in ciascun livello; infine, se si utilizza un matching vincolato geometricamente, anche i parametri di orientamento interno vengono opportunamente trasformati.

Lo scopo dell'approccio multi-risoluzione consiste nell'aumentare le corrispondenze fra i vari fotogrammi per ottenere una descrizione più puntuale e localmente corretta dei vari parametri nel matching finale che avviene, come ovvio, al livello di risoluzione originale.

Dai punti di legame, tramite interpolazione del campo di parallassi, si ricavano i parametri di traslazione per le patch sui diversi fotogrammi; qualora fra le diverse immagini intervengano trasformazioni non trascurabili si possono determinare anche i parametri di forma da introdurre nel matching: si considerano tutti i punti di legame presenti sul formato e si determina una triangolazione di Delaunay per individuare, per ciascun punto, i punti di legame più vicini; dalle coordinate di tali punti si stimano a minimi quadrati i parametri della trasformazione utilizzata nel matching, pesando eventualmente le singole osservazioni per mezzo di un criterio basato sulla distanza dal punto considerato:

$$w_i = \frac{1}{\sigma_0^2} \cdot \frac{1}{d^r} \quad (2.48)$$

Affinché tale stima risulti consistente è importante che la distribuzione dei punti e la loro interdistanza sia sufficientemente uniforme su tutto il formato: in caso contrario infatti si possono avere configurazioni geometriche critiche nelle zone di bordo o in zone con lacune di dati.

Dei punti riconosciuti come omologhi al livello di risoluzione minore, vengono accettati solo quelli con correlazione superiore ad una certa soglia (mantenuta moderatamente bassa in quanto in questo livello di risoluzione è raro ottenere valori elevati); essi vengono utilizzati, insieme ai punti di legame originari, per stimare a minimi quadrati i parametri iniziali del matching nel livello di risoluzione successivo. Poiché l'affidabilità dei punti può non essere la medesima, per assegnare il peso si è utilizzato il coefficiente di correlazione con diverse regole empiriche: detto ρ_0 una soglia minima di correlazione sono state sperimentate

$$\tilde{w}_i = \frac{\rho}{\rho_0} \quad (2.49)$$

$$\tilde{w}_i = \frac{(\rho - \rho_0)}{(1 - \rho_0)} \quad (2.50)$$

$$\tilde{w}_i = (\rho - \rho_0)^\lambda \quad (2.51)$$

Delle tre equazioni, la (2.51) ha fornito i risultati migliori con esponenti abbastanza elevati ($2 \div 4$). Le

variazioni sono comunque minime e non si può escludere, visto che la sperimentazione è stata condotta solo su due differenti casi di studio, che in differenti condizioni le tendenze non sia invertite. Sempre riferendoci ai risultati ottenuti finora risulta però interessante notare come l'uso di una qualsiasi delle precedenti equazioni fornisca risultati decisamente migliori se confrontato con i casi in cui la pesatura non viene effettuata. Ciò è dovuto principalmente al fatto che talvolta, soprattutto nei livelli più bassi della piramide, alcuni punti che verificano la soglia di accettazione sono in realtà errati e compromettono in parte le successive operazioni.

2.5.2. Metodi sequenziali

Un approccio differente per il matching denso è rappresentato dai metodi sequenziali: il termine sequenziale deriva dal fatto che il matching viene condotto utilizzando informazioni che vengono perfezionate nel corso della procedura. Partendo da un certo numero di punti di legame (punti seminali o *seed point*) si giunge a realizzare una fitta e densa rete di punti spostandosi, secondo un certo criterio, dal *seed point* che rappresenta un abbinamento preciso e sicuro, verso punti vicini, dei quali non si hanno informazioni; per tali punti vengono scelti dei parametri approssimati desunti da quelli stimati dal punto precedente (ovvero all'inizio dal *seed point*), che solitamente, grazie alla prossimità alla quale si trovano, sono generalmente poco errati; a questo punto l'algoritmo di matching stima i parametri di matching corretti creando, di fatto un nuovo *seed point*. Il processo viene iterato passando da punto a punto in una sorta di *region growing* fino a determinare le posizioni di tutti i punti che compongono il fotogramma.

Si deduce da quanto detto che il posizionamento e la distribuzione dei *seed point* giocano un ruolo essenziale per il risultato finale: devono infatti essere sufficientemente numerosi e distribuiti con maggior densità nelle zone più irregolari in cui i parametri del matching cambiano in maniera più improvvisa.

Le possibili strategie di matching definiscono come l'intera immagine venga successivamente elaborata nel corso della procedura: in particolare stabiliscono in che modo debba avvenire la crescita di punti a partire dalla rete di *seed point*. Un passo fondamentale, per ridurre i tempi di elaborazione e facilitare le operazioni andando ad analizzare solamente le aree del fotogramma in cui è presente contenuto informativo, consiste nella definizione nella *master image* della regione d'interesse (*region of interest* o ROI): molto spesso, infatti non è necessario studiare l'intero fotogramma ma solo una porzione di esso. Una volta definita la ROI sull'immagine di riferimento, essa è di fatto implicitamente definita anche sugli altri fotogrammi, in quanto vengono esclusi tutti i *seed point* che non le appartengono.

Per quanto riguarda le traiettorie di crescita dai punti seminali si è seguita l'impostazione suggerita in [D'Apuzzo 2003]: l'immagine, o meglio la porzione interessante di essa, viene suddivisa in tante regioni poligonali per mezzo di una tassellazione di Voronoi: come noto tale tecnica permette di individuare per ciascun punto seminale il luogo dei punti che presentano distanza minore da tale punto rispetto a tutti gli altri. In realtà la tassellazione può avvenire in maniera discreta, il che facilita notevolmente il processo:

dal momento che i punti da determinare solitamente vengono scelti utilizzando una griglia regolare, per ciascun nodo di tale griglia si determina il punto seminale più vicino.

Si può a questo punto inizializzare il processo di densificazione delle corrispondenze: si considera uno qualsiasi dei seed point; dalla coordinate dei seed point adiacenti (determinati per mezzo di una triangolazione di Delaunay) si determinano i parametri di una trasformazione affine fra i punti sul fotogramma di riferimento e le slave image: tali informazioni permettono di definire i parametri iniziali per i primi punti ottenuti partendo dal punto seminale; una delle possibili traiettorie di crescita è schematizzata in figura 2.10: ad ogni passo, ovvero per ogni nuovo punto elaborato, si utilizzano come parametri approssimati i valori stimati al passo precedente; è possibile utilizzare uno qualsiasi dei metodi di correlazione visti

precedentemente (semplice, LSM, MGCM); al termine della stima si valuta il grado di correlazione fra la template e la/le patch: se il risultato ottenuto è accettabile, nel senso che il coefficiente di correlazione è più alto di una certa soglia, si assume che il punto sia stato correttamente determinato e i parametri stimati vengono passati al punto successivo; nel caso contrario si può ritentare il processo di matching variando alcuni parametri (solitamente la posizione del punto o la dimensione della template); se infine, anche in questo caso, non si è in grado di stimare correttamente il punto, si procede al passo successivo utilizzando i parametri approssimati del punto precedente. Una volta esauriti tutti i punti relativi ad un seed point si passa ad un nuovo punto seminale e il processo viene reiterato fino alla copertura totale del fotogramma o della regione di interesse.

Dal momento che, quando la procedura incorre in un punto non determinabile, non si riesce a riottenere, nel continuo della traiettoria di matching, punti affidabili, solitamente al termine della procedura una o più regioni del fotogramma (o della ROI) risultano prive di punti: in tal caso, per mezzo di una finestra di ricerca, si determinano quali zone non siano state effettivamente determinate e si cerca di ottenere dei matching affidabili ora che, al termine della procedura, si hanno molte più informazioni sulle caratteristiche dei punti vicini. Tale procedura è, a nostro avviso, preferibilmente da condursi con metodi paralleli, in quanto nelle regioni da analizzare i metodi sequenziali hanno già mostrato di non essere affidabili: per ciascuna punto della griglia all'interno delle regioni non correlate, si determinano i parametri approssimati interpolando le informazioni ricavabili dai punti omologhi sul contorno della regione stessa e si procede con un nuovo

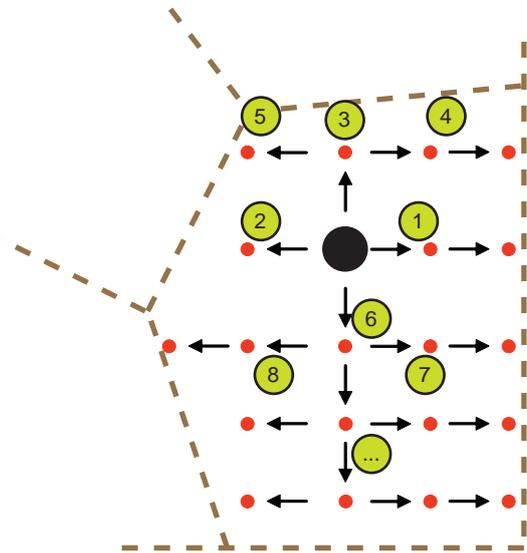


Figura 2.10: Schema di una delle possibili traiettorie di crescita nel metodo sequenziale: i numeri indicano la successione degli insiemi di punti presi in considerazione.

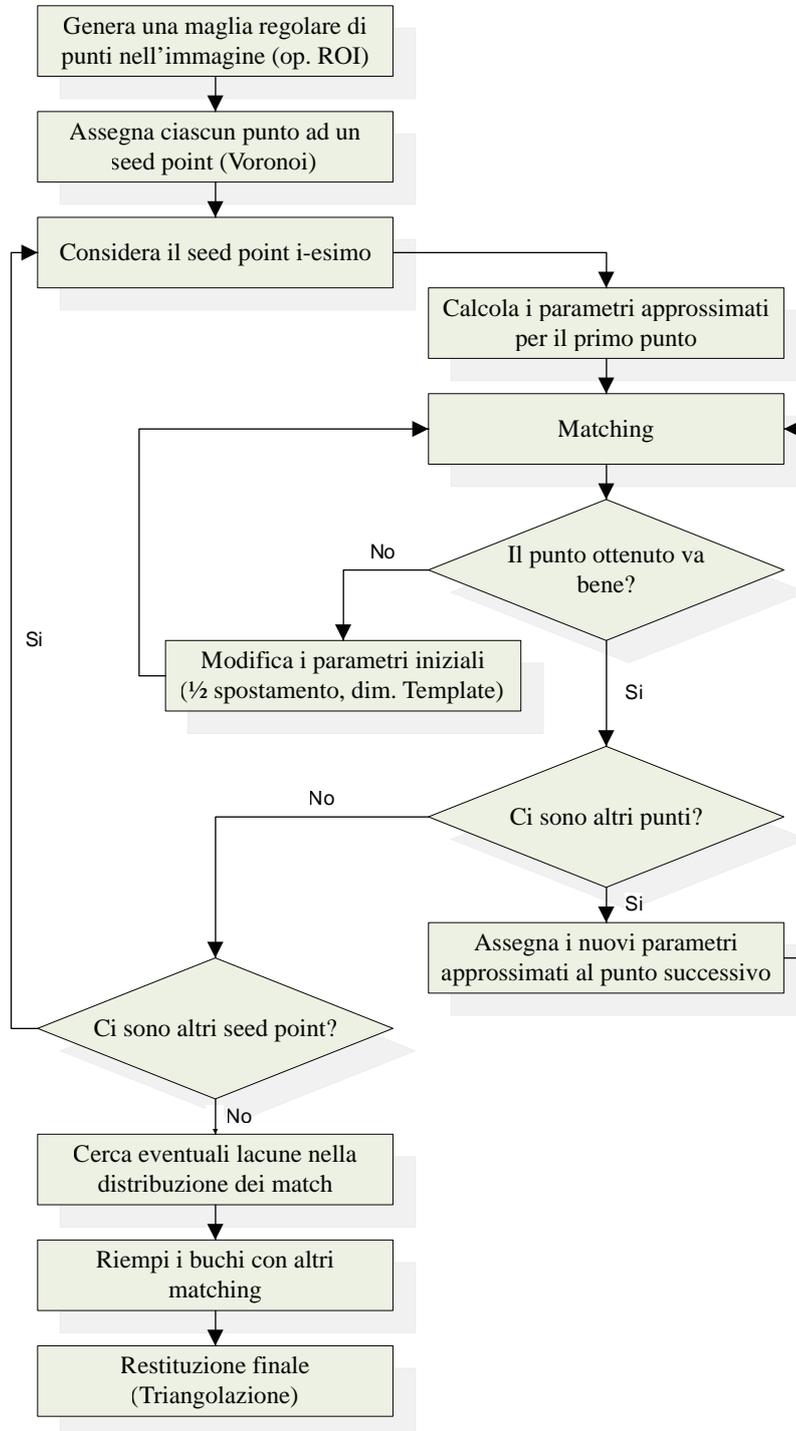


Figura 2.11: Schema a blocchi delle principali operazioni utilizzate nel metodo di matching denso sequenziale.

processo di stima.

Lo schema dell'intera procedura è riassunto in figura 2.11.

2.5.3. Metodi di rilassamento

Entrambe le impostazioni precedentemente illustrate mostrano caratteristiche che, a seconda dei casi, possono rendere stabile il processo o compromettere seriamente i risultati; i metodi sequenziali possono andare in crisi se, in alcune zone, durante la fase di crescita dai seed point, si incorre in punti in cui il matching non riesce a ottenere la corrispondenza cercata o stima dei parametri errati di cui però non è possibile valutare l'attendibilità; viceversa i metodi paralleli si dimostrano più robusti da questo punto di vista, nel senso che matching erronei in alcune zone del modello non vanno ad inficiare le procedure in altre regioni, ma richiedono un'ottima conoscenza iniziale dei parametri approssimati e non tendono ad usufruire di possibili informazioni messe in luce durante il procedimento (ciò non è del tutto vero se si utilizza l'approccio piramidale, nel qual caso si ha un passaggio di informazioni fra un livello di risoluzione e il seguente).

Un approccio differente, che permette in parte di sfruttare i punti di forza di entrambe le impostazioni, è rappresentato dai metodi di rilassamento ([Rosenfeld 1976], [Price 1985], [Brown 1992], [Zheng 2004]): al momento è ancora in fase di studio l'applicazione di tale metodologia all'interno del sistema sviluppato e pertanto, in questa sede, essi non vengono descritti.

2.6. Fotogrammetria o Laser Scanner?

Il mercato legato alle scienze del rilevamento è particolarmente fertile e la continua evoluzione della computer grafica, ormai da più di una decade, favorisce l'aumento nella richiesta di modelli digitali sempre più realistici e complessi. Al momento le due tecnologie che possono fornire questo tipo di informazione sono la fotogrammetria e il laser scanning: mentre la prima richiede però lo sviluppo di algoritmi piuttosto complessi per la restituzione automatica e competenza tecnica per una buona progettazione del blocco ed uno studio delle condizioni particolari di ciascun caso, la seconda è più immediata ed è in grado di fornire risultati in maniera più diretta e, sotto certi punti di vista, più affidabile. Viene dunque da chiedersi se il continuo sviluppo di questo tipo di tecnologia, unitamente alla diminuzione dei costi delle apparecchiature e del software di gestione, non significhi il lento ma inesorabile declino della fotogrammetria, almeno in questo campo applicativo. Al tempo stesso ci si interroga su quali siano effettivamente i limiti dei due approcci e quali possano essere i criteri oggettivi per la scelta di un metodo piuttosto dell'altro. La vastità dell'argomento è tale che in questa sede non è possibile addentrarsi eccessivamente nel soggetto; ciononostante ci sembra doveroso esprimere alcune osservazioni a riguardo rimandando alla letteratura per un quadro più completo. Numerose ricerche sono state condotte ([Beraldin 2004], [Habib 2004], [Alshawabkeh 2004], etc...) sulla possibile integrazione delle due metodologie: i tempi ridotti di acquisizione della nuvola di

punti non sono necessariamente compensati da quelli di restituzione delle primitive geometriche; al tempo stesso le caratteristiche tecniche degli strumenti attualmente in commercio non permettono di rinunciare alla presa di immagini per la realizzazione di ortofoto o modelli 3D fotorealistici. Come mostrato in [Kadobayashi 2004], in [Furkuo 2004] o in [Lee 2004] da un lato alcuni ritengono che l'integrazione di fotogrammetria e laser si limiti a questo aspetto, mentre altri caldeggiavano un grado di integrazione superiore. Ulteriori aspetti che devono essere presi in considerazione, a nostro avviso, riguardano la versatilità di una strumentazione nei confronti dell'altra, in relazione soprattutto al tipo di precisioni che può essere richiesto: da questo punto di vista diversi sono i lavori sviluppati negli ultimi anni che affrontano approfonditamente la tematica ([Fryer 2005], [Remondino 2004], [Boëlher 2004] o il già citato [Habib 2004]).

In figura 2.12 e 2.13 sono presentati due casi di studio illustrati in [Roncella 2005d] i cui risultati sono riassunti in tabella 2.1; per una descrizione più approfondita dei casi analizzati e delle caratteristiche dei rilievi si rimanda al testo citato. La fotogrammetria e le tecniche di laser scanning sono di fatto equivalenti dal punto di vista delle precisioni ottenibili: una adeguata progettazione del blocco fotogrammetrico permette infatti di ottenere il grado di precisione

richiesto, così come l'utilizzo di strumentazione laser adeguata permette di ottenere intervalli di precisione abbastanza ampi; da questo punto di vista le tecniche fotogrammetriche garantiscono una notevole versatilità se confrontate con le limitazioni della strumentazione laser: mentre con l'approccio fotogrammetrico la scala dell'oggetto influenza relativamente poco il tipo di strumentazione richiesto, nel senso che con una stessa camera ed un numero limitato di ottiche è possibile affrontare problemi caratterizzati da scale dimen-

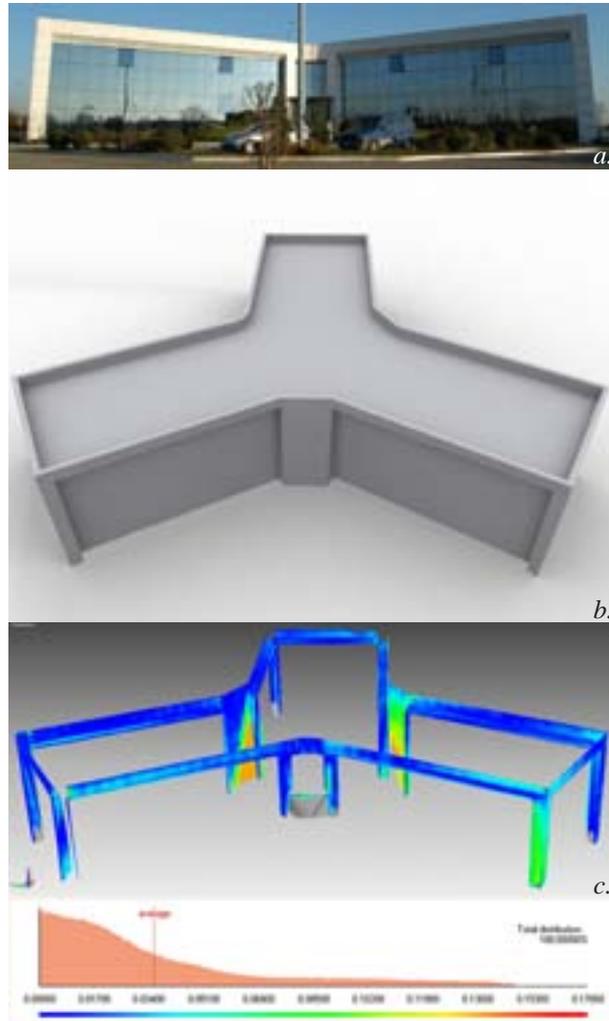


Figura 2.12: Uno dei due casi di studio considerati: (a.) L'edificio sede di Autocamionale della Cisa S.p.A.; (b.) Ricostruzione fotogrammetrica dell'edificio; (c.) Confronto fra modello restituito fotogrammetricamente e ottenuto tramite laser scanner. Le zone vetrate non sono state utilizzate in quanto il laser scanner non è stato in grado di restituirle.

sionali anche molto differenti, con strumentazione laser, i cui costi come tutti sappiamo sono tutt'altro che trascurabili, un sistema è in grado di garantire risultati accettabili solo entro un ben determinato intervallo di distanze.

La risoluzione sull'oggetto è, a parità di distanza, solitamente maggiore nel caso delle tecniche a scansione: sono oggi presenti sul mercato strumenti in grado di sviluppare risoluzioni angolari dell'ordine delle decine di μrad , mentre è facile verificare con semplici calcoli che le risoluzioni nel caso della fotogrammetria all'incirca dieci volte inferiori; supponendo ad esempio di utilizzare focali lunghe ($80\div 150\text{ mm}$) e considerando che, per non ottenere risultati troppo correlati, si può pensare di estrarre punti sull'immagine ogni $4\div 8$ pixel, con i sensori attualmente in commercio si ottengono risoluzioni angolari dell'ordine di $300\div 700\ \mu rad$. Inoltre, con ottiche così lunghe il campo inquadrato risulta solitamente piuttosto ristretto e possono essere richiesti diversi fotogrammi per coprire l'intera superficie da analizzare. Anche in questo senso, dunque, se sono richieste densità di punti elevati su una superficie piuttosto ampia e se esistono limitazioni per la realizzazione del blocco (ad esempio non è possibile avvicinarsi molto all'oggetto) le tecniche fotogrammetriche cedono il passo di fronte al laser a scansione.

Se al contrario la tipologia del rilievo richiede una buona maneggevolezza della strumentazione o è richiesto di raggiungere zone impervie, le attuali strumentazioni laser non si dimostrano molto versatili: il peso della strumentazione varia dai 10 ai 13 kg (Riegl, ILRIS,

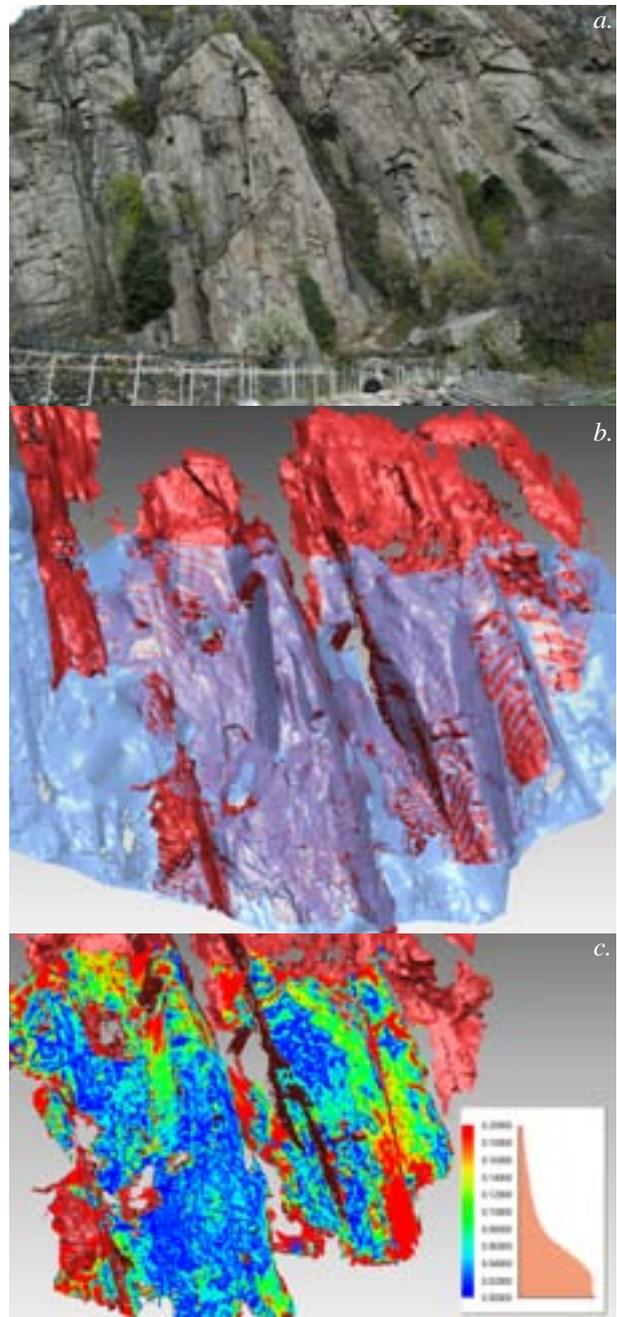


Figura 2.13: (a.) Una delle pareti di Arnad (vd. § 1.8.2.; (b.) Ricostruzione fotogrammetrica (in rosso) e mediante laser a scansione della parete (in azzurro) gentilmente messo a disposizione dal Politecnico di Torino; (c.) Confronto fra i due DSM. I risultati sono riassunti in tabella 2.1.

Leica, etc...) ma bisogna anche considerare il peso del treppiede, delle batterie e dei vari accessori, che portano spesso il peso complessivo della strumentazione intorno ai 25÷30 kg.

2.7. Simulazioni, Test e Applicazioni

In questo paragrafo vengono illustrati i risultati conseguiti con i metodi di matching sopra descritti, relativamente ad alcune simulazioni svolte nel corso dell'implementazione software e alle prime applicazioni a casi reali.

Si è posta attenzione innanzitutto ad una valutazione delle precisioni di misura delle coordinate pixel, a questo proposito motivata anche dallo sviluppo di un sistema di misura del campo di deformazione di superfici piane; tale metodologia è stata efficacemente applicata allo studio di provini di vario tipo e di vari materiali, soggetti a prove di rottura: essendo le precisioni attese particolarmente elevate, si è effettuato uno studio approfondito delle prestazioni raggiungibili dagli algoritmi illustrati.

Sono poi stati condotti alcuni test per verificare le differenze in termini di precisione e affidabilità fra i metodi geometricamente vincolati (MGCM) e quelli non condizionati (LSM): in tal caso, si è preferito utilizzare una serie di immagini generate artificialmente per poter controllare anche la successiva fase di restituzione.

Infine viene presentata l'applicazione dei metodi di matching denso per la generazione di nuvole di punti nello studio delle caratteristiche geometriche superficiali di diversi oggetti: come si vedrà la generalità delle tecniche permette la loro applicazione in diversi campi, dalla geotecnica, all'architettura, etc.

2.7.1. Precisioni di misura del least squares matching

Lo studio delle precisioni di misura ottenibili per mezzo del least squares matching è stato effettuato ponendosi in condizioni ideali, in cui tutte le ipotesi fatte nello sviluppo della metodologia sono approssimativamente verificate: che la superficie ripresa sia sostanzialmente piana e il cambiamento del punto di vista possa essere assimilato, in termini di trasformazione dei punti del piano immagine, ad una trasformazione affine, quanto meno in un ambito locale relativo alle dimensioni della template; che il contenuto informativo delle immagini, ovvero la caratterizzazione della tessitura rappresentata nel fotogramma, garantisca una buona determinabilità dei parametri.

Le prove sono state inizialmente condotte utilizzando immagini a scale di grigio generate artificialmente, costituite da cerchi neri su sfondo bianco; a tali immagini sono state applicate una serie di trasformazioni e sono state utilizzate, nelle diverse prove, dimensioni delle template differenti; per tenere in conto le possibili differenze fra le varie parti dell'immagine, i centri delle template sono stati presi, per ciascuna prova, su una griglia regolare che copriva approssimativamente tutto il formato, costituita da circa 40 nodi. La seguente tabella illustra i risultati salienti emersi:

Errori medi ed errori quadratici medi sui 40 nodi della griglia (<i>pixel</i>)		30 Pixel		20 Pixel	
		X	Y	X	Y
Traslazione:	Media	-0.00148	-0.00479	-0.00154	-0.00743
	RMSE	0.00493	0.00394	0.00677	0.00902
	1/RMSE	202.7	253.9	147.7	110.9
Traslazione*:	Media	-0.00136	-0.00461	-0.00101	-0.00693
	RMSE	0.00501	0.00589	0.00630	0.00837
	1/RMSE	199.4	169.8	158.6	119.5
Scalatura:	Media	0.00260	0.00011	0.00058	-0.00067
	RMSE	0.00373	0.00659	0.00342	0.00604
	1/RMSE	267.9	151.8	292.8	165.4
Rotazione:	Media	0.00241	0.00265	-0.00331	-0.00372
	RMSE	0.00420	0.00396	0.00796	0.00893
	1/RMSE	238.2	252.6	125.6	112.0

Tabella 2.1: Riepilogo dei dati emersi dallo studio delle precisioni di misura del LSM utilizzando differenti dimensioni di template e applicando trasformazioni differenti all'immagine originale.

Le prime due analisi riguardano trasformazioni di traslazione: per uniformità con le altre prove si è ricampionata l'immagine introducendo uno spostamento di 4÷5 pixel in entrambe le direzioni; nel primo caso si sono stimati tutti i parametri dell'affine nel LSM, mentre nel secondo (indicato con un asterisco in apice) si è svolta la procedura utilizzando solamente i parametri di traslazione, ovvero si è utilizzato un modello funzionale più aderente alla realtà: come si può osservare le precisioni tendono ad essere migliori di circa il 50% utilizzando la template di dimensioni maggiori; tuttavia, in quest'ultimo caso si presentano maggiori precisioni nel caso della stima a pieni parametri, mentre con template da 20 pixel si dimostra più accurato il modello non sovrapparametrizzato: ciò può essere in parte dovuto al fatto che, con il modello minimo, si ha una maggior sensibilità alla presenza di rumore (dovuto al ricampionamento).

Nel caso della trasformazione di scala (si è aumentata la dimensione dell'immagine del 5%) i risultati sono all'incirca analoghi, con precisioni lievemente superiori, mentre nella rotazione del piano immagine (di circa 2°) si nota maggiormente il vantaggio di utilizzare template di dimensioni maggiori: è abbastanza intuitivo che, avendo un maggior braccio, la stima dei parametri di rotazione riesca ad essere condotta in maniera più affidabile, conducendo a stime più corrette.

I valori ottenuti in questa serie di test hanno messo in luce precisioni di misura un po' troppo ottimistiche (all'incirca 1/200 di pixel nel posizionamento delle patch con template da 30 pixel) se confrontate con i valori solitamente presentati in letteratura che difficilmente superano il 1/100 di pixel. Bisogna però considerare che in questo caso specifico, pressoché tutte le ipotesi sulle quali si basa il metodo erano verificate: condizioni forse un po' troppo ideali per fornire una valutazione delle reali capacità operative del metodo.

Si è deciso dunque di sviluppare un'altra verifica, questa volta utilizzando sistemi di acquisizioni reali, in modo da tenere in conto tutte le possibili cause di errore (rumore del sensore, non perfetta pianeità della superficie, vibrazioni della camera, etc...). Si è quindi realizzata una superficie approssimativamente piana la quale, per permettere di avere una buona tessitura, è stata dipinta con vernici acriliche; in questo caso sono stati provati due diversi pattern: uno ancora a piccoli cerchi neri su sfondo bianco, l'altro a sottili linee intrecciate sempre su sfondo chiaro. La superficie ottenuta è stata applicata ad una slitta micrometrica con risoluzione da $10\ \mu\text{m}$ e precisione (dichiarata) di $1\ \mu\text{m}$; la fotocamera utilizzata per la prova, una Basler Af101 con ottica da $8\ \text{mm}$, è stata posizionata al di sopra della slitta, ad una distanza tale che la precisione dichiarata per lo strumento nel sistema oggetto, corrispondesse ad $1/1000$ di pixel in quello immagine.

Al termine delle riprese si sono processati i fotogrammi, andando a considerare nuovamente più punti per ciascuna immagine; per riportarsi alla scala terreno, in cui erano espressi gli spostamenti della slitta micrometrica, si sono considerate le posizioni iniziali e finali (cioè all'inizio ed alla fine della sequenza) dei punti e si è determinata una omografia monodimensionale (in quanto il movimento avveniva su una linea retta), che ha permesso di stimare le posizioni nel sistema oggetto di tutti i punti in tutti i fotogrammi intermedi. I risultati di entrambe le prove sono riportati in figura 2.14 (valori medi fra le due prove): come si può

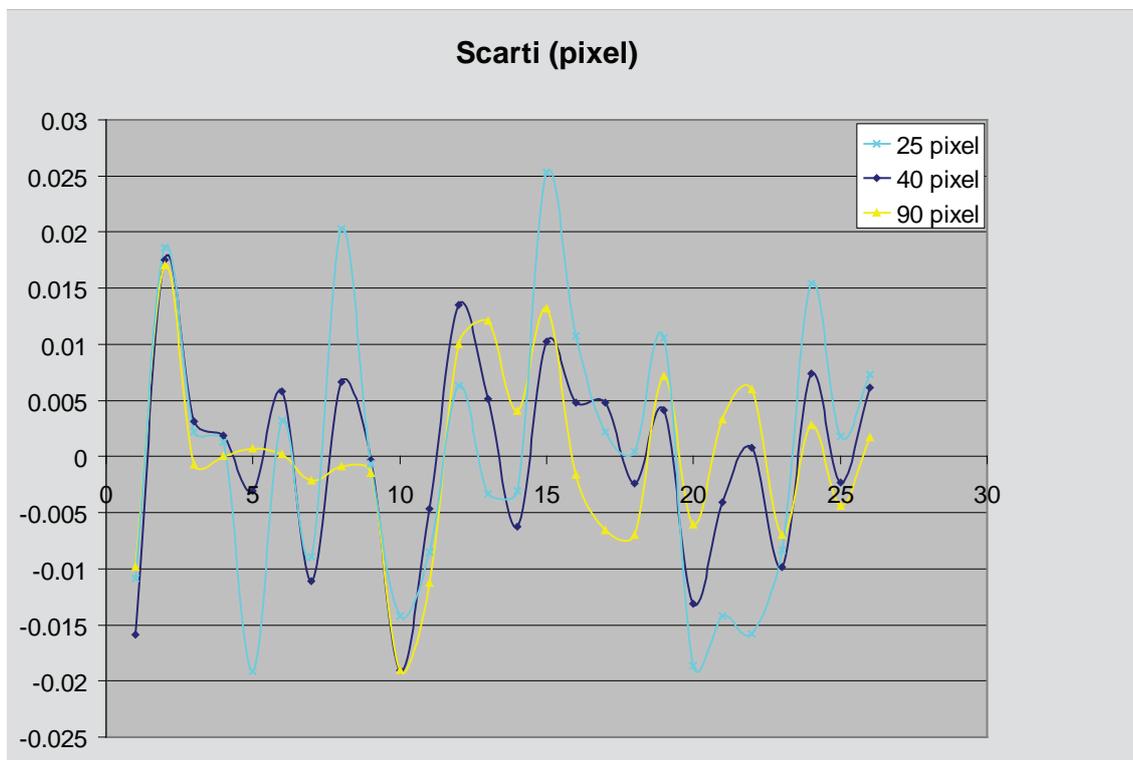


Figura 2.14: Differenze di posizione misurate in corrispondenza di spostamenti discreti della slitta micrometrica nel caso della tessitura a strisce.

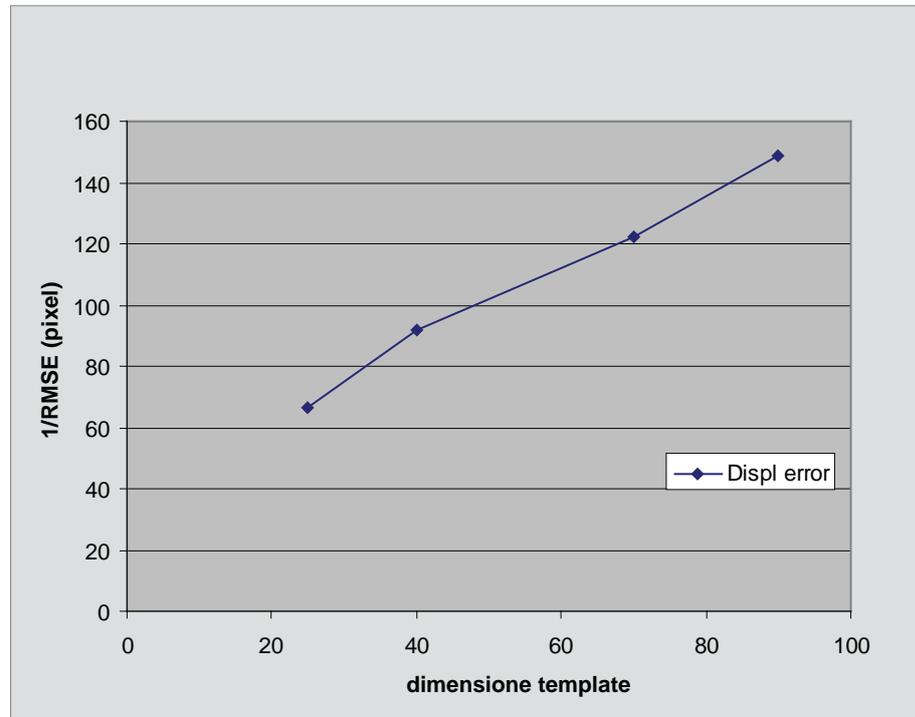


Figura 2.15: Errori di posizionamento medi in funzione della dimensione della template, nella prova eseguita con la slitta micrometrica.

notare, anche in questo caso si assiste ad un aumento delle precisioni di stima all'aumentare della dimensione della template; mediamente si può asserire che le precisioni ottenibili sono dell'ordine di $1/80 \div 1/120$ di pixel; tuttavia, il confronto fra le due prove (non riportato per necessità di sintesi) mostra come alcuni dei massimi residui si trovino all'incirca in corrispondenza delle stesse posizioni della slitta: ciò ci porta a credere che un tale sistematismo, essendo state le due prove condotte indipendentemente, possa essere solamente interpretato come una non perfetta calibrazione della slitta stessa; se ciò fosse vero precisioni mediamente superiori (dell'ordine di $1/150$ di pixel) sarebbero da aspettarsi.

In figura 2.15 viene infine mostrato l'andamento medio delle precisioni di posizionamento delle patch in funzione delle dimensioni della template: a questo riguardo bisogna ricordare che anche in questo caso la condizione di moto della slitta rende il caso ottimale, in quanto il tipo di trasformazione fra due fotogrammi successivi è del tutto assimilabile ad una traslazione rigida; in altri termini si assiste ad un continuo aumento delle precisioni di misura, anche per dimensioni non trascurabili della template, in virtù della particolarità della prova; in altri casi, come noto, la scelta di una template eccessivamente grande (più di 50 pixel) rende difficilmente verificata l'ipotesi di superficie dell'oggetto localmente piana, il che rende il modello affine non appropriato a interpretare il fenomeno, con conseguente riduzione delle precisioni ottenibili.

Viste le elevate prestazioni evidenziate dall'algoritmo si è messo a punto un sistema di misurazione per

l'analisi del campo deformativo di campioni di materiale soggetti a prove di rottura: lo studio della meccanica della frattura è un tema di grande attualità in quasi tutte le diverse branche dell'ingegneria; al momento i metodi per ottenere una descrizione accurata del campo di spostamenti a cui è soggetto un elemento in fase deformativa e fessurativa sono pochi e generalmente molto costosi (ESPI); inoltre molto spesso forniscono informazioni monodimensionali.

Quando un materiale raggiunge la sua fase plastica, gli spostamenti, e di conseguenza le deformazioni, tendono a diventare fenomeni localizzati in zone di particolare debolezza: per lo studio del fenomeno è quindi richiesta una metodologia che fornisca informazioni in tutti i punti della superficie del provino.

Riprendendo il materiale, durante una prova di carico, per mezzo di una fotocamera che permetta di scattare immagini in sequenza abbastanza rapidamente (solitamente sono sufficienti dai 2 ai 5 fotogrammi al secondo), applicando in seguito le tecniche di matching, si può arrivare ad ottenere una descrizione del campo di spostamenti a cui è stato soggetto il provino. Ovviamente, se l'approccio è monoscopico, il moto superficiale del provino deve essere approssimativamente piano, in modo da poter trasformare gli spostamenti nel piano immagine in corrispondenti spostamenti nello spazio oggetto.

Una volta ottenuto il campo di spostamenti, è possibile per mezzo di un approccio alle differenze finite, dal momento che i punti vengono determinati su una griglia, ottenere il tensore di deformazione in qualsiasi punto dell'immagine:

$$S = \begin{vmatrix} \frac{\Delta u_x}{\Delta x} & \frac{1}{2} \left(\frac{\Delta v_x}{\Delta x} + \frac{\Delta u_y}{\Delta y} \right) \\ \frac{1}{2} \left(\frac{\Delta v_x}{\Delta x} + \frac{\Delta u_y}{\Delta y} \right) & \frac{\Delta v_y}{\Delta y} \end{vmatrix} \quad (2.52)$$

in cui u e v sono le componenti di spostamento, rispettivamente in direzione x e in direzione y .

La (2.52) permette di determinare ciascuna componente di deformazione una volta noto il campo di spostamenti del provino: una rappresentazione a pseudo-colori in un test di trazione indiretta di provini in conglomerato bituminoso è riportato in figura 2.16.

