3D Skeleton-Based Body Pose Recovery

Clément Ménier, Edmond Boyer, Bruno Raffin

3D Skeleton-Based Body Pose Recovery

Clement Menier Edmond Boyer Bruno Raffin

GRAVIR–INRIA Rhône-Alpes
655, Avenue de l'Europe, 38334 Saint Ismier, France
E-mail: firstname.lastname@inrialpes.fr

Abstract

This paper presents an approach to recover body motions from multiple views using a 3D skeletal model. It takes, as input, foreground silhouette sequences from multiple viewpoints, and computes, for each frame, the skeleton pose which best fit the body pose. Skeletal models encode mostly motion information and allows therefore to separate motion estimation from shape estimation for which solutions exist; And focusing on motion parameters significantly reduces the dependancy on specific body shapes, yielding thus more flexible solutions for body motion capture. However, a problem generally faced with skeletal models is to find adequate measurements with which to fit the model. In this paper, we propose to use the medial axis of the body shape to this purpose. Such medial axis can be estimated from the visual hull, a shape approximation which is easily obtained from the silhouette information. Experiments show that this approach is robust to several perturbations in the model or in the input data, and also allows fast body motions or, equivalently, important motions between consecutive frames.

1. Introduction

An increasing number of virtual reality applications rely on marker-less interactions, for instance telepresence applications [16], or virtual object manipulation applications. This is, in most part, due to the fact that multi-view 3D modeling in real time becomes feasible, as demonstrated in recent works [8, 3]. However, models produced by such real-time methods are not necessarily rich enough to allow for complex interactions. In fact, information such as body part positions and velocities is often required by interaction applications. This *motion information* is related to, but different from, shape information for which efficient recovery solutions already exist. Our objective in this paper is therefore to focus on motion recovery, and in this way to provide a flexible and robust solution for body tracking from multiple views.

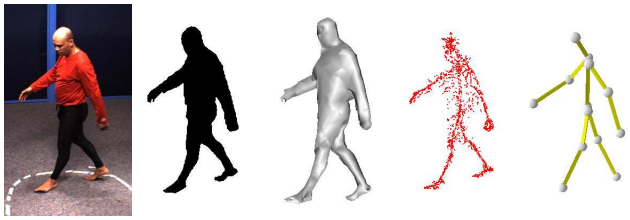


Figure 1. The tracking pipeline: (a) Color images ; (b) Silhouettes ; (c) Visual hulls ; (d) Medial axis points (d) ; (e) Skeleton pose.

Most marker-less motion tracking methods in computer vision fall into three categories. First, learning-based methods [1, 15] which rely on prior probabilities for human poses, and assume therefore limited motions. Second, model-free methods [9] which do not use any *a priori* knowledge, and recover articulated structures automatically. However, the articulated structure is likely to change in time, when encountering a new articulation for instance, hence making identification or tracking difficult. Third, model-based approaches which fit and track a known model using image information. In this paper, we aim at limiting as much as possible the required *a priori* knowledge, while keeping the robustness of the method reasonable for most interaction applications. Hence, our approach belongs to the third category.

Among model-based methods, a large class of approaches use an *a priori* surfacic or volumetric representation of the human body, which combines both shape and motion information. The corresponding models range from fine mesh models [6, 17, 4] to coarser models based on generalized cylinders [21, 12, 10], ellipsoids [8, 20] or other geometric primitives [11, 13, 14]. In order to avoid complex estimations of both shapes and motions as in [7], most approaches in this class assume known body dimen-

sions. However, this strongly limits flexibility and becomes intractable with numerous interaction systems where unknown persons are supposed to interact. A more efficient solution is to find a model which reduces shape information. To this purpose, a skeletal model can be used. This model does not include any volumetric information. Hence, it has fewer dependencies on body dimensions. In addition, limbs lengths tend to follow biological natural laws, whereas human shapes vary a lot among population.

