

# Fuzzy Qualitative Human Motion Analysis

Chee Seng Chan, *Student Member, IEEE*, and Honghai Liu, *Senior Member, IEEE*

**Abstract**—This paper proposes a fuzzy qualitative approach to vision-based human motion analysis with an emphasis on human motion recognition. It achieves feasible computational cost for human motion recognition by combining fuzzy qualitative robot kinematics with human motion tracking and recognition algorithms. First, a data-quantization process is proposed to relax the computational complexity suffered from visual tracking algorithms. Second, a novel human motion representation, i.e., qualitative normalized template, is developed in terms of the fuzzy qualitative robot kinematics framework to effectively represent human motion. The human skeleton is modeled as a complex kinematic chain, and its motion is represented by a series of such models in terms of time. Finally, experiment results are provided to demonstrate the effectiveness of the proposed method. An empirical comparison with conventional hidden Markov model (HMM) and fuzzy HMM (FHMM) shows that the proposed approach consistently outperforms both HMMs in human motion recognition.

**Index Terms**—Fuzzy qualitative reasoning, human motion analysis, image processing, intelligent robotics.

## I. INTRODUCTION

MOTION understanding is the ability to analyze human motion patterns, and, further, to produce high-level interpretations of these patterns. Human motion in-depth understanding plays a crucial role in a diverse spectrum of applications from surveillance-based suspicious-behavior recognition to monitoring of daily health care for elderly people. Fundamentally, human motion analysis systems consist of description and recognition of human motion: first, *extracting* relevant information through visual tracking that involves the detection of regions of interest in image sequences that are changing with respect to time and/or finding frame to frame correspondence of each region so that features of each region can reliably be extracted and second *modeling* this information as an abstraction of sensory data that should reflect a real-world situation. Finally, a *recognition* step aiming at determining the maximum similarity between an unobserved test sequence and prelearned motion models. Nevertheless, developing these algorithms is an immense challenge as it is a problem that combines the uncertainty associated with computational vision and the added whimsy of human behavior. For instance, the methodology of implementing such a system has been the focus of research in the past two decades, and many works have been conducted and published with a promising classification rate. Surprisingly,

Manuscript received August 9, 2008; revised December 3, 2008; accepted February 5, 2009. First published March 4, 2009; current version published July 29, 2009. This work was supported in part by the Royal Society under Grant 2008/R1 and Grant 2007/R2 and in part by the U.K. Engineering and Physical Science Research Council under Grant EP/G041377/1.

The authors are with the Institute of Industrial Research, University of Portsmouth, Portsmouth PO1 3QL, U.K. (e-mail: cheeseng.chan@port.ac.uk; honghai.liu@port.ac.uk).

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TFUZZ.2009.2016553

such a promising system has yet to be employed in practice. The underlying issue is that these existing methodologies are not computationally feasible, i.e., the complexity of such systems in terms of the processing time is too expensive to implement into real systems, mainly due to the tracking precision.

In visual tracking, particle filters [10] and its variants [11], [35] have been well developed in human-motion tracking research over the past decades, but its computational cost keeps them from practical applications. It can be seen that a standard particle filter has a computational complexity of  $O(2N)$  and time complexity of  $N \sum_{k=1}^m \tau_k$ , where  $N$  is the number of particles, and  $\tau_k$  is the cost of calculation  $p(z_k|x)$ . This is a major drawback as practical human motion analysis systems are required to run in real-time, or near real-time frame rates, whereas methods that do not use such approaches usually rely on the accuracy of motion sensors but seldom provide a measure of confidence of the results that are crucial to discriminate similar events in a noisy environment. On the other hand, representation of human motion is a very important and sometimes difficult aspect of an intelligent system. The representation is an abstraction of sensory data that should reflect a real-world situation and be compact and reliable. Probabilistic graph models have been the dominant methods in the field of human motion analysis systems [7]. Despite the fact that all these approaches have demonstrated success in modeling and recognizing complex activities, there is, however, a tendency to use the parameterization as a “black box,” i.e., these approaches highly depend on probabilities and intensive training to recognize all the activities. Thus, one needs to have a large number of training sequences with intensive training in order for each activity to be recognized correctly. This requires increasingly expensive computational power in that the complexities of these solutions in terms of processing time are proportional to the size of the training data.

Template matching is one of the earliest human motion analysis methods. Bobick and Davis [3] proposed a view-based approach to the representation and recognition of action using temporal templates. They made use of the binary motion energy image (MEI) and motion history image (MHI) to interpret human movement in an image sequence. First, motion images in a sequence were extracted by differencing, and these motion images were accumulated in time to form MEI. Then, the MEI was enhanced into MHI that was a scalar-valued image. Taken together, the MEI and MHI were considered to be a two-component version of a temporal template, i.e., a vector-valued image, in which each component of each pixel was some function of the motion at that pixel position. Finally, by representing the templates by its seven Hu-moments, a Mahalanobis distance was employed to classify the action of the subject by comparing it with the Hu-moments of prerecorded actions. Bradski and Davis [5] further contributed to the idea of MHI by proposing timed MHI (tMHI) for motion segmentation. tMHI allows for

determination of the normal optical flow. Motion is segmented relative to object boundaries and the motion orientation. Hu-moments are applied to the binary silhouette to recognize the pose. However, one of the main problems of template matching approaches is that the recognition rate of objects based on 2-D image features is low, due to the nonlinear distortion during perspective projection and the image variations with the view-point's movement. These algorithms are generally unable to recover the 3-D pose of objects. Moreover, the stability of dealing effectively with occlusion, overlapping, and interference of un-related structures is generally poor. Alternatively, model-based approaches have been adopted [8], [9], [14], [15] in which a human body model is constructed with prior knowledge. Wang *et al.* [34] presented an approach where contours are extracted and a mean contour is computed to represent the static contour information. Dynamic information is extracted by using a detailed model composed of 14 rigid body parts, each one represented by a truncated cone. Particle filtering [10] is used to compute the likelihood of a pose given an input image. However, there is a tradeoff between tracking precision against computational cost where a number of systems are based on incremental updates or searching around a predicted value [10], [31], [32].

Recently, hidden Markov models (HMMs), which are a family of popular parametric methods, have been formulated in human motion analysis systems. Common approaches consist of extracting low-level features by local spatiotemporal filtering on images and using the HMM on the collection of sequences of points to achieve activity recognition and classification tasks. The HMM considers this correlation between adjacent time instances by formulating a Markov process and assumes that the observation sequence is statistically determined by a hidden process that is composed of a fixed number of hidden states. For instance, Lv and Nevatia [21] decomposed the large joint space into a set feature space where each feature corresponds to a single joint or combination of related joints. An AdaBoost scheme was employed to detect and recognize each feature in the feature space, followed by an HMM system to recognize each action class based on the detected features. Leo *et al.* [16] attempted to classify actions at an archaeological site. A system that uses binary patches and an unsupervised clustering algorithm to detect human body postures was proposed. A discrete HMM is used to classify the sequences of poses into a set of four different actions. A mixed state statistical model for the representation of motion had been proposed in [6], i.e., the work decomposes a human behavior into multiple abstractions and represents the high-level abstraction by HMM built from phases of simple movements. Estimation and recognition of human behavior is performed with expectation-maximization approaches using particle filters [10] or structured variation inference techniques [26]. While all these solutions have demonstrated success in classifying complex activities, HMMs suffer from few drawbacks. First of all, an HMM relies on stochastic learning; they require extensive training. Therefore, one needs to have a large number of training sequences for each activity to be recognized correctly. This is not feasible as practical human-motion analysis systems must often work in real-time or near real-time frame rates. Second, for each activity to be recognized,

a separate HMM needs to be built. Hence, a solution that can flexibly handle the tradeoff between human motion precision and its computation efficiency are the next important step in a human-motion analysis system.

In this paper, we propose a fuzzy qualitative method to study human-motion analysis in order to reduce the computational complexity encountered by existing solutions, enabling such systems to be used in practical situations. First of all, we propose a solution to handle the tradeoff between computational efficiency and motion description precision in the visual tracking algorithm by applying a data-quantization process. During this process, we consider the rigid motion of each human body joint in a fuzzy qualitative description. Second, a novel human-motion representation, known as qualitative normalized template (QNT), is developed in terms of the fuzzy qualitative robot kinematics framework [19] to effectively represent human motion. The QNT is a template-based method instead of a statistical learning method; hence, large training datasets are not required. Instead, strong discriminative features can be derived from just a couple of example activities. Finally, empirical results show that our proposed solution outperforms existing solutions in human motion recognition by flexibly handling the tradeoff between human-motion precision and its computational efficiency.

The remainder of the paper is structured as follows. Section II derives the fuzzy qualitative human motion analysis, in particular, how both the data quantization and fuzzy qualitative robot kinematics framework are employed to study human motion in video sequences. Section III presents the experimental results and an empirical comparison with the HMM and FHMM in human-motion classification. Section IV concludes the paper with discussions and future works.

## II. FUZZY QUALITATIVE HUMAN-MOTION ANALYSIS

This section presents fuzzy qualitative description for human-motion analysis in video sequences. The culmination of algorithms is represented by a distributed video collection and processing system in which the basic tasks of human body modeling, human-motion tracking, representation, and recognition are performed in support of a single, underlying task, which is human-motion analysis.

