Accurately measuring human movement using articulated ICP with soft-joint constraints and a repository of articulated models

Lars Mündermann Stefano Corazza Thomas P. Andriacchi
Department of Mechanical Engineering
Stanford University, Stanford, CA 94305
{lmuender, stefanoc, tandriac}@stanford.edu

Abstract

A novel approach for accurate markerless motion capture combining a precise tracking algorithm with a database of articulated models is presented. The tracking approach employs an articulated iterative closest point algorithm with soft-joint constraints for tracking body segments in visual hull sequences. The database of articulated models is derived from a combination of human shapes and anthropometric data, contains a large variety of models and closely mimics variations found in the human population. The database provides articulated models that closely match the outer appearance of the visual hulls, e.g. matches overall height and volume. This information is paired with a kinematic chain enhanced through anthropometric regression equations. Deviations in the kinematic chain from true joint center locations are compensated by the soft-joint constraints approach. As a result accurate and a more anatomical correct outcome is obtained suitable for biomechanical and clinical applications. Joint kinematics obtained using this approach closely matched joint kinematics obtained from a marker based motion capture system.

1. Introduction

Human motion capture is a well established paradigm for the diagnosis of the patho-mechanics related to musculoskeletal diseases and the development and evaluation of rehabilitative treatments and preventive interventions. At present, the most common methods for accurate capture of three-dimensional human movement require a laboratory environment and the attachment of markers, fixtures or sensors on the skin's surface of the body segment being analyzed. These laboratory conditions can cause experimental artifacts. Comparisons of bone orientation from true bone embedded markers versus clusters of three skin markers indicate a worst-case root mean square artifact of 7° [17].

A technique for accurately measuring human body kinematics that does not require markers or fixtures placed on the body would greatly expand the applicability of human motion capture. To date, markerless methods are not widely available because the accurate capture of human movement without markers is technically challenging yet recent technical developments in computer vision provide the potential for markerless human motion capture for biomechanical and clinical applications [24, 6, 26]. Our current approach employs an articulated iterative closest point (ICP) algorithm with soft-joint constraints [1] for tracking human body segments in visual hull sequences (a standard 3D representation of dynamic sequences from multiple images). The soft-joint constraints approach extends previous approaches [5, 6] for tracking articulated models that enforced hard constraints on the joints of the articulated body. Subject-specific 3D articulated bodies were tracked in the visual hull sequences. Matching articulated bodies are obtained from a repository of human shapes and paired with a kinematic chain. Deviations in the kinematic chain from true joint center locations are compensated by the soft-joint constraints approach. As a result accurate and a more anatomical correct outcome is obtained suitable for biomechanical and clinical applications.

The notions of markerless motion capture have already appeared in literature. Following their review (Section 2), we outline our approach (Section 3). In Section 4, we present results. We conclude the paper with a discussion of the results, and problems open for further research (Section 5).

2. Previous work

The development of markerless motion capture systems originated from the fields of computer vision and machine learning, where the analysis of human actions by a computer is gaining increasing interest. Potential applications of human motion capture are the driving force of system development, and the major application areas are: smart surveillance, identification, control, perceptual interface, character animation, virtual reality, view interpolation, and motion analysis [24, 33]. A great variety of vision-based systems have been proposed for tracking human motion. These systems vary in the number of

1-4244-1180-7/07/$25.00 ©2007 IEEE

cameras used, the representation of captured data, types of algorithms, use of various models, and the application to specific body regions and whole body. Employed configurations typically range from using a single camera [12, 18] to multiple cameras [10, 14]. An even greater variety of algorithms has been proposed for estimating human motion including constraint propagation [28], optical flow [5, 35], stochastic propagation [13], search space decomposition based on cues [10], statistical models of background and foreground [34], silhouette contours [19], annealed particle filtering [9], silhouette based techniques [4, 6], shape-encoded particle propagation [25], simulated annealing [8] and fuzzy clustering process [23]. These algorithms typically derive features either directly in the single or multiple 2D image planes [13, 5] or, in the case of multiple cameras, at times utilize a 3D representation [10, 6] for estimating human body kinematics. The majority of approaches are model-based in which an a priori model with relevant anatomic and kinematic information is tracked or matched to 2D image planes or 3D representations. Different model types have been proposed including stick-figure [18], cylinders [12], super-quadrics [10], and CAD model [35]. Several surveys concerned with computer-vision approaches have been published in recent years, each classifying existing methods into different categories [24, 33].