Recovering motion using skeletal models has not been widely investigated. Theobalt *et al.* [23] propose an approach where a skeletal structure is fitted with the help of hand/feet/head tracking and voxel-based visual hull computation. However, volumetric dimensions are still required for the arms' and legs' limbs. Luck *et al.* [19] also propose a method where skeletal arms are fitted to a voxel-based visual hull of the upper body. The method still requires knowledge of the body radius, and suffers from inadequate captured volumetric data. Brostow *et al.* [5] have proposed a model-free method based on the extraction of a skeletal structure from the user's shape. Our approach relies on this idea of using a skeletal structure but differs in the method to extract it and in the use of an articulated model.

In this paper, we propose to use a skeletal model and hence, to focus on body motion parameters in the model parameters. In this way, we allow for adaptability to body sizes without sacrifying robustness or time complexity with respect to the aforementioned approaches. A difficulty in this context is to find a relevant data space in which to fit the skeletal model. Our main contribution lies in the combination of the skeletal model with specific input data in the form of 3D medial axis points. These points are obtained by computing the medial axis of the visual hull shape associated with the body silhouettes in the images. Figure 1 depicts the different steps of the method. All these steps can be, in the short term, achieved in real time, which makes the approach a good candidate for real time interaction applications.

§ 2 describes our skeletal articulated model and § 3 the associated measured data. § 4 presents the fitting and tracking process. § 5 reports on results obtained for real sequences and discusses on real time performance issues before concluding in § 6.

2. Skeletal Articulated Model

In this section, we describe the *a priori* articulated model representing a body pose. A great variety of models have been proposed in the literature. They rely on a kinematic chain adjoined with a shape model of the person (ellipsoids, quadrics, generalysed cylinders, *etc.*). Those models are thus specific to a particular user. We propose instead to use a 1D articulated model, therefore not including any volumetric information on the user.

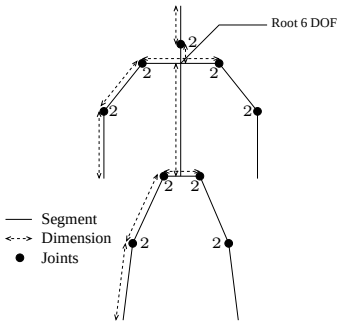


Figure 2. The skeletal articulated model.

This skeletal articulated model consists in a kinematic chain of segments. As interactive applications are usually only interested in the principle joints (elbows, shoulders, knees, legs and head), we limit our model to a set of 12 segments with those 9 joints (see figure 2). This leads to 24 degrees of freedom: 2 per joints and 6 for the root position and orientation. Note that other models, with higher fidelity to the human anatomy, could also be used if required by more demanding applications (e.g. graphics animations). For joints having 2 degrees of freedom, we chose a representation based on Euler angles. To avoid the classical discontinuity problems encountered with Eulerian parametrizations, we set the axis of rotation (where singularities occur) in the most unlikely direction (due to natural joint constraints for example). This proved to be sufficient in most of our experimentations. Other parametrizations, such as quaternions, would not necessarly give better results since they represent full 3D rotations (3 degrees of freedom).

3. Observed Skeleton Data

Another important element of the tracking process is the data which is considered as the measurement for the body pose, and to which the model is fitted. A great variety of data has been proposed in the literature for that purpose.

[14, 17, 6] use 2D cues such as silhouettes or contours. The body model is projected onto available image planes, and the fitting is achieved in the 2D image spaces. This has 2 major drawbacks: first, image features only affect the corresponding visible parts of the body model which must first be identified; second, skeletons are not invariant by projection, i.e. the 3D skeleton of a shape does not project onto the 2D skeletons of the projected shape, and thus fitting the

projection of a 3D skeleton to 2D skeletal data, such as 2D medial axis, would not make sense.

Other aproaches have proposed to directly use 3D cues. Most of them consider 3D data resulting from multi-view modeling methods such as Shape-From-Silhouette [4, 19] or stereo [11]. Such shape information is particularly well adapted when fitting shape models such as ellipsoids [8]. However it is not adapted to our approach since skeletal and shape information are of different nature and fitting our model to shape data would necessarly lead to inconsistencies.

