Visual tracking of independently moving body and arms

Markos Sigalas, Haris Baltzakis, Panos Trahanias
Institute of Computer Science
Foundation for Research and Technology – Hellas (FORTH)
E-mail: {msigalas, xmpalt, trahania}@ics.forth.gr

Abstract— Tracking of the upper human body is one of the most interesting and challenging research fields in computer vision and comprises an important component used in gesture recognition applications. In this paper a probabilistic approach towards arm and hand tracking is presented. We propose the use of a kinematics model together with a segmentation of the parameter space to cope with the space dimensionality problem. Moreover, the combination of particle filters with hidden Markov models enables the simultaneous tracking of several hypotheses for the body orientation and the configuration of each of the arms.

I. INTRODUCTION

VISION based detection and tracking of human body parts is a difficult goal with particular importance and a large number of application areas such as human-robot interaction, video surveillance, virtual environments and medical diagnosis. Among the above, an application with particular interest is hand gesture recognition which is also the target application behind this work.

Arm and hand tracking approaches for gesture recognition can be classified into two different families, depending on whether they involve tracking on the image plane or the human body pose space. Some of the most successful approaches of the first family detect hands as image blobs that share a specific characteristic, like skin color [10, 22], in each frame and temporally match blobs that occur in neighboring locations across frames. Hand shape can also be useful for the detection of the hand in sequential image frames either by using shape context descriptors [19, 20, 21] or through hand contour extraction and matching. The latter mainly uses real-time edge detection and similarity matching techniques for the generation of hand hypotheses [2, 4, 8, 9, 11, 13] or deformable contours, namely snakes [3]. Expanding this, hand detection and tracking can be enhanced by combining hand shape with other blob features, such as orientation and motion [27].

The second family of approaches (pose space approaches) can be further classified depending on whether they use a specific model or not. Non model-based approaches (e.g. [20]) commonly use training data to learn associations of training examples with corresponding known poses or learn specific mappings from observable image features to human pose space [21, 24]. Model-based (or generative) approaches perform tracking directly on the hidden pose space by generating measurement predictions according to a specific model and by comparing them with actual observed features with respect to specific error functions. Although model-based approaches are generally assumed to be more accurate than non model-based ones, they are usually more computationally expensive due to the large number of parameters that have to be tracked.

Different models require different image features to construct feature-model correspondences. Point and line features are employed in kinematic hand models to recover angles formed at the joints of the hand [7, 12, 15, 19]. Hand postures are then estimated, provided that the correspondences between the 3D model and the observed image features are well established.

Depending on the application at hand, various models have been proposed to detect the kinematics of the human body and/or arm/hand. For example Rehg and Kanade in [5] and Stenger et al. [23] use a full hand model with 27 Degrees of Freedom (DoFs), while Gavrila and Davis [8] use a total of 30 DoFs to model both the body and the arms. Simpler modeling can also come in hand. Goncalves et al. [6] model the arm with 7 DoFs while McCormick and Isard [17] represent user's hand as an articulated rigid object.

Some of the most successful tracking methods are the probabilistic ones. Kalman Filter is a widely used technique for hand pose estimation and tracking. For example, in Stenger et al. [18, 23], an unscented Kalman Filter is used in order to minimize the geometric error of the estimated hand model pose.

However, approaches based on Kalman Filters are limited by the unimodal nature of Gaussian densities which cannot be used to simultaneously track different solutions. An alternative method that has been used for model-based tracking is particle filtering [1]. Although particle filters offer significant advantages, including the ability to simultaneously track multiple solutions, the number of required particles increases dramatically with the number of tracked parameters increasing the computational cost and, thus, reducing performance. Therefore, a variety of techniques are often utilized to reduce the number of particles. For example, McCormick and Isard [17] adapt the partitioned sampling technique proposed in [16], to the problem of tracking articulated objects. In this way, they cope with the dimensionality problem by dividing the parameter space into M 'partitions' and then searching within each parameter partition separately.

