

Una strategia di matching multi-immagine per la ricostruzione accurata di superfici e oggetti 3D

Mattia Previtali, Luigi Barazzetti, Marco Scaioni

Politecnico di Milano, Dip. di Scienza e Tecnologia dell'Ambiente Costruito, P.za Leonardo da Vinci 32, Milano
Laboratorio IC&T, sez. Rilevamento, www.icet-rilevamento.lecco.polimi.it
mattia.previtali@mail.polimi.it, luigi.barazzetti, marco.scaioni@polimi.it

Riassunto

L'articolo descrive una procedura di correlazione multi-immagine per la ricostruzione automatica di superfici ed oggetti 3D mediante immagini digitali. Una ricostruzione approssimata della superficie dell'oggetto è dapprima creata con strategie di correlazione basate sull'estrazione di "patch". Successivamente, tale modello iniziale viene densificato con un algoritmo di *L.S. matching multi-immagine* (MGCM). Tale soluzione consente anche di migliorare la precisione e l'affidabilità dei punti ottenuti e di poter valutare statisticamente la consistenza della soluzione mediante una rigorosa propagazione della varianza. L'output della procedura proposta è una nuvola densa di punti in grado di rappresentare in dettaglio la superficie di un oggetto "free-form". Tale nuvola di punti viene quindi interpolata creando una superficie poligonale che potrà poi essere *texturizzata* ottenendo così un risultato foto-realistico.

Abstract

This paper describes a multi-image matching procedure for the automatic reconstruction of surfaces and 3D objects by using digital images. First, an approximate object surface is created by using matching strategies based on "patch" extraction. Then, this initial model is densified using a *L.S. multi-photo matching* (MGCM) algorithm. This solution allows to improve the precision and the reliability of the extracted points and gives the possibility of statistically evaluating the results' consistency through a rigorous variance propagation. The output of our proposed procedure is a dense point cloud describing "free-form" objects surface in an accurate way. The point cloud is then interpolated to obtain a mesh, which can be textured to obtain a photo-realistic result.

1. Introduzione

Negli ultimi anni un grande sforzo di ricerca è stato rivolto verso la ricostruzione automatica di superfici a partire da immagini digitali. Varie metodologie sono state sviluppate sia nel settore della fotogrammetria che in quello della Computer Vision (CV), ottenendo in quest'ultimo caso procedure in grado di gestire centinaia di immagini e fornire risultati eccellenti per la visualizzazione grafica (Furukawa *et al.*, 2010; Frahm *et al.*, 2010; Strecha *et al.*, 2010).

In ambito fotogrammetrico l'accuratezza metrica e l'affidabilità del prodotto rimangono gli interessi principali. L'uso di camere calibrate, l'acquisizione di immagini secondo un'opportuna geometria di presa e una rigorosa compensazione a stelle proiettive permettono di ottenere parametri di orientamento esterno accurati che consentiranno la ricostruzione dell'oggetto con tecniche di *matching multi-immagine*. L'orientamento di un blocco di immagini terrestri viene generalmente condotto in modalità manuale, anche se oggi giorno un interesse sempre maggiore è rivolto all'automazione con procedure che sfruttano la tessitura dell'oggetto stesso. Recentemente è stata anche rilasciata la nuova versione di PhotoModeler (dapprima chiamata 2010 e poi 2011), che può essere considerato il primo software commerciale, sviluppato per la fotogrammetria terrestre, in

grado di eseguire l'orientamento automatico di immagini senza *target* codificati. Alle stesso tempo, sono stati proposti differenti metodi sviluppati in ambito scientifico (ad esempio Barazzetti *et al.*, 2010; Del Pizzo & Troise, 2011; Deseilligny & Clery, 2011; Roncella *et al.*, 2011) sempre basati sull'estrazione di *feature* invarianti per scala, traslazione e rotazione.