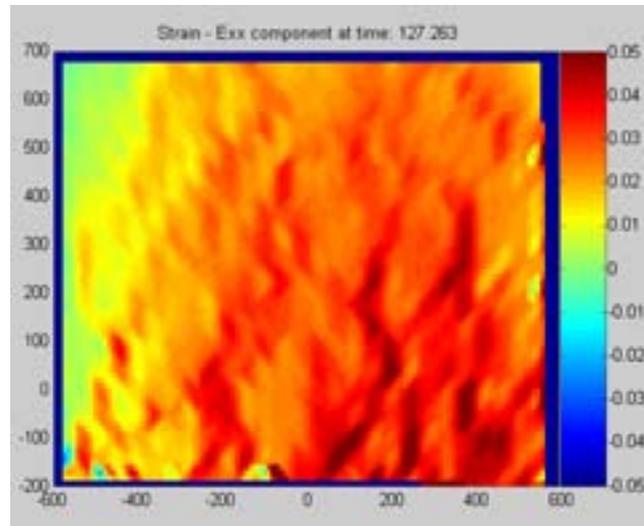


Figura 2.16: Rappresentazione del campo di deformazione ε_{xx} di un provino di conglomerato bituminoso in un test di trazione indiretta.

Inoltre, dalla propagazione della varianza, ricordando che

$$\begin{aligned}\Delta u_x(i, j) &= u(i+1, j) - u(i, j) \quad , \quad \Delta u_y(i, j) = u(i, j+1) - u(i, j) \\ \Delta v_x(i, j) &= v(i+1, j) - v(i, j) \quad , \quad \Delta v_y(i, j) = v(i, j+1) - v(i, j)\end{aligned}\quad (2.53)$$

è immediato ricavare, nota la precisione di misura degli spostamenti σ_u , e detti Δx e Δy i passi di maglia della griglia di punti, la corrispondente precisione nella stima del tensore:

$$\begin{aligned}\varepsilon_{xx} &= \frac{u(i+1, j) - u(i, j)}{\Delta x} \Rightarrow \sigma_{\varepsilon} = \frac{\sqrt{2}}{\Delta x} \sigma_u \\ \varepsilon_{yy} &= \frac{1}{2} \left(\frac{u(i, j+1) - u(i, j)}{\Delta y} + \frac{v(i+1, j) - v(i, j)}{\Delta x} \right) \Rightarrow \sigma_{\varepsilon} = \sqrt{\frac{1}{2\Delta^2 x} + \frac{1}{2\Delta^2 y}} \sigma_u = \frac{1}{\Delta x} \sigma_u\end{aligned}\quad (2.54)$$

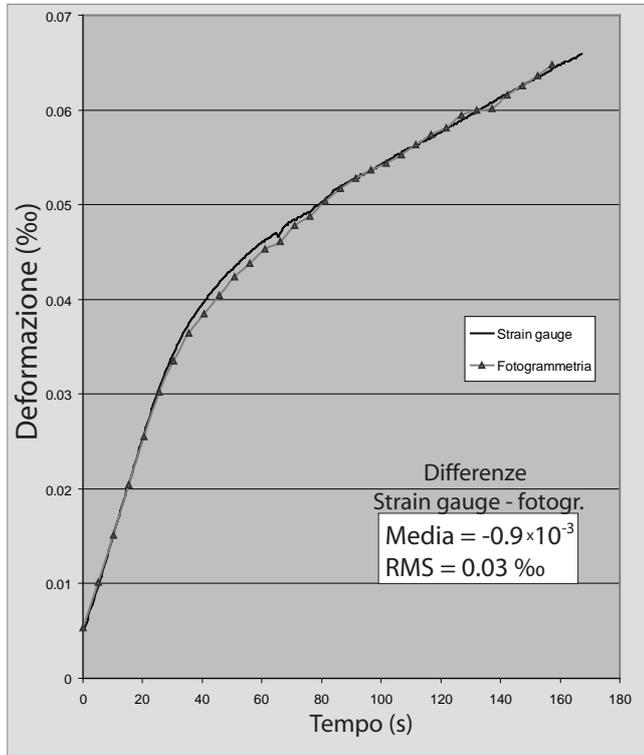


Figura 2.17: Deformazione di un provino d'acciaio in una prova di trazione monoassiale misurata per mezzo di uno strain gauge e con il metodo fotogrammetrico.

in cui nell'ultima uguaglianza si è supposto che $\Delta x = \Delta y$: la precisione di misura del tensore è direttamente proporzionale alla precisione di determinazione degli spostamenti ed inversamente proporzionale al passo di griglia.

Diversi test sono stati condotti per verificare se le precisioni di misura ottenute nelle prove di figura 2.14 e 2.15 fossero analoghe a quelle ottenibili nel caso reale, utilizzando materiali e condizioni di presa differenti. In figura 2.17 è mostrato il confronto fra i dati ottenuti per mezzo del metodo fotogrammetrico e uno strain gauge, nello studio di un provino d'acciaio in una prova di trazione monoassiale: dal momento che lo strain gauge fornisce una misura mediata su tutta la sua estensione dei valori di deformazione assiale, si è proceduto, per il confronto, a calcolare la media ad ogni istante dei dati puntuali ottenuti con il nostro metodo; l'utilizzo di un materiale rego-

lare come l'acciaio permette di avere un campo di deformazione all'incirca uniforme e, pertanto, il valore medio corrisponde approssimativamente a tutti i valori puntuali misurati.

In questo caso, le precisioni stimate con la (2.54), devono essere opportunamente modificate, in quanto

la misura del valore della deformazione deriva da n osservazioni indipendenti:

$$\sigma_{\bar{\varepsilon}} = \frac{1}{\sqrt{n}} \sigma_{\varepsilon} \quad (2.55)$$

Avendo utilizzato un passo di griglia di 35 pixel, supponendo una precisione di misura prossima a 1/90 di pixel (le template utilizzate nella prova erano grandi 40 pixel per lato), ed avendo utilizzato approssimativamente 300 punti per la stima, si ottiene dalla (2.54) e dalla (2.55) una precisione stimata della deformazione media pari a $\pm 0.026\%$, valore abbastanza vicino a quello osservato sperimentalmente.

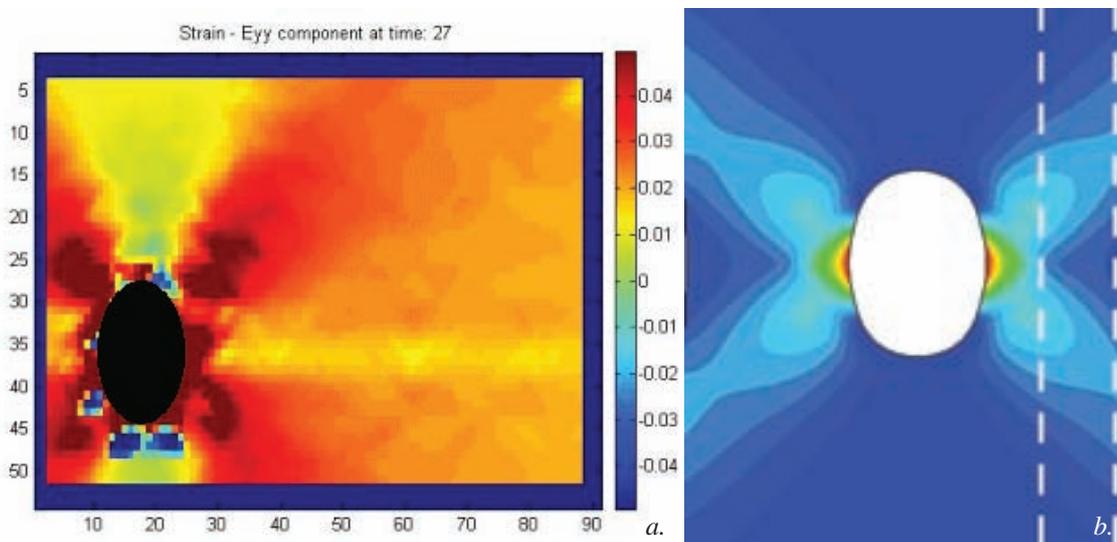


Figura 2.18: Campo di deformazioni di un provino d'acciaio in cui è stato praticato un foro in una prova di trazione monoassiale: (a.) metodo fotogrammetrico; (b.) simulazione numerica agli elementi finiti.

In figura 2.18 è rappresentato il campo di deformazione per la componente in x in una prova trazione di un provino in acciaio in cui è stato realizzato un foro, e la corrispondente mappa di deformazione teorica ottenuta mediante un modello agli elementi finiti: come si può notare facilmente, le caratteristiche qualitative dei due diagrammi sono molto simili.

Per maggiori dettagli sul metodo e per una più ampia panoramica di applicazioni si rimanda a [Roncella 2005e].

2.7.2. Least squares matching e MPGCM

Un altro aspetto abbastanza interessante da approfondire riguarda i vantaggi nell'utilizzo delle condizioni di collinearità, nell'utilizzo di algoritmi MGC: molto spesso l'utilizzo di tali algoritmi comporta un

aumento del carico computazionale richiesto dall'elaboratore, a causa delle dimensioni maggiori delle matrici normali nella stima a minimi quadrati, della necessità di svolgere operazioni aggiuntive e della maggior occupazione di memoria di tutte le immagini interessate dal processo. Si vuole cioè capire se l'utilizzo di algoritmi più raffinati presenti dei benefici rispetto al LSM, tali da giustificare i più lunghi tempi di elaborazione.

Per svolgere la prova si è utilizzato un modello digitale, presentato in § 1.8.1., che permettesse di valutare la precisione e l'affidabilità di restituzione ottenuta dai punti individuati con le due tecniche.

In particolare si è utilizzato l'algoritmo per il matching denso sequenziale presentato in § 2.5.2., sfruttando in un caso il LSM, nell'altro l'MGCM: in entrambe le prove si sono utilizzati tre fotogrammi, per cui, nel caso del LSM che affronta due immagini per volta, l'algoritmo è stato applicato due volte. Al termine si è proceduto alla restituzione della nuvola di punti triangolando a minimi quadrati e si è andato a verificare il grado di coerenza con la mesh poligonale originale del modello. I risultati sono rappresentati in figura 2.19 e 2.20. Sostanzialmente gli algoritmi, dal punto di vista qualitativo, raggiungono approssimativamente lo stesso risultato: le zone in cui si hanno differenze maggiori rispetto alla mesh originale sono le stesse utilizzando il LSM e l'MGCM; tuttavia, da un punto di vista quantitativo, come si può vedere in tabella 2.2, è quest'ultimo a fornire risultati migliori. Possiamo quindi concludere, sebbene sarebbe necessario un maggior numero di campioni per generalizzare i concetti, che, oltre ad essere più robusto ed affidabile in quanto fornisce sempre accoppiamenti geometricamente coerenti, l'MGCM è anche in grado, quando la tessitura dell'oggetto e le condizioni di ripresa sono ottimali, di fornire precisioni superiori.

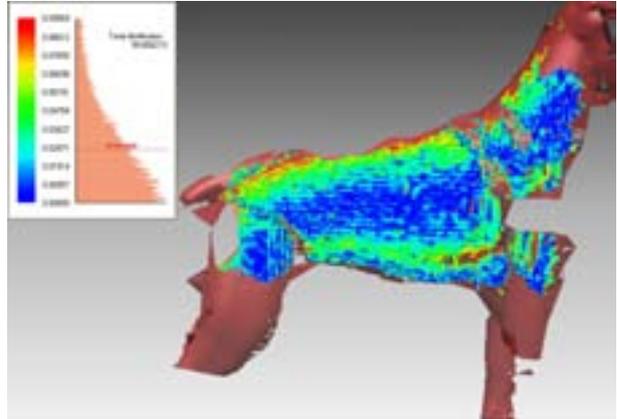


Figura 2.19: Confronto con la mesh originale del DSM ottenuto tramite LSM.

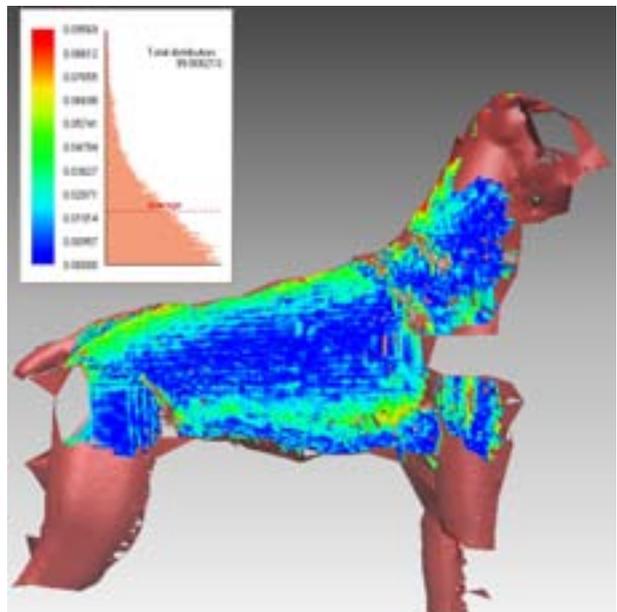


Figura 2.20: Confronto con la mesh originale del DSM ottenuto tramite MGCM.

Differenze Mesh-DSM (<i>mm</i>)		
	LSM	MGCM
Media	0.2762	0.2183
Dev std.	0.3361	0.1880
Err. Max	1.8629	0.9569

Tabella 2.2: Statistiche delle differenze misurate fra mesh originale e il DSM ottenuto con LSM e MGCM.

2.7.3. Generazione di nuvole di punti

Di seguito vengono presentate due applicazioni in cui è stata richiesta la determinazione della nuvola di punti che definisce la geometria dell'oggetto da analizzare: in questi casi si è fatto ricorso ai metodi di matching denso presentati in § 1.8.

Il primo esempio riguarda la restituzione della cupola dell'Ospedale Vecchio di Parma: l'edificio, di origine quattrocentesca, ha interessato la cronaca cittadina negli ultimi anni per la proposta, avanzata dall'amministrazione comunale, di cambiarne la destinazione d'uso (al momento esso ospita due biblioteche, l'archivio storico comunale e l'archivio di stato), per realizzare zone residenziali e destinate al commercio; inutile in questa sede, cercare di illustrare l'apocalittica diatriba suscitata da tale decisione.

Rimanendo nel più placido campo di studio delle caratteristiche della cupola, essa suscita notevole interesse in quanto presenta elementi costruttivi abbastanza insoliti e, a causa del deterioramento nel corso dei secoli, richiede al più presto degli interventi di risanamento e ristrutturazione.

Per studiarne la geometria si è pensato di studiare l'intero sviluppo intradossale, cercando di ottenere una descrizione non eccessivamente precisa (dell'ordine del centimetro) ma che permettesse di mettere in evidenza possibili fenomeni di distacco o di deformazione.

L'utilizzo di un laser a scansione, sicuramente più efficiente per uno studio di questo tipo, è stato scartato per la mancanza di risorse da impiegare nelle operazioni di rilievo.

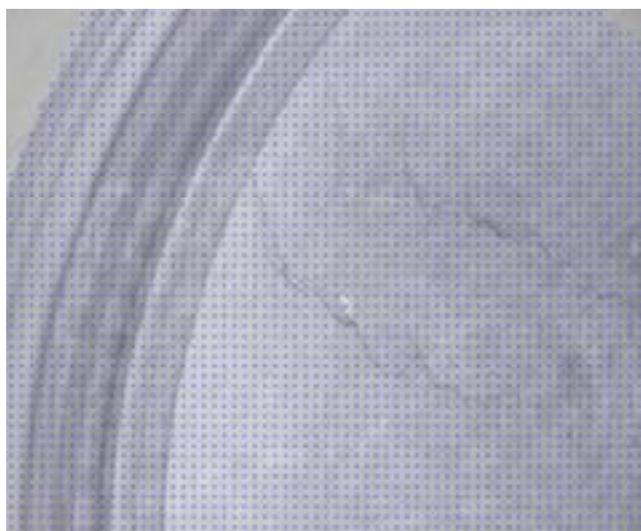


Figura 2.21: Particolare della cupola dell'Ospedale Vecchio di Parma con indicate le posizioni di misura; a parte le zone lesionate, la caratterizzazione di tessitura è decisamente povera. Mediamente sono stati misurati 850 punti al metro quadro.

La restituzione densa di punti della cupola è di notevole interesse se affrontata con metodi fotogrammetrici dal momento che, come si può vedere in figura 2.21, essendo l'intradosso intonacato, la tessitura superficiale risulta poco caratterizzata e può quindi mettere facilmente in mostra eventuali difficoltà dell'algoritmo a tracciare correttamente i punti.

E' stato realizzato un blocco fotogrammetrico con uno schema approssimativamente circolare ai piedi della cupola, costituito da 12 immagini ottenute per mezzo di una fotocamera Nikon D100 con ottica da 18 mm; per permettere analisi più dettagliate, da 4 postazioni differenti sono state scattate altre 32 immagini con un obiettivo da 105 mm. L'orientazione del blocco, avvenuto manualmente a causa dei

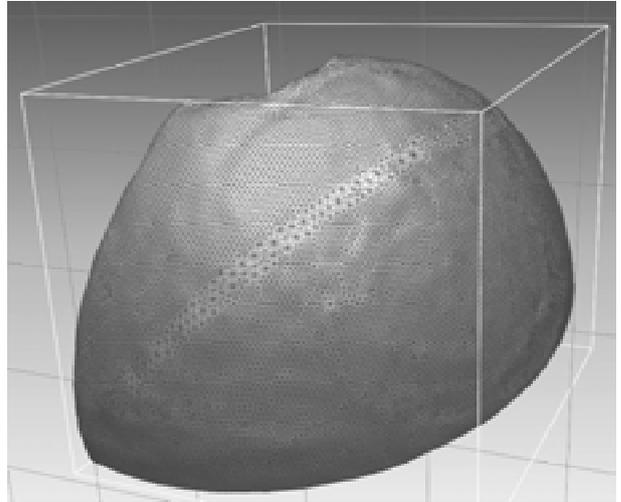


Figura 2.22: Meshing di una porzione della nuvola di punti ottenuta.

problemi di tessitura che non permettevano un buon esito delle tecniche di structure and motion, ha richiesto la collimazione di 112 punti terreno, visti almeno su tre fotogrammi. Un rilievo d'appoggio con stazione totale ha fornito 8 punti d'appoggio utilizzati per la georeferenziazione del blocco. Sui pochi punti di controllo rilevati ed individuabili con certezza sono stati ottenuti residui dell'ordine di $1 \div 1.5$ cm.

Per irrobustire il risultato si è generato un primo DSM per mezzo del metodo sequenziale illustrato in § 2.5.2.; come presumibile, il processo in molte zone non è riuscito a fornire risultati accettabili e sono state necessarie due successive procedure di infittimento dei punti per ottenere un sufficiente numero di informazioni. Da questa nuvola preliminare si è determinato un modello digitale della superficie, andando ad interpolare la geometria in quelle zone in cui erano ancora presenti delle lacune (da questo punto di vista, la regolarità geometrica dell'oggetto ha reso tale procedura abbastanza precisa). Si è quindi definita, su questa superficie di riferimento, una distribuzione omogenea di punti, successivamente proiettati su ciascun piano immagine: in questo modo si è potuto avviare un processo, questa volta parallelo, in cui si disponeva di parametri approssimati abbastanza precisi, che ha portato alla restituzione finale della cupola.

Come possibile vedere in figura 2.22, in cui per semplicità di rappresentazione si mostra solamente un settore della cupola, restano alcune zone, di estensione limitata, prive di punti per la completa assenza di elementi riconoscibili dal matching.

Per verificare la qualità dei risultati si è confrontata la geometria ottenuta dalla nuvola di punti, con una serie di check point determinati per mezzo di una stazione totale reflectorless. I risultati sono riassunti in tabella 2.3.

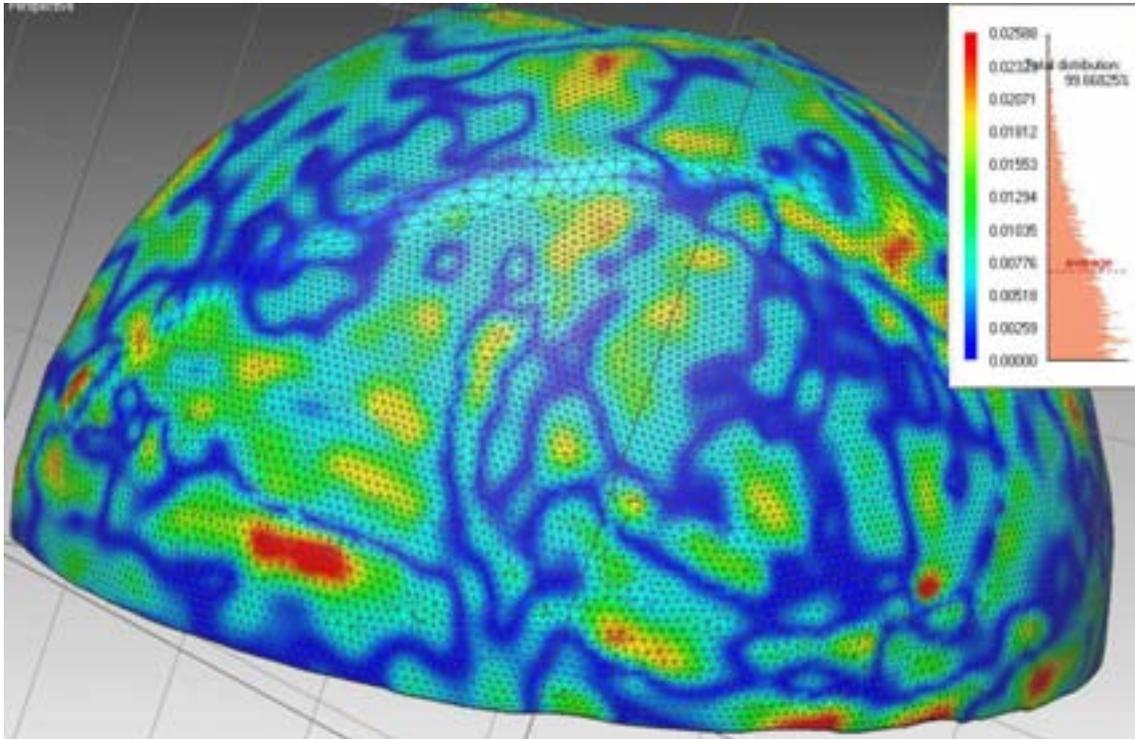


Figura 2.23: Distanze geometriche fra la superficie della cupola stimata tramite l'MGCM e la sfera interpolante i punti della nuvola. La distanza media risulta inferiore ad 1 cm.

Dall'analisi della figura 2.23, in cui sono rappresentate le distanze fra la nuvola di punti ottenuta fotogrammetricamente e una sfera che interpoli tali punti, si può facilmente notare come non solo la cupola dimostri di essere in buona salute (a parte qualche leggero distacco probabilmente di intonaco nelle zone rosse), ma anche di essere un ottimo esempio della perizia costruttiva delle maestranze rinascimentali: la cupola approssima una sfera geometricamente perfetta a meno di $1 \div 2$ cm.

Differenze check point (115) in cm	
Media	0.04
Dev. Std.	1.41
Err. Max	2.17

Tabella 2.3: Differenze misurate sul DSM da 115 check point ottenuti per mezzo di una stazione totale reflectorless.

L'ultimo esempio che intendiamo presentare, riguarda la restituzione della parete nord-ovest dell'ammasso roccioso che sovrasta l'eremo di Longeborgne (figura 2.24), di cui si è già parlato in § 1.8.2.: in questo caso i risultati di orientamento ottenuti dalla structure and motion sono stati utilizzati per generare le

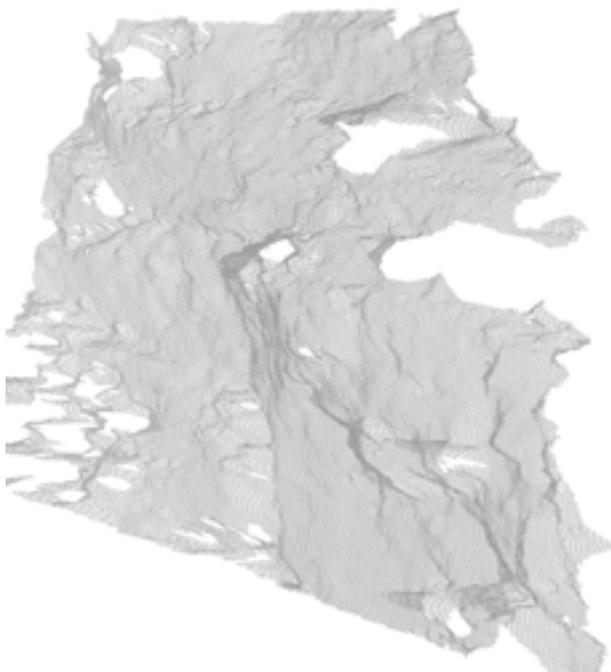


Figura 2.24: DSM della parete nord-ovest dell'ammasso di Longeborgne.

immagini epipolari rappresentate in figura 2.9; da tali immagini, per mezzo del metodo parallelo, utilizzando le informazioni dei numerosi punti di legame, si è, in maniera completamente automatica, ottenuta la nuvola di punti dell'intera parete, che risulta costituita da circa 100.000 punti.

Grazie alla stazione totale reflectorless con la quale si era realizzato il rilievo d'appoggio, si disponeva di una serie di 400 punti di controllo, collimati automaticamente dallo strumento in maniera casuale, che hanno permesso di ottenere una verifica delle precisioni di restituzione fotogrammetrica del DSM; i risultati sono riassunti in tabella 2.4.

Differenze check point (200) in cm	
Media	-0.95
Dev. Std.	2.78
RMSE	2.92
Err. Max	3.84

Tabella 2.4: Differenze misurate sul DSM da 200 check point ottenuti per mezzo di una stazione totale reflectorless.

CAPITOLO 3

Analisi di Point Clouds

L'analisi dei dati rilevati fotogrammetricamente con matching denso o per mezzo di laser a scansione, è complicata dall'enorme mole di dati; tale informazione è sovrabbondante quando si vogliono ricavare solamente alcune grandezze descrittive della geometria dell'oggetto (distanze, aree, volumi, forme, rapporti spaziali etc...); se si desidera una descrizione continua della superficie (ad esempio con una mesh poligonale) è richiesta un'operazione di segmentazione della nuvola; in altri casi ancora si vuole ottenere una descrizione semplificata della geometria dell'oggetto, riducendo la mole di dati; infine può essere richiesta l'integrazione dei dati puntuali con altre informazioni, ad esempio l'applicazione ad un modello digitale delle tessiture estratte da immagini.

La gestione di nuvole di punti di grandi dimensioni (decine di milioni) è oggi possibile grazie all'aumento delle prestazioni di calcolo e memorizzazione dei moderni calcolatori; ciononostante il modello originale deve essere semplificato per permettere una sua gestione snella nelle procedure successive (visualizzazione, analisi, modifica etc...). Tali operazioni richiedono un dispendio di tempo non trascurabile, in quanto al momento non si dispone ancora di tecniche automatiche sufficientemente robuste e, soprattutto nel caso di geometrie complesse, l'intervento umano è necessario per garantire risultati affidabili.

In altri termini, se con tecniche di laser scanning, o in maniera meno efficiente con la fotogrammetria, è possibile oggi ottenere nuvole di punti di notevoli dimensioni con una certa semplicità, sono i successivi passaggi che rendono complessa e a volte costosa la restituzione: la letteratura sull'argomento è molto ricca di spunti interessanti e la ricerca si sta sempre più orientando verso questo settore, a dimostrazione dell'esigenza di tecniche innovative.

Ricordiamo brevemente in questa sede solo alcuni settori applicativi: nel laser aereoportato da oltre un decennio si studiano metodi per l'estrazione e la catalogazione di oggetti, con particolare riferimento alla generazione di DTM e di modelli urbani digitali (3d city modelling): solo nel nostro paese sono diversi gli autori che si sono occupati dell'argomento ([Nardinocchi 2003], [Roggero 2002], [Crosilla 2005], [Brovelli 2004] etc...) seppur non si possano tralasciare i lavori di [Vosselman 2001], [Rottensteiner 2002], [Hyypä 2005] etc...; in ambito close-range, seppur il mercato (e quindi i relativi investimenti) siano lieve-

mente inferiori rispetto al caso precedente, sono numerosissime le applicazioni che richiedono lo sviluppo di tecniche di automazione: si spazia da strumenti per il change detection, ovvero di tecniche per lo studio dei cambiamenti in una scena tridimensionale ([Girardeau 2005], [Biasion 2004], etc.), all'estrazione di primitive geometriche [Rabbani 2005], al monitoraggio di strutture [Lindenbergh 2005], alla modellazione architettonica [El-Hakim 2005], [Abmayr 2005].

Nel presente capitolo verranno illustrate le tecniche sviluppate per l'analisi di nuvole di punti e per la loro segmentazione in zone piane, messe a punto nell'ambito di un progetto per l'analisi delle caratteristiche di discontinuità geo-strutturale di pareti rocciose, di cui ci si occuperà nel prossimo capitolo. L'obiettivo è da un lato sviluppare tecniche per ottenere informazioni geometriche d'insieme dalla nuvola di punti, dall'altro individuare le discontinuità della roccia (caratterizzate da superfici piane) per sviluppare analisi di stabilità. Sebbene messe a punto nel quadro dell'applicazione specifica, queste tecniche sono estendibili, con le dovute modifiche, ad altri ambiti applicativi. Inoltre, essendo il dato primario la nuvola di punti, è possibile applicarle sia in ambito fotogrammetrico sia in applicazioni laser scanning.

3.1. Analisi interattiva di nuvole di punti

Ancora oggi non sono molti i pacchetti software che permettano di analizzare le caratteristiche delle nuvole di punti con semplicità ed immediatezza: da un lato soluzioni a basso costo o gratuite sono di difficile reperimento, dall'altro le applicazioni professionali hanno solitamente costi molto elevati e forniscono funzionalità che l'utente finale non ha la necessità, a volte, la preparazione per utilizzare appieno. L'estrazione delle informazioni dalla nuvola di punti viene molto spesso gestita da chi ha realizzato il rilievo, mentre il committente non interviene sul dato grezzo originario.

Per superare tale limitazione e fondere sorgenti di dati complementari, diverse soluzioni sono state proposte negli ultimi anni: di queste quella probabilmente più conosciuta in Italia è la cosiddetta immagine solida [Bornaz 2003], che permette di integrare le informazioni radiometriche nella matrice di punti che costituisce un'immagine digitale, con informazioni relative alle coordinate tridimensionali dei punti in essa rappresentati.

Nel nostro caso si è sviluppato in maniera indipendente un sistema sostanzialmente analogo. Noti i parametri di orientamento esterno di un fotogramma utilizzato per la restituzione della nuvola, o di un'altra immagine orientata nello stesso sistema di riferimento, per mezzo del proiettore algebrico (1.40) si proiettano i punti della nuvola sul fotogramma, inserendoli in una struttura dati ottimizzata (*quad-tree*, si veda paragrafo 2.4.1). L'utente, individuando interattivamente il punto sull'immagine, è in grado di ricavare informazioni relative agli elementi che lo interessano: a seconda del contesto viene infatti eseguita dal programma una *point query*, per estrarre dalla nuvola il punto più vicino a quello indicato sul fotogramma, o una *window query* per estrarre un insieme di punti proiettati in una certa regione sul fotogramma. Nel

primo caso si individua la foglia dell'albero quaternario corrispondente al punto selezionato: in tale struttura il numero di punti è solitamente modesto (da una decina a un centinaio di punti) ed è quindi immediato individuare il punto più vicino a quello scelto; nel secondo si individuano tutte le foglie (che corrispondono ad aree rettangolari del fotogramma) che intersecano l'area individuata dall'utente: anche in questo caso, l'operazione di filtraggio dei punti che sono contenuti all'interno della regione selezionata è estremamente rapida in quanto in tutte le foglie completamente contenute nella regione di interesse non bisogna eseguire ricerche.

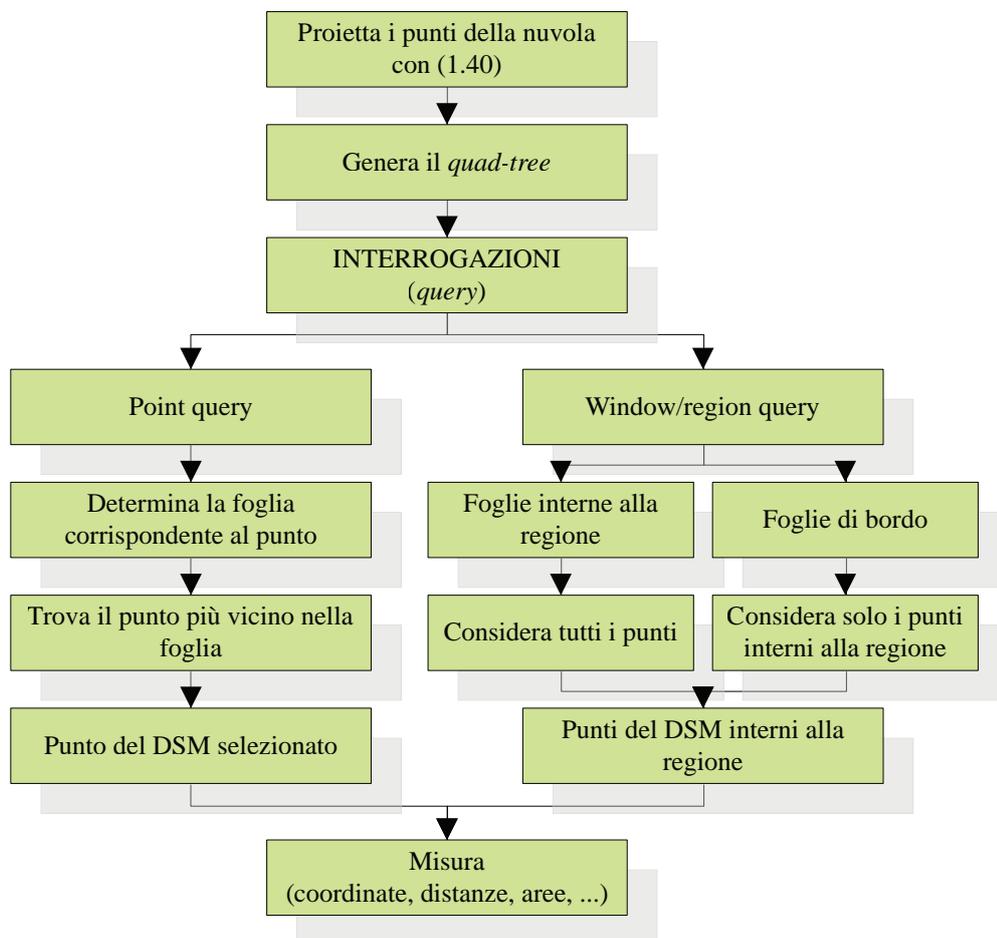


Figura 3.1: Rappresentazione schematica delle operazioni utilizzate nell'analisi di point cloud interattiva.

In tal modo si ottiene una struttura dati bi-direzionale che lega i punti della nuvola ai punti del fotogramma: qualsiasi tipo di informazione richiesta dall'utente (coordinate di punti, distanze, elenchi di punti, etc.) può essere efficacemente fornita. In figura 3.1 viene mostrato schematicamente il funzionamento del motore di ricerca dei punti, mentre in figura 3.2 viene mostrata l'individuazione di una regione di interesse



Figura 3.2: Selezione interattiva di porzioni di DSM.

del DSM su un fotogramma orientato.

Confrontando l'immagine solida con il metodo proposto si possono fare alcune considerazioni: innanzitutto, concettualmente i due metodi si equivalgono, in quanto si ottiene una rappresentazione della superficie del DSM sullo spazio immagine bidimensionale; nel primo caso la struttura dati è più compatta in quanto a ciascun pixel corrisponde un punto (interpolato) della nuvola di punti, il che permette una più semplice implementazione delle operazioni di interrogazione dei dati, a fronte di una procedura iniziale di generazione dell'immagine; nel secondo caso, al contrario, si ha il vantaggio di lavorare direttamente sul dato originale in quanto, salvo non venga richiesto dall'operatore, si lavora direttamente sulla nuvola di punti: tale aspetto risulta abbastanza importante per l'applicazione del metodo al caso dell'analisi delle superfici di discontinuità di pareti rocciose di cui si parlerà nel prossimo capitolo; al tempo stesso i tempi di elaborazione richiesti inizialmente sono minimi, in quanto la proiezione dei punti della nuvola sui fotogrammi è quasi istantanea (solo nel caso in cui i punti siano dell'ordine dei milioni sono richiesti un paio di secondi per svolgere l'operazione); nel metodo proposto, tuttavia, si ha una maggior complessità nella gestione dei dati, in quanto si devono contemporaneamente gestire i punti della nuvola e i punti sul fotogramma.

Il metodo è ancora in fase di sviluppo: al momento è possibile estrarre informazioni puntuali (coordinate), determinare distanze fra punti, informazioni areali (elenchi di punti), estensioni di superficie (al momento ottenute proiettando su un piano medio i punti interessati e determinandone il contorno chiuso con una *convex hull*), determinare il/i piani che approssimano un'insieme di punti per mezzo della procedura

Ransac (si veda il prossimo paragrafo), estrarre sezioni del DSM e ottenere, per interpolazione, una nuova nuvola di punti regolarizzata sulla maglia dei punti immagine (pixel).

3.2. Segmentazione automatica di piani

Sebbene la modalità interattiva permetta di avere un buon controllo della geometria da analizzare e consenta di svolgere diverse operazioni di misura sulla nuvola di punti, in molte applicazioni può essere richiesto, per la mole di dati da analizzare e per la semplicità e ripetitività delle operazioni da compiere, l'utilizzo di tecniche che presentino un grado di automazione maggiore. Una delle più frequenti operazioni da svolgere su una nuvola di punti è la sua scomposizione (o segmentazione) in una serie di superfici piane: ad esempio, l'analisi delle superfici di discontinuità nello studio della stabilità di pareti rocciose, di cui si tratterà nel prossimo capitolo, richiede di suddividere il DSM della parete in regioni approssimativamente piane, le quali possono fornire importanti informazioni sulle caratteristiche statiche e fessurative di un ammasso roccioso. Un'esigenza analoga è sentita in molti altri ambiti applicativi: si pensi per semplicità al caso del rilievo architettonico o di infrastrutture industriali nel reverse engineering.

L'esigenza di procedure automatiche è sempre più avvertita a fronte delle accresciute prestazioni delle nuove strumentazioni: poiché l'acquisizione si effettua in tempi ormai brevissimi, per trarre un reale beneficio da tali tecniche è di primaria importanza che anche i successivi passaggi della procedura di restituzione siano rapidi e non richiedano troppe risorse.

Di seguito vengono presentati due approcci differenti per ottenere la segmentazione automatica di una nuvola di punti generica: il primo è sostanzialmente basato su algoritmi di stima robusti (Ransac) ed è stato appositamente sviluppato per l'analisi di discontinuità; in tal senso si è posta particolare attenzione a che gli algoritmi non solo garantissero una scomposizione corretta della geometria, ma fornissero anche lo stesso tipo di risultato che otterrebbe un operatore umano impegnato nella stessa procedura. Determinare le superfici di discontinuità di un ammasso roccioso non significa, infatti, estrarre tutte le superfici piane possibili, ma effettuare tale estrazione rispettando una serie di condizioni, sostanzialmente legate agli aspetti statici indagati, che permettono di ottenere, al termine del processo, una corretta interpretazione strutturale dell'oggetto; in molti casi, ad esempio, se superfici adiacenti presentano approssimativamente la stessa giacitura, l'algoritmo deve essere in grado di interpretarle come un'unica superficie; viceversa, in certi casi, una superficie piana deve essere scomposta in sotto-regioni in funzione della sua forma, per garantire una corretta schematizzazione della geometria. Seppur applicabili ad altri ambiti d'indagine, dunque, tali algoritmi presentano alcune particolarità specifiche.

Si accenna infine brevemente ad un metodo più generale, che impiega la trasformata di Hough, sottolineando tuttavia che il quadro dei metodi di segmentazione di range data è ben più ampio di quello che verrà tratteggiato.

3.3. Segmentazione con metodi di stima robusti

Come già anticipato in §3.1. è possibile, per mezzo di algoritmi di stima robusti, individuare in un insieme di punti (nuvola di punti) le porzioni corrispondenti a zone piane della superficie su cui i punti sono stati rilevati. Si noti che anche nell'utilizzo interattivo può essere importante, estraendo un sottoinsieme di punti dal DSM complessivo, verificare se esso contenga punti tutti appartenenti ad un'unica superficie piana o a più superfici distinte. Un insieme di punti (al limite l'intera nuvola), o meglio le loro coordinate, può essere pensato come un campione di osservazioni in cui sono presenti, in percentuale incognita, una certa quantità di errori grossolani: in altre parole, volendo stimare un piano da un insieme di elementi puntuali, ci si aspetta che solo alcuni dei punti possano essere utilizzati per la procedura, mentre altri, che non appartengono alla superficie, debbano essere esclusi.

