

# A Markerless Motion Capture System to Study Musculoskeletal Biomechanics: Visual Hull and Simulated Annealing Approach

S. CORAZZA,<sup>1</sup> L. MÜNDERMANN,<sup>1</sup> A. M. CHAUDHARI,<sup>1</sup> T. DEMATTIO,<sup>2</sup> C. COBELLI,<sup>2</sup>  
and T. P. ANDRIACCHI<sup>1,3,4</sup>

<sup>1</sup>Department of Mechanical Engineering, Stanford University, 496 Lomita Mall, Durand B. 201, Stanford, CA; <sup>2</sup>Department of Information Engineering, University of Padova, Padova, Italy; <sup>3</sup>Bone and Joint Center, Palo Alto VA, Palo Alto, CA; and <sup>4</sup>Department of Orthopedic Surgery, Stanford University Medical Center, Stanford, CA

(Received 1 July 2005; accepted 29 March 2006; published online: 5 May 2006)

**Abstract**—Human motion capture is frequently used to study musculoskeletal biomechanics and clinical problems, as well as to provide realistic animation for the entertainment industry. The most popular technique for human motion capture uses markers placed on the skin, despite some important drawbacks including the impediment to the motion by the presence of skin markers and relative movement between the skin where the markers are placed and the underlying bone. The latter makes it difficult to estimate the motion of the underlying bone, which is the variable of interest for biomechanical and clinical applications. A model-based markerless motion capture system is presented in this study, which does not require the placement of any markers on the subject's body. The described method is based on visual hull reconstruction and an *a priori* model of the subject. A custom version of adapted fast simulated annealing has been developed to match the model to the visual hull. The tracking capability and a quantitative validation of the method were evaluated in a virtual environment for a complete gait cycle. The obtained mean errors, for an entire gait cycle, for knee and hip flexion are respectively  $1.5^\circ (\pm 3.9^\circ)$  and  $2.0^\circ (\pm 3.0^\circ)$ , while for knee and hip adduction they are respectively  $2.0^\circ (\pm 2.3^\circ)$  and  $1.1^\circ (\pm 1.7^\circ)$ . Results for the ankle and shoulder joints are also presented. Experimental results captured in a gait laboratory with a real subject are also shown to demonstrate the effectiveness and potential of the presented method in a clinical environment.

**Keywords**—Human motion capture, Musculoskeletal biomechanics, Visual hull, Simulated annealing.

## INTRODUCTION

Motion capture techniques are used over a very broad field of applications, ranging from digital animation for entertainment to biomechanical analysis for sport and clinical applications. Sport and clinical applications require excellent accuracy and robustness. Two other major requirements for clinical applications are to minimize the

time for patient preparation and the inter-observer variability. At present, using reflective markers on the skin is the most common technique.<sup>5,12,13</sup> Despite their precision and popularity, marker based methods have several limitations: (i) markers attached to the subject can influence the subject's movement, (ii) a controlled environment is required to acquire high-quality data, (iii) the time required for marker placement can be excessive, and (iv) the markers on the skin can move relative to the underlying bone, leading to what is commonly called skin artifact.<sup>4,14,32</sup> Several recent review articles have summarized the common shortfalls of skin based marker techniques.<sup>6,8,21</sup>

Markerless motion capture offers an attractive solution to the problems associated with marker based methods. However, the use of markerless methods to capture human movement for biomechanical or clinical applications has been limited by the complexity of acquiring accurate three-dimensional kinematics using a markerless approach. The general problem of estimating the free motion of the human body or more generally of an object without markers, from multiple camera views, is underconstrained without the spatial and temporal correspondence that tracked markers guarantee.

Model based approaches provide methods to address some of the complexities associated with a markerless approach. An *a priori* model of the subject, for example, can be used to strongly reduce the total number of degrees of freedom of the problem. Another option is to increase the number of cameras so that more measured data is available to solve for a given number of degrees of freedom. Thus the robustness of a markerless approach can be increased by increasing the number of cameras and by limiting the search space of possible body configurations to anatomically appropriate ones. This last strategy can be pursued by using a human model to identify the motion of the subject.

Several model based methods have been proposed in the past, modeling the human body or parts of it with rigid<sup>1,9,15,18,29,30</sup> or non rigid segments.<sup>19</sup> However, these

Address correspondence to S. Corazza, Department of Mechanical Engineering, Stanford University, 496 Lomita Mall, Durand B. 201, Stanford, CA 94305-4038. Electronic mail: stefano.corazza@stanford.edu

approaches have problems with accurate identification of three dimensional kinematics of the segments or use a limited number of body segments.<sup>7</sup> For what concerns the mathematical formulation of the model joints, exponential maps<sup>3</sup> is able to provide several advantages to previous approaches by simplifying the estimation of model pose and leading to robust identification of body kinematics.

Another important consideration in choosing an approach for markerless motion capture is the formulation of the cost function used to match the representation of the subject (2D silhouette, 3D visual hull, 2D features, etc.) to the model. While approaches utilizing only 2D information have been used,<sup>3</sup> most biomechanical applications require a 3D model. In the approach that we pursued we built the subject's 3D representation using shape-from-silhouette technique. Our group has developed several methods in the past to reconstruct the outer surface of the body.<sup>10,26</sup> In this study the 3D representation has been obtained using the algorithm described in.<sup>26</sup>

The other important consideration in designing an approach involves the choice of an optimization algorithm that will successfully minimize the cost function to allow the calculation of the subject's kinematics. This optimization is difficult because the cost function has many local minima and the search space has very high dimensionality, but it can be accomplished through the simulated annealing matching algorithm running on an exponential maps geometry formulation. Simulated annealing is a statistical computational method based on Boltzmann Sampling and the Metropolis Monte Carlo method.<sup>24</sup> In standard Monte Carlo simulation, the state of a system is randomly changed by sampling the search space. All such changes are accepted, whether or not the new state results in a reduction of the cost function. Thus the system may be in a high-cost state much of the time, and the simulation has to run for a very long time to properly sample all low-cost regions of the search space. The modification made by Metropolis and co-workers is that there should be a probabilistic acceptance of a Monte Carlo step. This modified algorithm is known as Metropolis Monte Carlo simulation, and it further evolved into the method known as simulated annealing. Simulated annealing has the capability to identify states defined by many degrees of freedom while consistently reducing the risk of getting trapped into local minima. Thus the integration of a method that includes a subject's model, visual hulls and simulated annealing as described above offers a potential framework for a markerless motion caption system.