While many existing computer vision approaches offer a great potential for markerless motion capture for biomechanical applications, these approaches have not been developed or tested for this applications. To date, qualitative tests and visual inspections are most frequently used for assessing approaches introduced in the field of computer vision and machine learning. Evaluating existing approaches within a framework focused on addressing biomechanical applications is critical. Approaches from the field of computer vision have previously been explored for biomechanical applications. These include markerless systems for the estimation of joint centers [15], tracking of lower limb segments [29], analysis of movement disabilities [19, 23], and estimation of working postures [30]. In particular, Persson [29] proposed a markerless method for tracking the human lower limb segments. Only movement in the sagittal plane was considered. Marzani et al. [23] extended this approach to a system consisting of three cameras. A 3D model based on a set of articulated 2D super-quadrics, each of them describing a part of the human body, was positioned by a fuzzy clustering process. These studies demonstrate the applicability of techniques in computer vision for automatic human movement analysis, but are lacking of a validation against marker-based data.

Previous work in the field of computer vision was an inspiration for our work on tracking an articulated model in visual hull sequences. Articulated models have been used for object tracking in video [12, 5, 36, 32] and in 3D data streams [20, 6]. Our soft-joint constraints approach allows small movement at the joint, which is penalized in least-squares terms. This extends previous approaches [5, 6] for tracking articulated models that enforced hard constraints on the kinematic structure (joints of the skeleton must be preserved).

3. Methods

The proposed approach consists of tracking a 3D subject-specific articulated model in 3D representations constructed from multiple camera views.

3.1. Data acquisition

Full body movement was captured using a marker-based and a markerless motion capture system simultaneously. The marker-based system consisted of an eight-Qualisys camera optoelectronic system monitoring 3D marker positions at 120 fps. The markerless motion capture system consisted of eight Basler VGA color cameras synchronously capturing images at 75 fps. Internal and external camera parameters and a common global frame of reference were obtained through offline calibration.

3.2. 3D representation

The subject was separated from the background in the image sequence of all cameras using an intensity and color threshold for the color cameras. The 3D representation was achieved through visual hull construction from multiple 2D camera views [22, 16, 7].

3.3. Articulated models

An articulated model is typically derived from a morphological description of the human body's anatomy plus a set of information regarding the kinematic chain and joint centers. The morphological information of the human body can be a general approximation (cylinders, super quadrics, etc.) or an estimation of the actual subject's outer surface. Ideally, an articulated model is subject-specific and created from a direct measurement of the subject's outer surface. The kinematic chain underneath an anatomic body can be manually set or estimated through either functional [4, 7] or anthropometric methods. The more complex the kinematic description of the body the more information can be obtained from the 3D representations. An optimal subject-specific articulated body can be created from a detailed full body laser scan with markers affixed to the subject's joints that were defined through manual palpation, but appears infeasible for practical scenarios.

A database of human shapes was recently proposed by Anguelov et al. [2]. Deformable models of human shapes were learned from 46 full body laser scans using principal component analysis and among other things non-rigid transformations D accounting for changes in body shape between different individuals were calculated. Our work builds on the work of Anguelov et al. [2] by processing a wide variety of evenly distributed human models from a template mesh based on modifications to the non-rigid transformations D. Anguelov et al. [2] proposed deforming triangle edges $v_{k,j} = x_{k,j} - x_{k,1}$, $j = 2,3$ (triangles p_k containing the points $(x_{k,1}, x_{k,2}, x_{k,3})$) of a template mesh using up to three transformations

$$R_k D_k Q_k v_{k,j}, j = 2, 3, \quad (1)$$

where R specifies a rigid pose transformation and Q a non-rigid transformation specifying changes in pose such as bulging of muscles. Components in the transformation matrix D represent very reasonable variation in weight and height, gender, abdominal fat and chest muscles, and bulkiness. We created a repository of articulated models through modifications of these components (Figure 1).

