

Building A Visual Tracking System for Home-Based Rehabilitation

Yaqing Tao and Huosheng Hu

Department of Computer Science, University of Essex
Wivenhoe Park, Colchester CO4 3SQ, United Kingdom
Email: ytao@essex.ac.uk, hhu@essex.ac.uk

Abstract—Visual tracking of human movement has attracted increasing attention recently because of its wide spectrum of applications, including athletic and clinical performance analysis, human-computer interface, surveillance, motion capture for games and animation. There are two main techniques in the visual tracking of the human movement community: marker-based tracking and marker-free tracking. This paper presents a visual tracking system, which exploits both marker-based and marker-free methods, to support the home-based rehabilitation program. It provides a useful and complete tracking system to help the patients' who sustain a stroke to recover and improve their mobility at a home environment.

Index Terms—Visual Tracking, Motion Capture, Rehabilitation, Decision Module, Multiple Image Features.

I. INTRODUCTION

Visual tracking of human movement has attracted increasing attention from researchers. The interest is motivated by its wide spectrum of applications. The home-based rehabilitation program falls into the clinical analysis application domain. A visual tracking system used in this area should be able to track and analyse the limb motion of the human body accurately and to direct the human's motion automatically. The standard methods for human movement research and clinical movement analysis are to use marker-based motion capture systems. Normally, a marker-based tracking system uses three or more cameras to capture either active or passive markers attached to the joints of human body. Reflective balls are the most frequently used markers, which is passive markers and can reflect infrared light. The 3D positions of these markers are calculated by stereovision triangulation and the trajectories of these 3D markers can be obtained by tracking these markers through image sequences. Then the skeleton motion of the human body can be inferred from the trajectories of the 3D markers by using inverse kinematics methods.

A lot of existing commercial human motion tracking and capture systems can be employed for this purpose because they rely on markers. However, these systems are

quite expensive because they require special cameras and equipments. Also, their use requires accurate positioning all the markers in the right joint positions, which may be difficult for a stroke patient to do so.

Due to the limitations of marker-based tracking systems, a lot of attempts have been made to design a marker-free tracking system for the human motion capture [8][1]. Such a human motion tracking system would have more applications than marker-based tracking because only conventional cameras are required instead of intrusive markers and special cameras. However, designing a system for tracking human motion in video sequences is a non-trivial task. There are a lot of existing difficulties [20][17], including depth ambiguities, occlusion, high dimensionalities, kinematics singularities and appearance deformation. Although marker-free motion capture has been studied widely, no universal solution has been obtained to be able to solve all the difficulties with reasonable computational effort and good accuracy.

To solve the occlusion problem in a human motion tracking system, a model-based approach is usually preferred by many researchers. The model can also direct the feature extraction procedure and some motion and physical constraints can be embedded in the model, which is helpful to find the configurations of human body in the tracking procedure by narrowing the search spaces. The simplest human body model is a skeleton model [5], where the body segments are approximated as lines connected by joints. In 2D motion tracking or labelling, the body segments are represented by 2D patches [14][4], which serve as templates in the tracking procedure to find the corresponding patches in the image sequences. The most frequently used human body models are 3D volumetric models. The body segments are represented by cylinders [19][11], tapered cones [6][10] or shape deformable super-quadric ellipsoids [20] connected at body joints. The posture of a human at each time instance is represented by the configuration of the model, while the configurations of human models at each time step are controlled by all the joint angles and the global positions of the human body. Therefore the task of tracking is to estimate the set of joint angle parameters from the image sequences over time.

The information about how the human body moves is called the dynamics of the human body. This information is used to predict the new model states according to the previous model states thus reduce the search time in the image and constrain the state spaces of model configurations. Traditional methods assume that the human motion is smooth motion, which model the human motion as constant velocity or acceleration. However, human motion varies more complicatedly and it's not sufficient to use constant velocity or acceleration assumptions. More realistic motion models such as linear dynamics models [14] or linearised dynamics models of non-linear models [10] are used in some systems. The human motion is essentially non-linear. In order to improve the accuracy, the non-linear dynamics models are employed by learning from the training data using principal component analysis (PCA) [19] [12]. Other systems use Hidden Markov Models (HMM).

