

Ricostruzione automatica di oggetti terrestri con immagini digitali

Luigi Barazzetti (*), Fabio Remondino (**), Marco Scaioni (*)

(*) Politecnico di Milano, Dip. B.E.S.T., P.za Leonardo da Vinci 32, Milano (luigi.barazzetti, marco.scaioni)@polimi.it

(**) Fondazione Bruno Kessler (FBK) - 3D Optical Metrology Group, via Sommarive 18, Trento, remondino@fbk.eu

Riassunto

La ricostruzione automatica di oggetti a partire da immagini digitali è un settore di ricerca di primaria importanza sia in Fotogrammetria che in Computer Vision. Oggi non è ancora disponibile un software commerciale in grado di ricostruire in modalità completamente automatica un modello 3D completo ed accurato di un oggetto senza ricorrere alla sua segnalizzazione con target. In questo lavoro viene presentata una metodologia per (i) l'orientamento automatico di un blocco di immagini terrestri e (ii) la ricostruzione 3D della superficie dell'oggetto mediante algoritmi di matching multi-immagine.

Abstract

Automatic image-based object reconstruction is a topic of primary importance in both Photogrammetry and Computer Vision. Nowadays, a software capable of creating an accurate and detailed 3D model in an automatic way is not available on the market, unless by using targeting. This paper presents an automated methodology (i) to orient a sparse image block and (ii) to reconstruct the object surface with multi-image matching techniques.

1. Introduzione

Il processo di ricostruzione e modellazione 3D di un generico oggetto terrestre può essere realizzato con immagini digitali oppure mediante sensori ottici attivi (laser scanner, sistemi a luce strutturata, ecc.). In quest'ultimo caso, le coordinate 3D dei punti vengono direttamente misurate dal sensore e sono quindi sin da subito disponibili per la modellazione 3D. L'uso di immagini richiede invece una fase di elaborazione necessaria per la derivazione delle coordinate 3D dei punti a partire da osservazioni 2D. Le immagini contengono tutte le informazioni sufficienti per la formazione di un modello 3D, ma la generazione automatica di un modello accurato e dettagliato dell'oggetto è un processo complesso, specialmente nel caso siano state impiegate lunghe basi di presa e/o immagini convergenti, come spesso avviene nei rilievi fotogrammetrici terrestri. In ogni caso, il vantaggio offerto dall'utilizzo delle immagini è notevole: mentre i sensori a scansione richiedono l'utilizzo di una strumentazione ingombrante, costosa e con una limitata flessibilità d'impiego, le immagini possono essere oggi acquisite anche con camere a basso costo e garantiscono prestazioni sufficienti per ottenere modelli 3D aventi un'accuratezza simile a quella offerta dai sensori attivi.

In questo articolo viene presentata una metodologia per la ricostruzione automatica di un oggetto a partire da un blocco di immagini terrestri acquisite con una camera digitale calibrata. L'applicazione di target non è necessaria, in quanto tutte le operazioni sono condotte utilizzando la tessitura dell'oggetto stesso. Tale metodo integra sia tecniche fotogrammetriche (al fine di ottenere una ricostruzione accurata) sia tecniche utilizzate nel settore della Computer Vision (per l'automazione della procedura). Il processo può essere suddiviso in due fasi successive (Fig. 1): (i) la determinazione dei parametri di orientamento esterno (OE) utilizzando punti omologhi nelle immagini estratti con operatori di interesse e (ii) la misura e ricostruzione della superficie dell'oggetto con tecniche avanzate di matching applicate simultaneamente alle immagini del blocco.

Il metodo permette l'analisi di un generico blocco di immagini ma, nel particolare caso di sequenza (ovvero quando è noto l'ordine delle immagini), una procedura per la riduzione del costo computazionale è stata sviluppata ed implementata (Barazzetti *et al.*, 2009).