Come accennato in diverse occasioni nel capitolo 1 un metodo estremamente efficiente per affrontare questo tipo di problemi è rappresentato dalla procedura Ransac, di cui si trattano gli aspetti teorici in appendice A (§ A.2.4.). Il modello funzionale da considerare è rappresentato dall'equazione in coordinate cartesiane di un piano:

$$\pi: ax + by + cz + d = 0 \quad (3.1)$$

Per la determinazione di tale modello occorrono tre punti: i parametri da determinare, in relazione alla (3.1), sono in apparenza quattro; essendo la quaterna (a, b, c, d) definita a meno di una costante moltiplicativa, il numero di incognite indipendenti è effettivamente pari a 3. Per risolvere il problema, molto spesso, si fissa uno dei parametri e si stimano i rimanenti; tuttavia, in tal modo, si esclude la possibilità di determinare alcune famiglie di piani: fissando ad esempio il parametro c , non è possibile determinare tutti i piani verticali, o fissando il parametro d si escludono tutte le superfici passanti per l'origine del sistema di riferimento. Per superare la limitazione si può considerare il problema

$$\|A\mathbf{x}\| = 0 \quad \text{con} \quad \|\mathbf{x}\| = 1 \quad (3.2)$$

in cui il vettore \mathbf{x} rappresenta la quaterna di parametri che individuano il piano, e la matrice disegno A , dati tre punti $\mathbf{p}_1 = (x_1, y_1, z_1)$, $\mathbf{p}_2 = (x_2, y_2, z_2)$, $\mathbf{p}_3 = (x_3, y_3, z_3)$, è semplicemente

$$\begin{vmatrix} x_1 & y_1 & z_1 & 1 \\ x_2 & y_2 & z_2 & 1 \\ x_3 & y_3 & z_3 & 1 \end{vmatrix} \quad (3.3)$$

Il problema (3.2) è risolvibile, come già visto, per mezzo della decomposizione ai valori singolari (§ A.2.1.). E' immediato verificare che in tal modo non si esclude nessun piano dello spazio.

Come funzione di controllo per definire il consenso del set di dati si può banalmente utilizzare la distan-

za di ciascun punto dal piano stimato:

$$d = \frac{|ax + by + cz + d|}{\sqrt{a^2 + b^2 + c^2}} \quad (3.4)$$

Tutte le operazioni finora viste richiedono pochissimi calcoli, e il fatto di utilizzare un campione minimo costituito da sole tre osservazioni richiede (si veda tabella A.1-A.2) poche iterazioni: il metodo è estremamente rapido ed efficiente.

Le buone prestazioni all'interno della procedura interattiva per filtrare le zone individuate dall'operatore, ci hanno suggerito di estendere l'algoritmo all'estrazione di tutti i piani presenti nella nuvola senza alcun intervento umano. In questo caso il procedimento deve essere modificato per permettere l'estrazione di più modelli funzionali, corrispondenti ai piani di discontinuità presenti (dei quali ovviamente non è noto a priori il numero): viene innanzitutto individuato il campione di massimo consenso; eventualmente si ristimano i parametri del piano con approcci a minimi quadrati, come illustrato in § 1.5.3. per il caso della stima della matrice fondamentale; si individuano quindi tutti gli elementi della nuvola che si dimostrano in accordo con tale modello e si isolano dal resto del campione; a questo punto è possibile applicare nuovamente la procedura alla ricerca di nuove primitive. Idealmente quindi, la nuvola viene segmentata gerarchicamente, ovvero i piani vengono estratti in ordine decrescente di cardinalità dei punti che li definiscono.

A causa della complessità geometrica dell'oggetto e del numero elevato di piani differenti e di punti che compongono il modello, Ransac non è però in grado di lavorare direttamente sulla totalità dei dati ed è necessario sviluppare una strategia che semplifichi il problema. Sempre in riferimento alle due tabelle A.1 e A.2, è facile accorgersi che, se la percentuale di outlier risulta troppo elevata, il numero di iterazioni richiesto per garantire con buona probabilità l'affidabilità dei risultati tende rapidamente a valori computazionalmente non affrontabili. Un metodo, in realtà abbastanza cautelativo, per stimare la percentuale di errori grossolani all'interno della nuvola deriva dal seguente ragionamento: dal momento che, con le successive estrazioni di entità planari, non solo il numero di punti ma anche il numero di superfici, tende ad assottigliarsi, la percentuale di outlier è da considerarsi maggiore all'inizio della procedura. Assumiamo per semplicità che n piani compongano il DSM e che tutti presentino la stessa numerosità: la percentuale di inlier verosimilmente presente è allora pari a $1/n$. Se il numero complessivo di piani è inferiore alla decina, si può applicare direttamente il metodo programmato nel caso interattivo; con l'aumentare del numero di superfici, tuttavia, il problema diventa ben presto computazionalmente impraticabile.

Per superare il problema, si è sviluppata una strategia multi-risoluzione brevemente schematizzata in figura 3.3: inizialmente viene generata una struttura a multi risoluzione in cui la geometria dell'oggetto in termini di numero di punti che la rappresentano viene via via semplificata. Partendo dal livello a risoluzione minore (generalmente costituito da 3.000-4.000 punti) vengono estratte macrozone che stabiliscono le strutture piane principali individuabili sull'oggetto. Per limitare il numero di superfici da considerare (e quindi la percentuale di errori grossolani) vengono utilizzate soglie di accettazione elevate.

In molti casi l'algoritmo tende, nonostante tale soglia, ad individuare un numero di piani eccessivamen-

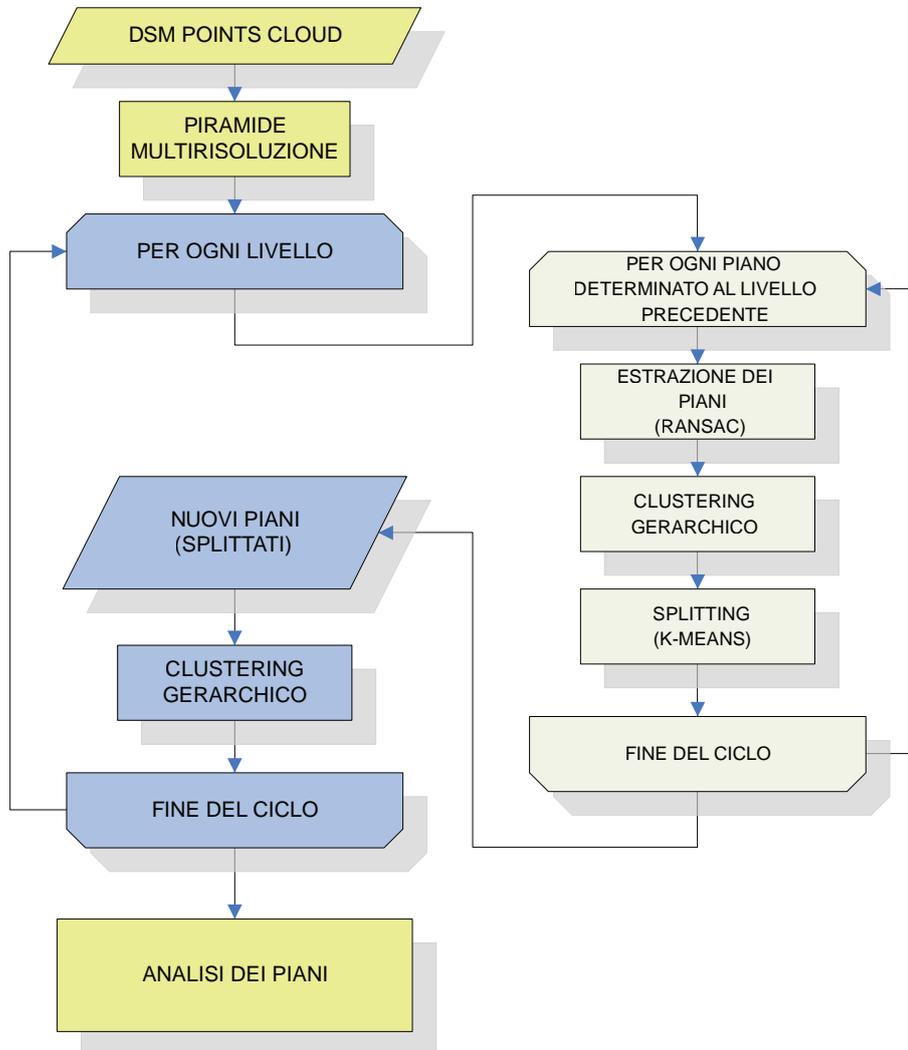


Figura 3.3: Principali operazioni per la segmentazione automatica di piani con metodi di stima robusti.

te elevato: molti piani adiacenti presentano giaciture assai simili e vengono distinti a causa di un non sufficiente condizionamento numerico e geometrico. Tali piani vengono analizzati in un successivo passaggio e per mezzo di un modulo di clustering gerarchico, vengono eventualmente aggregati per costituire un piano di dimensioni maggiori.

Tale aggregazione porta ad una ridefinizione della geometria delle zone piane che in taluni casi può non essere strutturalmente corretta (i piani assumono una forma che, seppur geometricamente coerente, non interpreta correttamente la superficie di discontinuità). Pertanto un'ulteriore fase di clustering, questa volta con un approccio basato su K -means, cerca dal piano riaggregato di ottenere, qualora ve ne fosse bisogno, una divisione in porzioni strutturalmente più compatta e coerente.

Al termine di questo ciclo di operazioni l'algoritmo prende in considerazione, una alla volta, ciascuna delle superfici estratte aumentando la risoluzione dei punti che la compongono. In tal modo si analizza solamente una porzione del DSM, sufficientemente distinta dalle adiacenti in quel livello della piramide ed in cui ragionevolmente il numero di piani presente non può essere eccessivamente elevato. Ancora una volta è necessario, dopo aver diviso tale superficie in porzioni con caratteristiche di giacitura simili (di volta in volta vengono diminuite le soglie di accettazione), procedere ad una fase di riaggregazione dei punti e ad una eventuale loro ri-suddivisione.

Al termine del ciclo, avendo preso in considerazione solamente i punti che giacevano su una macrozona dell'oggetto, bisogna verificare che ai bordi di tale macrozona non sia possibile aggregare punti o piani appartenenti a macrozone differenti: si va cioè a verificare che, nel livello di dettaglio precedente, si fosse delimitata la macrozona in maniera corretta. Un ultimo passaggio in cui viene determinato un ulteriore clustering di giaciture adiacenti provvede a risolvere tali incongruenze.

Il processo descritto viene iterato fino all'analisi del livello a massima risoluzione in cui vengono utilizzati tutti i punti che costituivano inizialmente il DSM. Nel seguito i passi dell'algoritmo verranno maggiormente descritti e analizzati.

3.3.1. Generazione del DSM multi-risoluzione

Per la generazione della piramide del DSM viene utilizzato un approccio basato su voxels (si veda figura 3.4): lo spazio tridimensionale viene suddiviso in porzioni cubiche di dimensioni prestabilite (generalmente viene scelta una dimensione tale che, nota la spaziatura media fra i punti, mediamente 4-8 punti siano contenuti in ciascun cubetto); per ciascun voxel si determina se esistono punti contenuti, nel qual caso ad essi viene sostituito il baricentro. Tale semplificazione non riduce sufficientemente il numero di punti del DSM: nei livelli successivi vengono aggregati 8 voxel adiacenti per generare un cubo di dimensioni maggiori; anche in questo caso ai baricentri di ciascun cubetto viene sostituito il baricentro dell'aggregazione. Il calcolo del centro di massa viene effettuato con una media pesata dal numero di punti associato a ciascun

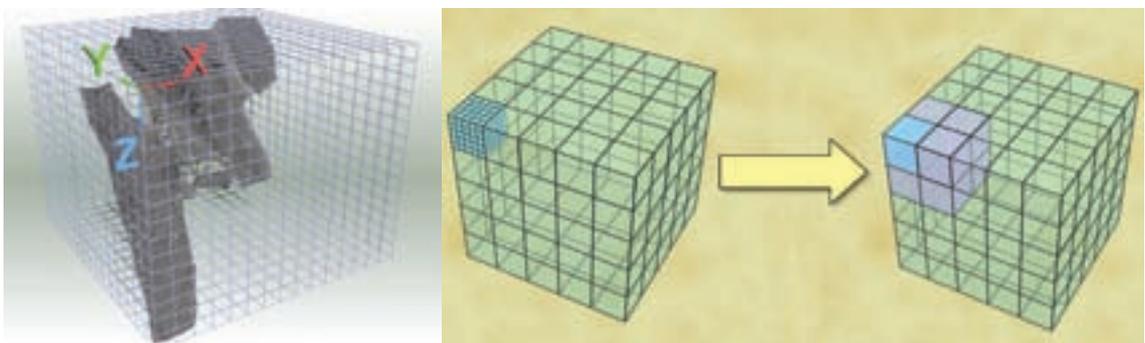


Figura 3.4: Voxelling e successiva aggregazione dei punti che compongono la nuvola.

baricentro: in tal modo la semplificazione della geometria tiene sempre in conto l'effettiva distribuzione dei punti al livello di risoluzione massimo. Come possibile notare in figura 3.5, tipicamente le variazioni di geometria da un livello al successivo sono piuttosto modeste.

Il processo di semplificazione, iterato per il numero di livelli desiderato, oltre a decimare la numerosità dei punti che costituiscono il DSM porta con sé un ulteriore vantaggio: ai livelli di risoluzione minore la geometria in esame subisce un sostanziale appiattimento, ovvero vengono diminuiti caratteri locali di convessità o concavità. Ciò si dimostra particolarmente utile in quanto, nelle prime fasi di segmentazione, ha maggior importanza la suddivisione in zone approssimativamente omogenee in termini di giacitura, e fenomeni locali troppo accentuati potrebbero instabilizzare la scelta di tali giaciture. Risalendo, nelle iterazioni successive, i livelli di risoluzione del DSM, le caratteristiche locali vengono recuperate e la segmentazione finale del DSM risulta pertanto corretta.

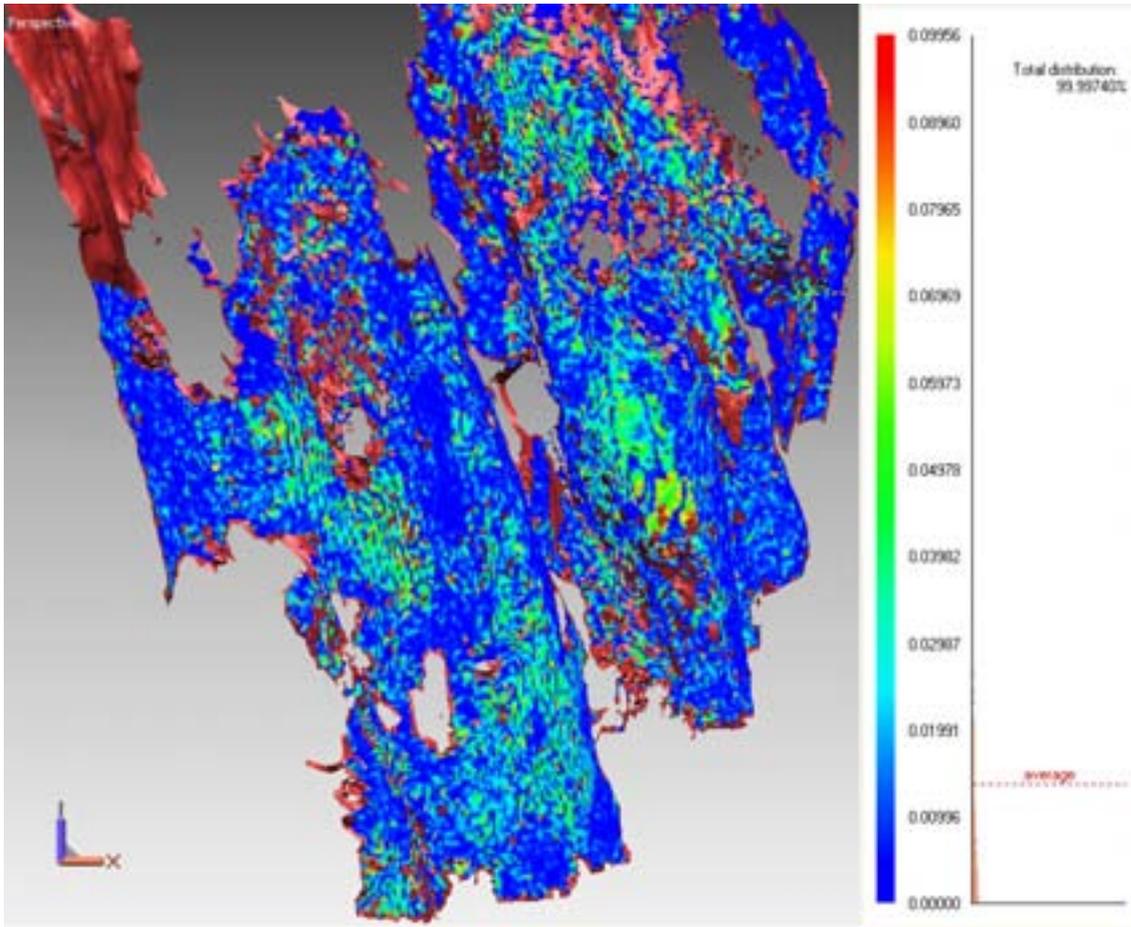


Figura 3.5. Distanze assolute fra i primi due livelli della piramide del DSM: la distanza media fra i due si assesta su valori di 1.5-2 cm con punte massime di circa 5 cm nei punti di maggior convessità.

Dall'analisi delle differenze geometriche, evidenziabili dal confronto fra livelli successivi del DSM, appare evidente come un tale approccio non porti ad un sconvolgimento effettivo della geometria: ed esempio, nel caso del sito di Arnad (vedi § 4.2), mediamente tali differenze sono contenute fra 2.5 cm (fra i livelli a massima risoluzione) e 8 cm (nei livelli più alti della piramide), con una deviazione standard massima fra il primo e ultimo livello che non supera i 12-15 cm.

3.3.2. Estrazione delle entità planari

L'estrazione dei piani (dal DSM complessivo nel livello a minor risoluzione o dalle singole macrozone negli altri livelli) è basato su RANSAC, scegliendo in maniera casuale tre punti purché non allineati e purché il triangolo costituito dai tre punti non sia troppo allungato (ovvero se nessuno dei suoi tre angoli risulta minore di $15 \div 20$ gradi). Per rendere più efficiente la scelta, le estrazioni casuali successive al primo punto, avvengono ad una distanza limitata da questi (solitamente entro $1/3 \div 1/5$ le dimensioni della nuvola considerata ovvero della macro-zone). In questa fase, come del resto in altre del processo in cui la qualità delle osservazioni dipende da diversi parametri (non allineamento dei punti e loro distanza reciproca, etc.), sono attualmente in fase di studio tecniche basate sulla logica *fuzzy*, di cui si parla brevemente in § A.4.

Come già precedentemente osservato, l'approccio multi-risoluzione consente una notevole ottimizzazione del carico computazionale: se da un lato ridurre il numero di punti significa dover calcolare meno distanze, la scelta di una soglia di accettazione alta nelle prime fasi, unitamente ad una geometria meno complessa, fa sì che la percentuale di outlier (e quindi il numero di iterazioni) sia contenuta quando viene considerata l'intera estensione del DSM.

Per ottenere che Ransac scelga terne di punti sulla superficie della parete e che gli inlier siano effettivamente parte della medesima superficie di discontinuità, occorre imporre vincoli aggiuntivi: considerando come unico criterio la distanza euclidea dal piano stimato punti anche molto distanti fra loro o formanti zone non tangenti il piano possono essere considerati inliers pur non essendolo nella realtà. Per tale motivo sono stati implementati tre criteri di ordine topologico che i punti devono soddisfare per essere classificati inliers:

- 1. Connessione dei punti:** i punti devono formare una regione compatta dello spazio, ovvero essere connessi: con una sorta di region growing, si verifica che ogni punto sia adiacente ad almeno un altro punto del piano, partendo da un seed point scelto in maniera casuale;
- 2. Forma delle regioni estratte:** la regione determinata dai punti scelti deve essere sufficientemente convessa e compatta nel senso che non deve presentare buchi o diramazioni troppo grossi: la condizione equivale a considerare la densità di punti intesa come rapporto fra numero di elementi e area del contorno convesso dell'insieme; se tale densità è inferiore ad una certa soglia il piano viene scartato.
- 3. Rapporti dimensionali della regione estratta:** la stessa regione deve soddisfare dei criteri di forma, non essendo ad esempio troppo sviluppata solo in una direzione: si calcolano i due momenti principali

di inerzia della distribuzione dei punti proiettati sul loro piano interpolante; se il rapporto fra il momento massimo e quello minimo è superiore ad una certa soglia il piano viene scartato.

Anche in questo caso sono attualmente in fase di studio predicati fuzzy che permettano di considerare la qualità di una regione estratta in funzione, contemporaneamente, di tutte le condizioni indicate.

Al termine di RANSAC i punti considerati inliers vengono utilizzati per una stima a minimi quadrati dei parametri del piano interpolante.

3.3.3. Clustering gerarchico delle giaciture

Come già detto, al termine della segmentazione dei piani, capita spesso di ottenere un numero eccessivo di piani, che pur essendo adiacenti e contraddistinti da giaciture quasi identiche, vengono però distinti dalla procedura. In parte ciò è dovuto all'approccio seguito da Ransac: quelli che possono essere individuati come due piani distinti stimando i parametri geometrici per mezzo di tre sole osservazioni, al termine della stima a minimi quadrati appaiono in effetti sostanzialmente coincidenti.

Ciò può causare dei problemi durante l'analisi geostatistica delle famiglie di discontinuità dal momento che, se un piano di notevoli dimensioni viene scomposto in un numero eccessivo di parti, queste possono mettere in ombra altre giaciture in cui un livello di segmentazione così fitto non ha avuto luogo. Per rimediare, al termine della scomposizione della geometria con Ransac, ciascun piano estratto viene confrontato con i vicini per verificare che alcuni di essi non appartengano ad una stessa giacitura.

Per clustering gerarchico [Anderberg 1973] si intende una procedura grazie alla quale una serie di dati viene suddivisa in campioni mutuamente escludentisi tali per cui dati in uno stesso cluster hanno caratteristiche simili, mentre dati in cluster differenti presentano forti disomogeneità; il termine gerarchico sta ad indicare che tale suddivisione in cluster avviene per passaggi successivi, aggregando di volta in volta un numero maggiore di dati. Vengono innanzitutto stabiliti degli accoppiamenti in base ad una certa metrica: nel nostro caso i parametri di orientamento del piano vengono usati come base di uno spazio parametrico bidimensionale in cui minore è la distanza fra punti distinti, maggiore è il parallelismo fra le due giaciture. Il fatto di considerare piani adiacenti esclude la possibilità che due o più dei piani analizzati siano approssimativamente paralleli ma localizzati in zone differenti dello spazio: in altre parole, per la metodologia seguita è sufficiente considerare unicamente le informazioni desumibili dalle normali ai piani (cioè dal loro orientamento).

Viene poi definita una funzione di collegamento che determina la maggiore o minore vicinanza fra coppie di punti:

$$L(r, s) = \min (dist (p_{ri}, p_{sj})) \quad (3.5)$$

in cui p_{ri} e p_{sj} sono rispettivamente l' i -esimo elemento nel cluster r e il j -esimo elemento nel cluster s . A tal riguardo possono essere definiti diversi criteri e diverse metriche per generare la funzione di collega-

mento (ovvero la funzione $dist$), ma nel nostro caso è proprio un criterio di minima distanza euclidea che fornisce le prestazioni migliori, come diversi test hanno messo in evidenza.

Vengono dunque presi in considerazione tutti gli accoppiamenti, fra punti del piano prima e fra cluster diversi in seguito, seguendo uno schema ad albero fintanto che non si ottiene un unico cluster che inglobi tutto il data set. Una volta scelto il livello di vicinanza desiderato, tale struttura ad albero (rappresentata schematicamente in figura 3.6) viene “tagliata” ad un preciso livello e la suddivisione in cluster risulta immediatamente definita. Il livello di clustering (ovvero il livello di distanza al quale si vuole considerare il clustering) può essere definito a priori dall’utente o può essere stabilito da considerazioni statistiche sui dati in esame; definito infatti il coefficiente di inconsistenza del collegamento k come:

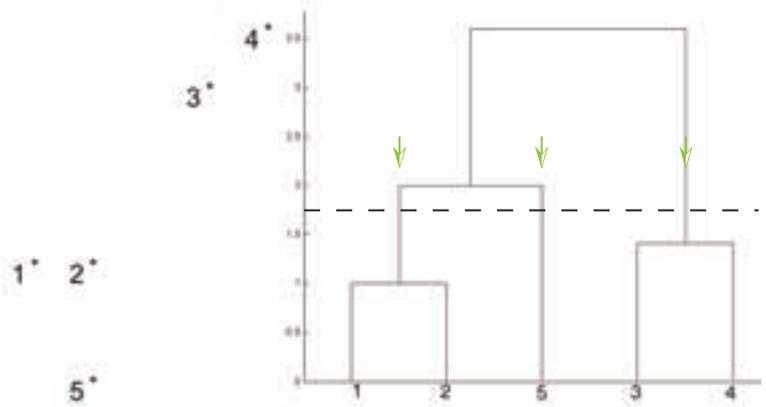


Figura 3.6: Un set costituito da 5 oggetti e il corrispondente albero gerarchico; la linea tratteggiata indica il “taglio” dell’albero che porta, in questo caso, ad ottenere 3 cluster.

$$Y(r, s) = \frac{(L(r, s) - \bar{M})}{S} \quad (3.6)$$

in cui $L(r, s)$ è la funzione di collegamento di cui alla (3.5), e \bar{M} ed S rappresentano rispettivamente la media e la deviazione standard di tutti i collegamenti al medesimo livello. Tale coefficiente è in grado di caratterizzare ciascun collegamento fra dati o cluster, confrontandolo con i collegamenti vicini, per capire se i dati inseriti nel cluster sono consistenti fra loro o meno. Se il coefficiente di inconsistenza da un lato fornisce una valutazione della soluzione di clustering a livello locale, è possibile, per mezzo del coefficiente cofenetico, ottenere una valutazione della soluzione a livello globale; tale coefficiente ha il significato di un coefficiente di correlazione in cui vengono presi in esame tutti i diversi cluster:

$$c = \frac{\sum_{i < j} (D_{ij} - \bar{D}) \cdot (L_{ij} - \bar{L})}{\sqrt{\sum_{i < j} (D_{ij} - \bar{D})^2 \cdot (L_{ij} - \bar{L})^2}} \quad (3.7)$$

dove D_{ij} è la distanza fra gli oggetti originali i e j , L_{ij} è il valore ottenuto dalla funzione di collegamento fra coppie di cluster contenenti gli oggetti i e j al momento del loro collegamento e \bar{L} e \bar{D} sono i valori

medi di L_{ij} e D_{ij} ; tale coefficiente assume il valore 1 per strutture fortemente clusterizzate e -1 per strutture disomogenee e casuali. Utilizzando le informazioni di carattere locale e globale fornite dalla (3.6) e dalla (3.7) è possibile decidere adattivamente quale sia il livello gerarchico che fornisce la decomposizione in cluster ottimale. Va però messo in evidenza che, essendo l'approccio basato sui rapporti che intercorrono fra i diversi punti (ovvero le distanze relative sul piano degli orientamenti) e non sui loro valori effettivi, se non viene imposta una soglia sulla massima differenza di orientamento fra coppie di piani, l'algoritmo può riconoscere strutture fortemente clusterizzate anche se contraddistinte da giaciture molto differenti fra loro. In questi termini viene quindi definita una soglia massima di distanza per la costituzione dei cluster.

3.3.4. Suddivisione della geometria

A volte la fase di aggregazione, dal momento che in essa vengono presi in esame solo criteri di adiacenza e giacitura, porta a costituire superfici che dalle forme insolite, che possono portare ad un'incorretta interpretazione della geostruttura dell'ammasso: si veda a riguardo la figura 3.7 in cui due piani paralleli palesemente divisi da un avvallamento vengono raggruppati in quanto verificano tutte le condizioni utilizzate dal clustering gerarchico. Come già anticipato tale fenomeno, seppur geometricamente corretto e concettualmente coerente con tutte le condizioni imposte, porta ad una non corretta interpretazione in fase di analisi strutturale dell'oggetto.

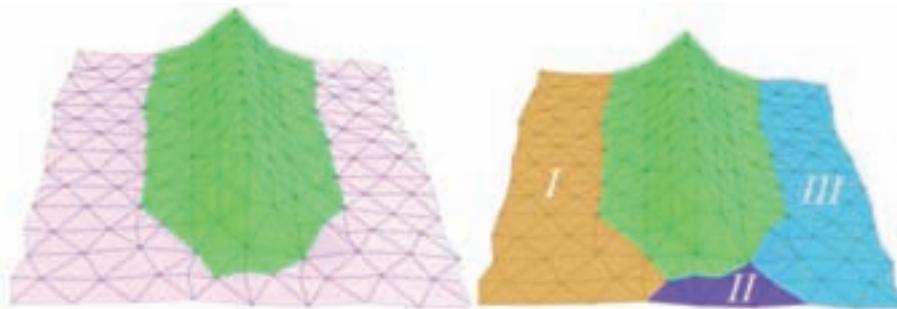


Figura 3.7: Aggregazione non corretta dal punto di vista strutturale (a sinistra) e conseguente suddivisione tramite k -means (a destra).

In questi casi è necessario tentare di ri-suddividere la geometria in un numero sufficienti di parti tali che le distinte regioni abbiano forma approssimativamente compatta. Per fare ciò si utilizza un differente algoritmo di clustering che prende il nome di k -means [Hartigan 1975]: in tale algoritmo tutte le osservazioni sono partizionate contemporaneamente in un certo numero di cluster mutuamente esclusivi; per mezzo di una procedura iterativa, i cluster vengono di volta in volta ridefiniti tramite lo spostamento di osservazioni da un campione ad un altro ricercando l'ottimizzazione di una funzione obiettivo. Nel nostro caso cerchiamo di rendere minima la distanza di ciascun cluster rispetto al centro di massa degli stessi: in altre parole ad ogni iterazione si cerca di diminuire la somma delle distanze di ciascun punto dal centroide del cluster al

quale appartiene il che equivale a ricercare la configurazione più compatta dei punti. Sebbene sia possibile definire a priori il numero di suddivisioni da utilizzare, un approccio di questo tipo porterebbe ancora una volta ad una eccessiva scomposizione della geometria; viceversa vengono tentate diverse scomposizioni differenti con numero di parti via via maggiori (generalmente fino a 5 suddivisioni) per poi decidere in base ad alcuni parametri di forma quale scomposizione si dimostri ottimale.

Nel prossimo capitolo verranno mostrati diversi esempi in cui è stata applicata la metodologia, per l'estrazione dei piani di discontinuità per l'analisi geostatistica di stabilità di pareti rocciose.

3.4. Segmentazione per mezzo della trasformata di Hough

Un metodo più diretto per l'individuazione di entità planari, in cui non si ha la necessità di utilizzare un approccio multi-risoluzione, è rappresentato dall'utilizzo della trasformata di Hough [Hough 1962], [Duda 1972]: il metodo, molto conosciuto nell'ambito della visione artificiale, viene generalmente utilizzato per l'individuazione, nelle immagini, di entità lineari; ciononostante esso è in grado di essere esteso a qualsiasi tipo di primitiva geometrica: l'utilizzo della trasformata per individuare cerchi si può già individuare in [Kimme 1975], mentre in [Vosselman 2001] e in [Sarti 2002] vengono prese in considerazione superfici piane; Rabbani e van den Heuvel hanno recentemente mostrato un'applicazione per l'estrazione di superfici cilindriche [Rabbani 2005].

Inoltre la trasformata generalizzata di Hough, di cui non si parlerà nel seguito, permette di ricercare feature alle quali è difficile dare un'espressione in termini analitici [Ballard 1981].

Per il momento si consideri, per semplicità, il caso dell'individuazione di entità lineari nel piano. La tecnica consiste nell'identificare specifici pattern presenti ad esempio in un'immagine proiettandoli (ovvero trasformandoli) in uno spazio opportunamente definito, detto spazio dei parametri o delle caratteristiche, nel quale è più semplice metterli in evidenza. Caratteristica della trasformazione è la capacità di addensamento dell'informazione: tutti i pixel che soddisfano il modello vengono convertiti in un unico punto dello spazio parametrico. Dal momento il riconoscimento del pattern avviene nello spazio dei parametri (ovvero nello spazio in cui l'informazione è addensata) e non nello spazio immagine (in cui l'informazione ha intrinsecamente carattere locale, ovvero è dispersa) la stima della configurazione avviene in maniera estremamente robusta, ovvero la stima risente poco di effetti di rumore o errori grossolani presenti nel set di dati.

Nel nostro caso lo spazio dei parametri da utilizzare può essere generato da una qualsiasi coppia di valori che individuino univocamente una retta ad esempio l'inclinazione θ rispetto ad un asse di riferimento e la distanza dall'origine ρ . Una retta nel piano xy è rappresentata in tal caso dall'equazione:

$$\rho = x \cos \theta + y \sin \theta \quad (3.8)$$

Nel piano θ, ρ la curva corrispondente ad un punto (x, y) è, come evidente dalla (3.xx), la somma di una

funzione seno e coseno di ampiezza rispettivamente y ed x . Vi è dualità tra i piani x, y e θ, ρ : ad ogni retta in x, y corrisponde un punto in θ, ρ , mentre ad un punto in x, y corrisponde una curva in θ, ρ .

Ipotizzando in base ad un qualche criterio (ad esempio il gradiente elevato dei toni di grigio) che un pixel giaccia su un contorno, ovvero appartenga ad una retta si disegna nello spazio dei parametri la curva corrispondente. Se numerosi pixel giacciono sulla medesima retta, nello spazio dei parametri le diverse curve si intersecheranno in corrispondenza della coppia θ, ρ che individua tale retta nel piano x, y . Operativamente viene allocata nello spazio θ, ρ una matrice di accumulazione costituita da tante righe e tante colonne quanto sono il numero di discretizzazioni rispettivamente per il parametro θ e per il parametro ρ . Per ciascun punto interessante vengono calcolati i valori di ρ corrispondenti ai diversi valori di θ e viene incrementato di una unità il corrispettivo elemento nella matrice di accumulazione. Una volta esauriti tutti i punti nel piano, la matrice di accumulazione viene analizzata alla ricerca di massimi locali con numerosità sufficientemente elevate. I massimi locali corrispondono ai parametri (θ, ρ) delle rette identificate.

La tecnica illustrata è tanto semplice quanto efficace: non solo è poco influenzata da errori grossolani o fenomeni di disturbo ma, anche in presenza di contorni frammentati in più porzioni, è in grado di determinare correttamente la linea comune a tutti i segmenti. Il metodo, nella forma lineare, è stato ad esempio utilizzato in [Roncella 2005b] per la determinazione automatica delle strisce di delimitazione della carreggiata stradale per mezzo di analisi di immagine.

L'impostazione può essere facilmente estesa al caso di superfici piane: in questo caso la funzione di accumulazione diventa:

$$\rho = (x \cos \theta + y \sin \theta) \cos \phi + z \sin \phi \quad (3.9)$$

in cui ρ è la distanza del punto dall'origine del sistema di riferimento, ϕ è l'angolo formato dalla normale al piano con l'asse delle z e θ l'angolo con quello delle x . Per mezzo della (3.9) un punto nello spazio (euclideo) tridimensionale si trasforma, al variare di θ e ϕ in una superficie trigonometrica nello spazio

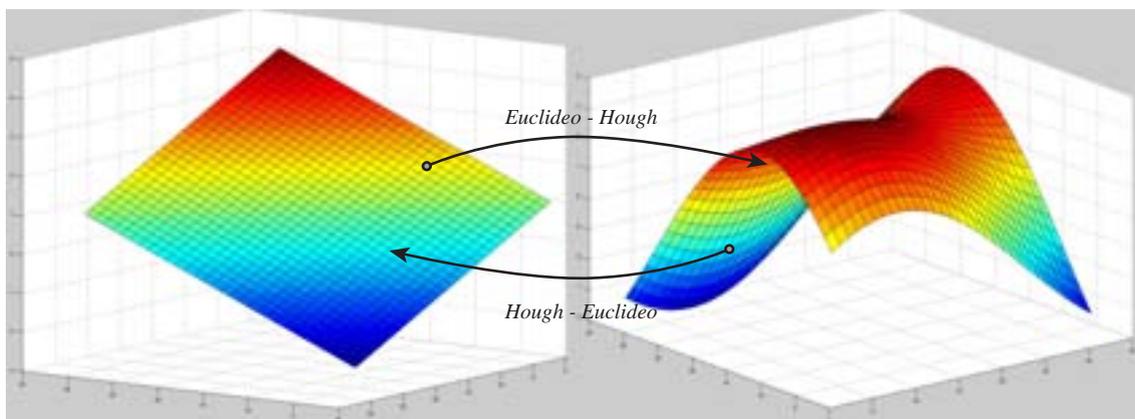


Figura 3.8: Trasformazioni fra spazio euclideo e spazio dei parametri: un punto nel primo corrisponde ad una superficie nel secondo e viceversa.

dei parametri; viceversa un punto nello spazio dei parametri rappresenta una superficie piana nello spazio euclideo come mostrato in figura 3.8.

La procedura di estrazione si svolge come segue: si determina un passo di discretizzazione per i due parametri θ e ϕ : in molti casi si possono scegliere passi relativamente ampi (2÷5 gradi); si analizza poi la nuvola per determinare il campo di variabilità del parametro ρ : il limite inferiore è sempre 0 (piano passante per l'origine del riferimento) mentre quello superiore può essere fissato pari alla distanza massima dall'origine dei punti della nuvola, o, con maggiore efficienza, applicare una trasformazione ai punti della nuvola in modo da far coincidere il centro del parallelepipedo che la racchiude (*bounding box*) con l'origine del sistema di riferimento; per ottimizzare ulteriormente la procedura si può orientare il suddetto parallelepipedo in modo che i suoi lati risultino paralleli agli assi principali d'inerzia della nuvola. Una volta stabilito un passo di discretizzazione anche per quest'ultimo parametro, si genera la corrispondente matrice (tridimensionale) di accumulazione, ponendo tutti i suoi elementi a 0; per ciascun punto della nuvola si determina, tramite la (3.9), al variare dei due parametri θ e ϕ , la superficie di accumulazione ρ (figura 3.9); tutti gli elementi della matrice di accumulazione che contengono punti di tale superficie vengono incrementati di una unità; una volta reiterato il procedimento per tutti i punti si analizza la matrice di accumulazione alla ricerca di massimi locali: i parametri ρ , θ , ϕ corrispondenti rappresentano i piani cercati. Per l'analisi dei massimi si può utilizzare un volume di ricerca di dimensioni appropriate (costituito ad esempio da $3 \times 3 \times 3$ elementi o, se richiesto un grado di approssimazione maggiore, anche più grande). Quando

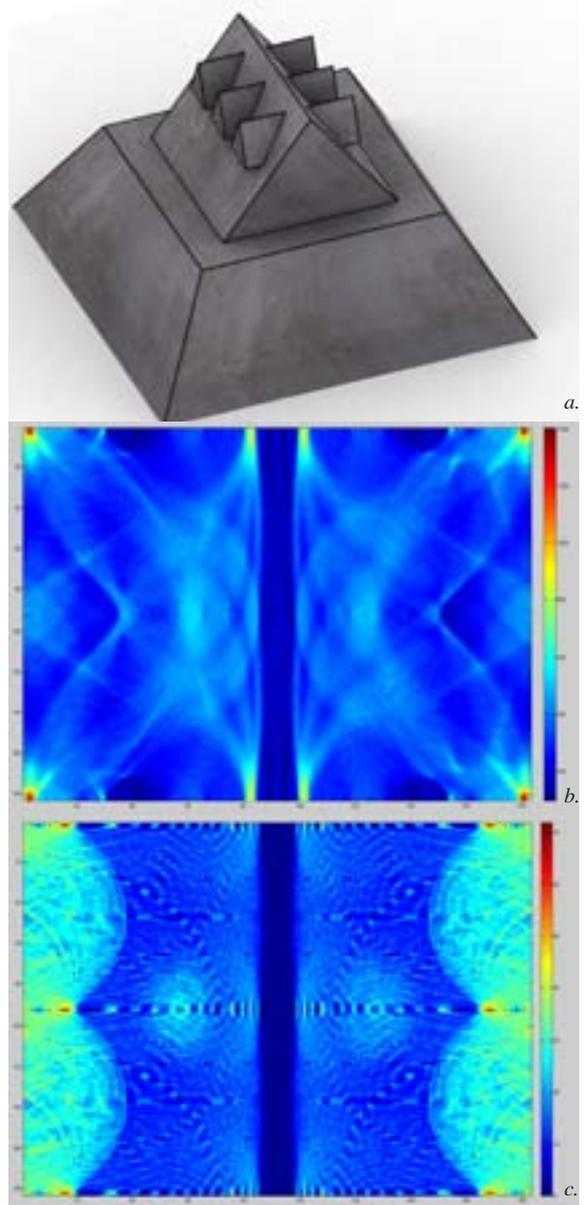


Figura 3.9: Spazio di accumulazione della trasformata di Hough: (a.) l'oggetto da analizzare; (b.) e (c.) sezioni dello spazio di accumulazione per due differenti valori di ρ .

parte di tale volume, analizzando elementi di bordo della matrice di accumulazione, si trova parzialmente al di fuori di quest'ultima nel piano $\theta\phi$, si possono aggiungere gli elementi sul bordo opposto, in modo da considerare anche piani che presentano parametri prossimi a $180^\circ/360^\circ$.

Se richiesto un grado di dettaglio superiore al livello di discretizzazione utilizzato per generare la matrice di accumulazione, si può valutare la posizione del piano nello spazio di Hough in corrispondenza di ciascun massimo locale, calcolando il centroide degli elementi di accumulazione all'interno del volume di ricerca centrato sul punto individuato.

