

This publication is an edited version of:

---

Diss. ETH No. 15271

**Surface Measurement and Tracking  
of Human Body Parts from  
Multi Station Video Sequences**

A dissertation submitted to the  
SWISS FEDERAL INSTITUTE OF TECHNOLOGY ZURICH

for the degree of  
Doctor of Technical Sciences

presented by  
NICOLA MICHELE D'APUZZO

Dipl. Masch. Ing. ETH

born July 11, 1970  
citizen of Bellinzona, TI

accepted on the recommendation of

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

Prof. Dr. Henrik Haggrén, co-examiner  
Helsinki University of Technology, Espoo, Finland

Prof. Dr. Hans-Peter Meinzer, co-examiner  
German Cancer Research Center, Heidelberg, Germany

Zurich 2003

---

IGP Mitteilung Nr. 81

Surface Measurement and Tracking of Human Body Parts  
from Multi Station Video Sequences

Nicola D'Apuzzo

Copyright © 2003

Institut für Geodäsie und Photogrammetrie  
Eidgenössische Technische Hochschule Zürich  
ETH Hönggerberg  
CH-8093 Zürich

Alle Rechte vorbehalten

ISSN 0252-9335

ISBN 3-906467-44-9

# *Abstract*

This work pertains to surface measurement and surface tracking of human body parts using video sequences acquired by multiple cameras. Traditionally, and even today, research and commercial applications were concentrated either on the measurement and modeling of the human face and body or on the capture of movement of the whole body and facial expressions. In this work, a method is presented to treat both the surface measurement aspects and the surface tracking aspects in a unique process. The proposed method can be applied to measure either the surface of a static human body part or the moving surface of a dynamic event. In the latter case, the 3-D data gained can be of two different types: surface measurement of the part of interest in the form of a 3-D point cloud for each recorded time step or surface tracking in form of a vector field of 3-D trajectories.

The process is composed of seven steps: (1) calibration of the system, i.e., establishing the internal and external orientation of the cameras and the parameters modeling the lens distortions; (2) acquisition of multi-image sets and/or multi-image sequences; (3) matching process, i.e., establishing correspondences in the multi-images; (4) computation of the 3-D point cloud for each matched multi-image set; (5) surface tracking in the multi-image sequences; (6) establishing a 3-D vector field of trajectories (position, velocity, acceleration); and (7) tracking key-points in the vector field of trajectories. In the case of surface measurement, only the first four steps are required.

The accurate measurement of a human body part starts with an adequate acquisition of the required data. In the case of static surface measurement are used multi-images (multiple images acquired from different positions in the space describing the same scene), while in the case of dynamic surface measurement and surface tracking are used multi-image sequences (multi-images acquired during a time interval). For the acquisition of multi-images different systems can be applied with differing levels of quality depending on the cameras used.

The orientation and calibration processes establish the position and orientation of the camera sensors in 3-D space, the parameters describing the internal geometry of the imaging device, and the parameters modeling the distortions caused by the optical system. A thorough determination of all the parameters is required for accurate measurement using photogrammetric techniques.

The goal of the automatic matching process is the determination of a dense set of corresponding points in the multi-images on the part of surface of interest. The process uses a stereo matcher based on least squares matching techniques. The automatic matching process begins by defining several seed points. Starting from them, a dense and robust set of corresponding points covering the entire interested region is generated. The 3-D coordinates of the matched points are then computed by forward ray intersection using the results of the calibration process. The strategy is customized to the characteristics of the surface of the human body. Moreover, it is designed to reduce the required processing time to a minimum.

The basic idea of the multi-image tracking process is tracking corresponding points in the multi-images through the sequence and computing their 3-D trajectories. Velocities and accelerations are also computed at each time step. The process is based on least squares matching techniques. These are applied to determine the spatial correspondences between the images acquired simultaneously from different views, as well as to determine the temporal correspondences between subsequent frames.

The proposed process can be used to track well defined points on the human body surface. Trajectories of single points, however, are not sufficient to understand and record the motion and movement of a human or the changes of the surface of human body parts. Accordingly, the tracking process is extended to simultaneously track a dense set of points belonging to a common surface. In this case, the result of the tracking process can be considered to be a 3-D vector field of trajectories.

To solve additional problems caused by occlusions, lack of texture, loss of tracked points and the appearance of new points, the concept of key-points is introduced. The key-points are 3-D regions, defined in the vector field of trajectories, whose size can vary and whose position is defined by their center of gravity. The key-points are interactively defined in a graphical user interface and tracked simply; the position in the next time step is computed as the mean value of the displacements of all the trajectories contained inside the 3-D region.

Graphical user interfaces were developed and implemented for all of the processes employed, including the multi-image acquisition, the calibration and orientation procedures, the automatic matching process and the visualization of the results as 3-D point clouds and 3-D vector field of trajectories.

To demonstrate the multiple functionality of the method, three different applications are presented: high accuracy measurement of human faces using five CCD cameras, measurement of a blood vessel branching casting fixed in a rotating frame using three CCD cameras and full body motion capture without markers using video sequences acquired by two or three synchronized CCD cameras.

