Applications of markerless motion capture in gait recognition

Martin Sandau

This review has been accepted as a thesis together with four previously published papers by University of Copenhagen 22nd of December 2014 and defended on 24th of April 2015

Tutor(s): Erik B Simonsen, Henrik Aanæs and Thomas B Moeslund

Official opponents: Michael S Andersen, Mark S Nixon and Rune Berg

Correspondence: Erik B Simonsen, Department of Neuroscience and Pharmacology, University of Copenhagen, Blegdamsvej 3, 2200 Copenhagen N. Denmark.

E-mail: msc.22@hotmail.com

Dan Med J 2016;63(3): B5211

THE FOUR ORIGINAL PAPERS ARE

P1 Sandau M, Heimbürger R V, Jensen KE, Moeslund T B, Aanæs H, Alkjær T, Simonsen E B., Reliable gait recognition using 3D reconstructions and random forests -An anthropometric approach, Journal of Forensic Sciences, 2016, In press

P2 Sandau M, Koblauch H, Moeslund T B, Aanæs H, Alkjær T, Simonsen E B. 2014. Markerless motion capture can provide reliable 3D gait kinematics in the sagittal and frontal plane. Journal of Medical Engineering & Physics, 2014, 36(9): 1168-1175

P3 Sandau M, Heimbürger R V, Villa C, Jensen K E, Moeslund T B, Aanæs H, Alkjær T, Simonsen E B., New equations to calculate 3D joint centres in the lower extremities, Journal of Medical Engineering and Physics, 2015, 37(10): 948–955

P4 Yang SXM, Christiansen MS, Larsen PK, Alkjær T, Moeslund TB, Simonsen EB, Lynnerup N., Markerless Motion Capture Systems for tracking of persons in Forensic Biomechanics: An overview. Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization, 2013, 2(1): 46-65

INTRODUCTION

The Danish police lack resources to maintain effective investigation of crimes. In 2013, 244 550 criminal cases were reported where 18 327 of the cases were classified as dangerous including murders, rapes, armed burglaries, violence, threats, arson, and several others [1]. The police filed charges in 92.7% of the dangerous cases indicating that the cases have been highly prioritised. However, the evidence collected and later presented by the prosecutors tends to be weak as only 56.7% of the cases were solved. The lack of resources is also supported by the fact that 73 370 burglaries were reported in 2013 where the police filed charges in 15.5% of the cases and only 8.5% of the cases were solved [1]. Additionally, according to a police report leaked by the newspaper BT in July 2014 larcenies below 1 000 000 DKR. are not investigated due to the lack of resources [2].

To meet these challenges in the future, the police investigation methods have to be optimised by automating trivial work and new effective investigation methods have to be developed. Automated back tracking of perpetrators through closed circuit television (CCTV) data could reduce the workload by the police, but also automatic recognition based on biometrics as fingerprints, DNA, facial features, body dimensions and gait would aid in this and strengthen the evidence as human bias will be minimised.

Today CCTV can provide strong evidence and plays an important role in criminal investigations [3]. However, recognition in CCTV is resource-demanding, as the police technicians have to collect the video data and extract the relevant sequences. In the cases where the perpetrators cannot be recognised by the face, forensic anthropologists have to analyse other biometric features in the images, which may also require a field calibration of the cameras as illustrated in Fig. 1.

Figure 1
Typical setup for camera calibration on a crime scene. Photo: P K Larsen

Various biometric features can be extracted and analysed from CCTV data. The shape of the ear, facial features, tattoos, birthmarks, eye colour and body dimensions are the classical examples but also gait is of increasing interest. For centuries gait has been known to be unique among individuals as Shakespeare can be cited for this in this work The Tempest from 1610-1611.

Cutting and Kozlowski [4] were the first to report scientific results on this as people recognised their friends by their walk when shown video recordings of markers mounted on the joints. The first time gait was used as evidence in Denmark was in 2002 where a bank robber had a highly characteristic gait [5, 6]. After this episode the interest of applying photogrammetric measures as evidence has increased but today these techniques are only applied in the most urgent cases due to the high resource demand and the varying strength of evidence caused by poor image quality and challenging perspectives [6-8]. The strength of evidence is also affected by the intra- and inter-observer variability of the gait kinematics and the body dimensions. This issue was studied by Larsen, Hansen [9] and the results showed that measures of segment lengths were not reliable for recognition as the intra- and inter-observer variability were high compared to the expected variability of the segment lengths within participants of similar stature.

To reduce the issues related to human interference and the resource demand, fully automatic recognition algorithms have been proposed [10-17]. The algorithms can be grouped into model-based and appearance-based methods. In model-based methods, kinematics are obtained by fitting an articulated limb segment model to silhouettes or 3D reconstructions extracted from video data [11, 18]. The articulated models comprise a surface mesh and an underlying stick model representing the skeleton. As the surface mesh is fitted, the underlying skeleton is subsequently used to estimate the locations of joint centres and the joint angles. Participants are then recognised by their gait based on the extracted kinematic parameters. The advantage of the model-based approaches is that joint angles can be analysed separately, but the methods reported are often presented without a quantification of the kinematic precision as no golden standards exist (P4). It is thus difficult to clarify whether it is the actual joint angles or participant specific biases that provide the kinematic differences between participants [10, 11]. In appearance-based methods, image features are extracted from the video sequence and mapped directly to a low-dimensional space for classification. Among the appearance-based methods, the Gait Energy Image [19] appears to be the most successful representative in gait recognition [16]. Appearance-based approaches provide promising results for recognition and compared to the model-based approaches they generally perform better [10-12, 16]. As these approaches consider gait as a holistic pattern, body dimensions and gait parameters cannot be interpreted separately, which makes it is difficult to explain which parameters provide distinct discrimination between participants contrary to the model-based methods.

Motion analysis with motion capture is applied for various other purposes as clinical analysis, sports analysis, animating motions and controlling avatars in computer games. However, motion capture systems perform differently according to the intended use. For instance, motion capture systems for gaming are highly dependent on estimating kinematics in real-time whereas motion capture in clinical biomechanics are dependent on high accuracy of the estimated kinematics and are less dependent on the processing time, which is similar to the needs in forensic sciences. Motion capture systems intended for clinical analysis can therefore provide inspiration for precise gait analysis in the forensic sciences with the exception that clinical biomechanics frequently allows the attachment of markers. A survey of formal 3D markerless motion capture systems was conducted in P4 with a discussion of their potential applications in the forensic sciences.

Probably the most popular device for markerless motion capture is the Kinect (Microsoft, Redmond, Washington, USA), which is a motion controller for video gaming. The Kinect is based on a hybrid motion capture method, which means that it is a combination of model-based and appearance-based methods [20, 21]. This enabled markerless motion capture in real time sufficiently precise for a wide range of applications including gaming and training for therapeutic rehabilitation. However, considerations in the data interpretation are required when applying the Kinect for clinical analysis as the precision is markedly lower than the traditionally applied marker-based systems [22].

Stanford University's BioMotion Laboratory has provided several essential contributions to precise measurement of 3D kinematics with markerless motion capture. Corazza, Mündermann [23] proposed to use participant specific articulated models that were fitted to primitive 3D reconstructions. The use of participant specific articulated models obtained from laser scans enabled precise estimations of sagittal plane joint angles. Mündermann, Corazza [24] expanded the approach by replacing the personal laser scans with models retrieved from a repository of laser scanned humans with varying body dimensions and used soft constraints on the joints of the articulated model to enable small movements of the joints. The generation of participant specific articulated models was further refined by Corazza, Gambaretto [25] and Corazza, Mündermann [26] also showed that combining the recent methods and adding rotational constraints to the articulated models enabled a high precision of the joint centre estimations. However, the precision of the 3D joint angles in the hip, the knee and the ankle were not quantified, which would have been interesting in regard to clinical applications as well as in forensic sciences.

To analyse the precision of the kinematics estimated by markerless motion capture systems, marker based motion capture systems are often applied as the 'true' reference. But several factors induce considerable errors including the placement of the markers, the regression equations used to predict the joint centres and the motion of the markers relative to the underlying bone. The errors regarded to the placement of the markers and the motion of the markers also called soft tissue artefacts are well known [27, 28], but quantifications of the errors regarding the regression equations to predict the knee and ankle joint centres have not been reported.

To summarise, current photogrammetric methods used to recognise perpetrators in CCTV are based on manual measures of body dimensions and gait kinematics. This is resource demanding and heavily affected by intra- and inter-observer variability. These issues can be minimised by using automatic recognition methods but the precision of the kinematics extracted with such methods are unknown and the reference techniques to quantify this are also affected by errors that have not been quantified. Therefore, this thesis provides three essential contributions to meet these challenges. The first contribution (P1) is a quantification of the intra- and inter-observer variability of annotations on 3D reconstructions and an analysis of the associated effect on recognition. The second contribution (P2) is a new model based markerless motion capture method to achieve automatic and precise estimation of gait parameters. The third contribution (P3) quantifies the errors of established marker based regression equations to predict joint centres and generates new equations corrected for biases.

In the first contribution, highly accurate 3D reconstructions of participants were annotated by eight expert observers to quantify the intra- and inter-observer variability of gait kinematics and

limb segment lengths and the associated effect on recognition. The recognition was performed with random forests [29], which provided easy interpretation of the parameter importance. The study shows how recognition can be performed reliably with minimal human interference in the future.

In the second contribution the accurate 3D reconstructions improved the precision of model-based markerless motion capture, which was validated for all 3D joint angles in the lower extremities. The study provides insight into how markerless motion capture systems might develop to acquire reliable gait recognition systems.

In the third study the error of established marker based methods to predict the joint centres in the lower extremities was quantified and a new unbiased set of equations were regressed, which also accounted for sex differences in pelvis anthropometrics. This was achieved with MRI and CT scans of ten males and ten females As marker-based motion capture is frequently applied for scientific analysis of gait, the increased accuracy will first of all contribute to the understanding of gait dynamics but will also provide more reliable validations of markerless motion capture systems as marker-based motion capture is frequently applied as the 'true' reference.

The purpose of the three contributions was to increase the reliability of recognition based on body dimensions and gait kinematics and to reduce the resource demand of criminal investigations in the future.

MATERIALS AND METHODS

This section is divided into five subsections: The DATA OVERVIEW section describes the data applied in the study. THE GAIT LABORATORY section describes the laboratory setup and the 3D reconstruction algorithms that were applied to synthesise the 3D reconstructions. THE VARIABILITY OF GAIT AND THE EFFECT ON RECOGNITION section describes how the variability of gait kinematics and limb segment lengths was quantified and compared with the intra- and inter-observer variability. It also describes how the importance of the kinematics and the limb segment lengths were quantified regarding recognition. The AUTOMATIC POSE ESTIMATION section describes how to extract joint angles in the lower extremities from the 3D reconstructions with high precision and with minimal human interaction. Finally, the NEW REGRESSION EQUATIONS section explains how the error of the existing marker based regression equations was quantified and how new accurate equations were regressed.

DATA OVERVIEW

The gait data set

Data from 16 male participants were recorded in the gait laboratory to obtain 3D reconstructions of them during normal walking. The participants were Caucasians aged 22-33 years and had no history of lower extremity pathology. The heights ranged from 1.69 to 1.89 m and the BMI ranged between 20.6 to 25.7 kg/m2. Data from each participant were recorded over two different days where the participants performed ten gait trials with a walking speed of 1.1 m/s (4.0 km/h) ±10% each day. The walking speed was measured by two photo cells and the participants were allowed to practice the desired walking speed and were given immediate oral feedback. The participants were dressed in leopard spandex to enhance the curvatures and the texture on the body as illustrated in Fig. 2.