The purpose of this study was to describe the development and to validate such a markerless motion capture system. The goals of this system are to have the advantage of not requiring the placement of markers or the design of an acquisition protocol, and to be potentially usable in any possible environment where a large number of synchronized calibrated cameras are available. From these multiple views the geometric representation of the human body is

reconstructed based on a visual hull concept, which was first described in.<sup>20</sup> Simulated annealing is used to match an *a priori* 3D model to the visual hull, and the subsequently calculated kinematics of the matched model are validated against ground truth in a virtual environment.

## METHODS

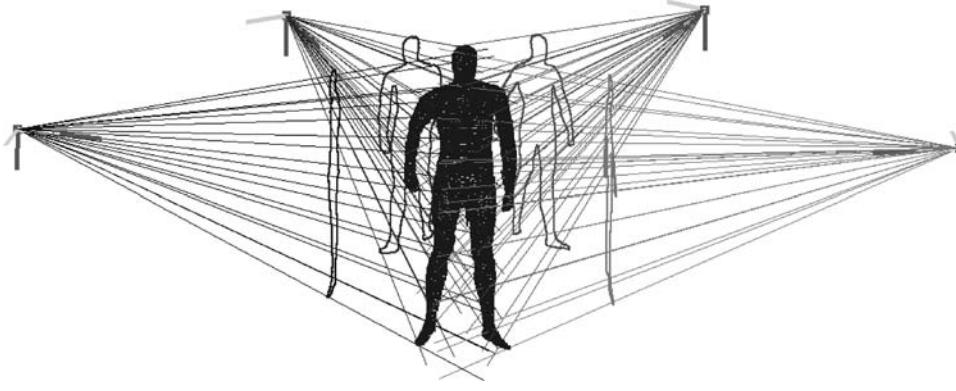
The developed method, described in the following sections, is used to track the motion of a human subject in both a virtual and a real environment. 16 cameras in the virtual environment and eight cameras in the experimental setup (same resolution 640 by 480 pixels) were used to obtain the subject's 3D representation. Tracking the 3D representation using the described matching algorithm leads to the extraction of the subject's kinematics. The following subsections describe the several steps required to achieve this goal.

### *Reconstruction of the Subject's Visual Hull*

The visual hull of an object, first extensively described in,<sup>19</sup> can be defined as the locally convex (over) approximation of the volume occupied by an object. The 3D representation of the motion of the subject across the motion capture volume consists of one visual hull for each instant in time captured by the camera system. The visual hull construction process, diagrammed in Fig. 1, consisted of the projection of the subject's silhouette from each of the camera planes back to the 3D volume. The intersection of the resulting cones in 3D space generated the subject's visual hull. The 2D silhouettes in the camera planes were obtained by foreground/background separation for every captured frame. In general, previously-described shape-from-silhouette methods reconstruct the subject's visual hull by dividing the 3D space into cubic voxels whose size is inversely proportional with the desired resolution.<sup>11, 18, 23, 28, 33</sup> The method used in the present work belongs to the same family of algorithms and its detailed description can be found in.<sup>26, 28, 33</sup> An extensive study on the influence of camera number, resolution and placement on reconstructed visual hull quality can be found in.<sup>26</sup> Its applicability in the typical *in vivo* experimental setup of a gait analysis laboratory was demonstrated.

### *Exponential Maps Formulation*

The adopted exponential maps formulation<sup>27</sup> guarantees a simple linear representation of the motion with the posture uniquely defined, avoiding the nonlinearities and singularities common to the Euler angles formulation. The exponential map for a twist describes the relative motion of two coordinate frames in space, and in the same way, every rigid transformation can be represented by a combination of twists. The exponential map formulation allows the representation of multiple twists by simply multiplying the exponentials of the transformation matrices together (1).



**FIGURE 1. Visual hull reconstruction concept. The silhouettes of the subject from different camera planes are back projected in space. Their intersection generates the visual hull, a locally convex over-approximation of the volume occupied by the subject's body.**

For example, for the lower limb, the rigid movement of the thigh with respect to the pelvis can be written as a single screw transformation (also known as a finite helical axis transformation) or as a combination of three twists. Going from the pelvis (defined in Eq. (1) as segment ( $a$ )) all the way down to the foot (defined in Eq. (1) as segment ( $b$ )), the final position of a point on the foot can be expressed as a function of the articular parameters of hip, knee and ankle joints through the multiplication of twists, as shown in Eq. (1). The pelvic coordinate system is in this case the reference,

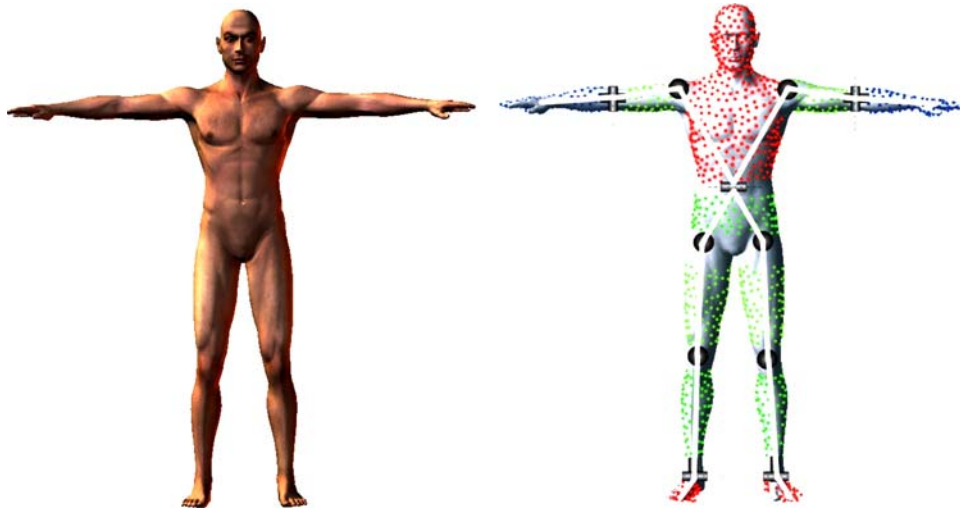
$$g_{ab}(\vartheta) = g_{ab}(0) \prod_{i=1}^n e^{\hat{\xi}_i \vartheta_i} \quad (1)$$