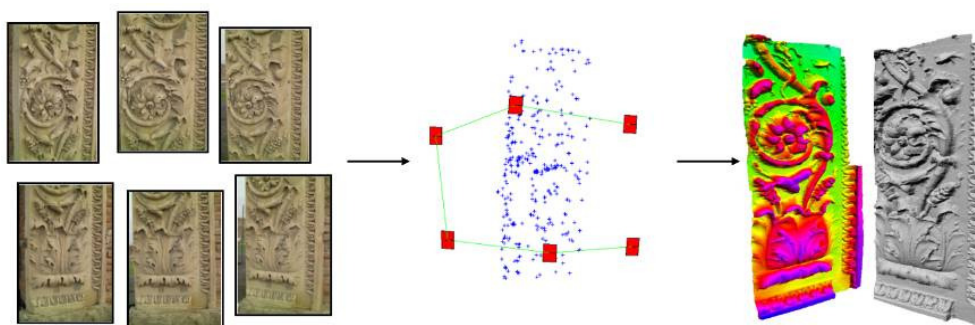


Figura 1 - La procedura automatica di ricostruzione 3D di oggetti o scene da immagini terrestri prevede (i) la determinazione dei punti omologhi e dei parametri di orientamento esterno e (ii) la misura e ricostruzione della superficie dell'oggetto attraverso un matching multi-immagine.

2. Orientamento delle immagini

L'analisi di un generico blocco o di una sequenza di immagini avviene mediante la sua suddivisione in coppie. Le operazioni di identificazione dei punti omologhi sono dapprima condotte tra le coppie di immagini, che vengono progressivamente concatenate in modo da determinare i punti di legame visibili su un numero il più possibile elevato di immagini. Questa procedura consente anche di determinare tutti i valori approssimati necessari per il calcolo finale della triangolazione a *stelle proiettive* dell'intero blocco.

Attualmente la fase più critica nell'orientamento automatico di un blocco di immagini terrestri riguarda l'estrazione automatica dei punti omologhi (si ricorda che i software commerciali possono eseguire tale operazione solo mediante opportuni target codificati). Nel caso in cui si voglia usare la naturale tessitura dell'oggetto o della scena, solamente alcune soluzioni a livello di ricerca sono disponibili. Inoltre, la particolare geometria di un blocco terrestre, con lunghe basi di presa e/o immagini convergenti, è estremamente variabile in funzione dell'oggetto e rende l'operazione di orientamento automatico delle immagini sicuramente più complessa rispetto alle applicazioni di fotogrammetria aerea (ove la TAA ha ormai raggiunto un notevole livello di sviluppo e automazione).

Nel settore della Computer Vision sono stati sviluppati vari operatori capaci di identificare lo stesso punto su immagini riprese da differenti punti di vista, anche con notevoli variazioni dell'assetto della camera. Nella metodologia sviluppata e riportata in questo articolo vengono utilizzati gli operatori SIFT (Lowe, 2004) e SURF (Bay et al., 2008), che estraggono sia punti di interesse (*detector*) che i relativi vettori che descrivono il singolo punto (*descriptor*). La determinazione dei punti omologhi avviene poi confrontando i descrittori (vettori di 128 elementi nella nostra implementazione) mediante criteri di minima distanza. Tale operazione può essere condotta secondo due modalità a discrezione dell'utente e, per una generica coppia di immagini I e J , si può procedere come segue:

- per ogni punto sulla prima immagine I della coppia, l'omologo viene determinato confrontando il suo descrittore con tutti i descrittori della seconda immagine J (cercando la minima distanza). Successivamente, si verifica che la distanza tra il primo candidato sia inferiore a quella dal secondo moltiplicata per una soglia t (ad esempio $t = 0.6$);

- con i descrittori della seconda immagine viene creata una struttura “kd-tree” (Beis & Lowe, 1997), che consente una riduzione del costo computazionale del confronto mediante un’organizzazione delle informazioni più efficiente.

Una volta che i punti omologhi sono stati identificati in tutte le coppie di immagini del blocco o della sequenza di immagini, viene eseguita una stima robusta della matrice fondamentale (o essenziale) per la rimozione di corrispondenze errate (Hartley & Zissermann, 2004). L’algoritmo per la stima della geometria epipolare offre tre differenti modalità di stima (LMedS, MAPSAC, RANSAC), in quanto non vi è un metodo riconosciuto universalmente come il migliore.