Figure 2
To enhance the body curvatures and the texture on the body the participants were dressed in leopard spandex

The 3D reconstructions from all 16 participants were applied in P1 and 10 of them were randomly included in the validation of the proposed markerless motion capture approach in P2. The 16 participants are 3D reconstructed in appendix.

The anthropometric reference data set

To quantify the errors in P1 and P3, 'true' anthropometric measurements were obtained from structural MRI scans and CT scans. In P3 the hip joint centre (HJC) regression equations were based on CT scans of the 10 male and the 10 female cadavers to account for sex differences. All participants were Caucasians aged 21-57 years with a body mass index (BMI) between 17.8-27.2 kg/m2. CT scans were preferred because CT provides excellent contrast between the bone and soft tissues and motion artefacts were avoided by using cadavers instead of living participants. The CT scans were performed using a Somatom Definition CT scanner (Siemens AG, Erlangen, Germany) with the following settings: 120 kV and 285 mAs and a spatial resolution at 3x3x3 mm3.

The 'true' segment lengths in P1 and the 'true' locations of the knee joint centres (KJC) and the ankle joint centres (AJC) in P3 were obtained from structural MRI scans. The MRI scans were obtained with a 1.0 Tesla, Harmony MRI scanner (Siemens AG, Erlangen, Germany). The Transmit-receiver body coil was used for excitation and signal detection. A gradient echo T1-weighted pulse sequence was used in order to provide the best possible tissue contrast between muscle and bone tissue. The spatial resolution was 0.8x0.8 mm2/pixel in the frontal plane slices and the slice thickness was 3 mm.

To obtain the full length of the long bones, the MRI scans were fused by 3D registration. The registration was performed by rigid transformation using the correlation coefficient of the overlapping anatomical structure as similarity measure and the simplex search method for optimisation [30]. To quantify the error caused by the 3D registration and inhomogeneity in the magnetic field, a femur phantom was constructed as illustrated on Fig. 3.

The phantom consisted of a polypropylene tube with a length of 60.0 cm and a diameter of 5.0 cm and a polyvinylchloride (PVC) ring with a length at 4.0 cm and a diameter at 12.0 cm fixed midway of the tube. The PVC ring was labelled with nine equally distributed elliptical markers consisting of fish oil with a length at 2.3

cm and a diameter at 1.0 cm. Along the first 30.0 cm of the tube, 15 markers were placed with 2.0 cm between the markers whereas 13 markers were placed at the other side of the ring also with 2.0 cm space between the markers. The 15 aligned markers on one side of the ring were used to provide a linear model. This was obtained with the first component from a Principal Component Analysis (PCA) where the two other components quantified the regression error together with the variability of the distances between the markers. To quantify the 3D registration error, the root-mean-squared-error (RMSE) was computed of the residual between the linear model and the 13 aligned markers in the transformed MRI scan.

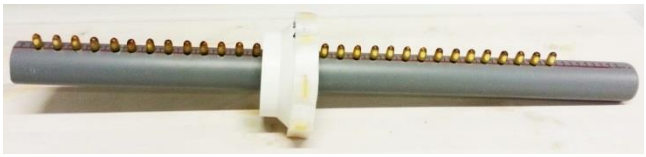

Figure 3
Femoral phantom to quantify the measurement error related to image registration and distortion of the MRI scans

Definition of the joint centres

In P1 and P3 the joint centres correspond to the centre of rotation (CoR). The CoR of the shoulder and the elbow joint were assumed to be the centre of caput humeri and trochlea humeri, respectively. The wrist was more complex consisting of the radio-lunate and lunate-capitate joint [31]. This was simplified to be a single joint located in the joint line between os lunatum and os capitatum as illustrated in Fig. 4.

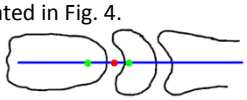

Figure 4
Sketch of the wrist with the medial axis (blue), the radio-lunate and lunate-capitate joints (green) and the simplified rotation centre (red) (P1)

With regard to the lower extremities, the hip joint centre was assumed to be in the centre of the femoral head. The knee joint centre in medial/lateral direction was assumed to be located in the midway between the epicondyles, whereas the flexion/extension axis was estimated similar to the method by Iwaki, Pinskerova [32] as image slices in the sagittal plane were obtained of the knee joint with flexion angles at 0 and 35 degrees. The tibias were superimposed in the images and the contact region on the tibial articular surface was labelled. Circles were then fitted to the tibio-femoral contact surface within the regions where the centre of the circles was considered as the instantaneous rotation centre[32]. The fixed CoR was defined as the mean of the circle centres as illustrated in Fig. 5. The same approach was applied to the AJC where the plane of intersection to the estimated axis was assumed to be in the midway between the medial and lateral side of trochlea tali.

THE GAIT LABORATORY

The following sections are describing the setup and the considerations regarding the hardware, the camera calibration, the software development and the validation of the 3D reconstructions obtained with the system.

The hardware setup

The setup was built in an already existing gait laboratory measuring 9.0x7.0x2.3 m3 with the Volume of Interest (VoI) in the centre of the room covering roughly 2.0x2.0x2.0 m3. To cover this volume, eight cameras were located in pairs with a vertical baseline between them as sketched in Fig. 6. The eight cameras were connected to one computer with four Matrox Radient eCL-DF frame grabbers (Matrox Electronic Systems Ltd., Montreal, Quebec, Canada), which enabled hardware synchronisation through external auxiliary I/O connectors. The computer had a BIG BANG – MARSHAL (B3) Series (MS-7670) ATX mainboard (Micro-Star Int'l Co. Ltd., New Taipei City 235, Taiwan) and 16 GB DDR3-1600 SDRAM.

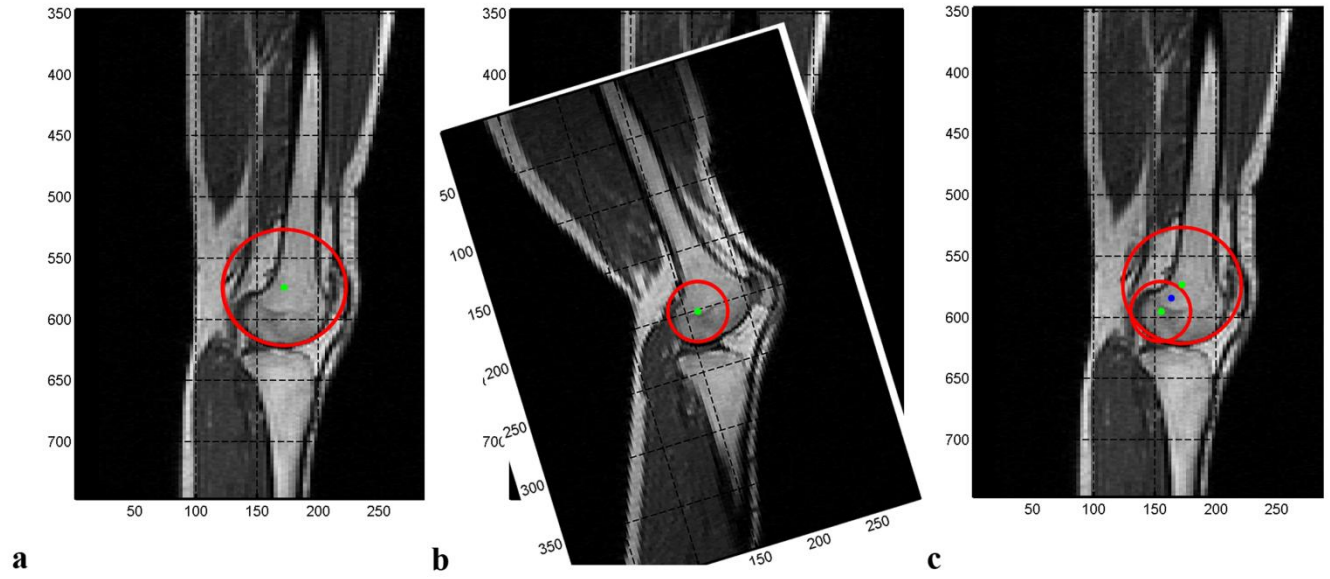

Figure 5
a: fitting a circle to the tibiofemoral contact surface of a stretched knee. b: superimposing shank of a flexed knee and fitting a new circle to the contact surface. c: the mean position of the instantaneous joint centres represents the fixed CoR (blue) (P3)

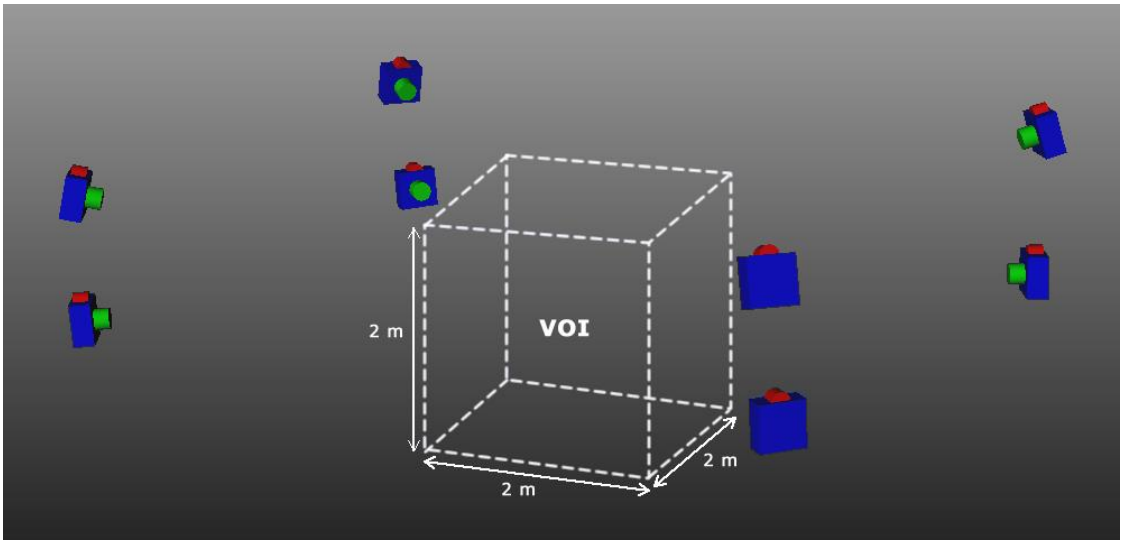

Figure 6
Reconstruction of the camera locations and the Volume of Interest (VoI) (P1)

Imaging of highly dynamic motions requires a short shutter time to avoid motion blur and a high sampling frequency. In addition, highly detailed 3D reconstructions require a high spatial resolution. The Camera Link cameras: Gazelle GZL-CL-41C6 (Point Grey Research Inc., Richmond, British Columbia, Canada) meets this with 2048x2048 pixels and a maximal frame rate at 142 fps. However, we could sustain a maximal frame rate at 135 fps with the current setup.

The imaging sensors in the cameras were monochromatic 1" CMOS with global shutter. Monochromatic 1" CMOS imaging sensors has a relatively high light affinity. This was preferred to obtain low noise images with a shutter time at 1 ms and 850 LUX illuminating the VOI. The global shutter was also preferred as rolling shutters induces motion artefacts.

The data size of the images was 4MB, resulting in 540 MB/s for each camera and 4.32 GB/s from eight cameras. As this requires a relatively large band width, it was necessary to use DDR3 SDRAM with a data transfer rate at 1066 MT/s in the computer memory.

The cameras were connected to the computer with 12 metre BitMaxx™ extended length cables (Components Express Inc, Woodbridge, Illinois, USA) as they were certified up to 85 MHz, which was required to run the Camera Link cameras in full 8-tap configuration (the highest performance).

The 1" format 25 mm fixed focal length lenses (Edmund Optics Inc., Barrington, New Jersey, USA) was mounted on the cameras to cover the VOI with optimal spatial resolution. The lens resolution at 100±20 line pairs (lp)/mm matched the spatial sensor resolution at 90 lp/mm).