The new state of the model is updated by combining the predicted state and the detected image features from the raw images. Many different image features have been used in the tracking systems such as edges [11], optical flow [14][19], silhouettes/contour, points and colour [21]. As Sidenbladh analysed in[17], different image features may fail to detect some specific motion information, or may be unavailable in some situations. In order to build up a robust human motion tracking system, multiple image features have been employed in some tracking systems and good performances have been achieved [6][20][2][16]. Because Different image features can compensate the disadvantages for each other, thus make the tracking systems robust and effective.

Some systems use Single Hypothesis methods such as Kalman filters [10] or local-optimisation methods [9] to estimate and propagate the model states. This method assumes that the state of human body is uniquely determined at each time instance. However, cluttered background, occlusion, depth ambiguities and kinematic singularities may cause the state of human body multi-modals. Different methods such as multiple cameras [3], simplified background [10] and multiple hypothesis methods have been used to relief these problems. Unlike the single hypothesis methods, multiple hypothesis methods calculate multiple estimates of the model states at each time step and propagate them through image sequences. The most popular multiple hypothesis algorithm is the Particle Filter (PF), which can be used in multi-modal situations [7][13]. Deutscher et al. [7] [6] use edge and contour information to track the whole body motion successfully by using an annealed particle filter. Sidenbladh et al. [18] use a Bayesian inference framework to fuse the human body model, motion dynamics and image features in different modalities to a consistent format and use the particle filter to estimate the model states based on those information. Image features including edge, ridge and optical flow are used to

calculate the image feature likelihood by using a learning method. The particle filter has gained a lot of interest recently and some good tracking results have been achieved so far [6][17]. But it's computational expensive and requires huge computer resource.

In this paper, we propose to build a visual tracking system for home-based rehabilitation. Both marker-based methods and marker-free methods are adopted in this system. Marker-based tracking is used to build up the motion templates, which are pre-stored in a database and serves as the ground truth. The marker-free method is employed to track the patients' motion.

Section II introduces the whole visual tracking framework for the system. And three major modules of the system, modelling of human motion, motion tracking and motion comparison, are described respectively in section III. Our preliminary experimental results are presented in section IV. Finally, the conclusions and future work are presented in section V.

II. A VISUAL TRACKING SYSTEM

Patients who sustain a stroke need take rehabilitation exercises to improve their mobility every day with the help of expert physiotherapists or well-trained carers. We propose to develop a visual tracking system, which serves as the physiotherapist, to support the rehabilitation program for the patients at home environment.

The proposed visual tracking system consists of three parts: a patient, video cameras and a PC. The configuration of the system is illustrated in Figure 1. The patient's motion is filmed by video cameras and the captured image sequences are input into the PC. The software in the PC processes and tracks the human motion filmed in the images automatically. Therapeutic instruction is feed back to the patient to direct his motion.

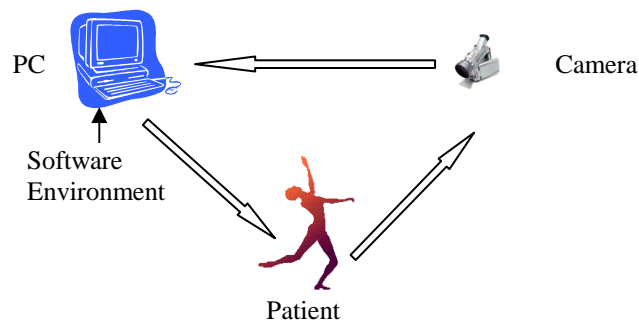


Figure 1 Configuration of the System

The software in the PC comprises of three modules: motion tracking module, database module and decision module. The three modules and the communication between different components in the whole tracking system are showed in Figure 2. The patients' rehabilitating exercise is captured by video cameras. The captured images are input into the PC and processed firstly by a motion-tracking module (see Fig.2 M2) to detect and track the human motion. Before the decision module (see Fig.2 M3), the tracked motion must be

recognized first, which means to category the type of the tracked motion, because this information is required to select the corresponding motion template from the database. After recognition, the motion-tracking module is linked to a decision-making module, which can judge whether the tracked motion is correct or not. In order to make the decision, the decision module uses both the therapeutic instruction and motion templates information derived from a database module to support the judgment. The database module is built up before the tracking and is pre-stored in the PC. The database consists of standard motion templates and therapeutic instructions. The combination of technology and knowledge modules in the whole visual tracking system will track the correctness of the patients' motion and measure the effectiveness of the patients' actions. The results will be feed back to the patients in audio or visual formats.