More recently, Brostow *et al.* [5] have proposed to use 3D skeletal information for motion analysis. They retrieve motion information directly from an extracted 1D skeletal structure. Their approach being model free, a great care is taken to obtain a very precise skeleton, leading to a very slow extraction (several minutes per frame). It is therefore not adapted for interactive systems, which is our main objective. We propose to use a less robust but faster skeleton extraction technic. The lack of precision in the skeleton extraction is compensated by the *a priori* knowledge (human articulated model).

In our approach, we assume that silhouettes, extracted from calibrated cameras with different viewpoints, are available. These silhouettes are obtained through standard background subtraction methods. From these silhouettes, we first compute their 3D equivalent, i.e., the visual hull [18]. To this purpose, we use an exact method [3] which computes a polyhedron in space. This shape exactly projects onto the silhouettes in the images and thus preserves all the silhouette information. It is then processed in order to extract its internal structure, namely a skeleton. This step, called skeletonization, has received considerable attention from the computational geometry community. Several definitions can be considered for skeletons but the most successful is certainly the medial axis [22]. The medial axis is defined as the locus of centers of closed balls that are maximal with respect to inclusion. In the case of a discrete surface, the process leading to a discrete approximation of the medial axis is sometimes called the Medial Axis Transform. An important drawback of the discrete medial axis comes from its sensitivity to noise (see figure 3(b)). However some works have tackled this issue and proposed algorithms that take into account input shape noise. Attali *et al.* [2] have proposed such an algorithm. The idea is first to compute a discrete medial axis and second to prune it in order to eliminate outliers. The algorithm proceeds then as follows:

1. Voronoi centers are computed from the mesh vertices. Note that we only consider centers lying inside the mesh (see figure 3(b)).
2. For each center C we retrieve its corresponding Delaunay tetrahedron (P_1, P_2, P_3, P_4) and compute:
   - its radius $\rho(C) = d(C, P_1)$
   - its bisector angle $\theta(C) = \max_{i \neq j}(\widehat{P_i C P_j})$.
3. Outliers are eliminated based on a minimal radius and bisector angle threshold.

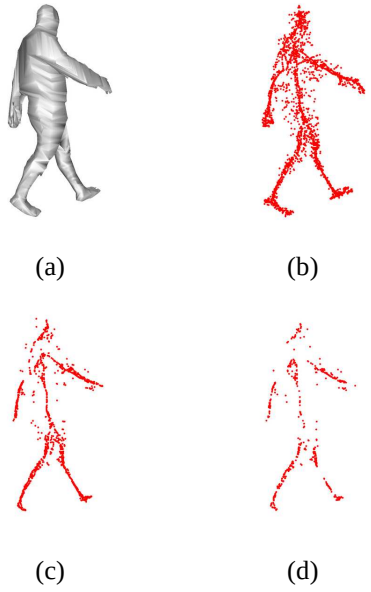
This results in a set of 3D points $\{X_0 \cdots X_n\}$ that we call the *Skeleton Data* (see figures 3(c) and 3(d)). Chosing a radius and bisector angle threshold consists in finding a tradeoff between the skeleton quality and the number of resulting points. Indeed the higher the thresholds are, the better the skeleton is but the fewer points are selected (see figure 3(d)). In practice we set the radius threshold at 4 cm and the bisector angle threshold around 160° (see figure 3(c)). It should be noticed here that the 3D medial axis is not a curve, as in 2D, but a surface. In practice, this has little impact on our approach for 2 reasons. First, the width of this surface, in the human case, is usually less or at most comparable (in the case of the torso) to the measurement noise. Second, the skeletal structure lies at the middle of the medial axis surface, therefore minimizing distances to the extracted medial axis points. Note that other skeletonization methods may be used, such as Brostow's method, as our fitting method is not specific to the medial axis but to the expense of the interactivity of the system.

4. Model Tracking

We have defined, in the previous sections, our skeletal articulated model and the observed skeleton data. In this section, we first define the generative model which explains the observations in function of the articulated model. We then present how this generative model is used in a fitting process which computes the maximum a posteriori estimate (MAP). Finally we discuss the important issue of pose tracking over sequences.