PhotoModeler® (Eos Systems Inc., Vancouver, Canada) was used for both intrinsic and extrinsic calibration. The intrinsic calibration was performed with a calibration board and the extrinsic calibration was performed with a 1.55x1.05x1.35 m3 frame with 42 coded targets.

Background subtraction

To speed up the image processing, the amount of pixels to be processed was reduced by segmenting the object of interest with background subtraction. Typical problems in background subtractions are regarded to variations in the background, inconsistent lighting and low contrast between object and background. Shadows also caused problems in the current setup and high contrast between object and background were also challenging as the lab environment was cluttered. However, accurate background subtraction was not crucial as the inclusion of background pixels in the segmentation resulted in a prolonged computation time and redundant 3D points near the reconstructed participant. On the other hand, parts of the participant that were erased by the background subtraction were not recovered in the 3D reconstruction. Therefore, to ensure full recovery of the participant, a dilation of the segmentation was performed according to van den Boomgaard and van Balen [33].

The PMVS algorithm

From the background subtracted multi-view images a 3D point cloud was reconstructed of the participant with the 'Patch based multi-view stereo algorithm' (PMVS) [34]. The PMVS algorithm processes 3D reconstructions in a three step procedure embracing matching, expanding and filtering. The image features to be matched were based on Harris corners [35] and Difference of Gaussians (DoG) [36]. After matching, an expansion step and a filtration step was iteratively executed to create a denser point cloud and to filter the outliers using constraints concerning visibility consistency and spread.

Matching

Corners are good image features because the intensity gradient is high in two directions, which can be extracted with DoG and Harris corners. DoG is a filtering technique whereas Harris corners are based on 2D structure tensor representations of the intensity gradients. The features extracted by these two methods were matched by finding a corresponding pixel with low photogrammetric discrepancy. The discrepancy was quantified by following cross-correlation based method: Given a reference image and another image in which the feature point is visible, the discrepancy is computed by 1) overlaying a 11x11 pixel2 grid on the feature point in reference image 2) Project the grid into the other image and sample the corresponding pixel colors 3) computing one minus the normalised cross-correlation between the grid and its projection. The grid represents a patch in 3D that is a local tangent plane approximation of a surface where the center and the normal of the patch is optimised and re-projected into the image planes to obtain minimal discrepancy. In this optimisation the

center of the patch is constrained to lie along the epipolar line. Potential matches of a feature point in a reference image are the features within n pixels from the epipolar lines in the other images where the best match is considered as a point correspondence.

Expansion

In the expansion step each image was divided into small cells (4x4 pixels2 in our study). The goal of the expansion step was hereby to reconstruct at least one patch within each cell. New patches were generated by taking existing patches and generate new ones in the nearby empty spaces that satisfy certain criteria. The first expansion criterion is that the expansion is unnecessary if a patch has already been reconstructed in the current cell. The second criterion is that even if no patches have been reconstructed, the expansion is unnecessary for a cell if there is a depth discontinuity viewed from the corresponding camera.

Filtering

Three filters were applied to remove outlier patches. The first filter relies on visibility consistency: Consider the patches that are in the same cell as the current patch. The current patch is filtered out if 1) the other patches cannot be considered as neighbouring in 3D and 2) the photogrammetric discrepancy is higher of the current patch than the photogrammetric discrepancy of the other patches. Secondly if the number of images in which the patch is visible is less than γ then the patch is filtered out as an outlier. Since the cameras were placed pairwise in the corners of the room in the current approach, γ was set to 2. The third filter is a weak form of regularisation: Consider the patches on its own and in the adjacent cells in all images where the current patch is visible. If the proportion of the surrounding patches that can be considered as neighbouring in 3D is less than 0.25, then it is removed as an outlier.

Closed Surface Reconstruction

The PMVS algorithm reconstructed a 3D point cloud that was used to synthesise a closed surface mesh with the Poisson Surface Reconstruction (PSR) algorithm by Kazhdan, Bolitho [37] as illustrated in Fig. 7. The closed surface was created to fill holes in the point cloud caused by self-occlusion. By this method, around 75 000 3D points were recovered from the body surface which corresponded to a point density less than one centimetre on average.

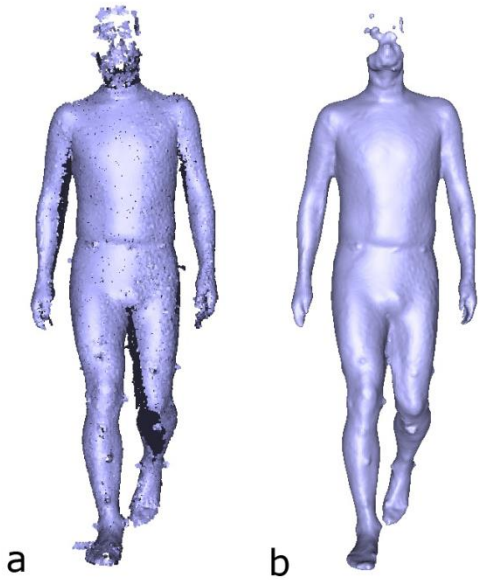

Figure 7

a: The 3D point cloud obtained with the PMVS algorithm. b: The closed surface mesh obtained with the PSR algorithm (P2)

Validation of the 3D reconstructions

The precision of the 3D reconstructions obtained from the laboratory setup was quantified by reconstructing a mannequin dummy, which was compared to a 3D reconstruction obtained by laser scan. The laser scanned 3D reconstruction was obtained with a FastSCAN, COBRATM C1 hand held laser scanner (Polhemus, Colchester, Vermont, USA), with a precision of < 1 mm. To avoid distortion in the laser scanned 3D reconstruction, all metal in the mannequin doll was removed. Like the participants, the doll was dressed in a snow leopard spandex suit to enhance the texture as illustrated in Fig. 8. The head was not included in this study as the head on the test participants was also ignored. The laser scan and the photogrammetric reconstruction was registered with the Iterative Closest Point algorithm [38]. The error metric was based on the distance between the nearest neighbour point correspondences in the reconstruction obtained by the current method and the reconstruction obtained by the laser scan.

Figure 8

Setup for validating the accuracy of the 3D reconstructions (P1)

THE VARIABILITY OF GAIT AND THE EFFECT ON RECOGNITION

The 3D reconstructions were annotated by eight expert observers to quantify the intra- and inter-observer variability of the extracted kinematics and body dimensions and to observe the associated effect on recognition. The observers annotated one gait trial from six participants who were randomly selected from the data set. The annotated landmarks included the wrist, the elbow, the shoulder, the hip, the knee, the ankle and the metatarsal head II. The landmarks were annotated in the frontal and the sagittal plane during heel strike (corresponding to the initial state in the gait cycle), mid-stance (corresponding to 15% in the gait cycle), toe-off (corresponding to 60% in the gait cycle) and mid-swing (corresponding to 73% in the gait cycle) in the gait trials as illustrated in Fig. 9. The segment lengths were defined as the distance between the estimated joint centres and the frontal and the sagittal plane kinematics were computed directly from the annotations in the two planes. The shoulder to shoulder angle and hip joint angle were computed in the global reference frame

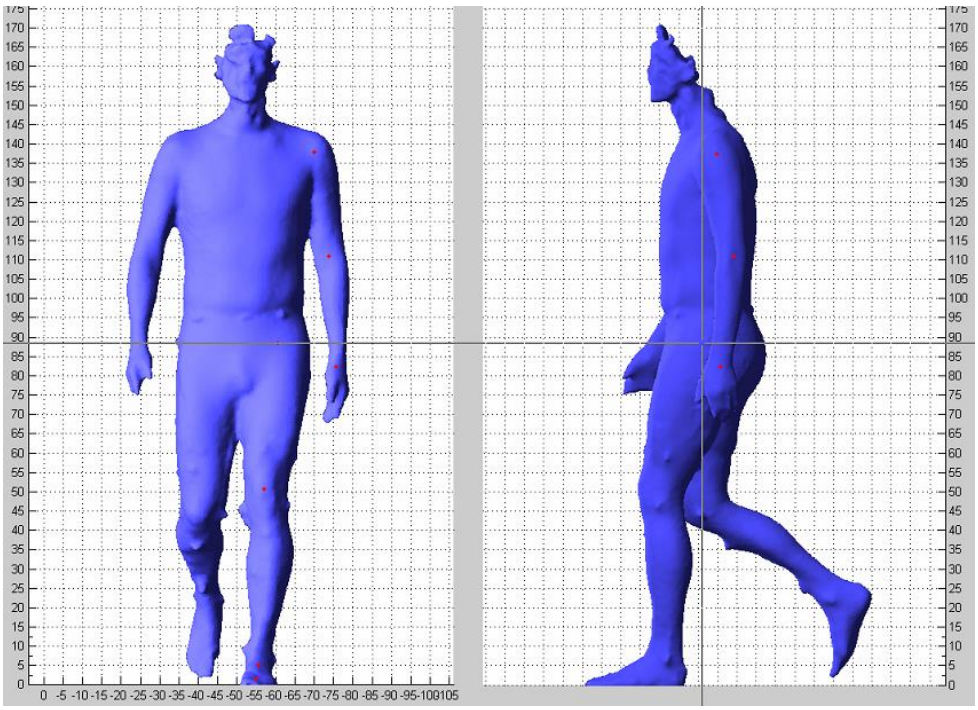

Figure 9
Screenshot from the annotation of the 3D reconstructions (P1)

whereas the elbow, knee and ankle joint angles were computed relative to the proximal limb segment.

Test statistics for a two-factor factorial with random factors were applied to analyse the variability in the data under the assumption that the extracted parameters y_{ij} could be described by the following linear model:

$$y_{ij} = \mu + \tau_i + \beta_j + \epsilon_{ij} \begin{cases} i = 1,2,\dots,I \\ j = 1,2,\dots,J \end{cases} \quad (1)$$

where I = 6 participants, J = 8 observers, μ is the mean parameter value and the model parameters τ, β and ϵ are participant specific bias, observer specific bias and the error, respectively. The model parameters was assumed to be independent random variables and the total variance of the parameters was

$$V(y_{ij}) = \sigma_\tau^2 + \sigma_\beta^2 + \sigma^2 \quad (2)$$

where σ_τ^2, σ_β^2 and σ^2 are the variance components of the participants, observers and the error, respectively. The error accounted for the variability of the annotated parameters when the same participant was annotated by the same observer. This is referred to as intra-observer variability. The participant component accounted for the variation of the parameters between participants and the observer component accounted for the variation between the observers also referred to as the inter-observer variability. The variance component was computed by a two-way ANOVA.

Recognition of the participants was performed using Random forests [29] and the discriminatory power of the parameters included were estimated by the out-of-bag variable importance [29]. Random forest classification was preferred because 1) random forests perform well and they are simple to tune, 2) the distribution of the votes from the trees allows estimation of a certainty score and 3) the parameter importance is estimated automatically. This provides substantial information that can support the decisions made by the forensic anthropologists. The recognition was based 3D reconstructions from 16 participants, which were annotated by the same expert. Five trials from each day were applied as test and training data respectively. The training data were applied to train the random forest, whereas recognition was performed on the test data.

AUTOMATIC POSE ESTIMATION

Automatic extraction of gait parameters is preferred to avoid human bias, which can be obtained with model based markerless motion capture. The pipeline of the proposed model-based motion capture system consists of four steps: 1) Data acquisition, 2) 3D reconstruction, 3) articulated model generation and 4) pose estimation. Data acquisition comprises the hardware and software required to capture the image data. The 3D reconstruction is the conversion of recorded image data to 3D surface meshes. Articulated model generation is the synthesis of the articulated model that will be fitted to the 3D reconstructions to estimate the poses. In pose estimation the pose is estimated by fitting the articulated model through numerical optimization. Step 1 and 2 are described in THE GAIT LABORATORY section whereas step 3 and 4 are described in the following sections.