Infine, una volta ottenuti i parametri dei piani, si possono, tornando a considerare il set di dati nello spazio euclideo, ricercare quei punti che distano meno di una certa soglia da ciascun piano, in modo da ottenere la porzione di nuvola corrispondente. La soglia deve essere impostata su valori non inferiori al passo di discretizzazione del parametro ρ , per evitare di escludere punti che hanno contribuito nello spazio dei parametri alla definizione della superficie. Al termine è comunque consigliabile ristimare il piano a minimi quadrati ed eventualmente ripetere il processo di determinazione dei punti, soprattutto se è richiesto un grado di precisione elevato o la discretizzazione in ρ , per motivi di occupazione di memoria, è stato mantenuta su valori alti.

CAPITOLO 4

Determinazione dei piani di discontinuità di pareti rocciose

In questo capitolo conclusivo si illustra con maggior dettaglio una serie di esempi di applicazioni delle tecniche di automazione sviluppate nel corso del presente lavoro. Per il suo carattere innovativo e poiché impiega la maggior parte delle predette tecniche, si è scelta la determinazione automatica dei piani di discontinuità geostrutturale nell'ambito dell'analisi di stabilità di corpi rocciosi.

Il tema è di grande attualità, come dimostra la ricca letteratura a riguardo, anche se solo negli ultimi anni si sta affrontando il problema con tecniche di rilievo indirette [Harrison 2000], per mezzo di stazioni totali [Feng 2001], fotogrammetria terrestre [Kemeny 2003], [Roncella 2005c], o per mezzo di laser a scansione [Monte 2004], [Slob 2004], [Biasion 2004]. Approcci automatici per la determinazione delle discontinuità sono di recente introduzione ([Roncella 2005c], [Slob 2005]) e, sebbene ancora la sperimentazione sia nelle prime fasi, l'interesse dimostrato dal mercato dimostra come la tematica possa svilupparsi nei prossimi anni.

4.1. Lo studio delle discontinuità negli ammassi rocciosi

Nonostante le apparenze, un ammasso roccioso è tipicamente un mezzo discontinuo, suddiviso in volumi elementari, con caratteristiche di resistenza che, globalmente, dipendono dalla presenza di giunti e discontinuità di diverso tipo. L'analisi di stabilità di un pendio in roccia, cioè la previsione della risposta tensionale e deformativa della massa rocciosa alle forze agenti, si appoggia a modelli matematici per schematizzarne il comportamento geomeccanico. La scelta del tipo di modellazione più idonea ad un determinato problema dipende da diversi fattori quali: l'orientazione delle discontinuità, la spaziatura, la persistenza delle fratture e la scala a cui il problema viene esaminato. I parametri geometrici, fisici e meccanici necessari alla descrizione dell'ammasso roccioso vengono raccolti tramite una serie di rilievi geostrutturali attraverso la selezione intelligente dei dettagli geologici e l'individuazione dei parametri importanti dal punto

di vista del comportamento meccanico. Osservando, ad un'opportuna scala, le superfici di affioramento di un complesso roccioso si constata la presenza di sistemi di fratture che talvolta delimitano chiaramente dei blocchi; in altri casi al contrario la loro esistenza non è affatto evidente.

Il comportamento statico e la stabilità degli ammassi dipendono fortemente dalle discontinuità strutturali presenti, ovvero da quegli elementi che ne interrompono la continuità e che, essendo superfici di debolezza statica, possono causare cedimenti (scivolamenti, ribaltamenti, crolli...). Le discontinuità possono essere rappresentate da giunti, ovvero superfici di rottura all'interno della massa rocciosa in cui non sono evidenti movimenti tra le due pareti che possono originarsi sia per trazione sia per taglio, da faglie, ovvero fratture o zone di frattura lungo cui c'è stato uno spostamento riconoscibile (da pochi centimetri ad alcuni chilometri) o, a scala microscopica, dalle scistosità cioè da caratteristiche della matrice rocciosa che rappresentano dei piani di debolezza derivanti dall'orientazione preferenziale dei minerali costituenti la roccia.

Le tracce dei giunti sugli affioramenti, nell'analisi geostrutturale, vengono generalmente rappresentate con segmenti di retta e le superfici sono di norma assimilate a elementi piani. Tale semplificazione rende numericamente affrontabile il problema, permettendo la determinazione dei parametri di fratturazione dell'ammasso.

Per caratterizzare la geometria fessurativa si considerano due livelli differenti: le fratture individuali, ovvero le discontinuità e le loro caratteristiche analizzate singolarmente, e i sistemi di fratture, in cui invece si analizza complessivamente l'insieme di discontinuità rilevate, cercando segni della presenza di sistemi di discontinuità, ovvero di insiemi (o famiglie) di piani di discontinuità aventi caratteristiche simili (giacitura, spaziatura...) e, pertanto, rappresentabili schematicamente con un piano medio, individuato con un'analisi statistica.

Le grandezze che caratterizzano le discontinuità, sono diverse:

1. **giacitura:** ovvero l'orientazione di un piano di discontinuità definita, secondo le convenzioni generalmente in uso, per mezzo di due angoli: l'inclinazione (*dip*), che rappresenta l'angolo che la retta di massima pendenza del piano di discontinuità forma con un piano orizzontale e la direzione di immersione (*dip direction*) cioè l'angolo tra la direzione del Nord e la proiezione della linea di massima pendenza sul piano orizzontale. In figura 4.1 sono schematicamente rappresentati tali parametri.
2. **persistenza:** ovvero l'estensione areale o la dimensione di una discontinuità entro un piano definita come la frazione di area della discontinuità; può essere approssimativamente quantificata osservando le lunghezze delle tracce di discontinuità

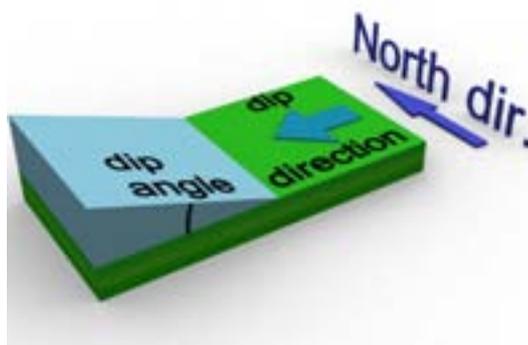


Figura 4.1: Definizione degli angoli di dip e dip direction

sulla superficie esposta.

3. **spaziatura:** la distanza, misurata lungo una linea retta d'orientamento prefissato, tra discontinuità adiacenti della stessa famiglia;
4. **rugosità;**
5. **resistenza di parete;**
6. **apertura:** definita come la distanza, misurata perpendicolarmente, tra le due pareti adiacenti di una discontinuità;
7. **riempimento;**
8. **presenza d'acqua;**
9. **numero di sistemi di discontinuità:** si tratta di un fattore determinante per la stabilità di un pendio roccioso assieme all'orientazione delle discontinuità relativamente al fronte dello stesso: se non esiste un numero sufficiente di sistemi di discontinuità, la probabilità d'instabilità è praticamente ridotta a zero; al contrario più sistemi ci sono, più si hanno possibili cinematismi all'interno dell'ammasso.
10. **dimensione dei blocchi:** che è un indicatore estremamente importante del comportamento della massa rocciosa; le misure dei blocchi sono determinate dalla spaziatura delle discontinuità, dal numero di sistemi e dalla persistenza delle discontinuità che delimitano i blocchi potenziali.

La manifestazione più evidente della strutturazione dei complessi rocciosi è l'esistenza di famiglie di discontinuità che presentano le medesime caratteristiche individuali e sono disposte con una certa regolarità nello spazio. Le diverse famiglie sono identificate da: un'orientazione preferenziale, una spaziatura media, in direzione perpendicolare all'orientazione, una lunghezza media e un'apertura media. Per fornire una descrizione scientifica del fenomeno gli studiosi solitamente cercano di individuare tali parametri rapportandoli a distribuzioni probabilistiche: ricorrere a leggi di probabilità conosciute permette di descrivere le famiglie con un numero limitato di parametri e di ipotizzare l'andamento della fratturazione dove le discontinuità non sono osservabili. Solitamente si studia pertanto la ripartizione delle giaciture dei piani individuando il polo di orientazione caratteristico della famiglia. Conoscendo unicamente la distribuzione dei parametri che definiscono le giaciture dei piani di discontinuità osservabili sulla superficie esterna dell'ammasso roccioso è cioè possibile, in termini probabilistici, caratterizzare approssimativamente il comportamento strutturale dell'ammasso. Gli altri parametri vengono in-



Figura 4.2: Rilievo tradizionale delle superfici di discontinuità condotto da un rocciatore professionista.

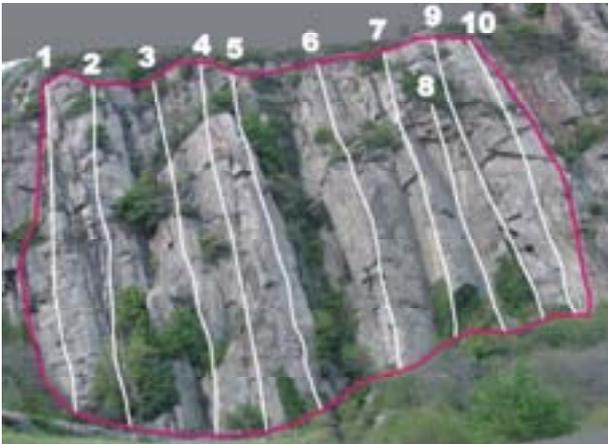


Figura 4.3: Esempio di rilievo condotto per *scanline* lungo la parete.

seriti in modelli numerici per simulare il comportamento statico della parete rocciosa, ed individuare con maggior precisione i possibili punti deboli.

Un progetto geotecnico riguardante strutture rocciose deve sempre prendere l'avvio con l'esecuzione di un "rilievo geostrutturale", rivolto alla raccolta di dati per la descrizione quantitativa delle discontinuità presenti nella massa rocciosa. Il rilievo si effettua in sito, utilizzando un equipaggiamento portatile, composto da una bussola e una livella accedendo direttamente all'ammasso roccioso (figura 4.2) lungo stendimenti, detti *scanline* (figura 4.3).

Tale metodologia, seppur estremamente diffusa, presenta innegabili limitazioni:

- gli ammassi rocciosi possono essere difficilmente raggiungibili, ed avere forme geometriche che rendono difficile, se non impossibile, l'accesso. L'acquisizione di dati in questo caso può essere un processo lungo e dispendioso;
- gli ammassi rocciosi molto fratturati, non supportati da opere di sostegno, possono comportare rischi per l'operatore (solitamente un rocciatore professionista, che segue le indicazioni del geologo incaricato per il rilievo);
- la scelta di elementi d'interpretazione significativi può essere difficile da eseguire in sito. Spesso, in fase di analisi dei risultati, l'interpretazione e l'elaborazione delle misurazioni conduce ad ambiguità che richiedono chiarimenti per mezzo di un ulteriore rilievo.

I metodi di rilievo tradizionali non consentono inoltre, in molti casi, di cogliere alcuni aspetti del sistema di fratturazione che sono indispensabili all'analisi della stabilità di un ammasso roccioso. In particolare, le relazioni esistenti tra le famiglie di discontinuità (definiti dalla gerarchia dei sistemi di fratturazione), evidenziati dalla tipologia di interruzione dei sistemi di fratturazione, possono essere molto importanti dal punto di vista del comportamento meccanico d'insieme dell'ammasso. Ciò è intuitivo se si pensa alla differenza tra la stabilità di un muro di mattoni incolonnati direttamente su quelli sottostanti e la solidità di un muro nel quale la posizione dei mattoni è alternata nelle diverse file.

Infine un rilievo tradizionale presenta costi molto elevati in quanto è richiesto un notevole dispendio di risorse umane non trascurabile, a fronte di una produttività estremamente ridotta: solitamente in una giornata di rilievi si è in grado di coprire pareti che si estendono per non più di 50÷70 m. Ciò rende pressoché impraticabile la possibilità di svolgere campagne di misurazione particolarmente estese e richiede un'attenta progettazione delle operazioni. Al contrario, per numerose amministrazioni ed enti pubblici, si

sente sempre più la necessità, sia per aspetti di protezione civile e del territorio, sia per aspetti legati allo sfruttamento turistico delle risorse montane, di avere un quadro il più possibile esteso delle caratteristiche statiche degli ammassi rocciosi.

Per superare tutte queste limitazioni sono stati proposti [Harrison 2000] metodi alternativi basati essenzialmente su un'analisi geometrica della superficie dell'ammasso (DSM) ricostruito con un rilievo fotogrammetrico. I parametri di dip e dip direction possono essere facilmente ottenuti una volta nota l'equazione del piano che rappresenta la discontinuità in un sistema di riferimento orientato a Nord; il piano può essere stimato da una serie di punti ben distribuiti sulla superficie di discontinuità, ad esempio per mezzo di una stima a minimi quadrati; il centro di massa di tali punti può fungere da indicatore della collocazione spaziale della discontinuità. Procedendo sistematicamente lungo lo sviluppo della parete rocciosa si genera un database delle discontinuità, descritte per mezzo della loro posizione e del loro orientamento, da impiegare nell'analisi statistica della loro distribuzione o nella costruzione geometrica del modello numerico. L'approccio descritto, sebbene in grado di superare tutte le limitazioni presenti nel rilievo diretto, deve soddisfare alcuni requisiti essenziali affinché tale metodologia possa a buon diritto essere considerata un reale passo avanti rispetto alle tecniche tradizionali: deve essere possibile ottenere punti misurabili su tutta l'estensione rocciosa considerata ed in particolar modo sulle principali o significative superfici di discontinuità; la precisione di misura in termini di dip e dip direction deve, con buona garanzia, essere almeno analoga a quella ottenibile con bussola geotecnica (alcuni gradi sessagesimali); il metodo deve in effetti presentare la soluzione ai problemi di accesso alla parete e di sicurezza degli operatori coinvolti; il metodo deve anche rappresentare una soluzione ottimale in termini di tempi e di risorse impiegate, nel senso di incrementare la produttività delle procedure di misura a costi sostanzialmente inferiori.

4.2. Un sistema fotogrammetrico per il controllo di stabilità dei versanti in roccia

Nell'ambito del progetto INTERREG III, n. 23: "Sviluppo di metodi per il riconoscimento e lo studio della propagazione delle valanghe di roccia" e in collaborazione con il gruppo di Meccanica delle rocce del nostro Dipartimento [Ferrero 2004], lo scrivente ha realizzato e sperimentato un sistema di misurazione altamente automatizzato per la stima delle caratteristiche di discontinuità degli ammassi rocciosi basato su tecniche fotogrammetriche e di image sequence analysis, il cui schema procedurale è brevemente riassunto in figura 4.4. Il sistema ha costi molto contenuti sia dal punto di vista dell'hardware richiesto (una camera digitale reflex con risoluzione di 5-6 Mpixel e ottiche calibrate da grandangolare a teleobiettivo), sia dal punto di vista del personale necessario: sul posto è sufficiente un singolo operatore, con un certo bagaglio di conoscenze tecniche sul rilievo fotogrammetrico e chiare indicazioni sulle modalità di presa delle immagini, mentre l'elaborazione è svolta via software, sia interattivamente sia automaticamente.

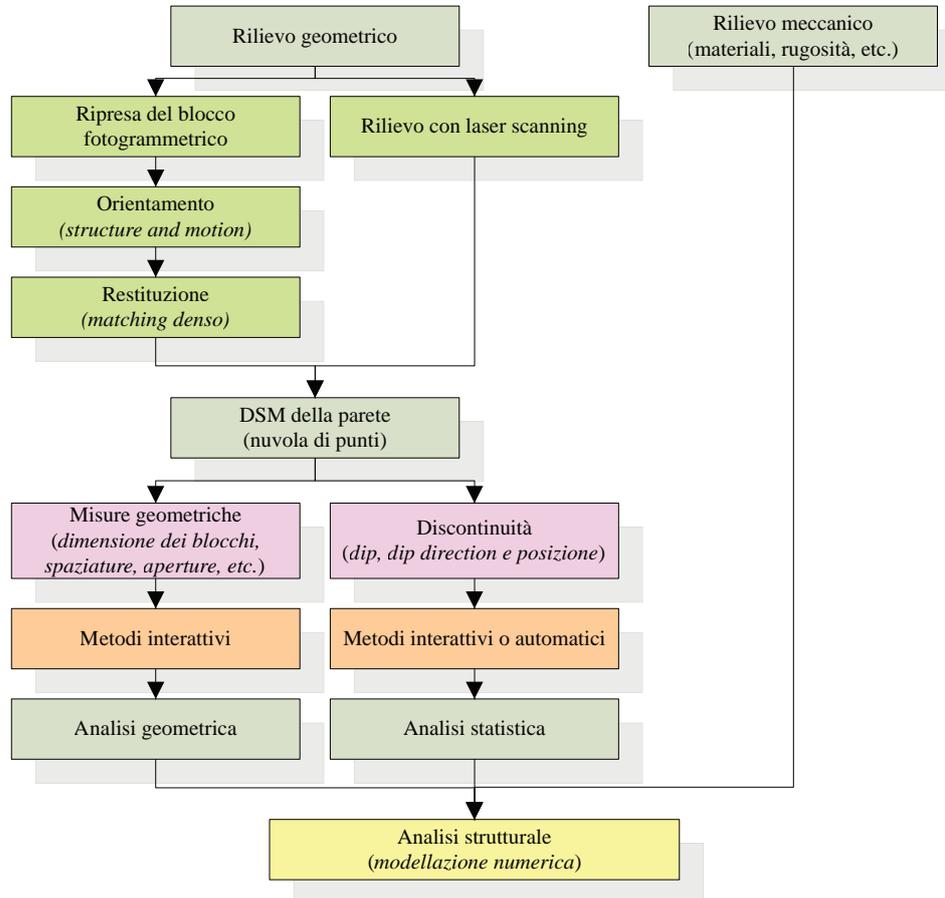


Figura 4.4: Schema procedurale per il rilievo, la modellazione e l'analisi dei fenomeni di instabilità di pareti rocciose.

4.2.1. Progetto e ripresa delle immagini

L'acquisizione delle immagini richiede una certa esperienza e sensibilità da parte dell'operatore: in particolare l'attenzione ad assicurare assenza o minimizzazione delle occlusioni, una buona copertura multi-immagine della superficie da rilevare, per garantire omogeneità di precisione e affidabilità del blocco.

Per la grande varietà di situazioni cui ci si può trovare di fronte si possono solo indicare intervalli di massima per i parametri geometrici di riferimento nella progettazione del blocco: dai 3 esempi sui quali si è lavorato si può desumere che, in relazione alle caratteristiche tipologiche della roccia ed all'ampiezza della struttura, la scala fotogramma deve consentire una risoluzione al suolo da 2-3 cm a 10-12 cm (densità di campionamento di 1 punto ogni 10-40 cm). Per quanto riguarda la precisione del DSM, essa andrebbe a rigori derivata da quella necessaria per i piani di discontinuità, dell'ordine di 4-5°. L'incertezza degli angoli dip e dip direction dipende infatti primariamente da 2 elementi:

- rapporto tra la precisione delle coordinate dei punti e la distanza media tra i punti;
- “gibbosità” naturale della superficie rocciosa.

Pertanto, porzioni di roccia sensibilmente piane e di forma pressoché quadrata si prestano meglio di altre, di forma allungata; a parità di forma, superfici estese permettono di scegliere i punti a distanze medie maggiori e quindi richiedono minor precisione alle coordinate dei punti. Risultati soddisfacenti sono stati raggiunti con precisioni tra 2 e 30 cm circa.

Per quanto riguarda la presa, per applicare con successo le tecniche di S&M, occorre effettuare riprese in sequenza, fino ottenere una base ottimale tra primo ed ultimo fotogramma; la base intermedia lungo la sequenza deve essere opportunamente proporzionata. Per contenere le occlusioni e assicurare coperture multi-immagine occorre cercare punti di presa diversi, variando se necessario la focale. Importante è inoltre evitare che famiglie di discontinuità siano occluse o mal determinabili (ad esempio perché viste molto di scorcio), di nuovo agendo sulla scelta dei punti di presa. Il fatto che, nella prospettiva del progetto, la ripresa sia affidata ad un geologo o comunque che vi sia stata una ricognizione preliminare, è sotto questo aspetto assai importante, in quanto l'esame visivo della struttura è sufficiente al geologo per cogliere gli elementi caratterizzanti da rilevare.

Se l'area è molto estesa, può essere necessario determinare punti di appoggio, per evitare deformazioni del blocco; in tal caso la soluzione ideale è adottare teodoliti reflectorless di grande portata, con immediata realizzazione della monografia relativa ad ogni punto. Un'alternativa potrebbe essere quella di riprendere le immagini con camera fissata a palina con antenna GPS, per vincolare il blocco sui centri di presa. Con strutture molto grandi o siti inaccessibili l'unica possibilità è la ripresa da elicottero, che richiede una preliminare ed attenta pianificazione delle riprese.

La ricostruzione della S&M si effettua in un sistema di riferimento arbitrario con scala arbitraria; l'analisi geotecnica richiede invece una corretta messa in scala del modello, il riferimento alla verticale e l'orientamento al Nord con l'approssimazione di qualche grado. Su questo aspetto non si è ancora definito un metodo applicabile in tutte le situazioni e probabilmente occorrerà, a seconda delle circostanze, optare per una tra varie soluzioni, diversa per impegno tecnologico, competenze richieste, livello di precisione e affidabilità.

Impiegando un elicottero, è possibile montare la camera solidale ad una antenna GPS tramite un elemento rigido [Vallet 2000]. In tal caso, interfacciando la camera al GPS, si possono inserire nella compensazione del blocco le posizioni dell'antenna agli istanti di presa e impiegarli come nella Triangolazione Aerea assistita da GPS [Forlani 1994], referenziando il blocco in WGS84. Esperimenti condotti a terra nel campus di Parma hanno confermato la praticabilità e precisione del metodo, a condizione di effettuare due strisciate ricoperte trasversalmente. Qualora la costellazione GPS sia povera e impedisca una soluzione con le fasi, una possibilità è quella di limitarsi al calcolo del fattore di scala dalle posizioni con misure di codice, usando i punti di presa agli estremi della strisciata.

Per rilievi da terra, si può operare con camera e GPS su palina, come accennato; in alternativa, il fattore

di scala e il riferimento alla verticale possono essere ottenuti con metodi speditivi, ad esempio riprendendo nella sequenza una stadia posta in verticale in più punti; il riferimento al Nord può di nuovo essere misurato con bussola a traguardo, materializzando due punti di presa. Altre soluzioni sono ovviamente possibili con strumentazione topografica (livello o teodolite): dal rilievo di punti d'appoggio e dei centri di proiezione all'imposizione di vincoli (orizzontalità, verticalità, distanza, dislivello) tra questi; nel quadro del progetto, tuttavia, si vorrebbero limitare sia le esigenze di strumentazione sia le competenze tecniche richieste.

4.2.2. Il software di elaborazione ed il flusso di informazioni

L'orientamento delle immagini e la generazione del DSM dell'ammasso roccioso sono sostanzialmente automatiche, se la geometria di presa e la tessitura della roccia hanno le caratteristiche adatte; la georeferenziazione invece, per la quale esistono più opzioni, è ovviamente interattiva e può essere effettuata con GPS o metodi speditivi. La segmentazione dei piani, la stima delle dip e dip direction possono essere effettuate sia interattivamente sia automaticamente (fatto salvo l'affinamento di alcuni parametri nell'algoritmo di segmentazione del DSM); si misurano poi manualmente gli altri parametri geometrici che caratterizzano la parete (spaziatura, dimensione dei blocchi etc.). Un'opportuna interfaccia permette poi l'immediato inserimento dei risultati nei programmi dedicati all'analisi statistica e strutturale dell'ammasso, rendendo l'intero processo estremamente snello.

4.2.3. Versanti in roccia: Fotogrammetria o Laser Scanner?

Dal momento che il metodo proposto da Harrison è essenzialmente geometrico, esso è indipendente dalla tecnologia con cui sono stati rilevati i dati: in alternativa alla fotogrammetria, quindi, può essere impiegato il laser a scansione terrestre. Obiettivo intermedio del progetto INTERREG è pertanto anche il confronto fra questi sistemi di misura (o altri adatti allo scopo), per valutare pro e contro di ciascuno. Sebbene una dettagliata analisi comparativa delle due tecniche sia estranea agli scopi del presente lavoro, sembra opportuno integrare le considerazioni generali del § 2.6. con alcune proprie del rilievo di pareti in roccia.

Il laser scanner si presenta come uno strumento di maggior semplicità d'uso (in questo contesto è richiesta una minor sensibilità nella progettazione del rilievo rispetto a quella di un equivalente blocco fotogrammetrico); in quanto strumento attivo, risente in misura assai minore o trascurabile della tessitura della roccia e delle condizioni di illuminazione (ombre) rispetto alla correlazione di immagini: precisione, risoluzione e affidabilità dei risultati sono legati a prestazioni hardware (peraltro in continuo miglioramento) piuttosto che a elaborazioni software con percentuali di successo variabili; è meno critica la presenza di occlusioni, poiché basta una sola visuale libera invece delle due necessarie in fotogrammetria e quindi vi è maggior libertà di scelta dei punti di stazione; il risultato del rilievo in campagna è direttamente la nuvola di punti, senza elaborazioni intermedie.

Per contro, sebbene destinati a calare in qualche misura con la diffusione crescente della tecnologia, i costi di acquisizione dell'hardware sono oggi inaccessibili alla maggior parte degli utenti e rimarranno di ben oltre un ordine di grandezza superiori a quelli della fotogrammetria (anch'essi in calo pur con risoluzione delle immagini in crescita); nel caso specifico, tuttavia, quello che appare come (a nostro avviso) un importante e forse decisivo vantaggio a favore della fotogrammetria è la sua maggior flessibilità operativa. Il laser infatti deve necessariamente operare da stazioni a terra sufficientemente agevoli da raggiungere (dato il costo e il peso delle apparecchiature); questo vale ovviamente anche per le prese fotogrammetriche da terra, se non si vogliono impiegare fotogrammetri-alpinisti. Tuttavia basta far mente locale e appare subito evidente che moltissime sono le situazioni in cui è indispensabile (per le condizioni di accessibilità alla parete, il mascheramento da parte della vegetazione, la distanza dalle più vicine strade o sentieri, la pendenza del versante, le visuali defilate od occluse) effettuare il rilievo impiegando un elicottero, dove le esigenze in termini di geometria di presa della fotogrammetria possono essere soddisfatte senza particolari difficoltà. Questo senza negare che in un numero certamente non piccolo di casi è ovviamente possibile operare anche da terra (e vi sono anche casi, in valli molto strette, in cui l'elicottero sarebbe a rischio e la fotogrammetria avrebbe maggiori difficoltà rispetto al laser nel trovare stazioni di presa soddisfacenti). Per impostare una campagna "a tappeto" in ambito regionale, l'impiego dell'elicottero ci sembra tuttavia una soluzione al tempo stesso economica e obbligata.

Per completare il quadro delle tecnologie disponibili, sembra corretto aggiungere una ulteriore possibilità, costituita da un laser a scansione da elicottero, con inerziale e GPS. Occorrerebbero tuttavia opportune modifiche all'installazione del telemetro, per consentire prese oblique od orizzontali. Non sembra che la strumentazione sul mercato, anche per valori di divergenza del raggio non trascurabile, abbia la possibilità di modificare i parametri di scansione per raggiungere le risoluzioni elevate richieste in queste applicazioni; infine, non va dimenticata la limitazione alla capacità operativa di questi sistemi con costellazioni GPS in vista spesso assai povere.

4.3. Sito campione: Arnad

La metodologia fotogrammetrica è stata messa a punto sul sito pilota ubicato nel comune di Arnad, in bassa Valle d'Aosta, sul versante idrografico sinistro della Dora Baltea, presso il fondovalle (in figura 4.5 viene rappresentato l'inquadramento cartografico del sito). Si tratta di una parete rocciosa estesa per circa 150 m in larghezza e 90 m in altezza.

La Corma di Machaby, struttura cui appartiene il settore studiato, ha le caratteristiche di un esteso dosso montonato la cui forma attuale è il risultato dell'erosione del ghiacciaio Baltea. Le sue dimensioni (circa 1000 m in lunghezza 400 m in altezza) la rendono una delle più rilevanti forme legate al glacialismo della Valle d'Aosta.

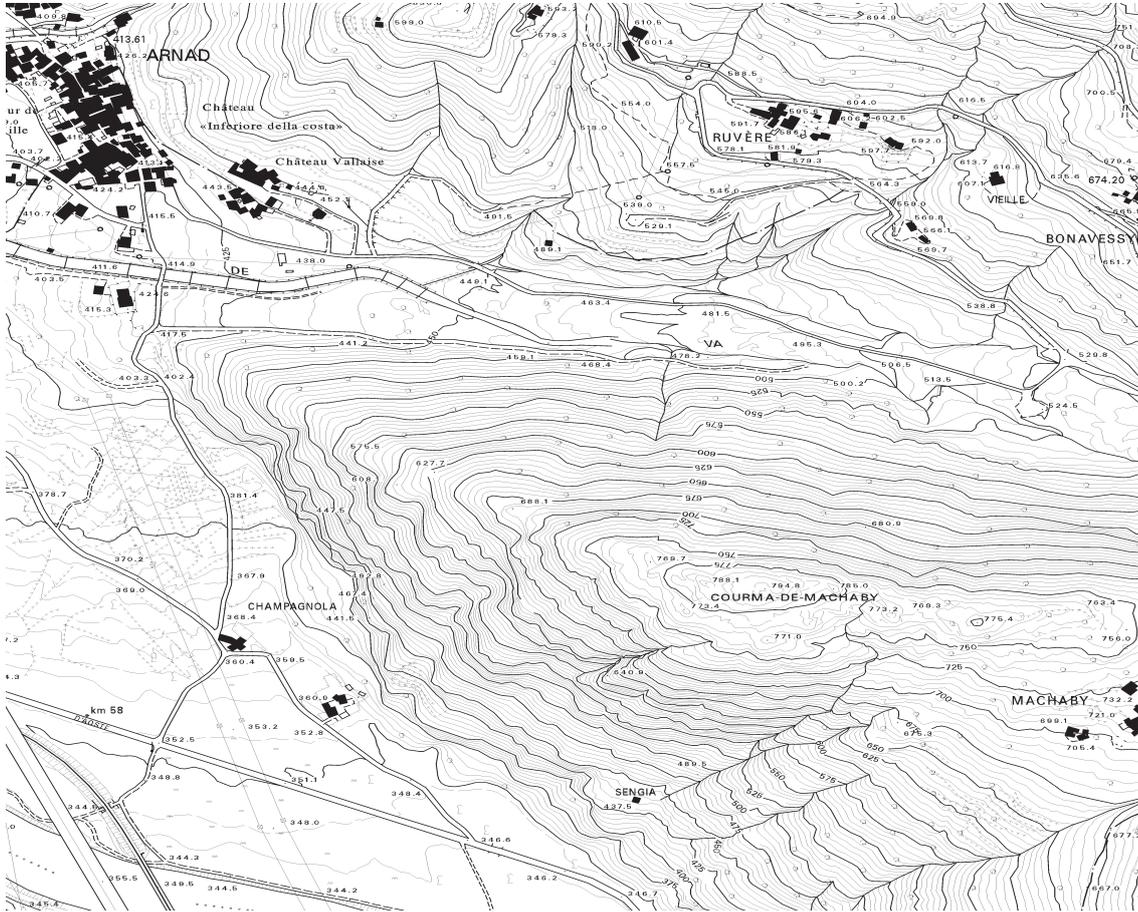


Figura 4.5: Inquadramento cartografico del sito di Arnad.

Sul versante sud-occidentale, rivolto verso l'asse vallivo della Dora, l'azione glaciale ha determinato la quasi completa asportazione della copertura pedologica mettendo in risalto il substrato, quasi ovunque levigato e striato, che costituisce una parete con superficie di circa 18 ha. Alla base si osservano limitati depositi di crollo, presenti in modo discontinuo e sviluppati, nei casi maggiori, a scala decametrica. Nel settore rilevato la parete presenta forme sub-verticali a pilastro separate talvolta da stretti canali legati ad elementi strutturali, nei quali non si osservano indizi rilevanti collegati all'azione delle acque di ruscellamento concentrato. Le litologie dell'intera struttura sono costituite dagli gneiss minuti di colore beige chiaro della Zona Sesia Lanzo; localmente si osservano lenti e bande anche decametriche di metabasiti, con andamento concorde con la scistosità (Sc).

Per il sito in esame si disponeva di due rilievi, condotti con metodi tradizionali su 190 punti, usati successivamente come confronto con il metodo fotogrammetrico. Le discontinuità sono costituite in prevalenza da fratture sviluppatesi in fase rigida, con rare e ridotte evidenze di scorrimento relativo delle parti. In alcuni sporadici casi sembra notarsi la dislocazione di alcune discontinuità ad opera di altre; si tratta

di situazioni di entità limitata e di difficile interpretazione a causa della copertura vegetale che la parete presenta in alcune aree fortemente tettonizzate. L'unica dislocazione di una certa rilevanza è localizzata in corrispondenza di una estesa discontinuità sub-orizzontale ben visibile su tutta la parete.

L'analisi dei parametri giacitureali ha consentito di identificare 5 famiglie principali di discontinuità e tre secondarie.

Nello studio del sito, si è effettuato anzitutto un rilievo topografico con una total station no-prism, determinando circa una trentina di particolari della parete, adeguatamente monografati, da impiegare come punti di appoggio e di controllo; diversi blocchi fotogrammetrici sono stati acquisiti in diverse campagne di rilievo: sono state ripresi circa 40 fotogrammi per mezzo di un elicottero, ottenendo due strisciate volate ad altezze di 40 e 80 metri rispetto al piano di base della roccia e a circa 60 metri di distanza dalla parete; tre differenti blocchi sono stati ottenuti a livello del terreno sia con prese pseudo-nadirali, sia con assetto convergente (lunghezza delle baseline approssimativamente di 10 m); infine una sequenza di 17 immagini a baseline ridotte (vd § 1.8.2) è stata ottenuta per testare il metodo di orientamento automatico. Nel seguito proprio quest'ultimo blocco verrà preso in considerazione in quanto di maggior interesse nella presente trattazione. Tutte le immagini sono state acquisite per mezzo di camera digitale Nikon D100 da 6 Mpixel (dimensione del pixel 7.8 μm) con obiettivo Nikkor da 18 mm. Si tratta di un prodotto (e altri equivalenti ma con costi inferiori sono apparsi nel frattempo sul mercato) con un prezzo ormai alla portata della maggioranza degli utenti, ovvero compatibile con la filosofia del progetto di impiegare strumentazione non specialistica e di costo contenuto.

La sequenza è stata elaborata coi moduli di structure and motion ottenendo automaticamente gli orientamenti relativi fra tutti i fotogrammi; utilizzando alcuni dei punti d'appoggio si è poi georeferenziato il blocco con un bundle block adjustment finale; si rimanda al paragrafo sopra citato per una descrizione più ampia dell'operazione, limitandosi in questa sede a riportare solamente i risultati ottenuti dal confronto con i check point.

Arnad (23 check point)			
	X (cm)	Y (cm)	Z (cm)
Media	3.1	-0.3	3.1
RMSE	6	8.2	9.1
Dev. Std.	5.7	9.2	9.5

Tabella 4.1: Risultati ottenuti dal confronto con i check point nel rilievo fotogrammetrico della parete di Arnad.

Una volta noti gli orientamenti assoluti della sequenza si è ottenuta una ricostruzione densa della superficie della parete rocciosa: i punti di legame, sebbene numerosi, non sono ovviamente sufficienti a garantire una buona risoluzione per le successive fasi di segmentazione della parete essendo, fra l'altro, distribuiti non uniformemente. Utilizzando un approccio sviluppato inizialmente all'ETH di Zurigo [D'Apuzzo 2003], che si era già dimostrato estremamente efficiente in questo tipo di applicazioni [Roncella 2005c] è stato svi-

luppato un modulo software in grado di fornire una ricostruzione ad elevata densità di punti della superficie in esame, come illustrato nel paragrafo 2.5.2. L'utilizzo delle equazioni di collinearità, ovvero dell'MGCM, garantisce una geometria finale più coerente ed ha il vantaggio di ottenere direttamente, al termine della procedura, le coordinate tridimensionali di tutti i punti, senza dover ricorrere a operazioni di triangolazione. La spaziatura fra punti adiacenti nell'immagine di riferimento è stata posta a 4 pixel (pari a 12-15 cm sull'oggetto) per un totale di circa 300.000 punti che descrivono la superficie.

Al termine della procedura si sono determinate le famiglie di discontinuità sul modello ottenuto: inizialmente si è utilizzato il metodo interattivo, brevemente descritto in § 3.1, in cui, delineata sull'immagine un'area, una procedura automatica seleziona i punti del DSM che appartengono a regioni assimilabili ad un piano e calcola dip e dip direction per ciascuna regione: il censimento dei piani è avvenuto pertanto con la scelta manuale delle varie porzioni di parete selezionando per ciascuna il fotogramma con l'inquadratura più adatta e avendo cura di non selezionare la stessa zona più di una volta, per non condizionare l'analisi statistica della distribuzione dei poli e quindi l'individuazione delle famiglie di discontinuità. Per capire se il risultato dipendesse in qualche misura dalla dimensione delle aree selezionate, si sono ripetute due volte le misure della medesima parete. Inizialmente sono state selezionate sulle immagini piccole porzioni di DSM (figura 4.6), in modo da avere risultati particolareggiati, correndo peraltro il rischio di avere una maggiore dispersione dei risultati durante la successiva trattazione statistica degli stessi; il numero di piani così ottenuto (177) è dello stesso ordine di grandezza di quello ottenuto col rilievo classico (190). Successivamente sono state contornate porzioni di DSM maggiori, utilizzando un unico fotogramma per individuare le regioni di interesse. Al termine dell'operazione e della successiva elaborazione dei dati si sono ottenuti complessivamente 111 piani.

Il metodo interattivo è stato anche utilizzato per determinare empiricamente i parametri da introdurre nel filtraggio totalmente automatico delle entità planari dal DSM: osservando figura 4.7 è facile notare come una soglia di accettazione troppo ampia nella procedura Ransac conduca, di fatto, a non considerare alcune famiglie di discontinuità che risultano importanti; viceversa una soglia troppo piccola può portare l'algoritmo a suddividere eccessivamente la superficie del modello; ciò conduce, nella successiva analisi statistica delle giaci-

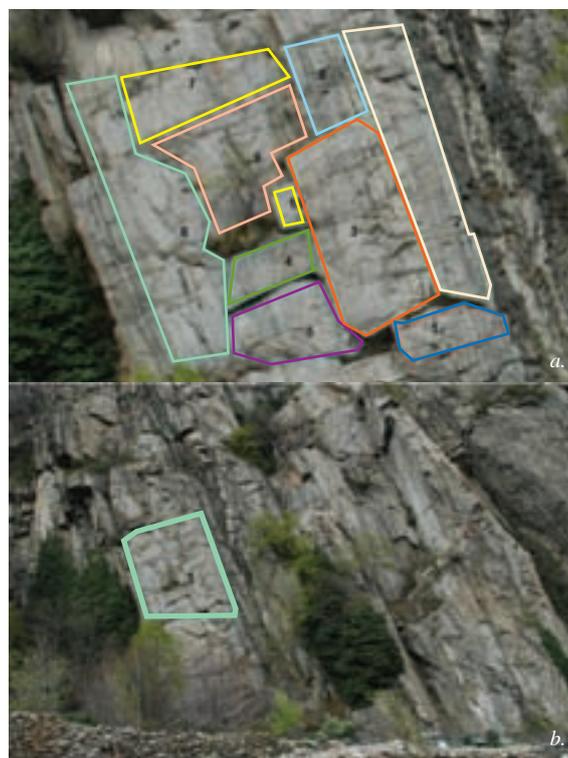


Figura 4.6: Suddivisione del DSM in micro-zone (a.) e in macro-zone (b.)

ture ad un effetto indesiderato: le zone piane che hanno maggiore estensione o che non sono perfettamente piane (presentano cioè delle micro-ondulosità) vengono suddivise in molte parti e vanno a mascherare altre zone, dal momento che, nell'analisi statistica, si considera la numerosità relativa a ciascuna famiglia. Per ovviare al problema, a meno di andare a specificare pesi differenti per ciascun piano, bisogna individuare una segmentazione corretta della superficie: analizzando ad esempio i risultati ottenuti con soglie differenti in figura 4.7, si è determinato che quella ottimale sia compresa fra 10 e 15 cm.