### A. Human Body Modeling

In principle, one must perceive a human motion before modeling and interpreting it, which means that an appearance model is needed. In this paper, we simplify a human body into a collection of hierarchical structure skeleton composed of segments and limbs linked together in a kinematic chain. The model parameters are given by  $\phi_t = [(\phi_t^{pos})^T, (\phi_t^{vel})^T]^T = [t_t^T, \alpha_t^T, \theta_t^T, \dot{t}_t^T, \hat{\alpha}_t^T, \hat{\theta}_t^T]^T$ , where  $t_t$  and  $\alpha_t$  represent the translation and rotation that map the body into the world coordinates system, and  $\theta_t$  represents the relative angles between all pairs of connected limbs. Parameters  $\dot{t}_t^T$ ,  $\hat{\alpha}_t^T$ , and  $\hat{\theta}_t^T$  represent the corresponding velocities. In addition, an event or action is defined as the temporal movement of a human body segment in a short time period and is represented by a state representation



Fig. 1. Sample human motions used from the database provided by [2] and [29]. Trajectories from six landmarks (shoulder, elbow, wrist, hip, knee, and ankle) on the human body are tracked over time using the condensation algorithm.

in a normalized fuzzy qualitative unit circle. Further, an activity is defined as a combination-ordered sequence of the entire participating body segment’s movement over time, which is restricted by the motion constraints of the human body. For instance, the sequences of events for a walking activity include segment events for the foot, lower leg, and thigh and joint events for the ankle, knee, and hip. All these sequences occur in the leg that is moving forward, while the leg that supports the body will show no such events, and the walking activity is defined as a combination of these sequences of events.

### B. Human-Motion Tracking

The condensation algorithm [10] has been employed to estimate the posterior probability distribution over human pose, given a sequence of observation measurement. It is assumed that all elements  $\phi_{i,t} \in \phi_t$  are independent, and all parameters were initialized manually with a Gaussian distribution at time  $t = 0$ . In order to restrict the set of admissible motions and reduce the ambiguities in the estimation, each joint is allowed only for one single DOF, which is a rotation around its axis. This constraint is justified by the fact that typically, the motion of the limbs can be approximated as planar around an axis perpendicular to the direction of the motion. Fig. 1 shows examples of sets of trajectories from different limb motions in the case of real videos. The condensation algorithm is one of dominant tracking algorithms; for detailed human-motion tracking algorithms, see [1], [22], and [33].

The condensation algorithm discussed herein has the advantage that the particle representation can represent distributions that are difficult to model analytically. The performances of such algorithms have not, however, been fully evaluated under circumstances specifically to real-time vision systems, where there exists a certain tradeoff between motion description precision and computational efficiency. For instance, the condensation algorithm is an approximation technique by representing the posterior density as a set of samples of the state space with as-

sociated likelihood weights  $\omega_t^i, i \in \{1, \dots, N\}$ . The sample set approximation of the current posterior density  $p(X_t|O_{1:t})$  can be obtained via

$$p(X_t|O_{1:t}) \approx \sum_{i=1}^N \omega_t^i \delta(X_t - x_t^i) \quad (1)$$

where  $X_t$  denotes the multivariate state at time  $t$ ,  $O_{1:t}$  denotes the sequence of observation measurements within the time range  $[1, t]$ , and  $\delta(X_t - x_t^i)$  denotes the Dirac delta function; the prior is approximated as

$$p(X_t|O_{1:t-1}) \approx \sum_{i=1}^N \omega_{t-1}^i \delta(X_t - x_{t-1}^i). \quad (2)$$

Weight  $\omega_{t-1}^i$  is determined such that  $\omega_{t-1}^i \propto p(O_{t-1}|X_{t-1}^i)$ ,  $\sum_{i=1}^N \omega_{t-1}^i = 1$ .

On examining (1) and (2), both equations are almost accurate if one employs a sufficient large number of particles. In other words, if the number of particles is infinite, i.e.,  $N \rightarrow \infty$ , the right-hand sides of (1) and (2) become identical to the left-hand sides. In reality, however, using an infinite number of particles is not practical, especially for real-time processing. Nevertheless, if relatively low numbers of particles are employed, the tracking system will fail and, hence, greatly affect human-motion analysis systems during the representation and recognition stages.

### C. Data Quantization

Data quantization is a process adopted from Liu and Coghill [20], in which the unit circle of the conventional trigonometry has been fuzzified by the introduction of fuzzy qualitative quantity spaces for its orientation and translation components; the data quantization is a process that mapped quantitative information qualitatively with respect to the configuration in the fuzzy qualitative unit circle. The fuzzy qualitative quantity space is a set of overlapped fuzzy numbers whose individual distance among them is defined by a predefined metric. Besides, each fuzzy number is a finite and convex discretization of the real number line by default. Four tuple fuzzy numbers (i.e.,  $[a, b, \alpha, \beta]$ ) and its arithmetic are employed to describe the characteristic of each state in the fuzzy qualitative unit circle. Such a representation has been selected as it provides good compositionality and high resolution [30].

In this paper, we employed the data-quantization process to represent the predicted motion parameters  $\phi_t$  qualitatively. The advantage of the data-quantization process is that each of the fuzzy qualitative quantity space in the fuzzy qualitative unit circle that is a finite and convex discretization of the real number line will be able to model the tracking error when relatively low numbers of particles are selected to perform the visual tracking algorithm, i.e., we consider the motion of each joint as a collection of time series describing the joint angles as they evolve over time. This is achieved through the visual tracking algorithm discussed in Section II-B. However, as shown in Fig. 2, it is evident that about 1000 particles are needed in the condensation algorithm, with approximately 60 min of training time

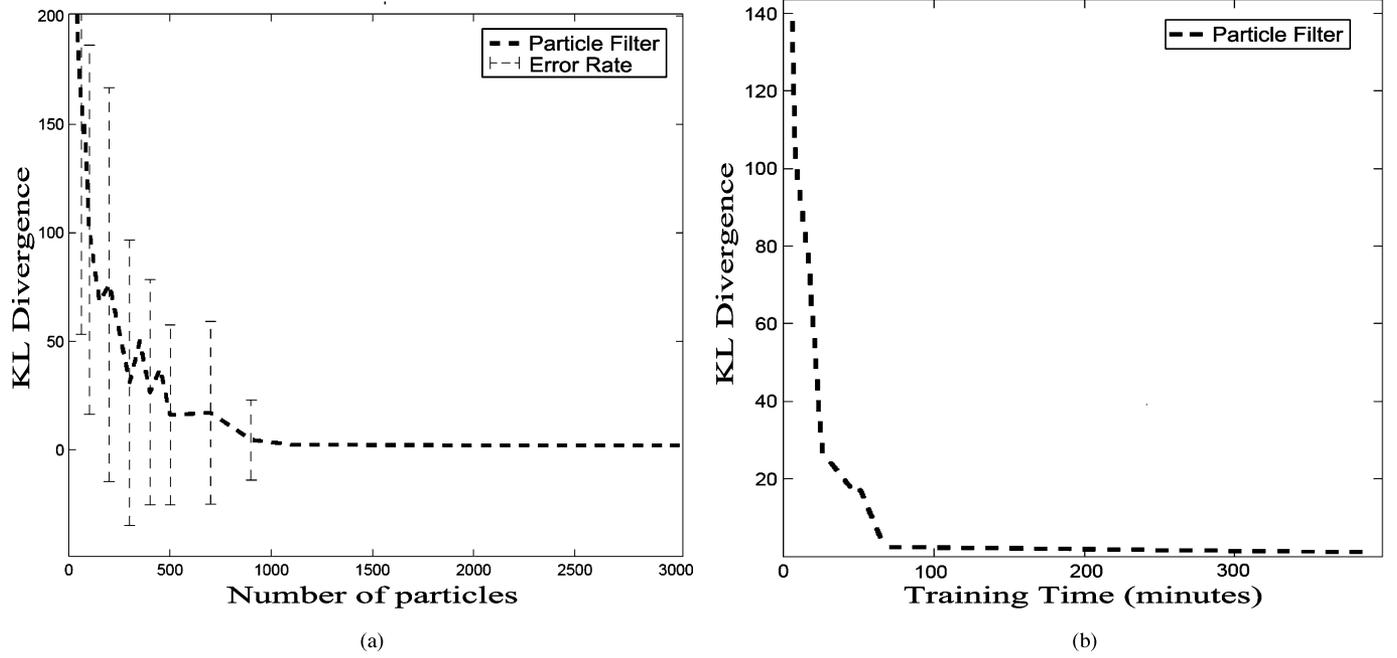


Fig. 2. Number of particles versus training time. (a) Referring to the error rate result, it is evident that about 1000 of particles are needed in order to achieve the required tracking precision. (b) Training time of the condensation algorithm employing 1000 particles requires about 60 min.

required in order to achieve the required tracking precision. This is not feasible for a real-time vision-based human-motion system.