Articulated model generation

This study focused on the lower extremities as forensic gait analysis as well as clinical gait analysis often only concern this [8, 11, 18]. The articulated model thus consisted of the pelvis, thigh, shank and foot. Articulated models can either be synthesised by fully automatic approaches [25] or by manual annotation of the joint centres like the proposed method in THE VARIABILITY OF GAIT AND THE EFFECT ON RECOGNITION section. However, as the proposed method was compared to marker-based motion capture, joint centres and reference frames of the segments were transferred directly from the marker based on a 3D reconstruction of the participant in a standing pose to obtain similar initial conditions between the two methods. The 3D reconstruction was automatically segmented into the limb segments by associating each 3D point to the nearest line segment going through the joint centres. The synthesis of the articulated model is shown in Fig. 10.

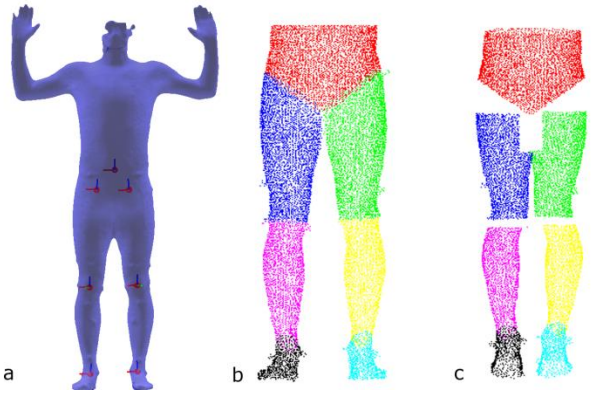

Figure 10
Synthesis of the articulated model. a: 3D reconstruction with labeled joint centers and orientations of the limb segments; b: segmentation of the 3D model; c: the articulated model with all joint angles set to zero and points near the knee and the hip joint removed (P2)

Pose estimation

Pose estimation was performed by fitting the articulated model to the 3D reconstructions using a hierarchical Iterative Closest Point (ICP) algorithm. The algorithm was hierarchical in the sense that the pose of the proximal limb segments was estimated before the pose of the distal segment. The order of the pose estimation was therefore: Pelvis, thigh, shank and foot.

The cost function that was minimised in the ICP algorithm was based on the Euclidian distances between the point correspondences defined by the nearest neighbour criteria

$$\min_{\mathbf{p}_r \epsilon \mathbb{R}^3} \| f(\boldsymbol{p}_a) - \boldsymbol{p}_r \|_2 \quad (3)$$

where pa is a point in the articulated model and pr is a point in the 3D reconstruction. The function $f(\boldsymbol{p}_a)$ is the transformation function defined by

$$f(\boldsymbol{p}_a) = \boldsymbol{R}\boldsymbol{p}_{\boldsymbol{a}} + \boldsymbol{t} \quad (4)$$

where R is the rotation matrix and t is the translation vector. Since rotation and translation is relative to the local coordinate system of the proximal limb segment, the resulting rotation and translation is calculated by

$$\boldsymbol{R}_{total} = \boldsymbol{R}_{proximal} \cdot \boldsymbol{R}_{distal} \quad (5)$$

$$\boldsymbol{t}_{total} = \boldsymbol{t}_{proximal} + \boldsymbol{t}_{distal} \quad (6)$$

The final cost function for each limb segment is hereby

$$\min_{\boldsymbol{R}_{distal} \in \mathbb{R}^3, \boldsymbol{t}_{distal} \in \mathbb{R}^3} \frac{1}{N_a} \sum_{a=1}^{N_a} \min_{\mathbf{p}_r \epsilon \mathbb{R}^3} \| f(\boldsymbol{p}_a) - \boldsymbol{p}_r \|_2^2 \quad (7)$$

where Na is the number of points in the articulated model and Rdistal and tdistal are the variables to be optimised with 6 DoF for each limb segment resulting in 42 DoF for the articulated model in total. ICP was iterated according to Besl and McKay [38] until convergence. To speed up the computation time, kd-trees were used to find the nearest neighbour correspondences between the articulated model and the 3D model. The articulated model fitted to a 3D reconstruction is illustrated in Fig. 11.

Figure 11
Pose estimation by fitting an articulated model (red) to a 3D reconstruction (cyan) (P2)

ICP requires a qualified starting guess. Fit n-1 was therefore used as starting guess in frame n. To make a qualified starting guess for frame one, it was assumed that the participant was standing upright facing the direction of motion. The initial translation of the pelvis of the articulated model was roughly estimated by aligning the sacral landmark of the articulated model with the mean position of the points of 3D reconstruction in frame one.

From the relative orientations of the segmental reference frames in the fitted articulated model, joint kinematics could be extracted directly with the kinematic equations by Vaughan, Davis [39]

Experimental work

The markerless motion capture system was validated on ten gait trials from ten randomly selected participants. The kinematics obtained from the markerless approach were compared to data obtained from a marker-based motion capture system. The marker setup consisted of 15 markers placed on sacrum, the left and right anterior superior iliac spine (ASIS), the medial and lateral femoral epicondyles, the medial and lateral malleolus the calcaneal bone and the metatarsal head V. The hip joint centres were predicted with the regression equations by Davis, Ounpuu [40] and the knee and ankle joint centres were predicted by the midpoint between the epicondyle markers and the malleoli markers, respectively. The orientation of the limb segments was defined by a plane intersecting the proximal joint centre and the two markers defining the distal joint centre. The local coordinate system of the ankle was defined according to Vaughan, Davis [39]. To use the same joint angle conventions as the marker based model, the joint angles were computed using the equations by Vaughan, Davis [39]. The 3D positions of the predicted joint centres from the marker-based system were transferred to the 3D reconstruction of the articulated model to obtain comparable data. The joint angles were low-pass filtered at 10 Hz (zero-phase-lag filtered with a 6th order Butterworth filter with cut off frequency at 10 Hz) to reduce high frequency noise on both the marker based and markerless results. Only the joint angles from the left leg are presented in the present study, since the differences between the left and right leg were negligible.

Statistics

The mean difference, understood as the mean of the difference between the joint angles obtained with the present approach and the marker-based system, was quantified with its

standard deviation (SD) to investigate the bias and the variability. The Root-Mean-Square Deviation (RMSD) and the Range of Motion (RoM) averaged over all the participants were also quantified in order to provide a relative measure of the differences.

The average difference was also obtained for heel strike (HS), mid-stance (MST) (corresponding to the maximum knee flexion angle in the stance phase), toe off (TO) and mid-swing (MSW) (corresponding to the maximum knee flexion in the swing phase). The level of significance was evaluated with a paired, two-tailed students t-test with a level of significance at 95%. MATLAB's statistics toolbox (Math Works Inc. MA, USA) was used for this purpose.

NEW REGRESSION EQUATIONS

In P3 the errors of the joint centre predictions obtained with the Conventional Gait Model (CGM) were quantified and a new set of equations were regressed to correct for biases. The CGM predicts the joint centres in the lower extremities based on anthropometric measures. The regression equations were originally proposed by Davis, Ounpuu [40] and Kadaba, Ramakrishnan [41] and the most common modification is described by Vicon® [42]. The new set of equations were regressed with the same anthropometric measures as those applied in the regression equations described by and Vicon® [42] but with the pelvic depth added as a parameter and was defined as the distance between the midpoint of the ASIS-markers and the sacral marker.

The new hip joint centre regression equation was obtained by forward-stepwise selection [43]. The forward-stepwise selection was performed in three steps: 1) Start with a model equal to zero. 2) For each variable not in the model, try to include it and check if it contributes to the reduction of the sum of squared error of the regression using students t-test. If any variables have p-values less than p=0.05, add the one with the smallest p-value and repeat this step, else, go to step 3. 3) For each variable in the model, test for sex differences in the coefficients for each model variable. If any variables have p-values less than p=0.05, add the one with the smallest p-value and repeat this step, else, end. The differences between sexes were tested in a separate step to avoid falling into non-plausible local minima.

The sensitivity of the regression coefficients was evaluated with Cook's Distance [44]. A high sensitivity could be caused by insufficient samples, outliers or the coefficient being close to zero. The sample with the largest Cook's Distance in the data set was removed and the regression rebuilt. In case a coefficient changed more than 10%, the variable with the highest p-value was excluded from the regression. By reducing the model, a little bias was sacrificed to reduce the sensitivity of the predicted coefficients. The level of significance of the new coefficients was quantified with the p-value of the t-statistics. Differences of the mean errors between regression equations were tested with a one-way ANOVA with Tukey's correction and a level of significance of 95%.

Only one anthropometric parameter was included in the existing regression equations for the KJC and the AJC. Therefore, new regression equations for these joints were regressed with a General linear model (GLM) using the same parameters but in new local reference frames. Constants in the regression equations were not included to avoid over-fitting. The new reference frames were defined by the proximal joint centre, the proximal wand marker and the lateral epicondyle/malleolus marker. The x-axis was defined by the normal plane formed by the markers pointing anteriorly. The z-axis was parallel to the vector pointing from the lateral epicondyle/malleolus marker to the proximal joint centre. The y-axis was orthogonal to the x-axis and z-axis pointing medially.

The prediction errors of regression equations are reported relative to the reference frame of the proximal limb segment. The pelvic reference frame is defined with a right handed coordinate system so the z-axis is the normal of the plane defined by the ASIS markers and the sacral marker pointing superiorly. The y-axis is a vector parallel to the ASIS markers and the x-axis is perpendicular to these axes pointing anteriorly. For the thigh, the reference frame is defined so that the HJC and the KJC define the z-axis, lateral epicondyle marker and the KJC define the y-axis and the x-axis is perpendicular to these pointing in the anterior direction. In this study the y-axis is pointing medially for both knees to make interpretation of the results easier. The reference of the calf is defined so that the KJC and the AJC define the z-axis, the malleolus marker and the AJC define the y-axis and the x-axis is perpendicular to these axes pointing in the anterior direction. Again, the y-axis is pointing medially for both ankles to facilitate interpretation of the results. The prediction error ($\boldsymbol{\varepsilon}$) is hereby defined by

$$\boldsymbol{\varepsilon} = M(\boldsymbol{y}) - \boldsymbol{y} \qquad (8)$$

where M(y) is the joint centres predicted with the regression equations and (y) is the 'true' joint centres extracted from the scans.

RESULTS & DISCUSSION

THE ANTHROPOMETRIC REFERENCE DATA

A MRI scanned femur phantom was applied to quantify the measurement error caused by the 3D registration of the MRI scans and the inhomogeneous magnetic field. The registration is illustrated in Fig. 12 with the regressed linear model applied to quantify the error. The SD residual of the regressed linear model was estimated to 1.2 mm and the Root-mean-squared-error (RMSE) of the 3D registration was estimated to 3.9 mm. The RMSE of the 3D registration was high compared to the residual of the regressed linear model. However, compared to average bone lengths of adults that usually span between 40 and 50 cm, the prediction error is smaller than one percent, which is acceptable.

The result of a 3D registration of four MRI scans of the lower extremities is illustrated in Fig. 13. Note the mismatch between the overlapping images near the image borders at the proximal end of the thigh, which is a result of the inhomogeneous magnetic field. This could have a severe effect on the anthropometric measures but the error quantification obtained from the femur phantom showed that the distortion is limited near the long bones where the anthropometric measures was extracted.