Per quanto riguarda invece la fase di ricostruzione automatica della superficie, nel settore della fotogrammetria dei vicini sono da tempo disponibili alcune soluzioni commerciali (per es. ImageMaster, PhotoModeler Scanner e ShapeScan). Alcuni software derivano direttamente dalla fotogrammetria aerea (es. CLORAMA - Remondino *et al.*, 2008 o LPS eATE - www.erdas.com) e, di conseguenza, nella loro formulazione originaria permettono l'estrazione di un modello digitale della superficie (DSM). Tale rappresentazione 2.5D è adeguata per scopi cartografici ma ha dei limiti evidenti nella modellazione di oggetti tridimensionali.

Lo scopo del presente articolo è la descrizione di una metodologia di ricostruzione delle superfici che sia completamente 3D e che garantisca i requisiti di accuratezza e di analisi statistica dei risultati necessari per un rilievo fotogrammetrico. La metodologia sviluppata (Previtali *et al.*, 2011) si compone essenzialmente di due fasi: innanzitutto viene generato un modello approssimato della superficie dell'oggetto mediante l'algoritmo proposto da Furukawa *et al.* (2010); successivamente, tale modello viene raffinato e densificato con tecniche di *area-based matching* (Grün e Baltsavias, 1988; Baltsavias, 1991). L'impiego di una tecnica multi-immagine permette di ridurre al minimo il numero di errori grossolani e contemporaneamente offre risultati più precisi incrementando il numero di osservazioni. Nel corso del presente lavoro l'attenzione è stata posta verso oggetti "free-form", aventi cioè una geometria rappresentabile con l'interpolazione automatica di nuvole di punti. Non appartengono a questa categoria di oggetti tutti quelli per i quali è necessaria l'estrazione di features al fine di modellare correttamente l'oggetto (ad esempio una facciata architettonica).

2. La procedura di ricostruzione della superficie

Come evidenziato nello schema di Figura 1, è necessario disporre di un blocco di immagini adeguato, sia in termini di geometria di presa delle immagini che di risoluzione delle stesse.

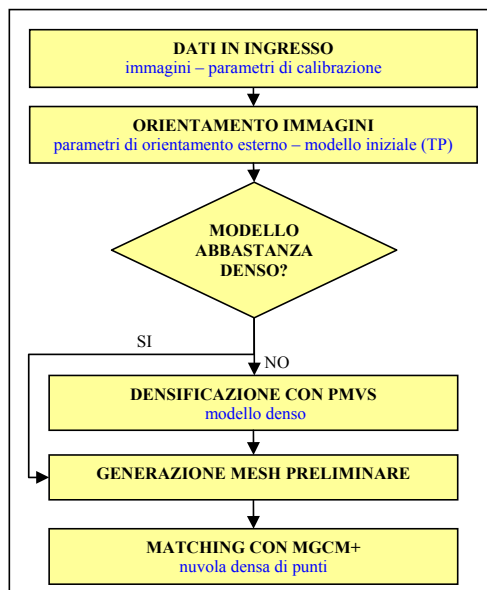


Figura 1 – Schema della procedura di ricostruzione sviluppata.

In particolare, al fine di ottenere un modello finale sufficientemente dettagliato, è bene che ogni porzione dell'oggetto sia visibile in almeno 3-4 immagini in modo da sfruttare le potenzialità del matching multi-immagine. E' inoltre necessario disporre dei parametri di orientamento esterno delle immagini, a cui viene rimossa la distorsione indotta dall'obbiettivo tramite un preliminare progetto di calibrazione della camera. Come già detto, l'orientamento può essere eseguito sia in modalità manuale sia mediante algoritmi di orientamento automatico, anche se ai fini del presente lavoro la seconda procedura ha il vantaggio di fornire molti più punti di legame, che formano quindi una nuvola 3D utilizzabile per la generazione di un modello iniziale della superficie. Nel caso invece si preferisca un orientamento in modalità manuale (oppure se i punti di legame ricavati con gli algoritmi automatici non ricoprono in modo sufficientemente denso l'oggetto) il modello iniziale della superficie viene ricavato mediante l'algoritmo proposto da Furukawa & Ponce (2010). Il principale vantaggio di questa metodologia è la sua capacità di operare direttamente in 3D senza richiedere un modello approssimato. La conversione dei parametri di orientamento esterno fotogrammetrici nel formato tipico utilizzato in CV è descritta in Barazzetti (2010).

