# Markerless human motion capture through visual hull and articulated ICP

**Lars Mündermann**
Biomechanical Engineering
Stanford University
Stanford, CA 94305
*lmuender@stanford.edu*

**Stefano Corazza**
Biomechanical Engineering
Stanford University
Stanford, CA 93405
*stefanoc@stanford.edu*

**Thomas. P. Andriacchi**
Biomechanical Engineering
Stanford University
Stanford, CA 94305

Bone and Joint Center
Palo Alto VA
Palo Alto, CA 94304
*tandriac@stanford.edu*

Orthopedic Surgery
Stanford Medical Center
Stanford, CA 94305

## Abstract

A next critical advancement in human motion capture is the development of a non-invasive and marker-free system. In this study a marker-free approach based on an articulated iterative closest point (ICP) algorithm with soft joint constraints was evaluated using the HumanEva dataset. Our tracked results accurately overlay the visual data. As the ground truth marker placement and methodology for calculating joint centers is not sufficiently documented, accurate quantitative comparison of error is impossible. Nevertheless, average difference and standard deviation between corresponding joints are reported to give an "upper bound" to the error of the proposed algorithm. The standard deviation for most joints is around 1 to 2 cm.

## 1 Introduction

Human motion capture is a well established paradigm for the diagnosis of the patho-mechanics related to musculoskeletal diseases, the development and evaluation of rehabilitative treatments and preventive interventions for musculoskeletal diseases. 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-based markers indicate a worst-case root mean square artifact of 7° [1, 2].

A next critical advancement in human motion capture is the development of a non-invasive and marker-free system [3]. A technique for human body kinematics estimation 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 [3, 4]. Our current approach employs an articulated iterative closest point (ICP) algorithm with soft joint constraints [5]

for tracking human body segments in visual hull sequences (a standard 3D representation of dynamic sequences from multiple images) [3]. The soft joint constraints approach extends previous approaches [6, 7] for tracking articulated models that enforced hard constraints on the joints of the articulated body. Two adjacent body segments each predict the location of the common joint center. The deviation between the two predictions is penalized in least-squares terms.

The purpose of this study was to evaluate the performance of our markerless motion capture approach using the HumanEva dataset [8].

## 2 Methods

The subject was separated from the background in the image sequence of all cameras using an intensity threshold for the grayscale cameras, and an intensity and color threshold for the color cameras [9]. The 3D representation was achieved through visual hull construction from multiple 2D camera views [10-12].

A 3D articulated body was tracked in the 3D representations using an articulated body from a repository of subject-specific articulated bodies [13] 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.

a) b)

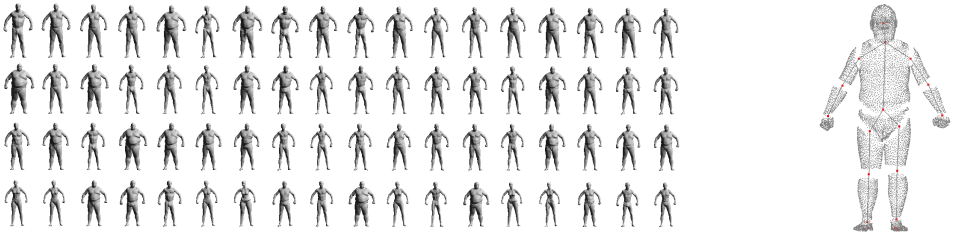

Figure 1: a) Selected bodies from repository of subject-specific articulated bodies. b) Articulated body consisting of 15 body segments and 14 joint for subject S1.

The subject's pose was roughly matched based on a motion trajectory to the first valid frame in the motion sequence and subsequently tracked automatically over the gait cycle. The motion trajectory was calculated as a trajectory of center of volumes obtained from the 3D representations for each frame throughout the captured motion.

## 3 Results

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 [14, 15]. Accurate background segmentation in the greyscale cameras is challenging (Figure 2). Difficulties in the accurate fore/background separation in the grayscale cameras BW1 and BW4 prevented the use of all cameras. As a result, visual hulls were created using only five cameras (BW2, BW3, C1, C2, and C3) resulting in a rather crude approximation (Figure 2).

a) b)

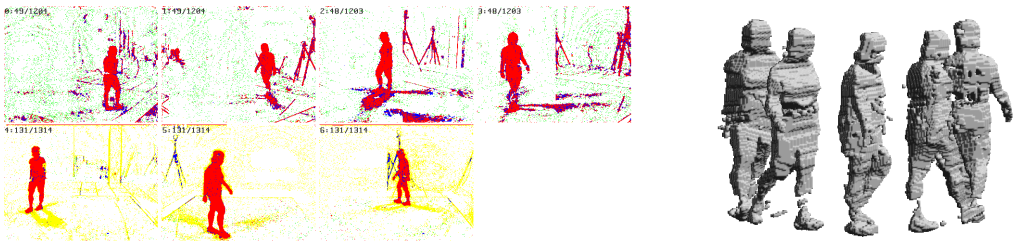

Figure 2: a) Selected foreground (red) in grayscale cameras (top row) and color cameras (bottom row). b) Visual hulls for selected frames (54, 104, 154, 204 and 254) of image sequence S1\Walking_1.

