

Image-based modeling for object and human reconstruction

Doctoral Thesis

Author(s):

Remondino, Fabio

Publication date:

2006

Permanent link:

<https://doi.org/10.3929/ethz-a-005211924>

Rights / license:

In Copyright - Non-Commercial Use Permitted

IMAGE-BASED MODELING FOR OBJECT AND HUMAN RECONSTRUCTION

A dissertation submitted to the
SWISS FEDERAL INSTITUTE OF TECHNOLOGY (ETH) - ZURICH

for the degree of
Doctor of Technical Sciences

presented by
FABIO REMONDINO
Dipl.Ing., Politecnico di Milano
born 04.02.1974
citizen of Italy

accepted on the recommendation of

Prof. Dr. Armin Grün, examiner
ETH Zurich, Switzerland

Prof. Dr. Petros Patias, co-examiner
Aristoteles University, Thessaloniki, Greece

Zurich 2006

ABSTRACT

The topic of this research is the investigation of the image-based approach for the 3D modeling of close-range scenes, static objects and moving human characters. Three-dimensional (3D) modeling from images is a great topic of investigation in the research community, even if range sensors are becoming more and more a common source and a good alternative for the generation of 3D information. The interest in 3D modeling is motivated by a wide spectrum of applications, such as video games, animation, navigation of autonomous vehicles, object recognition, surveillance and visualization. In particular, the production of 3D models from existing images or old movies would allow the generation of new scenes involving objects or human characters who may be unavailable for other modeling techniques.

Techniques for 3D modeling have been rapidly advancing over the past few years although most of them focus on single objects or specific applications such as architecture or city mapping. Nowadays the accurate and fully automated reconstruction of 3D models from image data is still a challenging problem. Most of the current approaches developed to recover accurate 3D models are based on semi-automatic procedures, therefore the introduction of new reliable and automated algorithms is one of the key goals in the photogrammetric and vision communities. In fact fully automated image-based approaches generally do not work under certain image network configuration or are not reliable enough for some applications, like cultural heritage documentation. Automated image-based methods require good features in multiple images and very short baselines between consecutive frames to extract dense depth maps and complete 3D models. But these requirements are not satisfied in some practical situations, due to occlusions, illumination changes and lack of texture. Automated processes often end up with areas containing too many features that are not all needed for the object modeling and areas with very few features to produce a complete and detailed model. Automated reconstruction methods generally do not report good accuracy, limiting their use for applications that require only nice-looking 3D models. Furthermore, post processing operations are often required, which means that the user interaction is still needed. Therefore fully automated procedures are generally limited in finding point correspondences and camera poses while for the surface measurement phase the user interaction is generally preferred, in particular for architectures.

The image-based modeling of an object should be meant as the complete process that starts from the acquisition system and ends with a virtual model in three dimensions visible interactively on a computer. The photogrammetric modeling pipeline consists of few well known steps: calibra-

tion and orientation, surface measurement and point cloud generation, structuring and modeling of the object geometry, visualization and analysis. Different efforts have been done to increase the level of automation within these steps and broaden the use of the image-based modeling technology. So far, however, the efforts to completely automate the processing, from the image acquisition to the output of a 3D model, are not always successful or not applicable in many 3D modeling projects.

In this dissertation different techniques developed to analyze existing sequence of images and partially automate the process of constructing digital 3D models of static objects or moving human characters are reported. In particular the work investigates if automated and markerless sensor orientation is feasible and under which conditions, if it is possible to recover complete and detailed 3D models of complex objects using automated measurement procedures, which kind of (3D) information can be retrieved from existing image data as well as the capabilities or limits of photogrammetric algorithms in dealing with uncalibrated images and zooming effects. For the investigations, sets of available or self-acquired images, as well as frames digitized from existing monocular videos are used.

The possibility to automatically orient an image sequence heavily depends on the type of images, acquisition and scene. Compared to other research approaches, the developed method for the automated tie point extraction and image orientation relies on accurate feature location achieved using least squares matching measurement algorithm and a statistical analysis of the matched and adjustment results. The reported examples demonstrate its capabilities also for the orientation of images acquired under a wide baseline. A photogrammetric bundle adjustment is always employed to recover the camera parameters and the 3D object coordinates.

On the other hand, the analysis of moving human characters using a monocular video is based on a deterministic approach together with constraints and assumptions on the imaged scene as well as on the human's shape and movement. The developed photogrammetric pipeline can accommodate different input data and different types of human motions. The resulted 3D characters and scene information can be used for visualization or animation purposes or in biometric applications with medium accuracy requirements.

For the automated tie point extraction phase, programs for the feature extraction and the relative orientation between image pairs and triplets were implemented, together with a graphical tool to display the recovered correspondences and epipolar geometry. Concerning the human reconstruction from monocular videos, programs were developed to recover 3D models from single images and combine them under the same reference system in case of image sequence analysis.

RIASSUNTO

Il tema di questa ricerca è l'investigazione del processo di modellazione tridimensionale di oggetti o persone a partire da immagini. La modellazione tridimensionale (3D) basata su immagini è un argomento di ricerca molto investigato anche se recentemente sensori attivi (come il laser scanner) sono molto usati e a volte preferiti alle immagini nei progetti di modellazione tridimensionale a computer.