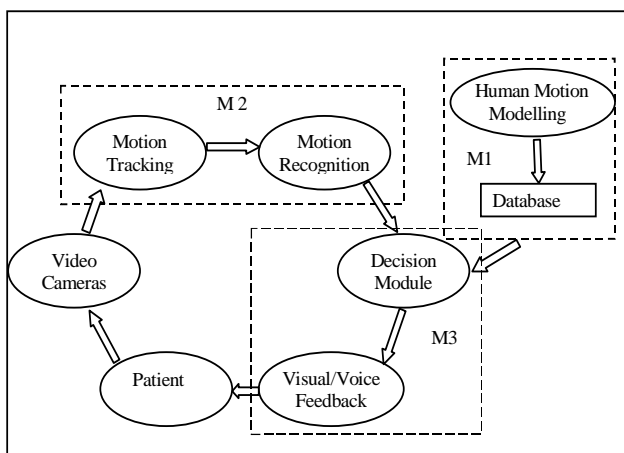


Figure 2 Modules of the Software

The three modules are described respectively in the following section.

III. BUILDING THREE MODULES

A. Modelling of Human Motion

In order to know whether the patients are doing their rehabilitating exercise in the right way, a quantitative comparison between the patients' activities and the correct motion templates is very important. The correct rehabilitating motion is pre-stored in a database and serves as a reference template. We consider the rehabilitating motions that are taken by some physiotherapist as the correct motion. In order to track the physiotherapist's motion as accurately as possible, the marker-based motion tracking method is adopted. There has been a lot of commercial marker tracking systems, which is widely used in athletic and gait analysis. In this section, we will discuss how to build the database by using a commercial system, namely Qualisys.

Small retro-reflective ball markers are attached to the performer's joints (see Figure 3 (a)), which can reflect infrared light and it's easy to detect and locate them in the 2D images filmed by different synchronized cameras

because they are bright points in the image planes (see Figure 3 (b)(c)(d)). The number of markers can vary from 20 to 30, which depends on the degrees of freedom of a performer or patient to be tracked. But it's better to keep the number of markers as low as possible to avoid the confusion of marker association. The motion of the performer here is captured by three digital cameras fixed in the scene.

The marker points in 2D images can be extracted using image processing techniques and 3D position of every marker (see Figure 4) is calculated from corresponding 2D marker point in each camera plane by using the stereo correspondence method. Then each 3D marker is tracked from one frame to the next frame and the correspondence is established through the image sequence. The marker trajectories can be obtained in this step.

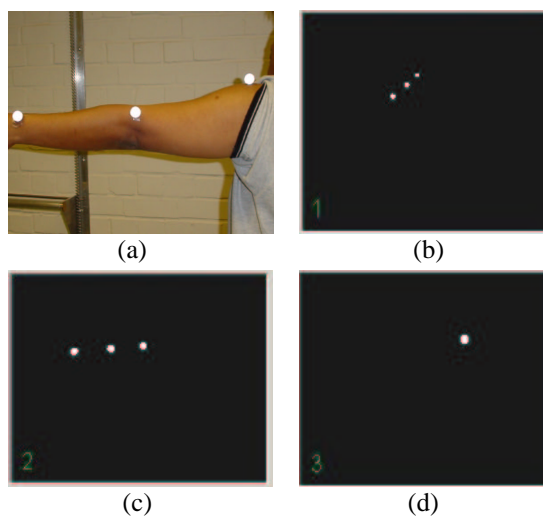


Figure 3 (a) Markers attached to the joints of the performer. (b)(c)(d) Marker points captured from three cameras.