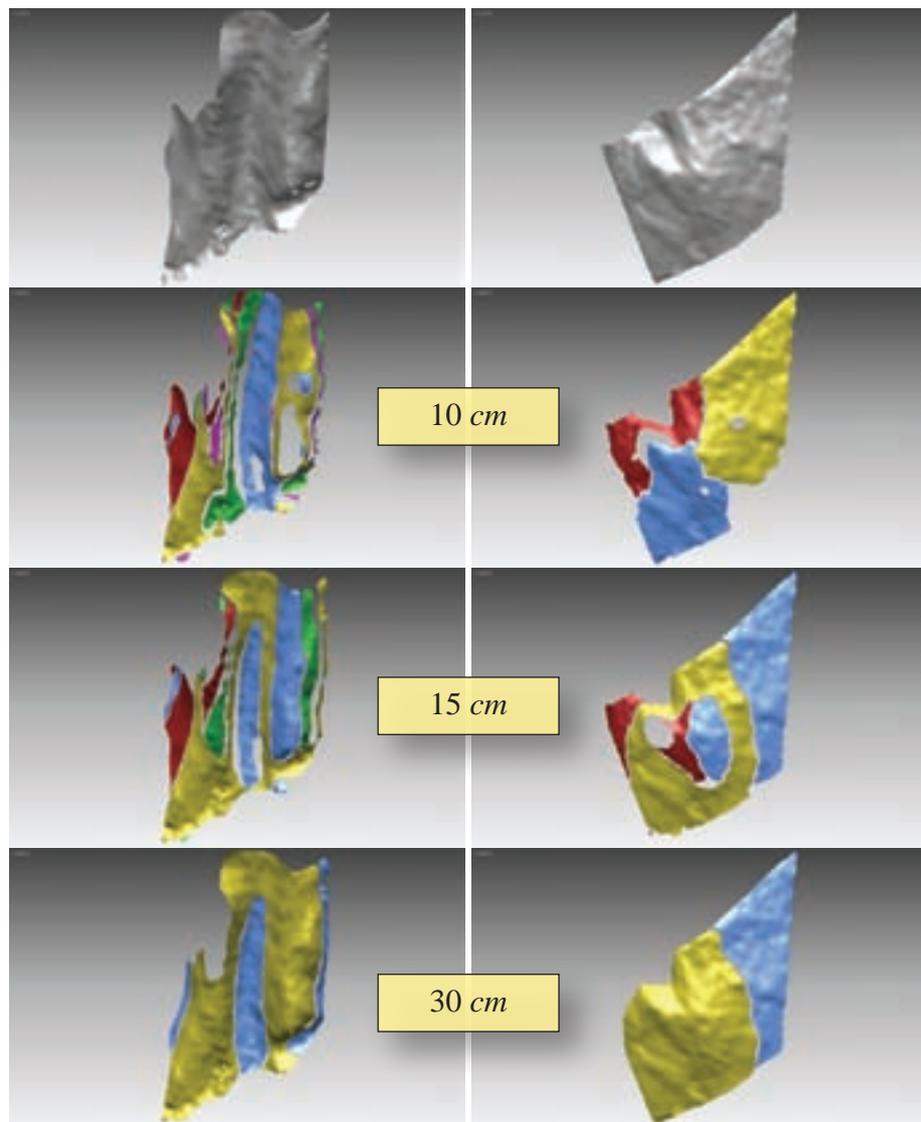


Figura 4.7: Prove di segmentazione per determinare il livello di soglia ottimale in relazione al livello di dettaglio che si vuole ottenere: (a sinistra) analisi di una zona con diverse giaciture marcate; (a destra) analisi di una zona gobba.

L'intera nuvola di punti è stata poi elaborata con l'algoritmo di segmentazione automatica (§ 3.3), utilizzando nella fase finale gli stessi parametri geometrici usati nelle prove precedenti. Il tempo di calcolo richiesto è stato di circa 30 minuti ed ha portato alla definizione di 43 zone differenti. La suddivisione finale è rappresentata in figura 4.8.

Di seguito vengono riportati i risultati ottenuti con le quattro differenti metodologie: per rappresentare ed analizzare i dati di giacitura dei piani di discontinuità individuati si utilizza solitamente un grafico polare che rappresenta la distribuzione sulla sfera di Gauss degli orientamenti di ciascun elemento; i grafici di figura 4.9 devono quindi essere intesi come una rappresentazione polare dell'emisfero settentrionale di tale sfera: i punti visibili rappresentano i poli corrispondenti a ciascun piano, ovvero l'intersezione fra la retta con direzione parallela alla normale alla superficie e passante per l'origine del sistema di riferimento e la sfera di Gauss; le curve rappresentate nello stesso diagramma



Figura 4.8: Segmentazione del DSM ottenuta per mezzo dell'algoritmo automatico.

rappresentano invece l'intersezione fra il piano medio relativo ad una certa famiglia, considerato passante nell'origine e la sfera stessa: per semplicità e sintesi di rappresentazione si determina solitamente una clusterizzazione dei poli di ciascun piano e per ciascun cluster si determina un piano medio (eventualmente pesando opportunamente ciascun polo) in modo da ottenere una giacitura caratteristica della famiglia. Si ricorda che in questo ambito sono richieste precisioni nell'individuazione di tali giaciture caratteristiche di qualche grado (solitamente 4° - 8°) in quanto lo scopo principale dell'operazione è mettere in luce, dal punto di vista qualitativo, i vari sistemi di discontinuità. In questo senso, dunque, le numerose cause di possibile errore o disturbo delle misure, sia nel caso di rilievi tradizionali che di quelli da noi sviluppati, divengono trascurabili a fronte delle tolleranze ammesse.

In figura 4.9.a. vengono illustrati i risultati ottenuti con il rilievo tradizionale, in 4.9.b. e 4.9.c. quelli ottenuti con approccio interattivo rispettivamente individuando micro-zone e macro-zone; infine in figura 4.9.d. vengono illustrati i risultati emersi con il metodo di estrazione automatico.

Si può innanzitutto osservare che la scelta di utilizzare un singolo fotogramma, scegliendo porzioni di

DSM ampie, ha fornito gli stessi risultati ottenuti selezionando manualmente le diverse micro-zone sui fotogrammi in cui erano meglio esposte, nonostante la selezione risulti meno accurata e si corra il rischio di non rilevare alcune famiglie secondarie. Queste famiglie sono ben rappresentate nel caso di selezione di micro-zone, con lo svantaggio però di avere una dispersione maggiore attorno ai valori centrali. Stesso discorso può essere fatto per i risultati ottenuti completamente in automatico.

Dal confronto con il rilievo tradizionale, si ritrovano tutte le famiglie principali ad eccezione di una famiglia che, essendo sub-orizzontale, non è rilevata in quanto occlusa sui fotogrammi ripresi da una quota inferiore e coi punti di presa sostanzialmente giacenti in un piano orizzontale. Selezionando specificamente le diverse zone del DSM che sembrano appartenere a famiglie sub-orizzontali, con un rilievo di dettaglio ottenuto con fotogrammi in cui sia possibile utilizzare algoritmi di matching, ovvero in cui non risultino troppo defilate o scure tali zone, si perviene in effetti ad isolare questo tipo di discontinuità.

Al tempo stesso, rispetto al rilievo tradizionale, è stata individuata una ulteriore famiglia di discontinuità, sub-verticale, che non era stata messa in evidenza nei rilievi con tecnica classica. Tale famiglia appare ben rappresentata nelle analisi statiche e la sua localizzazione

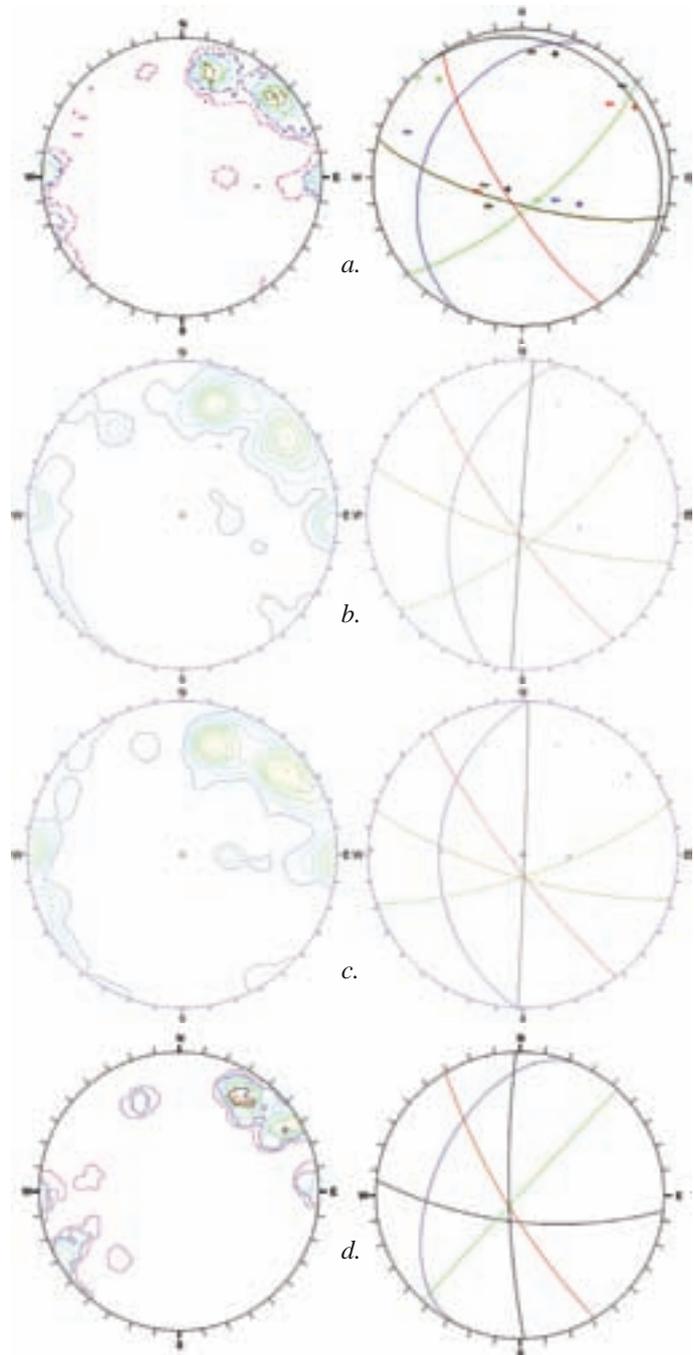


Figura 4.9: Stereogrammi ottenuti con le quattro modalità: (a.) rilievo tradizionale; (b.) rilievo interattivo con micro-zone; (c.) rilievo interattivo con macro-zone; (d.) rilievo automatico.

nelle immagini rappresentanti il sito mostra che si tratta di un sistema non trascurabile: le fratture riconducibili a questa famiglia contribuiscono alla formazione delle strutture colonnari che caratterizzano la porzione destra della parete in esame. Nella precedente analisi dei dati ottenuti col rilievo geologico classico, la famiglia non è stata individuata come sistema a sé stante poiché risultava indistinguibile da un'altra, essendo questi sistemi molto simili e variando l'orientamento con una certa continuità sulla parete.

4.4. Sito campione: Longeborgne

Il secondo caso di studio analizzato riguarda l'ammasso roccioso sovrastante l'Ermitage de Longeborgne, situato tra le Alpi Cernesi e le Alpi Pennine, nel comune di Bramois, nel cantone Vallese della Svizzera. In figura 4.10 viene fornito l'inquadramento cartografico del sito, negli ultimi decenni soggetto a numerosi fenomeni di distacco, che rappresentano un rischio elevato per gli edifici che sorgono ai suoi piedi.

La parete è stata scelta come sito di controllo per la validità del metodo sviluppato: al termine della campagna di analisi, che si concluderà probabilmente a metà 2006, verrà realizzato indipendentemente un rilievo ed uno studio con tecniche tradizionali, per verificare se i risultati emersi siano sostanzialmente gli

stessi e valutare quale metodologia (se il laser o la fotogrammetria) risulti più competitiva. Al momento, pertanto, non si dispone di informazioni relative alle famiglie di discontinuità che interessano la parete, ed è possibile solamente mostrare i risultati ottenuti con le tecniche manuali ed automatiche.

Le pareti rocciose di Longeborgne sono costituite principalmente da quarziti risalenti al periodo Triasico, appartenenti alla "Nappe des Pontis". Si tratta di conglomerati piuttosto massicci, i quali alternano strati di sabbia a stratificazioni incrociate. Queste ultime, più sensibili alle alterazioni che si verificano per ragioni mineralogiche, divengono friabili e polverulente, provocando la formazione di cavità o grotte, le quali caratterizzano la micro-morfologia della parete.

A piccola scala, la morfologia del luogo è caratterizzata da numerosi speroni di roccia separati da faglie orientate in direzione WSW-ENE. La stratificazione principale è molto sviluppata, regolare e in direzione



Figura 4.10: Inquadramento cartografico del sito campione di Longeborgne.

orizzontale. La scistosità è discreta e si confonde con i piani di stratificazione. Le altre discontinuità che interessano la roccia sono disposte verticalmente: l'ammasso roccioso risulta quindi essere costituito da un accatastamento di blocchi dalla forma di parallelepipedo che conferiscono alla roccia un aspetto di "pile in equilibrio".

La frammentazione della roccia è direttamente legata alla granulometria del materiale: dentro gli strati di sabbia grossolana o di conglomerati, le diaclasi (o fratturazioni) risultano essere piuttosto persistenti e sufficientemente spaziate da tagliare l'ammasso roccioso in blocchi, che possono raggiungere le dimensioni di alcuni metri cubi, mentre negli strati di sabbia più fine le diaclasi sono molto più limitate come espansione e poco spaziate tra loro, così da lasciar spazio ad una micro-fessurazione dell'ordine dei millimetri. Il processo di erosione delle pareti è condizionato in gran parte dalle variazioni reologiche che si verificano in seno alle pareti di roccia: gli strati più teneri si corrodono progressivamente fino a causare la destabilizzazione degli strati superiori. In base alla qualità della roccia e alla disposizione strutturale, in alcune pareti più che in altre si genera il fenomeno del distacco e della caduta dei blocchi di roccia; tuttavia le tracce di una recente attività di questo genere sono visibili in ogni parte dell'ammasso (per esempio i segni lasciati dal crollo su alcune pareti o la presenza di blocchi recentemente franati).

Di seguito si parlerà unicamente della parete nord-ovest dell'ammasso in quanto per quella sud-ovest non si dispone ancora di parametri conclusivi. Le dimensioni approssimative di tale parete sono di circa 30 m in larghezza e 25 m in altezza. La difficoltà di accesso a zone in cui fosse possibile avere una buona visibilità della parete ha comportato alcune limitazioni nella realizzazione del blocco fotogrammetrico: esso è costituito da tre differenti strisciate parallele costituite da 4, 5 e 6 fotogrammi ad una distanza media dalla parete rispettivamente di 48 m, 53 m e 57 m ed è stato realizzato per mezzo di una fotocamera digitale Nikon D100 con focale da 18 mm.

Anche in questo caso il blocco è stato risolto utilizzando la structure and motion, ottenendo gli orientamenti relativi in maniera completamente automatica; un rilievo topografico, realizzato in due fasi successive, ha reso disponibili 27 punti d'appoggio con monografia e circa 200 punti presi automaticamente con una stazione totale reflectorless da utilizzare per il confronto con il DSM. Grazie a tali punti si è potuto georeferenziare e verificare il blocco fotogrammetrico. Rimandiamo al § 1.8.2. per

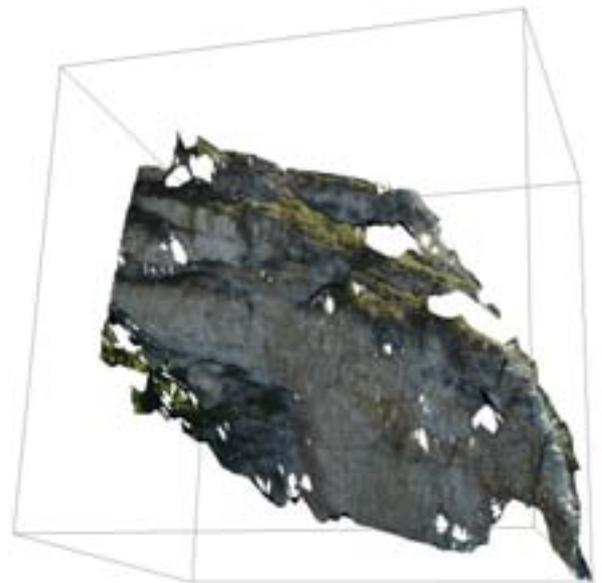


Figura 4.11: DSM della parete nord-ovest dell'ammasso roccioso di Longeborgne con applicata la tessitura fotografica.

l'analisi dei risultati emersi.

Longeborgne (18 check point)			
	X (cm)	Y (cm)	Z (cm)
Media	0.4	0.2	-0.1
RMSE	1.1	1.6	1.2
Dev. Std.	1.1	1.6	1.2

Tabella 4.2: Risultati ottenuti dal confronto con i check point nel rilievo fotogrammetrico della parete di Longeborgne.

Come nel caso precedente i dati di orientamento sono stati utilizzati per generare, tramite procedure di matching, il DSM della parete (figura 4.11): in particolare, in questo caso, si è deciso di generare due immagini epipolari e, per mezzo dell'algoritmo parallelo sviluppato, generare un DSM iniziale della parete; al termine, con il modulo di MGCM, utilizzando 6 fotogrammi distinti e i parametri geometrici desunti nel precedente passaggio, si è ristimata la posizione di tutti i punti. Utilizzando un passo di griglia di 4 pixel sull'immagine master, si sono ottenuti circa 100.000 punti terreno spaziatamente sull'oggetto di 8 cm. La qualità dei dati è stata verificata confrontando il DSM con i 200 punti determinati con la stazione totale.

Distanze (cm) DSM-TS	
Media	-0.95
Dev. Std.	2.78
RMSE	2.92
Err. max	3.84

Tabella 4.3: Distanze fra i 200 punti di controllo ottenuti con stazione totale e il DSM generato tramite la fotogrammetria.

I valori emersi sono in linea con le precisioni di misura attese.

Per quanto riguarda la segmentazione della geometria della parete e la conseguente individuazione dei piani di discontinuità, si è provveduto, come nel caso precedente, a svolgere un'analisi iniziale per definire correttamente i parametri di soglia del metodo. Si è così messa in luce la presenza di quattro zone distinte con caratteristiche in parte differenti; per tale motivo, nella determinazione delle discontinuità condotta con il metodo interattivo, si è preferito analizzare ciascuna singolarmente, per avere maggiore controllo sui risultati: sono stati estratti rispettivamente per ciascuna zona 39, 28, 26 e 33 micro-aree (complessivamente dunque 126 regioni); al tempo stesso, come nel caso precedente, si è verificata l'efficienza di Ransac estraendo solamente 54 macro-aree per analizzare se l'algoritmo è in grado di risalire alle superfici corrette; infine si è analizzato l'intero DSM con l'algoritmo di segmentazione automatica che ha permesso, con gli stessi parametri di soglia all'ultimo livello di elaborazione, di estrarre 120 regioni differenti.

L'unione di tutti i risultati (figura 4.12) porta ad un mascheramento di alcune famiglie che, seppur costituite da pochi piani, risultano interessanti per l'analisi geo-strutturale. E' ovvio che, non avendo proceduto, come nel caso interattivo, ad individuare ed analizzare ciascuna zona separatamente, è inevitabile ottenere tale risultato: in generale è l'operatore che decide come svolgere le analisi e soprattutto quali regioni siano da individuare; se, come in questo caso, si decide di analizzare separatamente alcune porzioni del DSM,

L'unione di tutti i risultati (figura 4.12) porta ad un mascheramento di alcune famiglie che, seppur costituite da pochi piani, risultano interessanti per l'analisi geo-strutturale. E' ovvio che, non avendo proceduto, come nel caso interattivo, ad individuare ed analizzare ciascuna zona separatamente, è inevitabile ottenere tale risultato: in generale è l'operatore che decide come svolgere le analisi e soprattutto quali regioni siano da individuare; se, come in questo caso, si decide di analizzare separatamente alcune porzioni del DSM,

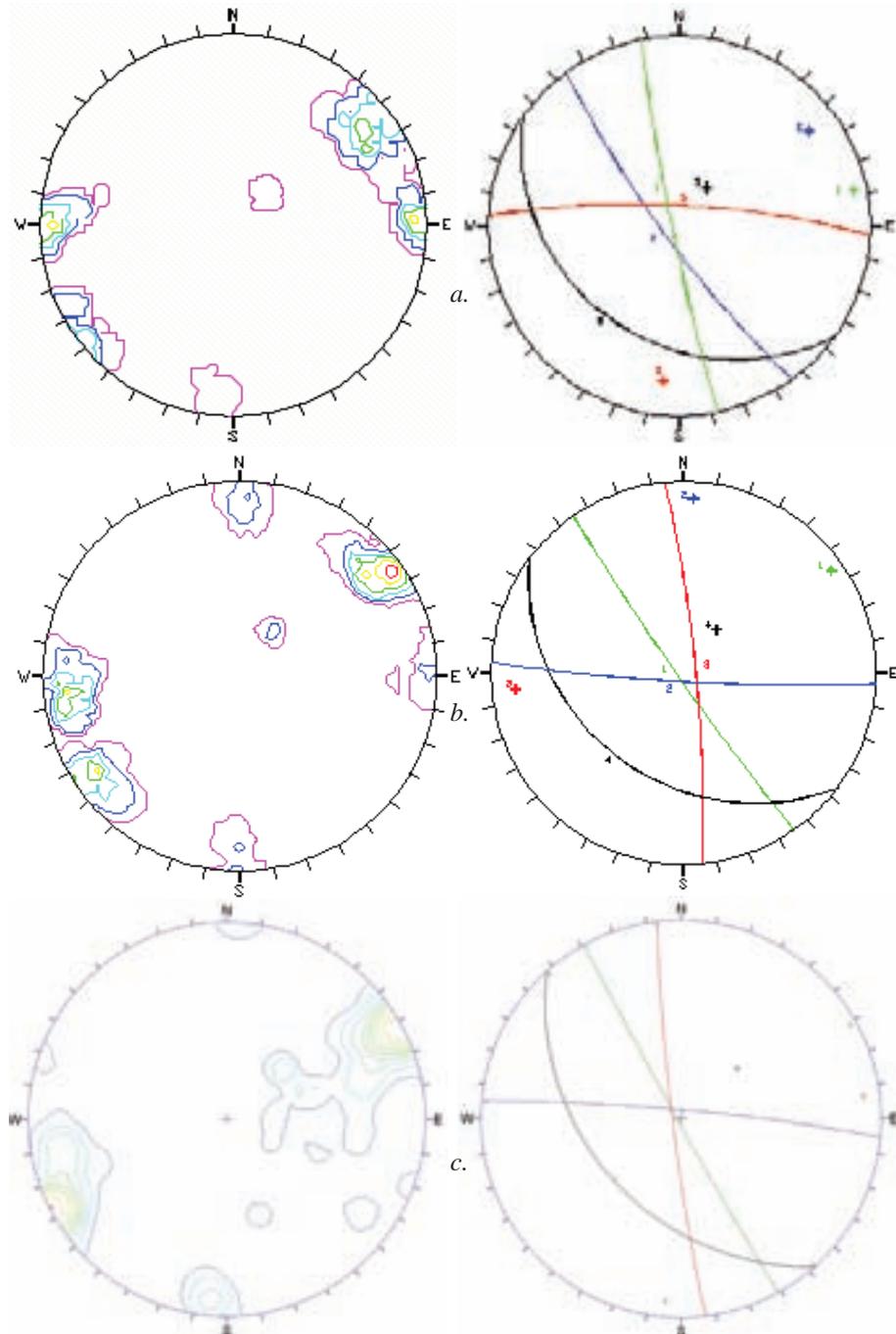


Figura 4.12: Stereogrammi ottenuti dall'analisi del sito di Longeborgne: (a.) rilievo interattivo con micro-zone; (b.) rilievo interattivo con macro-zone; (c.) rilievo automatico.

può accadere che in alcune si vada a individuare un numero maggiore di piani di quelli richiesti o selezionati nelle altre porzioni: ciò, come si è visto, porta ad un possibile mascheramento di alcune informazioni importanti. Il metodo automatico, intervenendo sulla nuvola di punti nella sua interezza, è meno soggetto a questo tipo di problematiche, fermo restando che, se alcune peculiarità statiche della parete sono evidenziate da poche superfici, è solo l'abilità dell'operatore che permette di metterle in luce, nonostante l'analisi numerica e statistica tenda a non riconoscerle.

Come nel caso precedente l'individuazione di micro-zone e macro-zone è sostanzialmente equivalente. È facile osservare che l'estrazione automatica, tende a mettere in luce molti più piani di discontinuità, in quanto va a considerare anche quelli esclusi nella procedura interattiva, pur fornendo sostanzialmente lo stesso risultato in termini di famiglie presenti nell'ammasso; lievissime differenze (2-3 gradi) possono essere riscontrate nel confronto fra le varie figure, ad eccezione soltanto del piano sub-orizzontale che, nella soluzione automatica presenta dip direction di 10° maggiori verso nord: oltre al fatto che più il piano medio tende ad essere orizzontale, più tale parametro perde di importanza, si osserva come la tecnica automatica abbia messo in luce molti piani con dip direction più settentrionale rispetto alle soluzioni interattive.

Al momento, non disponendo di un rilievo tradizionale da usare come controllo, non è possibile stabilire quale delle due soluzioni sia maggiormente corretta: l'elevato grado di somiglianza fra i vari stereogrammi dimostra, comunque, che anche la procedura automatica dimostra una buona affidabilità.

4.5. Sito campione: Granier

Il Monte Granier forma lo sperone Nord del massiccio della Chartreuse e si trova al confine tra i dipartimenti dell'Isère e della Savoia, nel comune di Chambéry, in Francia. Nel novembre del 1248 un crollo, il cui volume è stimato tra i 300 e i 500 milioni di metri cubi, si produsse sulla parete Nord.

Le vittime, secondo la stima di uno storico del XIX secolo (Trèmpier, 1870) furono all'incirca 5000. Il dislivello percorso dalla massa principale del distacco fu di circa 1500 metri. La distanza massima percorsa dai blocchi fu di diversi chilometri, fino a 7.5 km, in gran parte sul piano pressoché orizzontale della valle sottostante.

La massa in movimento, invece di formare un accumulo ai piedi della parete, colò e si allargò estendendosi sui terreni della valle, ricoprendo una superficie totale di 23 km quadrati. Al termine della colata, la forma assunta mostrava un rigonfiamento frontale, morfologia tipica di un accumulo dovuto solamente alla gravità (detriti asciutti). La concomitanza dei fenomeni di distacco e di colata caratterizzarono la catastrofe del Granier. I blocchi localizzati nella parte più a valle della colata, rappresentano, secondo Gogue e Pachoud nel 1972 e successivamente secondo Pachoud nel 1991, soltanto la centesima parte della massa totale franata, ovvero dai 3 ai 5 milioni di metri cubi. Il resto del volume sarebbe costituito soprattutto da pietrame e da materiale fine.



Figura 4.13: Inquadramento cartografico del sito del Granier.

Al giorno d'oggi, il monte Granier forma un bastione di massa calcarea dalle pareti verticali alte circa 300 m che appoggia su livelli marno-calcarei affioranti sulla parete Nord per circa 550 m di altezza e aventi inclinazione di 15° verso Sud-Est. La fessurazione interessa tutta la massa dei calcari, caratterizzati da fratture verticali ben visibili sulla faccia Nord. Tale massiccio è caratterizzato inoltre da numerose gallerie estese anche per parecchi chilometri, provocate dal fenomeno del carsismo, le quali sono state in passato utilizzate dagli speleologi per lo studio dell'ammasso roccioso. D'altra parte sono state riscontrate all'interno della massa rocciosa delle fratture caratterizzate da una notevole apertura ma non di origine carsica, che presentano cioè un allontanamento delle pareti di una discontinuità. L'osservazione di queste fratture, che interessano l'intera massa calcarea, dalla superficie superiore ai livelli marno-calcarei, permette di constatare la decompressione attuale del massiccio, che potrebbe essere dovuta ad un rilassamento della base marnosa, e l'instabilità risultante.

Questa decompressione dell'ammasso roccioso e l'apertura delle fratture possono essere dovute al rilassamento della base marnosa, oltre alla pendenza sfavorevole. Una destabilizzazione improvvisa con conseguente crollo di una massa stimata nell'ordine di 500000 metri cubi potrebbe venire innescata da un aumento eccezionale della pressione idraulica alla base dell'ammasso oppure da una scossa sismica. In tal

caso, la zona che verrebbe interessata dal crollo è di difficile identificazione e richiederebbe uno studio completo (rilievi, simulazioni, modellazioni tridimensionali, ecc...).

A differenza del sito analizzato precedentemente, inizialmente non è stato possibile realizzare un rilievo topografico della parete, e la restituzione è avvenuta in un sistema di riferimento arbitrario. Si è lavorato quindi sulle immagini digitalizzate di un precedente rilievo fotogrammetrico con camera metrica UMK 10 effettuato da elicottero: le dimensioni della parete e la difficoltà di accesso rendono infatti economicamente vantaggioso questo tipo di approccio. La camera (dimensione del formato 160x115 mm) era dotata di un'ottica da 100 mm ed era accompagnata da un certificato di calibrazione per le distorsioni ottiche. I fotogrammi sono stati digitalizzati in un primo momento alla risoluzione di 14 $\mu\text{m}/\text{pixel}$, e in seguito, dal momento che era emerso dagli studi preliminari che si poteva utilizzare un livello di dettaglio minore, alla risoluzione di 20 $\mu\text{m}/\text{pixel}$.

L'orientazione delle immagini, viste le dimensioni ragguardevoli (circa 8000x6000 pixel), è avvenuta manualmente, in quanto il software per la structure and motion presentava problemi di gestione di memoria; inoltre, come è possibile notare in figura 4.14, le basi di presa fra le diverse coppie di fotogrammi utilizzate per generare il modello sono piuttosto lunghe: viene meno quindi una delle ipotesi di base delle tecniche S&M, ovvero il blocco non aveva caratteristiche adatte a garantire risultati accettabili.

Il DSM dell'intera parete (rappresentato in figura 4.15) è stato ottenuto utilizzando quattro diverse coppie di fotogrammi, in assetto pseudonadirale rispetto alla parete. Per ciascuna coppia si è ottenuta la nuvola di punti corrispondente e, al termine del procedimento, le varie nuvole, per mezzo di un software commerciale, sono state allineate ed unite. Il modello complessivo risulta costituito all'incirca da 3.5 milioni di punti con interdistanze medie di circa 10 cm. In un primo

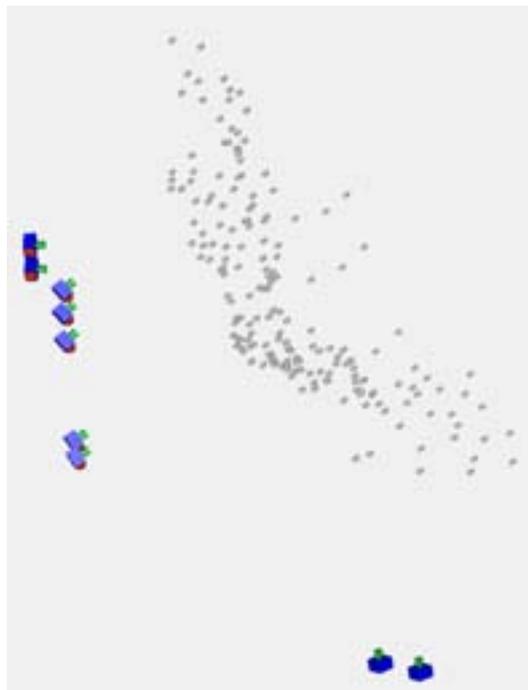


Figura 4.14: Posizioni di presa e punti di legame del blocco realizzato per il sito del Granier.

momento, non disponendo di punti d'appoggio, si è effettuata una georeferenziazione di massima del DSM utilizzando le informazioni desumibili dalla cartografia alla scala 1:25000; mentre l'orientamento al nord del modello (importante per la determinazione dei parametri di giacitura dei piani di discontinuità, è stato ricavato con precisione sufficiente (circa un paio di gradi) per gli scopi dello studio, la determinazione del fattore di scala dei modelli non era sufficientemente affidabile (errori di circa 2÷5 m) per poter estrarre, con il grado di dettaglio richiesto, altre informazioni (spaziatura tra le discontinuità, dimensioni dei blocchi,

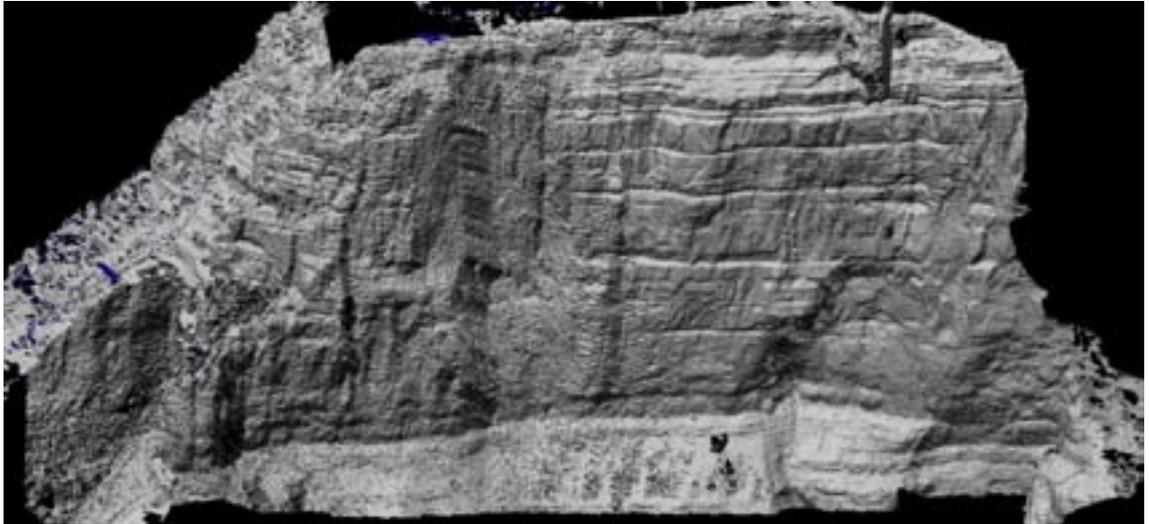


Figura 4.15: DSM della parete nord del complesso del Granier, ottenuto dall'unione di quattro differenti modelli.

etc.).

In seguito, servendosi di quattro punti d'appoggio misurati topograficamente, gentilmente forniti dall'Université Joseph Fourier di Grenoble, è stato possibile georeferenziare il modello, stimando errori di posizionamento dell'ordine di $10\div 20$ cm, più che adeguati alle richieste dei geologi.

Per quanto riguarda lo studio delle discontinuità, si è proceduto come nel caso precedente: una analisi preliminare ha definito le soglie di accettazione per l'estrazione dei piani. In questo caso, essendo molto più ampia la zona da considerare ed assai inferiore la risoluzione delle immagini sull'oggetto, si sono adottati valori di soglia intorno ai $30\div 40$ cm, che hanno permesso di mettere in evidenza i macrosistemi di discontinuità della parete.

Sempre in analogia a quanto fatto negli altri casi, si sono individuate manualmente le famiglie di discontinuità, analizzando separatamente ciascuna nuvola (non si è cioè utilizzato il DSM complessivo) e individuando di volta in volta sia macro-zone che micro-zone. Come mostrato in figura 4.16.a e 4.16.b, anche in questo caso, i risultati con i due metodi sono sostanzialmente equivalenti; dal confronto con la figura 4.16.c, in cui sono rappresentati i risultati ottenuti con il metodo automatico, si nota che il numero di famiglie individuabili è decisamente maggiore: con il metodo interattivo, infatti, viste le dimensioni dell'oggetto da analizzare, si sono estratte informazioni solamente in zone che sembravano significative, e non si è coperta interamente la superficie dell'ammasso. L'esempio mette in mostra due aspetti: da un lato il fatto che il metodo automatico è in grado di estrarre correttamente le stesse informazioni che vengono determinate da un operatore umano; dall'altro il fatto che il metodo interattivo, per essere condotto correttamente, deve sempre analizzare tutta l'estensione della superficie rocciosa. Il limitare l'analisi a zone apparentemente significative, probabilmente, è corretta quando si dispone di un elevato livello di dettaglio sulle immagini,

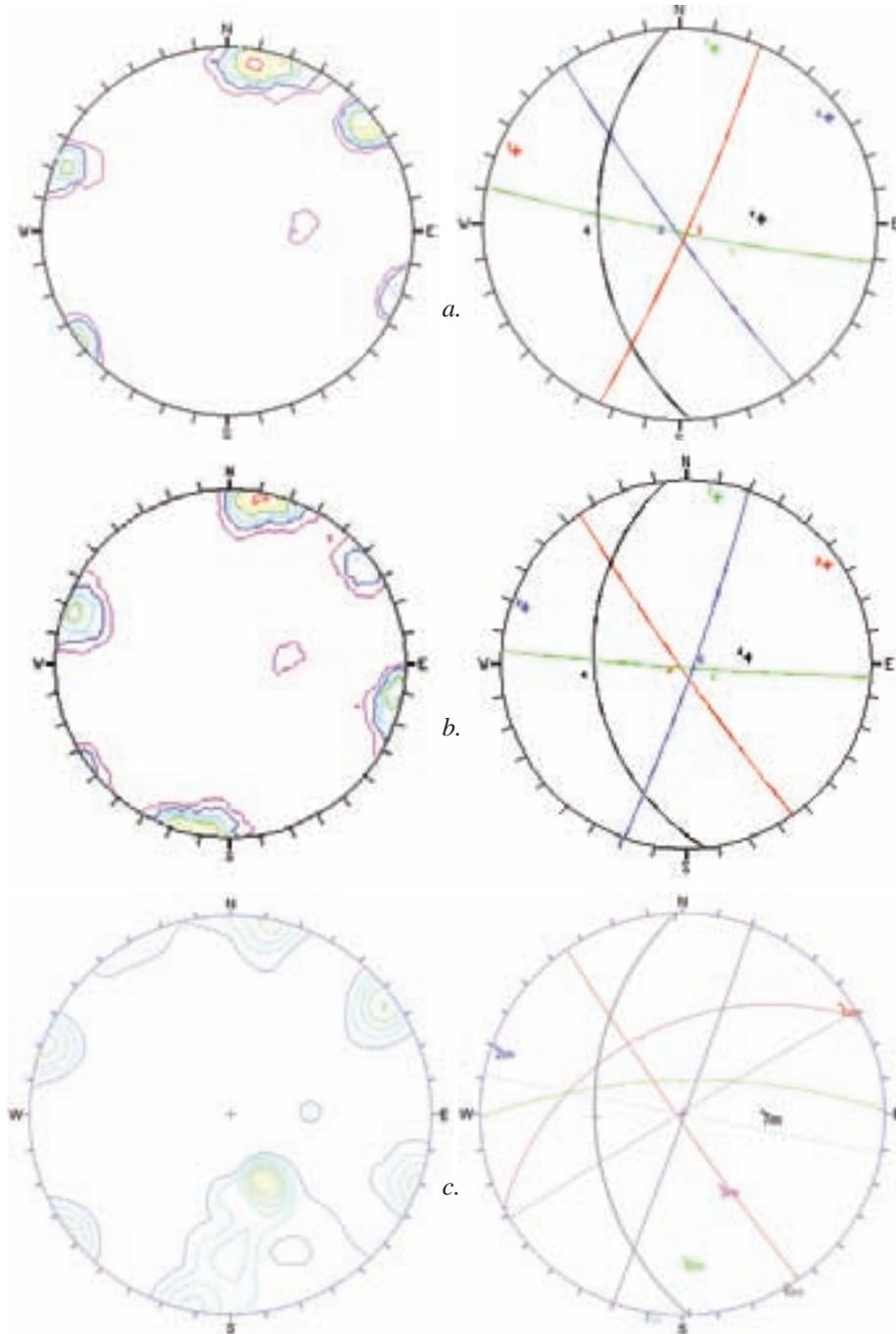


Figura 4.16: Stereogrammi ottenuti dall'analisi del sito del Granier: (a.) rilievo interattivo con micro-zone; (b.) rilievo interattivo con macro-zone; (c.) rilievo automatico.

che permette di avere una visione abbastanza chiara della morfologia della parete, ed, al tempo stesso, non si affrontano casi eccessivamente estesi, in cui un'analisi qualitativa sommaria dell'aspetto della roccia non assicura l'individuazione di tutte le possibili discontinuità significative per l'analisi.

Anche in questo caso, come nel precedente, deviazioni di qualche grado fra una soluzione e l'altra sono da ritenersi del tutto trascurabili.

Sebbene la sperimentazione sia solo all'inizio, e i tre siti analizzati non formino certamente un campione sufficiente per la valutazione delle capacità dei metodi sviluppati, i risultati sono molto incoraggianti.

Conclusioni

Oggetto della presente dissertazione è stato lo sviluppo e l'applicazione di tecniche di automazione in fotogrammetria dei vicini (close-range) con particolare riguardo, in campo geotecnico, al loro impiego nell'analisi di stabilità di pareti rocciose.

La motivazione per un impegno in questo ambito è nata dalla constatazione che, di fronte agli sviluppi tecnologici nel settore del rilevamento, che vedono altre strumentazioni e nuove esigenze affacciarsi, solo un progresso nel senso dell'automazione può tener vivo il patrimonio di conoscenza accumulato nel settore della fotogrammetria. Perché ciò sia possibile occorre realisticamente prendere atto anzitutto delle affinità con il settore della visione artificiale e dell'elaborazione di immagini e riconoscere che concetti nuovi o ripresentati in chiave moderna e assai promettenti per le applicazioni del rilevamento stanno maturando o sono già patrimonio di questi gruppi. Fa riflettere in particolare il fatto che non solo tali metodi ed innovazioni sono, come naturale, nel settore proprio dell'elaborazione delle immagini (in cui la fotogrammetria digitale ha pur dato nel decennio scorso contributi di assoluto rilievo, come il LSM e il MGCM) ma anche e soprattutto sulla geometria adatta per descrivere la presa dei fotogrammi e sui metodi di stima, considerati patrimonio indiscusso dei fotogrammetri e dei geodeti.