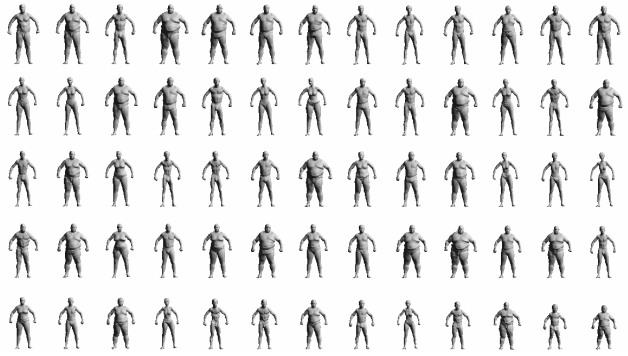

Figure 1: Selected articulated models from a repository of articulated models calculated from a database of human shapes.

The resulting meshes provide matching articulated bodies consisting of the same number of triangles (labeling is preserved to easily approximate joint areas). It is difficult to identify automatically an accurate kinematic chain of a human as a function of anthropometric parameters such as height and/or volume because anthropometric parameters vary widely for people of similar stature and morphology (Figure 2, [11]). Figure 2a shows the variation for the upper arm length for a human population. A similar variation exists in our repository of articulated models matching the general trend. Graphs for other body segments showed similar results. As a result, a kinematic chain only approximates a human subject and accurate tracking requires a solution with soft-joint constraints that allows small movements at the joints.

a) b)

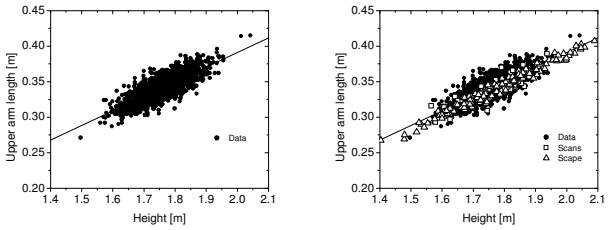

Figure 2: a) Correlation of upper arm length as a function of height (1774 men and 2208 women, adapted from [11]). b) Compared to our repository of articulated models.

In our approach, a 3D articulated body was tracked in the 3D representations using an articulated body from a repository of subject-specific articulated bodies that would match the subject closest based on a volume and height evaluation (Figure 1). The lack in detailed knowledge of the morphology and kinematic chain of the tracked subjects was adjusted by allowing larger inconsistencies at the joints. The articulated body consisted of 15 body segments (head, trunk, pelvis, and left and right arm, forearm, hand, thigh, shank and foot) and 14 joints connecting these segments.

3.4. Articulated ICP

Our articulated ICP algorithm is a generalization of the standard ICP algorithm [3, 31] to articulated models. The objective is to track an articulated model Y in a sequence of visual hulls. The articulated model Y is represented as a discrete sampling of points $x_1, \ldots, x_M$ on the surface, a set of rigid parts $p_1, \ldots, p_P$, and a set of joints Q connecting the segments. Each measurement (visual hull) is represented as a set of points $Z = z_1, \ldots, z_K$, which describes the appearance of the person at that time. For each frame of the sequence, an alignment T is computed, which brings the surfaces of Y and Z into correspondence, while respecting the model joints Q. The alignment T consists of a set of rigid transformations T_j, one for each rigid part p_j.