$\vartheta = (\vartheta_1, \dots, \vartheta_8)$  is the state vector (scalar angles) for a kinematic chain with eight degrees of freedom and  $n$  is the number of degrees of freedom, in this case equal to

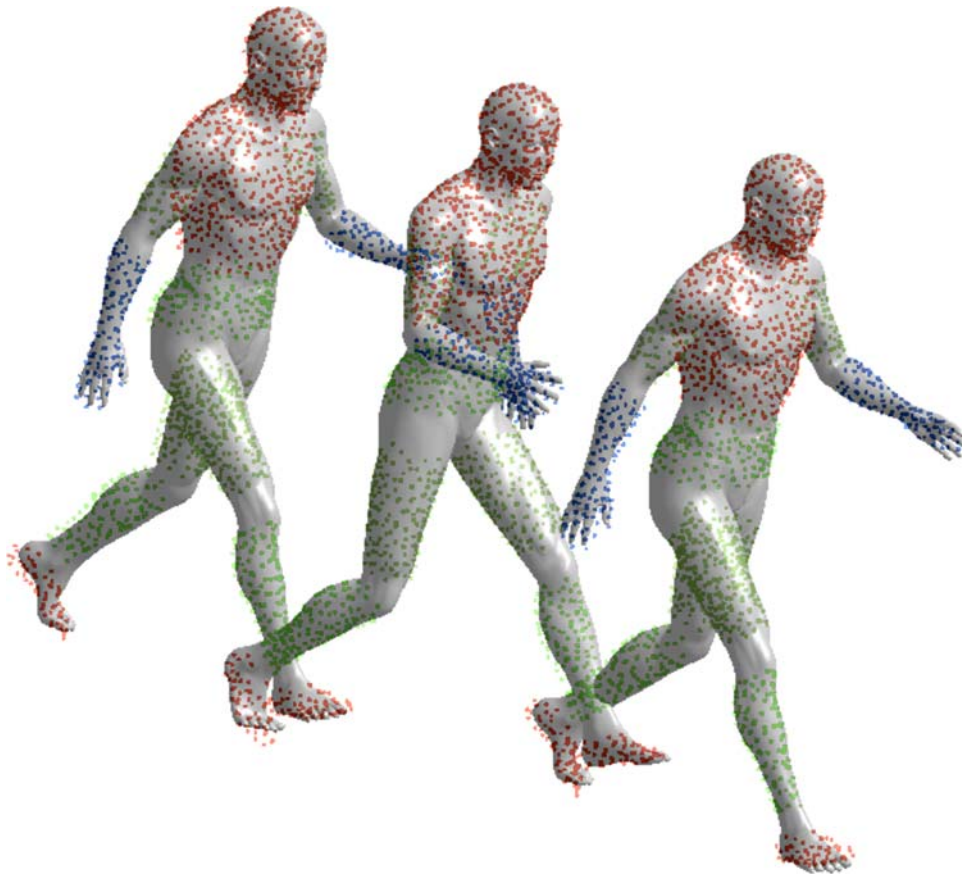
8. ( $\vartheta_1, \vartheta_2, \vartheta_3$ ) represents the three rotational degrees of freedom of the thigh with respect to the pelvis, ( $\vartheta_4, \vartheta_5, \vartheta_6$ ) represent the three rotational degrees of freedom of the shank with respect to the thigh, and ( $\vartheta_7, \vartheta_8$ ) represent the two rotational degrees of freedom of the foot with respect to the shank.  $g_{ab}(0)$  represents the transformation matrix from  $a$  to  $b$  coordinate frame in the initial configuration, i.e. with all state variables equal to zero. The general twist matrixes  $\hat{\xi}_i$  are defined for each joint relative to the parent segment, following the formulation described in.<sup>27</sup>

The final transformation from one body configuration to another one is given by a matrix  $T$  defined as follows:

$$K'_{p \times 4N} = K_{p \times 4N} T = K_{p \times 4N} \begin{bmatrix} g_1 & 0 & \dots & 0 \\ 0 & g_2 & \dots & 0 \\ \dots & \dots & g_i & \dots \\ 0 & 0 & \dots & g_N \end{bmatrix}_{4N \times 4N} \quad (2)$$



**FIGURE 2. (a) Poser model in the reference pose, (b) 33 DOF full body model. Points in highly deformable regions have been removed, as most clearly seen at the hips.**



**FIGURE 3.** Results of the matching algorithm (*colored points*) applied to the virtual environment sequence superimposed over the virtual character (*gray surface*).

where  $g_i$  is the  $4 \times 4$  matrix representing the generic rigid transformation for segment  $i$  with respect to the parent segment.  $T$  is a  $4N \times 4N$  square matrix in which  $N$  is the total number of segments in the human body model.  $K$  matrices contain the  $p$  visual hull points in homogeneous coordinates.

#### *Full Body Model*

The model contains morphological information (surface with 1600 points) and kinematics information about how the model can move. The morphological information came from a reference pose (Fig. 2a). Then the model was segmented into the different parts corresponding to the 12 main anatomical segments shown in Fig. (2b): pelvis, thighs, shanks, feet, arms, forearms, and combined torso and head. For the real application with human subjects, the morphological information was obtained from a laser scan of the subject, providing an accurate description of the body's outer surface that was then manually segmented.

The kinematic model (depicted by the lines connecting joints in Fig. 2b) includes the full body and has 33 degrees of freedom (DOF). Joints were modeled as ball-

socket joints or as simple hinge joints. In particular, for the lower limbs, the hip and knee were modeled as spherical joints with three degrees of freedom in rotation (flexion-extension, adduction-abduction, internal-external rotation), while the ankle was modeled as a double hinge joint having two rotational DOF (plantar-dorsi flexion, in-eversion). For the upper body the movement between the torso and the pelvis was modeled as a simple hinge-joint with one rotational DOF (flexion at the 5th lumbar), the shoulder was modeled as a spherical joint (flexion-extension, internal-external rotation, adduction-abduction) and the elbow was modeled as a double hinge joint having two rotational DOF (flexion-extension and pronation-supination). The remaining six degrees of freedom described the rigid body translation and rotation in space of the root segment, the pelvis. The geometrical formulation of the model is open in the sense that any joint model can be modified independently without the need to readjust the others. More complex joint models may take into account both rotational and translational behavior using the same mathematical structure, by using the appropriate formulation of a particular joint that allows the translation along the particular twist axis. The completed model was created by rigidly joining the





**FIGURE 4.** Results of the matching algorithm applied to the virtual environment sequence. Points visual hull (top, in blue) and matched model (bottom) for a gait cycle, in the sagittal plane.

morphological representations of each segment (each a set of points describing a surface) to the corresponding rigid segments of the underlying kinematic model. Motion was constrained to anatomically consistent ranges.

