GMM-QNT Hybrid Framework for Vision-based Human Motion Analysis

Chee Seng Chan and Honghai Liu

Abstract— **The understanding of human behaviour in video is a challenging task in that the same behaviour might have several different meanings depending upon the scene and task context in which it is performed. While human seem to perform scene interpretations without effort, this is a formidable and yet unsolved task for artificial vision systems. One of the main reasons is that there exists a gap between low-level vision at signal level and high-level representation of activities at symbolic level. In this paper, we present an intelligent connection framework using Gaussian Mixture Model-based clustering (GMM) to bridge the low-level vision data and the Qualitative Normalised Templates (QNT) - a symbolic representation for human motion based on fuzzy qualitative robot kinematics, which could link the former with domain-dependent scenarios. The proposed method has been applied to the recognition of eight types of human motions and an empirical comparison with fuzzy hidden Markov-based human motion recognition system.**

I. Introduction

Understanding human behaviour in video is essential in numerous applications, e.g. crime and healthcare industries. Technically, it is evident that the state of the art in vision sensing hardware has sufficiently met the applications requirements for the purpose of public safety and healthcare. The hardware involves are hundreds of thousands of networked cameras located in specific regions and locations. On the other hand, developing algorithms capable of recognising human behaviour in certain constraints is the major bottleneck. A typical software package of such a system usually consists of motion segmentation, motion representation, action recognition and behaviour description. It is anticipated that an ideal human motion analysis system automatically generates either a quantitative or qualitative description for understanding the involved human motion behaviour given a set of visual sequences.

To this extend, significant amounts of research have been conducted and contribute to human behaviour understanding at different systematic levels in the past two decades. For instance, Wilson and Bobick [17] proposed a modified Hidden Markov Model (HMM) approach to activity recognition. They introduced the term parametric gestures defined as gestures that exhibit a systematic spatial variation. As an example, the paper cited a pointing gesture, where the relevant parameter is a two-dimensional direction. The standard HMM method of gesture recognition was extended by including a global parameter-driven variation in the output probabilities of the HMM states. Then they formulated an Expectation-Maximisation (EM) method for training this parameter driven HMM. During testing, a similar EM algorithm simultaneously maximises the output likelihood of the parametrically driven HMM for the given motion sequence, whilst estimating the quantifying parameters. EM training is based on learning from samples, and the parametric aspect in terms of direction is based on prior knowledge of certain gestures. This approach assumes the availability of pre-segmented gestures. Oliver et al. [14], on the other hand, employed coupled HMM [3] to model pedestrian activity for surveillance and analysing actions which occur between two pedestrians. In this model, two (or more) HMM are coupled, with the state of each at time t affecting the state at time $t+1$. They trained their model on synthetic data, and a mixture of synthetic and real data.

Vaswani et al. [16] used Kendall's statistical shape theory [9] where nonlinear dynamical models are used to characterise the shape variation over time. An activity is recognised if it agrees with the learnt parameters of the shape and associate dynamics. Gonzalez et al. [6] employed the point distribution model to model the variability of joint angle setting of a stick figure model. An action space, aSpace, is trained by giving a set of joint angle setting coming from different individuals by showing the same action. Ivanov and Bobick [8] present an automatic surveillance system that labels events and interactions by using syntactic constraints. Their goal is to label person-vehicle interactions such as pick-up, drop-off, exit, and enter in an open parking lot. In a similar way in Ayers and Shah [1], they built an environment map and selected region of interests. Their system is composed of a tracking module, an event generator, and a parser. The tracker tracks any moving objects, and the event generator maps the object tracks onto a set of predetermined discrete events. The event generator uses an environment map (e.g. the scene model of the parking lot) as contextual information to assign visual changes to discrete events in the parking lot. The parser uses an activity grammar to parse the sequence of discrete events into meaningful labels of interactions between the person and the vehicle. This approach seems adequate for grouping into meaningful labels the discrete events distributed sparsely along a lengthy sequence of visual surveillance data. That is, the syntactic method makes it possible to extract meaningful interactions from the heterogeneous sequence that is composed of, and intermingled with, several different processes to which HMM methods can not be applied.

While all these approaches have demonstrated success in recognising complex human activities in research context,

Chee Seng Chan and Honghai Liu are with the Institute of Industrial Research, University of Portsmouth, Portsmouth PO1 3QL, U.K. (e-mails: {cheeseng.chan;honghai.liu}@port.ac.uk.)

978-1-4244-3597-5/09/$25.00 ©2009 IEEE

in real world however, the ability to model and process video information computationally lags far behind. Several key issues have slowed the advance in automated video processing: (1) lack of robust method for tracking human activity in video. For instance, there is a trade-off between computational cost and tracking precision. 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 τ_k is the cost of calculating $p(z_k|x)$; and (2) there is a gap between the low-level image/video processing and high-level representation of human activities at symbolic level. That is, it requires the integration of semantic knowledge of the real world and the processing of low-level video signal from camera input.