Similar to ICP, our algorithm iterates between two steps. In the first step, each point x_i on the model is associated to its nearest neighbor $z_{c[i]}$ among the sensor measurement points Z, where $l_{[i]}$ defines the mapping from the index of a surface point x_i to its rigid part index. In the second step, given a set of corresponding pairs $(x_i, z_{c[i]})$, a set of transformations T is computed, which brings them into alignment. It is assumed that an initial estimate of the transformation T is given (for example, by borrowing the solution for the previous frame in the sequence). A newly obtained set of transformations can then subsequently be used to bootstrap the process. The second step is defined

by an objective function of the transformation variables given as $F(T) = H(T) + G(T)$. The term $H(T)$ ensures that corresponding points (found in the first step) are aligned (Figure 3a), given as

$$H(r,t) = w_H \sum_{i=1}^{M} \left\| R(r_{l[i]})x_i + t_{l[i]} - z_i \right\|^2 \qquad (2)$$

where the transformation T_j of each rigid part p_j is parameterized by a 3x1 translation vector t_j and a 3x1 twist coordinates vector r_j (twists are standard representations of rotation [21], and $R(r_{l[i]})$ denotes the rotation matrix induced by the twist parameters $r_{l[i]}$. The term $G(T)$ ensures that joints are approximately preserved (Figure 3b), where each joint $q_{i,j}$ can be viewed as a point belonging to parts p_i and p_j simultaneously. The transformations T_i and T_j were forced to predict the joint consistently

$$G(r,t) = w_G \sum_{(i,j)\in Q} \left\| R(r_{l[i]})q_{i,j} + t_{l[i]} - R(r_{l[j]})q_{i,j} - t_{l[j]} \right\|^2 \qquad (3)$$

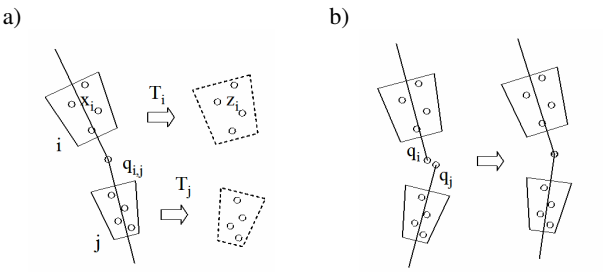

Figure 3: (a) Point-to-point associations used to define the energy $H(T)$. (b) Illustration of the joint mismatch penalty $G(T)$.

Linearizing the rotations around their current estimate in each iteration resulted in a standard least-squares function over the transformation parameters (r,t)

$$\arg\min_{r,t} \left\| A \cdot \begin{bmatrix} r \\ t \end{bmatrix} - b \right\|^2 \Rightarrow \begin{bmatrix} r \\ t \end{bmatrix} = \left(A^T \cdot A\right)^{-1} \cdot A^T b\,, \qquad (4)$$

where A is the Jacobian of the cost function calculated in present state and b is the configuration state derived from Equations 2 and 3. Decreasing the value of w_G allows greater movement at the joint, which potentially improves the matching of body segments to the sensor measurement. The center of the predicted joint locations (belonging to adjacent segments) provides an accurate approximation of the functional joint center.

The articulated model was roughly aligned to the first valid frame in the motion sequence based on a motion trajectory obtained from the center of volumes of the 3D representations and subsequently tracked automatically over the motion sequence.

4. Results

Figure 4 shows an articulated model from the repository of articulated models matched to a visual hull. The lack in detailed knowledge of the kinematic chain of the tracked subjects yielded inconsistencies at the joints, which were adjusted in a post-processing step.

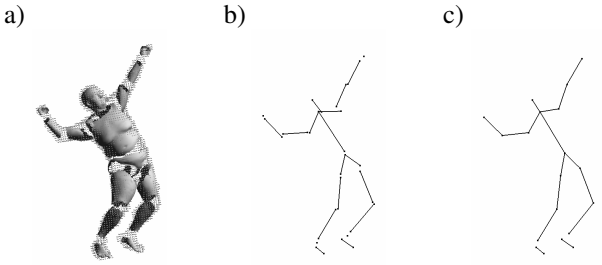

Figure 4: a) Articulated model matched to visual hull. b) Soft-joint constraints approach allows movement at the joints. c) Consistent kinematic chain.