Surface points close to the joints in the model were removed to minimize the influence of tissue deformation that occurs around the joints during movement.

#### *Matching Process by Simulated Annealing*

The matching process consisted of the minimization of a cost function in a continuous domain describing the quality of matching between the model and the visual hull cloud of points (made of about 2500 points). This matching was done for each time frame in order to identify the whole motion of the subject. Since all degrees of freedom were matched simultaneously the search space was 33-dimensional (number of DOF in the kinematic model). Gradient-based methods were not appropriate to solve such a high-dimensional problem due to the large number of local minima in which the algorithm could get trapped. Instead, we adopted a

stochastic approach called simulated annealing that is an extension of the original Metropolis Monte Carlo method. This class of methods has been refined during last decades and has the capability of climbing up local minima until the desired matching accuracy is achieved.<sup>16,17,31,34</sup>

#### *Simulated Annealing*

The implemented simulated annealing method uses the acceptance function (3) proposed by Metropolis,<sup>24</sup> which is a function of the parameter  $T$  and of the value of the cost function  $f$ . The parameter  $T$ , commonly called temperature due to the analogy of the optimization process with the chemical process of annealing, is a function that decreases as the iteration number increases.

$$A(x, y, T) = \min \left\{ 1, e^{\frac{f_y - f_x}{T}} \right\} \quad (3)$$

Moving from current state  $x_i$  to next state  $x_{i+1}$ , the step is accepted or not depending on (4) where  $p$  is sampled from a uniform distribution  $[0, 1]$  and the value  $k_{i+1}$  is a state

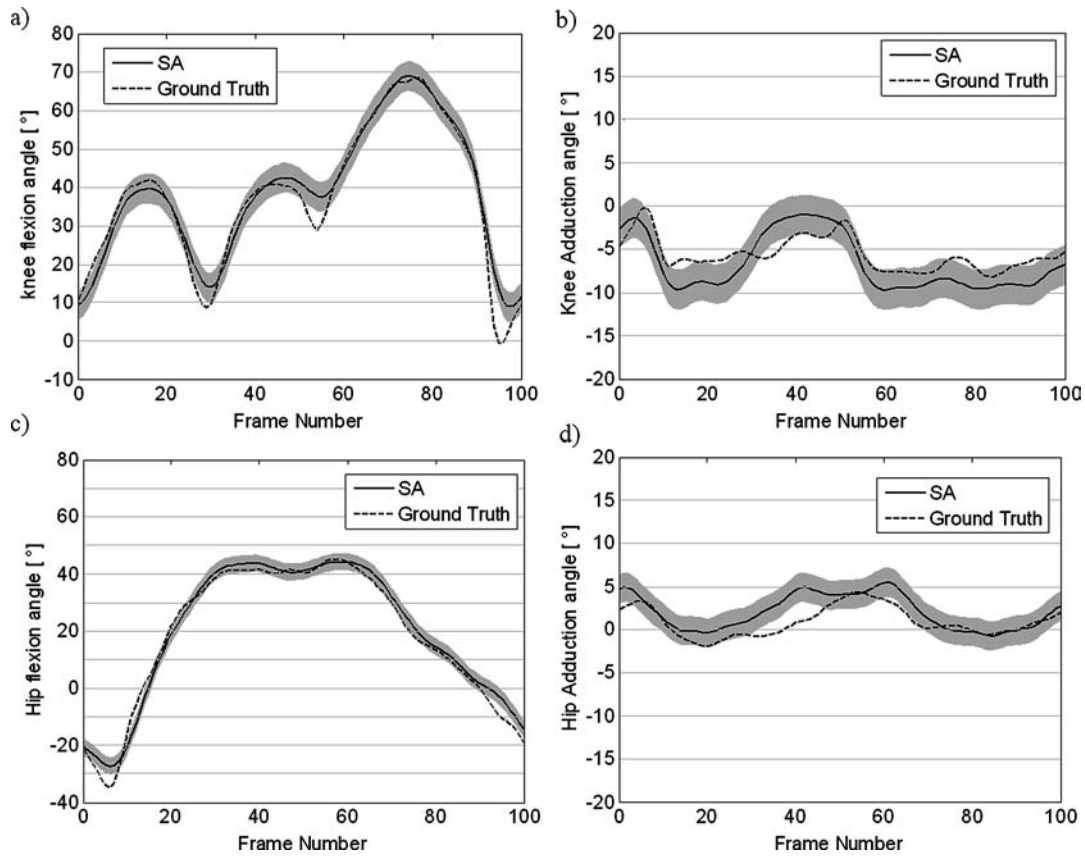


FIGURE 5. Comparison between ground truth provided by the virtual environment model and the results from the matching algorithm (gray shaded area indicates  $\pm$  one standard deviation) for (a) knee flexion, (b) knee adduction, (c) hip flexion, and (d) knee adduction.

sampled from a chosen distribution (see next paragraph).

$$x_{i+1} = \begin{cases} y_{i+1} = x_i + k_{i+1} & \text{if } p \leq A(x_i, y_{i+1}, T_i) \\ x_i & \text{otherwise} \end{cases} \quad (4)$$