The articulated body was tracked successfully through individual video sequences. Our tracked results (red, Figures 3 and 4) accurately overlays the visual data.

a) Frame 72 b) Frame 224 c) Frame 860

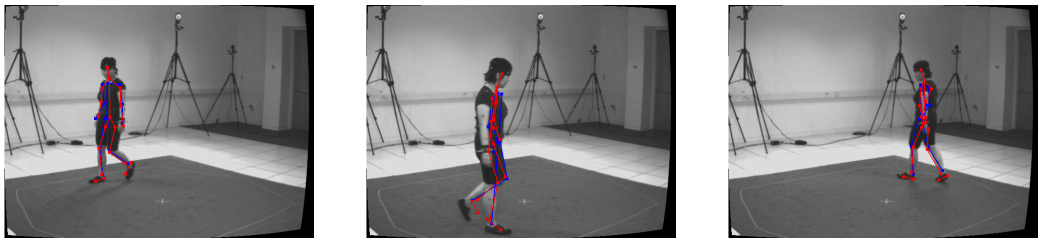

Figure 3: Selected frames of camera BW3 of trial S1\Walking_1 with overlaid marker-based data (blue) and our markerless tracking results (red).

a) Frame 254 b) Frame 370 c) Frame 458

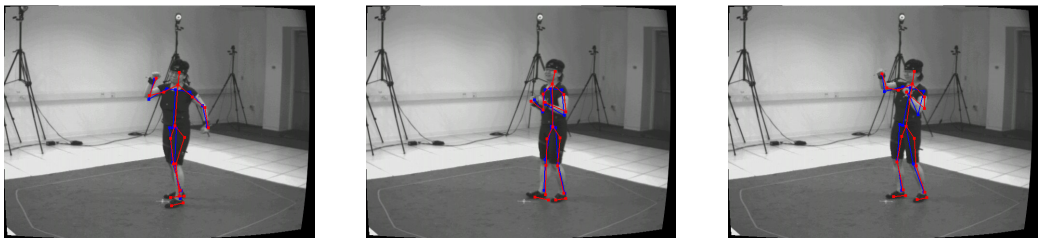

Figure 4: Selected frames of camera BW3 of trial S1\Boxing_1 with overlaid marker-based data (blue) and our markerless tracking results (red).

Comparison between the joints provided by the marker-based system and calculated by our approach yielded truthful overlay, but differences in defining joint centers in the kinematic chain in the selected articulated model and marker-based system cause an offset among corresponding joints. As the ground truth marker placement and methodology for calculating joint centers is not sufficiently documented, accurate quantitative comparison of error is impossible.

Nevertheless, average difference and standard deviation calculated from the Euclidian distances between corresponding joints are reported to give an upper bound to the error of the proposed algorithm (Figure 5, Table 1). However, the results do not reflect the true potential of the proposed algorithm due to differences in defining joint centers.

a) b)

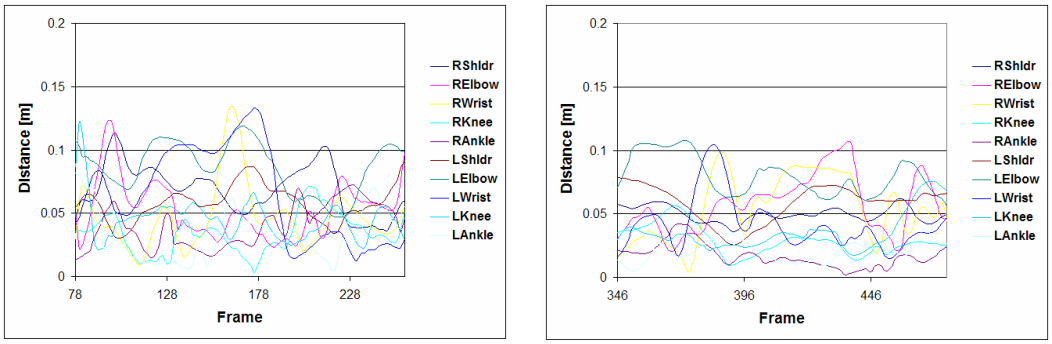


Figure 5: Euclidian distance between corresponding joints from the marker-based and markerless information. a) S1\Walking_1. b) S1\Boxing_1.