The quality of visual hulls depends on numerous aspects including camera calibration, number of cameras, camera configuration, imager resolution and the accurate fore/background segmentation in the image sequences [27]. The accuracy of visual hulls also depends on the human subject's position and pose within the investigated viewing volume [27]. Simultaneous changes in position and pose result in decreasing the accuracy of visual hulls. Configurations with as few as four cameras provided accurate tracking results. Joint centers in the visual hull sequences were predicted with an average accuracy that matches the in-plane camera accuracy of magnitude of approximately 1 cm (Figure 5).

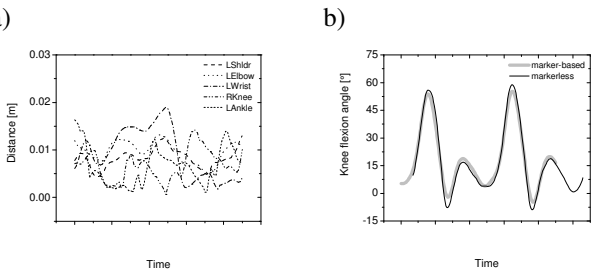

Figure 5: a) Euclidian distance of joint centers obtained from marker-based and markerless system. b) Knee flexion angle.

The joint angles (sagittal and frontal plane) for the knee calculated as angles between corresponding axes of neighboring segments were used as preliminary basis of comparison between the marker-based and markerless systems. The accuracy of sagittal and frontal plane knee