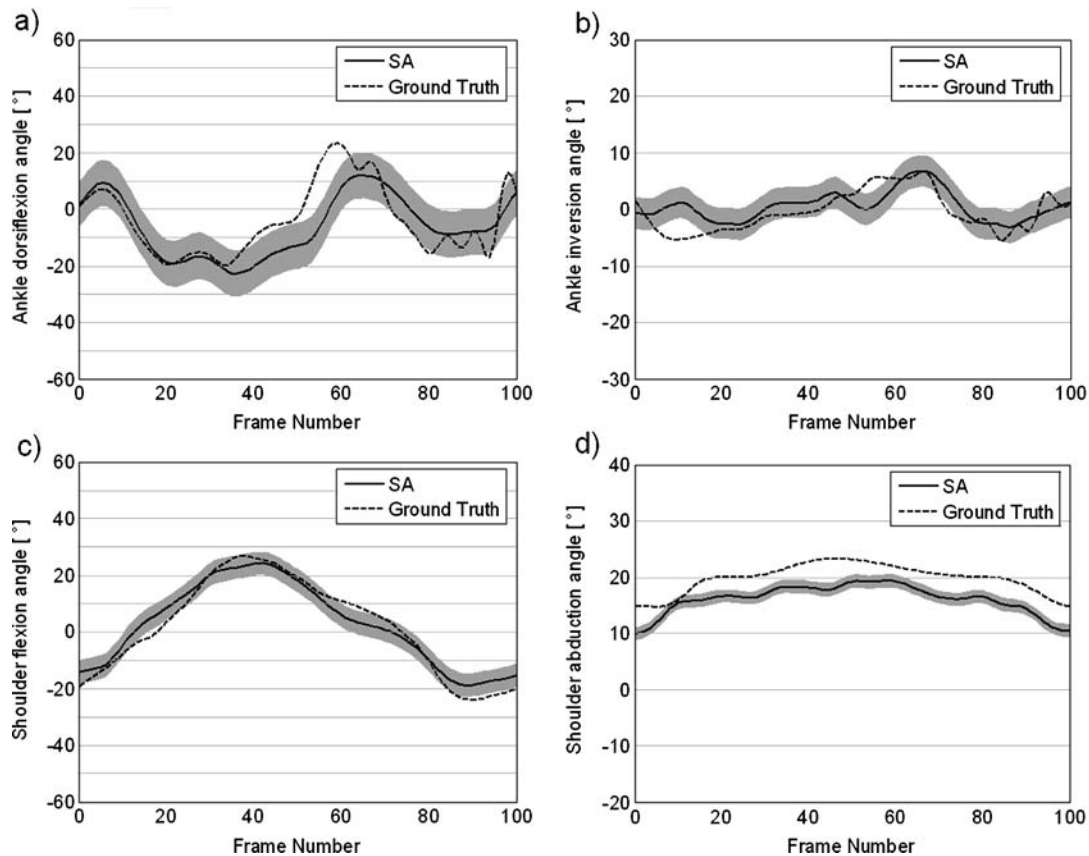
Since the parameter  $T$  plays an important role in the acceptance function, several authors have proposed different formulations for its decreasing function (cooling schedule) and the corresponding sampling distribution for  $k_{i+1}$ , in order to improve the performances of the algorithm that normally has a high computational cost.<sup>16,17,24,31,34</sup> The formulation used in this work is described in<sup>34</sup> which samples  $k_{i+1}$  from a Cauchy distribution. Sampling in this way allows the algorithm to visit each region with positive Lebesgue measure infinitely often when a cooling schedule proportional to  $T_0/i$  is adopted, where  $T_0$  is a large enough constant and  $i$  is the number of iterations. To assure better capabilities for climbing up local minima (as demonstrated in simulated trials<sup>10</sup>), in this work the parameter  $T$  is not decreased linearly with respect to the number of iterations but depends also on the value of the cost function. An extensive and complete description of the general simulated annealing method can be found in.<sup>22</sup>

### The Cost Function

An appropriate choice of the cost function is one of the core requirements for successful and robust matching. Two clouds of points need to be matched, one that is articulated using a kinematic model and the other coming from visual hull reconstruction. The latter has in general a non constant number of points through the sequence frames, and there is no correspondence between points in different time frames even though they are equally spaced in a 3D voxel structure. The chosen cost function for this work (5) was a variation on the Hausdorff distance and has been shown to be very robust even if computationally demanding.

$$\text{COST}(A, B) = \sum_{\forall a \in A} \min_{\forall b \in B} \left\{ \|a - b\| \right\} \quad (5)$$

The cost function used here differs from the original formulation of the Hausdorff distance since it sums every single contribution, instead of taking just the maximum between the minimal distances between pairs of points. This modification increases robustness to possible outliers. As is the case for the original Hausdorff distance, this cost function is not commutative. Intuitively, one could state that a low value of  $\text{COST}(A, B)$  guarantees that all the points of set  $A$



**FIGURE 6.** Comparison between ground truth provided by the virtual environment model and the results from the matching algorithm (gray shaded area indicates  $\pm$  one standard deviation) for (a) ankle dorsiflexion, (b) ankle inversion, (c) shoulder flexion, and (d) shoulder abduction.

are not very far from their closest point of set  $B$ . However, it does not guarantee that all points of set  $B$  are not very far from their closest point of set  $A$ . In the first frame of the sequence, the visual hull points are set  $A$ , while the model points are set  $B$ , since the two sets may not be close to each other (visual hull-to-model formulation). For subsequent frames the next visual hull frame is always very close to the previous matched model state, so the cost function is changed to the model-to-visual hull formulation, which guarantees better accuracy because it is less sensitive to phantom volumes in the visual hull. Phantom volumes are defined as a large local deviation from the real subject's outer body surface resulting from the use of too few cameras. In our case phantom volumes generate points of set  $B$  (visual hull) far from their closest point of set  $A$  (model) that are neglected since the cost function—based on  $COST(A, B)$ —only accounts for the distance between points of set  $A$  from their closest point of set  $B$  (model-to-visual hull).

#### *Motion Data: Virtual Environment*

In order to provide data with a real ground truth, a virtual character was animated with known kinematics

using *Poser*<sup>®</sup> software (by *Curious Labs*, CA—USA). In the virtual sequence a male subject walks along a straight line mimicking a gait analysis sequence. Since the animation software uses Euler angles formulation, the internal-external rotations of each joint were set to zero to avoid cross talk between rotations along different axes. Sixteen virtual cameras were uniformly distributed in a most favorable hemispherical configuration<sup>26</sup> around the virtual character. Images from each camera were taken at every frame of the gait sequence. Silhouettes were extracted from each camera image and then processed to create the visual hulls that feed the matching algorithm presented in the previous sections.

#### *Motion Data: Experimental*

To demonstrate the effectiveness and potential of the method for biomechanical applications, a running sequence of a human subject was captured using 8 color video cameras with a resolution of 640 by 480 pixels and a frame rate of 75 frames/s. A running sequence is more challenging than gait analysis for the tracking algorithm since it involves higher velocities and accelerations of

the anatomical segments. The acquisition was done in a standard gait analysis laboratory environment, i.e. without altering the background or lightning conditions. The sequence was processed with the same algorithms described in this section. The subject's model was created using a 3D laser scan (Whole Body 3D scanner Model WBX by Cyberware—USA, accuracy within 1mm and about 15 seconds scanning time). The 11 joint centers of a 33 degree of freedom model were manually identified on the model obtained from the laser scanner.