In this work, we present a probabilistic model-based

This work is partially supported by the FP6-Information Society Technologies program, European Commission INDIGO research project (Contract No. IST-045388).

method for tracking human arms using stereo visual input. For modeling the human arm we have employed a model similar to the one presented in [26], which uses 9 DoFs in total for modeling both hands (4 DoFs for each arm and one DoF for the orientation of the human torso). Moreover, in order to reduce the complexity of the problem and to meet the increased computational requirements of the task at hand, we split the model space into three different partitions and perform tracking separately in each of them. More specifically, we employ a Hidden Markov Model (HMM) to track the orientation of the human torso in the 1D space of all possible orientations and two different sets of particles to track the four DoFs associated with each of the two hands using a particle filtering approach.

The proposed approach is perfectly suited to the problem examined here, where the track of each arm is assumed to be statistically independent from the track of the other arm given the orientation of the torso. The orientation of the torso is assumed to be independent of the arm, although its exact value is required to realize a useful observation model. Our approach has been extensively tested on various datasets and the results were very promising.

II. Approach Overview

The proposed methodology is depicted in the diagram of Fig. 1. Initially, skin-color blobs are extracted on both input 2D images and the depth of the blobs centroid, is estimated using triangulation. Knowing the 3D location of the extracted blobs, those corresponding to human heads are easily separated from those corresponding to hands by using simple heuristics, while the size of arm limbs can be estimated proportionally to the height by utilization of anthropometric measures.

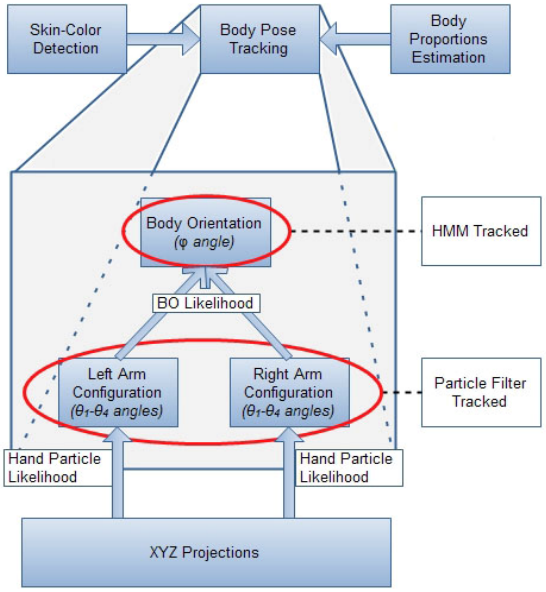

Fig. 1. Block diagram of the proposed tracking methodology.

Based on the extracted proportions, a 4 DoF kinematic model is assumed for each arm and both arms are assumed to be mounted on a rigid human body located on the ground plane at the locations constrained by the 3D coordinates of the head. The orientation of the human body, i.e. its orientation φ around the vertical axis (see Fig. 2) is assumed to be unknown.

In reality, the targeted problem is a 9 parameters estimation problem. These parameters generate a set of possible solutions which are propagated over time with the use of human body dynamics. However, tracking in a 9D space requires an excessively large number of particles which cannot be handled by a single tracker. To cope with this issue, i.e. dimensionality reduction, the parameter space is fragmented into three sub-spaces; a 1D parameter space for body orientation angle and two 4D spaces, one for each hand.

Body orientation angle φ is appropriately quantized (50 quantization levels) and tracked over time by means of an HMM. For every possible solution, a separate particle filter tracker is employed for each arm, as shown in Fig. 1. The result of each particle filter is used to estimate the observation probability, which is subsequently employed to update the HMM.

In the following sections more details regarding each of the above described modules are given.

III. Skin-Color Detection

As already mentioned, the first step of the proposed approach is to detect skin-colored areas within the input images. For this purpose we use a technique similar to the one in [22]. Initially, the foreground area of the image is extracted, as described in [14]. Then, foreground pixels are characterized according to their probability to depict human skin and then grouped together into blobs using hysteresis thresholding and connected components labeling.