Table 1: Average difference (avg) and standard deviation (stdev) calculated from the Euclidian distances between corresponding joints of the marker-based and markerless information.

| | S1\Walking_1 | | S1\Boxing_1 | |
|---|---|---|---|---|
| | avg [m] | stdev [m] | avg [m] | stdev[m] |
| RShldr | 0.069 | 0.02 | 0.051 | 0.009 |
| RElbow | 0.053 | 0.023 | 0.062 | 0.025 |
| RWrist | 0.047 | 0.025 | 0.054 | 0.028 |
| RKnee | 0.034 | 0.016 | 0.027 | 0.011 |
| RAnkle | 0.04 | 0.017 | 0.015 | 0.009 |
| LShldr | 0.058 | 0.014 | 0.058 | 0.017 |
| LElbow | 0.087 | 0.02 | 0.082 | 0.022 |
| LWrist | 0.064 | 0.037 | 0.045 | 0.022 |
| LKnee | 0.046 | 0.017 | 0.035 | 0.016 |
| LAnkle | 0.033 | 0.02 | 0.025 | 0.009 |

## 4 Discussion

The results presented here demonstrate the feasibility of 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.

The presented approach has been successfully applied to simple walking and running sequences and more complex motion sequences such as a cricket bowl, handball throw and cart wheel. The obtained accuracy compared to marker-based systems [3].

Utilizing more cameras to enhance the quality of the visual hulls and a better knowledge of the morphology and kinematic chain of the tracked subjects allowing stronger consistencies at the joints would improve the accuracy.

**Acknowledgments**

Funding provided by NSF #03225715 and VA #ADR0001129.

**References**

1. Leardini, A., et al., *Human movement analysis using stereophotogrammetry Part 3: Soft tissue artifact assessment and compensation.* Gait and Posture, 2005. **21**: p. 221-225.
2. Cappozzo, A., et al., *Position and orientation in space of bones during movement: experimental artifacts.* Clinical Biomechanics, 1996. **11**: p. 90-100.
3. Mündermann, L., S. Corazza, and T.P. Andriacchi, *The evolution of methods for the capture of human movement leading to markerless motion capture for biomechancial applications.* Journal of Neuroengineering and Rehabilitation, 2006. **3**(6).
4. Moeslund, G. and E. Granum, *A survey of computer vision-based human motion capture.* Computer Vision and Image Understanding, 2001. **81**(3): p. 231-268.
5. Anguelov, D., L. Mündermann, and S. Corazza. *An Iterative Closest Point Algorithm for Tracking Articulated Models in 3D Range Scans*. in *Summer Bioengineering Conference*. 2005. Vail, CO.
6. Bregler, C. and J. Malik. *Tracking people with twists and exponential maps*. in *Computer Vision and Pattern Recognition*. 1997.
7. Cheung, G., S. Baker, and T. Kanade. *Shape-from-silhouette of articulated objects and its use for human body kinematics estimation and motion capture*. in *IEEE Conference on Computer Vision and Pattern Recognition*. 2003. Madison, WI: IEEE.
8. Sigal, L. and M.J. Black, *HumanEva: Synchronized Video and Motion Capture Dataset for Evaluation of Articulated Human Motion.* Technical Report CS-06-08, 2006.
9. Haritaoglu, I. and L. Davis, *W4: real-time surveillance of people and their activities.* IEEE Transactions on Pattern Analysis and Machine Intelligence, 2000. **22**(8): p. 809-830.
10. Martin, W. and J. Aggarwal, *Volumetric description of objects from multiple views.* IEEE Transactions on Pattern Analysis and Machine Intelligence, 1983. **5**(2): p. 150-158.
11. Laurentini, A., *The Visual Hull concept for silhouette base image understanding.* IEEE Transactions on Pattern Analysis and Machine Intelligence, 1994. **16**: p. 150-162.
12. Cheung, K., S. Baker, and T. Kanade, *Shape-From-Silhouette Across Time Part I: Theory and Algorithm.* International Journal of Computer Vision, 2005. **62**(3): p. 221-247.
13. Anguelov, D., et al., *Scape: Shape completion and animation of people.* ACM Transaction on Graphics, 2005. **24**(3).
14. Mündermann, L., et al., *Most favorable camera configuration for a shape-from-silhouette markerless motion capture system for biomechanical analysis.* SPIE-IS&T Electronic Imaging, 2005. **5665**: p. 278-287.
15. Mündermann, L., et al., *Conditions that influence the accuracy of anthropometric parameter estimation for human body segments using shape-from-silhouette.* SPIE-IS&T Electronic Imaging, 2005. **5665**: p. 268-277.