The rigid motion of each joint is mapped into a fuzzy qualitative description through the data-quantization process so that each joint movement is represented by the fuzzy qualitative orientation states and fuzzy qualitative translation states in which the quantitative dynamic characteristic of motion data resides in the normalized fuzzy qualitative unit circle. A fuzzy qualitative unit circle is constructed using (3), shown below, where the orientation and translation components in the conventional unit circle are being replaced by fuzzy qualitative quantity spaces

$$\begin{aligned} \lim_{s \rightarrow s_o=12} C_t(s) &= \text{QS}(qp_t) \\ \lim_{r \rightarrow r_o=16} C_o(r) &= \text{QS}(qp_\theta) \end{aligned} \quad (3)$$

where  $s$  is the number of states resides in the  $x$ - $y$  translation, while  $r$  is the number of states resides on the orientation in the fuzzy qualitative unit circle, i.e.,  $s$  and  $r$  represent the number of translation and orientation states employed in the quantity spaces to represent the fuzzy qualitative unit circle, respectively. As  $s \rightarrow s_o$  and  $r \rightarrow r_o$ , the limits of  $C_t(s)$  and  $C_o(r)$  will approach to a set of  $s_o$  qualitative states for a translation component and a set of  $r_o$  qualitative states for an orientation component. The range of  $s$  and  $r$  are application-dependent. Empirically, we selected the translation  $s$  and orientation  $r$  as  $s = 12$  and  $r = 16$  four-tuple fuzzy numbers, respectively, as shown in Fig. 3.

The fuzzy qualitative quantity space  $Q$  of the fuzzy qualitative unit circle consists of an orientation component  $Q^a$  and a

translation component  $Q^d$ , and it can be described as follows:

$$\begin{aligned} Q^a &= \{\text{QS}_a(\theta_i)\}, \quad \text{where } i = 1, 2, \dots, m \\ Q^d &= \{\text{QS}_d(l_j)\}, \quad \text{where } j = 1, 2, \dots, n \end{aligned} \quad (4)$$

where  $\text{QS}_a(\theta_i)$  denotes the state of an angle  $\theta_i$ ,  $\text{QS}_d(l_j)$  denotes the state of a distance  $l_j$ , and  $m$  and  $n$  are the number of the elements of the two components. The position measurement of  $P(\text{QS}_a(\theta_i), \text{QS}_d(l_j))$  determined by both the characteristics of the fuzzy membership functions of  $\text{QS}_a(\theta_i)$  and  $\text{QS}_d(l_j)$ . The geometric meaning of fuzzy qualitative trigonometry is demonstrated in a proposed fuzzy qualitative unit circle, in which the motion is described by an orientation component and a translation component. The level of resolution in the fuzzy qualitative unit circle can be adjusted by setting the fuzzy number.

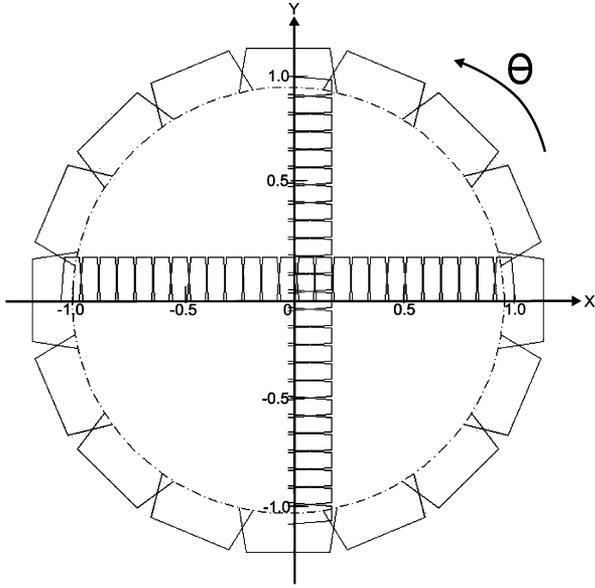
Further, with respect to the constructed fuzzy qualitative unit circle, the predicted motion parameters  $\phi_t$  obtained from Section II-B are mapped into its corresponding fuzzy qualitative states, as given in

$$\begin{cases} qp_t^i | qp_t^i \in [0, l_{i1}, l_{i2}, \dots, l_{i(r_i-1)}, l_{ir_i}] \\ qp_\theta^i | qp_\theta^i \in [0, q\theta_{i1}, q\theta_{i2}, \dots, q\theta_{i(s_i-1)}, 2\pi] \end{cases} \quad (5)$$

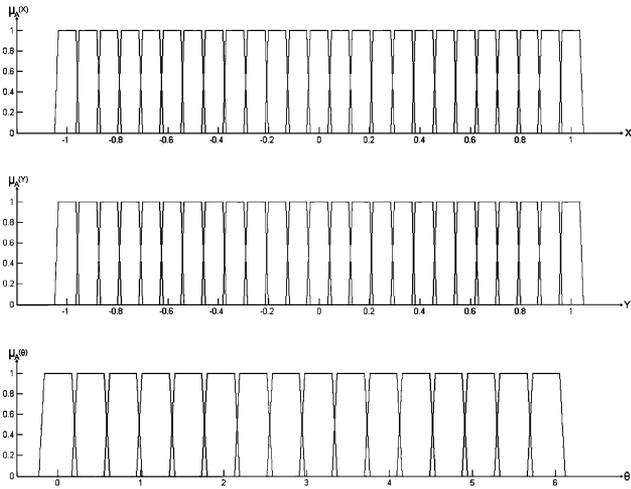
where

$$\begin{aligned} qp_j^i &= \frac{l_i j}{s_i}, & q\theta_k^i &= \frac{2\pi k}{r_i} \\ 0 &\leq qp_1^i \leq qp_2^i \leq \dots \leq qp_{(s_i-1)}^i \leq l_i \\ 0 &\leq q\theta_1^i \leq q\theta_2^i \leq \dots \leq q\theta_{(r_i-1)}^i \leq 2\pi. \end{aligned}$$

Each of these states is the corresponding region in which the quantitative dynamic characteristic of motion data resides in the fuzzy qualitative unit circle. Due to the fact that a motion component comprises evenly distributed normalized numeric data,



(a)



(b)

Fig. 3. Fuzzy qualitative unit circle with resolution  $s = 12$  and  $r = 16$ , respectively. (a) Description of the Cartesian translation and orientation in the conventional unit circle are replaced by a quantity space. (b) Element of the quantity space for every variable in the fuzzy qualitative unit circle is a finite and convex discretization of the real number line.

fuzzy numbers have the same shapes by default. For simplicity, we have  $b - a = \kappa_0 \alpha$  and  $\alpha = \beta$ , and the membership value of the crossing point of adjacent fuzzy numbers is 0.5 by default. Hence, a function for generating a quantity space including  $n$  normalized qualitative states can be obtained shortly, where  $[a_i, b_i, \alpha_i, \beta_i]$  denotes the  $i$ th fuzzy qualitative state  $QS(i)$

$$QS(i) = \begin{cases} [0, \kappa_0 \alpha, 0, \beta], & i = 1 \\ [(\kappa_0 + 1)(i - 1)\alpha, (\kappa_0 + 1)i\alpha - \alpha, 0, \beta] & i = 2, \dots, n - 1 \\ [1 - \kappa_0 \alpha, 1, \alpha, 0], & i = n \end{cases} \quad (6)$$

TABLE I  
QUANTIZATION RESULTS WITH  $s = 12$  AND  $r = 16$

| Joints       | Before Quantisation (°) | After Quantisation (State) |
|--------------|-------------------------|----------------------------|
| $J_{1,t=1}$  | 40                      | $Q^a(2)$                   |
| $J_{2,t=1}$  | 48                      | $Q^a(2)$                   |
| $J_{3,t=1}$  | 46                      | $Q^a(2)$                   |
| $\vdots$     | $\vdots$                | $\vdots$                   |
| $J_{1,t=13}$ | 65                      | $Q^a(3)$                   |
| $J_{2,t=13}$ | 18                      | $Q^a(1)$                   |
| $J_{3,t=13}$ | 120                     | $Q^a(6)$                   |
| $\vdots$     | $\vdots$                | $\vdots$                   |
| $J_{1,t=24}$ | 46                      | $Q^a(2)$                   |
| $J_{2,t=24}$ | 46                      | $Q^a(2)$                   |
| $J_{3,t=24}$ | 50                      | $Q^a(2)$                   |

where  $\alpha = [1/(n(\kappa_0 + 1) - 1)]\kappa_0$  is a threshold parameter to define the shape of the fuzzy numbers, and  $\kappa_0$  and  $n$  are chosen by applications or by a learning algorithm.