2.1. Orientamento di un blocco

Un blocco di N immagini è composto da $(N^2-N)/2$ coppie. Per ciascuna di esse vengono dapprima determinati i punti omologhi come descritto in precedenza e successivamente le varie coppie vengono concatenate mediante un semplice confronto delle coordinate immagine dei punti di legame già localizzati. Al fine di migliorare la misura dei punti di legame dopo la concatenazione delle coppie, può essere applicato un algoritmo di L.S. Matching (LSM, Gruen, 1985) che consente una maggiore precisione di misura delle coordinate immagine.

In Fig. 2 sono riportati i risultati relativi all’orientamento di un blocco di 23 immagini (3872×2592 pixel) su 3 strisciate. In totale, sono state analizzate 253 combinazioni e, mediamente, per ogni immagine (elaborata alla risoluzione originale) sono stati estratti circa 23000 punti (con un tempo di elaborazione di circa 17” per immagine). La comparazione dei descrittori mediante il kd-tree ha richiesto circa 5” per ogni coppia e l’elaborazione complessiva del blocco è stata condotta in meno di 3^h. La stessa procedura ripetuta senza kd-tree avrebbe richiesto circa 435” per la singola coppia e un tempo di elaborazione di oltre 30^h (in tutte le analisi presentate nell’articolo è stato utilizzato un Intel Core 2 CPU con 3 GB di RAM). E’ inoltre importante evidenziare come parte del codice sia stato sviluppato in ambiente MATLAB, il che implica un notevole aumento dei tempi di elaborazione. L’orientamento del blocco è stato eseguito mediante una classica triangolazione fotogrammetrica *a stelle proiettive*, in modo tale da ottenere una soluzione precisa ed affidabile, a differenza di quelle offerte dalle tecniche di orientamento sviluppate nel settore della Computer Vision (“Structure from Motion”), adatte alla formazione di modelli con soli scopi di visualizzazione.

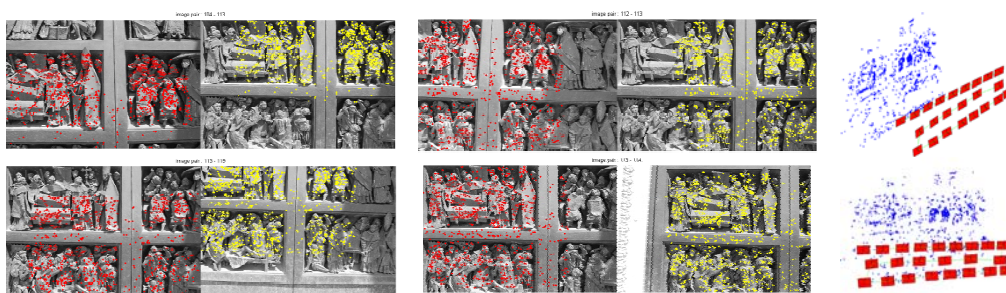


Figura 2 – Coppie di immagini con relativi punti omologhi estratti automaticamente ed i risultati dell’orientamento.

2.2. Orientamento di una sequenza

Per il particolare caso di sequenza ordinata di immagini (Fig. 3) è stata sviluppata una procedura ad-hoc in grado di ridurre notevolmente il costo computazionale. Tale metodo si basa sulla suddivisione della sequenza in triplette di immagini. Ogni tripletta T_i è formata dalla immagini $\{I_i, I_{i+1}, I_{i+2}\}$ (per una sequenza di N immagini sono necessarie $N-2$ triplette). Il matching avviene tramite le due coppie esterne di immagini della singola tripletta $C_i^1 = \{I_i, I_{i+1}\}$ e $C_i^3 = \{I_{i+1}, I_{i+2}\}$ in modo tale da determinare i loro punti omologhi tramite il confronto dei descrittori. Il matching della

rimanente coppia $C_i^2 = \{I_i, I_{i+2}\}$ può essere invece eseguito mediante l'analisi delle coordinate immagine delle due immagini esterne (I_i e I_{i+2}) che compaiono anche nell'immagine I_{i+1} . Dopo aver eseguito il matching per tutte le triplette, è possibile eseguire una progressiva concatenazione considerando che una generica tripletta T_i e la successiva $T_{i+1} = \{I_{i+1}, I_{i+2}, I_{i+3}\}$ hanno due immagini in comune, e quindi un confronto condotto sui soli valori numerici delle coordinate immagine risulta sufficiente. Il grande vantaggio offerto da tale metodo consiste nella riduzione del costo computazionale che, se nel caso di blocchi sparsi dipende da N^2 , in tal modo dipende da N .