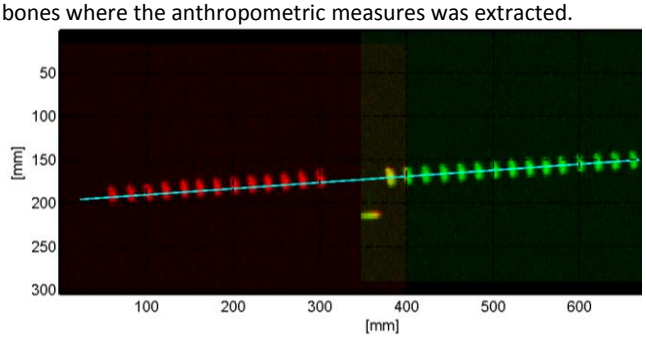

Figure 12
Sagittal view of the MRI scanned femur phantom. Green intensity image is the scan of the 15 markers used for the regression. The red intensity image is the scan of the 13 markers used for prediction. The cyan line represents the linear regression that was estimated from the green markers

Figure 13
MRI of the lower extremities consisting of four registered scans

THE GAIT LABORATORY

The accuracy of the 3D reconstructions obtained with the proposed laboratory setup was quantified with a laser scan of a mannequin dummy as a reference. The RMSE were estimated to 1.76 mm, the mean error to 1.26 mm and the median error to 0.97 mm and the distribution of the errors on the 3D reconstruction is also visualised in a heat map in Fig 14. The results show that the proposed setup was capable of reconstructing human bodies with a high precision allowing anatomical landmarks to appear distinct on the surface. However, holes caused by self-occlusion as in the areas near the armpit and the groin causes considerable errors, which might confuse expert observers in the annotations of the nearby joint centres. The accuracy achieved with the multi-view stereo approach outperforms the Shape-From-Silhouette approaches, which was validated for human motion analysis by Mündermann, Corazza [45] and is probably the most applied algorithm in laboratory setups intended for this purpose [14, 23, 46-48]. Comparing the current setup with other laboratories using stereo vision based algorithms for human motion analysis [49] the improvement gained by the current setup is the reduced number of cameras from 20 to 8 and a frame rate increased from 30 fps to 135 fps. However, the leopard spandex suit is a drawback for the current approach, which makes it infeasible for surveillance but the suit can be replaced by projection of a random or pseudorandom pattern onto the participant with an infrared projector. A similar accuracy may be achievable with other depth sensors as the Kinect 2® (Microsoft, Redmond, Washington, USA). Such alternatives may therefore provide new opportunities in surveillance to obtain reliable gait recognition in the future.

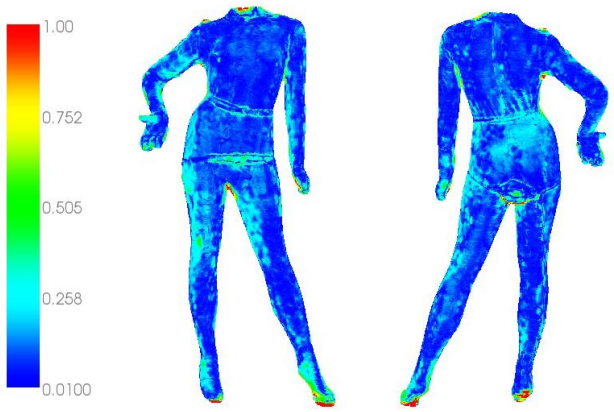

Figure 14
The error of the stereo vision based 3D reconstruction. The error equals the distance between the nearest neighbour point correspondences in the current 3D reconstruction and the laser scan (P1)

THE VARIABILITY OF GAIT AND THE EFFECT ON RECOGNITION

The 3D reconstructions obtained from the laboratory setup were applied for recognition based on limb segment lengths and gait kinematics. The validity of the parameters is dependent on the variability between participants and the measurement errors related to the inter- and intra-observer variability. The inter- and intra-observer variability and the systematic errors of the segment lengths are illustrated in Fig. 15. The results show that the observers generally underestimated the thigh length, the forearm tended to be overestimated whereas the error of the shank was small compared to the other segment lengths. The shank also provided the highest agreement between observers.

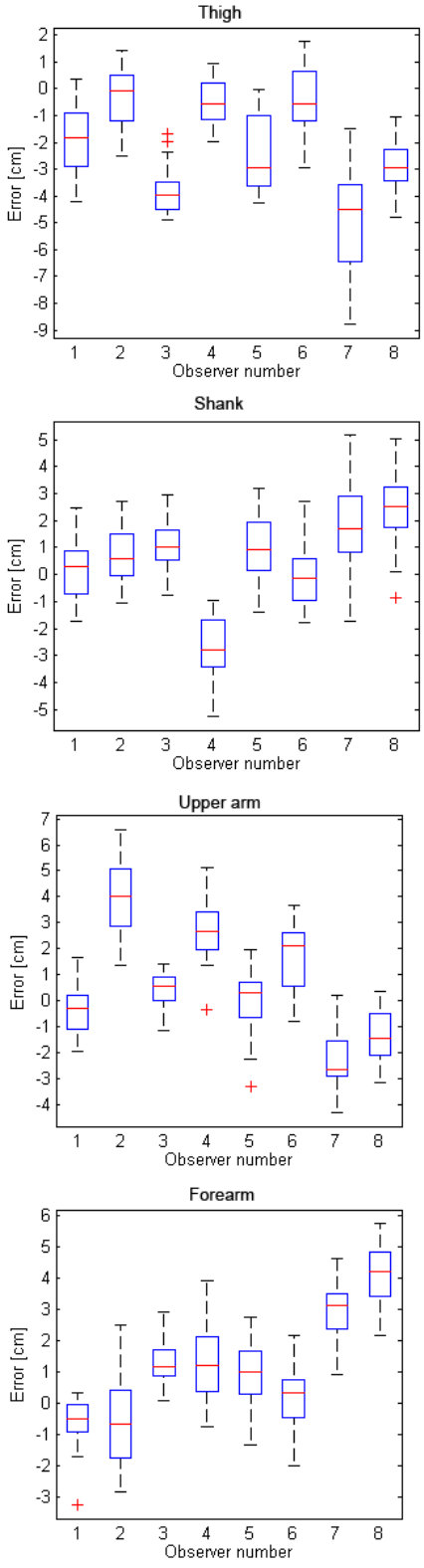

Figure 15
Box plots of the segment length errors. Errors are relative to the segment lengths measured from the MRI scans. Positive errors are overestimation and negative errors are underestimation. The red central mark is the median, the edges of the boxes are the 25th and 75th percentiles, the whiskers extend approximately to +/–2.7σ corresponding to 99.3% coverage if the data are normally distributed (P1)

The pie charts in Fig. 16 show that the majority of the parameters are dominated by inter observer variability, which means that the parameters are difficult to compare between observers but also the intra-observer component is large compared to the participant component. The shank length in general and the length between the shoulders measured at heel strike were less affected compared to the other parameters. Compared to the results by Larsen, Hansen [9] the inter-observer variability of the estimated segment lengths was markedly reduced by the current approach but this was still not sufficient to obtain valid comparisons between observers.

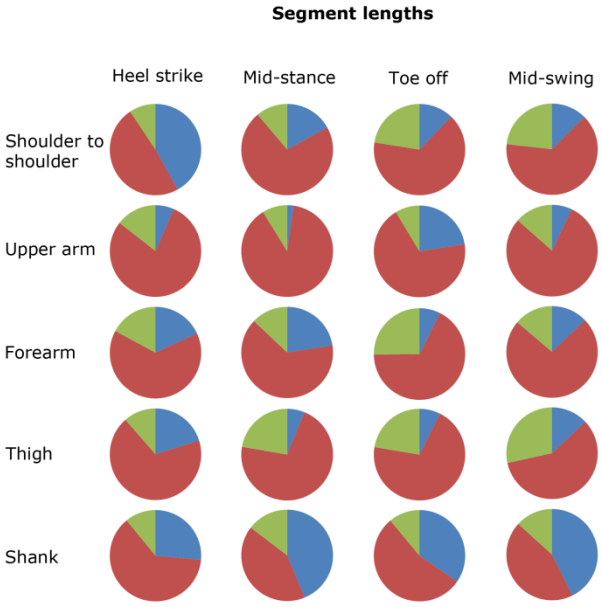

Figure 16
The pie charts of the variance components of the segment length measured during heel strike, mid-stance, toe off and mid-swing. The red, green and blue colour represent the variance components of inter observers, intra observers and participants, respectively (P1)

Considering the sagittal kinematics illustrated in Fig. 17, ankle joint angles are not comparable between observers and neither are any of the parameters in the mid-stance phase. On the contrary, parameters during heel strike and toe-off provided high reliability overall. Regarding the frontal kinematics illustrated in Fig. 18, the elbow joint angles are difficult to estimate precisely and the shoulder joint angles and the hip joint angles are generally not comparable between observers. With the exception of sagittal ankle and frontal elbow joint kinematics, the observers were highly consistent in extraction of segment lengths and kinematics but they clearly had a different intuitive understanding of the joint centre locations resulting in a high inter-observer variability of the segment lengths in particular. Training the experts or providing a common understanding of the joint centre locations among the observers might have a reducing effect on the inter-observer variability.

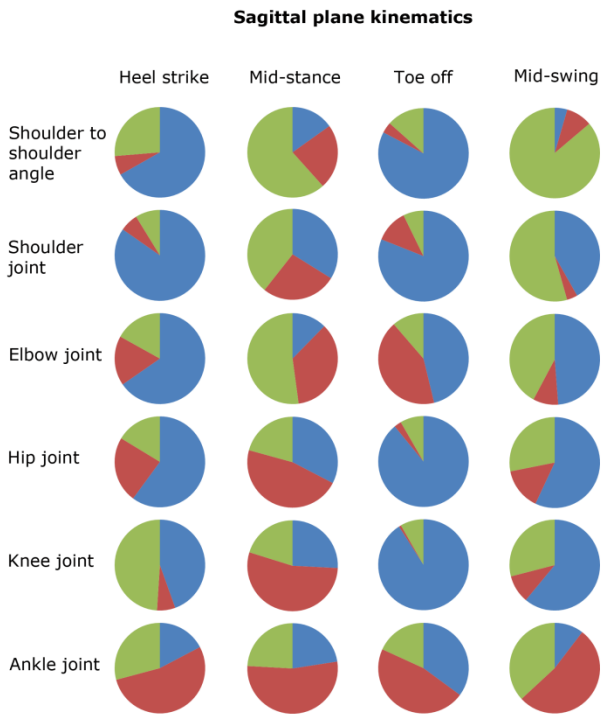

Figure 17
The pie charts of the variance components of the sagittal plane kinematics measured during heel strike, mid-stance, toe off and mid-swing. Red, green and blue represent the variance components of inter observers, intra observers and subjects, respectively (P1)

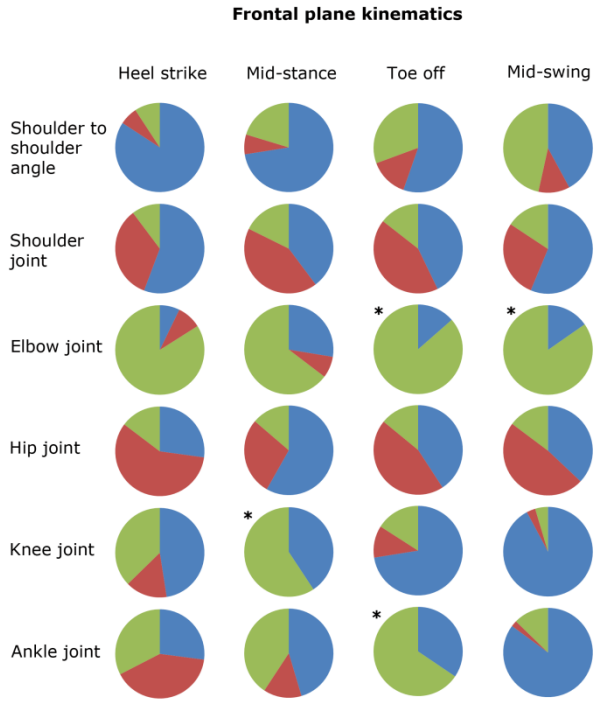

Figure 18
The pie charts of the variance components of the sagittal plane kinematics measured during heel strike, mid-stance, toe off and mid-swing. Red, green and blue represent the variance components of inter observers, intra observers and subjects, respectively. * indicates negative variance component for the inter observer factor, which was set to zero [50] (P1)

As the expert observers annotated the landmarks on the 3D reconstructions from two fixed views as illustrated in Fig. 5 the effect of using 3D reconstructions might be limited compared to a synchronised multi-view CCTV system with cameras directed frontally and sagittally to the participants. However, these criterions are far from trivial to meet in practice.