La mesh di triangoli così ottenuta viene quindi utilizzata come modello in input per l'algoritmo MGCM. Questo combina (i) il *Least Squares matching* (LSM) tradizionale (Grün, 1985) basato sull'intensità dei toni di grigio e (ii) le equazioni di collinearità usate come vincolo geometrico per la determinazione delle coordinate oggetto di ciascun punto della nuvola. Il LSM classico tiene in considerazione la deformazione geometrica esistente tra un'immagine "master" di riferimento e le immagini di servizio dette "slave" (Baltasvias, 1991). L'introduzione delle equazioni di collinearità e la possibilità di eseguire il *matching* simultaneamente tra diverse immagini consente di ridurre in modo significativo il numero di valori anomali e di aumentare l'affidabilità interna delle corrispondenze tra i punti omologhi sulle varie immagini. L'MGCM presenta alcuni vantaggi rispetto ad altri algoritmi di correlazione di immagine solitamente utilizzati in fotogrammetria. Rispetto al LSM tradizionale, in cui è possibile correlare le sole coppie di immagini, l'MGCM consente di sfruttare la ridondanza offerta dall'impiego di più immagini simultaneamente. Pertanto i problemi di ambiguità legati a tessiture ripetitive possono essere più facilmente risolti, rendendo così la soluzione più affidabile. In ogni caso, anche l'MGCM (nella sua formulazione originaria) presenta dei punti deboli, ovvero (i) la necessità di avere a disposizione valori approssimati per quanto riguarda le coordinate oggetto del punto considerato, le posizioni dei rispettivi punti omologhi nelle varie immagini e i parametri delle trasformazioni geometriche implementate (affini); (ii) la scelta delle immagini tra le quali eseguire la correlazione.

3. L'algoritmo proposto: MGCM+

Come evidenziato nel precedente paragrafo, l'algoritmo MGCM ha alcuni limiti che in questo lavoro si è cercato di risolvere: la generazione di valori approssimati dei parametri e la scelta delle immagini tra cui eseguire il matching. La soluzione sviluppata ha portato all'implementazione di un algoritmo chiamato MGCM+.

3.1 La generazione dei parametri approssimati

Un aspetto da tenere in considerazione riguarda la necessità di una serie di valori approssimati per la posizione dei punti oggetto e le corrispondenti posizioni approssimate nelle varie immagini. Nella metodologia sviluppata la gestione di questo punto critico avviene mediante l'impiego di un modello preliminare dell'oggetto costituito da una *mesh* poligonale generata a partire dall'interpolazione delle coordinate 3D dei punti di legame (se sufficientemente fitti), oppure della serie di punti orientati (*patch*) estratti dal software PMVS. Siccome l'algoritmo deve poi essere utilizzato per la ricostruzione di oggetti 3D, una triangolarizzazione 2.5D (ad esempio quella di Delaunay) non è sufficiente. Si è quindi fatto ricorso all'algoritmo, basato sull'interpolazione con il metodo di Poisson, proposto da Kazhdan *et al.* (2006). Sulla *mesh* così ottenuta vengono quindi identificati una serie di punti, la cui distanza corrisponde al passo di ricostruzione che si vuole ottenere nella nuvola finale. Tali punti rappresentano delle posizioni approssimate per la nuvola 3D

finale. Riproiettando quindi i punti sulle varie immagini a disposizione (per mezzo delle equazioni di collinearità), è possibile estrarre anche i valori approssimati per la posizione dei rispettivi punti omologhi sulle varie immagini da utilizzare.

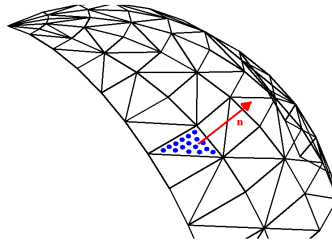


Figura 2 – Posizione approssimata dei punti sulla mesh ottenuta suddividendo i triangoli del modello approssimato.