4.1. The generative model

In order to retrieve the pose of the user at a given time t, we must define the relationship between the *a priori* articulated model and the observed data. A first solution would be to characterize the similarity between the skeleton dataset of points $\{X_0, \cdots, X_n\}$ and a skeletal model S based on the distance of each point to its closest articulated segment $s \in S$ as in the following joint probability:



(a) (b) (c) (d)

Figure 3. (a) Exact visual hull obtained. (b) Internal voronoi centers yielding a noisy skeleton. (c) Skeletonization after pruning with $r > 4$ cm and $\theta > 160°$: most outliers are removed. (d) Skeletonization after pruning with $r > 5$ cm and $\theta > 170°$.

$$P(\{X_i\}\, S) = P(S) \times \prod_{i=0}^{n} P(X_i|S)\,, \qquad (1)$$

where: $P(X_i|S) = \mathcal{N}(d(X_i, S), \sigma^2)$ and $d(X_i, S) = \min_{s \in S} d(X_i, s)$, with $d()$ representing the Euclidean distance.

However, maximizing the corresponding posterior distribution $P(S|\{X_i\})$ leads to difficulties. Indeed, the attachment of a point to a segment is subject to change during the fitting process, generating inconsistencies and gradient discontinuities. To solve this issue, we introduce hidden variables a_i, one for each point, representing the segment attached to point X_i. The joint probability of the observed data and the pose becomes then:

$$P(\{X_i\}\,\{a_i\}\, S) = P(S) \times \prod_{i=0}^{n} P(a_i|S) \times \prod_{i=0}^{n} P(X_i|a_i\, S),$$

where:

- $P(S)$ is the prior distribution of the pose. In our case we make the assumption of an uniform distribution. However it could account for joint constraints and/or knowledge on given poses (splits are less probable than standing positions for example).
- $P(a_i = j|S)$ represents the *a priori* on the attachment with the sole knowledge of the pose. We set it proportional to the length of the corresponding segment s_j. Note that with our model, the segment lengths are fixed. Hence, this prior distribution does not depend on the pose.
- $P(X_i|a_i = j\ S)$ represents the probability that point X_i belongs to the limb corresponding to the segment s_j. We model it as a standard gaussian $\mathcal{N}(d(X_i, s_j), \sigma_j^2)$. Note that with an ideal skeletonization algorihm, all σ_j should be identical (uniform noise). However in practice, skeletonization methods lead to higher noise on the torso than on the arms or the legs. The variances σ_j are therefore set to approximately 1 cm, except for the torso where it is set to approximately 3 cm.

Finding the best pose consists then in maximizing the following posterior:

$$\begin{aligned} P(S|\{X_i\}) &\propto \textstyle\sum_{\{a_i\}} P(\{X_i\}\,\{a_i\}\, S), \\ &\propto \textstyle\prod_{i=0}^{n} \sum_{a_i} P(X_i a_i S), \\ &\propto P(S) \textstyle\prod_{i=0}^{n} \sum_{a_i} P(a_i|S) P(X_i|a_i S). \end{aligned}$$

Unlike the first solution (1), this posterior is well adapted for maximization as all its derivatives are continuous ($\mathcal{C}^\infty$ function). This posterior is also more robust as it marginalizes over all possible point to segment attachments instead of considering the single possible attachment from a point to its closest segment.

4.2. Fitting

In order to find the above MAP and as classical when dealing with hidden variables, we use an expectation maximization approach where:

- **The E step** consists in the computation of the expectation terms $E(a_i = j)$ for the current estimated pose $\overline{S}$:

$$\begin{aligned} E(a_i = j) &= P(a_i = j|X_0 \cdots X_n\, \overline{S}), \\ &= P(a_i = j|X_i\, \overline{S})\,, \\ &= \tfrac{P(a_i = j\, X_i\, \overline{S})}{\sum_{a_i} P(a_i\, X_i\, \overline{S})}\,; \end{aligned}$$

- **The M step** consists in finding the pose S maximizing:

$$F(S) = \textstyle\sum_{i=0}^{n} \sum_{a_i} E(a_i) \times log P(a_i\, X_i\, S).$$