Random forests were applied for recognition and ranking of the parameter importance. The random forest was built with 250 trees as the out-of-bag recognition error hereby reached convergence. The fraction of correct recognitions as a function of the number of parameters included in the recognition is illustrated in Fig. 19. Parameters were included according to the ranking of the parameter importance listed in Fig. 20 in decreasing order. The results show that single parameters have a relatively weak discriminatory power with up to 40% correct recognitions when recognition is based on one gait trial and this was not considerably improved when based on five gait trials. With inclusion of ten parameters, all participants were correctly recognised based on five trials and the correct recognition rate for one gait trial converged around 30 parameters with over 99% correct recognitions.

The ranking of the parameter importance listed in Fig. 20 shows that shank lengths, shoulder to shoulder angles, shoulder joint angles and the knee joint angles were among the most important parameters and both the frontal and the sagittal parameters were considerable important for recognition. The importance of the shank length and the shoulder joint angles were surprising since no previous studies have reported these parameters to be of considerable importance in human recognition and contrary to Larsen, Hansen [9] our findings showed that shank lengths are comparable under the assumption that the same observer performs the measurements on all the participants.

Studying the inter- and intra-observer variability and the associated effect on recognition are important in understanding the reliability of gait recognition and such study had to be conducted with a controlled experimental setup to reduce the impact from other factors. Challenges related to gender, adiposity, clothing, lighting conditions etc. were therefore not considered in this study. These parameters are highly relevant and have to be addressed in future studies to reflect the practical issues related to surveillance.

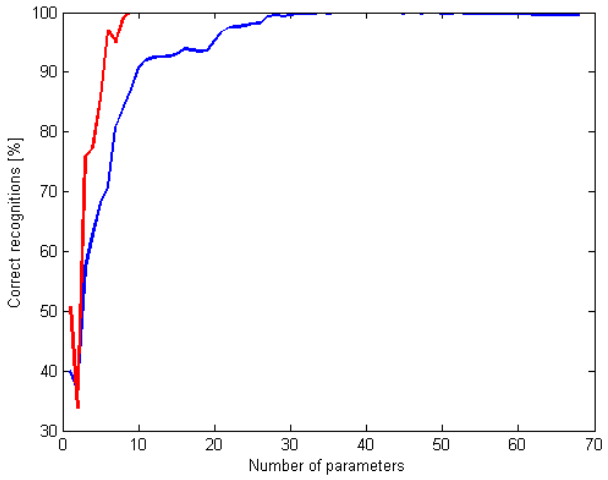

Figure 19
Correct recognitions as function of number of features included in the random forest. Parameters were included according to their importance in decreasing order. The blue line and the red line are recognition based on one and five gait trials, respectively (P1)

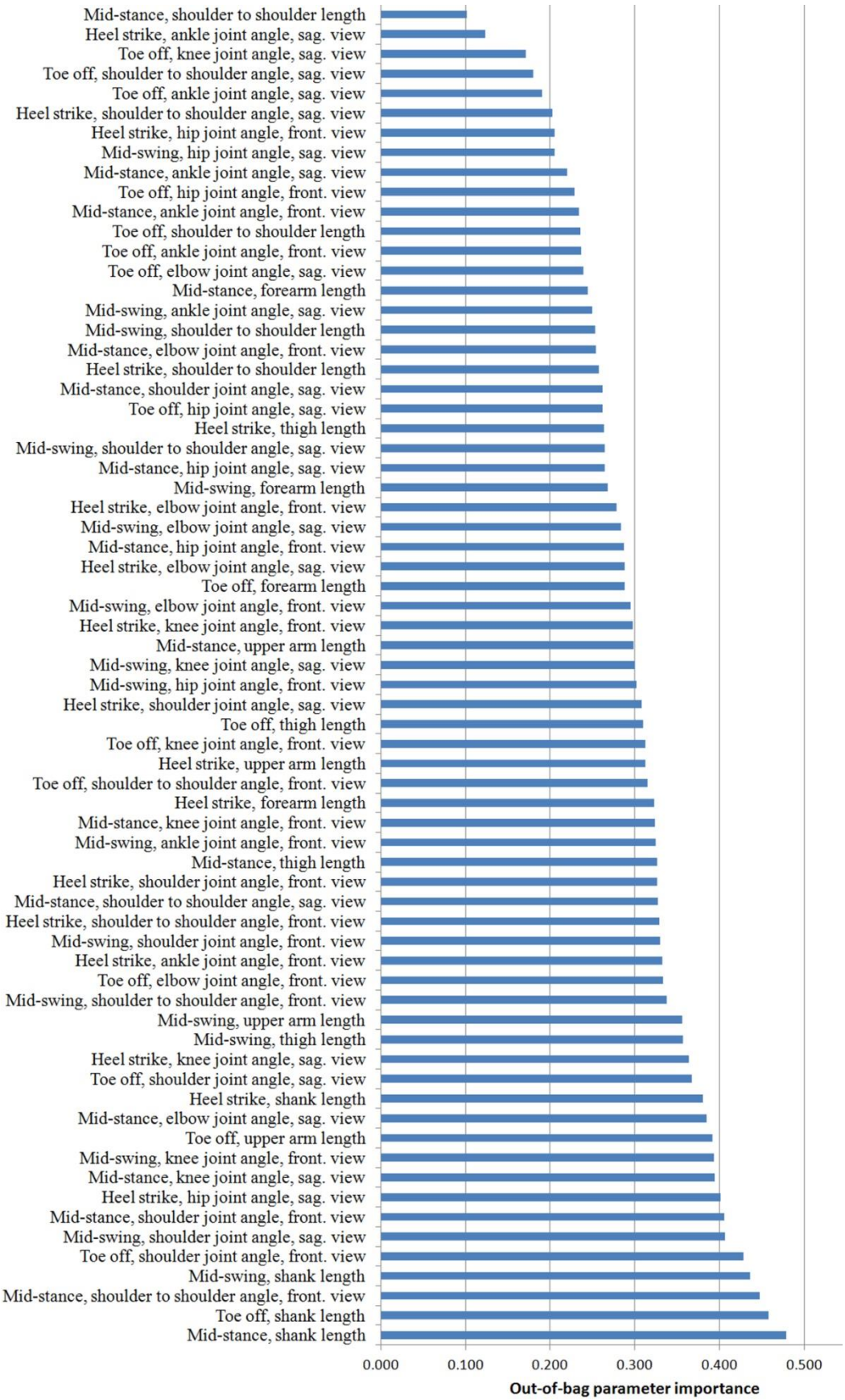

Figure 20

The out-of-bag parameter importance is ranked in increasing order, such the most important parameter is in the bottom of the list. Out-of-bag parameter importance is the out-of-bag error increase as a fraction of the error before the parameter was randomly permuted (P1)

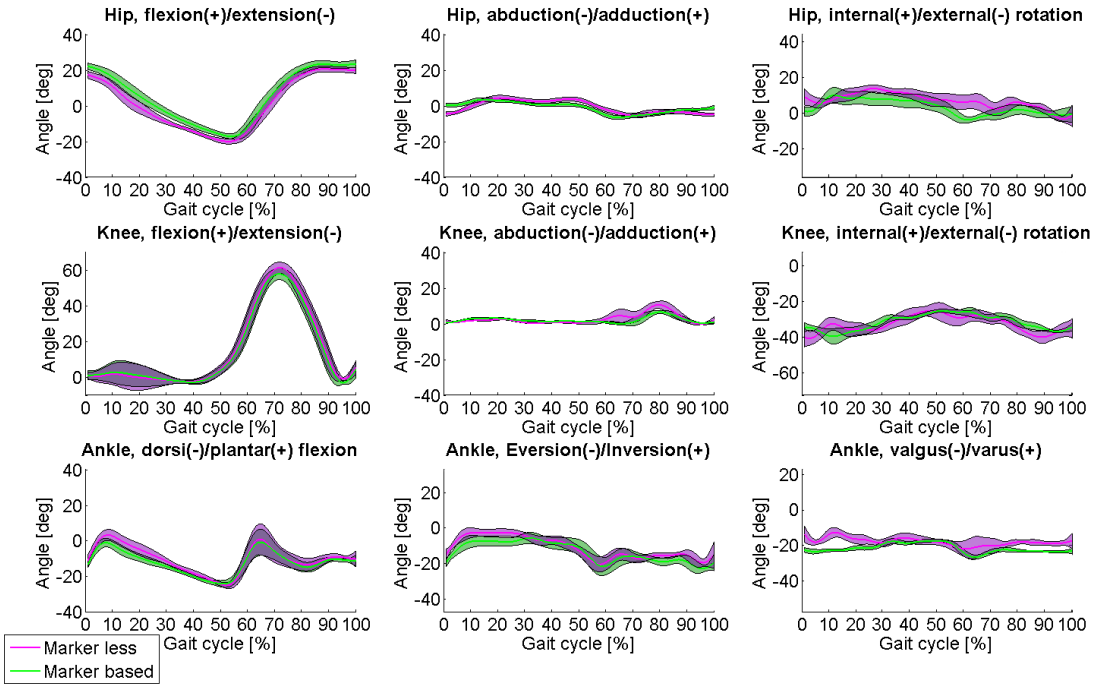

Figure 21
Mean and 95% confidence intervals for the joint angles from an arbitrary test participant (P2)

AUTOMATIC POSE ESTIMATION

The poses were also estimated automatically on the 3D reconstructions with markerless motion capture (P2) and the results were compared to a marker-based motion capture system. The mean difference and the SD of the joint angles were sampled during heel strike, mid-stance (corresponding with maximum knee flexion angle during the first 60% of stance), toe-off and mid-swing (corresponding with maximum knee flexion angle during the mid-swing), which are illustrated in Table 1. All joint angles were significantly different from the marker-based system with exception of a few angles that differed less than 0.3 degrees. However, the SD was generally high compared to the mean difference with a slight exception of the knee joint angles.

It was noticed in the results presented in Table 1 that the knee FE angle was overestimated during the whole gait cycle compared to the marker-based approach. This difference was likely to be induced by STA affecting the markers on the femoral condyles. Reinschmidt, vandenBogert [51] and Benoit, Ramsey [52] have both reported an underestimation of the knee joint angle when using skin markers where Benoit et al. reported a mean and SD of the absolute rotational error of 2.8 (2.6), 2.4 (2.0), 2.7 (2.4) degrees during HS, MST and TO, respectively. Similarly, the mean and SD of the differences between the markerless and the marker based system were 2.8 (1.1), 1.8 (1.7) and 3.6 (1.5) degrees in the three phases, respectively, which closely resembled the errors of the marker based system. Hence, the current approach was likely to be more accurate for knee FE.

The dorsi/plantar flection in the ankle joint were also very promising in relation to previous markerless approaches [23] as the highest obtainable mean error was -1.0 (2.7) degrees compared to 3.5 (8.2) degrees.