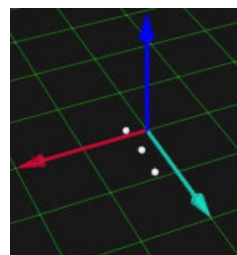


Figure 4 3D Markers

After obtaining the 3D marker trajectories for the entire image sequences, the skeleton motion of the performer can be inferred. The human motion is essentially the motion of the underlying skeleton, so it's sufficient to use the skeleton motion as the motion template in the database. The skeleton motion obtained from the Qualisys system is controlled by the 3D marker trajectories.

B. Motion Tracking

Computer vision systems allow in principle for touch-free motion tracking. It is desirable to track and recognize human motions only from video sequence footages instead

of using intrusive markers. Remarkable tracking results in 3D tracking and reconstruction of human motion by using multiple cameras have been achieved so far [6][3][9]. Sidenbladh [17] and Sminchisescu [20] even successfully track 3D human motion in a monocular image sequences. However, video-based human motion tracking problem is still far from being solved because of the difficulties addressed in section I. The existing tracking systems and algorithms are all based on some constraints to make the problems tractable, such as the constraints on the appearance of human and background (special clothes and background), multiple cameras and human body models (recover depth information and reduce occlusion), human motion models (reduce high dimensionality), etc. A robust system should be able to work in an unconstrained environment and uses as less constraints as possible.

We propose to develop a human motion capture method that can track the motion accurately for medical analysis in a marker-free environment. Tracking using only a single camera is desirable, but it's a highly underdetermined problem especially the lost of depth information. In order to recover the human motion accurately and efficiently, two video cameras are employed in the system. The motion-tracking module is formulated in an analysis-by-synthesis framework, which was introduced by O'Rourke et al. [15].

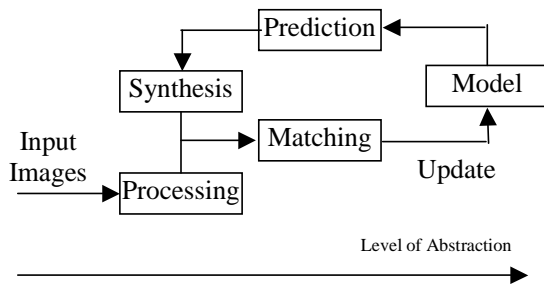


Figure 5 Analysis by Synthesis Framework

The framework consists of five components: prediction, synthesis, matching, image processing and the model. At each time step t , the prediction component uses the previous states of the model up to time t to predict the new states of the model at time $t+1$. Since the input images and the model are represented at different abstraction levels, the synthesis component is to translate the predicted states of the model from high level (state spaces) to low level (image plane). Thus the extracted image features of the image-processing component can be compared and combined at the same abstraction level with the predicted image features. Finally, the matching component updates the states of the model using the prediction and segmented image at time step $t+1$. The model interacts with system components actively during the tracking period. This is an iteration of the tracking in the motion-tracking module. Different techniques and algorithms have been developed in each component to achieve the tracking task. The proposed methods used in our tracking system are as follows:

1) Shape Deformable Human Model

One of the difficulties in human motion tracking systems is the changing appearance of different subjects. Most tracking systems assume that the shape of the human model used in the systems is constant [11][6][10][17]. This constraint simplifies the modelling task. Thus the state of the model is determined only by the joint angles and the global position parameters of the model. If the set of state parameters of the model at time step t is represented by Φ_t , for each time instance, once Φ_t is determined, the state of the model is captured. However, this method is limited to track specific subject. In order to extend such systems to track variable subjects, a shape deformable human model is proposed, which allows to obtaining the length and shape of the limbs from the image sequences. The limbs' shape parameters are also included into the state parameters set Φ_t . Using deformable model makes the tracking systems robust, but the bad news is that this method makes the dimensionality of parameter spaces even high; an efficient method must be found to relief the burden of the searching problem. We propose to represent the shape of limbs by super-quadratics, which is allowed to deform its shape according to the different subjects being tracked by tapering and blending [20]. If possible, we also want to try other deformable models, compare them with the deformable super-quadratics and select the better one.