For each of the fuzzy qualitative state representation of the motion parameters, we normalized them within the fuzzy qualitative unit circle  $[-1 \ 1]$  using (7). Therefore, the qualitative and quantitative representations are linked together, which paves the way to connect numerical image sequences with symbolic natural language description

$$\begin{cases} QS(qp_l) = qp_l | qp_l \in \left[ \frac{ql_1}{ql}, \frac{ql_2}{ql}, \dots, \frac{ql_{s-1}}{ql}, 1 \right] \\ QS(qp_\theta) = qp_\theta | qp_\theta \in \left[ \frac{q\theta_1}{2\pi}, \frac{q\theta_2}{2\pi}, \dots, \frac{q\theta_{r-1}}{2\pi}, 1 \right] \end{cases} \quad (7)$$

where  $x$ - $y$  translation states  $qp_l$  are normalized by the average length of the human body segment  $ql$ , and the orientation states  $qp_\theta$  are normalized to  $2\pi$ . Hence, (5) can be rewritten as

$$\begin{cases} qp_l^i | qp_l^i \in \left[ \frac{qp_1^i}{\sum_{i=1}^n l_i}, \frac{qp_2^i}{\sum_{i=1}^n l_i}, \dots, \frac{qp_{(s_i-1)}^i}{\sum_{i=1}^n l_i}, \frac{l_i}{\sum_{i=1}^n l_i} \right] \\ qp_\theta^i | qp_\theta^i \in \left[ \frac{q\theta_1^i}{2\pi}, \frac{q\theta_2^i}{2\pi}, \dots, \frac{q\theta_{(r_i-1)}^i}{2\pi}, 1 \right]. \end{cases} \quad (8)$$

The visual tracking parameters of each body joint are mapped into the normalized fuzzy qualitative unit circle in order to achieve the fuzzy qualitative description. It can be seen that each motion parameter  $\phi_t$  is represented by the corresponding state region of a fuzzy qualitative state in the normalized unit circle. For example, the predicted joint motion parameters of  $J_1$ ,  $J_2$ , and  $J_3$  at time instant  $t = 1$  are represented by the same fuzzy qualitative states after the data-quantization process, as shown in Table I. In other words, if we have a sufficient number of particles to meet the resolution needed in the normalized fuzzy qualitative unit circle, the data-quantization process will be able to model the uncertainty under the constraints of the limitation in visual tracking algorithms and to correctly preserve the underlying motion description for motion representation and recognition process.

In order to quantify the accuracy of the estimated motion parameter  $\phi_t$  before and after the proposed data-quantization

TABLE II  
COMPARISON OF THE ESTIMATED MOTION PARAMETERS WITH THE GROUND TRUTH POSITIONS BEFORE AND AFTER DATA-QUANTIZATION PROCESS

| Particles<br>( $N$ ) | Joint | Error Rate             |                       | Computational<br>Complexity |
|----------------------|-------|------------------------|-----------------------|-----------------------------|
|                      |       | Before<br>Quantisation | After<br>Quantisation |                             |
| 800                  | Thigh | 6.5%                   | 0%                    | $\mathcal{O}(2N)$           |
|                      | Knee  | 15.6%                  | 0%                    |                             |
| 2000                 | Thigh | 5.1%                   | 0%                    | $\mathcal{O}(2N)$           |
|                      | Knee  | 7.5%                   | 0%                    |                             |

process, Table II shows a comparison of the estimated motion parameter  $\phi_t$  obtained from the condensation algorithm with  $N = 800$  and  $N = 2000$  to its ground truth position before and after the proposed data-quantization process. It can be seen that the accuracy of the estimated motion parameters  $\phi_t$  after the proposed data quantization process is typically much higher. Also, it can be noted that the error rates of both the condensation algorithms ( $N = 800$  and  $N = 2000$ ) are similar. This demonstrates the effectiveness of the proposed data-quantization process in terms of flexibly handling the tradeoff between human-motion precision and its computational efficiency. This is an essential step in order to move the human-motion analysis systems into practical use.

#### D. Human-Motion Representation

The data-quantization process is addressed in previous sections with a focus on representing the motion of individual joints of the human body qualitatively in order to relax the computational complexity in visual tracking algorithms. It is evident that human motion is a combination of ordered sequences of all the independent movements [4]. In order to build a reliable human motion template that reflects the real-world situation, the fuzzy qualitative robot kinematics framework has been employed to construct a novel motion template called a QNT. We model a human skeleton structure as articulated rigid bodies with kinematic chains. It is assumed that the human motion is rigid so that the priority is given to solely on the joint movements, and the work will not be distracted by effects of muscle stretch and reflex during a movement.

Based on fuzzy qualitative robot kinematics [19], the motion of each body joint  $i$  is represented in terms of twist representation, as shown in (9), since it provides a simpler solution and leads to a compact 3-D linear representation of a motion model [25]

$$\xi = [v_1 \ v_2 \ v_3 \ \omega_1 \ \omega_2 \ \omega_3] \quad (9)$$

where  $\xi$  is a 3-D fuzzy qualitative unit vector that points in the direction ranges of the rotation axis. The amount of rotation is specific with a fuzzy qualitative angle state  $\theta$  multiplied by the twist  $\xi\theta$ , whereas the  $v$  component determines the location range of the rotation axis and the amount of translation along this axis. For instance, in order to realize the motion performed by each human body at time  $t$ , first, we define a base body reference frame  $F_0$  that is attached to the base body and a spatial

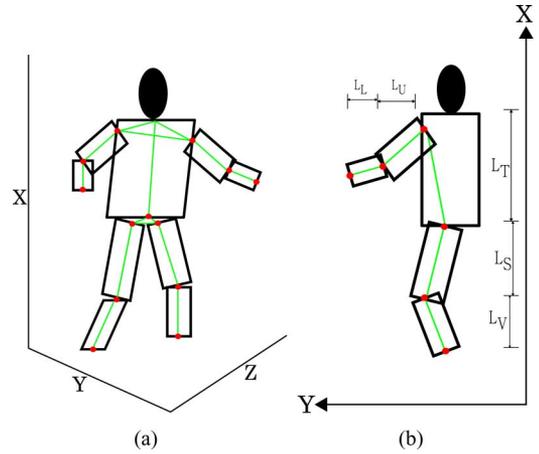


Fig. 4. Proposed human model where each segment (i.e., limb) of a human is represented by patches connected by joints. (a) 3-D human body. (b) 2-D human body.

reference frame  $F_a$ , which is static and coincides with  $F_0$  at time  $t$ . By considering a single kinematics chain of two human body segments connected to the base frame, we parameterize the orientation between these connected components in terms of the angle of rotation around the axis of the object coordinate frame  $\theta$ . This rotation axis in the object frame can be represented by a 3-D fuzzy qualitative unit vector  $\omega_1$  along the axis and a fuzzy qualitative point  $q_1$  on the axis. The twist representation can be described as follows for revolute joint 1

$$\xi_1 = \begin{bmatrix} -\omega_1 \times q_1 \\ \omega_1 \end{bmatrix}. \quad (10)$$

The transformation of fuzzy qualitative point  $q_1$  from  $F_a$  coordinates to the base frame  $F_0$  can be obtained as

$$g(\theta_1) = g_1 g(0) = e^{\xi_1 \theta_1} g(0). \quad (11)$$

For a kinematics chain of  $K$  bodies, its motion of the  $k$ th body is represented by joint  $\theta_k$ , and each joint is described by a twist  $\xi_k$ . The forward kinematics  $g_K(\theta_1, \theta_2, \dots, \theta_k)$  therefore can be computed by the individual twist motion for each joint  $e^{\xi_k \theta_k}$ , and the transformation between the base frame  $g(0)$  and Frame  $F_k$  can be obtained as

$$g(\theta_1, \theta_2, \dots, \theta_k) = e^{\xi_1 \theta_1 + \xi_2 \theta_2 + \dots + \xi_k \theta_k} g(0). \quad (12)$$

As described, typical human motion is a combination of ordered sequences of all the independent movements performed by each of the human body joint and can be considered as a function with respect to time. Thus, for a continuous representation in time period  $T$ , where  $T \in (1, \dots, m)$ , the compact version of the transformation can be written as

$$g_K(\theta_1, \theta_2, \dots, \theta_k) = [e^{\xi_1 \theta_1 + \xi_2 \theta_2 + \dots + \xi_k \theta_k} g(0)]_{1 \times m}.$$

All the performed activities captured in the video data are front-to-parallel with the camera plane in this paper; all the joints therefore have an axis orientation parallel to the  $Z$ -axis on the camera plane. Therefore, only half of the human model is employed to construct the QNT, as shown in Fig. 4.

The base body reference frame  $F_0$  is located at the hip, and by employing the normalized qualitative representation and the concept of fuzzy qualitative robot kinematics [18], the product of exponential maps for the arm kinematics chains with respect to base frame  $g(0)$  over time T can be obtained as

$$g_{\text{arm}}(\text{QS}(\theta_1, \theta_2, \theta_3)) = [e^{\xi_1 \theta_1 + \xi_2 \theta_2 + \xi_3 \theta_3} g_{\text{arm}}(0)]_{1 \times m}$$

where

$$g_{\text{arm}}(0) = \begin{bmatrix} \mathbf{I} & \begin{bmatrix} \mathbf{L}_T \\ \mathbf{L}_U + \mathbf{L}_L \\ 0 \\ 1 \end{bmatrix} \\ \mathbf{0} & \end{bmatrix} \quad (13a)$$

$$\omega_{\text{arm}} = [\omega_1 \quad \omega_2 \quad \omega_3] = \begin{bmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{1} & \mathbf{1} & \mathbf{1} \end{bmatrix} \quad (13b)$$

$$q_{\text{arm}} = [q_1 \quad q_2 \quad q_3] = \begin{bmatrix} \mathbf{0} & \mathbf{L}_T & \mathbf{L}_T \\ \mathbf{0} & \mathbf{0} & \mathbf{L}_U \\ \mathbf{0} & \mathbf{0} & \mathbf{0} \end{bmatrix} \quad (13c)$$

$$\xi_{\text{arm}} = [\xi_1 \quad \xi_2 \quad \xi_3] = \begin{bmatrix} \mathbf{0} & \mathbf{0} & -\mathbf{L}_U \\ \mathbf{0} & -\mathbf{L}_T & -\mathbf{L}_T \\ \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{1} & \mathbf{1} & \mathbf{1} \end{bmatrix} \quad (13d)$$

where

$$\mathbf{0} = [0 \quad 0 \quad 0 \quad 0]; \quad \mathbf{1} = [1 \quad 1 \quad 0 \quad 0].$$

