Medical Engineering & Physics 36 (2014) 1168–1175

Contents lists available at ScienceDirect

Medical Engineering & Physics

journal homepage: www.elsevier.com/locate/medengphy

Editor's comment: In this paper the authors have developed a markerless motion capture approach for measuring lower extremity 3D gait kinematics. The technique was developed using a custom made data acquisition system and reconstruction algorithms combined with a subject specific unconstrained articulated model. While not describing 3D motion quite as accurately as a marker based system it provides an insight into how markerless systems might be developed.

Mark Pearcy, Associate Editor

Markerless motion capture can provide reliable 3D gait kinematics in the sagittal and frontal plane

Martin Sandau [a,b,*], Henrik Koblauch [a], Thomas B. Moeslund [c], Henrik Aanæs [d], Tine Alkjær [a], Erik B. Simonsen [a]

[a] Department of Neuroscience and Pharmacology, University of Copenhagen, 2200 Copenhagen N, Denmark
[b] The Danish Institute of Fire and Security Technology, 2650 Hvidovre, Denmark
[c] Department of Architecture, Design and Media Technology, Aalborg University, 9200 Aalborg Sv, Denmark
[d] Department of Informatics and Mathematical Modelling, Technical University of Denmark, 2800 Kgs. Lyngby, Denmark

ARTICLE INFO

Article history:
Received 21 February 2014
Received in revised form 1 July 2014
Accepted 2 July 2014

Keywords:
Markerless
Motion capture
Tracking
Gait analysis
Biomechanics
3D reconstruction
Dense point cloud
Stereo vision
Photogrammetry
Iterative closest point
Articulated model

ABSTRACT

Estimating 3D joint rotations in the lower extremities accurately and reliably remains unresolved in markerless motion capture, despite extensive studies in the past decades. The main problems have been ascribed to the limited accuracy of the 3D reconstructions. Accordingly, the purpose of the present study was to develop a new approach based on highly detailed 3D reconstructions in combination with a translational and rotational unconstrained articulated model. The highly detailed 3D reconstructions were synthesized from an eight camera setup using a stereo vision approach. The subject specific articulated model was generated with three rotational and three translational degrees of freedom for each limb segment and without any constraints to the range of motion. This approach was tested on 3D gait analysis and compared to a marker based method. The experiment included ten healthy subjects in whom hip, knee and ankle joint were analysed. Flexion/extension angles as well as hip abduction/adduction closely resembled those obtained from the marker based system. However, the internal/external rotations, knee abduction/adduction and ankle inversion/eversion were less reliable.

1. Introduction

Motion capture is widely applied in fields as entertainment, sports and clinical gait analysis. In entertainment, motion capture is used to animate 3D characters in movies and games, while sports use it for injury prevention and improving performances. In clinical analysis, motion capture is applied for diagnosis and treatment of patients with neural or musculoskeletal dysfunctions. A more widespread use of motion capture and analysis has been limited because the available methods for accurate capture of 3D human movement require a laboratory environment and the attachment of markers that are associated with several drawbacks. The placement of markers is time consuming and people feel uncomfortable with markers attached. It requires knowledge and skills in human anatomy and small variations in the placement of the markers can induce large variations in crosstalk between joint angles [1]. In addition, *soft tissue artefacts* (STA) result in inaccuracies for all 3D joint angles which makes estimation of joint angles with small dynamic ranges such as internal/external (IE) rotations of the segments less reliable [2–4]. Lighting conditions are also crucial to track markers. Markerless motion capture has therefore been

* Corresponding author at: DBI, Security Consulting, Jernholmen 12, 2650 Hvidovre, Denmark. Tel.: +45 61 20 16 65.
E-mail address: msc@dbi-net.dk (M. Sandau).

http://dx.doi.org/10.1016/j.medengphy.2014.07.007
1350-4533/

studied intensively during the past decades, despite the challenges in combining the knowledge of biomechanics, computer vision, computer graphics and numerical optimization [5–10].

Markerless systems are divided into appearance based and model based approaches [8]. The appearance based approaches learn image features such as silhouettes and map them directly into a multidimensional space of poses. These approaches need training data closely resembling the motions to be analysed. Model based approaches use primitive representations of the human anatomy with an underlying kinematic chain that are fitted to either silhouettes from each camera view or *visual hull* (VH) models to estimate the poses [11–20]. Hybrid methods have also been proposed recently [21,22]. The Kinect (Microsoft, Redmond, Washington, USA) is the most popular example which has enabled low cost markerless motion capture in real time with sufficient accuracy for wide range of applications including gaming and therapeutic rehabilitation. However, the accuracy is markedly lower than marker based systems and consideration in the data interpretations is therefore required when applying the Kinect for clinical analysis [23].