2) Example-Based Motion Dynamics

Humans move in a complex way. If the dynamics of the human body is available before tracking, this prior information can be used to predict and propagate the states of the body model from previous time instances to the next time instances. We propose to build up the motion dynamics by learning from the training data. Although this kind of motion dynamics will limit the tracking system to only track the learned motion types, it's applicable for home-based rehabilitation because the patients are required to do a series of pre-defined rehabilitating exercise. The most frequently used learning technique is principle component analysis [16][2][19]. We are going to use this technique to build the motion models from training data.

3) Synthesis

The predicted state of model is in 3D, while the detected image features are 2D measurements. By using the perspective camera models, the 3D state of model can be projected into 2D image planes. After this stage, the predicted model state is at the same level as the detected image features. Thus this information can be used to update the new state of the model.

4) Multiple Image features

The image features that can be obtained through image processing components include different modalities. Original tracking systems use only one kind of image features. As we know, every image feature suffers from its

weakness. For example, optical flow usually results in tracking drift; edge information may cause observation ambiguity and contour fails to offer the details within the object outline. By realizing the different properties of these image features, using multiple image features in a tracking system seems to be a reasonable way to improve the system's tracking performance. Some researchers have successfully used multiple image features in their tracking systems [20] [17][2]. The most useful image features are edges, contour, skin-colour and optical flow. We will investigate all these image features and combine some or all of them into our tracking system to support the tracking.

5) Bayesian Model-Based Tracking

Model-Based human motion tracking is essentially to estimate the state parameters set Φ_t of the model based on a sequence of images I_T at time t . I_T is the image sequence up to time t . A useful way to fuse the multiple image features, the deformable human model and the motion model is the Bayesian rule [17]:

$$p(\Phi_t | I_T) = \kappa p(I_t | \Phi_t) p(\Phi_t | I_{T-1}) \quad (1)$$

where I_t is the image at time t and κ is a normalizing constant. $p(\Phi_t | I_T)$ is the probability distribution of the state parameters Φ_t based on the image sequence. It is updated by the observed image features $p(I_t | \Phi_t)$ and the predicted state of the model $p(\Phi_t | I_{T-1})$. Equation (1) can be written as equation (2) according to the Bayesian inference. As we can see, the predicted state of the model is derived from evolving previous model state $p(\Phi_{t-1} | I_{T-1})$ according to the motion dynamics $p(\Phi_t | \Phi_{t-1})$.

$$p(\Phi_t | I_T) = \kappa p(I_t | \Phi_t) \int p(\Phi_t | \Phi_{t-1}) p(\Phi_{t-1} | I_{T-1}) d\Phi_{t-1} \quad (2)$$

It is clear the Bayesian method integrates different tracking sources into one framework naturally. The Bayesian tracking method is carried out by using a particle filter. Some previous methods that use Kalman Filters assume that the distribution of the state of the model $p(\Phi_t | I_T)$ is Gaussian (uni-modal), which means the task of the tracking is to find the maximum of the distribution. However, it is not truth when tracking is in a cluttered environment or occlusion occurs, where the distribution is multiple-modals. The particle filter is proved powerful in these situations, which digitises the continuous distribution using an importance sampling. The continuous distribution is approximated by a set of particles S_t^i weighted by $\pi_t^i, (1 \leq i \leq N)$. So instead estimating only one state of the model at each time step, multiple states of the model are estimated by using the particle filter. The expectation of these particles or the one that has the biggest weight is selected as the ultimate estimation at each time. The problem with this approach is that it's highly computational complexity; so efficient implementation method is required.

C. Motion comparison

There are two properties that can be used to determine the motion quality. The first one is to match the motion trajectories of body joints on the patient's body with the trajectories of corresponding body joints on the body template in the database. The second one is to compare the joint angles between two connected body segments of the patient to the corresponding joint angles of the template in the database. The advantages of using the joint trajectory and joint angle are that they are translation and rotation invariant and independent of view direction. Since the state of the model is expressed by joint angles in the proposed system, the second method is employed in our system.

IV. EXPERIMENTAL

To develop a complete tracking system for the home-based rehabilitation, the three modules discussed above need to be built up. Our preliminary experiment involves the construction of the motion templates, which are pre-stored in the database module.