The probability that a pixel depicts a skin-colored area, given its observed color c, is computed according to the Bayes rule as:

$$P(s \mid c) = P(c \mid s)P(s) / P(c), \quad (1)$$

where $P(s)$ and $P(c)$ are the prior probabilities of skin pixels and pixels having color c respectively. Color c is assumed to be a 2D variable encoding the U and V components of the YUV color space. $P(c|s)$ is the prior probability of observing color c for skin colored regions. All three components in the right side of the above equation are computed offline during training.

After probabilities have been assigned to each image pixel, hysteresis thresholding and connected components labeling is used to extract solid skin color blobs. Finally, size filtering is also applied in order to ensure that small, isolated blobs are eliminated.

IV. Kinematic Model

The proposed model is depicted in Fig. 2. According to this model, as already mentioned, the whole human body is

assumed to be rigid with only one unknown degree of freedom: the orientation φ.

The human arms are assumed to be attached to the human body at fixed locations corresponding to the shoulders. Both arms are modeled by a 4 DoF kinematic model, similar to the one presented in [26]. Angles θ_1 to θ_3 represent the 3 DoFs of the shoulder and θ_4 that of the elbow. L_1 and L_2 are the lengths of the upper arm and the forearm, respectively.

The length of the limbs can be calculated proportionally to the observed human's (actor's) height. To preserve simplicity, we assume that during initialization the head is the topmost detected skin color blob. Since the input is provided by a stereo pair, its 3D location, and therefore the height, can be estimated by triangulation. Having estimated the height of the actor in the scene, the anthropometric measurements presented in [25], are used to calculate the lengths L_1 and L_2.

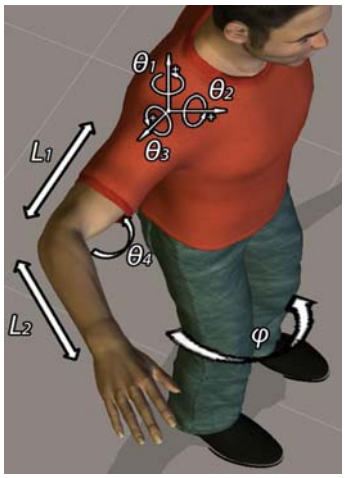

Fig. 2. Arm's Kinematic model

TABLE 1
DENAVIT-HARTENBERG PARAMETERS

i	α_{i-1}[deg]	a_{i-1}[m]	d_i[m]	θ_i[deg]
1	90	0	0	θ_1-90
2	-90	0	0	θ_2+90
3	90	0	L_1	θ_3+90
4	-90	0	0	θ_4-90
5	0	L_2	0	0

The location in space for each joint is calculated with respect to the base joint, namely the shoulder. In general, the transformation from joint P_{i-1} with coordinates $[x_{i-1}\ y_{i-1}\ z_{i-1}]^T$ to joint P_i with coordinates $[x_i\ y_i\ z_i]^T$ is given by:

$$P_i = {}^{i-1}_{i}T \cdot P_{i-1}, \tag{2}$$

where

$${}^{i-1}_{i}T = \begin{bmatrix} {}^{i-1}_{i}R & {}^{i-1}_{i}t \\ 0 & 1 \end{bmatrix} \tag{3}$$

is the transformation matrix, ${}^{i-1}_{i}R$ is the rotation matrix and ${}^{i-1}_{i}t$ is the translation vector.

By using the Denavit-Hartenberg parameterization (see Table 1), the rotation matrix and the translation vector which transform joint *i-1* into joint *i* can be expressed as:

$${}^{i-1}_{i}R = \begin{bmatrix} \cos(\theta_i) & -\sin(\theta_i) & 0 \\ \sin(\theta_i)\cos(\alpha_{i-1}) & \cos(\theta_i)\cos(\alpha_{i-1}) & -\sin(\alpha_{i-1}) \\ \sin(\theta_i)\sin(\alpha_{i-1}) & \cos(\theta_i)\sin(\alpha_{i-1}) & \cos(\alpha_{i-1}) \end{bmatrix} \tag{4}$$

and

$${}^{i-1}_{i}t = \begin{bmatrix} a_{i-1} \\ -d_i \sin(\alpha_{i-1}) \\ d_i \cos(\alpha_{i-1}) \end{bmatrix}. \tag{5}$$

For non sequential joints, the corresponding transformation is given by the product of all intermediate transformations as:

$${}^{0}_{N}T = {}^{0}_{1}T\,{}^{1}_{2}T \ldots {}^{N-1}_{N}T\,. \tag{6}$$

Therefore, for example, the elbow and the hand locations are computed with respect to the shoulder, by applying transformations ${}^{0}_{3}T$ and ${}^{0}_{5}T$ respectively.