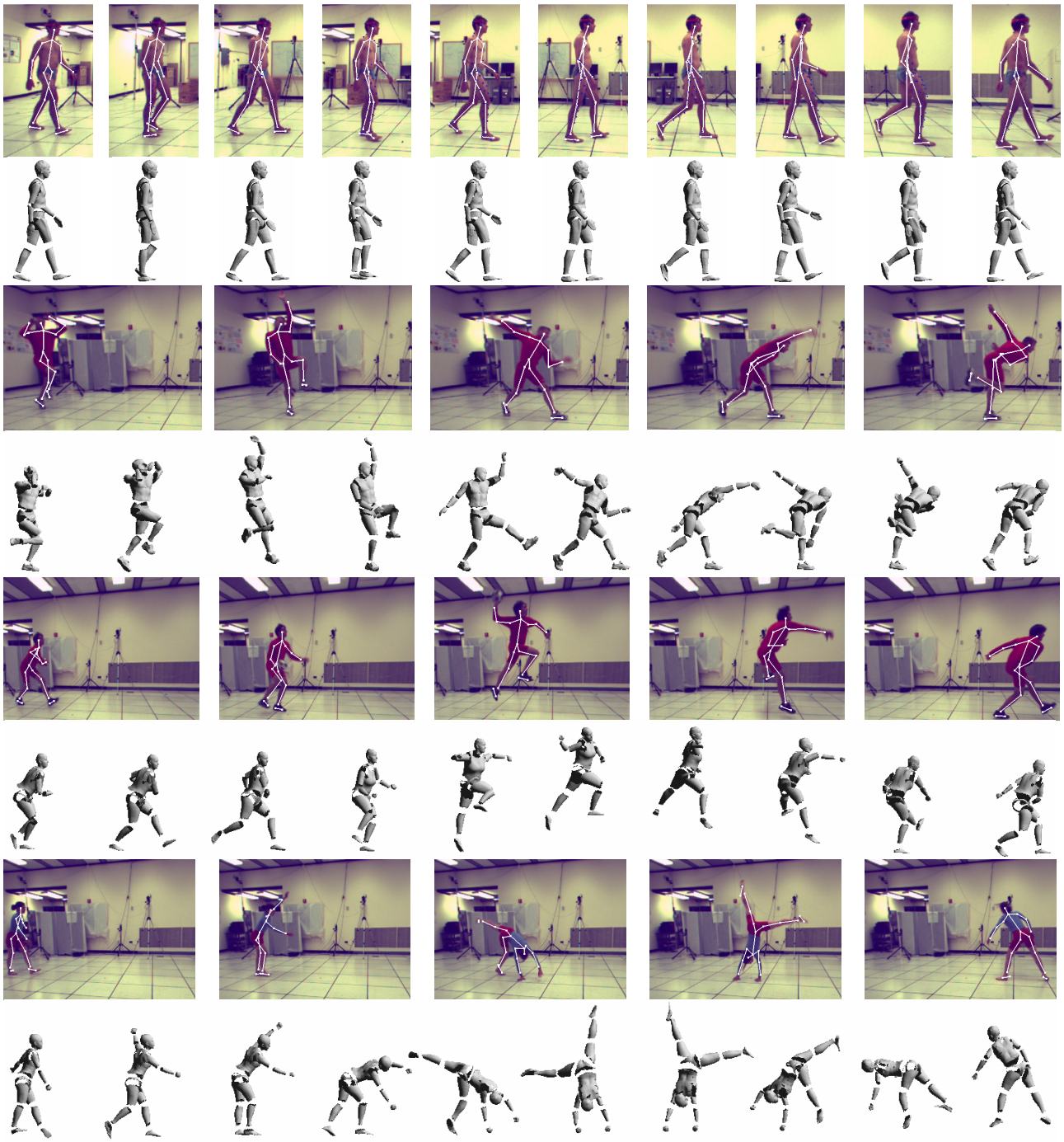

Figure 6: Selected frames of motion sequences with our markerless tracking results overlaid and the corresponding sequence of articulated models. Top to bottom: walking, cricket bowl, handball throw, and cart wheel.

joint angles calculated from experiments was 2.3 ± 1.0° and 1.6 ± 0.9°, respectively.

5. Discussion

The results presented here demonstrate the feasibility of accurately measuring 3D human body kinematics using a markerless motion capture system on the basis of visual hulls. The employed algorithm yields great potential for accurately tracking human body segments. The algorithm does not enforce hard constraints for tracking articulated models. The employed cost function consists of two terms, which ensure that corresponding points align and joint are approximately preserved. The emphasis on either term can be chosen globally and/or individually, and thus yields

more anatomically correct models. Moreover, the presented algorithm can be employed by either fitting the articulated model to the visual hull or the visual hull to the articulated model. Both scenarios will provide identical results in an ideal case. However, fitting data to the model is likely to be more robust in an experimental environment where visual hull only provide partial information due to calibration and/or segmentation errors.

Future work should focus on a more rigorous matching of articulated models to the visual hull sequence and a comparison of true bone embedded markers versus the markerless results.

Acknowledgements

Funding provided by NSF #03225715 and VA #ADR0001129.

References

[1] D. Anguelov, L. Mündermann, et al. An Iterative Closest Point Algorithm for Tracking Articulated Models in 3D Range Scans. SBC, Vail, CO, 2005.

[2] D. Anguelov, P. Srinivasan, et al. Scape: Shape completion and animation of people. TOG 24(3):408-416, 2005.

[3] P. Besl and N. McKay. A method for registration of 3D shapes. TPAMI 14(2):239-256, 1992.

[4] A. Bottino and A. Laurentini. A silhouette based technique for the reconstruction of human movement. CVIU 83:79-95, 2001.

[5] C. Bregler and J. Malik. Tracking people with twists and exponential maps. CVPR, San Juan, Puerto Rico, 1997.

[6] G. Cheung, S. Baker, et al. Shape-from-silhouette of articulated objects and its use for human body kinematics estimation and motion capture. CVPR, Madison, WI, 2003.

[7] G. Cheung, G., S. Baker, et al. Shape-From-Silhouette Across Time Part I & II. IJCV 62(3):221-247 & 63(3):225-245, 2005.

[8] S. Corazza, L. Mündermann, et al. A markerless motion capture system to study musculoskeletal biomechanics: visual hull and simulated annealing approach. Ann Biomed Eng 34(6):1019-1029, 2006

[9] J. Deutscher, A. Blake, et al. Articulated body motion capture by annealed particle filtering. CVPR, Hilton Head, SC, 2000.