3.2 Scelta delle immagini da correlare

L'altro punto estremamente critico dell'algoritmo MGCM è costituito dalla scelta delle immagini tra le quali calcolare la correlazione, ovvero la distinzione tra l'immagine "master" e le immagini "slave". Poiché nella maggior parte dei casi l'implementazione di MGCM riguarda le applicazioni aeree, il problema potrebbe essere risolto semplicemente selezionando manualmente come immagine "master" quella centrale all'interno della sequenza e considerando come immagini "slave" le altre. Tale approccio è valido nel caso di immagini con assetto normale, in cui la superficie da misurare può essere considerata come un modello 2.5D. Nelle applicazioni terrestri questa soluzione non può essere considerata come percorribile, almeno per un paio di ragioni:

- (i) in un blocco terrestre che riprende un oggetto a 360° non è possibile definire una immagine centrale della sequenza;
- (ii) a causa delle occlusioni esistenti non ha senso utilizzare tutte le immagini a disposizione per la correlazione: operando in tale modo si rischierebbe di incrementare il costo computazionale e di incorrere in errori grossolani.

Per queste ragioni è necessario operare una selezione tra le immagini, partendo dal problema associato alla definizione dell'immagine "master". Per ogni posizione approssimata sulla *mesh* è possibile definirne la normale ed utilizzare come immagine di riferimento quella che presenta l'asse ottico più prossimo alla normale del punto sulla *mesh* (Fig. 3). Tale definizione di immagine "master" varia quindi da triangolo a triangolo ed è in grado di adattarsi alla forma dell'oggetto e di seguirne i cambiamenti.

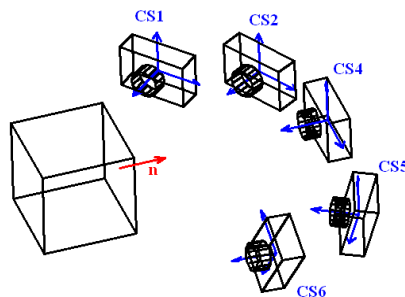


Figura 3 – Scelta dell'immagine master mediante la stima delle normali alla superficie: in questo caso verrà utilizzata l'immagine CS4 per la ricostruzione della faccia.

Il secondo problema riguarda la selezione delle immagini di servizio. Un'ottimizzazione della scelta è necessaria e diviene evidente se si considera il caso di un blocco di immagini riprese attorno ad un oggetto come quello rappresentato in Figura 4. In una situazione di questo tipo, nel caso in cui si stia considerando come "master" l'immagine CS1, non ha evidentemente senso considerare l'immagine CS13 come "slave" a causa delle auto-occlusioni generate dall'oggetto. Solo un numero ridotto di immagini dovrà essere utilizzato in questo caso, ad esempio le immagini CS2-CS3-CS4 in quanto riprendono la stessa porzione di oggetto.

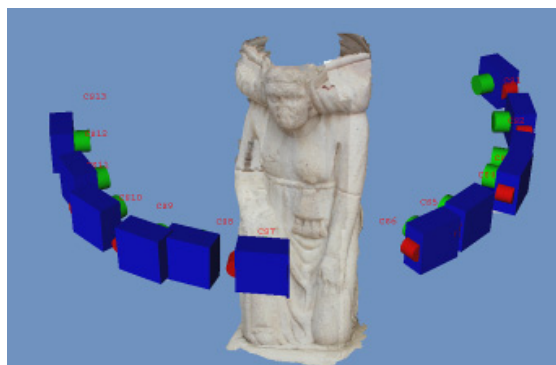


Figura 4 – Geometria di acquisizione delle immagini di un oggetto tridimensionale.

Tale ottimizzazione viene nuovamente operata mediante l'adozione del modello approssimato. Vengono infatti considerate come potenziali "slave" solo le immagini per cui l'angolo formato tra la normale della posizione approssimata e la normale al piano immagine sia inferiore a 90° . Per angoli superiori le immagini risultano essere ocluse e non vengono considerate come potenziali "slave" (Fig. 5).