Fig. 21 illustrating the average and 95% confidence interval for an arbitrary test participant shows that the variability of the trials was similar for the two approaches with a slight exception of knee IE rotation and ankle valgus/varus, which indicates that observed variation between the gait trials is mainly related to the variability in gait. The hip IE rotation seemed to be more internally rotated about toe off (40-60% of the gait cycle), which also appeared for other participants in various degrees.

McGinley, Baker [53] reviewed the session to session (or day to day) variability of kinematic measurements for marker-based systems and reported that the SD of repeated marker-based

Table 1
The mean difference and the STD of the rotation angles during heel strike, mid-stance, toe-off and mid-swing. Those who are labelled with * was not significant at a level of significance at 95% (P2)

Joint	Angle	Heel strike [°]	Mid-stance [°]	Toe-off [°]	Mid-swing [°]
Hip	Flexion/extension	0.4 (2.3)	-1.6 (2.3)	-0.2 (2.2)*	1.4 (2.7)
	Abduction/ adduction	2.1 (1.4)	0.5 (1.4)	-0.4 (1.3)	0.6 (1.2)
	Internal external rotation	-2.1 (5.7)	0.1 (4.5)*	0.8 (4.7)	0.1 (4.2)*
Knee	Flexion/extension	2.8 (1.1)	1.8 (1.7)	3.6 (1.5)	5.4 (2.0)
	Abduction/ adduction	0.4 (1.5)	0.0 (2.1)*	0.9 (3.1)	-0.2 (3.5)*
	Internal external rotation	0.8 (5.5)	2.5 (4.0)	2.5 (4.1)	0.8 (4.2)
Ankle	Dorsiflexion/plantar flexion	-0.5 (2.1)	0.9 (2.0)	-1.0 (2.7)	-0.7 (2.0)
	Eversion/inversion	3.9 (2.8)	1.5 (2.5)	1.7 (2.9)	1.5 (2.5)
	Valgus/varus	2.7 (2.8)	4.0 (2.6)	1.4 (4.5)	3.4 (2.6)

measurements was between 2 to 5 degrees and inconsistent marker placement was considered the main error source. By using the proposed markerless system, issues regarded to this would be eliminated by applying the same articulated model from session to session under the assumption that the body shape remains unchanged between sessions. However, a few disadvantages are related to the system: 1) Fitting articulated models assume limb segments to be rigid, which means that deformations caused by muscle contractions and wobbling masses interfere with accurate pose estimation. In the present study the thigh and the pelvis segment seemed to suffer from this in particular. This is likely to be the most critical factor when estimating the IE rotations in the hip and the knee joint. The patella and the shank are reliable indicators for the orientation of the thigh [54, 55]. Integrating this in the pose estimation might improve the accuracy but has to be explored further. 2) The joint centres and the orientation of the limb segments have to be annotated manually on a static 3D reconstruction when generating the participant specific articulated model. This is somewhat time consuming but still more efficient than the physical placement of the markers.

Inter-observer variability will not be eliminated either, due to the manual annotation. This could be avoided with a fully automated method for articulated model generation as proposed by Corazza, Gambaretto [25]. 3) The proposed approach is computationally heavy with a computation time of approximately three minutes for each frame on an Intel® Core™ i7 2.67 GHz CPU. This was mainly caused by the PMVS algorithm and the PSR algorithm, which together accounted for two minutes computation. Distortion correction and the pose estimation did also contribute to the long computation time. It should be noted that all processing was performed in MATLAB with the exception of the stereo algorithms. GPU processing or C++ compilation would greatly reduce the computation time. 4) The system has a limited record time, because only RAM and on-board-memory in the cameras could store the data with sufficient speed. This system has 7.5 GB RAM dedicated for data storage resulting in recordings up to 3.3 seconds.

Limitations of the present study design are mainly that only healthy young men were used as participants and only one fixed walking velocity was used. However, to the best of our knowledge, difficulties with patients as opposed to healthy participants concerning the use of marker-based systems have not been reported. Accordingly, it may be assumed that the same will be the case for markerless approaches. Extremely obese participants are known to increase inaccuracies with maker-based systems, but this may not necessarily be the case with markerless systems, which have no problems with hidden markers and assumably less problems with wobbling masses. Regarding slower and faster walking velocities, it cannot be predicted whether or not this could have any influence of the performance of a markerless system, but it seems rather unlikely. However, future research should be directed towards the possible limitations mentioned.

NEW REGRESSION EQUATIONS

The new regression equations for the CGM were estimated by forward-stepwise selection and GLM with the following results:

$$\begin{aligned} p_{\text{Hip, male}} = p_{\text{mid ASIS}} &- 0.324(\text{Pelvis depth})\mathbf{x}_{\text{pelvis}} \\ &\pm 0.095(\text{Leg length})\mathbf{y}_{\text{pelvis}} \\ &- 0.092(\text{Leg length})\mathbf{z}_{\text{pelvis}} \end{aligned} \quad (9)$$

$$\begin{aligned} p_{\text{Hip, female}} = p_{\text{mid ASIS}} &- 0.324(\text{Pelvis depth})\mathbf{x}_{\text{pelvis}} \\ &\pm 0.103(\text{Leg length})\mathbf{y}_{\text{pelvis}} \\ &- 0.092(\text{Leg length})\mathbf{z}_{\text{pelvis}} \end{aligned} \quad (10)$$

$$\begin{aligned} p_{\text{Knee}} = p_{\text{Femoral epicondyle}} &- 0.024(\text{Knee diameter})\mathbf{x}_{\text{thigh}} \\ &+ 0.484(\text{Knee diameter})\mathbf{y}_{\text{thigh}} \\ &+ 0.133(\text{Knee diameter})\mathbf{z}_{\text{thigh}} \end{aligned} \quad (11)$$

$$\begin{aligned} p_{\text{Ankle}} &= p_{\text{Lateral malleolus}} + 0.86(\text{Malleolus width})\mathbf{x}_{\text{calf}} \\ &+ 0.570(\text{Malleolus width})\mathbf{y}_{\text{calf}} \\ &- 0.115(\text{Malleolus width})\mathbf{z}_{\text{calf}} \end{aligned} \quad (12)$$

The prediction errors provided by the original and the new regression equations are listed in Table 2. The results showed significant errors with RMSE up to 28.1 mm, 9.0 mm and 14.6 mm for the existing hip, knee and ankle regression equations. The observed error for the KJC is relatively small compared to the STA which is typically up to 20 mm in the knee region [27]and because the fixed joint centre is a simplification of an instantaneous joint centre [56]. Regarding the KJC and the AJC, the error may be acceptable in a comparison with a markerless system as the kinematic effect may be limited. However, as the errors were most dominant in the proximal/distal direction, they may have a critical effect on the computations of the kinetics as the errors propagate through the segments [57, 58]. The new regression equations significantly reduced the RMSE to 9.2 mm, 3.5 mm and 3.6 mm for the HJC, KJC and AJC, respectively.

The marker size is not accounted for in the new regressions. However, in the modifications of the CGM these corrections are quite simple to implement, because the corrections primarily concern a single axis. It is therefore recommended to correct for the

Table 2
The prediction errors provided by the original regression equations in the CGM and the new regression equations (P3)

Equations	**x** Ant.(+)/post.(-) [mm]	**y** Lat.(+)/med.(-) [mm]	**z** Sup.(+)/inf.(-) [mm]	RMSE total [mm]
Hip equations				
Original equation	17.4 (5.6)/1.9·10^{-11}	-13.2 (11.3)/4.5·10^{-5}	-11.1 (5.2)/1.0·10^{-8}	28.1
New equation	0.3 (6.5)/8.6·10^{-1}	-0.7 (5.2)/5.4·10^{-1}	0.5 (3.7)/ 5.7·10^{-1}	9.2*
Knee equations				
Original equation	2.9 (3.3) / 2.2·10^{-2}	0.0 (0.7) /9.1·10^{-1}	-7.8 (1.2) /8.2·10^{-9}	9.0
New parameters	0.1 (3.3) / 8.8·10^{-1}	0.0 (1.0) / 4.6·10^{-1}	0.0 (1.2) / 8.7·10^{-1}	3.5*
Ankle equations				
Original equation	-5.4 (1.3) / 3.8·10^{-7}	-5.2 (1.8) /9.6·10^{-6}	12.0 (2.9) /3.4·10^{-7}	14.6
New equation	0.0 (1.5) / 9.7·10^{-1}	0.1 (1.8) / 8.3·10^{-1}	0.0 (3.0) / 9.7·10^{-1}	3.6*

marker size by adding the marker radius posteriorly in the proposed HJC regression and medially in the proposed KJC and AJC regressions.

As marker-based motion capture is frequently applied for scientific analysis of gait in 3D, this improvement will provide higher accuracy of gait analysis and thus contribute to the understanding of gait dynamics. The contribution is also expected to provide basis for more reliable comparisons of markerless and marker based motion capture systems in future.

CONCLUSIONS

The contributions of this work occasioned following conclusions:

3D reconstructions are feasible for biometric measurements in forensic sciences as alternative to video data obtained from CCTV. However, uncontrolled factors as gender, clothing, adiposity, lighting conditions, incomplete and noisy 3D reconstructions might challenge the reliability and has to be addressed in the future.

Limb segment lengths, ankle joint kinematics and elbow joint kinematics annotated by different observers are difficult to compare. Despite this, gait recognition can be obtained with high reliability also when the data have been annotated by different observers.

Not only kinematic parameters from the lower extremities are important for gait recognition but also shoulder kinematics and calf length are reliable parameters to consider.

Markerless motion capture systems can provide highly reliable estimation of all FE angles and hip AA angles which can be applied for gait recognition. Large differences in IE rotations were found and these rotations remain to be addressed in the future. However, IE rotations are not necessarily important in gait recognition as existing methods in gait recognition are generally based on joint angles measured in a frontal view plane and a sagittal view plane, which are invariant to IE rotations.

As the bone dimensions are different between males and females, using different regression equations for the sexes provide more accurate predictions of the hip joint centres.

The prediction errors of the CGM were surprising and could have a critical effect on the computed kinetics but these errors were significantly reduced by the new regression equations.

SUMMARY

This thesis is based on four manuscripts where two of them were accepted and two were submitted to peer-reviewed journals. The experimental work behind the thesis was conducted at the Institute of Neuroscience and Pharmacology, University of Copenhagen. The purpose of the studies was to explore the variability of human gait and to conduct new methods for precise estimation of the kinematic parameters applied in forensic gait analysis.

The gait studies were conducted in a custom built gait laboratory designed to obtain optimal conditions for markerless motion analysis. The setup consisted of eight synchronised cameras located in the corners of the laboratory, which were connected to a single computer. The captured images were processed with stereo vision based algorithms to provide accurate 3D reconstructions of the participants.

The 3D reconstructions of the participants were obtained during normal walking and the kinematics were extracted with manual and automatic methods. The kinematic results from the automatic approach were compared to marker based motion capture to validate the precision. The results showed that the proposed markerless motion capture method had a precision comparable to marker-based methods in the frontal plane and the sagittal plane. Similar markerless motion capture methods could therefore provide the basis for reliable gait recognition based on kinematic parameters.

The manual annotations were compared to the actual anthropometric measurements obtained from MRI scans and the intra- and inter-observer variability was also quantified to observe the associated effect on recognition. The results showed not only the kinematics in the lower extremities were important but also the kinematics in the shoulders had a high discriminatory power. Likewise, the shank length was also highly discriminatory, which has not been previously reported. However, it is important that the same expert performs all annotations, as the inter-observer variability was high compared to the variability between the participants.