## *Riassunto*

Questo lavoro tratta la misurazione e il tracking di superfici di parti del corpo umano utilizzando videosequenze registrate con più telecamere. Finora, la ricerca e le applicazioni commerciali in questo settore si sono concentrate unicamente sulla misurazione e modellizzazione del viso e del corpo umano e sulla registrazione del movimento del corpo umano o di espressioni del viso. In questo lavoro, invece, viene presentato un metodo che integra in un unico processo le procedure sia di misurazione di superfici sia di registrazione del movimento del corpo umano. Infatti, il metodo proposto può venire utilizzato per la misurazione di superfici statiche di parti del corpo umano, come pure per la misurazione di superfici in movimento. In quest'ultimo caso, i dati tridimensionali ottenuti possono produrre una misurazione della superficie della parte interessata sotto forma di nuvola di punti tridimensionali per ogni passo temporale oppure un tracking della superficie sotto forma di campo vettoriale di traiettorie tridimensionali.

L'intero processo è composto dai seguenti passaggi: (1) calibrazione del sistema, determinazione dell'orientamento interno ed esterno delle telecamere, inclusi i parametri di modellizzazione delle distorsioni ottiche; (2) acquisizione di multi-immagini e/o sequenze di multi-immagini; (3) processo di matching, determinazione di corrispondenze nelle multi-immagini; (4) calcolo delle nuvole di punti tridimensionali per ogni gruppo di multi-immagini sottoposte al processo di matching; (5) tracking di superfici nelle sequenze di multi-immagini; (6) determinazione del campo vettoriale tridimensionale di traiettorie (posizione, velocità, accelerazione); (7) tracking di punti chiave nel campo vettoriale di traiettorie. Nel caso della misurazione di superfici, sono richiesti solamente i primi quattro passaggi.

La misurazione accurata di parti del corpo umano inizia con un'adeguata acquisizione di dati. Nel caso di misurazioni statiche di superfici vengono utilizzate delle multi-immagini (immagini multiple che descrivono la stessa scena da posizioni differenti), mentre nel caso di misurazioni dinamiche o di tracking di superfici vengono utilizzate delle sequenze di multi-immagini (multi-immagini acquisite durante un intervallo di tempo). Per l'acquisizione di multi-immagini vari sistemi possono essere utilizzati con differenti livelli di qualità a dipendenza delle telecamere impiegate.

I processi di orientamento e di calibrazione hanno lo scopo di determinare la posizione e l'orientamento dei sensori, stabilire i parametri che descrivono la geometria interna delle telecamere e ottenere i parametri necessari per la modellizzazione delle distorsioni

generate dai sistemi ottici. La determinazione precisa di tutti i parametri è necessaria per ottenere un'accurata misurazione.

L'obiettivo del processo automatico di matching consiste nella determinazione di punti corrispondenti nelle multi-immagini nella parte di superficie interessata. Il processo automatico di matching utilizza uno stereo-matcher costruito con tecniche ai minimi quadrati. La strategia sviluppata per il processo automatico di matching consiste nella definizione di alcuni punti base, dai quali viene determinato automaticamente un denso gruppo di punti corrispondenti coprendo l'intera parte interessata. Le coordinate tridimensionali dei punti corrispondenti sono successivamente calcolate tramite intersezione dei raggi utilizzando i risultati del processo di calibrazione. La strategia è stata sviluppata interamente considerando le caratteristiche della superficie del corpo umano ed è stata implementata in modo da ridurre al minimo il tempo di elaborazione. L'idea base del processo di tracking consiste nel seguire punti corrispondenti nelle multi-immagini durante la sequenza completa e nel calcolare poi le relative traiettorie tridimensionali. Velocità e accelerazione possono pure venire calcolate per ogni passo temporale. Il processo è costruito su tecniche di matching ai minimi quadrati. Queste vengono utilizzate per la determinazione delle corrispondenze spaziali fra immagini acquisite contemporaneamente da diversi punti di vista come pure per la determinazione di corrispondenze temporali fra immagini consecutive.

Il processo proposto può essere utilizzato per il tracking di punti ben definiti sul corpo umano. Le traiettorie tridimensionali di singoli punti non sono sufficienti per comprendere e descrivere il movimento di una persona o i cambiamenti della superficie di una parte del corpo umano. Per questo motivo, il processo di tracking viene esteso per seguire simultaneamente un denso gruppo di punti appartenenti a una superficie comune e il risultato può venire considerato come un campo vettoriale di traiettorie tridimensionali.

Per risolvere problemi causati da occlusioni, mancanza di texture, perdita di punti e apparizioni di nuovi punti, vengono introdotti i punti chiave. I punti chiave sono regioni tridimensionali di dimensioni variabili definiti nel campo vettoriale di traiettorie e la loro posizione è definita dal centro di gravità. Essi vengono scelti e posizionati interattivamente in una interfaccia grafica ed il processo di tracking avviene in maniera semplice: la posizione nel prossimo passo temporale è calcolata come il valore medio dello spostamento di tutte le traiettorie contenute nella regione tridimensionale.

Interfacce grafiche sono state sviluppate per tutti i processi, compresi l'acquisizione delle multi-immagini, la calibrazione e l'orientamento, il processo automatico di matching e la visualizzazione dei risultati come nuvola di punti tridimensionali e/o campo vettoriale di traiettorie tridimensionali.

Per dimostrare le funzionalità multiple del metodo proposto, sono presentate tre differenti applicazioni: misurazione ad alta precisione del viso umano tramite cinque telecamere CCD, misurazione di un calco arteriale tramite tre telecamere CCD e registrazione del movimento del corpo umano senza l'utilizzo di punti marcanti, tramite due o tre telecamere CCD sincronizzate.