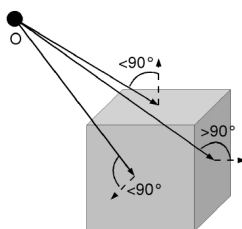


Figura 5 – Principio di selezione delle immagini slaves.

Anche tra le "slave" così individuate viene poi eseguita una selezione ulteriore. Infatti, in molti casi le deformazioni prospettiche possono essere molto consistenti da diventare un serio problema per il LSM tradizionale, che è basato su una trasformazione affine. Per questa ragione, tra le potenziali "slaves" vengono selezionate e successivamente elaborate solo quelle per cui le ipotesi del LSM possano essere considerate come effettivamente valide. In particolare, per ciascuna posizione approssimata viene considerata la deformazione geometrica del triangolo della *mesh* a cui il punto appartiene sulle potenziali "slave". Tale triangolo viene proiettato sia sull'immagine "master" sia sulle potenziali "slave" e vengono poi valutati due parametri di forma:

1. l'area dei triangoli così formati;
2. le variazioni degli angoli delle *mesh* triangolari.

In particolare se l'area di un triangolo su una potenziale "slave" è inferiore al 50% all'area del rispettivo triangolo sull'immagine "master", si ritiene che le deformazioni sulla "slave" dovute a deformazioni prospettiche o a variazioni di scala siano eccessive per la corretta applicazione del LSM. In alcuni casi, importanti variazioni prospettiche possono lasciare sostanzialmente invariata l'area dei triangoli proiettati sulle immagini e per questa ragione vengono tenute in considerazione anche le deformazioni angolari dei triangoli: se (per angoli omologhi) le variazioni angolari superano il 40% l'immagine viene rigettata.

3. Sperimentazione della metodologia sviluppata

La procedura sviluppata è stata sperimentata su alcuni blocchi di immagini. Tutti i blocchi sono costituiti da immagini ad alta risoluzione (almeno 10 Mpixel), che sono state elaborate impiegando la risoluzione originale. L'orientamento esterno è stato eseguito in modalità automatica mediante il software ATiPE (Barazzetti *et al.*, 2010) che permette di estrarre una serie di punti omologhi che sono stati successivamente utilizzati come punti di legame per la compensazione a stelle proiettive del blocco. Una descrizione dei blocchi è riportata nella Tabella 1, dove sono anche indicate le caratteristiche dei sensori impiegati. In particolare, oltre alle camere professionali sono state utilizzate anche camere "compatte" dai costi più contenuti.





Blocco 1	Blocco 2	Blocco 3	Blocco 4
Samsung ST45 6.3 mm 5 immagini 4000×3000 pixel	Olympus FE46 6.3 mm 13 immagini 3968×2976 pixel	Olympus FE46 6.3 mm 15 immagini 3968×2976 pixel	Nikon D80 20 mm 18 immagini 3872×2592 pixel
			

Tabella 1 – Blocchi di immagini utilizzati per testare la procedura sviluppata.

Il primo blocco riguarda un bassorilievo di larghezza 90 cm situato nel Foro Romano a Roma. Per questo oggetto potrebbe anche essere sufficiente una modellazione di tipo 2.5D, basata su un DSM dell'oggetto anziché una *mesh* di triangoli (Fig. 6a). La nuvola di punti ottenuta da 5 immagini convergenti è costituita da circa 500.000 punti. Al termine dell'elaborazione, nella nuvola di punti non sono stati evidenziati evidenti errori grossolani grazie alla robustezza dell'algoritmo MGCM+. Infine, l'interpolazione della nuvola ha fornito una *mesh* di triangoli (Fig. 6b) e un modello fotorealistico (Fig. 6c). Per questo esempio è stato eseguito un confronto con i risultati ottenuti da un *software* commerciale (Leica Photogrammetry Suite LPS – eATE) in grado di estrarre nuvole di punti. Questo software è stato sviluppato per applicazioni cartografiche mediante fotogrammetria aerea ed è basato sull'algoritmo *semi-global matching* (Hirschmüller, 2008). Entrambe le *mesh* sono state allineate usando l'algoritmo di co-registrazione ICP implementato in Geomagic Studio. L'intervallo della barra d'errore è ± 13 mm, mentre lo s.q.m. delle differenze è risultato ± 0.7 mm (Fig. 6d).