This paper focuses on the latter and presents an intelligent connection framework using Gaussian Mixture Model-based clustering (GMM) to bridge the low-level vision data extracted from image/video sequences to our previous research, Qualitative Normalised Templates (QNT) [5] which could link the former with domain-dependent scenarios. QNT is a novel, symbolic representation for human activity that based on fuzzy qualitative robot kinematics. Empirical results on the two available databases and a comparison with the Fuzzy Hidden Markov Model (FHMM) approach have shown the effectiveness of the proposed method.

The rest of the paper is structured as follows. Section II revisits the QNT. Section III introduces the GMM algorithm. Section IV presents the experimental results and a comparison with the FHMM. Section V concludes the paper with discussions and future work.

II. Qualitative Normalised Templates

In this section, the QNT [5] - a novel, symbolic representation for human motion based on fuzzy qualitative robot kinematics [10] and template matching is revisited.

A. Video Sequence

From the perspective of robotics, a human motion skeleton can be constructed in terms of robot kinematics models of body segments; it then convert human motion analysis into a conventional robotics problem. Moreover, modelling the human skeleton as rigid parts linked in a kinematics structure is relatively easy to automatically detect and track in real videos. First of all, a reduced kinematic model (half of the human skeleton along a sagittal section, since the other half follows by symmetry) is constructed as illustrated in Fig. 1.

The model has been initialised in the first frame of the video sequence by specifying the geometry of the links and by clicking their position in the image. Once this initialisation step is completed, the system performs the tracking approach similar to [7]. That is, on each video sequence, the time trajectory of the projection of 6 body joints, i_s where $s \in \{1,2,\cdots,6\}$ are considered onto the image plane.

B. Qualitative Data

Secondly, the time trajectory of each body joints projection, i_s are mapped into a set of discrete symbol representations, *qualitative states* based in the modified unit circle [11] by a quantisation process:

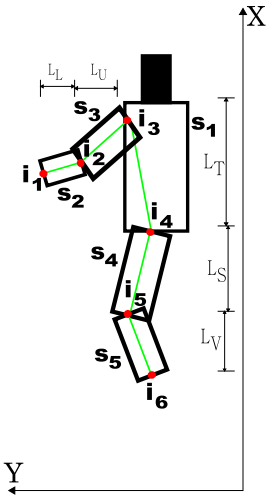


Fig. 1. The proposed 1/2 reduced kinematic model of human skeleton along a sagittal plane.

$$\begin{aligned}\lim_{s\to s_o} C_t(s=12) &= QS(qp_l)\\ \lim_{r\to r_o} C_o(r=16) &= QS(qp_\theta)\end{aligned} \tag{1}$$

s is the number of qualitative states that reside in the x-y translation and *r* is the number of qualitative states that reside on the orientation in the modified unit circle. As $s \to s_o$ and $r \to r_o$, 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 number of 4-tuple fuzzy numbers of the translation, *s* and orientation, *r* in the modified unit circle are application dependent.

In order to make the representation works on the same platform, a normalisation process within the modified unit circle [-1 1] is conducted,

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

where x-y translation states, qp_l are normalised by the average length of the human body segment and the orientation states, qp_θ are normalised to 2π. Now, each body joints, i_s is represented by the corresponding state region of a fuzzy qualitative state in the modified, normalised unit circle.

C. QNT

As discussed earlier, from the perspective of robotics, a human motion skeleton can be constructed in terms of robot kinematics models of body segments; it then convert human motion analysis into a conventional robotics problem. The motion of each body joint, i_s is represented in terms of twist representation, ξ

$$\xi = \begin{bmatrix} \upsilon_1 & \upsilon_2 & \upsilon_3 & \omega_1 & \omega_2 & \omega_3 \end{bmatrix}^T \quad (3)$$

where ω is a 3D 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, θ multiplied by the twist, $\xi\theta$. While the υ component determines the location range of the rotation axis and the amount of translation along the same axis.

For instance, the product of exponential maps for the arm kinematics chains with respect to a base frame $g(0)$ over time T ($T \in (1, \cdots, m)$) can be obtained as below,

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

where

$$g_{arm}(0) = \begin{bmatrix} \mathbf{I} & \begin{bmatrix} \mathbf{L_T} \\ \mathbf{L_U} + \mathbf{L_L} \\ 0 \end{bmatrix} \\ \mathbf{0} & \mathbf{1} \end{bmatrix} \quad (4a)$$

$$\omega_1 = \omega_2 = \omega_3 = \begin{bmatrix} \mathbf{0} \\ \mathbf{0} \\ \mathbf{0} \end{bmatrix} \quad (4b)$$

$$q_1 = \begin{bmatrix} \mathbf{0} \\ \mathbf{0} \\ \mathbf{0} \end{bmatrix} \qquad q_2 = \begin{bmatrix} \mathbf{L_T} \\ \mathbf{0} \\ \mathbf{0} \end{bmatrix} \qquad q_3 = \begin{bmatrix} \mathbf{L_T} \\ \mathbf{L_U} \\ \mathbf{0} \end{bmatrix} \quad (4c)$$