The product of exponential mapping for leg kinematics chains with respect to the same base frame over T is given as

$$g_{\text{leg}}(\text{QS}(\theta_4, \theta_5)) = [e^{\xi_4 \theta_4 + \xi_5 \theta_5} g_{\text{leg}}(0)]_{1 \times m}$$

$$g_{\text{leg}}(0) = \begin{bmatrix} \mathbf{I} & \begin{bmatrix} -\mathbf{L}_S - \mathbf{L}_V \\ \mathbf{0} \\ \mathbf{0} \\ 1 \end{bmatrix} \\ \mathbf{0} & \end{bmatrix} \quad (14a)$$

$$\omega_{\text{leg}} = [\omega_4 \quad \omega_5] = \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \\ \mathbf{1} & \mathbf{1} \end{bmatrix} \quad (14b)$$

$$q_{\text{leg}} = [q_4 \quad q_5] = \begin{bmatrix} \mathbf{0} & -\mathbf{L}_S \\ \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \quad (14c)$$

$$\xi_{\text{leg}} = [\xi_4 \quad \xi_5] = \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & -\mathbf{L}_S \\ \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \\ \mathbf{1} & \mathbf{1} \end{bmatrix}. \quad (14d)$$

An activity is defined as a combination of ordered sequences of all the independent movements in terms of the human body

segments [4], [12]; hence, for any given activity, its corresponding QNT is derived as

$$\text{QNT} = g_{\text{arm}} \oplus g_{\text{leg}}. \quad (15)$$

It is evident that the QNT can be represented by strong discriminative features derived from small set of example activities, as it is the advantage in terms of real-time performance over a statistical method that requires large training data. The proposed QNT method is a supervised learning method, and its training phase algorithm is provided in Algorithm 1.

---

**Algorithm 1** QNT BASED HUMAN MOTION RECOGNITION - TRAINING PHASE

---

**Require:** Define a human motion skeleton and generate symbolic QNTs as in equation 15

**Ensure:**  $t_{\text{tracking}} < t_{ct}$  {Tracking computational time threshold  $t_{ct}$ }

**Ensure:**  $t_{\text{matching}} < t_{tm}$  {Template match computational time threshold  $t_{tm}$ }

SET  $m_{\text{tracking}}$  {Set particle number for the condensation algorithm}

$\phi_t = [(\phi_t^{\text{pos}})^T, (\phi_t^{\text{vel}})^T]^T \Leftarrow$  CONDENSATION TRACKING

SET  $FQT_O, FQT_T$  {Set precision parameters for the fuzzy qualitative unit circle.}

$\hat{\phi}_t \Leftarrow$  DATA QUANTISATION {Project and normalize  $\phi_t$  as equation 8}

TEMPLATE MATCHING {Compare  $\hat{\phi}_t$  with the symbolic QNTs}

**return**  $m_{\text{tracking}}, FQT_O$  and  $FQT_T$

---

### III. EXPERIMENTS

In this section, we present the performance of the proposed approach under different conditions such as tracking errors, size of training data, the choice of training data, and comparisons with the HMM (i.e., HMM) and fuzzy HMM (FHMM).

#### A. Datasets and Preprocessing

We conducted experiments on two public databases: the KTH database [29] and Weizmann database [2]. Sample images from all the datasets are shown in Figs. 5 and 6, in which some activities are somewhat similar in the sense that limbs have similar motion paths; this high degree of similarity makes the discrimination more challenging. In addition, all the actors have different physical characteristics and perform activities differently in both motion styles and speeds.

Three datasets were created for the validation purpose. First, we have dataset S1: 225 video streams of three human motions in three planar view scenarios from each of the 25 subjects were selected from the KTH database. The selected activities are walking, running, and jogging. The aim here is to evaluate the efficacy of the QNT in that walking, running, and jogging are motions that exhibit similar movements but are dramatically different in motion meaning. Second, we have dataset S2: 55 video streams from the six human motions are employed from the Weizmann database. The selected activities are bending, walking, jacking, jumping, one-hand waving (wave1), and two hands waving (wave2). The objective is

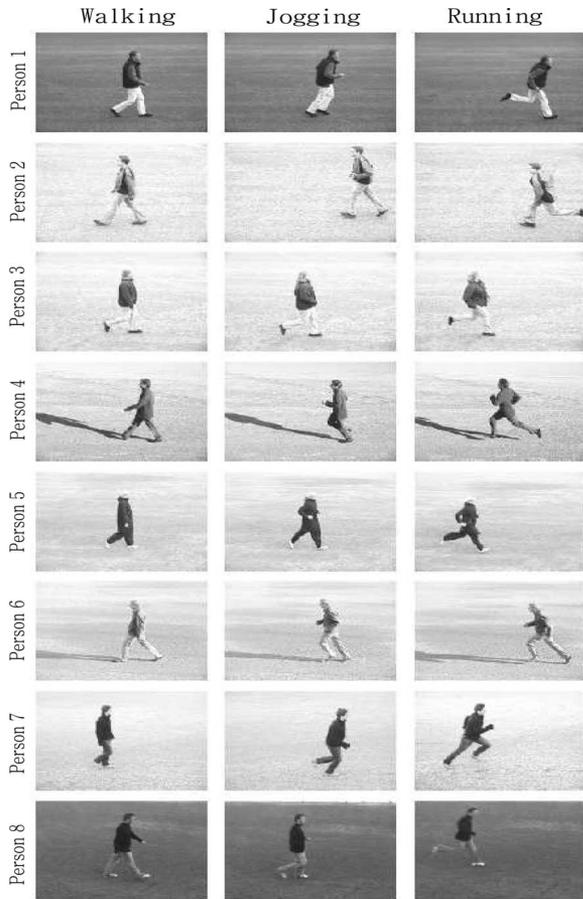


Fig. 5. Example sequences from dataset with four human activities performed by eight different people. Each activity in the dataset is repeated four times by each person, where for walking, jogging, and running, each activity is repeated twice in each directions, i.e., from the left to the right and *vice versa*. The whole dataset contains  $4 \times 8 \times 4 = 128$  sequences and is a subset of a larger database presented in [29]. All sequences were obtained with a stationary camera with frame rates of 25 frames/s and with subsampling of the spatial resolution to  $160 \times 120$  pixels.

to test the effectiveness of the proposed approach in distinguishing a wide variety of human motion that are performed by different subjects. Finally, we have dataset S3: All video streams of S1 and only the walking of S2 are selected. The purpose is to test the generality of the QNT in differentiating the same human motion from different environments.

For each video sequence created in the datasets, we defined a five DOF kinematic human skeleton structure. The motion parameter  $\phi_t$  for the six landmarks points on the proposed human model as in Fig. 4, including the reference landmark point, are obtained using a visual tracking approach [10]. The numbers of particles are selected as 400, 800, and 2000, respectively. Further,  $\phi_t$  are mapped into their associated sequence discrete symbolic representation (i.e., qualitative states in the fuzzy qualitative unit circle). The level of resolution in the fuzzy qualitative unit circle is set to  $s = 12$  and  $r = 16$  for translation and orientation components, respectively. Finally, the QNTs for each human motion are constructed using the fuzzy qualitative robot kinematics algorithm [19], and human motion recognition is carried out afterwards. Fig. 7 shows the visualization of the activity model derived from eight different subjects.

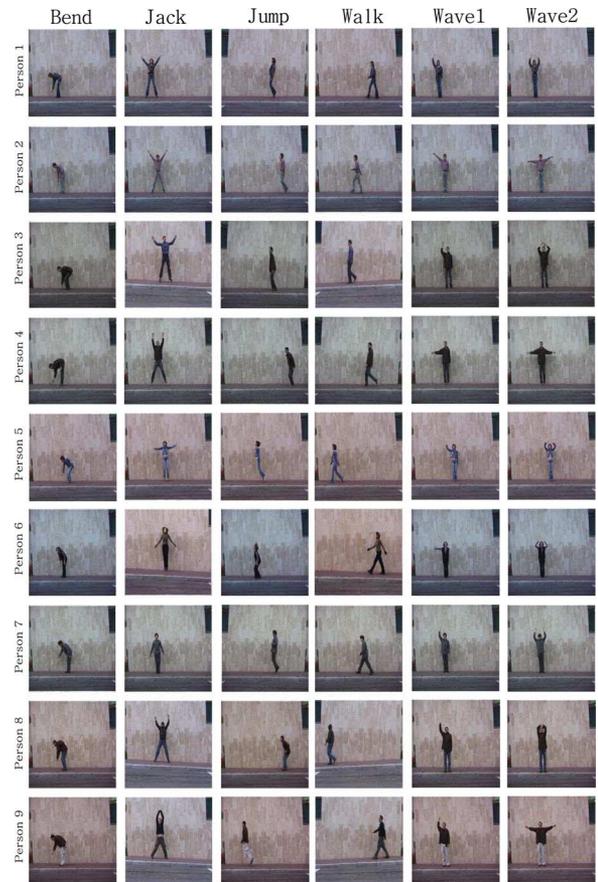


Fig. 6. Example sequences from dataset with six human activities performed by nine different persons. The whole dataset contains  $6 \times 9 = 54$  sequences and is a subset of a larger database presented in [2].