The described method was validated in a virtual environment and qualitatively tested in experimental conditions. Using the virtual environment permitted the evaluation of the accuracy of extracting human body kinematics while excluding errors due to experimental artifacts (e.g. due to camera calibration errors, errors in background subtraction, etc.), thus obtaining the true potential of the method with the given camera setup (16 cameras,  $640 \times 480$  pixel resolution). A Kalman filter was used to smooth results and improve the quality of derivatives.

## RESULTS

The motion obtained in the virtual environment for the model (colored points) and the original character compared favorably (Fig. 3). In Fig. 4 the visual hulls and the matched model in the sagittal plane are shown as point clouds. The joint angles for the walking sequence are known and are compared with the ones obtained from the matching algorithm. Motions at the hip, knee, ankle and shoulder show good agreement between the virtual character kinematics and the matched model results (Figs. 5 and 6). Moreover, the algorithm does not drift, as shown by the fact that the errors do not increase with frame number.

The errors for the hip, knee, ankle and shoulder joints through the entire gait sequence are reported in Table 1. Good results are obtained in terms of mean absolute errors for flexion and adduction of hip, knee and shoulder. Errors are slightly bigger for the ankle joint, mainly due to the poor ratio between camera resolution and dimensions of the foot. The high degree of axial symmetry of the thighs and shanks makes it difficult for the algorithm to track internal-external rotation, leading to lower accuracy. This symmetry results in mean errors of  $3.9^\circ (\pm 4.1^\circ)$  and  $2.7^\circ (\pm 4.7^\circ)$  for hip and knee internal-external rotations, respectively.

Unlike most other tracking algorithms, the presented method does not require accurate initialization of the model to match the first frame. A rough rigid body positioning (as shown in Fig. 7, left) of the model in a reference frame is enough to have a consistent matching of the first frame of the sequence (Fig. 7, right). The computational time for solving for the entire sequence in this non-optimized version is on the order of few hours, since no specific hardware or optimized software has been used.

**Table 1. Summary of the validation results for joint angles at the hip, knee, ankle and shoulder.**

	Mean error ( $^\circ$ )	Standard deviation ( $^\circ$ )	RMS error ( $^\circ$ )
Hip flexion/ext	2.0	3.0	3.6
Hip adduction/abd	1.1	1.7	2.0
Knee flexion/ext	1.5	3.9	4.2
Knee adduction/abd	2.0	2.3	3.1
Ankle plantar/dorsifl	3.5	8.2	9.0
Ankle inversion/ev	4.7	2.8	5.9
Shoulder flexion/ext	1.2	4.2	4.4
Shoulder adduction/abd	3.8	1.2	4.0

The experimental results relative to a running sequence of a male subject are presented in Fig. 8. The sequence was processed with the algorithms described in the methods section. The effectiveness of the tracking results on the visual hull are shown in Fig. 8 where the point clouds representing the different anatomic segments consistently overlay the visual hull of the subject (shown in gray).

## DISCUSSION

The proposed method has been quantitatively validated for several joints. An effective tracking capability has been shown even for smaller body segments like the feet, which are normally neglected by other approaches. The results with this data also demonstrate the robustness of the approach, since the performance of the matching process did not deteriorate with frame number, a common problem for most feature-tracking based approaches.<sup>1,2</sup> Unlike in those approaches, a bad initial guess will only increase the computational time necessary to obtain the desired matching because in each frame the model is being matched to the absolute position of the visual hull rather than to the change in the video images from the previous frame to the current one.

In the kinematic model 33 degrees of freedom have been modeled, including 3 rotational degrees of freedom for the hip, knee and shoulder together with 2 degrees of freedom for the ankle (dorsi-plantarflexion, in-eversion). This set of degrees of freedom is appropriate for most biomechanical studies of the lower limb. In fact, having for example 3 DOF at the knee is important in order to investigate the secondary rotations which are a crucial point in understanding injury and disease mechanisms. One critical aspect of developing a successful algorithm for biomechanical analysis of many different activities is the measurement of internal-external rotation of the hip and the knee, which is quite noisy due to the almost cylindrical symmetry of the thigh and shank. This aspect represents the main limitation of the presented method that remains to be addressed in the future. Nevertheless, an increase in camera resolution alone would provide more accurate 3D reconstruction of the



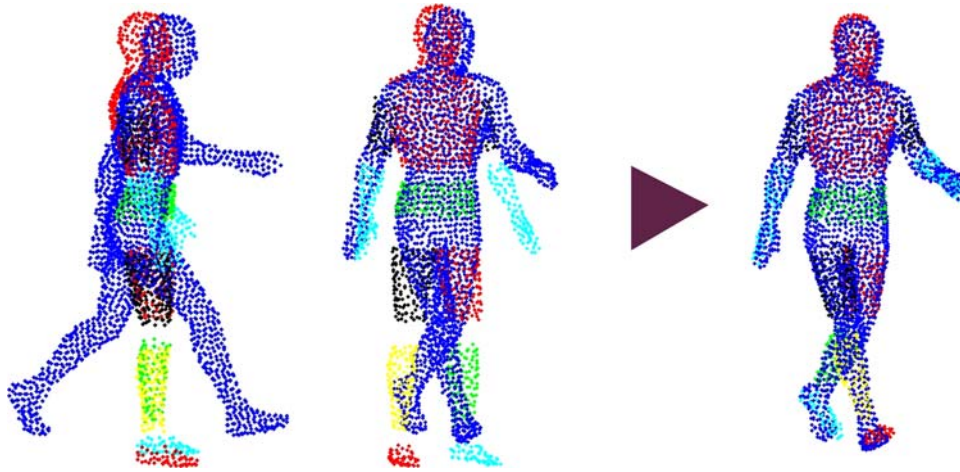


FIGURE 7. Matching of the first frame. The visual hull point cloud is shown in blue, while the different segments of the model are shown in other colors. The algorithm does not require a good initialization of the model in order to achieve the first matching (*right*).