Developping $P(a_i X_i S) = P(S)P(a_i|S)P(X_i|a_i\, S)$, we notice that the two first terms are constants. $P(S)$ is supposed uniform and the prior distribution on a_i does not depend on the pose. This leads to maximizing:

$$F(S) \propto \sum_{i=0}^{n} \sum_{a_i} E(a_i) \times logP(X_i|a_i\, S)$$

This is equivalent to minimizing its negated form:

$$\sum_{i=0}^{n} \sum_{a_i=j} E(a_i = j) \times \frac{d(X_i, s_j)^2}{2\sigma_j^2}$$

This formula defines a least squares problem. We use the well known Levenberg-Marquardt minimization algorithm as it is well adapted to this type of problem.

4.3. Tracking

The fitting process recovers a single pose at a given frame. To recover the motion of the user, we need to describe how we obtain the pose S_{t+1} at frame $t+1$ knowing the previous poses. This "propagation" problem consists in predicting a likely position S'_{t+1}. This prediction is used as an initial guess in the minimization process resulting in the final pose S_{t+1}. This prediction is commonly based on a dynamic model such as constant velocity or constant acceleration. Those models are efficient in modeling displacements of objects with relatively stable velocity. This condition generally implies a small ratio between the applied forces and the mass of the object. If this condition is valid for the root position and orientation of the body, it is clearly not valid for arms or legs. Their motions can be very erratic. In such cases tracking without dynamic model ($S'_{t+1} = S_t$) is a good solution as our experiments will demonstrate. A better solution would be to consider that the recovered velocity is noisy and incorporate a noise model in the propagation process with a particle filtering or belief propagation algorithm for example. In our experiments however particle filtering with up to 1000 particles did not improve results while significantly increasing the computational cost. We therefore seldom use it. Using non parametric belief propagation could lead to better results but again this would make the tracking process too slow for interactive systems.

5. Results

The body tracking method presented in the previous sections has been implemented and tested on various sequences of natural motions like walking in any direction. In this section, we present the corresponding results and discuss the robustness of our tracking method. We also discuss an important issue which is time performance through computations cost.

5.1. Data Acquisition

Image sequences were acquired using 6 firewire cameras shooting 780×580 images at 27 fps. These cameras are electronically triggered to ensure synchronization between images. Silhouettes are obtained through a standard background subtraction method. Results shown here are based on 2 sequences. The first one consists in a person walking in circle and lasts 15 seconds (around 400 frames), corresponding to 2 walking circles. The second one consists in a person performing a rapid kick in the air. It lasts 4 seconds with only 30 frames corresponding to the kick itself. Dimensions of the model were manually set, with an error of approximately 10%.

5.2. Tracking Results

Tracking results on the walking sequence are presented in figure 4. Validation is done by visually checking each frame. Only 6 frames out of 400 were found partially mistracked. Those 6 frames are organized in 2 groups of 3 consecutive frames, the 2 groups corresponding to the same situation in the sequence but at different times. In this situation, an elbow joint was found away from its real position (see figure 4-frame 290 for instance). This situation is due to visibility problems which result in skeleton data outliers between the torso and the arm that are wrongly attached to the arm, making the elbow moving toward the torso. Note that such a situation could probably be avoided by using temporal consistency through a dynamic model, again to the price of computational cost.

Tracking results on the kicking sequence are presented in figure 5. This sequence was used to evaluate the robustness of the approach to large motion between consecutive frames, or in other words to fast motions with respect to the acquisition frame rate. As shown by the results, the approach behaves well in such situation, even without prediction between consecutive frames, validating in that case the fact that no dynamic model is used.

5.3. Robustness

An important aspect for body tracking algorithms concerns their robustness to all types of noise. In our case, the main sources of errors are coming from noises in the input data as well as in the model parameters. Both are discussed in this section.

Noisy input data


Frame 10 Frame 50 Frame 90


Frame 130 Frame 170 Frame 210


Frame 250 Frame 290 Frame 330

Figure 4. Skeleton poses at different times for the walking sequence.