Per quanto riguarda la ricostruzione della Structure and Motion, il lavoro svolto e l'esperienza fatta mostrano che l'estrazione ed il matching di features è una fase decisiva per il successo del metodo e che la ricerca si sta orientando verso operatori sempre più sofisticati, con proprietà di invarianza che li rendono molto più robusti rispetto a variazioni del punto di vista. Con tutto ciò, i risultati presentati nel primo capitolo con l'orientamento automatico di una sequenza di Mobile Mapping composta da quasi 100 coppie di immagini dimostrano, al di là dei problemi certamente ancora da risolvere e dei miglioramenti possibili, che la selezione ed il matching di corrispondenze possono affrontare ambienti pochissimo strutturati (nel caso, essenzialmente campi e strada, ovvero tessiture povere o ambigue per definizione).

Non meno importante è il ricorso a tecniche di stima robusta, in particolare RANSAC e le sue varianti, capaci di lavorare con percentuali di errori grossolani (inevitabili) che renderebbero del tutto inutilizzabili

i risultati dei minimi quadrati.

Queste tecniche, per filtrare gli errori, devono tuttavia determinare i parametri di un modello fisico o geometrico che discrimini le osservazioni accettabili dagli errori. Entra qui in gioco la novità dell'approccio della geometria proiettiva, in cui è possibile formulare il tradizionale vincolo di complanarità ed il più complesso legame tra i parametri di orientamento di una terna di immagini in forma lineare, senza quindi richiedere la conoscenza di parametri approssimati. Si tratta di un risultato, come evidente dalla trattazione nel primo capitolo, che non si consegue a poco prezzo né senza una districarsi in una complessità algoritmica che pare talvolta perfino ingiustificata. La potenza del metodo risiede però proprio nel fatto che non è richiesta la conoscenza di alcuna informazione preliminare sulla struttura del blocco o della scena e nella sua generalità, che consente sia applicato nel rilievo architettonico come nell'analisi di sequenze, in campo archeologico, medico, cinematografico etc...

Vi sono tuttavia anche dei limiti al percorso delineato nella geometria proiettiva. Per ritrovare risultati metricamente validi e corretti, occorre infatti tornare al classico bundle adjustment ed alle equazioni di collinearità; imporre o sfruttare condizioni e vincoli su parametri validi in uno spazio non è semplice. Sembra quindi in un certo senso incompleto (e forse lo è, almeno per un fotogrammetra) l'approccio delineato; si tratta di una linea di sviluppo possibile del lavoro che può arricchire questo percorso, ad esempio studiando il bundle adjustment nello spazio proiettivo e le sue proprietà.

Se in passato la ricostruzione di superficie era svolta per curve di livello e punti quotati con restituzione manuale, oggi questa operazione è svolta dal calcolatore in modo estremamente più rapido (e, con gli opportuni distinguo, preciso) grazie allo sviluppo ormai quasi ventennale di algoritmi di matching sub-pixel. Una scommessa di questo lavoro, senza pretendere di aver dato una risposta definitiva, è stata quella di puntare sul matching vincolato geometricamente. Idea estremamente illuminata e concettualmente brillante, non ha a nostro avviso finora trovato la diffusione che merita, perchè unisce in modo semplice la precisione di localizzazione con la congruenza geometrica. Non tutti gli aspetti della sua gestione, in particolare i pesi, sono ancora ben assestati, ma il potenziale, specie nel close range, è certamente molto elevato. Per sfruttarlo, il matching denso richiede di essere però programmato con estrema attenzione sia nel calcolo dei parametri approssimati, sia nella correzione delle distorsioni, sia nel ricampionamento epipolare delle immagini. Anche qui, spazi di miglioramento sono possibili, a fronte della grande varietà di situazioni che la geometria di presa e la forma dell'oggetto presentano, rendendo di volta in volta assai utile questo o quell'accorgimento. Il fatto che, in un ambiente non semplice come quello delle pareti in roccia i risultati siano del tutto confrontabili con quelli di un laser a scansione terrestre, come mostrato nel capitolo 2, mostrano che il metodo è solido.

La sfida più aperta è forse quella relativa all'ultima delle fasi del procedimento, ovvero la classificazione dei dati: un tempo lasciati all'interpretazione dell'operatore, oggi devono per quanto possibile essere etichettati da un algoritmo. Si tratta di un settore per certi aspetti nuovo, per certi altri di una costola del-

l'immagine processing, tuttavia in forte espansione. Anche se i contributi più innovativi di questo lavoro sono stati raggiunti proprio in questo settore, la presentazione dei metodi è stata ristretta a quelli adottati o in fase di implementazione, ovvero senza coprire che in minima parte la varietà delle tecniche, spesso application-oriented, per la segmentazione. I risultati in buon accordo con quelli tradizionali ottenuti nei siti test col metodo automatico per la determinazione di piani di discontinuità messo a punto, sebbene la sperimentazione non possa ancora dirsi conclusa, indicano che la metodologia sviluppata è certamente promettente.

Appendice A

A.1. Cenni di geometria proiettiva

Lo studio delle proiezioni geometriche e il tentativo di definirne gli aspetti in un quadro teorico coerente ed unitario, risale agli inizi del Rinascimento con gli studi condotti dai due architetti più importanti dell'epoca: Filippo Brunelleschi (1377-1446) e Leon Battista Alberti (1404-1472). In particolare quest'ultimo si dedicò alla elaborazione di regole teoriche che permettessero un approccio metodologico al problema della geometria proiettiva. Ciononostante la matematica del Rinascimento non aveva ancora sviluppato i presupposti per consentire una trattazione analitica e gli studi dell'Alberti, riassunti nella sua opera "*De pictura*", riguardano principalmente la geometria descrittiva ed il disegno. Molti altri illustri artisti (ricordiamo Leonardo da Vinci, Piero della Francesca, Albrecht Durer) nel corso dei secoli hanno studiato i processi di formazione di un'immagine e le regole con le quali percepiamo il mondo che ci circonda; bisogna però aspettare il XVII secolo con gli studi sulle coniche di Desargues (1593-1662) prima che le proiezioni facciano il loro ingresso nel mondo della matematica (il periodo coincide con la nascita della geometria analitica ad opera di Renè Descartes (1596-1650)). Il primo passo verso uno studio organico della geometria proiettiva viene però intrapreso solo nel 1822 da Jean-Victor Poncelet (1788-1867) con la sua opera "*Traité des propriétés projectives des figures*". E' proprio nei primi trent'anni dell'800 che la matematica e la fisica entrano nella loro era moderna abbattendo assiomi che duravano da più di due millenni e ponendo le basi per la rivoluzione scientifica dei successivi due secoli. Ricordiamo che Poncelet poteva vantare contemporanei del calibro di Johann Gauss (1777-1855), Sadi Carnot (1796-1832), Friedrich Bessel (1784-1846) e Nikolai Lobachevsky (1792-1856) l'inventore delle geometrie non euclidee, per non parlare di molti altri illustri fisici e matematici come Poisson, Green, Mobius, Jacobi, Legendre, Navier.

Nel seguito vengono richiamati concetti e nozioni fondamentali della geometria proiettiva, legati alle tecniche sviluppate nella nostra ricerca. Tentare di delineare un quadro bibliografico, in virtù delle numerosissime opere realizzate sull'argomento negli ultimi due secoli (sarebbero fin troppe quelle realizzate negli

ultimi dieci anni) è molto difficile; sia allora sufficiente ricordare in questa sede solo alcuni testi di carattere generale: [Springer 1964], [Semple 1979], [Mundy 1992], [Faugeras 1993].

A.1.1. Spazi vettoriali proiettivi

Per introdurre gli spazi vettoriali proiettivi partiamo con un esempio: consideriamo innanzitutto lo spazio vettoriale $\mathbb{R}^2 \setminus \{0\}$ (ovvero il piano cartesiano privato dell'origine); due vettori di tale spazio si dicono *collineari* se appartengono alla stessa retta per l'origine, ovvero se sono uno multiplo dell'altro attraverso uno scalare non nullo, o ancora se esiste uno scalare non nullo λ tale che le componenti di un vettore siano uguali al prodotto delle componenti dell'altro scalate di λ . E' facile dimostrare che tale relazione, che indicheremo con il simbolo \sim , è una relazione di equivalenza, in quanto è riflessiva, simmetrica e transitiva:

$$X \sim Y \iff Y = \lambda X, \quad \text{con } \lambda \in \mathbb{R} \setminus \{0\} \quad (\text{A.1})$$

$$\text{Dove } X = \begin{pmatrix} x_0 \\ x_1 \end{pmatrix} \text{ e } Y = \begin{pmatrix} y_0 \\ y_1 \end{pmatrix}.$$

Possiamo allora definire la retta proiettiva \mathbb{P}^1 come l'insieme quoziente (ossia l'insieme delle classi di equivalenza) dei vettori di $\mathbb{R}^2 \setminus \{0\}$ rispetto alla relazione di collinearità.

E' facile osservare che, essendo \mathbb{P}^1 un sottospazio di $\mathbb{R}^2 \setminus \{0\}$, chiuso rispetto alla somma ed alla moltiplicazione per uno scalare, esso è a sua volta uno spazio vettoriale. Si nota inoltre che lo spazio proiettivo ha dimensione inferiore di uno rispetto al corrispondente spazio vettoriale che lo ha generato. Al tempo stesso si può definire la proiezione canonica $\pi: \mathbb{R}^2 \setminus \{0\} \longrightarrow \mathbb{P}^1$ come l'applicazione che ad un vettore X associa la sua direzione. L'immagine di X risulterà $\pi(X) = (x_0 : x_1)$ in cui ":" sottolinea che, per la classe di equivalenza, quello che conta è il rapporto fra x_0 e x_1 . Tali componenti prendono il nome di coordinate omogenee dello spazio proiettivo. Per quanto visto, punti che presentano coordinate omogenee proporzionali, appartengono alla stessa classe di equivalenza e rappresentano dunque il medesimo elemento della retta proiettiva.

Si può quindi interpretare geometricamente lo spazio \mathbb{P}^1 come l'insieme dei punti generati dall'intersezione fra la stella di rette di \mathbb{R}^2 con centro nell'origine ed una qualsiasi retta non passante per l'origine stessa (figu-

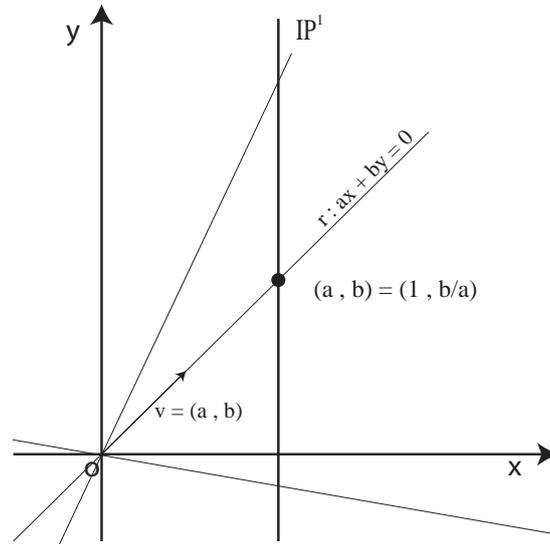


Figura A.1: Rappresentazione geometrica della retta proiettiva \mathbb{P}^1 .

ra A.1). Si consideri per semplicità che tale retta coincida con la retta di equazione $r : x = 1$: lo spazio \mathbb{P}^1 oltre a contenere tutti i punti di intersezione con il fascio di rette per l'origine (di coordinate in $\mathbb{R}^2 (1, x_1/x_0)$) contiene anche il punto corrispondente alla retta della stella coincidente con l'asse y $(0, k)$ che non genera alcuna intersezione nello spazio reale. In altre parole lo spazio proiettivo rappresenta un completamento dello spazio affine nel senso che l'operazione di "omogeneizzazione" delle coordinate di uno spazio di dimensione n (nel nostro caso $n = 1$) corrisponde ad immergere tale spazio in uno spazio vettoriale $n+1$ dimensionale e considerare, in luogo dei punti, i sottospazi vettoriali di dimensione 1 entro tale spazio, rappresentati dalle classi di equivalenza.

Generalizzando i concetti finora esposti possiamo dare la seguente:

Definizione A1: Sia V uno spazio vettoriale di dimensione finita sul campo C . Si definisce Spazio Proiettivo su V , in simboli $\mathbb{P}(V)$, l'insieme dei sottospazi vettoriali di V . Se W è un sottospazio vettoriale di V , indicheremo con σW il corrispondente elemento dello spazio proiettivo $\mathbb{P}(V)$, e chiameremo dimensione dell'elemento σW il numero intero

$$\dim \sigma W := \dim_c W - 1$$

Analogamente la dimensione dello spazio proiettivo sarà definita come la dimensione del corpo che viene utilizzato per definire lo spazio meno uno.

A questo punto non è difficile considerare spazi proiettivi di dimensione maggiore di 1: lo spazio proiettivo \mathbb{P}^2 può essere pensato generato dalle classi di equivalenza definite dall'insieme dei vettori dello spazio $\mathbb{R}^3 \setminus \{0\}$, ovvero generato dall'intersezione della stella di rette di \mathbb{R}^3 con centro nell'origine con un generico piano non passante per $(0, 0, 0)$. Anche in questo caso le rette che giacciono su un piano parallelo al piano di proiezione scelto non hanno un'intersezione reale nello spazio euclideo ma vanno a definire un sottospazio (di dimensione 1) dello spazio vettoriale \mathbb{P}^2 che prende il nome di retta all'infinito o retta ideale. Tale retta contiene tutti i punti all'infinito del piano che corrispondono ciascuno ad una retta di \mathbb{R}^3 senza intersezione con il piano di proiezione.

Proseguendo con l'analogia geometrica appena descritta, possiamo considerare le rette del piano proiettivo, come le rette generate dall'intersezione fra la stella di piani con centro nell'origine ed un generico piano non passante per tale punto. L'equazione del generico piano appartenente alla stella risulta:

$$ax + by + cz = 0 \tag{A.2}$$

Un altro modo particolarmente efficace per visualizzare lo spazio proiettivo \mathbb{P}^2 è quello di considerare la cosiddetta sfera unitaria: dal momento che le coordinate omogenee sono definite a meno di un fattore di proporzionalità, possiamo eliminare un grado di libertà nella definizione di tali coordinate ammettendo, ad esempio, che la loro norma sia uguale ad 1; in altre parole si può sempre considerare il vettore di coordinate omogenee $\frac{1}{\sqrt{x_0^2 + x_1^2 + x_2^2}}(x_0, x_1, x_2)$; tali punti descrivono in \mathbb{R}^3 la superficie di una sfera con centro

nell'origine e raggio unitario (vedi figura A.2).

In altre parole lo spazio proiettivo \mathbb{IP}^2 è topologicamente equivalente ad una sfera: su tale sfera è possibile individuare tutte le primitive di cui abbiamo finora parlato e condurre operazioni di tipo geometrico. La sfera unitaria, o sfera di Gauss, è infatti molto usata, soprattutto laddove le coordinate dei punti tendono all'infinito (ovvero il punto tende a divenire improprio): in tal caso è possibile sostituire al piano proiettivo che ha estensione illimitata, una superficie, quella sferica appunto, limitata.

Anche per quanto riguarda l'equazione (A.2) è interessante notare che terne di parametri proporzionali

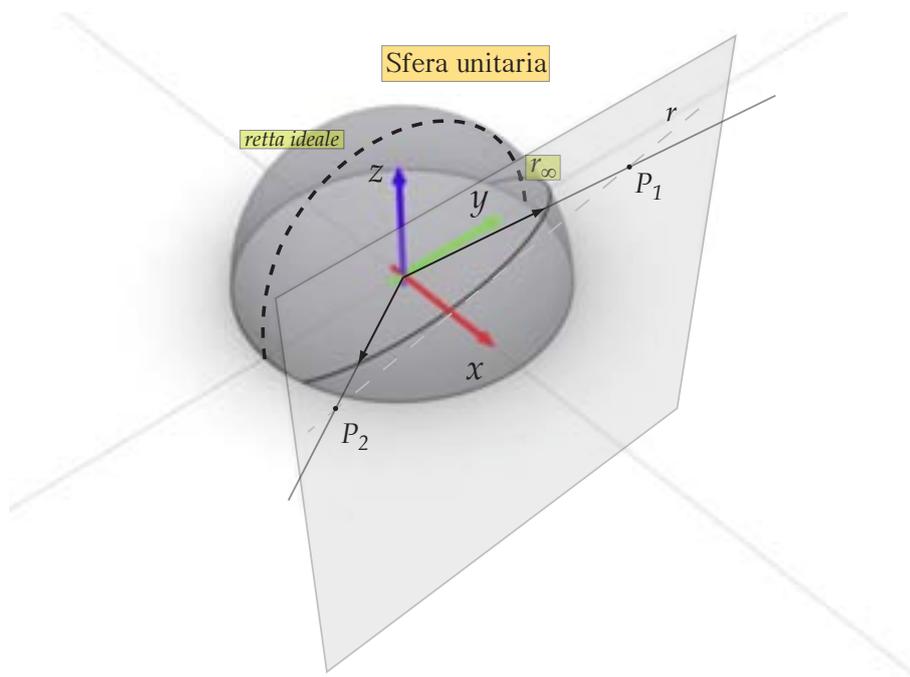


Figura A.2: Rappresentazione della sfera unitaria (viene per comodità rappresentato solo l'emisfero superiore). E' possibile notare come rette nel piano proiettivo individuino circonferenze di diametro massimo sulla sfera; l'intersezione di tale circonferenza con il meridiano corrispondente alla retta ideale coincide con il punto improprio della retta stessa.

individuano piani coincidenti. In altri termini è possibile anche in questo caso andare a definire una relazione di equivalenza (non più di *collinearità* bensì di *complanarità*) i cui elementi (ovvero le classi di equivalenza) assumono il significato di rette dello spazio proiettivo. Non deve a questo punto stupire il fatto che lo stesso tipo di notazione (cioè una terna di numeri reali) possa essere usata nel piano \mathbb{IP}^2 sia per riferirsi ad un punto sia per riferirsi ad una retta; si può infatti dimostrare che la similitudine fra le due primitive geometriche non è solamente una similitudine a livello analitico ma anche a livello concettuale. Si può infatti dimostrare il principio di dualità della geometria proiettiva, in base al quale a qualsiasi teorema relativo alla geometria proiettiva bidimensionale corrisponde un teorema duale che può essere derivato scambiando i

ruoli che punti e rette hanno nel teorema originale.

Il fatto di avere due diverse primitive geometriche individuate dallo stesso tipo di notazione algebrica (ovvero una terna di numeri reali) sebbene possa generare confusione, porta proprio a ribadire il legame di dualità tra punti e rette: è di volta in volta l'utilizzatore a comprendere se la terna di coordinate omogenee che sta utilizzando ha, per lui, il significato di punto o di retta del piano proiettivo.

Analizziamo allora il comportamento di punti e rette nello spazio proiettivo \mathbb{P}^2 : in geometria euclidea, utilizzando un sistema di riferimento cartesiano, si è soliti valutare la condizione di appartenenza di un punto ad una retta per mezzo di un'equazione lineare:

$$ax + by + c = 0 \quad (\text{A.3})$$

Nel piano proiettivo, al contrario, l'appartenenza di un punto ad una retta può essere interpretato come la complanarità fra il vettore (o la retta) di \mathbb{R}^3 cui corrisponde il punto ed il piano cui corrisponde la retta. Essendo il vettore l , rappresentato dalla terna di parametri (a, b, c) dell'equazione (A.2), parallelo alla normale a tale piano, la condizione può essere espressa, indicando con x il vettore relativo al punto, come:

$$x \cdot l = x^T l = 0 \quad (\text{A.4})$$

Ci accorgiamo a questo punto che, svolgendo il prodotto scalare, e considerando le già più volte sottolineate proprietà di equivalenza dei punti di \mathbb{P}^2 , le equazioni (A.3) e (A.4) sono formalmente identiche. Ciò ci porta a concludere che il piano proiettivo (o più in generale uno spazio proiettivo n -dimensionale) è, dal punto di vista geometrico, formalmente identico ad uno spazio geometrico euclideo di ugual dimensione, con il vantaggio che l'operazione di omogeneizzazione delle coordinate permette il completamento dello spazio con gli elementi geometrici all'infinito.

L'estensione agli spazi proiettivi di dimensione maggiore di due segue di fatto le linee precedentemente tratteggiate: in particolare lo spazio proiettivo \mathbb{P}^3 sarà costituito da elementi caratterizzati da una quaterna di componenti omogenee, che rappresenteranno i punti dello spazio o le loro entità duali che in questo caso risultano piani dello spazio. I risultati che discuteremo nel seguito potranno dunque essere riferiti a queste due diverse primitive geometriche.

A.1.2. Struttura degli spazi proiettivi

In virtù di quanto visto nel precedente paragrafo, si vadano a definire alcune operazioni fondamentali nello spazio proiettivo, cominciando a considerare l'operazione di intersezione fra due rette. Richiamando il risultato ottenuto per mezzo dell'equazione (A.4) e rammentando la proprietà commutativa del prodotto vettoriale misto, è possibile determinare il punto di intersezione x di due rette l ed l' per mezzo di:

$$x = l \times l' \quad (\text{A.5})$$

Sfruttando inoltre il principio di dualità in analogia alla precedente equazione è possibile determinare la retta l passante per due punti x ed x' :

$$l = x \times x' \quad (\text{A.6})$$

Se le due rette nell'equazione (A.5) sono parallele, ovvero hanno le prime due componenti proporzionali, il prodotto vettoriale fornisce un punto x caratterizzato dall'aver l'ultima componente uguale a zero ($x = (a, b, 0)$). Tale punto, come già visto, è un punto improprio (o ideale o all'infinito), nel senso che non può essere immaginato nello spazio bidimensionale finito. Si è dunque mostrato come nello spazio proiettivo il quinto postulato di Euclide (o postulato delle parallele), che, come sappiamo, ha sempre creato non pochi problemi ai matematici fin dalla sua formulazione, di fatto non valga: due rette parallele si intersecano sorprendentemente in un punto, sebbene tale punto debba essere immaginato all'infinito. Si vedrà fra breve che, pur essendo un'entità concettualmente astratta, un punto ideale (così come una retta ideale o qualsiasi altro elemento ideale della geometria proiettiva) assume un significato di grande concretezza nelle operazioni proiettive: quando, ad esempio, il piano proiettivo è soggetto ad un qualche tipo di trasformazione, può accadere che punti ideali si trasformino in punti non ideali, cioè punti al finito.

La retta che contiene tutti i punti ideali prende il nome di retta all'infinito o retta ideale: dall'equazione (A.4) e dalla forma generica che assumono i punti all'infinito, tale retta deve necessariamente essere:

$$l_{\infty} = (0, 0, 1) \quad (\text{A.7})$$

I risultati finora ottenuti, ed in particolare le equazioni (A.5) ed (A.6) non trovano facilmente riscontro nello spazio proiettivo \mathbb{IP}^3 : in questo caso, infatti, le entità duali dello spazio sono il punto e il piano; mentre continua a valere la relazione di appartenenza (A.4) non è possibile sfruttare il prodotto vettoriale misto per determinare il piano passante per tre punti (o dualmente l'intersezione fra tre piani). Sfruttando la già citata (A.4) è possibile individuare un sistema di equazioni che in forma matriciale assume la forma:

$$\begin{bmatrix} x_1^T \\ x_2^T \\ x_3^T \end{bmatrix} \pi = 0 \quad (\text{A.8})$$

Ovvero il piano π rappresenta lo spazio nullo (monodimensionale nel caso in cui il rango del sistema sia uguale a tre) del sistema omogeneo in (A.8). In questo caso, qualora i tre punti fossero allineati (cioè collineari), il sistema ammette ∞^2 soluzioni che rappresentano i punti del fascio di piani con generatrice la retta passante per i tre punti; se i tre punti, infine, sono coincidenti, il sistema ammette ∞^3 soluzioni corrispondenti ai piani della stella con centro nel punto.

Come di consueto, per mezzo del principio di dualità è possibile determinare l'intersezione di tre piani

per mezzo del sistema omogeneo:

$$\begin{bmatrix} \boldsymbol{\pi}_1^T \\ \boldsymbol{\pi}_2^T \\ \boldsymbol{\pi}_3^T \end{bmatrix} \mathbf{x} = 0 \quad (\text{A.9})$$

In questo caso, qualora due dei tre piani fossero paralleli ma non coincidenti si ottiene un punto ideale corrispondente alla direzione della retta intersezione fra due dei piani non paralleli che compongono il sistema. Se tutti i piani risultano paralleli non coincidenti il rango della matrice si abbassa a 2 e si ottengono ∞^2 soluzioni corrispondenti a ∞^1 punti ideali distinti e allineati che compongono la retta ideale corrispondente alla giacitura comune ai tre piani.

Analoghe considerazioni possono essere fatte nel caso in cui due dei tre piani (o tutti e tre) sono coincidenti.

Infine, in analogia a quanto fatto per il piano proiettivo, si determina l'equazione del piano ideale nello spazio proiettivo \mathbb{IP}^3 come:

$$\boldsymbol{\pi}_\infty = (0, 0, 0, 1) \quad (\text{A.10})$$

A.1.3. Coniche nello spazio proiettivo

Avendo dimostrato che un generico spazio proiettivo n -dimensionale è uno spazio vettoriale definito sul campo \mathbb{R}^{n+1} in cui è definita la somma vettoriale e il prodotto scalare, anch'esso permette la definizione di applicazioni bilineari e di forme quadratiche. Possiamo allora dare le seguenti definizioni:

Definizione A2: Una quadrica in $\mathbb{IP}^n(\mathbb{R})$ è una forma quadratica a coefficienti reali, a meno di un fattore di proporzionalità diverso da zero. Lo spazio nullo di tale forma quadratica, costituito da punti di \mathbb{IP}^n è detto supporto della quadrica.

Definizione A3: Una quadrica di \mathbb{IP}^2 prende il nome di *conica*.

Si vedrà fra breve come le coniche abbiano un ruolo fondamentale per la comprensione delle proprietà dello spazio proiettivo.

Una conica nello spazio euclideo \mathbb{R}^2 è individuata da un'equazione di secondo grado in coordinate cartesiane:

$$ax^2 + bxy + cy^2 + dx + ey + f = 0 \quad (\text{A.11})$$

Nel piano \mathbb{IP}^2 è possibile ottenere una corrispondente equazione, in questo caso di tipo omogeneo:

$$ax_1^2 + bx_1x_2 + cx_2^2 + dx_1x_3 + ex_2x_3 + fx_3^2 = 0 \quad (\text{A.12})$$

a cui è possibile associare una forma bilineare simmetrica e quindi una forma quadratica. Tale forma qua-

dratica in notazione matriciale assume la forma:

$$\mathbf{x}^T \mathbf{C} \mathbf{x} = 0 \quad (\text{A.13})$$

in cui gli elementi della matrice C , con riferimento all'equazione (A.12), risultano

$$C = \begin{vmatrix} a & b/2 & d/2 \\ b/2 & c & e/2 \\ d/2 & e/2 & f \end{vmatrix} \quad (\text{A.14})$$

In tal senso una conica nel piano proiettivo è topologicamente equivalente ad una conica nello spazio reale. Dalla (A.14) deriva che, per definire compiutamente una conica, in virtù del fatto che essa è definita a meno di una costante di proporzionalità, è sufficiente la conoscenza di almeno cinque punti che le appartengano. Inoltre, per mezzo della (A.4), è semplice dimostrare che la tangente alla conica C in un suo punto \mathbf{x} , risulta:

$$\mathbf{l} = \mathbf{C}\mathbf{x} \quad (\text{A.15})$$

Grazie al principio di dualità si può definire una conica per mezzo delle rette ad essa tangenti; la conica in questo caso prende il nome di conica involuppo o conica duale:

$$\mathbf{l}^T \mathbf{C} \mathbf{l} = 0 \quad (\text{A.16})$$

Sempre per mezzo del principio di dualità il punto di tangenza alla conica duale C^* sulla retta tangente \mathbf{l} risulta:

$$\mathbf{x} = \mathbf{C}^* \mathbf{l} \quad (\text{A.17})$$

Ma il principio di dualità ci permette anche di approfondire maggiormente il rapporto che intercorre fra coniche, rette e punti: dato infatti un punto generico \mathbf{x} ed una conica C è possibile definire la retta polare \mathbf{l} di \mathbf{x} rispetto a C come:

$$\mathbf{l} = \mathbf{C}\mathbf{x} \quad (\text{A.18})$$

La retta polare \mathbf{l} gode della proprietà che interseca la conica C nei due punti di tangenza delle rette tangenti condotte da \mathbf{x} . Nel caso in cui \mathbf{x} appartenga alla conica, come abbiamo già avuto modo di vedere (equazione (A.15)), la polare \mathbf{l} coincide con la tangente e il punto di tangenza prende il nome di punto doppio.

Per quanto precede è possibile quindi dare la seguente

Definizione A4: Due punti \mathbf{x} e \mathbf{y} dello spazio proiettivo \mathbb{IP}^2 si dicono coniugati rispetto alla conica C se e solo se:

$$\mathbf{x}^T C \mathbf{y} = 0 \quad (\text{A.19})$$

analogamente due rette l ed m dello spazio proiettivo \mathbb{P}^2 si dicono coniugate rispetto alla conica duale C^* se e solo se:

$$l^T C^* m = 0 \quad (\text{A.20})$$

Dall'equazione (A.18) è semplice verificare che, se due punti (o due rette) sono coniugati, ciò equivale a dire che uno dei due giace sulla retta (o punto) polare dell'altro.

Passando ad uno spazio proiettivo di dimensione maggiore, possiamo dare la seguente definizione:

Definizione A5: Una forme quadratica di \mathbb{P}^3 prende il nome di *quadrica*.

I concetti illustrati possono essere facilmente estesi a questa tipologia di forme quadratiche.

A.1.4. Applicazioni lineari e proiettività

Definizione A6: Dicesi *proiettività* $h(x)$ una applicazione invertibile da \mathbb{P}^2 in \mathbb{P}^2 tale che, dati tre punti allineati x_1, x_2 e x_3 , anche le loro immagini $h(x_1), h(x_2)$ e $h(x_3)$ risultano allineate.

Queste trasformazioni, che prendono anche il nome di *collinearità* o *omografie*, formano un'importante famiglia di trasformazioni geometriche della quale fanno parte anche le traslazioni, le rotazioni e le trasformazioni affini. Si può dimostrare che in un generico spazio proiettivo \mathbb{P}^n tali trasformazioni sono sempre rappresentate da una applicazione lineare. Consideriamo per semplicità il caso del piano proiettivo \mathbb{P}^2 , sia H la matrice associata a tale applicazione ed l la retta passante per i tre punti per x_1, x_2 e x_3 ovvero tale che (equazione A.4):

$$l^T x_i = 0, \quad \forall i = 1, \dots, 3 \quad (\text{A.21})$$

E' semplice allora verificare che:

$$l^T H^{-1} H x_i = 0, \quad \forall i = 1, \dots, 3 \quad (\text{A.22})$$

e che, di conseguenza le immagini dei tre punti risultano allineate, in virtù di (A.4) sulla retta

$$l' = H^{-T} l \quad (\text{A.23})$$

L'equazione (A.23) mostra al tempo stesso in che modo, analiticamente parlando, è possibile trasformare una retta del piano per mezzo di una proiettività. Anche nel caso delle omografie l'applicazione lineare è definita a meno di un fattore moltiplicativo; pertanto, dei 9 elementi della matrice H , solo 8 sono linearmente indipendenti e quindi necessari per la definizione di una generica trasformazione. Come per punti e rette, la collinearità può essere applicata anche alle coniche; in tal caso si ottiene:

$$\begin{aligned} 0 &= \mathbf{x}^T C \mathbf{x} = \mathbf{x}'^T [H^{-1}]^T C H^{-1} \mathbf{x}' \\ &\Rightarrow C' = H^{-T} C H^{-1} \end{aligned} \quad (\text{A.24})$$

mentre per la conica duale vale:

$$C'' = H C^* H^T \quad (\text{A.25})$$

Le proiettività possibili comprendono famiglie di trasformazioni differenti, che possono essere classificate basandosi su criteri di invarianza di determinate proprietà [Klein 1939]. Una prima suddivisione in classi può essere fatta fra trasformazioni che preservano la forma, e si parlerà allora di trasformazioni conformi (traslazioni, rotazioni, similitudini) e quelle che non la preservano, e si parlerà allora di trasformazioni non conformi (trasformazioni affini, prospettività). Lo studio di tali trasformazioni e delle loro proprietà è un tema centrale di tutta la fotogrammetria e più in generale di tutte le scienze geomatiche; si è visto nel capitolo 1 come nella structure and motion uno dei problemi maggiori sia ricavare una trasformazione che riporti la restituzione da uno spazio proiettivo generico ad uno spazio euclideo. Vengono pertanto di seguito brevemente descritte le varie famiglie di trasformazioni.

A.1.4.1. Isometrie

Le isometrie sono trasformazioni che preservano la distanza euclidea fra i punti. La loro forma generica è data da:

$$H = \begin{vmatrix} \varepsilon \cos \theta & -\sin \theta & t_x \\ \varepsilon \sin \theta & \cos \theta & t_y \\ 0 & 0 & 1 \end{vmatrix} \quad (\text{A.26})$$

in cui ε è un coefficiente che può valere ± 1 : nel caso in cui $\varepsilon = +1$ l'isometria conserva l'orientamento, nel caso contrario $\varepsilon = -1$ la trasformazione attua una riflessione. Una sottoclasse delle isometrie è la classe delle rototraslazioni. E' facile notare che per definire un'isometria (a meno del coefficiente di riflessione ε) sono richiesti 3 parametri (2 per la traslazione e 1 per la rotazione).

A.1.4.2. Similitudini

Le similitudini sono trasformazioni che pur conservando la forma non sono invarianti rispetto alla distanza euclidea fra punti. In altre parole una similitudine è la composizione di un'isometria seguita da una variazione di scala isotropa. La loro forma generale è:

$$H = \begin{vmatrix} \varepsilon s \cos \theta & -s \sin \theta & t_x \\ \varepsilon s \sin \theta & s \cos \theta & t_y \\ 0 & 0 & 1 \end{vmatrix} \quad (\text{A.27})$$

Un importante sottoinsieme delle similitudini è rappresentato dalle rototraslazioni con variazione di scala. Per definire una generica similitudine occorrono 4 parametri (2 per la traslazione, 1 per la rotazione, e 1 per il fattore di scala).

A.1.4.3. Trasformazioni affini

Una trasformazione affine è la più generica trasformazione lineare delle coordinate non omogenee che descrivono un oggetto: essa può essere interpretata come la composizione di una rotazione, una scalatura anisotropa ed una traslazione. In coordinate omogenee essa è espressa da:

$$H = \begin{vmatrix} a_{11} & a_{12} & t_x \\ a_{21} & a_{22} & t_y \\ 0 & 0 & 1 \end{vmatrix} \quad (\text{A.28})$$

Una trasformazione affine può anche essere pensata come un'iniziale rotazione di un angolo ϕ pari a quello presente fra la direzione principale di variazione di scala e gli assi coordinati, una trasformazione di scala anisotropa di questi ultimi, una rotazione di $-\phi$, una rotazione finale θ ed una traslazione; in termini analitici il minore di ordine due nella (A.28) è ottenuto come:

$$A = \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} = R(\theta) \cdot R(-\phi) \cdot D \cdot R(\phi) \quad (\text{A.29})$$

in cui la matrice D è una matrice diagonale che rappresenta la trasformazione anisotropa di scala:

$$D = \begin{vmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{vmatrix} \quad (\text{A.30})$$

Risulta evidente, dall'analisi delle (A.28), (A.29) ed (A.30) che complessivamente una trasformazione affine presenta sei gradi di libertà. E' abbastanza semplice verificare che una trasformazione affine gode della proprietà di mantenere il parallelismo fra rette (se si trasformano due rette parallele, la loro intersezione rimane all'infinito) e mantiene il rapporto fra le lunghezze e le aree (in quanto la trasformazione di scala agisce dipendentemente solo dalla direzione).

A.1.4.4. Omografia

Un'omografia, detta anche trasformazione proiettiva, rappresenta la più generica trasformazione lineare che si possa applicare in coordinate omogenee: la sua rappresentazione matriciale assume la forma

$$H = \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} = \begin{vmatrix} A & \mathbf{t} \\ \mathbf{v}^T & v \end{vmatrix} \quad (\text{A.31})$$

I gradi di libertà sono pari ad otto, essendo H definita a meno di una costante di proporzionalità. Si ricava poi facilmente dalla (A.31) la classica formulazione non omogenea ([Krauss 1982]) della proiezione centrale di un piano:

$$x = \frac{a_{11}\xi + a_{12}\eta + a_{13}}{a_{31}\xi + a_{32}\eta + a_{33}} \quad y = \frac{a_{21}\xi + a_{22}\eta + a_{23}}{a_{31}\xi + a_{32}\eta + a_{33}} \quad (\text{A.32})$$

Un'omografia risulta invariante solamente ai bi-rapporti fra lunghezze, ovvero al rapporto di rapporti di lunghezze.

L'estensione delle trasformazioni viste al caso tridimensionale è abbastanza immediato: in tal caso l'applicazione lineare è rappresentata da una matrice 4x4 che va ad operare su grandezze omogenee tridimensionali: tutte le osservazioni fatte ed i metodi per il calcolo dei gradi di libertà di ciascuna trasformazione possono essere applicati alla stessa maniera.

A.2. Metodi di stima

Nel corso della trattazione si è più volte fatto riferimento a diverse metodologie di stima, soprattutto nel caso della structure and motion, per le quali sono richieste efficienza computazionale, stabilità numerica e, possibilmente, semplicità di implementazione. Si è visto inoltre come in molti casi si debba assumere una percentuale elevata di errori grossolani all'interno del set di dati.

A.2.1. Sistemi di equazioni e decomposizione ai valori singolari.

La decomposizione ai valori singolari (o Singular Value Decomposition, SVD), nota in letteratura con il nome di teorema di Eckart [Eckart 1939], è una delle più potenti e più utilizzate fattorizzazioni di matrici. Essa permette di scomporre una generica matrice A , costituita da m righe ed n colonne, nel prodotto di tre matrici:

$$A = UDV^T \quad (\text{A.33})$$

in cui U e V sono matrici quadrate rispettivamente di dimensione $m \times m$ ed $n \times n$ e D è una matrice rettangolare $m \times n$ con elementi non nulli solo sulla diagonale principale; gli elementi diagonali di D altro non sono che le radici quadrate degli autovalori della matrice $A^T A$; solitamente gli algoritmi per eseguire la decomposizione provvedono al riordino dei termini delle tre matrici in modo che in D i valori singolari siano inseriti in ordine decrescente dall'alto verso il basso: tale aspetto è importante per l'implementazione degli algoritmi di stima.

Gli aspetti teorici legati alla decomposizione, il suo significato geometrico e le modalità di implementazione trascendono la presente trattazione e si rimanda ai testi specializzati [Press 1988] per un maggior

approfondimento.

Quello che rende la decomposizione così interessante è che essa permette di ottenere stime equivalenti a quelle ottenibili con un approccio a minimi quadrati (e quindi di massima verosimiglianza) in maniera estremamente semplice ed efficiente: si consideri inizialmente un sistema di equazioni lineari in n incognite:

$$\mathbf{Ax} = \mathbf{b} \quad (\text{A.34})$$

E' noto che, se il numero di equazioni linearmente indipendenti m è maggiore del numero di incognite il sistema non ammette soluzioni; quello che si vuole ottenere è una stima del vettore \mathbf{x} per cui risulti minima la norma dei residui:

$$\|\mathbf{Ax} - \mathbf{b}\| = \min \quad (\text{A.35})$$

Sviluppando la decomposizione ai valori singolari della matrice A si ottiene:

$$\|\mathbf{Ax} - \mathbf{b}\| = \|\mathbf{UDV}^T \mathbf{x} - \mathbf{b}\| = \min \quad (\text{A.36})$$

Ricordando che U e V sono due matrici ortonormali, e quindi se applicate ad un vettore generico non ne modificano la norma, è possibile, detto $\mathbf{y} = \mathbf{V}^T \mathbf{x}$ ottenere:

$$\|\mathbf{U}^T (\mathbf{UDV}^T \mathbf{x} - \mathbf{b})\| = \|\mathbf{DV}^T \mathbf{x} - \mathbf{U}^T \mathbf{b}\| = \min \quad (\text{A.37})$$

$$\|\mathbf{Dy} - \mathbf{U}^T \mathbf{b}\| = \|\mathbf{Dy} - \mathbf{b}'\| = \min \quad (\text{A.38})$$

in altre parole si riconduce il problema iniziale alla stima del vettore \mathbf{y} , come definito in precedenza, che rende minima la quantità (A.38).

Ordinando gli elementi diagonali della matrice D in senso decrescente, il sistema complessivo assume la forma:

$$\begin{vmatrix} d_1 & & & \\ & d_2 & & \\ & & \ddots & \\ & & & d_n \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \end{vmatrix} \begin{vmatrix} |y_1| \\ |y_2| \\ \vdots \\ |y_n| \end{vmatrix} = \begin{vmatrix} b'_1 \\ \vdots \\ b'_n \\ b_{n+1} \\ \vdots \\ b_m \end{vmatrix} \quad (\text{A.39})$$

E' facile dunque accorgersi che la soluzione ottimale è ottenuta ponendo $y_i = b'_i/d_i$; a questo punto, per ottenere la stima nel sistema di riferimento originario, è sufficiente calcolare $\mathbf{x} = \mathbf{Vy}$.