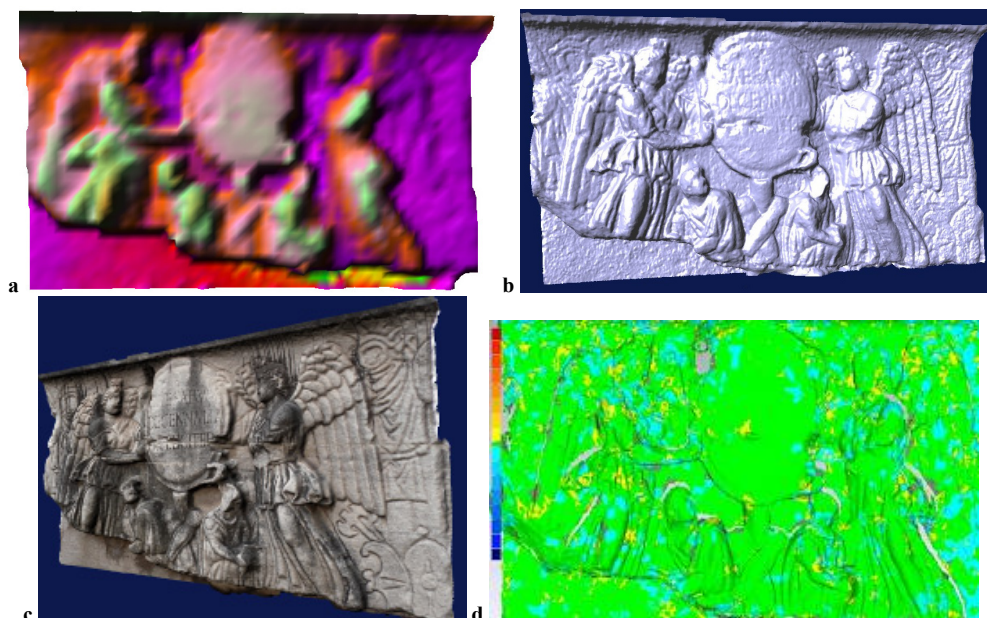


Figura 6 – Blocco 1. DTM approssimato (a), mesh finale (b), modello texturizzato (c) e grafico delle differenze (d).

I blocchi 2 e 3 riguardano due piccole sculture poste ai lati dell'entrata della chiesa di Santa Maria Maggiore (Bergamo). Le immagini sono state riprese attorno a ciascuna delle due statue, determinando quindi immagini caratterizzate da forti occlusioni imputabili alla geometria dell'oggetto stesso. La ricostruzione di tali oggetti nella loro complessità non può quindi che procedere attraverso una procedura completamente 3D come quella sviluppata. In questo caso, come modelli approssimati, sono state utilizzate alcune *mesh* ottenute dall'interpolazione di una serie di *patch* generate con PMVS (Fig. 7a-7d). Le due nuvole di punti sono costituite per il blocco 1 da circa 1M di punti e per il blocco 2 da circa 1.2M di punti. Per ciascuna statua è stata ricavata una *mesh* (Fig. 7b-7e) e un modello finale *texturizzato* (Fig. 7c-7f).

Una particolare attenzione è stata posta in questo caso al problema del *texture-mapping*. Per evitare che la tessitura venisse applicata anche a porzioni dell'oggetto occluse, è stata implementata una analisi di *visibilità* che consente di definire, per ciascuna immagine, solo i triangoli effettivamente visibili. Tale analisi viene condotta analizzando la normale al piano dell'immagine e la normale di ogni triangolo della *mesh*: se l'angolo formato tra le due normali è superiore a 90° , il triangolo viene considerato come occluso. Di seguito in Figura 8 sono riportati i risultati dell'analisi di visibilità.