Despite the increasing interest in markerless motion capture and the drawbacks of the markers, marker based systems are still preferred for clinical 3D gait analysis because the accuracy of the available markerless systems are limited and the kinematic validations are inadequate [24]. Corazza et al. [25–27] and Mündermann et al. [28–30] have achieved high accuracy with markerless motion capture by using subject specific articulated models synthesized from either laser scans of the concerned test subject or from a repository of laser scanned subjects. The poses were estimated by fitting the models to VH representations of each frame. VH models are fast to process and easy to implement but they tend to overestimate the volume of the reconstructed object, because the method cannot carve concave structures into the 3D reconstructions. The overestimation of the 3D volume makes the thigh and shank axial symmetric, which complicates accurate estimation of the IE rotations and the kinematics of the ankle joint [25].

Multi-view stereo algorithms are capable of producing highly detailed 3D reconstructions [31] and have been applied to markerless motion capture recently [32,33]. However, these approaches were not combined with pose estimation and were therefore not applicable for kinematics.

The purpose of the present study was to develop a novel markerless approach to address the problem of obtaining accurate kinematics in the lower extremities. The proposed approach was compared with a marker based method although the latter also suffers from limitations [2,34,35]. The contributions of the current study are thus (1) an approach to acquire highly detailed 3D reconstructions of human motion with high temporal resolution and low costs. (2) Accurate pose estimation without any simple articulations of the joints or constraints. (3) A validation on 3D gait analysis including all rotations in the hip, knee and ankle joint.

2. Methods

The pipeline of our model based motion capture consists of four steps: Data acquisition, 3D reconstruction, articulated model generation and pose estimation. Data acquisition comprises the technical requirements and the process that facilitates video capture. The 3D reconstruction is the conversion of recorded image data to 3D surface meshes. Articulated model generation is the step where the model is synthesized before the pose can be estimated. In pose estimation the pose is estimated by fitting the articulated model through numerical optimization. The four steps are described in detail in the following subsections.

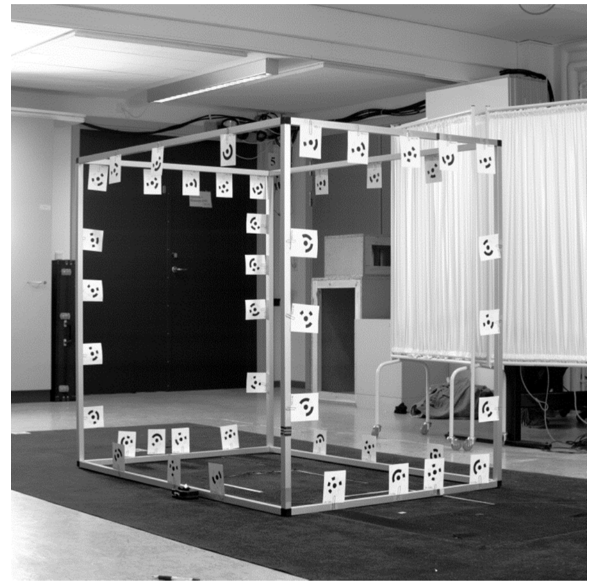

Fig. 1. Extrinsic camera calibration.

2.1. Data acquisition

The experimental setup consisted of eight monochrome *Camera Link* cameras placed in the corners of a $7 \times 9\,m^3$ laboratory. The cameras were hardware synchronized and operated at a frame rate of 75 frames per second. The cameras had a pixel resolution of 2048×2048 pixel covering a volume of roughly $2 \times 2 \times 2\,m^3$.

In order to avoid motion blur, the shutter time was set to one millisecond. Diffuse light sources were used to obtain 850 LUX in the volume of interest.

Intrinsic and extrinsic camera calibrations were performed using PhotoModeler® (Eos Systems Inc. Vancouver, Canada). Intrinsic calibration was performed with a calibration board and extrinsic calibration was performed by attaching 42 coded targets to a $1.55 \times 1.05 \times 1.35\,m^3$ calibration frame located in the centre of the volume of interest as illustrated in Fig. 1.

2.2. 3D reconstruction

A 3D point cloud was synthesized using the *patch based multi-view stereo* algorithm (PMVS) by Furukawa and Ponce [36]. To reduce the computation time, background subtraction was performed on each frame. This was complicated because the images were monochrome and the background was cluttered. However, subtracting all the background was not necessary to obtain satisfactory results of the 3D reconstruction. From the 3D point clouds a closed surface mesh was constructed by the *Poisson surface reconstruction* (PSR) algorithm by Kazhdan et al. [37] as illustrated in Fig. 2. The purpose of creating a closed surface was to fill holes in the point cloud caused by self-occlusion. The PSR algorithm had difficulties in proper reconstruction of the foot when partially occluded. The final point cloud consisted therefore of the raw 3D points synthesized by the PMVS algorithm merged with the points synthesized by the PSR algorithm.

Stereo algorithms use the texture of the surface to find point correspondences between two or more images. To enhance both texture and curvature of the body, the test subject was dressed in a full body snow leopard spandex suit as illustrated in Fig. 3. Projecting a random pattern onto the subject with a projector could also have been an alternative option to the suit. This enabled a reconstruction with around 75.000 3D points that corresponded to a point sampling at less than one centimetre.