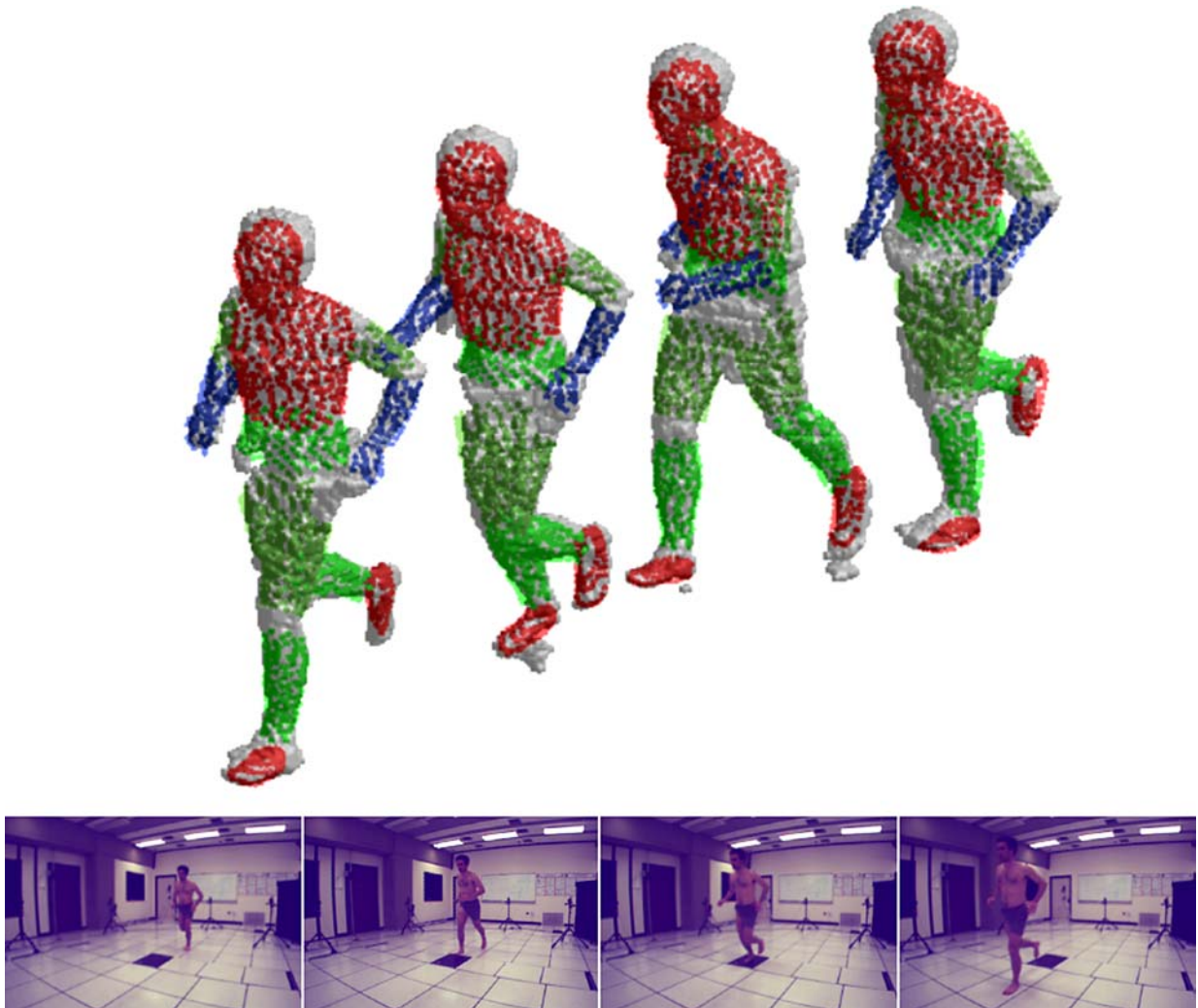


FIGURE 8. Result of the matching algorithm (*color points*) applied to an experimental data sequence (*gray surface*). On the bottom the corresponding video images of the running sequence from one camera are shown.

subject.<sup>25</sup> This increase would improve the overall tracking accuracy and reduce the problem of internal-external rotation of the hip and knee because the existing axial asymmetries in the thigh and shank would be better represented. This is one of the major challenges that will need to be addressed in the future since assessing true motions in transverse plane can be very relevant from a biomechanical point of view. The described algorithm, being based on the entire shape instead of a small number of single points, shows good performance even with camera resolutions of an order of magnitude lower than current stereophotogrammetric systems for marker-based motion analysis. This approach also offers a great potential for the reduction of skin artifact error. Instead of relying on just a few markers it is based on the tracking of few hundred points per segment, naturally averaging the skin artifact phenomenon across the segment during the matching process.

Markerless motion capture also guarantees a great reduction of the amount of time for subject preparation compared to marker based methods since no time for marker placement is required. Moreover inter-operator variability is eliminated since no trained operator is needed to accurately place markers. On the other hand, the processing time is longer since the creation of the model is required. Future work must automate the setting up of the model and address how foreground/background separation and camera calibration affect the accuracy of the results, which have not been directly addressed in the virtual environment validation. However, the experimental data presented already show good qualitative results, demonstrating the effectiveness and potential of the method for use in a clinical environment such as a gait laboratory. Moreover, tracking a running sequence, which is in general more challenging than walking due to increased velocities of the subject, demonstrated the robustness of the method. Overall, in the authors' opinion, the system can provide sufficient accuracy for biomechanical research and has great potential for future improvements, both in the creation of the model and in the matching algorithm. It is our intention to provide in the future an experimental comparison/validation against currently available techniques such as marker based methods.

Two further extremely valuable advantages of the presented algorithm are: (i) apart from a rough rigid body positioning, it does not need to be initialized at all, being able to go from a reference pose to the first frame pose consistently, as shown in Fig. 7; (ii) it directly provides joints centers and segment volume information during motion that can be used for a more accurate calculation of the subject's kinetics.