V. ARM HYPOTHESIS TRACKING

The above-defined kinematic equations along with the camera perspective transformations are used in order to project hypotheses from the hidden parameter space onto the observation space and, therefore, evaluate each of the hypotheses. More specifically, the kinematic equations are used to transform the rotations of the human body and the angles of the arm joints to the 3D coordinates of each joint (shoulder, elbow and hand) while the camera projection transformations transform the resulting 3D coordinates to image coordinates.

As stated previously, arm hypotheses are being tracked by independent sets of particles. A particle filter is a sequential sampling method for approximating an unknown distribution (that of the hidden parameters) using the observed features. The belief of the filter is expressed as a set of samples, namely the particles, distributed in the whole state-space. The denser the particles in a certain region of the state-space are, the higher the confidence of the corresponding hypothesis is.

To generate and maintain the particles, the Sampling/Importance Resampling (SIR) algorithm, introduced by Rubin [1], is utilized. According to SIR, instead of sampling the true distribution $P(\Theta_k \mid y_1, y_2, ..., y_k)$ (where Θ_k represents the estimated configuration for the time instant k and $y_1, y_2, ..., y_k$ are the observations up to that time), samples are drawn from the so-called proposal distribution $\pi(\Theta_k \mid y_1, y_2, ..., y_k)$. To compensate with this difference, each sample m is also assigned a weight w^m which is computed according to the Importance Sampling Principle:

$$w_k^m = \frac{P(\Theta_k^m \mid y_1, y_2, ..., y_k)}{\pi(\Theta_k^m \mid y_1, y_2, ..., y_k)}. \quad (7)$$

By choosing the proposal distribution to be the transition prior $P(\Theta_k^m \mid \Theta_{k-1}^m)$, the weights can be computed as:

$$\begin{aligned} w_k^m &= \frac{P(\Theta_k^m \mid y_1, y_2, ..., y_k)}{\pi(\Theta_k^m \mid y_1, y_2, ..., y_k)} \\ &\approx \frac{P(y_k \mid \Theta_k^m) P(\Theta_k^m \mid \Theta_{k-1}^m)}{P(\Theta_k^m \mid \Theta_{k-1}^m)} w_{k-1}^m \\ &= P(y_k \mid \Theta_k^m) w_{k-1}^m \end{aligned}, \quad (8)$$

which corresponds to the likelihood of the observations for the particle *m* at time instant *k*.

To avoid degenerate situations in which large number of samples have weights close to zero, after a few iterations, SIR also includes a resampling step which ensures that unlikely samples are replaced with more likely ones.

In our implementation, the likelihood $P(y_k \mid \Theta_k)$ of each particle is determined from the back-projection errors of the elbow and the hand with reference to the centroids of the detected skin color blobs – assuming that only one actor exists in the scene at any time and since head is determined, the rest of the blobs are expected to be hands.

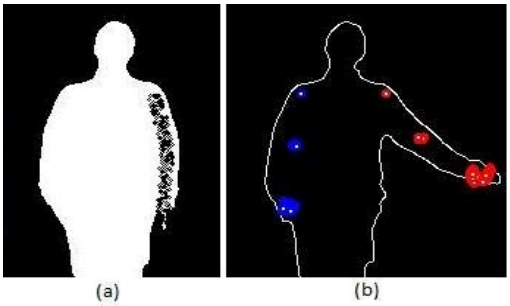

Fig. 3. (a) Projections of intermediate points (small black circles) onto the foreground image. (b) Shoulders are forced to lie close to the foreground edge.