In Figura 3 sono riportati i risultati relativi all'orientamento esterno di una sequenza composta da 33 immagini (3872×2592 pixel), elaborate anche in questo caso senza alcuna compressione. Dopo l'estrazione dei punti omologhi, la stima robusta della matrice fondamentale è stata necessaria per rimuovere automaticamente dei punti anomali dovuti al cambiamento della scena fotografata (es. persone e piccioni). La particolare geometria di acquisizione ad anello rende opportuna l'aggiunta di una tripletta virtuale $T_v = \{I_{N-1}, I_N, I_1\}$ che consente di chiudere il percorso con un conseguente miglioramento dei risultati durante la compensazione finale ai minimi quadrati. L'intera elaborazione ha richiesto circa 30'.

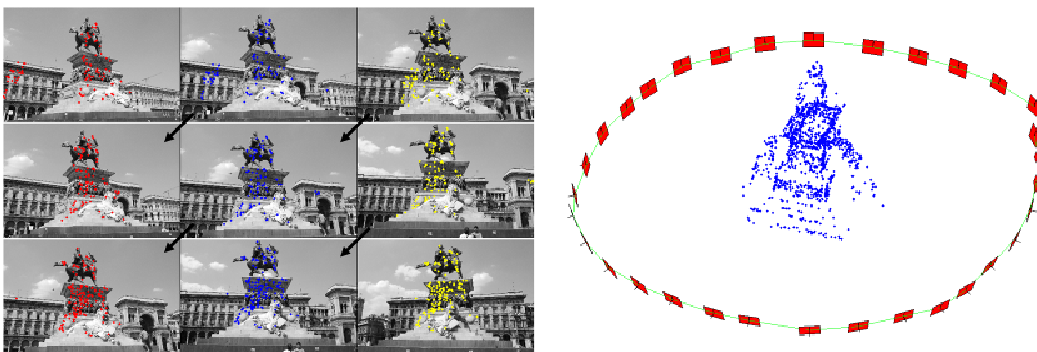


Figura 3 - Tre triplette consecutive con i rispettivi punti omologhi e i risultati relativi all'orientamento esterno.

3. Una strategia per migliorare la precisione dell'orientamento

La qualità dell'orientamento eseguito con il metodo sviluppato può essere verificata mediante un confronto con le attuali tecniche utilizzate per l'orientamento di blocchi terrestri. In tal caso, lo stato dell'arte consiste in misure interattive eseguite da un operatore esperto. Scelte due sequenze di riferimento che sono state orientate per via automatica e manuale, le statistiche della compensazione a stelle proiettive sono state confrontate. Nelle tabelle 1 e 2 sono riportati alcuni risultati ($\sigma_{x,y,z}$ sono le precisione medie teoriche) che mostrano come l'orientamento eseguito per via manuale sia migliore di quello ottenuto per via automatica (combinazione features SIFT o SURF con LSM). Tali risultati sono dovuti alla molteplicità dei singoli punti: infatti l'operatore umano è capace di misurare lo stesso punto di legame su un numero superiore di immagini rispetto a quanto ottenibile con gli operatori SIFT e SURF.

Per migliorare la precisione dell'orientamento automatico è dunque stata sviluppata una procedura basata sull'operatore FAST (Rosten & Drummond, 2006). La ripetibilità di tale operatore è superiore rispetto a SIFT e SURF, tuttavia, non disponendo di un descrittore, non consente di determinare i punti omologhi mediante la procedura proposta. Una nuova procedura di matching che utilizza i parametri di OE calcolati è stata quindi implementata. L'approccio utilizzato consiste in un LSM vincolato alle linee epipolari. I nuovi punti possono poi essere utilizzati per un nuovo processamento del blocco utilizzando come prima approssimazione i parametri di OE stimati in precedenza. I vari risultati sperimentali hanno mostrato una precisione dell'orientamento simile o addirittura superiore rispetto alle misure manuali, e sono alquanto simili a quelli proposti nelle Tabelle 1 e 2.