The MRI scans were also applied to estimate the errors of existing marker-based regression equations to predict the joint centres. The errors in the HJC and the AJC were surprisingly high, which may affect the computations of the joint kinetics and thus the understanding of gait dynamics. On the other hand, the effect on the kinematics would be limited and thus the existing regression equations provide a reliable basis to validate markerless motion capture methods as long as the limitations regarding STA and the placement of the markers are considered in the data interpretation. New regression equations corrected the estimated bias and they also accounted for the significant sex differences in pelvis.

ACKNOWLEDGEMENTS

Funding for this PhD thesis was granted from the Danish Agency for Science, Technology and Innovation. The project was conducted as an Industrial PhD by The Danish Institute of Fire and Security Technology.

A special thanks to:
Erik B Simonsen, Henrik Aanæs, Thomas B Moeslund, Tine Alkjær, Niels Lynnerup, Peter K Larsen, Karl Erik Jensen, Rikke V Heimbürger, Chiara Villa, Peter Raffalt, Henrik Koblauch, Carsten Damgaard,

who have made substantial contributions to the conception and design of the study, the acquisition of data, analysis and interpretation of data, revising the manuscripts critically or supported my work in other ways.

Also thanks to:
- PAR Scientific A/S
- The Neural Control of Movement Laboratory, University of Copenhagen
- The Visual Analysis of People Laboratory, Aalborg University
- The Image Analysis and Computer Graphics Section, DTU

for their hospitality and support of the project.

REFERENCES

1. Højbjerg JH, Hasselgaard O, Vollmer CK. Politiets og Anklagemyndighedens årsrapport. Copenhagen 2014:49.
2. Krog TN, Jørgensen U. BT afslører politi-rapport: Her gør politiet intet - tyve får frit spil. BT. 2014 20th july.

3. Crhristiansen K, Sørensen SN. TV-overvågning. In: Råd DK, ed. Glostrup 2005:25.
4. Cutting J, Kozlowski L. Recognizing friends by their walk: Gait perception without familiarity cues. *Bull Psychon Soc* 1977; 9: 353-6.
5. Lynnerup N, Vedel J. Person identification by gait analysis and photogrammetry. *Journal of Forensic Sciences* 2005; 50: 112-8.
6. Larsen PK, Simonsen EB, Lynnerup N. Gait Analysis in Forensic Medicine. *Journal of Forensic Sciences* 2008; 53: 1149-53.
7. Bouchrika I, Goffredo M, Carter J, Nixon M. On Using Gait in Forensic Biometrics. *J Forensic Sci*: Blackwell Publishing Ltd 2011:882-9.
8. Larsen PK, Lynnerup N, Henriksen M et al. Gait Recognition Using Joint Moments, Joint Angles, and Segment Angles. *Journal of Forensic Biomechanics* 2010; 1: 7.
9. Larsen PK, Hansen L, Simonsen EB, Lynnerup N. Variability of Bodily Measures of Normally Dressed People Using PhotoModeler® Pro 5. *Journal of Forensic Sciences* 2008; 53: 1393-9.
10. Ariyanto G, Nixon MS. Marionette mass-spring model for 3D gait biometrics. 2012: 354-9.
11. Ariyanto G, Nixon MS. Model-based 3D gait biometrics. *2011 International Joint Conference on Biometrics, .* Washington, DC IEEE 2011:1-7.
12. Bouchrika I, Goffredo M, Carter J, Nixon M. On Using Gait in Forensic Biometrics. *Journal of Forensic Sciences* 2011; 56: 882-9.
13. Matovski DS, Nixon MS, Mahmoodi S, Carter JN. The Effect of Time on Gait Recognition Performance. *Information Forensics and Security, IEEE Transactions on* 2012; 7: 543-52.
14. Seely RD, Samangooei S, Middleton L et al. The University of Southampton Multi-Biometric Tunnel and introducing a novel 3D gait dataset. *2nd IEEE International Conference on Biometrics: Theory, Applications and Systems*. Arlington, VA IEEE 2008:1 - 6.
15. Guan Y, Li C-T. A robust speed-invariant gait recognition system for walker and runner identification. *the 6th IAPR International Conference on Biometrics*: IAPR 2013:1-8.
16. Guan Y, Li C-T, Hu Y. Robust Clothing-Invariant Gait Recognition. *Proceedings of the 2012 8th International Conference on Intelligent Information Hiding and Multimedia Signal Processing, Iih-msp 2012 — 2012, pp 321-324* 2012:321-4.
17. Huang PS, Harris CJ, Nixon MS. Recognising humans by gait via parametric canonical space. *Artificial Intelligence in Engineering* 1999; 13: 359-66.
18. Ariyanto G, Nixon MS. Marionette mass-spring model for 3D gait biometrics. 5th IAPR International Conference on Biometrics (ICB), 2012; 2012 March 29 2012-April 1 2012; 2012. p. 354-9.
19. Han J, Bhanu B. Individual recognition using gait energy image. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 2006; 28: 316-22.
20. Shotton J, Fitzgibbon A, Cook M et al. Real-Time Human Pose Recognition in Parts from Single Depth Images. *2011 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* 2011: 1297-304.
21. Taylor J, Sharp T, Fitzgibbon A, Shotton J. The Vitruvian manifold: Inferring dense correspondences for one-shot human pose estimation. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition* 2012: 103-10.
22. Lligadas X, Susín A, Fernández-Baena A. Biomechanical validation of upper-body and lower-body joint movements of kinect motion capture data for rehabilitation treatments. *Proceedings of the 2012 4th International Conference on Intelligent Networking and Collaborative Systems, INCoS 2012* 2012: 656-61.
23. Corazza S, Mündermann L, Chaudhari AM et al. A Markerless Motion Capture System to Study Musculoskeletal Biomechanics: Visual Hull and Simulated Annealing Approach. *Ann Biomed Eng* 2006; 34: 1019-29.
24. Mündermann L, Corazza S, Andriacchi TP. Accurately measuring human movement using articulated ICP with soft-joint constraints and a repository of articulated models. *2007 IEEE Conference on Computer Vision and Pattern Recognition, Vols 1-8* 2007: 2550-5.
25. Corazza S, Gambaretto E, Mündermann L, Andriacchi TP. Automatic Generation of a Subject-Specific Model for Accurate Markerless Motion Capture and Biomechanical Applications. *IEEE Trans Biomed Eng* 2010; 57: 806-12.
26. Corazza S, Mündermann L, Gambaretto E et al. Markerless Motion Capture through Visual Hull, Articulated ICP and Subject Specific Model Generation. *International Journal of Computer Vision* 2010; 87: 156-69.
27. Peters A, Galna B, Sangeux M et al. Quantification of soft tissue artifact in lower limb human motion analysis: A systematic review. *Gait & Posture* 2010; 31: 1-8.
28. Della Croce U, Cappozzo A, Kerrigan C, Lucchetti L. Bone Position and Orientation Errors: Pelvis and Lower Limb Anatomical Landmark Identification Reliability. *Gait & Posture* 1997; 5: 156-7.
29. Breiman L. Random Forests. *Machine Learning* 2001; 45: 5-32.
30. Lagarias JC, Reeds JA, Wright MH, Wright PE. Convergence properties of the Nelder-Mead simplex method in low dimensions. *SIAM JOURNAL ON OPTIMIZATION* 1998; 9: 112-47.
31. Ferris BD, Stanton J, Zamora J. Kinematics of the wrist: Evidence for two types of mevement. *Journal of Bone & Joint Surgery, British Volume* 2000; 82-B: 242-5.
32. Iwaki H, Pinskerova V, Freeman M. Tibiofemoral movement 1: the shapes and relative movements of the femur and tibia in the unloaded cadaver knee. *Journal of Bone & Joint Surgery, British Volume* 2000; 82: 1189-95.
33. van den Boomgaard, van Balen. Methods for fast morphological image transforms using bitmapped binary images. *CVGIP: Graphical Models and Image Processing* 1992; 54: 252-8.
34. Furukawa Y, Ponce J. Accurate, dense and robust multi-view stereopsis. 2010; 32:
35. Harris C, Stephens M. A combined corner and edge detector. Alvey vision conference; 1988: Manchester, UK; 1988. p. 50.

36. Marr D, Hildreth E. Theory of Edge Detection. *Proceedings of the Royal Society B: Biological Sciences* 1980; 207: 31.
37. Kazhdan M, Bolitho M, Hoppe H. Poisson surface reconstruction. *ACM International Conference Proceeding Series* 2006; 256: 61-70.
38. Besl PJ, McKay ND. A Method for Registration of 3-D Shapes. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 1992; 14: 239-56.
39. Vaughan CL, Davis BL, O'Connor JC. Dynamics of human gait. Champaign, Illinois: Human Kinetics Publishers 1992.
40. Davis RB, III, Ounpuu S, Tyburski D, Gage JR. A Gait Analysis Data Collection and Reduction Technique. *Human Movement Science* 1991; 10: 575-88.
41. Kadaba M, Ramakrishnan H, Wootten M et al. Repeatability of kinematic, kinetic, and electromyographic data in normal adult gait. *Journal of Orthopaedic Research* 1989; 7: 849-60.
42. Vicon®. Bodybuilder for Biomechanics. Oxford: Oxford Mertics Ltd 2002.
43. Hastie T, Tibshirani R, Friedman J et al. The elements of statistical learning. New York: Springer 2009.
44. Cook RD. Detection of Influential Observations in Linear Regression. *Technometrics* 1977; 19: 4.
45. Mündermann L, Corazza S, Chaudhari AM et al. Most favorable camera configuration for a shape-from-silhouette markerless motion capture system for biomechanical analysis. SPIE - The International Society for Optical Engineering 2005:278-87.
46. Aguiar Ed, Stoll C, Theobalt C et al. Performance capture from sparse multi-view video. *ACM Trans Graph* 2008; 27: 1-10.
47. Vlasic D, Baran I, Matusik W et al. Articulated mesh animation from multi-view silhouettes. *ACM Trans Graph* 2008; 27: 1-9.
48. Starck J, Hilton A. Surface Capture for Performance-Based Animation. *Computer Graphics and Applications, IEEE* 2007; 27: 21-31.
49. Kun L, Qionghai D, Wenli X. Markerless Shape and Motion Capture From Multiview Video Sequences. *IEEE Transactions on Circuits and Systems for Video Technology* 2011; 21: 320-34.
50. Montgomery DC. Design and analysis of experiments. 5th rev. ed. New York: John Wiley & Sons 2008.
51. Reinschmidt C, vandenBogert AJ, Lundberg A et al. Tibiofemoral and tibiocalcaneal motion during walking: external vs. skeletal markers. *Gait & Posture* 1997; 6: 98-109.
52. Benoit DL, Ramsey DK, Lamontagne M et al. Effect of skin movement artifact on knee kinematics during gait and cutting motions measured in vivo. *Gait & Posture* 2006; 24: 152-64.
53. McGinley JL, Baker R, Wolfe R, Morris ME. The reliability of three-dimensional kinematic gait measurements: A systematic review. *Gait & Posture* 2009; 29: 360-9.
54. Lamoreux LW. Errors in thigh axial rotation measurements using skin mounted markers. *Journal of Biomechanics* 1992; 25.
55. Wren TAL, Do KP, Hara R, Rethlefsen SA. Use of a patella marker to improve tracking of dynamic hip rotation range of motion. *Gait and Posture* 2008; 27: 530-4.
56. Drake RL, Vogl AW, Mitchell AWM, Gray H. Gray's Anatomy for Students: Churchill Livingstone/Elsevier 2010.
57. Stagni R, Leardini A, Cappozzo A et al. Effects of hip joint centre mislocation on gait analysis results. *Journal of Biomechanics* 2000; 33: 1479-87.
58. Schutte MSL. The Effects of Variability of Placement of the Knee Alignment Device on Kinematic Data. *Gait & Posture* 1996; 4.