More specifically, the likelihood $P(y_k \mid \Theta_k^m)$ of particle *m* is computed as the product of two components:

$$P(y_k \mid \Theta_k^m) = p_1^m \cdot p_2^m. \quad (9)$$

The purpose of the first component p_1^m is to assign larger likelihoods to particles with end points close to the detected hand blobs. For this purpose, the first component is computed according to the distance of its hand projected coordinates from the closest skin color blob. Therefore, for a particle *m,* the first weight component is calculated as:

$$p_1^m = \prod_{cm=1}^{2} \frac{1}{1 + \alpha \cdot d(x_{cm}^m, c_{cm,b})^2}, \quad (10)$$

where *cm* is the index of the camera, *b* the currently examined skin color blob, *d* the distance between the projected coordinates of the hand and the skin color blob centroid and α is a constant (0.01 in our experiments), expressing the steepness of the weight decrease.

The second component p_2^m is used to assign higher likelihood to particles that produce solutions that are more likely with respect to the projections of points belonging to the whole arm's volume. This is achieved by projecting several points of each hypothesis onto the foreground of the image. By considering upper arm and forearm as rigid objects, one should expect that the projections of the points between two joints of a correctly hypothesized arm should lie into the foreground of the image (Fig. 3a). Therefore, p_2^m for the particle m is computed as:

$$p_2^m = (n / N) / p_{2\,\max}^m, \quad (11)$$

where *n* is the number of points that project inside the foreground area of the image, *N* is the total number of projected points and $p_{2\,\max}^m$ is the normalization factor.

After having computed the likelihood for all particles, the resampling step takes place. In order to enhance tracker's performance, the number of particles of each tracker is allowed to vary according to the confidence of the filter. That is, we start with a large number of particles (approximately 2000 for each arm hypothesis). The number of particles is gradually reduced as the filter converges to a solution and becomes confident of the actual state of the arm.

VI. Body Orientation Tracking

The above described particle filtering process is responsible for tracking each arm's configuration "within" a body's orientation. In other words, for each of the HMM states (i.e. body orientations) two sets of independent particle filters –one for each arm- are generated and tracked over time. The result of each particle filter is used to update the probability of the corresponding HMM state. This means that the weights of the particles are used to calculate the observation likelihood for a particular body orientation state. Therefore the *belief* of the HMM that the body orientation angle φ is *r* at time instant *t+1,* given all observations O_1, O_2,…,O_{t+1} up to that time can be expressed as:

$$\begin{aligned} P(\varphi_{t+1} = r \mid O_1, O_2, ..., O_{t+1}) &= \alpha P(O_{t+1} \mid \varphi_{t+1} = r) \\ &\sum_{k=1}^{N} P(\varphi_t = k \mid O_1, O_2, ..., O_t) P(\varphi_{t+1} = r \mid \varphi_t = k) \end{aligned}, \quad (12)$$

where $P(O_{t+1} \mid \varphi_{t+1} = r)$ is the likelihood of observation O_{t+1} given that the orientation is *r*, $P(\varphi_t = k \mid O_1, O_2, ..., O_t)$ is the belief of the HMM for the previous time instant and $P(\varphi_{t+1} = r \mid \varphi_t = k)$ is the transition probability of moving from rotation *k* to rotation *r* at time instant *t+1*. Usually the value of the transition probability is high for close to *r* hypotheses and steeply decreases for the distant ones. Finally α is a normalization factor so that all state probabilities sum up to 1.

In our implementation the likelihood is calculated as:

$$P(O_{t+1} \mid \varphi_{t+1} = r) = \mu_{r,w^m} \cdot s_r \,. \quad (13)$$

The first component, μ_{r,w^m}, is the mean of the weights of the particles belonging to body orientation r and represents the likelihood of the observations derived from the arm projections, while s_r is a confidence factor with respect to the observations derived directly from the orientation hypotheses. Shoulders are considered as non moveable point joints and their locations are univocally calculated according to the corresponding body orientation angle. Since shoulders are the outmost point of the upper body (ignoring arms), s_r ensures that the shoulder projections lie within an allowed area, close to the edges of the foreground area (Fig. 3b) and can be expressed as:

$$s_r = (D_{\max} - d_s) / (D_{\max} - D_{\min}) \,, \quad (14)$$

where d_s is the distance of the shoulder projection from the foreground edge and $D_{\max}$ and $D_{\min}$ are the maximum and the minimum distance allowed, respectively. By imposing this constraint, it is guaranteed that only those hypotheses which represent the actual pose will obtain high confidence, additionally increasing the efficiency of the proposed tracker.