Nel caso in cui il sistema abbia una deficienza di rango, ovvero che il numero di equazioni linearmente indipendenti sia minore del numero di incognite, come noto esso ammette ∞^{n-r} soluzioni, dove con r si è indicato il rango della matrice A : in questo caso la matrice D presenta $n-r$ valori singolari nulli; la soluzione

del sistema risulta pertanto data da tutti i vettori ottenuti da

$$\mathbf{x} = V\mathbf{y} + \lambda_{r+1}\mathbf{v}_{r+1} + \dots + \lambda_n\mathbf{v}_n \quad (\text{A.40})$$

in cui \mathbf{y} è il vettore definito come $y_i = b'_i/d_i$ ($i = 1, \dots, r$) e $y_i = 0$ per $i = r+1, \dots, n$, λ_n è un parametro reale e \mathbf{v}_n sono le ultime $n-r$ colonne della matrice V , che rappresentano una base dello spazio vettoriale ortogonale alla funzione da minimizzare.

Di notevole interesse è anche l'applicazione della SVD per la risoluzione di sistemi di equazioni omogenei: per determinare le soluzioni non banali si può considerare il seguente problema:

$$\|A\mathbf{x}\| = \min \quad \text{con} \quad \|\mathbf{x}\| = 1 \quad (\text{A.41})$$

Seguendo gli stessi procedimenti sviluppati in (A.37), (A.38) e (A.39) si arriva ad ottenere

$$\left\| \begin{array}{cccc} d_1 & & & \\ & d_2 & & \\ & & \ddots & \\ & & & d_n \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \end{array} \right\| \begin{array}{c} |y_1| \\ |y_2| \\ \vdots \\ |y_n| \end{array} = \min \quad (\text{A.42})$$

E' facile allora verificare che il vettore \mathbf{y} di norma unitaria che verifica la condizione di rendere minimo il modulo del vettore dei residui, ricordando che gli elementi diagonali della matrice D sono disposti in ordine decrescente, è il vettore

$$\mathbf{y} = (0 \ 0 \ \dots \ 1)^T \quad (\text{A.43})$$

Infine, il vettore soluzione $\mathbf{x} = V\mathbf{y}$, altro non è che l'ultima colonna della matrice V .

Per concludere si consideri il problema di stima incontrato ad esempio nella soluzione del tensore trifocale (§ 1.6.2.): in tal caso il problema era formulato come

$$\|A\mathbf{x}\| = \min \quad \text{con} \quad \|\mathbf{x}\| = 1 \quad e \quad \mathbf{x} = C\hat{\mathbf{x}} \quad (\text{A.44})$$

L'ultima condizione della (A.44) equivale a dire che il vettore \mathbf{x} è generato dalla base vettoriale rappresentata dalle colonne della matrice C ; tali colonne non costituiscono, in generale, una base ortonormale dello spazio vettoriale in cui è definito \mathbf{x} : si decomponga dunque C per mezzo della SVD ottenendo $C = UDV^T$, in cui, se r è il rango della matrice C , la matrice diagonale ha r elementi non nulli; è immediato verificare che, la matrice U' costituita dalle prime r colonne di U , rappresenta una base dello spazio vettoriale di \mathbf{x} equivalente al precedente ma ortonormale: si consideri in altre parole che $\mathbf{x} = U'\mathbf{x}'$ per un certo vettore \mathbf{x}' tale che $C\mathbf{x} = U'\mathbf{x}'$. A questo punto il problema (A.44) può essere riformulato come

$$\|AU'\mathbf{x}'\| = \min \quad \text{con} \quad \|U'\mathbf{x}'\| = 1 \quad (\text{A.45})$$

E' immediato verificare che, non influenzando U' sulla norma del vettore \mathbf{x}' , la seconda condizione è equiva-

lente a $\|U'\mathbf{x}'\| = \|\mathbf{x}'\|$; in altre parole si è ricondotto il problema a

$$\|B\mathbf{x}'\| = \min \quad \text{con } \|\mathbf{x}'\| = 1 \quad (\text{A.46})$$

in cui $B = AU'$, che risulta in tutto equivalente a quello espresso in (A.41). La soluzione al problema, una volta ottenuto il vettore \mathbf{x}' risulta banalmente $\mathbf{x} = U'\mathbf{x}'$.

Nel caso in cui venga infine richiesto anche il calcolo del vettore $\hat{\mathbf{x}}$, della seconda condizione della (A.44) si può ragionare come segue: da tale condizione è facile ricavare che

$$\hat{\mathbf{x}} = C^+ \mathbf{x} = C^+ U' \mathbf{x}' \quad (\text{A.47})$$

in cui con G^+ si è indicata la matrice pseudo-inversa di C ($C^+ = (C^T C)^{-1} C^T$); ottenendo la decomposizione ai valori singolari anche di quest'ultima matrice $C^+ = VD^+U^T$, è possibile scrivere

$$\hat{\mathbf{x}} = VD^+U^T U' \mathbf{x}' \quad (\text{A.48})$$

che può essere semplificata come

$$\hat{\mathbf{x}} = V'D'^{-1}\mathbf{x}' \quad (\text{A.49})$$

in cui V' rappresenta la matrice ottenuta dalle prime r colonne di V e D' è il minore rxr che contiene i termini non nulli di D .

Per quanto riguarda la complessità di calcolo che richiede la SVD, data una matrice $m \times n$ è stato dimostrato in [Golub 1989] che il numero di operazioni richieste è proporzionale al valore $4m^2n + 8mn^2 + 8n^3$; la maggior parte delle implementazioni della decomposizione, come ad esempio quella in [Press 1988], permette tuttavia di calcolare, se richiesto, solamente le matrici D e V , in quanto gli elementi di D vengono determinati al termine della procedura: in tal modo il numero di operazioni richieste dal calcolatore per effettuare la decomposizione divengono solamente $4mn^2 + 8n^3$; si osservi, in questo caso, che il termine in cui m compare al quadrato sparisce dall'equazione: come si è visto solitamente il vettore soluzione di un sistema può essere ottenuto dalla/dalle ultime colonne della matrice V ; al tempo stesso, mentre n è un valore solitamente contenuto, in quanto rappresenta il numero di parametri utilizzati nella stima, il numero di equazioni m può assumere valori molto elevati (si pensi al caso della stima della matrice fondamentale in cui si vogliono stimare 8 parametri, ma il più delle volte il numero delle corrispondenze è di alcune centinaia). E' facile allora accorgersi che, in questo caso, la complessità numerica del problema diventa rapidamente, all'aumentare del valore di m , meno impegnativa: in tutti i casi presentati nella presente trattazione, in cui si è ricorso ad una decomposizione ai valori singolari in cui fosse possibile escludere il calcolo della matrice U , si è ricorso a questo metodo, per rendere computazionalmente più efficiente l'implementazione.

A.2.2. Il metodo di Levenberg-Marquadt

Il metodo di Levenberg-Marquadt, per la risoluzione di problemi di stima non lineari a minimi quadrati, è una variante del metodo di Newton. Si consideri, ad una generica iterazione, la stima delle correzioni Δ da apportare ai parametri iniziali \mathbf{p}_0 ottenuta linearizzando per mezzo di uno sviluppo in serie di Taylor arrestato al primo ordine la funzione (non lineare) che descrive il modello funzionale delle osservazioni:

$$\Delta = (J^T P J)^{-1} J^T P (\mathbf{Y} - \mathbf{y}_0) = N \cdot J^T P (\mathbf{Y} - \mathbf{y}_0) \quad (\text{A.50})$$

La matrice normale $J^T P J$ rappresenta geometricamente l'iper-piano tangente alla varietà lineare che descrive, nello spazio dei parametri, la funzione obiettivo da minimizzare, calcolata nel punto corrispondente ai parametri iniziali. Nell'impostazione di Levenberg-Marquadt essa viene modificata, sostituendo ai suoi elementi diagonali N_{ii} , la quantità $N'_{ii} = (\lambda+1) N_{ii}$; il termine λ assume inizialmente valori piccoli (dell'ordine di 10^{-3}): se la soluzione del sistema porta ad una riduzione dell'errore, ovvero si assiste ad una diminuzione del valore della funzione obiettivo che si vuole minimizzare, il valore λ viene ridotto (solitamente di un fattore 10); se al contrario si assiste ad un incremento dei residui il suo valore viene aumentato (sempre di un fattore 10).

In altre parole il metodo di Levenberg-Marquadt è un metodo adattivo: diventa praticamente equivalente a quello di Newton laddove non emergono instabilità o difficoltà di convergenza e quindi non occorre condizionamento della matrice normale; è invece più vicino ad un metodo iterativo come il gradiente coniugato in presenza di minimi locali o di oscillazioni della soluzione. In tal caso cerca di aumentare la diagonalizzazione della matrice, il che in linea di massima stabilizza il sistema.

Nella maggioranza dei casi si consegue un aumento della velocità di convergenza alla soluzione, oltre a rendere più stabile la stima di fronte a minimi locali: per tale motivo è stata implementata nel calcolo del tensore trifocale (si veda § 1.6.2.) nel quale solitamente si presentano problemi di questo tipo.

A.2.3. Line search

Un altro metodo che si dimostra abbastanza robusto di fronte a problemi di convergenza nella soluzione di sistemi non lineari a minimi quadrati, quando si dispone di parametri iniziali non ottimali e quindi la convergenza del metodo di Newton non è garantita, è rappresentato dal line search, un algoritmo deterministico di ottimizzazione che consente un miglior controllo della ricerca del minimo. Ad ogni iterazione per la non linearità, per mezzo della (A.50), si analizza iterativamente se la soluzione

$$\mathbf{x} = \mathbf{x}_k + \alpha \mathbf{d}_k \quad (\text{A.51})$$

in cui \mathbf{x}_k rappresenta il vettore di parametri iniziali all'iterazione k considerata, α un valore reale positivo e $\mathbf{d}_k = \Delta$ rappresenta la direzione in cui si sta muovendo la soluzione a minimi quadrati, presenta una riduzione più significativa dei residui per un determinato valore di α ; in altre parole, si considera la funzione obiettivo parametrizzata per mezzo di α , una volta noti i parametri iniziali e la soluzione (A.50):

$$\varphi(\alpha) = f(\mathbf{x}_k + \alpha \mathbf{d}_k) \quad (\text{A.52})$$

e si valuta per quale valore di α essa risulti minima: ciò equivale a considerare come corretta la direzione della soluzione ottenuta con i minimi quadrati tradizionali nello spazio dei parametri, ma verificando però se un differente modulo del vettore correzione non porti a riduzioni più significative dei residui.

Per studiare la funzione (A.52) si può ricorrere al metodo di Armijo che consiste nel calcolare inizialmente la riduzione della norma dei residui fra un'iterazione e la successiva assumendo $\alpha = 1$ (ovvero considerando la stessa soluzione dei minimi quadrati tradizionali); se tale riduzione risulta significativa si fa l'ipotesi che il minimo della funzione (A.52) sia prossimo al valore di α utilizzato e si procede con una nuova iterazione del sistema; se al contrario tale riduzione è inferiore ad un certo valore, o addirittura, con la soluzione calcolata, la norma dei residui tende ad aumentare, si moltiplica α per un coefficiente di contrazione $\sigma \leq 0.5$ e si ricalcola la (A.52); il processo viene iterato fino al raggiungimento di una riduzione accettabile o al raggiungimento di un valore del parametro α inferiore ad una certa soglia (solitamente $0.01 \div 0.001$), nel qual caso si conclude che non è stato possibile individuare un parametro α che garantisca l'ottimizzazione della funzione.

La riduzione richiesta può essere espressa dalla condizione di Armijo [Armijo 1966], che deriva dalla prima condizione di Wolfe ([Wolfe 1969]) per garantire la convergenza globale alla soluzione:

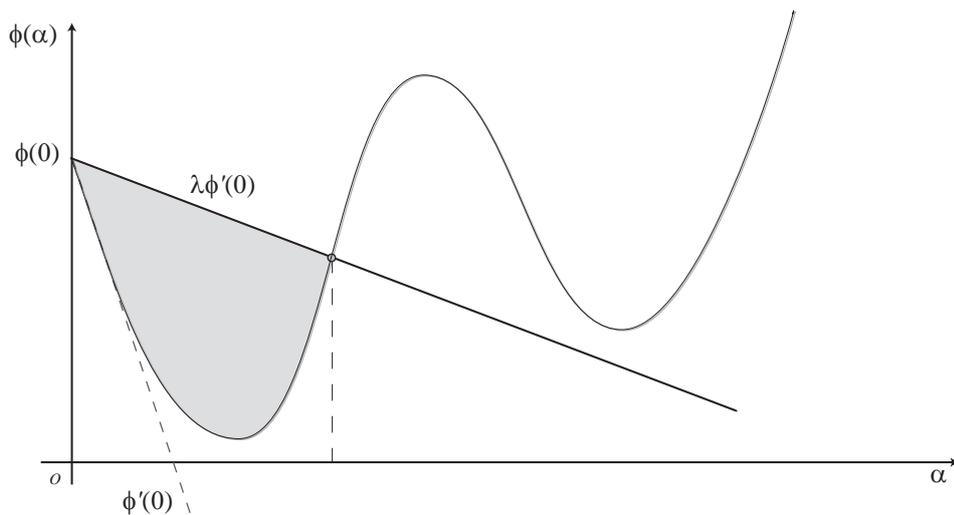


Figura A.3: Rappresentazione della condizione di Armijo nel line-search.

$$\varphi(\alpha) \leq \varphi(0) + \lambda\alpha\varphi'(0) \quad (\text{A.53})$$

in cui λ è un parametro compreso fra 0 e 1 che definisce il grado di ristrettezza della regola (A.53); la precedente relazione, riferendoci alla figura A.3, assume il significato che, una riduzione della funzione (A.52), per essere significativa, deve essere tale che il nuovo punto stimato deve trovarsi al di sotto della retta passante per il punto $(0, \varphi(0))$, ovvero il punto corrispondente all'iterazione precedente, e avente pendenza pari a $\lambda\varphi'(0)$, che, essendo λ minore di uno, risulta meno inclinata della tangente alla funzione obiettivo nel punto $\alpha = 0$. Se questo è vero la diminuzione dei residui osservata corrisponde effettivamente ad un movimento della soluzione verso il minimo assoluto.

Il valore che assume la derivata in (A.53) può essere efficacemente calcolato per mezzo della matrice Jacobiana calcolata nel sistema a minimi quadrati:

$$\varphi'(0) = J_k^T Pr_k \quad (\text{A.54})$$

in cui r_k rappresenta il vettore dei residui.

L'utilizzo delle tecniche di line-search, è di recente introduzione nelle discipline legate alla fotogrammetria: in [Börlin 2002] esse vengono applicate al LSM [Grün 1985], in cui si mostra che, pur partendo da posizioni approssimate tutt'altro che ottimali, il metodo raggiunge la soluzione corretta; in [Börlin 2003] e in [Börlin 2004] lo stesso tipo di impostazione viene utilizzato per la soluzione di blocchi fotogrammetrici mediante l'approccio a stelle proiettive in cui però il metodo non sembra fornire lo stesso grado di ottimizzazione rispetto al tradizionale approccio dei minimi quadrati.

L'introduzione di un modulo di line-search all'interno di una procedura a minimi quadrati, oltre a non presentare grosse difficoltà in termini di implementazione, genera un incremento dello sforzo computazionale del tutto trascurabile quando il problema è stabile: pertanto, la sua implementazione e il suo utilizzo è sempre auspicabile.

A.2.4. Metodi di stima robusti

E' noto che metodi di stima a minimi quadrati sono fortemente influenzati dalla presenza di errori grossolani (outlier) nelle osservazioni: non solo la stima dei parametri risulta falsata, ma spesso non si è neppure in grado di mettere in luce, dallo studio dei residui, quale osservazione sia effettivamente errata e vada quindi eliminata dal processo, a causa del mascheramento degli errori. Se poi la percentuale di outlier è elevata, il metodo di stima è sostanzialmente inutilizzabile.

L'algoritmo RANSAC (Random Sample Consensus [Fischler 1981]), sviluppato nella sua prima formulazione da Fischler e Bolles nel lontano 1981 è in grado di riconoscere sottoinsiemi di dati coerenti

con un modello assegnato, a partire da set di dati contenenti percentuali significative di errori grossolani (teoricamente prossimi al 100%). Il suo approccio è simmetrico a quello dei minimi quadrati tradizionali: invece di ottenere una stima dei parametri per mezzo di tutte le osservazioni disponibili, si determinano tali parametri per mezzo di un set minimo di equazioni; si va poi a verificare quali elementi del set di dati si dimostra coerente con il modello funzionale determinato e si calcola di conseguenza il numero di inlier, che rappresenta il consenso a tale modello. Ripetendo il processo un certo numero di volte, si assume che i parametri corretti (cioè che sono stati stimati a partire da osservazioni prive di errori grossolani) siano quelli che forniscono il consenso massimo. Formalmente: dato un modello che dipende da n parametri e noto un set di dati P (con numerosità $p > n$), si estrae casualmente da P un subset S_1 di n elementi necessari per determinare un primo modello M_1 . Il modello così ottenuto va a definire il subset S_1^* di punti appartenenti a P che sono all'interno di una predeterminata soglia d'errore da M_1 . Il set S_1^* è detto *consensus set* di S_1 .

Si noti che utilizzando come criterio il residuo delle equazioni di osservazione non è detto che all'interno della soglia di accettazione non siano presenti errori grossolani. Al tempo stesso si ritiene che se il subset minimo S_1 non è affetto da errori, tutti i dati che mostrano consenso a tale subset possono considerarsi coerenti. Inoltre si stabilisce per via probabilistica quante estrazioni di subset siano necessarie per ottenere con una certa probabilità almeno un campione privo di errori grossolani, nota la loro percentuale. Detta quindi p la probabilità che si vuole raggiungere di estrarre, nel corso delle iterazioni, almeno un subset costituito unicamente da inlier, ϵ la percentuale di outlier nelle osservazioni, si può ragionare come segue: la probabilità, con una singola estrazione, di ottenere un inlier risulta $(1-\epsilon)$; escludendo tale punto dal set ed eseguendo ulteriori estrazioni la probabilità che tutte le n osservazioni determinate non rappresentino errori grossolani è data da

$$P_{inlier} = (1 - \epsilon)(1 - \epsilon_2) \dots (1 - \epsilon_n) \quad (A.55)$$

in cui ϵ_i rappresenta la percentuale di outlier alla i -esima estrazione; per semplicità, considerando che solitamente il numero di osservazioni è piuttosto elevato in confronto al numero di estrazioni richieste per definire il subset minimo, si può pensare di assumere tali valori come costanti e pari all' ϵ iniziale; in tal caso la (A.51) diventa:

$$P_{inlier} = (1 - \epsilon)(1 - \epsilon) \dots (1 - \epsilon) = (1 - \epsilon)^n \quad (A.56)$$

E' immediato allora ricavare che la probabilità di estrarre un campione di n elementi in cui almeno una delle osservazioni è un errore grossolano risulta

$$P_{outlier} = 1 - P_{inlier} = 1 - (1 - \epsilon)^n \quad (A.57)$$

Considerando N iterazioni, ovvero N estrazioni casuali di subset minimi dal campione delle osservazioni, la probabilità di aver ottenuto sempre almeno un errore grossolano diventa

$$P_{outlier} = (1 - (1 - \epsilon)^n)^N \quad (A.58)$$

La stessa probabilità può essere confrontata con la probabilità p di avere, al termine delle iterazioni, almeno un set libero da errori grossolani:

$$(1 - p) = (1 - (1 - \varepsilon)^n)^N \quad (\text{A.59})$$

da cui è facile ricavare il numero minimo di iterazioni richieste per garantire con probabilità p che almeno un subset minimo non fosse affetto da osservazioni errate:

$$N = \frac{\log(1 - p)}{\log(1 - (1 - \varepsilon)^n)} \quad (\text{A.60})$$

Il numero di iterazioni richieste, oltre a dipendere direttamente dalla probabilità ammessa di non avere outlier in almeno un'estrazione, tende ad aumentare proporzionalmente con la percentuale di outlier e con le dimensioni del subset minimale: di seguito, in tabella A.1 e A.2, vengono presentati i valori numerici desumibili dalla (A.56).

p = 99%		Percentuale di outlier ε						
		5%	10%	20%	30%	50%	70%	85%
dimensione n	2	2	3	5	7	16	49	202
	3	2	4	6	11	34	168	1'362
	4	3	4	9	17	71	566	9'094
	5	3	5	12	25	145	1'893	60'642
	6	3	6	15	37	292	6'315	404'292
	7	4	7	20	54	587	21'055	2'695'296
	8	4	8	25	78	1'177	70'188	17'968'654

Tabella A.1: Numero di iterazioni minime assumendo la probabilità $p = 99\%$.

p = 95%		Percentuale di outlier ε						
		5%	10%	20%	30%	50%	70%	85%
dimensione n	2	1	2	3	4	10	32	132
	3	2	2	4	7	22	109	886
	4	2	3	6	11	46	368	5'916
	5	2	3	8	16	94	1'231	39'448
	6	2	4	10	24	190	4'108	262'998
	7	2	5	13	35	382	13'696	1'753'331
	8	3	5	16	50	765	45'658	11'688'879

Tabella A.2: Numero di iterazioni minime assumendo la probabilità $p = 95\%$.

Dall'analisi delle tabelle si nota che, con bassa presenza di outlier (10-20%), il numero di iterazioni ri-

chiesto è modesto e non richiede sforzi computazionali particolari, indipendentemente dalla dimensione del subset minimo; con l'aumento della percentuale di errori grossolani esso cresce invece esponenzialmente, mentre considerare probabilità minori di avere set esenti da errori non comporta una sostanziale diminuzione delle iterazioni necessarie; è quindi preferibile mantenere tale valore elevato (99%).

Spesso non si ha una stima a priori della percentuale di outlier presenti nel campione di osservazioni: se per valori bassi, in termini computazionali, è praticamente ininfluente impostare tale valore in maniera conservativa, quando la percentuale attesa supera il 40-50%, per limitare i tempi di elaborazione è importante averne una stima corretta. Per risolvere il problema è sufficiente utilizzare un approccio adattivo: si inizializza il sistema con un percentuale di outlier prossimo al 100%; al termine di ogni iterazione si calcola il consenso del set stimato: se la percentuale di inlier è maggiore di quella di massimo consenso fino a quel momento determinato, si ricalcola la percentuale di outlier e il numero di iterazioni necessarie: se esso è già stato raggiunto si termina il procedimento, altrimenti si prosegue. In questo modo, non solo non occorre specificare inizialmente la percentuale attesa di outlier, ma si è certi che il metodo non esegue iterazioni supplementari quando il numero di ripetizioni necessarie è stato raggiunto.

Un'ulteriore osservazione riguarda la soglia di accettazione per stabilire se un'osservazione deve essere considerata inlier o outlier: il valore di tale soglia determina direttamente la suddivisione delle due classi e la numerosità reciproca ed è quindi di fondamentale importanza. Al tempo stesso, in molte applicazioni, essa rappresenta l'unico parametro di cui si debba dare, a priori, una stima sufficientemente accurata. Come si è visto nei numerosi esempi trattati il valore della soglia di accettazione viene impostato basandosi sull'esperienza (ad esempio per la stima del tensore trifocale o della matrice fondamentale, con camere calibrate, si suppone che gli operatori di interesse forniscano buoni risultati e si imposta tale soglia a $1 \div 1.5$ pixel) o su prove effettuate su porzioni del campione (come nella determinazione dei piani di discontinuità).

Al termine della procedura è consigliabile stimare un nuovo modello M (ad esempio a minimi quadrati) usando il subset di punti di massimo consenso. Tramite metodi di individuazione di errori grossolani, di significatività dei parametri e di data snooping è possibile ottenere una valutazione statisticamente più raffinata e corretta della presenza di errori grossolani e una eventuale ulteriore eliminazione di osservazioni "sospette".

L'algoritmo Ransac, pur essendo estremamente potente e versatile, presenta l'innegabile svantaggio di semplificare in maniera eccessiva la natura del problema: il set di dati di partenza viene suddiviso in due sole categorie: inlier ed outlier; in altre parole il consenso di un subset viene valutato solamente sul numero di dati che sembrano essere, secondo una certa soglia di accettazione, in linea con un modello funzionale prestabilito per mezzo di un numero minimo di punti del campione. Tale numerosità non è però direttamente legata con quale sia effettivamente il consenso con il modello funzionale, ovvero quanto bene i dati considerati inlier si dimostrino in accordo con il modello funzionale stimato. Si pensi, per esempio, al caso in cui due subset presentano lo stesso numero di inlier (e quindi lo stesso consenso): non è possibile, con il criterio utilizzato, mettere in mostra quale dei due subset si dimostri più affidabile e vada dunque utilizzato.

Al tempo stesso, quando la divisione fra inlier e outlier non risulta così netta, può accadere che subset di dati meno numerosi rappresentino una soluzione più corretta rispetto ad altri con un maggior numero di punti che verificano la soglia di accettazione. Sono dunque nati nuovi metodi, basati su Ransac, che valutano però il consenso del set di dati utilizzando grandezze statistiche più significative.

In particolare è opportuno considerare, come suggerito in [Torr 1998], come parametro di consenso la somma dei residui:

$$C = \sum_i \rho(\varepsilon_i^2) \quad (\text{A.61})$$

in cui

$$\rho(\varepsilon^2) = \begin{cases} \varepsilon^2 & \text{se } \varepsilon^2 < T^2 \\ T^2 & \text{se } \varepsilon^2 \geq T^2 \end{cases} \quad (\text{A.62})$$

In tal modo il consenso viene stabilito in base alla bontà dell'adattamento al modello funzionale scelto oltre che alla numerosità di dati coerenti con esso.

A.3. Metodi di resezione

Nella ricostruzione metrica di un blocco fotogrammetrico con tecniche di structure and motion si alternano operazioni di intersezioni in avanti e resezione (si veda § 1.7): fissata la posizione assoluta di due fotogrammi per mezzo della matrice essenziale, bisogna riportare tutte le altre immagini che compongono il blocco nello stesso sistema di riferimento per mezzo dei punti di legame; note allora le posizioni nel sistema oggetto di tali punti, si possono stimare i parametri di orientamento esterno delle camere per mezzo di algoritmi di resezione. La letteratura scientifica, a riguardo, è ricchissima di spunti interessanti: solitamente il problema viene affrontato per via iterativa (essendo non lineare) come ad esempio in [Lowe 1991] e in [Yuan 1989], o tramite metodi diretti come illustrato in [Heralick 1991]. Entrambi i metodi presentano però delle limitazioni: quelli iterativi richiedono la conoscenza di parametri approssimati sufficientemente vicini alla soluzione corretta; i metodi diretti utilizzano solitamente il numero minimo di equazioni (3 o 4 punti) e sono pertanto sensibili al rumore delle osservazioni o all'eventuale presenza di errori grossolani nel set di dati utilizzato.

Nella structure and motion i punti di legame da cui si può ottenere la stima degli orientamenti per resezione sono numerosi (da alcune decine ad alcune centinaia) e potrebbero fornire una soluzione molto più ridondante e quindi stabile. Il metodo sviluppato, pertanto, si ispira a quello di [Quan 1999], in cui è possibile fare uso, per la stima diretta dei parametri di orientamento, di un numero qualsiasi di osservazioni.

Considerando il triangolo formato dal centro di proiezione e da due punti di cui si conoscono le coordinate immagine e terreno, dette x_i e x_j le distanze fra il centro della camera e i due punti, dal teorema di Carnot si ricava

$$d_{ij}^2 = x_i^2 + x_j^2 - 2x_i x_j \cos \theta_{ij} \quad (\text{A.63})$$

in cui θ_{ij} è l'angolo sotteso dai due raggi di proiezione; tale angolo, come mostrato in [Hartley 2000], può essere ottenuto per mezzo della matrice camera, della quale si è parlato in § 1.4., per mezzo dell'equazione

$$\cos \theta_{ij} = \frac{\mathbf{u}_i^T \mathbf{C} \mathbf{u}_j}{(\mathbf{u}_i^T \mathbf{C} \mathbf{u}_i)^{1/2} (\mathbf{u}_j^T \mathbf{C} \mathbf{u}_j)^{1/2}} \quad (C = (KK^T)^{-1}) \quad (\text{A.64})$$

in cui \mathbf{u}_i e \mathbf{u}_j sono le coordinate omogenee dei punti immagine. E' allora possibile scrivere, dalla (A.63), l'equazione che lega due distanze fra punti oggetto e il centro di proiezione e l'interdistanza fra gli stessi punti nel sistema terreno:

$$f_{ij}(i, j) = x_i^2 + x_j^2 - 2x_i x_j \cos \theta_{ij} - d_{ij}^2 \quad (\text{A.65})$$

Mettendo a sistema almeno tre condizioni del tipo (A.65)

$$\begin{cases} f_{12}(x_1, x_2) = 0 \\ f_{23}(x_2, x_3) = 0 \\ f_{13}(x_1, x_3) = 0 \end{cases} \quad (\text{A.66})$$

è possibile, per sostituzione mediante la risultante di Sylvester, ottenere una equazione polinomiale in una sola delle variabili di (A.66). Tale equazione assume la forma:

$$g(x_i) = a_5 x_i^8 + a_4 x_i^6 + a_3 x_i^4 + a_2 x_i^2 + a_1 = 0 \quad (\text{A.67})$$

in cui i parametri a_i sono ottenuti dalla succitata risultante di Sylvester. Ponendo $x = x_i^2$ si ottiene dalla (A.67) un'equazione di quarto grado:

$$g(x) = a_5 x^4 + a_4 x^3 + a_3 x^2 + a_2 x + a_1 \quad (\text{A.68})$$

la quale ammette come noto 4 soluzioni (che possono essere tutte e quattro reali, due reali e due complesse, o tutte complesse).

I metodi diretti tradizionali, per determinare quale delle soluzioni debba essere accettata, ricalcolano la stessa grandezza utilizzando un altro punto (ad esempio sostituendo x_3 con x_4 nella (A.66)) e scelgono le soluzioni comuni. Calcolando la (A.68) con tutti gli n punti disponibili, per le $n' = \frac{(n-1)(n-2)}{2}$ possibili combinazioni in cui compare x_i , si ottiene un sistema nella sola incognita x del tipo

$$\begin{vmatrix} a_1^1 & a_2^1 & a_3^1 & a_4^1 & a_5^1 & 1 \\ a_1^2 & a_2^2 & a_3^2 & a_4^2 & a_5^2 & x \\ \vdots & \vdots & \vdots & \vdots & \vdots & x^2 \\ a_1^{n'} & a_2^{n'} & a_3^{n'} & a_4^{n'} & a_5^{n'} & x^3 \\ & & & & & x^4 \end{vmatrix} = 0 \quad (\text{A.69})$$

che può essere risolto con la SVD.

Considerando le diverse potenze di x come variabili indipendenti, il valore ottenuto per la distanza non sarà univoco: in altri termini il rapporto $x_i/x_{(i-1)}$ non ha sempre lo stesso valore al variare di i ; per ottenere una stima del quadrato della distanza x , si possono tuttavia mediare tali valori, in quanto, salvo in presenza di molti errori grossolani, approssimativamente essi si equivalgono. Reiterando il procedimento per tutti i punti a disposizione si determina la loro distanza dal centro di presa, e pertanto, muovendosi lungo il raggio di proiezione, la loro posizione nel sistema di riferimento immagine.

A questo punto, per ottenere i parametri di trasformazione che permettono di portare il sistema di riferimento immagine nel sistema di riferimento oggetto si può procedere come suggerito in [Horn 1987]: innanzitutto si stima una traslazione da applicare al sistema di riferimento immagine ottenuta come differenza fra le coordinate del baricentro dei punti in coordinate oggetto e terreno: si applica poi tale traslazione a tutti i punti ottenuti nei precedenti passaggi in modo che i due baricentri abbiano le stesse coordinate. Se desiderato, calcolando il momento centrale di inerzia di tutti i punti nei due sistemi, si può anche determinare un fattore di scala:

$$\lambda = \frac{\sum_i \|q_i - G\|}{\sum_i \|p_i - G\|} \quad (\text{A.70})$$

in cui con p e q si sono indicate le coordinate rispettivamente nel sistema terreno e nel nuovo sistema immagine e G le coordinate del baricentro comune; è importante osservare che la stima di tale parametro può, a seconda dei casi, essere pressoché ininfluenza, migliorare la stima o peggiorarla drasticamente: se il set di dati è abbastanza compatto e le osservazioni sono in qualche modo affette da noise di misura, un fattore comune di scala λ può correggere eventuali inesattezze derivate dalla (A.69); viceversa, se alcuni punti sono molto distanti dagli altri, solitamente essi presentano precisioni di posizionamento molto meno accurate e il parametro λ può quindi falsare notevolmente il risultato. Infine, per mezzo dei quaternioni [Sansò 1973], è possibile ricavare gli angoli di rotazione. I quaternioni, estensione tridimensionale dei numeri complessi, permettono di rappresentare le rotazioni nello spazio. Un generico quaternione può essere rappresentato mediante la forma:

$$\hat{q} = q_0 + iq_x + jq_y + kq_z \quad (\text{A.71})$$

in cui i, j e k rappresentano l'unità complessa.

E' possibile dimostrare che una generica rotazione nello spazio di un angolo θ attorno ad un generico asse individuato dal versore $\boldsymbol{\omega} = (\omega_x \ \omega_y \ \omega_z)^T$ può essere rappresentato dal quaternione (a norma unitaria)

$$\dot{q}(\theta, \boldsymbol{\omega}) = \cos \frac{\theta}{2} + \sin \frac{\theta}{2} (i\omega_x + j\omega_y + k\omega_z) \quad (\text{A.72})$$

Il vantaggio di questa notazione consiste proprio nel fatto che, un qualsiasi quaternione, per rappresentare correttamente una rotazione nello spazio, deve unicamente avere norma unitaria.

Dalla (A.72), con la notazione utilizzata nella (A.71), è possibile ricavare la corrispondente matrice di rotazione per mezzo della matrice di Rodriguez:

$$R = \begin{vmatrix} q_0^2 + q_x^2 - q_y^2 - q_z^2 & 2(q_x q_y - q_0 q_z) & 2(q_x q_z + q_0 q_y) \\ 2(q_x q_y + q_0 q_z) & q_0^2 - q_x^2 + q_y^2 - q_z^2 & 2(q_y q_z - q_0 q_x) \\ 2(q_x q_z - q_0 q_y) & 2(q_y q_z + q_0 q_x) & q_0^2 - q_x^2 - q_y^2 + q_z^2 \end{vmatrix} \quad (\text{A.73})$$

Come illustrato in [Sansò 1973] e in [Horn 1987] è possibile determinare la rotazione ottimale risolvendo un problema agli autovalori: dette x_{im} e y_{im} le componenti m -esime dell' i -esimo punto, rispettivamente nello spazio immagine e in quello oggetto e indicata con s^{mn} la quantità

$$s^{mn} = \sum_i x_{im} y_{in} \quad (\text{A.74})$$

Il quaternione corrispondente alla rotazione cercata può essere ottenuto normalizzando l'autovettore corrispondente al massimo autovalore della seguente matrice simmetrica:

$$S = \begin{vmatrix} s^{11} + s^{22} + s^{33} & s^{23} - s^{32} & s^{31} - s^{13} & s^{12} - s^{21} \\ \text{simm} & s^{11} - s^{22} - s^{33} & s^{12} + s^{21} & s^{31} + s^{13} \\ \text{simm} & \text{simm} & -s^{11} + s^{22} - s^{33} & s^{23} + s^{32} \\ \text{simm} & \text{simm} & \text{simm} & -s^{11} - s^{22} + s^{33} \end{vmatrix} \quad (\text{A.75})$$

Per maggiori dettagli sugli aspetti teorici legati alla metodologia si rimanda ai succitati testi.

Va infine precisato che la soluzione con i quaternioni è equivalente a quella che si otterrebbe con un sistema a minimi quadrati (risolto ad esempio con il metodo di Newton); tuttavia la determinazione diretta dei punti oggetto nel sistema immagine, può non essere del tutto corretta, a causa di effetti di rumore sulle osservazioni: una volta ottenuti i parametri di orientamento esterno si consiglia quindi di effettuare sempre una ulteriore stima della resezione mediante metodi non lineari (ad esempio a minimi quadrati).

A.4. Logica fuzzy

La logica fuzzy ([Zadeh 1965], [Zadeh 1973]) nasce come soluzione alternativa alla logica bivalente, ossia quella aristotelica. Da oltre duemila anni i problemi sono affrontati e risolti con l'ausilio della logica del "sì o no", del "vero o falso", e, sebbene fin dalle sue origini alcuni filosofi abbiano messo in luce alcuni suoi punti deboli (ricordiamo i famosi paradossi di Zenone e le più recenti critiche di Russell), il suo predominio è tuttora indiscusso.

Solitamente nell'implementazione di un algoritmo, i flussi decisionali vengono orientati calcolando parametri, confrontandoli con soglie e verificando se la condizione è vera (1) o falsa (0); con il crescere della complessità dei problemi, formulare condizioni strette porta spesso a semplificazioni ed ipotesi aggiuntive; anche il modo di operare della mente umana solitamente prevede un maggior grado di variabilità nell'affrontare le decisioni.

Si è visto nel capitolo 4 che è complesso fissare una serie di parametri di soglia (alcuni anche fortemente correlati) legati all'algoritmo di estrazione automatica dei piani di discontinuità. A tale fine si sta sviluppando un approccio basato proprio sulla logica fuzzy: al momento tuttavia non si è ancora ottenuto un livello di sviluppo sufficiente per presentare risultati, ma si è ritenuto opportuno dare alcuni brevi cenni sull'argomento.

Come la logica tradizionale, bivalente, si basa sugli insiemi classici, la logica fuzzy si basa sugli insiemi fuzzy. Un insieme fuzzy è un insieme di oggetti nel quale non c'è un confine ben preciso o definito tra gli oggetti che vi appartengono e quelli che non vi appartengono. Il concetto alla base di tale definizione è quello di appartenenza: ad ogni elemento è associato un grado di appartenenza all'insieme, espresso da un numero compreso nell'intervallo $[0, 1]$, dove 0 e 1 indicano rispettivamente il minore ed il maggiore grado di appartenenza, mentre tutti i valori intermedi indicano appartenenza "parziale". Gli insiemi classici sono visti quindi come un caso particolare di insiemi fuzzy.

In [Zadeh 1973] viene osservato che gli elementi chiave del pensiero umano non sono numeri, ma "etichette" di insiemi fuzzy. Una delle capacità più sorprendenti del cervello umano, tutt'ora non riproducibile dall'intelligenza artificiale, è quella di riassumere informazioni; un riassunto per sua natura è un'approssimazione, e il cervello umano trae vantaggio da questa tolleranza all'imprecisione attraverso la codificazione delle informazioni più rilevanti rispetto a determinata necessità, racchiudendo queste informazioni in insiemi fuzzy che vengono successivamente etichettati. Per questo la logica fuzzy rappresenta un ottimo strumento di gestione della polivalenza e della vaghezza del linguaggio naturale, pur ammettendo una struttura formale che ne permette una successiva rappresentazione numerica.

In generale un concetto si dice fuzzy, quando corrisponde ad una classe di oggetti che non hanno confini ben precisi, ad esempio, bello, giovane, amico, vicino, ecc..., mentre altri concetti non lo sono: sposato, maschio, fratello, ecc... Si può notare che generalmente i concetti fuzzy sono più complessi dei concetti che

non lo sono (pensiamo ad esempio al concetto “amico” e “fratello”).

In tal senso, si riprenda uno degli esempi sviluppati in precedenza: la determinazione delle superfici di discontinuità. Esso si presenta come un problema in cui diversi fattori concorrono a stabilire se una determinata porzione di DSM debba essere considerata elemento di discontinuità oppure debba essere aggregato con altri elementi o, ancora, debba essere scomposto in elementi più semplici: tali fattori, come si è visto in § 3.3., possono essere la forma dell'elemento, la sua estensione, i rapporti con gli elementi adiacenti. In tal senso, finora è stato abbastanza complesso inserire regole che consentissero una segmentazione corretta della geometria, utilizzando la logica booleana tradizionale: tutti gli elementi precedentemente illustrati concorrono infatti alla definizione di “buona” o “corretta” superficie di discontinuità e tale giudizio di merito risente notevolmente della limitazione di poter assumere solamente due valori distinti, quando sarebbe più corretto considerare un intervallo di differenti possibilità.