## B. Results and Analysis

Human-motion recognition aims to recognize the type of activities performed by people in test sequences against results from training sequences. The datasets were divided into training sets and test sets with respect to subjects, such that the same person will not appear in the test and training sequences simultaneously. As the size of the datasets is relatively small, random permutations of the training and testing sets are considered, and the recognition rates were averaged. For each of the datasets S1, S2, and S3, we repeated the process 50 times and evaluated the performance of the proposed method; the correct classification rate (CCR) is used to justify the recognition rate. CCR is the ratio of correctly classified number of activities to the total number of the same activity. For activity classification, we adopted the nearest-neighbor classifier, where the Euclidean metric was used as the distance measure.

The recognition results for each dataset are shown in Tables III–V, respectively. The conclusions are drawn as follows. First, the percentages of correct classification of the proposed approach are acceptable for all three datasets. The mean of classification accuracy for each dataset is higher than 80%. The recognition rates of the three datasets with different number of particles  $N$  in the condensation algorithm are almost similar. This demonstrated that the proposed data-quantization process

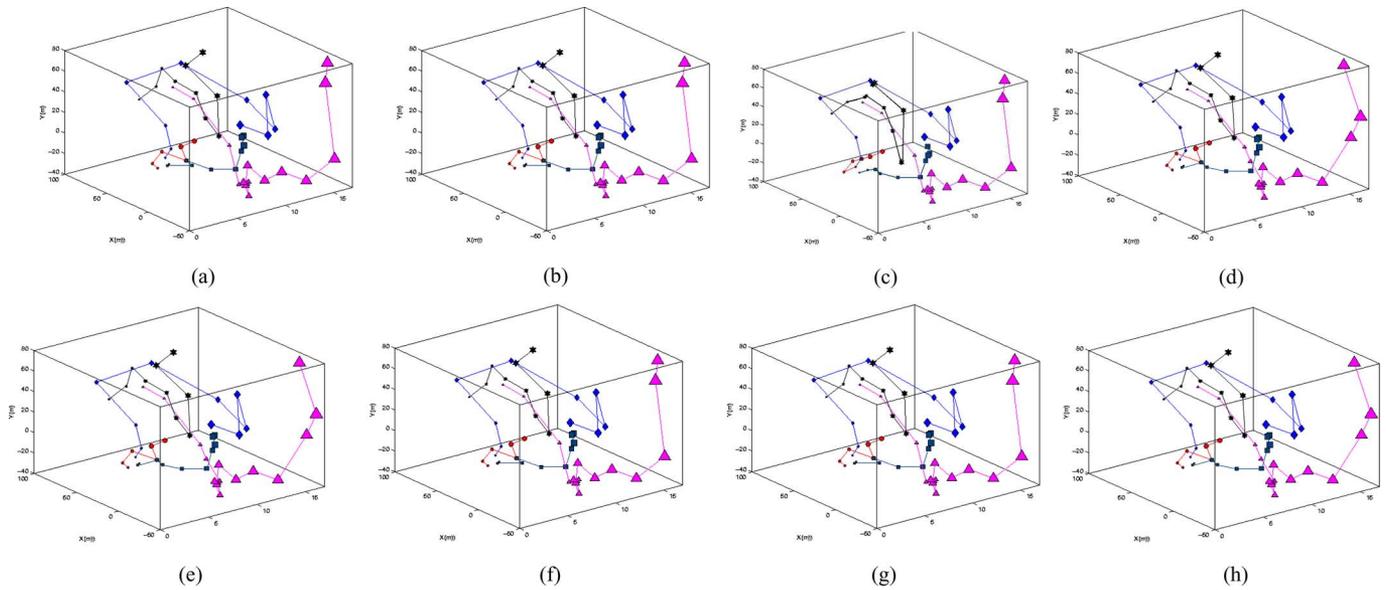


Fig. 7. Visualization of activities manifold: each of the five activities from eight subjects in quantity space. (a) Subject 1. (b) Subject 2. (c) Subject 3. (d) Subject 4. (e) Subject 5. (f) Subject 6. (g) Subject 7. (h) Subject 8.

TABLE III  
RECOGNITION RATE FOR S1

| Particles | Walking | Running | Jogging |
|-----------|---------|---------|---------|
| (400)     | 80%     | 81%     | 92%     |
| (800)     | 82%     | 81%     | 92%     |
| (2000)    | 80%     | 81%     | 93%     |

TABLE IV  
RECOGNITION RATE FOR S2

| Particles | Bending | Walking | Jacking | Jumping | Wave1 | Wave2 |
|-----------|---------|---------|---------|---------|-------|-------|
| (400)     | 100%    | 100%    | 100%    | 100%    | 100%  | 100%  |
| (800)     | 100%    | 100%    | 100%    | 100%    | 100%  | 100%  |
| (2000)    | 100%    | 100%    | 100%    | 100%    | 100%  | 100%  |

TABLE V  
RECOGNITION RATE FOR S3

| Particles | Walking  | Running | Jogging | Wave1 | Wave2 |
|-----------|----------|---------|---------|-------|-------|
| (400)     | 86%(86%) | 81%     | 92%     | 100%  | 100%  |
| (800)     | 86%(86%) | 81%     | 92%     | 100%  | 100%  |
| (2000)    | 86%(86%) | 81%     | 92%     | 100%  | 100%  |

has sufficiently relaxed the tradeoff between tracking precision and computational cost in the visual tracking algorithm. Second, the QNT are informative without necessary data lost, particularly in dataset S1, where the three activities exhibit very similar movement but have substantially different meanings. The QNT misclassified only a small number of subjects given by the three tested human motions in S1. Actually, the misclassified data are even hardly distinguishable from a human perspective. A confusion matrix in Fig. 8 is shown to analyze which activity has been incorrectly classified. It has shown that the algorithm misclassified some of the walking and running activities. Third, we showed the true positive fraction (TPF) and false positive fraction (FPF) via the leave-one-out rule in the verification model. This is to test the capability of a pattern classifier to verify whether a new measurement belongs to certain claimed class. Fig. 9 shows the receiver operating characteristic (ROC) curves of the three activities (i.e., jogging, running, and walking) in the datasets. The reason we only chose these three activities is that these three activities exhibit similar actions, and it is interesting to see how the proposed templates coped. Finally, Table V shows the generality results of the QNT in terms of differentiating the same human motion from different environments, i.e., the constructed walking QNT from dataset S1 is employed to recognize the walking data in S2 and *vice versa*. Similar recognition rates were achieved from both set of experiences. This illustrates that the proposed QNT are generic and insensitive to different motion styles, speeds across different human anatomy, and environments.

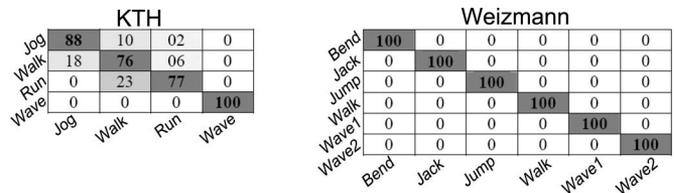


Fig. 8. Confusion matrices. Comparison of activity classification results in the KTH and Weizmann database.

C. Quantitative Comparison

A comparison was carried out between the dominant recognition methods (i.e., an HMM [28] and an FHMM [23]) and the proposed method. A four-state left-right discrete HMM is selected for the comparison, the preprocessing steps were conducted as given in [27], and the FHMM was conducted as in [24]. The number of states is empirically determined, and it is observed that an increase to a larger number of states did not result in any performance gains. Each model was trained using 1%, 20%, and 50% of randomly selected instances of human motions, and the rest were employed as testing data. The comparison results are provided in Tables VI–VIII, respectively. It is observed that on the three data sets, the QNT outperforms

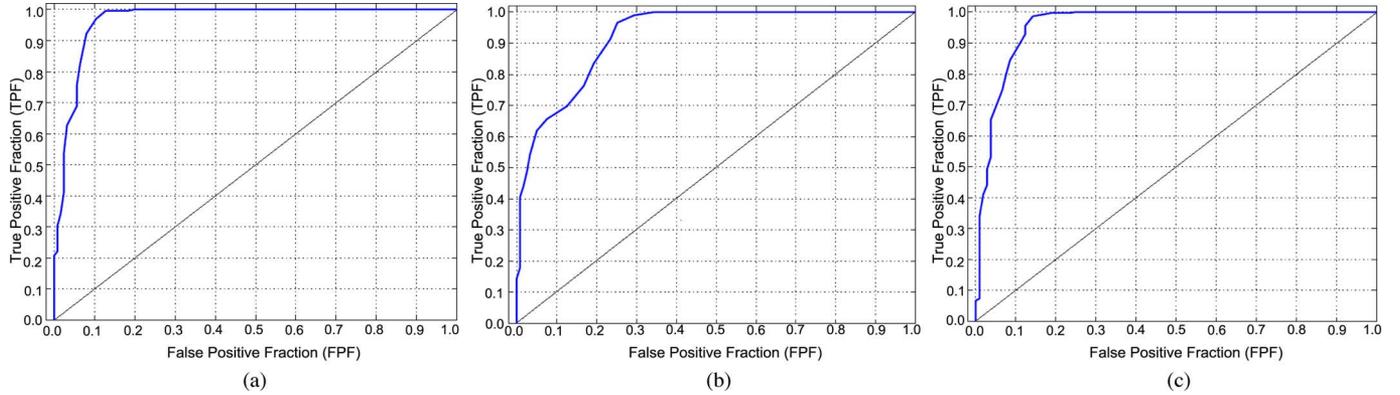


Fig. 9. ROC curves for three activities that exhibit similar actions. The dotted diagonal line shows the random prediction. (a) ROC curve for kicking. The area under the ROC curve is 0.966026. (b) ROC curve for running. The area under the ROC curve is 0.923142. (c) ROC curve for walking. The area under the ROC curve is 0.957291.