VII. Results

The above described approach has been extensively tested on several real image sequences. The target of our experiments was to confirm the performance of the proposed method on simple as well as complex and ambiguous cases, including moving actor and simultaneous movement of both arms, parallel to or towards the cameras. In all examined cases, the tracker performed accurately, leading to very promising results that are presented in this section.

The first presented example includes a static actor, moving one hand (Fig. 4) or both hands simultaneously (Fig. 5). The images in the left column of these figures show the input image taken from one camera of the stereo pair, together with the projections of the most confident particles. On the right column, a 3D view of the solutions that correspond to the particles with the largest weights is depicted. The topmost white sphere represents the location of the actor's head, while the colored spheres represent the location of the shoulders, the elbows and the hands. As can be easily observed, the tracker is converged to the actual pose of the actor.

Fig. 6 depicts frames from another sequence where the actor is also allowed to rotate his torso. The state of the HMM that tracks the rotations angle φ, is also shown in the form of the probability histogram in the right side of the corresponding image. Each cell of the histogram corresponds to a separate rotation φ. The value within each cell corresponds to the likelihood probability $P(\varphi_{t+1} = r \mid O_1, O_2, ..., O_{t+1})$ as computed in (12).

More specifically, in the initial frames of the sequence (e.g. Fig. 6a) the confidence about the orientation φ is spread among a large number of possible solutions (Fig. 6b). As time passes, the HMM becomes more confident about the actual pose of the human. This is reflected in the histograms of the rotations angle (Fig. 6d, f) which become more confident about the true angle φ.

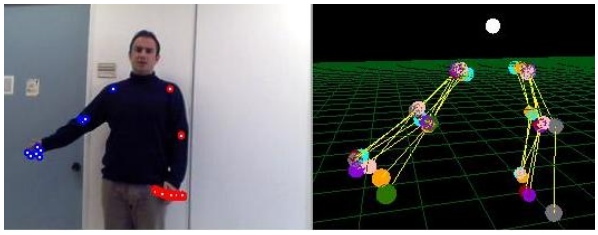

Fig. 4. The actor raises his hand parallel to the camera. Left: Input image with particles projections. Right: Particles with largest weights. The white sphere depicts the head while the colored ones represent the shoulder, the elbows and the hands.

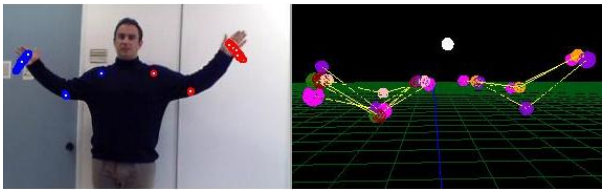

Fig. 5. The actor waves both hands.