The details of tracking procedure by using the commercial tracking system, Qualisys, has been discussed in section III. Here, a motion template of an arm is built. The passive markers are attached to the body joints of the performer (see Figure 3 (a)). And the performer does the motion of arm flexion and extension. As we can see from Figure 6 (top-bottom view), the 3D skeleton postures of the arm in different time instance are showed. The three markers in these images correspond to the marker attached to the shoulder, the elbow and the wrist respectively. The capture rate is 120HZ and 5 sec, totally 600 frames are captured in this experiment. Figure 6 shows only a part of the image sequences, in which the motion is the arm extension. The flexion motion is simply the inverse of extension.

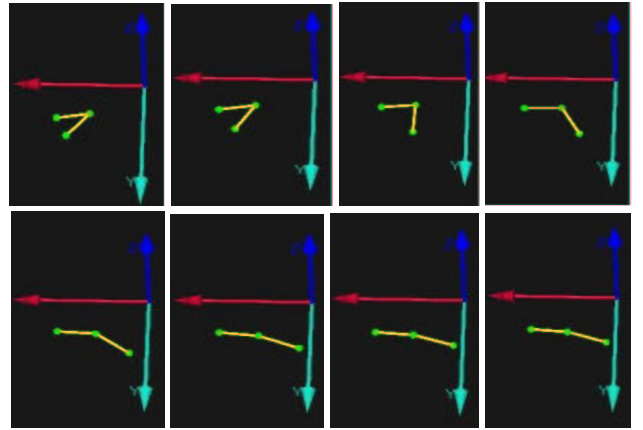


Figure 6 the Motion of Arm Extension. 3D skeleton extension motion of an arm is showed at frames 28, 48, 68, 88, 108, 128, 148, 150.

Based on the markers data above, the joint angle between upper arm and fore arm can be calculated. Figure 7 (a) illustrates the trajectory of joint angle from frame 28 to frame 150. The whole joint angle trajectory is showed

in Figure 7 (b). As we can see two cycles of arm flexion and extension are captured in the 600 frames. This information is required in the decision module, so it is used as the motion template of arm flexion and extension to build the database module. It is easy to build other motion templates by using this method.

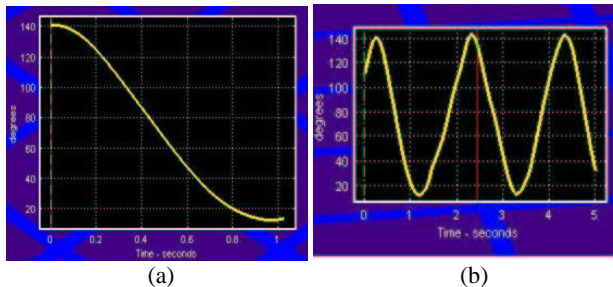


Figure 7 the Trajectory of Joint Angle. (a) Joint Angle Trajectory of Arm Extension. (b) Joint Angle Trajectory of the Whole Image Sequence

V. CONCLUSIONS AND FUTURE WORK

A visual tracking system for the home-based rehabilitation program is presented in this paper. Traditionally, patients' rehabilitation needs the help of physiotherapists or expert carers. By using this system, it is possible for the patients who sustain a stroke to do their rehabilitation exercise at a home environment without the need of physiotherapists' presence. Compared to commercial marker-based tracking systems, the proposed system uses conventional, cheaper equipments and offers the user complete freedoms. Thus it has more applications than the marker-based tracking system, such as surveillance tracking and human computer interface. The problems with the system are that it is computational expensive, only suitable to track slow motion. Especially the initialisations of the tracking are achieved manually. Future work will be focused on the following aspects:

- Build up a complete tracking system as proposed in Fig 2.
- Speed up the tracking algorithm and make the system run in real-time.
- Design a method to initialise the tracking system automatically.

ACKNOWLEDGMENT

The authors would like to thank Harun, Hafizah H to let us film her and thank Dr Martin Sellens and Prof. Ralph Beneke in the Biological Sciences Department to allow us to use their Qualisys system to collect data.

REFERENCES

[1] Aggarwal, J. K., and Cai, Q. "Human Motion Analysis: A Review". *Computer Vision & Image Under.:* CVIU. 1999.
 [2] Bowden, R., Mitchell, T. A., and Sarhadi, M. "Reconstructing 3D Pose and Motion from a Single Camera View". In *Proc. BMVC*, John N. Carter & Mark S. Nixon Eds, University of Southampton, Vol. 2, pp 102-108, Southampton, Sept. 1998.