Figura 7 – Blocchi 2 (in alto) e 3 (in basso): modello approssimato (a-b), mesh finale (c-d), modello texturizzato (e-f).



Figura 8 – Analisi di visibilità per il blocco 3. Vista dalla posizione della camera (a), vista laterale (b).

L'ultimo blocco (4) riguarda un'applicazione geotecnica. Lungo la strada provinciale 65 nel comune di Esino Lario (LC) una parete rocciosa è stata sottoposta ad un monitoraggio in diverse epoche per monitorarne la stabilità e prevenire potenziali rischi per la circolazione stradale. Per questa tipologia di applicazioni l'accuratezza metrica del modello è di primaria importanza. Inoltre, al fine di eseguire delle comparazioni tra le diverse acquisizioni, le immagini devono essere co-registrate tra le varie epoche facendo ricorso ad una serie di punti di controllo incorporati all'interno

della compensazione del blocco. In particolare, per l'epoca iniziale sono state acquisite 22 immagini ottenendo una nuvola di punti costituita da circa 3Mdi punti (Fig. 9a-9b). Il confronto tra la nuvola di punti così ottenuta e quella ricavata con un rilievo laser scanning terrestre, a seguito della rimozione degli elementi di disturbo come la vegetazione, ha rilevato una discrepanza di circa ± 5 mm, pari cioè alla precisione nominale del laser scanner utilizzato (Riegl LMS-Z420i).

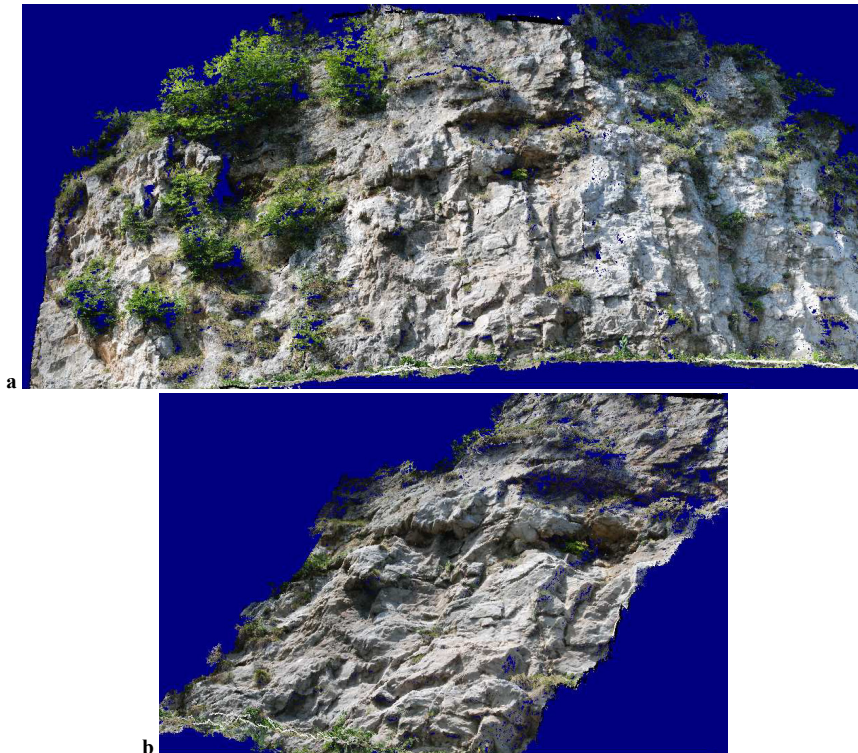


Figura 9 – Blocco 4: nuvola di punti texturizzata della parete in roccia.