Sequences are not taken in specific environments, such as blue rooms, resulting in noisy silhouettes as obtained by background subtraction (see figure 6). Our approach is


Frame 70 Frame 80 Frame 90

Figure 5. Recovered body poses for the kicking sequence.

robust to those errors in different ways. First, notice that since the visual hull algorithm used is exact with respect to silhouettes, it does not add any additional noise but filters instead silhouette errors which are inconsistent in different views (or false positives). Second, the medial axis is pruned which allows for some errors in the shape estimation.

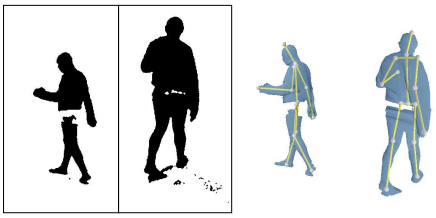


Figure 6. Left, examples of noisy silhouettes in the sequence. Right, result of the skeleton pose estimation with these silhouettes (2 different viewpoints).

Robustness to model errors

To test errors in the *a priori* model, noise was introduced in the dimensions of the model used for the walking sequence. The tracking performs correctly (only few partial mistracked frames) up to 20% of error. For higher noise, the number of mistracked frames increases: 30 frames out of 400 are mistracked in the walking sequence with 30% of error in the model. This robustness to model dimensions errors and the fact that the ratio between a limb size and the height of a person does almost not vary among the global population enables the model to be determined by only the height of the human body. This idea is currently being validated on a set of sequences acquired with users presenting different morphologies.

5.4. Real Time Issues

One of the main constraints imposed by interactive applications is real-time performances associated with low latencies. In this section, we discuss this issue for the two main steps of our method.

Skeleton Data Computation: As demonstrated in [3], the visual hull computation can be achieved in real time with a latency of 70 ms. The skeletonization cost lies essentially in the voronoi cells computation. This takes about 60 ms for 2000 surface points on an Opteron 2GHz. Distributing its computation allows this process to run at 30 frames per second but does not reduce its latency. Real time performance – less than 30 ms – is likely to be achieved in a year with the growth of computational power.

Tracking: Our tracking takes about one second per frame. Most of the time is spent computing distances from points to the model segments. This could be reduced by considering that only the 2 or 3 closest segments are relevant. This would reduce the computational cost by a factor of 5. Note also that this implementation is only an experimental prototype. Code optimization could significantly reduce computational cost. Moreover the *a posteriori* function can largely benefit from parallelization on multiple CPUs, as skeleton data input points can be treated independantly.

6. Conclusion

We have presented a 3D tracking algorithm that focus on motion parameters and relaxes dependencies on body shapes. It is based on a skeletal articulated model which is fitted to 3D skeleton data points. Those points lie on the medial axis of the visual hull, as obtained from silhouettes in multiple views. Experimental results on real sequences have been presented. They demonstrate the robustness of our approach to different aspects such as silhouette noise or dimension errors. This approach is relatively fast and should reach real time performances in the near future.

Several issues still remain to be addressed. First temporal consistency could be taken into account. One solution could be to integrate it directly in the generative model by changing $P(S)$ by $P(S_t|S_{t-1})$ corresponding to the probablistic dynamic model. Additionnaly, the uniform hypothesis on $P(S)$ could be changed to allow various joint constraints and to ensure that the skeletal model lies inside the visual hull. Second, the robustness of the tracking can be improved. In particular, the points to segments association could be more efficient if the visual hull containment constraint was taken into account. This would prevent attachment between torso points and arms for example. Also multiple cues such as color information (appearance model) or head/hand 3D tracking could be integrated in the process.

References

[1] A. Agarwal and B. Triggs. 3d human pose from silhouettes by relevance vector regression. In *Proceedings of CVPR'O4, Washington, (USA)*, pages II 882–888, June 2004.

[2] D. Attali and A. Montanvert. Modeling noise for a better simplifi cation of skeletons. In *Proceedings of ICIP, Lausanne (Switzerland)*, 1996.

[3] Authors.

[4] A. bottino and A. Laurentini. A silhouette based technique for the reconstruction of human movement. *Computer Vision and Image Understanding*, 83:79–95, 2001.