TABLE VI  
COMPARISON WITH THE HMM AND FHMM

|    | HMM with<br>1%<br>training data | HMM with<br>20%<br>training data | HMM with<br>50%<br>training data | FHMM with<br>1%<br>training data | FHMM with<br>20%<br>training data | FHMM with<br>50%<br>training data | QNT with<br>1%<br>training data |
|----|---------------------------------|----------------------------------|----------------------------------|----------------------------------|-----------------------------------|-----------------------------------|---------------------------------|
| S1 | 54%                             | 75%                              | 77%                              | 58%                              | 77%                               | 78%                               | 85%                             |
| S2 | 62%                             | 88%                              | 91%                              | 70%                              | 91%                               | 100%                              | 100%                            |
| S3 | 68%                             | 72%                              | 72%                              | 75%                              | 78%                               | 82%                               | 88%                             |

Average classification rate employing different training data sizes.

both the HMM and FHMM. Note that the QNT employed in this experiment was constructed from 400 particles with 1% training data, while the best models in the HMM and FHMM were employed for this comparison.

The simulation results have confirmed that the effectiveness of the HMM and FHMM models are significantly dependant on the accuracy of the training data and the amount of training data employed. For instance, in Table VI, the classification rate of the HMM and FHMM using 1%, 20%, and 50% training data had a big impact of the recognition rate, whereas the QNT is fairly consistent. One of the main reasons is that the proposed solution is not a statistical learning method; hence, it does not require large training data. Instead strong discriminative features are required from example activities. A further analysis by employing one subject sequentially as the only training data and the rest as testing data results in Table VII shows that the choice of the selection has a significant influence on the recognition rate in the HMM and FHMM. The worst and the best achieved differ by approximately 50% for the HMM and 48% for the FHMM, while the QNT only differs by approximately 2%. Besides, both the HMM and FHMM are also notoriously sensitive to the precision of the training data. This is notable from Table VIII as  $N = 400$ , and the average percentage of successful recognition is only 51% for the three datasets. However, at a much higher resolution of the tracking algorithm, the average percentage of successful recognition rate increases to more than 80%. Bear in mind that the high recognition rate is achieved at the cost of significant computational power.

#### D. Complexity Analysis

The condensation algorithm is an approximation technique by representing the posterior density as a set of samples of the state

TABLE VII  
COMPARISON WITH THE HMM AND FHMM

|    | HMM            | FHMM            | QNT               |
|----|----------------|-----------------|-------------------|
| S1 | 54% (38%)[88%] | 62% (45%)[88%]  | 85% (84%)[86%]    |
| S2 | 75% (64%)[98%] | 84% (72%)[100%] | 100% (100%)[100%] |
| S3 | 67% (32%)[80%] | 69% (36%)[84%]  | 88% (88%)[88%]    |

Average classification rate with using each subject as training data once, and tested against the remaining. The worst achieved is in brackets () and the best in square [].

space with associated likelihood weights  $\omega_t^i$ ,  $i \in \{1, \dots, N\}$ . The sample set approximation of the current posterior density  $p(X_t|O_{1:t})$  can be obtained via

$$p(X_t|O_{1:t}) \approx \sum_{i=1}^N \omega_t^i \delta(X_t - x_t^i) \quad (16)$$

where  $\delta(X_t - x_t^i)$  denotes the Dirac delta function. Therefore, the complexity of a standard condensation algorithm has a computational complexity of  $\mathcal{O}(2N)$ , where  $N$  is the number of particles.

In the meantime, the forward, backward, and Viterbi algorithms are the central elements of HMM training and testing. All have the same computational complexity [13], and the equation for the standard forward algorithm is

$$\alpha_j(t) = \left( \sum_i \alpha_i(t-1) a_{ij} \right) b_j(o_t) \quad \forall j, t. \quad (17)$$

The complexity of the standard algorithm is normally given as  $\mathcal{O}(M^2T)$ , where  $M$  is the total number of unique states in the HMM, and  $T$  is the number of observations.

In Table IX, a comparison of the computational complexity of the HMM, FHMM, and QNT method is shown. The QNT

TABLE VIII  
COMPARISON WITH HMM AND FHMM

|    | HMM with<br>featured data<br>400 particles | HMM with<br>featured data<br>800 particles | HMM with<br>featured data<br>2000 particles | FHMM<br>featured data<br>400 particles | FHMM<br>featured data<br>800 particles | FHMM<br>featured data<br>2000 particles | QNT with<br>featured data<br>400 particles |
|----|--|--|---|--|--|---|--|
| S1 | 38%  | 78%  | 82%   | 39%                                    | 81%                                    | 83%                                     | 85%  |
| S2 | 61%  | 80%  | 88%   | 72%                                    | 93%                                    | 100%                                    | 100%                                       |
| S3 | 54%  | 83%  | 88%   | 51%                                    | 84%                                    | 88%                                     | 88%  |

Average classification rate employing different tracking resolutions.

TABLE IX  
COMPARISON OF ALGORITHMS COMPUTATIONAL COMPLEXITIES

| Algorithm                         | Complexity                            |
|-----------------------------------|---------------------------------------|
| Particle Filtering + Standard HMM | $\mathcal{O}(2N) + \mathcal{O}(M^2T)$ |
| Particle Filtering + FHMM         | $\mathcal{O}(2N) + \mathcal{O}(M^2T)$ |
| QNT                               | $\mathcal{O}(2N)$                     |

TABLE X  
COMPARISON OF ALGORITHM COMPUTATIONAL COMPLEXITIES VERSUS  
RECOGNITION RATE

|    | Algorithm                         | Complexity                                | Recognition Rate |
|----|-----------------------------------|---|------------------|
| S1 | Particle Filtering + Standard HMM | $\mathcal{O}(4048)$<br>$N=2000, M=4, T=3$ | 82%              |
|    | Particle Filtering + FHMM         | $\mathcal{O}(4048)$<br>$N=2000, M=4, T=3$ | 83%              |
|    | QNT                               | $\mathcal{O}(800)$<br>$N=400$             | 85%              |
| S2 | Particle Filtering + Standard HMM | $\mathcal{O}(4096)$<br>$N=2000, M=4, T=6$ | 88%              |
|    | Particle Filtering + FHMM         | $\mathcal{O}(4096)$<br>$N=2000, M=4, T=6$ | 100%             |
|    | QNT                               | $\mathcal{O}(800)$<br>$N=400$             | 100%             |
| S3 | Particle Filtering + Standard HMM | $\mathcal{O}(4048)$<br>$N=2000, M=4, T=3$ | 88%              |
|    | Particle Filtering + FHMM         | $\mathcal{O}(4048)$<br>$N=2000, M=4, T=3$ | 88%              |
|    | QNT                               | $\mathcal{O}(800)$<br>$N=400$             | 88%              |

The best recognition rate achieved by each system is selected.

method outperformed both the HMM and FHMM in twofolds. First of all, the proposed solution does not need to employ a large amount of particles  $N$  to perform tracking as the proposed data quantization process can account for tracking error, as shown in Table II. Second, the proposed method is not a statistical learning method; hence, the algorithm does not need to loop until convergence. Furthermore, Table X shows a comparison of the recognition rate of the HMM, FHMM, and our proposed method against the computational complexity. It is notable that in order to achieve the recognition rate with our proposed method on each data set S1, S2, and S3, the HMM and FHMM computational complexities are at least five times higher than our proposed method.

#### IV. CONCLUDING REMARK

In this paper, a fuzzy qualitative approach has been proposed to solve real-time vision-based human motion analysis. It has integrated human motion tracking and recognition algorithms with fuzzy qualitative robot kinematics. The simulation results have shown that the proposed method outperforms the dominant recognition methods, i.e., the HMMs and FHMMs. It is demonstrated that the proposed method can be applied to recognize a range of human motions for a real-time human motion recognition system from the computation perspective. The proposed algorithm has achieved average computational time at about 0.010591 s per individual human motion recognition; on the

other hand, the frame rate of a closed-circuit television (CCTV) image sequence is usually reduced to 10–15 frames per second when in use. That the proposed recognition method is at least ten times faster than the frame rate of CCTV image sequences is effective enough to be implemented in a real-time vision-based human motion analysis. It is clearly evident that the proposed method can meet real-time requirement for human motion recognition. However, there are still some problems that keep the proposed method from real-world application. First, initial parameters of an image sequence are manually edited in human motion tracking. It has been confirmed that it is one of the most challenging problems for the computer vision community. Second, issues related to object tracking algorithms such as partial occlusion, clutter, and environmental lighting changes still remain as open problems to further apply the proposed method to real-world application. Additionally, we are trying to overcome the aforementioned image processing problems by understanding scenario-context information without accurate human motion extraction. For instance, we are currently developing multiple QNTs from different viewpoints to construct a pseudo 3-D motion template, which allows reasoning the human motion analysis under environmental uncertainty such as lighting and clutter background, which relax the requirements on accurate initial image parameters and uncertainty caused by the surroundings. We aim to develop a compositional model that uses predominantly knowledge-based techniques to translate among high-level human motion scenarios, Gaussian mixture models, and filtered numerical data of human motions [17].