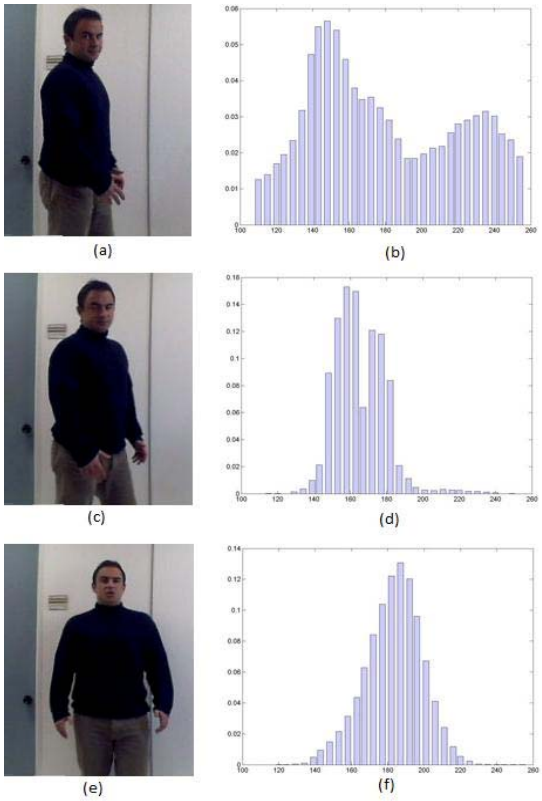

Fig. 6. The actor rotates his torso (left column).The histograms of the right column depict the transition between possible orientation solutions.

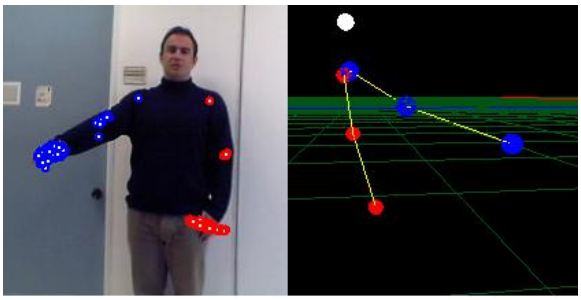

Fig. 7. Arm movement towards the camera. The side view of the hypothesis tracking depicts the success in ambiguous cases.

Finally, in the third example, the actor moves his arm towards the camera (Fig. 7). Since, in this work, the 3D position of the hand is used only implicitly, this case is difficult because of the ambiguities concerning the hand's depth and, consequently, the actual configuration of the arm. Another source of ambiguity is raised from the fact that the apparent image foreground which is used to compute the second term, p_2^m, in the likelihood equation (9) appears as a rigid area, making the discrimination between possible solutions more difficult. Nevertheless, as shown in the side-view of the hypothesized pose in the right image of Fig. 7, the proposed tracker was able to cope successfully with the above described difficulties, qualifying and tracking over time solutions which represent the actual configuration of the actor's arm.

VIII. Discussion and Future Work

In this paper, a probabilistic approach towards visual arm tracking has been presented. In order to cope with the high dimensionality of the problem, the 9 parameter model space is decomposed into three subspaces, one 1D for the body orientation and two 4D subspaces, each representing the joint configuration of a single arm. The first is tracked by the use of a HMM while the latter by the use of two different sets of particle filters. Experimental results have demonstrated the effectiveness of the proposed approach in a variety of cases that include ambiguous situations. As already mentioned, the motivation behind this work is to develop a tracker that will function as the core component of a gesture recognition system, intended for application in a mobile robot that operates in public places. To proceed with the application at hand, the work presented in this paper has to be enhanced in a multitude of ways, each constituting a direction for our future research. More specifically, recognizing that foreground subtraction will not be able to perform adequately on images taken by a moving robot, a first goal is to eliminate the algorithm's dependence on foreground information in favor of more robust information like color or shape. Finally, another direction of future research is towards addressing situations where more than one actor perform on the same scene.

References

[1] Rubin, D., *Using the SIR algorithm to simulate posterior distributions.* Bayesian statistics, 1988. **3**: p. 395-402.

[2] Krueger, M., *Artificial Reality II*. 1991, Addison-Wesley.

[3] Cootes, T. and C. Taylor. *Active shape models: Smart Snakes*. in *British Machine Vision Conference*. 1992.

[4] Krueger, M., *Environmental technology: making the real world virtual.* Communications of the ACM, 1993. **36**(7): p. 36-37.

[5] Rehg, J. and T. Kanade. *DigitEyes: vision-based hand tracking for human-computerinteraction*. in *IEEE Workshop on Motion of Non-Rigid and Articulated Objects*. 1994.

[6] Goncalves, L., et al. *Monocular tracking of the human arm in 3D*. in *International Conference on Computer Vision*. 1995.