[10] D. Gavrila and L. Davis. 3-D model-based tracking of humans in action: a multi-view approach. CVPR, San Francisco, CA, 1996.

[11] I. Haritaoglu and L. Davis. W4: real-time surveillance of people and their activities. TPAMI 22(8):809-830, 2000.

[12] D. Hogg. Model-based vision: A program to see a walking person. Image and Vision Computing 1(1):5-20, 1983.

[13] M. Isard and A. Blake. Visual tracking by stochastic propagation of conditional density. 4th European Conference on Computer Vision, Cambridge, UK, 1996.

[14] I. A. Kakadiaris and D. Metaxes. 3D human body model acquisition from multiple views. IJCV 30:191-218, 1998.

[15] H. Lanshammar, T. Persson, et al. Comparison between a marker-based and a marker-free method to estimate centre of rotation using video image analysis. Second World Congress of Biomechanics, 1994.

[16] A. Laurentini. The Visual Hull concept for silhouette base image understanding. TPAMI 16:150-162, 1994.

[17] A. Leardini, A., L. Chiari, et al. Human movement analysis using stereophotogrammetry Part 3: Soft tissue artifact assessment and compensation. Gait and Posture 21:221-225, 2005.

[18] H. J. Lee and Z. Chen. Determination of 3D human body posture from a single view. CVGIP 30:148-168, 1985.

[19] L. Legrand, F. Marzani, et al. A marker-free system for the analysis of movement disabilities. Medinfo 9(2):1066-1070, 1998.

[20] M. Lin. Tracking articulated objects in real-time range image sequences. ICCV 1:648-653, 1999

[21] Y. Ma, Y., S. Soatto, et al. An invitation to 3D vision, Springer Verlag, 2004.

[22] W. Martin and J. Aggarwal. Volumetric description of objects from multiple views. TPAMI 5(2):150-158, 1983.

[23] F. Marzani, E. Calais, et al. A 3-D marker-free system for the analysis of movement disabilities - an application to the legs. IEEE Trans Inf Technol Biomed 5(1):18-26, 2001.

[24] G. Moeslund and E. Granum. A survey of computer vision-based human motion capture. CVIU 81(3):231-268, 2001.

[25] H. Moon, R. Chellappa, et al. 3D object tracking using shape-encoded particle propagation. ICCV, Vancouver, BC, 2001.

[26] L. Mündermann, S. Corazza, et al. The evolution of methods for the capture of human movement leading to markerless motion capture for biomechancial applications. JNER 3(6), 2006.

[27] L. Mündermann, A. Mündermann, et al. Conditions that influence the accuracy of anthropometric parameter estimation for human body segments using shape-from-silhouette. SPIE Electronic Imaging 5665:268-277, 2005.

[28] J. O'Rourke and N. I. Badler. Model-based image analysis of human motion using constraint propagation. TPAMI 2:522-536, 1980.

[29] T. Persson. A marker-free method for tracking human lower limb segments based on model matching. Int J Biomed Comput 41(2):87-97, 1996.

[30] S. Pinzke and L. Kopp. Marker-less systems for tracking working postures - results from two experiments. Applied Ergonomics 32(5):461-471, 2001.

[31] S. Rusinkiewicz and M. Levoy. Efficient variants of the ICP algorithm. 3DIM, 2001.

[32] L. Sigal, M. Isard, et al. Attractive people: Assembling loose-limbed models using non-parametric belief propagation. NIPS, 2003.

[33] L. Wang, W. Hu, et al. Recent Developments in Human Motion Analysis. Pattern Recognition 36(3):585-601, 2003.

[34] C. R. Wren, A. Azarbayejani, et al. Pfinder: Real-time tracking of the human body. TPAMI 19(7):780-785, 1997.

[35] M. Yamamoto and K. Koshikawa. Human motion analysis based on a robot arm model. CVPR, Maui, HW, 1991.

[36] S. Yu, R. Gross, et al. Concurrent object recognition and segmentation with graph partitioning. NIPS, 2002.