1	6 immagini - 2816×2112 (pix) dimensioni: 1,2×1×1 (m)				
	manuale	Feat.+LSM	Feat.+LSM & FAST+LSM	FAST+LSM	
3D points	26	61	602	541	
σ_0 (pix)	0,28	0,32	0,28	0,28	
RMS (pix)	0,25	0,35	0,32	0,32	
σ_x (mm)	0,17	0,29	0,27	0,13	
σ_y (mm)	0,14	0,45	0,43	0,24	
σ_z (mm)	0,35	0,71	0,37	0,33	
2	9 immagini - 1856×1392 (pix) dimensioni: 7×1,5×2 (m)				
	manuale	Feat.+LSM	Feat.+LSM & FAST+LSM	FAST+LSM	
3D points	20	185	586	401	
σ_0 (pix)	0,45	0,29	0,54	0,56	
RMS (pix)	0,44	0,28	0,60	0,69	
σ_x (mm)	2,3	2,6	2,3	1,2	
σ_y (mm)	1,5	1,4	1,3	0,9	
σ_z (mm)	3,8	5,9	5,9	3,3	

Tabelle 1 e 2 - Statistiche relative all'orientamento mediante le features SIFT o SURF ed i corner FAST. In entrambi i casi è stato anche utilizzato il LSM.

4. Ricostruzione automatica della superficie dell'oggetto

I parametri di orientamento esterno stimati in modalità automatica sono poi utilizzati per la misurazione e ricostruzione della superficie dell'oggetto o scena in esame mediante un algoritmo di matching multi-immagine basato sulla procedura di Gruen e Baltsavias (1988). L'algoritmo, sviluppato inizialmente da Zhang e Gruen (2004) per immagini satellitari è stato poi adattato per altri sensori ed immagini convergenti terrestri (Remondino *et al.*, 2008) ed incluso nel software CLORAMA. Il software è ora venduto dalla spin-off 4DiXplorer dell'ETH di Zurigo (www.4dixplorer.com). CLORAMA esegue un matching combinando operatori capaci di individuare differenti features nelle immagini (spigoli, bordi, ecc.) e tecniche di matching (cross-correlazione, LSM, matching relazionale, B-Spline matching). In Fig. 4 sono riportati i risultati relativi all'orientamento ed alla ricostruzione delle sequenze analizzate nelle tabelle 1 e 2, mentre in Fig. 5 sono proposti ulteriori esempi nei quali sono state usate anche immagini molto convergenti.

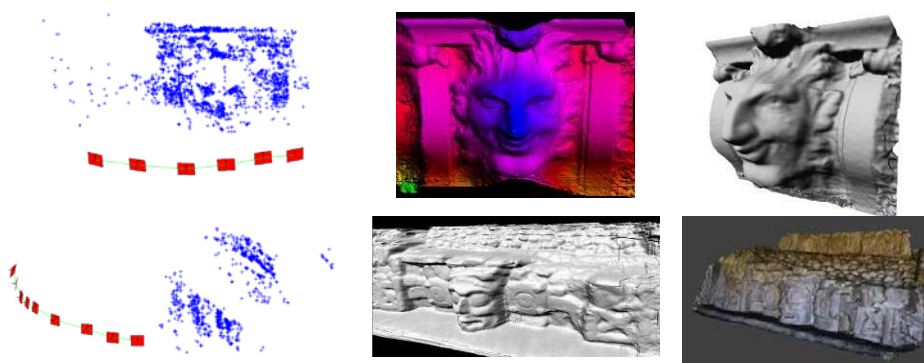


Figura 4 - La ricostruzione delle sequenze proposte nelle tabelle 1 e 2.