[7] Rehg, J. and T. Kanade. *Model-based tracking of self-occluding articulated objects*. in *Intl. Conf. of Computer Vision*. 1995.

[8] Gavrila, D. and L. Davis. *3-D model-based tracking of humans in action: a multi-view approach*. in *Proceedings of IEEE Computer Vision and Pattern Recognition (CVPR)*. 1996.

[9] Pavlovic, V., R. Sharma, and T. Huang. *Gestural interface to a visual computing environment for molecularbiologists*. in *International Conference Automatic Face and Gesture Recognition*. 1996.

[10] Birk, H., T. Moeslund, and C. Madsen. *Real-Time Recognition of Hand Alphabet Gestures Using Principal Component Analysis*. in *Proceedings of Scandinavian Conference on Image Analysis*. 1997.

[11] Utsumi, A. and J. Ohya. *Direct Manipulation Interface using multiple cameras for hand gesture recognition*. in *SIGGRAPH*. 1997.

[12] Shimada, N., et al. *Hand gesture estimation and model refinement using monocularcamera-ambiguity limitation by inequality constraints*. in *IEEE Intl. Conf. on Automatic Face and Gesture Recognition*. 1998.

[13] Utsumi, A. and J. Ohya. *Image segmentation for human tracking using sequential-image-basedhierarchical adaptation*. in *Proceedings of IEEE Computer Vision and Pattern Recognition (CVPR)*. 1998.

[14] Stauffer, C. and W. Grimson. *Adaptive background mixture models for real-time tracking*. in *Computer Vision and Pattern Recognition*. 1999.

[15] Wu, Y. and T. Huang. *Capturing Human Hand Motion: A Divide-and-Conquer Approach*. in *IEEE Intl. Conf. on Computer Vision*. 1999.

[16] MacCormick, J. and A. Blake, *A Probabilistic Exclusion Principle for Tracking Multiple Objects.* International Journal of Computer Vision, 2000. **39**(1): p. 57-71.

[17] MacCormick, J. and M. Isard, *Partitioned Sampling, Articulated Objects, and Interface-Quality Hand Tracking.* Proceedings of European Conference on Computer Vision, 2000: p. 3-19.

[18] Stenger, B., P. Mendonca, and R. Cipolla. *Model-based hand tracking using an unscented kalman filter*. in *British Machine Vision Conference*. 2001.

[19] Wu, Y., J. Lin, and T. Huang. *Capturing natural hand articulation*. in *International Conference on Computer Vision*. 2001.

[20] Shakhnarovich, G., P. Viola, and T. Darrell. *Fast pose estimation with parameter-sensitive hashing*. in *Ninth IEEE International Conference on Computer Vision*. 2003.

[21] Agarwal, A. and B. Triggs. *3D human pose from silhouettes by relevance vector regression*. in *Computer Vision and Pattern Recognition*. 2004.

[22] Argyros, A. and M. Lourakis, *Real-Time Tracking of Multiple Skin-Colored Objects with a Possibly Moving Camera. Proceedings of European Conference on Computer Vision*, 2004: p. 368-379.

[23] Stenger, B., P. Mendonca, and R. Cipolla. *Model-Based 3D Tracking of an Articulated Hand*. in *IEEE Computer Vision and Pattern Recognition* 2004.

[24] Tian, T., R. Li, and S. Sclaroff. *Articulated pose estimation in a learned smooth space of feasible solutions*. in *IEEE Workshop on Learning in CVPR*. 2005.

[25] G. Elert, a.s. *Size of a Human: Body Proportions*. The Physics Factbook 2006 [cited 2009 Tuesday, February 24]; Available from: http://hypertextbook.com/facts/2006/bodyproportions.shtml.

[26] Tsetserukou, D., et al. *Development of a Whole-Sensitive Teleoperated Robot Arm using Torque Sensing Technique*. 2007: IEEE Computer Society Washington, DC, USA.

[27] Baltzakis, H., et al. *Tracking of human hands and faces through probabilistic fusion of multiple visual cues*. in *International Conference on Computer Vision Systems (ICVS)*. May 2008. Santorini, Greece.