#### REFERENCES

- [1] J. K. Aggarwal and Q. Cai, "Human motion analysis: A review," *Comput. Vis. Image Understanding*, vol. 73, no. 3, pp. 428–440, 1999.
- [2] M. Blank, L. Gorelick, E. Shechtman, M. Irani, and R. Basri, "Action as space-time shapes," in *Proc. IEEE Int. Conf. Comput. Vis.*, Beijing, China, 2005, vol. 2, pp. 1395–1402.
- [3] A. Bobick and J. Davis, "Real-time recognition of activity using temporal templates," in *Proc. IEEE CS Workshop Appl. Comput. Vis.*, 1996, pp. 39–42.
- [4] A. Bobick and A. Wilson, "A state-based technique for the summarization and recognition of gesture," in *Proc. Int. Conf. Comput. Vis.*, 1995, pp. 382–388.
- [5] G. Bradski and J. Davis, "Motion segmentation and pose recognition with motion history gradients," *Mach. Vis. Appl.*, vol. 13, no. 3, pp. 174–184, Jul. 2002.
- [6] C. Bregler, "Learning and recognizing human dynamics in video sequences," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, Washington, DC, 1997, pp. 568–574.
- [7] H. Buxton, "Learning and understanding dynamic scene activity," *Image Vis. Comput.*, vol. 21, no. 1, pp. 125–136, 2003.
- [8] C. S. Chan, H. Liu, and D. Brown, "Recognition of human motion from qualitative normalized templates," *J. Intell. Robot. Syst.*, vol. 48, no. 1, pp. 79–95, Jan. 2007.

- [9] Z. Chen and H.-J. Lee, "Knowledge-guided visual perception of 3-d human gait from a singleimage sequence," *IEEE Trans. Syst., Man Cybern.*, vol. 22, no. 22, pp. 336–342, Mar./Apr. 1992.
- [10] M. Isard and A. Blake, "Condensation: Conditional density propagation for visual tracking," *Int. J. Comput. Vis.*, vol. 29, no. 1, pp. 5–28, 1998.
- [11] Z. Jing and S. Sclaroff, "Segmenting foreground objects from a dynamic textured background via a robust Kalman filter," in *Proc. 9th IEEE Int. Conf. Comput. Vis.*, Washington, DC, 2003, pp. 44–50.
- [12] G. Johansson, "Visual motion perception," *Sci. Amer.*, vol. 232, no. 6, pp. 76–88, 1975.
- [13] M. Johnson, "Capacity and complexity of hmm duration modeling techniques," *IEEE Signal Process. Lett.*, vol. 12, no. 5, pp. 407–410, May 2005.
- [14] S. Ju, M. Black, and Y. Yacoob, "Cardboard people: A parameterized model of articulated image motion," in *Proc. 2nd Int. Conf. Autom. Face Gesture Recognit.*, Killington, VT, 1996, pp. 38–44.
- [15] I. Karaulova, P. Hall, and A. Marshall, "A hierachical model of dynamics for tracking people with a single video camera," in *Proc. Br. Mach. Vis. Conf.*, 2000, pp. 262–352.
- [16] M. Leo, T. D'Orazio, I. Gnoni, P. Spagnolo, and A. Distanto, "Complex human activity recognition for monitoring wide outdoor environments," in *Proc. 17th IAPR Int. Conf. Pattern Recognit.*, 2004, vol. 4, pp. 913–916.
- [17] H. Liu, "A fuzzy qualitative framework for connecting robot qualitative and quantitative representations," *IEEE Trans. Fuzzy Syst.*, vol. 16, no. 6, pp. 1522–1530, Dec. 2008.
- [18] H. Liu and D. Brown, "An extension to fuzzy qualitative trigonometry and its application to robot kinematics," in *Proc. IEEE Int. Conf. Fuzzy Syst.*, 2006, pp. 1111–1118.
- [19] H. Liu, D. Brown, and G. Coghill, "Fuzzy qualitative robot kinematics," *IEEE Trans. Fuzzy Syst.*, vol. 16, no. 3, pp. 802–822, Jun. 2008.
- [20] H. Liu and G. Coghill, "Fuzzy qualitative trigonometry," in *Proc. IEEE Int. Conf. Syst., Man Cybern.*, 2005, pp. 1291–1296.
- [21] F. Lv and R. Navatia, "Recognition and segmentation of 3d human action using HMM and multi-class adaboost," in *Proc. Eur. Conf. Comput. Vis.*, Graz, Austria, 2006, vol. 4, pp. 359–372.
- [22] T. B. Moeslund and E. Granum, "A survey of computer vision-based human motion capture," *Comput. Vis. Image Understanding*, vol. 81, no. 3, pp. 231–268, 2001.
- [23] M. Mohamed and P. Gader, "Generalized hidden Markov models. i. Theoretical frameworks," *IEEE Trans. Fuzzy Syst.*, vol. 8, no. 1, pp. 67–81, Feb. 2000.
- [24] M. Mohamed and P. Gader, "Generalized hidden Markov models. ii. Application to handwritten word recognition," *IEEE Trans. Fuzzy Syst.*, vol. 8, no. 1, pp. 82–94, Feb. 2000.
- [25] R. M. Murray, S. Shastry, Z. Li, and S. S. Sastry, *A Mathematical Introduction to Robotic Manipulation*. Boca Raton, FL: CRC, 1994.
- [26] V. Pavlovic, J. Rehg, and J. MacCormick, "Impact of dynamic model learning on classification of human motions," in *Proc. Int. Conf. Comput. Vis. Pattern Recognit.*, 2000, pp. 788–795.
- [27] H. Qiang and C. Debrunner, "Individual recognition from periodic activity using hidden Markov models," in *Proc. Workshop Human Motion*, 2000, pp. 47–52.
- [28] L. Rabiner, "A tutorial on hidden Markov models and selected applications in speech recognition," *Proc. IEEE*, vol. 77, no. 2, pp. 257–284, Feb. 1989.
- [29] C. Schuldt, I. Laptev, and B. Caputo, "Recognizing human actions: A local svm approach," in *Proc. Int. Conf. Pattern Recognit.*, Hong Kong, 2004, vol. 3, pp. 32–36.
- [30] Q. Shen and R. Leitch, "Fuzzy qualitative simulation," *IEEE Trans. Syst., Man Cybern.*, vol. 23, no. 4, pp. 1038–1061, Jul./Aug. 1993.
- [31] H. Sidenbladh, M. Black, and D. Fleet, "Stochastic tracking of 3d human figures using 2d image motion," in *Proc. 6th Eur. Conf. Comput. Vis.*, London, UK: Springer-Verlag, 2000, pp. 702–718.
- [32] A. Sundaresan, R. Chellappa, and A. Roy Chowdhury, "Multiple view tracking of humans modeled by kinematic chains," in *Proc. Int. Conf. Image Process.*, Singapore, 2004, vol. 2, pp. 1009–1012.
- [33] L. Wang, W. Hu, and T. Tan, "Recent developments in human motion analysis," *Pattern Recognit.*, vol. 36, no. 3, pp. 585–601, 2003.
- [34] L. Wang, T. Tan, H. Ning, and W. Hu, "Silhouette analysis-based gait recognition for human identification," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 25, no. 12, pp. 1505–1518, Dec. 2003.
- [35] L. Yuan, A. Haizhou, T. Yamashita, L. Shihong, and M. Kawade, "Tracking in low frame rate video: A cascade particle filter with discriminative observers of different lifespans," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, Washington, DC: IEEE Comput. Soc., 2007, pp. 1–8.



**Chee Seng Chan** (S'05) received the B.Eng. degree from the University of Multimedia, Cyberjaya, Selangor, Malaysia, in 2003 and the M.Sc. and Ph.D. degrees from the University of Portsmouth, Portsmouth, U.K., in 2004 and 2008, respectively.

He is currently with the Institute of Industrial Research, University of Portsmouth. His current research interests include human motion analysis, artificial intelligence, qualitative reasoning, knowledge-based theory, and applications.



**Honghai Liu** (M'02–SM'06) received the Ph.D. degree in robotics from King's College London, London, U.K., in 2003.

He joined the University of Portsmouth, Portsmouth, U.K., in September 2005. He previously held research appointments at the University of London, London, U.K., and the University of Aberdeen, Aberdeen, U.K., and industrial appointments in system integration. He has authored or coauthored more than 100 research papers. His current research interests include computational intelligence methods and

applications with a focus on those approaches that could make contributions to the intelligent connection of perception to action.

Dr. Liu has received three Best Paper Awards.