4. Conclusioni

Nel presente articolo è stata presentata una procedura per la ricostruzione automatica di oggetti e superfici in fotogrammetria dei vicini con una tecnica di correlazione multi-immagine (MGCM+). Lo scopo del presente lavoro è stato quello di realizzare un algoritmo in grado di ricostruire oggetti “free-form” a partire da immagini ad alta risoluzione riprese aventi caratteristiche adeguate per quanto riguarda la geometria di presa e la copertura dei fotogrammi. Il risultato finale è una procedura completamente automatica, a partire dall'orientamento delle immagini fino alla generazione di una nuvola densa di punti e alla sua interpolazione.

Un aspetto importante di questo lavoro è la combinazione di alcune tecniche sviluppate nell'ambito della *computer vision* con altre nate in quello fotogrammetrico, riguardanti sia la fase di orientamento delle immagini sia la ricostruzione densa del modello finale.

Bibliografia

- Baltsavias, E.P., 1991. Multiphoto Geometrically Constrained Matching. Ph. D. thesis, Inst. of Geodesy and Photogrammetry, ETH Zurich, Switzerland, Mitteilungen No. 49, 221 pp.
- Barazzetti, L., 2010. A trip to Rome: physical replicas of historical objects created in a fully automated way from photos. Proc. of HCITOC. *Lecture Notes in Computer Science*, 6529: 63-72 (2011).
- Barazzetti L., Remondino, F. and Scaioni M., 2010. Extraction of accurate tie points for automated pose estimation of close-range blocks. *ISPRS Technical Commission III Symposium on Photogrammetric Computer Vision and Image Analysis*, 6 pp.
- Del Pizzo, S., Troisi, S. (2011). Automatic orientation of image sequences in cultural heritage. *IAPRS&SIS*, 38(5/W16), on CD-ROM.
- Frahm, J. et al. 2010. Building Rome on a cloudless day. In: Proc. of ECCV 2010, 14 pp.
- Furukawa, Y. and Ponce, J., 2010. Accurate, dense, and robust multi-view stereopsis. *IEEE Trans. PAMI*, 32(8): 1362-1376.
- Furukawa, Y., Curless, B., Seitz, S.M. and Szeliski, R., 2010. Towards Internet-scale Multi-view Stereo. Proc. of IEEE Conf. CVPR'10, 8 pp.
- Hirschmüller, H., 2008. Stereo processing by Semi-Global Matching and Mutual Information. *IEEE Trans. PAMI*, 30(2): 328-341.
- Grün, A., 1985. Adaptive least square correlation – a powerful image matching technique. *S. Afr. J. of Photogrammetry, Remote Sensing and Cartography*, 14(3), pp 175-187.
- Grün, A. and Baltsavias, E.P. 1988. Geometrically Constrained Multiphoto Matching. *PE&RS*, 54(5), pp. 663-671.
- Kazhdan, M., Bolitho, M. and Hoppe, H., 2006. Poisson surface reconstruction, *Proceedings of the fourth Eurographics symposium on Geometry processing 2006*, Eurographics Association, pp. 61-70.
- Previtali, M., Barazzetti, L., Scaioni, M., 2011. Multi-step and multi-photo matching for accurate 3D reconstruction. *Photogrammetric Image Analysis*, 5-7 October, Munich, 6 pp.
- Remondino, F., El-Hakim, S.F., Grün, A. and Zhang, L., 2008. Turning images into 3-D models - development and performance analysis of image matching for detailed surface reconstruction of heritage objects. *IEEE Signal Processing Magazine*, 25(4), pp. 55-65.
- Pierrot Deseilligny, M., Clery, I., 2011. Apero, an open source bundle adjustment software for automatic calibration and orientation of set of images. *IAPRS&SIS*, 38(5/W16), on CD-ROM.
- Roncella, R., Re, C. and Forlani, G., 2011. Performance evaluation of a structure and motion strategy in architecture and cultural heritage. *IAPRS&SIS*, 38(5/W16), on CD-ROM.
- Strecha, C., Pylvanainen, T. and Fua, P., 2010. Dynamic and Scalable Large Scale Image Reconstruction. Proc. of CVPR'10, 8 pp.