[5] G. J. Brostow, I. Essa, D. Steedly, and V. Kwatra. Novel skeletal representation for articulated creatures. In *Proceedings of ECCV'04, Prague*, pages Vol III: 66–78, 2004.

[6] J. Carranza, C. Theobalt, M Magnor, and H.-P. Seidel. Free-viewpoint video of human actors. *Proc. ACM Siggraph'03*, San Diego, USA, pages 569–577, July 2003.

[7] G. Cheung, S. Baker, and T. Kanade. Shape-From-Silhouette of Articulated Objects and its Use for Human Body Kinematics Estimation and Motion Capture. In *Proceedings of CVPR'O3, Madison, (USA)*, 2003.

[8] G. Cheung, T. Kanade, J.-Y. Bouguet, and M. Holler. A real time system for robust 3d voxel reconstruction of human motions. In *Proceedings of CVPR'00, Hilton Head Island, (USA)*, volume 2, pages 714 – 720, June 2000.

[9] C.-W. Chu, O. C. Jenkins, and M. J. Matarić. Markerless kinematic model and motion capture from volume sequences. In *Proceedings of CVPR'03, Madison, (USA)*, pages 475–482, 2003.

[10] I. Cohen, G. Medioni, and H. Gu. Inference of 3d human body posture from multiple cameras for vision-based user interface. In *5th World Multi-Conference on Systemics, Cybernetics and Informatics, Orlando*, 2001.

[11] Quentin Delamarre and Olivier D. Faugeras. 3d articulated models and multi-view tracking with silhouettes. In *Proceedings of ICCV'99, Kerkyra, (Greece)*, pages 716–721, 1999.

[12] J. Deutscher, A. Blake, and I. Reid. Articulated body motion capture by annealed particle fi ltering. In *Proceedings of CVPR'00, Hilton Head Island*, pages 2126–2133, 2000.

[13] T. Drummond and R. Cipolla. Real-time tracking of highly articulated structures in the presence of noisy measurements. In *Proc. of ICCV'01, Vancouver*, pages 315–320, 2001.

[14] D. M. Gavrila and L. S. Davis. 3-d model-based tracking of humans in action: a multi-view approach. In *Proceedings of CVPR'96, San Francisco)*, pages 73–80. IEEE Computer Society, 1996.

[15] K. Grauman, G. Shakhnarovich, and T. Darrell. Inferring 3d structure with a statistical image-based shape model. In *Proceedings of ICCV'03, Nice, (France)*, pages 641–648, 2003.

[16] M. Gross, S. Wuermlin, M. Naef, E. Lamboray, C. Spagno, A. Kunz, E. Koller-Meier, T. Svoboda, L. Van Gool, K. Strehlke S. Lang, A. Vande Moere, and O. Staadt. Blue-c: A spatially immersive display and 3d video portal for telepresence. In *ACM SIGGRAPH 2003*, San Diego, 2003.

[17] I. Kakadiaris and D. Metaxas. Model-based estimation of 3d human motion. *IEEE Transactions on PAMI*, 22(12):1453–1459, december 2000.

[18] A. Laurentini. The Visual Hull Concept for Silhouette-Based Image Understanding. *IEEE Transactions on PAMI*, 16(2):150–162, February 1994.

[19] J. Luck, D.Small, and C. Q. Little. Real-time tracking of articulated human models using a 3d shape-from-silhouette method. In *Roboton Vision*, pages 19–26, 2001.

[20] Ivana Mikić;, Mohan Trivedi, Edward Hunter, and Pamela Cosman. Human body model acquisition and tracking using voxel data. *IJCV*, 53(3):199–223, 2003.

[21] A. Senior. Real-time articulated human body tracking using silhouette information. In *proceedings IEEE Workshop on Visual Surveillance/PETS, Nice, France*, october 2003.

[22] J. Serra. *Image Analysis and Mathematical Morphology, Volume I*. Academic Press, 1982.

[23] C. Theobalt, M. Magnor, P. Schüler, and H.-P. Seidel. Combining 2d feature tracking and volume reconstruction for online video-based human motion capture. *International Journal of Image and Graphics*, 4(4):563–584, October 2004. Pacifi c Graphics 2002.