L'interesse per la modellazione a computer è spinto dalla grande gamma di applicazione esistenti, come i video giochi, la navigazione automatica di veicoli, il riconoscimento automatico di oggetti, la sorveglianza, la documentazione e la visualizzazione. In particolare, la ricostruzione di modelli 3D a partire da vecchie immagini o video permette di generare nuove scene virtuali inedite e rappresentare persone che non possono essere digitalizzate con altri sistemi attuali perché non disponibili o defunte.

Le attuali tecniche per la modellazione 3D da immagini stanno rapidamente evolvendosi, ma le ricerche sono spesso focalizzate su oggetti o applicazioni specifiche. La completa automazione del processo di modellazione basata su immagini è comunque ancora una grande sfida per i ricercatori. Attualmente la maggior parte dei processi vede l'interazione di un operatore perciò lo sviluppo di nuove procedure per aumentare l'automazione nella ricostruzione virtuale è uno dei punti chiave delle attuali ricerche della fotogrammetria e della computer vision. Infatti metodi di modellazione che vengono dichiarati completamente automatici non funzionano con molte sequenze di immagini e non sono affidabili per certe applicazioni come la documentazione di beni culturali, dove precisione e dettaglio sono fattori molto importanti. Metodi completamente automatici necessitano una buona tessitura nelle foto, piccole basi tra i centri di presa, assenza di occlusioni e costante luminosità nelle immagini. Ma nei casi pratici questi requisiti non possono essere spesso mantenuti, portando al fallimento del processo automatico o alla generazione di modelli virtuali grezzi. Spesso questi processi estraggono molte corrispondenze in zone non necessarie per le fasi successive della modellazione e riportano precisioni molto scarse, limitando il loro utilizzo ad applicazioni che richiedono modelli 3D per visualizzazione e non per una precisa documentazione.

La modellazione di un oggetto in tre dimensioni è un processo complesso che dovrebbe essere inteso come una serie di passaggi che iniziano con l'acquisizione dei dati e terminano con la visualizzazione del modello a computer. La modellazione con metodi fotogrammetrici consiste

di pochi ma ben noti passi: calibrazione ed orientamento delle immagini, misurazione di punti sulla superficie dell'oggetto, generazione di una superficie poligonale, tessitura e visualizzazione. Recentemente molti sforzi sono stati fatti per aumentare il livello di automazione di questi passaggi e allargare l'utilizzo della tecnologia di modellazione basata sulle immagini. Purtroppo gli sforzi non hanno ancora prodotto molti risultati affidabili e tuttora i risultati migliori sono quelli ottenuti con metodi interattivi.

In questo lavoro di ricerca, diversi metodi sviluppati per analizzare immagini o video esistenti e parzialmente automatizzare alcune fasi del processo di costruzione di modelli 3D sono presentati e discussi. In particolare, scene con oggetti statici e persone in movimento sono state considerate. Il lavoro cerca di automatizzare la fase di ricerca dei punti omologhi tra le immagini (senza l'utilizzo di target) e di ricavare modelli virtuali da immagini già esistenti, analizzando le potenzialità e i limiti della fotogrammetria nell'analisi di vecchi filmati o immagini, dove i parametri della camera sono assenti o variazioni di focale sono presenti. Per le investigazioni, immagini disponibili in Internet o acquisite personalmente e filmati monoculari digitalizzati da vecchie videocassette sono utilizzati.

La possibilità di orientare (e calibrare) automaticamente una sequenza di immagini dipende fortemente dal tipo di dati a disposizione, dalla procedura d'acquisizione e dalla scena ripresa. Se paragonato ad altri approcci sviluppati nella comunità scientifica, il lavoro fatto è basato su punti d'interesse correlati con procedure ai minimi quadrati e supportato da analisi statistiche. Il metodo è affidabile anche per immagini acquisite con una larga base tra i centri di presa. Il metodo di compensazione a stelle proiettive (bundle adjustment) è utilizzato per ricavare i parametri incogniti della camera e le coordinate oggetto dei punti estratti automaticamente.

Per l'analisi e restituzione di persone in movimento usando filmati esistenti, è stato sviluppato un processo deterministico, supportato da vincoli e ipotesi sulla scena ripresa e sulla forma della persona. Il processo fotogrammetrico è in grado di analizzare diversi tipi video e diversi tipi di movimenti delle persone. I modelli 3D delle persone e le informazioni sulla scena sono ricavati principalmente per scopi di visualizzazione e per applicazione biomediche che non richiedono altissime precisioni.

Per l'estrazione automatica di punti omologhi tra le immagini e il loro orientamento relativo, diversi programmi sono stati sviluppati, insieme ad un tool per la visualizzazione grafica dei risultati. Per la ricostruzione dei movimenti delle persone, è stato sviluppato un software in grado di estrarre modelli 3D dalle immagini singole e poi combinarli tra loro nello stesso sistema di riferimento.