#### ACKNOWLEDGMENTS

Funding provided by NSF#03225715 and VA#ADR0001129

#### REFERENCES

- <sup>1</sup>Balan, A. O., L. Sigal, and M. J. Black. A quantitative evaluation of video-based 3D person tracking. *Proc. IEEE VS-PETS* 349–356, 2005.
- <sup>2</sup>Bottino, A., and A. Laurentini. A Silhouette based technique for the reconstruction of human movement. *Comput. Vis. Image Understand.* 83:79–95, 2001.
- <sup>3</sup>Bregler, C., and J. Malik. Tracking people with twists and exponential maps. *Proc. IEEE CVPR* 8–15, 1998.
- <sup>4</sup>Cappozzo, A., F. Catani, A. Leardini, M. G. Benedetti, and U. Della Croce. Position and orientation in space of bones during movement: experimental artifacts. *Clin. Biomech.* 11:90–100, 1996.
- <sup>5</sup>Cappozzo, A., F. Catani, U. Della Croce, and A. Leardini. Position and orientation in space of bones during movement: anatomical frame definition and orientation. *Clin. Biomech.* 10:171–178, 1995.
- <sup>6</sup>Cappozzo, A., U. Della Croce, A. Leardini, and L. Chiari. Human movement analysis using stereophotogrammetry Part 1: theoretical background. *Gait Post.* 21:186–196, 2004.
- <sup>7</sup>Cheung, G. K. M., S. Baker, and T. Kanade. Shape-from-silhouette across Time Part II: Applications to human modeling and markerless motion tracking. *Int. J. Comp. Vis.* 63(3):225–245, 2005.
- <sup>8</sup>Chiari, L., U. Della Croce, A. Leardini, and A. Cappozzo. Human movement analysis using stereophotogrammetry Part 2: Instrumental errors. *Gait Post.* 21:197–211, 2004.
- <sup>9</sup>Concalves, L., E. D. Bernardo, E. Ursella, and P. Perona. Monocular tracking of the human arm in 3rd Proceedings of the ICCV'95, pp. 764–770, 1995.
- <sup>10</sup>Corazza, S., and C. Cobelli. An accurate model-based approach for markerless motion capture. Proceedings of the Medicon, Italy, 2004.
- <sup>11</sup>Corazza, S., E. Alexander, A. Chaudhari, C. Cobelli, and T. Andriacchi. Surface from silhouette reconstruction for markerless motion capture. In Proceedings of the 7th Symposium Comp. Methods in Biomech., Madrid Spain, 2004.
- <sup>12</sup>Davis, R. B., III, S. Ounpuu, D. Tyburski, and J. R. Gage. A gait analysis technique data collection and reduction. *Human Mov. Sci.* 4:575–587, 1991.
- <sup>13</sup>Frigo, C., A. Pedotti, L. C. Deming, D. C. Kerrigan, and M. Rabuffetti. Functionally oriented and clinically feasible quantitative gait analysis method. *Med. Biol. Eng. Comp.* 36(2):179–185, 1998.
- <sup>14</sup>Fuller, J., L. J. Liu, M. C. Murphy, and R. W. Mann. A comparison of lower-extremity skeletal kinematics measured using skin- and pin-mounted markers. *Human Mov. Sci.* 16:219–242, 1997.
- <sup>15</sup>Gavrila, D. M., and L. S. Davis. Towards 3-D model-based tracking and recognition of human movement: A multi-view approach. In Proceedings of the International Workshop on Automatic Face and Gesture Recognition, Zurich, 1995.
- <sup>16</sup>Ingber, L. Simulated annealing: practice versus theory. *Math. Comp. Model.* 18(11):29–57, 1993.
- <sup>17</sup>Ingber, L. Very fast simulated re-annealing. *J. Math. Comp. Model.* 12:967–973, 1989.
- <sup>18</sup>Ju, S. X., M. J. Black, and Y. Yacoob. Cardboard people: a parameterized model of articulated motion. In Proceedings of the 2nd International Conference on Automatic face- And Gesture Recognition, Vermont USA. pp. 38–44, 1996.
- <sup>19</sup>Kakadiaris, I. A., and D. Metaxas. Model-based estimation of 3D human motion with occlusion based on active multi-viewpoint selection. *Proc. IEEE CVPR* 81–87, 1996.

- <sup>20</sup>Laurentini, A. The visual hull concept for silhouette based image understanding. *IEEE PAMI* 16(2):150–162, 1994.
- <sup>21</sup>Leardini, A., L. Chiari, U. Della Croce, and A. Cappozzo. Human movement analysis using stereophotogrammetry Part 3: Soft tissue artifact assessment and compensation. *Gait&Posture* 21:212–225, 2004.
- <sup>22</sup>Locatelli, M. Simulated annealing algorithms for continuous global optimization: convergence conditions. *J. Optim. Theory Appl.* 104(1):121–133, 2000.
- <sup>23</sup>Matusik, W., C. Buehler, R. Raskar, S. Gortler, and L. McMillan. Image-based visual hulls. *Proceedings of the ACM SIGGRAPH*. pp. 369–374, 2000.
- <sup>24</sup>Metropolis, N., A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller. Equation of state calculations by fast computing machines. *J. Chem. Phys.* 21:1087–1092, 1953.
- <sup>25</sup>Mündermann, L., A. Mündermann, A. Chaudhari, and T. P. Andriacchi. Conditions that influence the accuracy of anthropometric parameter estimation for human body segments using shape-from-silhouette. *SPIE-IS&T Electron. Imag.* 5665:268–277, 2005.
- <sup>26</sup>Mündermann, L., S. Corazza, A. Chaudhari, E. J. Alexander, and T. P. Andriacchi. Most favourable camera configuration for a shape-from-silhouette markerless motion capture system for biomechanical analysis. *IS&T/SPIE Electron. Imag.* 5665:278–287, 2005.
- <sup>27</sup>Murray, R. M., Z. Li, and S. S. Sastry. *A Mathematical Introduction to Robotic Manipulation*, Boca Raton, FL USA: CRC Press, 1994.
- <sup>28</sup>Potmesil, M. Generating octree models of 3D objects from their silhouettes in a sequence of images. *CVGIP* 40:1–29, 1987.
- <sup>29</sup>Regh, J. M., and T. Kanade. Model-based tracking of self-occluding articulated objects. *Proc. IEEE CVPR* 612–617, 1995.
- <sup>30</sup>Rohr, K. Incremental recognition of pedestrians from image sequences. *Proc. IEEE CVPR* 8–13, 1993.
- <sup>31</sup>Salhi, S., and N. M. Queen. A hybrid algorithm for identifying global and local minima when optimizing functions with many minima. *Eu. J. Oper. Res.* 155:51–67, 2002.
- <sup>32</sup>Sati, M., J. A. De Guise, S. Larouche, and G. Drouin. Quantitative assessment of skin marker movement at the knee. *The Knee* 3:121–138, 1996.
- <sup>33</sup>Szeliski, R. Rapid octree construction from image sequences. *CVGIP Image Understand.* 58(1):23–32, 1993.
- <sup>34</sup>Szu, H., and R. Hartley. Fast simulated annealing. *Phys. Lett. A* 122:157–162, 1987.