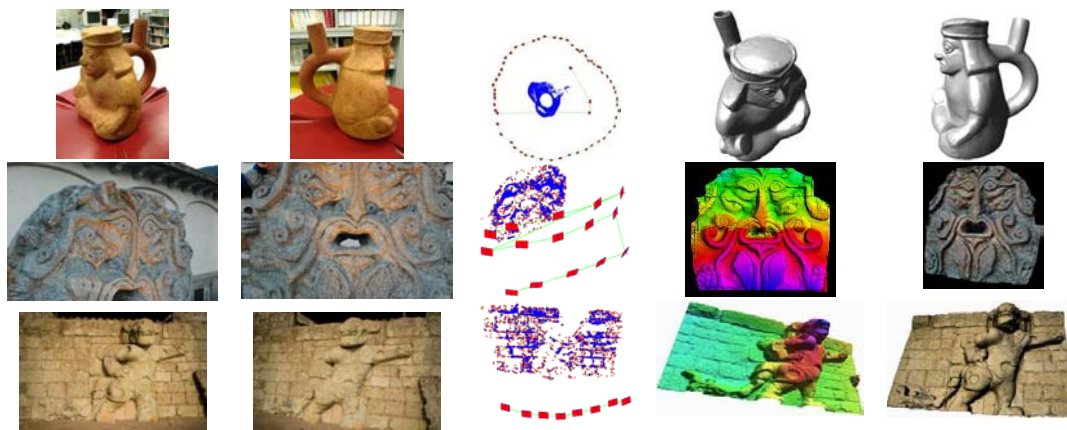


Figura 5 – Alcune immagini utilizzate per il rilievo e i modelli creati con la metodologia proposta.

5. Conclusioni

La ricostruzione 3D automatica di oggetti da immagini digitali può essere efficacemente ottenuta dall'integrazione della fotogrammetria con le tecniche di Computer Vision. L'uso congiunto di varie procedure, combinato con lo sviluppo di alcune nuove strategie proposte nel presente articolo, permette la creazione di modelli accurati e dettagliati in modalità automatica. Sebbene numerosi sensori a scansione siano oggi disponibili e abbiano già raggiunto un significativo livello di maturità tecnologica, in molti progetti la loro scarsa flessibilità ed i costi, congiunti a problematiche legate al trasporto, ne limitano ancora l'impiego. Da questo punto di vista, le immagini costituiscono una valida alternativa con costi nettamente più contenuti. Appare tuttavia evidente come sia necessario disporre di algoritmi automatici per la formazione di un prodotto in termini di accuratezza e completezza simile ai modelli realizzabili con strumenti a scansione.

Riferimenti bibliografici

- Barazzetti L., Remondino F., Scaioni M. (2009), Combined use of photogrammetric and computer vision techniques for fully automated and accurate 3D modelling of terrestrial objects, *Videometrics, Range Imaging and Applications X*, Proc. of SPIE Optics+Photonics, Vol. 7447, 2-3 August, San Diego, CA, USA
- Bay H., Ess. A., Tuytelaars T., Van Gool L. (2008), SURF Speeded Up Robust Features, *CVIU*, 110(3): 346-359
- Beis J., Lowe D. (1997), Shape indexing using approximate nearest-neighbor search in high-dimensional spaces, *Proc. CVPR*, pp. 1000-1006
- Gruen A. (1985), Adaptive least square correlation: a powerful image matching technique, *South African Journal of PRS and Cartography*, 14(3): 175-187
- Gruen A., Baltasavias E. (1988), Geometrically Constrained Multiphoto Matching, *PE&RS*, 54(5): 633-641
- Hartley R., Zisserman A. (2004), *Multiple View Geometry in computer vision*. Cambridge University Press
- Lowe D. (2004), Distinctive image features from scale-invariant keypoints, *IJCV*, 60(2): 91-110
- Remondino F., El-Hakim S., Gruen A., Zhang L. (2008), Development and performance analysis of image matching for detailed surface reconstruction of heritage objects, *IEEE Signal Processing Magazine*, Vol. 25(4), pp.55-65, July 2008
- Rosten E., Drummond T. (2006), Machine learning for high-speed corner detection, *Proc. of ECCV2006*, pp. 430-443
- Zhang L., Gruen A. (2004), Automatic DSM generation from linear array imagery data. *IAPRS&SIS*, 35(3), pp. 128-133