[3] Bregler, C., and Malik, J. "Tracking People with Twists and Exponential Maps". In *IEEE International Conference on Computer Vision and Pattern Recognition*, 1998.
 [4] Cham, T. J., and Rehg, J. M. "A Multiple Hypothesis Approach to Figure Tracking". *Computer Vision & Pattern Recognition*, Ft. Collins, CO, pages 239-245, June 1999.
 [5] Chen, Z., and Lee, H. J. "Knowledge-guided visual perception of 3D human gait from a single image sequence". *IEEE Trans. On Systems, Man, and Cybernetics*, 22(2):336-342, 1992.
 [6] Deutscher, J., Blake, A., and Reid, I. "Articulated Body Motion Capture by Annealed Particle Filtering". *IEEE Conf. on Computer Vision and Pattern Recognition*, Vol. 2, pp.126-133.
 [7] Deutscher, J., North, B., Bascl, B., and Blake, A. "Tracking through singularities and discontinuities by random sampling". *Proc. Int. Conf. Computer Vision*, 1144-1149 (1999).
 [8] Gavrilu, D. M. "The Visual Analysis of Human Movement: A Survey". *Computer Vision and Image Understanding*, vol.73, no1, pp.82-98, 1999.
 [9] Gavrilu, D. M., and Davis, L. S. "3D Model-based Tracking of Humans in Action: a Multi-view Approach". *Proc. of IEEE Conference on Computer Vision and Pattern Recognition*, pp. 73-80, San Francisco, U.S.A., 1996.
 [10] Goncalves, L., Bernardo, E. D., Ursella, E., and Perona, P. "Monocular Tracking of the Human arm in 3D". *ICCV95*.
 [11] Hogg, D. "Model-based vision: a program to see a walking person". *Image and Vision Computing*, 1(1): 5-20, 1983.
 [12] Howe, N. R., Leventon, M. E., and Freeman, W. T. "Bayesian Reconstruction of 3D Human Motion from Single-Camera Video". *Advances in Neural Information Processing Systems (NIPS)*, Vol. 12, View 0820, Nov.1999
 [13] Isard, M. and Blake, A. "CONDENSATION -- conditional density propagation for visual tracking". *Int. J. Computer Vision*, 29, 1, 5--28, (1998).
 [14] Ju, S. X., Black, M. J., and Yacoob, Y. "Cardboard people: A parameterised model of articulated motion". *2nd Int. Conf. on Automatic Face- and Gesture-Recognition*, Killington, Vermont, Oct 1996, pp.38-44.
 [15] O'Rourke, J., and Badler, N. "Model-based image analysis of human motion using constraint propagation". *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2(6):522-536, 1980.
 [16] Ong, E. and Gong, S. "Tracking Hybrid 2D-3D Human Models from Multiple Views". *Proceedings of IEEE International Workshop on Modelling People*, 20 September 1999, Corfu, Greece.
 [17] Sidenbladh, H. "Probabilistic Tracking and Reconstruction of 3D Human Motion in Monocular Video Sequences". PhD Thesis TRITA-NA-0114, ISBN 91-7283-169-3, Dept. of Numerical Analysis & Comp. Sci., KTH, Sweden 2001.
 [18] Sidenbladh, H. and Black, M. "Learning Image Statistics for Bayesian Tracking". In *IEEE International Conference on Computer Vision*, 2001.
 [19] Sidenbladh, H., Black, M. and Fleet, D. "Stochastic Tracking of 3D Human Figures Using 2D Image Motion". In *European Conference on Computer Vision*, 2000.
 [20] Sminchisescu, C. "Estimation Algorithms for Ambiguous Visual Models Three-Dimensional Human Modelling and Motion Reconstruction in Monocular Video Sequences". PhD Thesis, Institute National Politechnique de Grenoble (INRIA), July 2002.
 [21] Wren, C., Azarbayejani, A., Darrel, T. and Pentland, A. "Pfinder: Real-time tracking of the human body". In *Proc. SPIE*, Bellingham, WA, 1995.