Appendice B

B.1. Matrice camera

$$P = KR |I \quad -\tilde{C}| = K |R \quad -R\tilde{C}| = |M \quad -M\tilde{C}| = \begin{vmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ a_{31} & a_{32} & a_{33} & a_{34} \end{vmatrix} = |p_1 \quad p_2 \quad p_3 \quad p_4| = \begin{vmatrix} p^1 \\ p^2 \\ p^3 \end{vmatrix}$$

Centro di proiezione:

$$C \text{ t.c. } PC = 0 \quad (SVD)$$

Matrice di rotazione e matrice di calibrazione:

$$\text{decomposizione } RQ: \quad M = KR$$

Equazione del piano principale:

$$P^{3T} X = 0$$

Punti di fuga degli assi X, Y, Z sul fotogramma:

$$v_x = p_1, \quad v_y = p_2, \quad v_z = p_3$$

Proiezione diretta di un punto oggetto X :

$$x = PX$$

Proiezione inversa di un punto (raggio di proiezione):

$$X(\lambda) = P^+ x + \lambda C$$

Proiezione inversa di una retta l (piano di proiezione):

$$\pi = P^T l$$

B.2. Matrice fondamentale

$$F = \begin{bmatrix} f_{11} & f_{12} & f_{13} \\ f_{21} & f_{22} & f_{23} \\ f_{31} & f_{32} & f_{33} \end{bmatrix}$$

Linee epipolari:

$$l' = Fx \quad , \quad l = F^T x'$$

Epipoli:

$$e \text{ ed } e' \text{ t.c. } Fe = 0 \quad , \quad F^T e' = 0$$

Condizione di corrispondenza epipolare fra punti:

$$x'^T Fx = 0$$

Epipoli e matrici camera:

$$e' = P'C \quad e = PC'$$

Fondamentale da matrici camera generiche:

$$F = [e']_{\times} P'P^+$$

Fondamentale da matrici camera canoniche ($P = K \begin{bmatrix} I & | & 0 \end{bmatrix}$, $P' = \begin{bmatrix} M & | & m \end{bmatrix}$):

$$F = [e']_{\times} M = M^{-T} [e]_{\times} \quad (e' = m \quad , \quad e = M^{-1}m)$$

Matrice essenziale:

$$E = K'^T FK$$

Omografia fra rette epipolari:

$$H = [e']_{\times} F + e'a^T$$

B.3. Tensore trifocale

Tensore dalle camere canoniche ($P = [I | 0]$, $P' = [A | \mathbf{a}_i]$ e $P'' = [B | \mathbf{b}_i]$):

$$T_i = \mathbf{a}_i \mathbf{b}_i^T - \mathbf{a}_i \mathbf{b}_i^T$$

Epipoli \mathbf{e}' ed \mathbf{e}'' :

$$\mathbf{e}' \text{ ed } \mathbf{e}'' \text{ t.c. } \mathbf{e}'^T | \mathbf{u}_1 \ \mathbf{u}_2 \ \mathbf{u}_3 | = 0, \ \mathbf{e}''^T | \mathbf{v}_1 \ \mathbf{v}_2 \ \mathbf{v}_3 | = 0 \quad \text{con } \mathbf{u}_i \text{ e } \mathbf{v}_i \text{ t.c. } \mathbf{u}_i^T T_i = 0 \text{ e } T_i \mathbf{v}_i = 0$$

Matrici fondamentali:

$$F_{21} = [\mathbf{e}']_{\times} [T_1, T_2, T_3] \mathbf{e}'' \quad \text{e} \quad F_{31} = [\mathbf{e}'']_{\times} [T_1^T, T_2^T, T_3^T] \mathbf{e}'$$

Matrici camera (P in forma canonica):

$$P' = [[T_1, T_2, T_3] \mathbf{e}'' \ | \ \mathbf{e}'] \quad P'' = |(\mathbf{e}'' \mathbf{e}''^T - I) [T_1^T, T_2^T, T_3^T] \mathbf{e}' \ | \ \mathbf{e}''|$$

Incidenza retta-retta-retta:

$$(\mathbf{l}^T [T_1, T_2, T_3] \mathbf{l}'') [\mathbf{l}]_{\times} = \mathbf{0}^T \quad \text{(matriciale)}$$

$$(\mathbf{l}, \boldsymbol{\varepsilon}^{ris}) \mathbf{l}'_j \mathbf{l}''_k T_i^{jk} = 0_s \quad \text{(tensoriale)}$$

Incidenza retta-punto-retta:

$$\mathbf{x}^T (\mathbf{l}^T [T_1, T_2, T_3] \mathbf{l}'') = \mathbf{l}^T \left(\sum_i \mathbf{x}^i T_i \right) \mathbf{l}'' = 0 \quad \text{(matriciale)}$$

$$\mathbf{x}^i \mathbf{l}'_j \mathbf{l}''_k T_i^{jk} = 0 \quad \text{(tensoriale)}$$

Incidenza retta-punto-punto:

$$\mathbf{l}^T \left(\sum_i \mathbf{x}^i T_i \right) [\mathbf{x}'']_{\times} = \mathbf{0}^T \quad \text{(matriciale)}$$

$$\mathbf{x}^i \mathbf{l}'_j (\mathbf{x}''^k \boldsymbol{\varepsilon}_{kqs}) T_i^{jq} = 0_s \quad \text{(tensoriale)}$$

Incidenza punto-punto-punto:

$$[\mathbf{x}']_{\times} \left(\sum_i \mathbf{x}^i T_i \right) [\mathbf{x}'']_{\times} = \mathbf{0}_{3 \times 3} \quad \text{(matriciale)}$$

$$\mathbf{x}^i (\mathbf{x}'^j \boldsymbol{\varepsilon}_{jpr}) (\mathbf{x}''^k \boldsymbol{\varepsilon}_{kqs}) T_i^{pq} = 0_{rs} \quad \text{(tensoriale)}$$

Bibliografia

- [Abdel-Aziz 1971] Y. I. Abdel-Aziz and H. M. Karara. *Direct linear transformation from comparator coordinates into object space coordinates in close-range photogrammetry*. Symposium on Close-Range Photogrammetry. Falls Church, American Society of Photogrammetry, pp. 1-18, 1971.
- [Abmayr 2005] T. Abmayr, F. Härtl, M. Reinköster and C. Fröhlich. *Terrestrial Laser Scanning - Applications In Cultural Heritage Conservation and Civil Engineering*. IAPRS Vol. 36, 5/W17, Mestre, Italy, 2005.
- [Alshwabkeh 2004] Y. Alshwabkeh and N. Haala. *Integration of Digital Photogrammetry and Laser Scanning for Heritage Documentation*. IAPRS, Vol. 35(5), Istanbul, Turkey, 2004.
- [Alvarez 1997] L. Alvarez and F. Morales. *Affine morphological multiscale analysis of corners and multiple junctions*. International Journal of Computer Vision, 2(25), pp. 95–107, 1997.
- [Anderberg 1973] M. R. Anderberg. *Cluster analysis for applications*. Probability and Mathematical Statistics, Academic Press, NY, 1973.
- [Armijo 1966] L. Armijo. *Minimization of functions having continuous partial derivatives*. Pacific J. Math., vol. 16, pp. 1-3, 1966.
- [Asada 1986] H. Asada and M. Brady. *The curvature primal sketch*. IEEE Transaction on Pattern Analysis and Machine Intelligence, 8(1), pp. 2-14, 1986.
- [Aschwanden 1992] P. Aschwanden and W. Guggenbuhl. *Experimental results from a comparative study on correlation-type registration algorithms*. Robust Computer Vision, W. Förstner and S. Ruwiedel, Eds. Wichmann, 1992, pp. 268-289.
- [Ballard 1981] D. H. Ballard. *Generalizing the Hough transform to detect arbitrary shapes*. Pattern Recognition, vol. 13, n. 2, pp. 111-122, 1981.
- [Baltsavias 1991] E. P. Baltsavias. *Multiphoto Geometrically Constrained Matching*. Ph. D. thesis, Institute of Geodesy and Photogrammetry, ETH Zurich, Switzerland, Mitteilungen No. 49, 1991.
- [Baltsavias 1996] E. P. Baltsavias, H. Li, S. Mason, A. Stefanidis and M. Sinning. *Comparison of Two Digital Photogrammetric Systems with Emphasis on DTM Generation: Case Study Glacier Measurement*. IAPRS, Vol. 31, Part B4, Vienna, pp. 104-109, 1996.
- [Baumberg 2000] A. Baumberg. *Reliable feature matching across widely separated views*. Conference on Computer Vision and Pattern Recognition, Hilton Head Island, South Carolina, USA, pp. 774–781, 2000.
- [Beardsley 1996] P. A. Beardsley, P. H. S. Torr and A. Zisserman. *3D model acquisition from extended image*

- sequences*. 4th European Conference on Computer Vision, LNCS 1065, Cambridge, pp. 683-695, 1996.
- [Beraldin 2004] J. A. Beraldin. *Integration of Laser Scanning and Close-Range Photogrammetry - The Last Decade and Beyond*. IAPRS, Vol. 35(7), Istanbul, Turkey, pp. 972-983, 2004.
- [Beyer 1992] H. A. Beyer. *Geometric and radiometric analysis of a CCD-camera based photogrammetric close-range system*. Ph.D. thesis. Institute of Geodesy and Photogrammetry, ETH Zurich, Switzerland, Mitteilungen No. 51, 1992.
- [Biasion 2004] A. Biasion, L. Bornaz and F. Rinaudo. *Monitoraggio di eventi franosi in roccia con tecniche di laser-scanner*. Bollettino Sifet, no. 3, pp. 47-60, 2004.
- [Boehler 2004] W. Boehler and A. Marbs. *3D scanning and Photogrammetry for Heritage Recording: a Comparison*. 12th Int. Conf. on Geoinformatics, Gävle, Sweden, 2004.
- [Börlin 2002] N. Börlin. *Improving the robustness of least squares template matching with a line-search algorithm*. IAPRS, Vol. 34/5 WG3, 2002.
- [Börlin 2003] N. Börlin, P. Lindström and J. Eriksson. *A globally convergent Gauss-Newton algorithm for the bundle adjustment problem with functional constraints*. A. Gruen and H. Kahmen (eds.), Optical 3-D Measurement Techniques VI, vol. 2, Wichmann-Verlag, pp. 269-276, 2003.
- [Börlin 2004] N. Börlin, P. Grussenmeyer, J. Eriksson and P. Lindström. *Pro and Cons of Constrained and Unconstrained Formulation of the Bundle Adjustment Problem*. IAPRS, Vol. 35(3), Istanbul, Turkey, 2004.
- [Bornaz 2003] L. Bornaz and S. Dequal. *The solid image: A new concept and its applications*. IAPRS Vol. 34, 5/W12, Ancona, 2003.
- [Brown 1971] D. Brown. *Close-range camera calibration*. Photogrammetric Engineering and Remote Sensing, 37(8), pp 855-866, 1971.
- [Brown 1992] L. G. Brown. *A Survey of Image Registration Techniques*. ACM Computing Surveys, vol. 24(4), pp. 325-376, 1992.
- [Brown 2002] M. Brown and D. G. Lowe. *Invariant features from interest point groups*. British Machine Vision Conferenc, Cardiff, pp. 656-665, 2002.
- [Brovelli 2004] M. A. Brovelli, M. Cannata and U. M. Longoni. *LIDAR data filtering and DTM interpolation within GRASS*. Transaction in GIS, Blackwell Publishing Ltd., 2004.
- [Bruck 1989] H. A. Bruck, S. R. McNeill, M. A. Sutton and W. H. Peters. *Digital image correlation using Newton-Raphson method of partial differential correction*. Experimental Mechanics, vol. 29, pp. 261-267, 1989.
- [Carlsson 1995] S. Carlsson. *Duality of reconstruction and positioning from projective views*. IEEE Workshop on Representation of Visual Scenes, Boston, 1995.
- [Carlsson 1998] S. Carlsson and D. Weinshall. *Dual computation of projective shape and camera positions from multiple images*. International Journal of Computer Vision, 1998.
- [Chaplin 1998] B. A. Chaplin and M. A. Chapman. *A procedure for 3D motion estimation from stereo image sequences for a mobile mapping system*. International Archives of Photogrammetry and Remote Sensing, Vol. 32(3W1), 1998.
- [Chaplin 2001] B. A. Chaplin and M. A. Chapman. *Collection of roadway asset data along an obstructed urban corridor*. 3rd International Symposium on Mobile Mapping Technology, Cairo, Egypt. Session 2:5, 2001.
- [Cooper 1996] M. A. R. Cooper and S. Robson. *Theory of close range photogrammetry*. Close Range Photogrammetry and Machine Vision, K. B. Atkinson (Ed.), Whittles Publishing, Scotland, pp. 9-51,
-

- 1996.
- [Crosilla 1983] F. Crosilla. *Procrustean transformation as a tool for the construction of a criterion matrix for control networks*. Manuscripta geodetica, vol. 8, 1983.
- [Crosilla 1998] F. Crosilla and D. Visintini. *External Orientation of a Mobile Sensor System for Cartographic updating by dynamic vision of digital map points*. Bollettino di Geodesia e Scienze Affini, 1, pp. 41-68, 1998.
- [Crosilla 2005] F. Crosilla, D. Visintini and F. Sepic. *A segmentation procedure of LiDAR data by applying mixed parametric and nonparametric models*. IAPRS 36 (3W19), pp. 132-137, 2005.
- [Crowley 1981] J. Crowley. *A representation for visual information*. PhD. thesis, Carnegie Mellon University, 1981.
- [D'Apuzzo 2003] N. D'Apuzzo. *Surface Measurement and Tracking of Human Body Parts from Multi Station Video Sequences*. Ph. D. thesis, Institute of Geodesy and Photogrammetry, ETH Zurich, Switzerland, Mitteilungen No. 81, 2003.
- [Da Silva 1998] J.F.C. Da Silva and R. A. Oliveira. *Triangulation of a sequence of terrestrial digital images*. International Archives of Photogrammetry and Remote Sensing, Vol. 32(2), pp. 273-277, 1998.
- [Da Silva 2003] J.F.C. da Silva, R. A. Oliveira and B. A. Gallis. *Development of a Low-Cost Mobile Mapping System: a South American Experience*. The Photogrammetric Record, Vol.18 (101) pp. 5 -26, 2003.
- [Deriche 1993] R. Deriche and G. Giraudon. *A computational approach for corner and vertex detection*. International Journal of Computer Vision, 10(2), pp. 101–124, 1993.
- [Dhond 1989] U. Dhond and J. Aggarwal. *Structure from Stereo - a Review*. IEEE Transactions on Systems, Man and Cybernetics, 19(6), 1989.
- [Duda 1972] R. O. Duda and P. E. Hart. *Use of the Hough Transformation to Detect Lines and Curves in Pictures*. Comm. ACM, vol. 15, pp. 11–15, 1972.
- [Eckart 1939] C. Eckart and G. Young. *A principal axis transformation for non-Hermitian matrices*. Bulletin of the American Mathematical Society, vol. 45, pp. 118-121, 1939.
- [Edelman 1997] S. Edelman, N. Intrator and T. Poggio. *Complex cells and object recognition*. Non pubblicato: <http://kybele.psych.cornell.edu/~edelman/archive.html>.
- [El-Hakim 2005] S. El-Hakim, E. Whiting, L. Gonzo and S. Girardi. *3D Reconstruction of Complex Architectures from Multiple Data*. IAPRS Vol. 36, 5/W17, Mestre, Italy, 2005.
- [Faugeras 1990] O. Faugeras, P. Fua, B. Hotz, R. Ma, L. Robert, M. Thonnat and Z. Zhang. *Quantitative and Qualitative Comparison of some Area and Feature-Based Stereo Algorithms*. Robust Computer Vision - Quality of Vision Algorithm, W. Förstner and S. Ruwiedel, pp. 1-26, Eds. Wichmann, 1990.
- [Faugeras 1992a] O. D. Faugeras. *What can be seen in three dimensions with an uncalibrated stereo rig?* Proc. European Conference on Computer Vision, LNCS 588, pp. 563-578. Springer-Verlag, 1992.
- [Faugeras 1992b] O. D. Faugeras, Q. Luong, and S. Maybank. *Camera self-calibration: Theory and experiments*. European Conference on Computer Vision, LNCS 588, pp. 321-334. Springer-Verlag, 1992.
- [Faugeras 1993] O.D. Faugeras. *Three-Dimensional Computer Vision: a Geometric Viewpoint*. MIT Press, 1993.
- [Feng 2001] Q. Feng, P. Sjogren, O. Stphansson and L. Jing. *Measuring fracture orientation at exposed rock faces by using a non-reflector total station*. Engineering Geology, vol. 59, pp. 133-146, 2001.
- [Ferrero 2004] A. M. Ferrero, G. Forlani and I. H. Voyat. *Rock slope stability analysis based on photogrammetric surveys*. IX International Symposium on Landslides, Rio de Janeiro, 2004.
-

- [Fishler 1981] M. Fischler and R. Bolles. *Random sample consensus: a paradigm for model fitting with application to image analysis and automated cartography*. Commun. Assoc. Comp. Mach., Vol. 24(3), pp. 81-95, 1981.
- [Förstner 1987] W. Förstner and E. Gülch. *A fast operator for detection and precise location of distinct points, corners and centres of circular features*. Intercommision Conference on Fast Processing of Photogrammetric Data, Interlaken, Switzerland, pp. 281-305, 1987.
- [Fraser 2001] C. S. Fraser. *Photogrammetric Camera Component Calibration: A Review of Analytical Techniques*. Calibration and Orientation of Camera in Computer Vision, A. Gruen and T. S. Huang (Eds.), Springer-Verlag Berlin Heidelberg, pp. 95-121, 2001.
- [Fryer 2005] J. Fryer, J. H. Chandler and S. El-Hakim. *Recording and Modeling an Aboriginal Cave Painting: With or Without Laser Scanning?* IAPRS Vol. 36, 5/W17, Mestre, Italy, 2005.
- [Forkuo 2004] E. K. Forkuo and B. King. *Automatic Fusion of Photogrammetric Imagery and Laser Scanner Point Clouds*. IAPRS, Vol. 35(4), Istanbul, Turkey, 2004.
- [Forlani 1986] G. Forlani. *Sperimentazione del nuovo programma CALGE dell'ITM*. Bollettino SIFET no. 2, pp. 63-72, 1986.
- [Forlani 1994] G. Forlani and L. Pinto. *Experiences of combined block adjustment with GPS data*. International Archives of Photogrammetry and Remote Sensing, vol. 30 Part 3/1, Monaco, pp. 219-226, 1994.
- [Forlani 2005] G. Forlani, R. Roncella and F. Remondino. *Structure and Motion Reconstruction of Short Mobile Mapping Image Sequences*. Optical 3-D Measurement Techniques, vol. 1 pp. 265-274, Vienna, Austria, 2005.
- [Golub 1989] G. H. Golub and C. F. Van Loan. *Matrix Computations*. The John Hopkins University Press, Baltimore, MD, second edition, 1989.
- [Grün 1985] A. Grün. *Adaptive least squares correlation - a powerful image matching technique*. South African Journal of Photogrammetry, Remote Sensing and Cartography, Vol. 14 (3), pp.175-187, 1985.
- [Grün 1988] A. Grün and E. P. Baltsavias. *Geometrically Constrained Multiphoto Matching*. PERS, vol. 54, no. 5, pp 663-671, 1988.
- [Grün 2003] A. Grün, F. Remondino and L. Zhang. *Modeling and visualization of the Great Buddha statue in Bamiyan, Afghanistan*. IAPRS, Vol. 34 Part 5/W10, Tarasp-Vulpera, Switzerland, 2003.
- [Girardeau 2005] D. Girardeau-Montaut, M. Roux, R. Marc and G. Thibault. *Change detection on point cloud data acquired with a ground laser scanner*. IAPRS 36 (3W19), pp. 85-90, 2005.
- [Guarnieri 2005] A. Guarnieri, F. Pirotti, M. Pontin and A. Vettore. *Combined 3D Surveying Techniques for Structural Analysis Applications*. IAPRS Vol. 36, 5/W17, Mestre, Italy, 2005.
- [Habib 2004] A. Habib, M. Ghanma and M. Tait. *Integration of LIDAR and Photogrammetry for Close Range Applications*, IAPRS, Vol. 35(5), Istanbul, Turkey, 2004.
- [Harris 1988] C. Harris and M. Stephens. *A combined corner and edge detector*. Alvey Vision Conference, pp. 147-151, 1988.
- [Hartigan 1975] J. Hartigan. *Clustering Algorithms*, Wiley, 1975.
- [Harrison 2000] J. P. Harrison and T. R. Reid. *A semi-automated methodology for discontinuity trace detection in digital images of rock mass exposures*. International Journal of Rock Mechanics & Mining Sciences, vol. 37, pp. 1073-1089, 2000.
- [Hartley 1992] R. I. Hartley, R. Gupta, and T. Chang. *Stereo from uncalibrated cameras*. Proc. IEEE Conference on Computer Vision and Pattern Recognition, 1992.
-

-
- [Hartley 1993] R. I. Hartley. *Cheirality invariants*. DARPA Image Understanding Workshop, pp. 745-753, 1993.
- [Hartley 1994a] R. I. Hartley. *Projective reconstruction from line correspondence*. IEEE Conference on Computer Vision and Pattern Recognition, 1994.
- [Hartley 1994b] R. I. Hartley. *Euclidean reconstruction from uncalibrated views*. J. Mundy, A. Zisserman, and D. Forsyth (eds.), Applications of Invariance in Computer Vision, LNCS 825, pp. 237-256. Springer-Verlag, 1994.
- [Hartley 1997a] R. Hartley and P. Sturm. *Triangulation*. Computer Vision and Image Understanding: CVIU, 68(2), pp. 146–157, 1997.
- [Hartley 1997b] R. Hartley. *In defense of the eight-point algorithm*. IEEE Trans. Pattern Anal. Machine Intell., vol. 19, pp. 580-593, 1997.
- [Hartley 2000] R. I. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, Cambridge, 2000.
- [Heitger 1992] F. Heitger, L. Rosenthaler, R. von der Heydt, E. Peterhans and O. Kuebler. *Simulation of neural contour mechanism: from simple to end-stopped cells*. Vision Research, 32(5) pp 963-981, 1992.
- [Heraldick 1991] R.M. Haralick, C. Lee, K. Ottenberg and M. Nölle. *Analysis and Solutions of the Three Point Perspective Pose Estimation Problem*. Proc. IEEE Conf. Computer Vision and Pattern Recognition, pp. 592-598, Maui, Hawaii, 1991.
- [Horn 1987] B. K. P. Horn. *Closed-form solution of absolute orientation using unit quaternions*. Journal of the Optical Society of America, vol. 4, pp. 629-642, 1987.
- [Horn 1990] B. K. P. Horn. *Relative orientation*. International Journal of Computer Vision, vol. 4, pp. 59-78, 1990.
- [Hough 1962] P. V. C. Hough. *Method and means for recognizing complex patterns*. U. S. Patent 3.069.654, 1962.
- [Huang 1989] T. S. Huang and O. D. Faugeras. *Some properties of the E-matrix in two-view motion estimation*. IEEE Transactions on Pattern Analysis and Machine Intelligence, (11), pp. 1310-1312, 1989.
- [Huang 1994] T. S. Huang and A. Netravali. *Motion and Structure from Feature Correspondences: a Review*. Proceedings of the IEEE, 82(2), 1994.
- [Hyypä 2005] H. Hyypä, X. Yu, J. Hyypä, H. Kaartinen, S. Kaasalainen, E. Honkavaara and P. Rönholm. *Factors affecting the quality of DTM generation in forested areas*. IAPRS 36 (3W19), pp. 85-90, 2005.
- [Kadir 2001] T. Kadir and M. Brady. *Scale, saliency and image description*. International Journal of Computer Vision, 45(2), pp. 83-105, 2001.
- [Kadobayashi 2004] R. Kadobayashi, N. Kochi, H. Otani and R. Furukawa. *Comparison and Evaluation of Laser Scanning and Photogrammetry and their Combined Use for Digital Recording of Cultural Heritage*. IAPRS, Vol. 35(5), Istanbul, Turkey, 2004.
- [Kemeny 2003] J. Kemeny and R. Post. *Estimating three dimensional rock discontinuity orientation from digital images of fracture traces*. Computers and Geosciences, vol. 29, pp. 65-77, 2003.
- [Kim 2004] S. Kim, K. Choi, S. Lee, J. Choi, T. Hwang, B. Jang and J. Lee. *A Bimodal Approach for Land Vehicle Localization*. ETRI Journal, vol. 26 (5), pp. 497-500, 2004.
- [Kimme 1975] C. Kimme, D. Ballard and J. Sklansky. *Finding circles by an array of accumulators*. Commun. ACM18(2), pp. 120–122, 1975.
-

- [Klein 1939] F. Klein. *Elementary Mathematichs from an Advanced Standpoint*. Macmillan, New York, 1939.
- [Koenderink 1984] J. J. Koenderink. *The structure of images*. Biological Cybernetics, vol. 50, pp. 363-396, 1984.
- [Kraus 1982] K. Kraus. *Photogrammetry*. Ummmler/Bonn, vol. 1-2, Germany, 1982.
- [Lee 2004] I. Lee and Y. Choi. *Fusion of Terrestrial Laser Scanner Data and Images for Building Reconstruction*. IAPRS, Vol. 35(5), Istanbul, Turkey, 2004.
- [Lewis 1995] J. P. Lewis. *Fast normalized cross-correlation*. Vision Interface, pp.120-123, 1995.
- [Lindeberg 1994] T. Lindeberg. *Scale-space theory: A basic tool fro analysing structures at different scales*. Journal of Applied Statistics. Vol. 21 (no. 2), pp 225-270, 1994.
- [Lindenbergh 2005] R. Lindenbergh, N. Pfeifer and T. Rabbani. *Accuracy analysis of the Leica HDS3000 and feasibility of tunnel deformation monitoring*. IAPRS 36 (3W19), pp. 24-29, 2005.
- [Longuet Higgins 1981] H. C. Longuet Higgins. *A computer algorithm for reconstructing a scene from two projections*. Nature, pp. 133-135, 1981.
- [Lowe 1991] D. Lowe. *Fitting Parameterized Three-Dimensional Models to Images*. IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 13, no. 5, pp. 441-450, 1991.
- [Lowe 1999] D. G. Lowe. *Object recognition from local scale-invariant features*. International Conference on Computer Vision, Corfù, pp. 1150-1157, 1999.
- [Lowe 2004] D. G. Lowe. *Distinctive image features from scale-invariant keypoints*. International Journal of Computer Vision, 2004.
- [Lu 2000] H. Lu and P. D. Cary. *Deformation measurements by digital image correlation: implementation of a second-order displacement gradient*. Experimental Mechanics, Vol. 40, pp. 393-400, 2000.
- [Luong 1996] Q. T. Luong and O. Faugeras. *The fundamental matrix: theory, algorithms and stability analysis*. International Journal of Computer Vision, 17(1), pp 43-76, 1996.
- [Martin 1995] J. Martin and J.L. Crowley. *Experimental comparison of correlation techniques*. In Proc. International Conf. on Intelligent Autonomous Systems, 1995.
- [Medioni 1987] G. Medioni and Y. Yasumoto. *Corner detection and curve representation using cubic B-splines*. Computer Vision, Graphics and Image Processing, volume 39, pp. 267-278, 1987.
- [Mikolajczyk 2001] K. Mikolajczyk and C. Schmid. *Indexing based on scale invariant interest points*. 8th International Conference on Computer Vision, Vancouver, Canada, pp. 525-531, 2001.
- [Mikolajczyk 2003] K. Mikolajczyk and C. Schmid. *A performance evaluation of local descriptors*. IEEE Computer Society Conference on Computer Vision and Pattern Recognition - Volume 2, 2003.
- [Mohr 1993] R. Mohr, F. Veillon, and L. Quan. *Relative 3D reconstruction using multiple uncalibrated images*. IEEE Conference on Computer Vision and Pattern Recognition, pp. 543-548, 1993.
- [Mohr 1996] R. Mohr and B. Triggs. *Projective geometry for image analysis*. Technical report, International Society for Photogrammetry and Remote Sensing, Vienna Congress, July 1996. WG III/2 Tutorial.
- [Monte 2004] J. M. Monte. *Rock Mass characterization using laser scanner and digital imaging data collection techniques*. Master Degree Thesis, University of Arizona, 2004.
- [Moravec 1977] H. P. Moravec. *Towards automatic visual obstacle avoidance*. Proceedings of the 5th International Joint Conference on Artificial Intelligence, Cambridge, Massachussets, 1977.
- [Mundy 1992] J. Mundy and A. Zisserman. *Geometric Invariance in Computer Vision*. MIT Press, 1992.
- [Nardinocchi 2003] C. Nardinocchi, G. Forlani and P.Zingaretti. *Classification and Filtering of laser data*.
-

- IAPRS, vol. 34 Part. 3/W13, pp. 79-86, 2003.
- [Pilu 1997] M. Pílu. *Uncalibrated stereo correspondences by SVD*. Proc. CVPR, pp. 261-266, 1997.
- [Pollefeys 1996] M. Pollefeys, L. Van Gool, and A. Oosterlinck. *The modulus constraint: a new constraint for self-calibration*. International Conference on Pattern Recognition, pp. 31-42, 1996.
- [Pollefeys 1999] M. Pollefeys, R. Koch and L. Van Gool. *A simple and efficient rectification method for general motion*. International Conference on Computer Vision, Corfu, pp. 496-501, 1999.
- [Press 1988] W. Press, B. Flannery, S. Teukolsky and W. Vetterling. *Numerical Recipes in C*. Cambridge University Press, 1988.
- [Price 1985] K.E. Price. *Relaxation matching techniques - a comparison*. PAMI, 7(5), pp. 617-623, 1985.
- [Quan 1999] L. Quan and Z. Lan. *Linear N-Point Camera Pose Determination*. IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 21, no. 7, pp. 1-1999.
- [Rabbani 2005] T. Rabbani and F. van den Heuvel. *Efficient Hough transform for automatic detection of cylinders in point clouds*. IAPRS 36 (3W19), pp. 60-65, 2005.
- [Remondino 2005] F. Remondino, A. Guarnieri and A. Vettore. *3D Modeling of Close-Range Objects: Photogrammetry or Laser Scanning?* Videometrics IX – Electronic Imaging – IS&T/SPIE 17th Annual Symposium, pp. 216-225, 2005.
- [Reisfeld 1995] D. Reisfeld, H. Wolfson and Y. Yeshurun. *Context free attentional operators: the generalized symmetry transform*. International Journal of Computer Vision, vol. 14, pp. 119-130, 1995.
- [Roggero 2002] M. Roggero. *Object segmentation with region growing and principal component analysis*. International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences, 34(3A), pp. 289-294, 2002.
- [Roncella 2005a] R. Roncella, G. Forlani and F. Remondino. *Photogrammetric bridging of GPS outages in mobile mapping*. Videometrics IX – Electronic Imaging – IS&T/SPIE 17th Annual Symposium, 2005.
- [Roncella 2005b] R. Roncella and G. Forlani. *Determinazione automatica della larghezza della carreggiata dall'analisi di una sequenza stereo*. Convegno nazionale SIFET, Mondello (PA), 2005.
- [Roncella 2005c] R. Roncella, G. Forlani and F. Remondino. *Photogrammetry for geological applications: automatic retrieval of discontinuity orientation in rock slopes*. Videometrics IX – Electronic Imaging – IS&T/SPIE 17th Annual Symposium, pp. 17-27, 2005.
- [Roncella 2005d] R. Roncella and G. Forlani. *Photogrammetry and Laser Scanner: a Metric Reliability Comparison*. Workshop Italy-Canada, Padova, Italy, 2005.
- [Roncella 2005e] R. Roncella, E. Romeo and G. Forlani. *Image Based Microstructural Analysis in Fracture Mechanics. Optical 3-D Measurement Techniques*, vol. 2, pp. 51-60, Vienna, Austria, 2005.
- [Rosenfeld 1976] A. Rosenfeld, R. Hummel and S. Zucker. *Scene labeling by relaxation operations*. IEEE Transactions on Systems, Man and Cybernetics 6, pp. 420-433, 1976.
- [Rottensteiner 2002] F. Rottensteiner and C. Briese. *A new method for building extraction in urban areas from high-resolution LIDAR data*. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, vol. 34(3A), pp. 295-301, 2002.
- [Sansò 1973] F. Sansò. *An exact solution of the roto-translation problem*. Photogrammetria, vol. 29, 1973.
- [Sarti 2002] A. Sarti and S. Tubaro. *Detection and characterisation of planar fractures using a 3d Hough transform*. Signal Process. 82(9), pp. 1269-1282, 2002.
- [Schaffalitzky 2001] F. Schaffalitzky and A. Zisserman. *Viewpoint invariant texture matching and wide baseline stereo*. 8th International Conference on Computer Vision, Vancouver, Canada, pp. 636-643, 2001.

- [Schaffalitzky 2002] F. Schaffalitzky and A. Zisserman. *Multi-view matching for unordered image sets*. 7th European Conference on Computer Vision, Copenhagen, Denmark, pp. 414-431, 2002.
- [Schmid 2000a] C. Schmid and A. Zisserman. *The geometry and matching of lines and curves over multiple views*. International Journal on Computer Vision, 40(3), 199-234, 2000
- [Schmid 2000b] C. Schmid, R. Mohr and C. Bauckhage. *Evaluation of Interest Point Detectors*. In International Journal of Computer Vision, 37(2), 151-172, 2000.
- [Scott 1991] G. Scott and H. Longuet-Higgins. *An algorithm for associating the features of two patterns*. Proc. Royal Society London, volume B244, pp 21-26, 1991.
- [Seiz 2003] G. Seiz. *Ground and satellite-based multi-view determination of 3D cloud geometry*. Ph. D. thesis, Institute of Geodesy and Photogrammetry, ETH Zurich, Switzerland, Mitteilungen No. 80, 2003.
- [Semple 1979] J. G. Semple and G. T. Kneebone. *Algebraic Projective Geometry*. Oxford University Press, 1979.
- [Slob 2004] S. Slob, R. Hack. *3D terrestrial laser scanning as a new field measurement and monitoring technique*. Engineering Geology for Infrastructure Planning in Europe - A European Perspective. R. Hack, R. Azzam and R. Charlier (eds.), Berlin, Springer Verlag, pp. 179-190, 2004.
- [Slob 2005] S. Slob, R. Hack, B. van Knapen, K. Turner and J. Kemeny. *A method for automated discontinuity analysis of rock slopes with 3D laser scanning*. Transportation Research Board - 84th annual meeting, Washington, D.C., 2005.
- [Spetsakis 1991] M. E. Spetsakis and J. Aloimonos. *A multi-frame approach to visual motion perception*. International Journal of Computer Vision, 16(3), pp. 245-255, 1991.
- [Springer 1964] C. E. Springer. *Geometry and Analysis of Projective Spaces*. Freeman, 1964.
- [Stewénius 2005] H. Stewénius, F. Schaffalitzky and D. Nistér. *How hard is 3-view triangulation really?* 10th International Conference on Computer Vision, Beijing, China, 2005.
- [Sutton 1988] M. A. Sutton, S.R. McNeill, J. Jang and M. Babai. *Effects of Subpixel Image Restoration on Digital Correlation Error*. Journal of Optical Engineering, Vol. 27 (10), pp. 870-877, 1988.
- [Tao 1999] C. Tao, M. A. Chapman, N. El-Sheimy, B. Chaplin. *Towards automated processing of mobile mapping image sequences*. International Archives of Photogrammetry and Remote Sensing, Vol. 32(2W1), 1999.
- [Torr 1995] P. H. S. Torr, A. Zisserman and D. W. Murray. *Motion clustering using the trilinear constraint over three views*. In R. Mohr and C. Wu (eds), Europe-China Workshop on Geometrical Modelling and Invariants for Computer Vision, pp. 118-125, Xidan University Press, 1995.
- [Torr 1997] P. H. S. Torr and D. W. Murray. *The Development and Comparison of Robust Methods for Estimating the Fundamental Matrix*. International Journal of Computer Vision, pp. 1-33, Kluwer Academic Publishers Boston, 1997.
- [Torr 1998] P. H. S. Torr and A. Zisserman. *Robust computation and parametrization of multiple view relations*. U Desai, editor, ICCV6, pp. 727-732. Narosa Publishing House, 1998.
- [Torr 2002] P. H. S. Torr. *A Structure and Motion Toolkit in Matlab - Interactive Adventures in S and M*. Microsoft Research Technical Report MSR-TR-2002-56, 2002.
- [Triggs 2000] Bill Triggs, P. McLauchlan, Richard Hartley and A. Fitzgibbon. *Bundle Adjustment - A Modern Synthesis*. B. Triggs, A. Zisserman, R. Szeliski (eds.), Vision Algorithms: Theory and Practice, vol. 1883, pp. 298-372, Springer-Verlag, 2000.
- [Tuytelaars 1999] T. Tuytelaars and L. V. Gool. *Content-based image retrieval based on local affinity invariant regions*. International Conference on Visual Information Systems, pp. 493-500, 1999.
-

-
- [Tuytelaars 2000] T. Tuytelaars and L. V. Gool. *Wide baseline stereo matching based on local, affinely invariant regions*. 11th British Machine Vision Conference, University of Bristol, UK, pp. 412–425, 2000.
- [Ullman 1979] S. Ullman. *The interpretation of Visual Motion*. MIT Press, Cambridge, Massachusetts, 1979.
- [Vallet 2000] J. Vallet, J. Skaloud, O. Koelbl and B. Merminod. *Development of a Helicopter-based integrated system for avalanche and hazard management*. International Archives of Photogrammetry and Remote Sensing. Vol. 33 part B2, Amsterdam, pp. 565-572, 2000.
- [VonSanden 1908] H. von Sanden. *Die Bestimmung der Kernpunkte in der Photogrammetrie*. PhD thesis, University of Göttingen, 1908.
- [Vosselman 2001a] G. Vosselman and H. G. Maas. *Adjustment and filtering of raw laser altimetry data*. OEEPE Official Publication, vol. 40, pp. 62-72, 2001.
- [Vosselman 2001b] G. Vosselman and S. Dijkman. *3D building model reconstruction from point clouds and ground planes*. IAPRS 34 (3W4), pp. 37-43, 2001.
- [Yuan 1989] J.S.C. Yuan. *A General Photogrammetric Solution for the Determining Object Position and Orientation*. IEEE Trans. Robotics and Automation, vol. 5, no. 2, pp. 129-142, 1989.
- [Wallis 1976] R. Wallis. *An approach to the space variant restoration and enhancement of images*. Proc. Of Symposium on Current Mathematical Problems in Image Science, Naval Postgraduate School, Monterey, California, 1996.
- [Weinshall 1995] D. Weinshall, M. Werman and A. Shashua. *Shape descriptors: Bilinear, trilinear and quadrilinear relations for multi-point geometry and linear projective reconstruction algorithms*. IEEE Workshop on Representation of Visual Scenes, Boston, pp. 58-65, 1995.
- [Weng 1988] J. Weng, N. Ahuja and T. S. Huang. *Closed-form solution and maximum likelihood: A robust approach to motion and structure estimation*. IEEE Conference on Computer Vision and Pattern Recognition, 1988.
- [Witkin 1983] A. P. Witkin. *Scale-space filtering*. 8th International Joint Conference on Artificial Intelligence, Karlsruhe, pp. 1019-1022, 1983.
- [Wolfe 1969] P. Wolfe. *Convergence conditions for ascent methods*. SIAM Review, vol. 11, pp. 226-235, 1969.
- [Zadeh 1965] L. A. Zadeh. *Fuzzy Sets*. Journal of Information and Control, vol. 8, pp. 338-353, 1965.
- [Zadeh 1973] L. A. Zadeh. *Outline of a New Approach to the Analysis of Complex Systems and Decision Processes*. IEEE Trans. Systems, Man and Cybernetics, vol. 3 pp. 28-44, 1973.
- [Zhang 1992] Z. Zhang, J. Zhang, X. Wu, H. Zhang. *Global Image Matching with Relaxation Method*. International Colloquium on Photogrammetry, Remote Sensing and Geographic Information Systems, Wuhan, China, pp. 175-186, 1992.
- [Zheng 2004] S. Zheng, Z. Zhang and J. Zhang. *Image Relaxation Matching Based on Feature Points for DSM Generation*. Geo-Spatial information science, vol.7 no. 4, pp. 243-248, 2004.
-

Ringraziamenti

Al termine di questo percorso formativo e lavorativo mi sento in dovere di ringraziare tutte le persone che hanno contribuito, con i loro insegnamenti, consigli e con il loro supporto morale, alla mia crescita professionale e personale. Un ringraziamento particolare va al Prof. Gianfranco Forlani per la sua guida instancabile, per i numerosissimi spunti intellettuali, per la pazienza dimostrata, per la libertà concessami in questi tre anni nello sviluppare i temi di ricerca che mi interessavano maggiormente e per non avermi mai fatto mancare occasioni di impegno. Ringrazio la Prof.ssa Anna Maria Ferrero e l'Ing. Iris Voyat per aver cercato di farmi capire, nonostante la mia scarsa attitudine alla disciplina della meccanica delle rocce, cosa significhi svolgere analisi di stabilità di pareti rocciose, e avermi offerto numerosi spunti interessanti di riflessione e di applicazione delle tecniche che ho implementato; l'Ing. Fabio Remondino dell'ETH di Zurigo per la fitta corrispondenza, la collaborazione nella stesura di numerosi articoli, e alcune dritte veramente indispensabili per la buona riuscita del mio lavoro; l'Ing. Leandro Bornaz e il Politecnico di Torino per avermi fornito i dati laser della parete di Arnad; il Prof. Niclas Börnin per averci fornito la sua implementazione Matlab di line-search applicato al LSM.

Un ringraziamento particolare va poi a tutto il Dipartimento di Ingegneria Civile dell'Università di Parma per avermi offerto uno spettro di possibili applicazioni per le tecniche implementate pressoché illimitato e per aver creduto così tanto in me in questi ultimi anni; alla regione Valle D'Aosta per il supporto offerto per la ricerca; al Prof. Livio Pinto, all'Ing. Carla Nardinocchi e all'Ing. Marco Scaioni per le numerose chiacchierate e per avermi fatto capire, insieme al Prof. Forlani, cosa significhi condurre una seria attività accademica.

Un ringraziamento finale va poi a tutte le persone a me care che, a modo loro, mi hanno aiutato in questo percorso, pur essendo estranee al mondo dell'università: l'elenco sarebbe veramente troppo lungo e credo esse sappiano quanto io sia loro riconoscente.