$$\xi_1 = \begin{bmatrix} -\omega_1 \times q_1 \\ \omega_1 \end{bmatrix} = \begin{bmatrix} \mathbf{0} \\ \mathbf{0} \\ \mathbf{0} \\ \mathbf{0} \\ \mathbf{0} \\ \mathbf{1} \end{bmatrix} \qquad \xi_2 = \begin{bmatrix} \mathbf{0} \\ -\mathbf{L_T} \\ \mathbf{0} \\ \mathbf{0} \\ \mathbf{0} \\ \mathbf{1} \end{bmatrix} \qquad \xi_3 = \begin{bmatrix} -\mathbf{L_U} \\ -\mathbf{L_T} \\ \mathbf{0} \\ \mathbf{0} \\ \mathbf{0} \\ \mathbf{1} \end{bmatrix} \quad (4d)$$

The product of exponential mapping for leg kinematics chains with respect to the same base frame over T ($T \in (1, \cdots, m)$) is given as below,

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

where

$$g_{leg}(0) = \begin{bmatrix} \mathbf{I} & \begin{bmatrix} -\mathbf{L_S} - \mathbf{L_V} \\ \mathbf{0} \\ \mathbf{0} \end{bmatrix} \\ \mathbf{0} & \mathbf{1} \end{bmatrix} \quad (5a)$$

$$\omega_4 = \omega_5 = \begin{bmatrix} \mathbf{0} \\ \mathbf{0} \\ \mathbf{1} \end{bmatrix} \quad (5b)$$

$$q_4 = \begin{bmatrix} \mathbf{0} \\ \mathbf{0} \\ \mathbf{0} \end{bmatrix} \qquad q_5 = \begin{bmatrix} -\mathbf{L_S} \\ \mathbf{0} \\ \mathbf{0} \end{bmatrix} \qquad \xi_4 = \begin{bmatrix} \mathbf{0} \\ \mathbf{0} \\ \mathbf{0} \\ \mathbf{0} \\ \mathbf{0} \\ \mathbf{1} \end{bmatrix} \qquad \xi_5 = \begin{bmatrix} \mathbf{0} \\ -\mathbf{L_S} \\ \mathbf{0} \\ \mathbf{0} \\ \mathbf{0} \\ \mathbf{1} \end{bmatrix} \quad (5c)$$

where

$$\mathbf{0} = \begin{bmatrix} 0 & 0 & 0 & 0 \end{bmatrix}; \quad \mathbf{1} = \begin{bmatrix} 1 & 1 & 0 & 0 \end{bmatrix}$$

Finally for any given activity, c its corresponding QNT is derived as:

$$QNT_c = g_{arm} \oplus g_{leg} \quad (6)$$

III. Gaussian Mixture Model

GMM plays an important role in clustering with a focus on density estimation and pattern recognition due to its analytical tractability, asymptotic properties, and ease of implementation.

A mixture model of *k* Gaussian components in dimension *D* is given as follows,

$$p(O_j) = \sum_{k=1}^{K} p(k)p(O_j|k) \quad (7)$$

where it denotes the density of O_j of the *k* components. Each component's density $p(O_i|k)$ is a normal probability distribution,

$$\begin{aligned} P(O_j) &= \sum_{k=1}^{K} p(k)p(O_j|\mu_k, \textstyle\sum_k) \\ &= \frac{p(k)}{(2\pi)^{D/2}\left|\sum_k\right|^{1/2}} \exp\left\{-\frac{1}{2}(O_j - \mu_k)^T \textstyle\sum_k^{-1}(O_j - \mu_k)\right\} \end{aligned} \quad (8)$$

where μ_k and $\sum_k$ represents the *k*the component center and covariance, respectively.

Due to the maximum likelihood estimate of the parameters of GMM cannot be solved analytically, EM algorithm has been employed. It is an iterative algorithm, and adjusts the parameters in the model by iterating over an E-steps and M-steps alternatively until convergence. The E-step and M-step are given as follows,

- E-step : The posterior probability of O_j is computed as

$$p(k|O_i) = \frac{p(k)p(O_i|\mu_k, \sum_k)}{\sum_{k=1}^{K} p(k)p(O_i|\mu_k, \sum_k)}$$

- M-step: The data class membership distribution is updated by

$$\mu_k = \frac{\sum_{i=1}^{N} p(k|O_i)O_i}{\sum_{k=1}^{K} p(k|O_i)}$$

$$\sum\nolimits_k = \frac{\sum_{i=1}^{N} p(k|O_i)(O_i - \mu_k)(O_i - \mu_k)^T}{\sum_{k=1}^{K} p(k|O_i)}$$

$$p(k) = \frac{1}{N}\sum\nolimits_{i=1}^{N} p(k|O_j)$$

IV. Experiments

In this section, the performance of the GMM as a bridge to link the low-level vision data and QNT; and a comparison to FHMM-based human motion analysis system are presented.

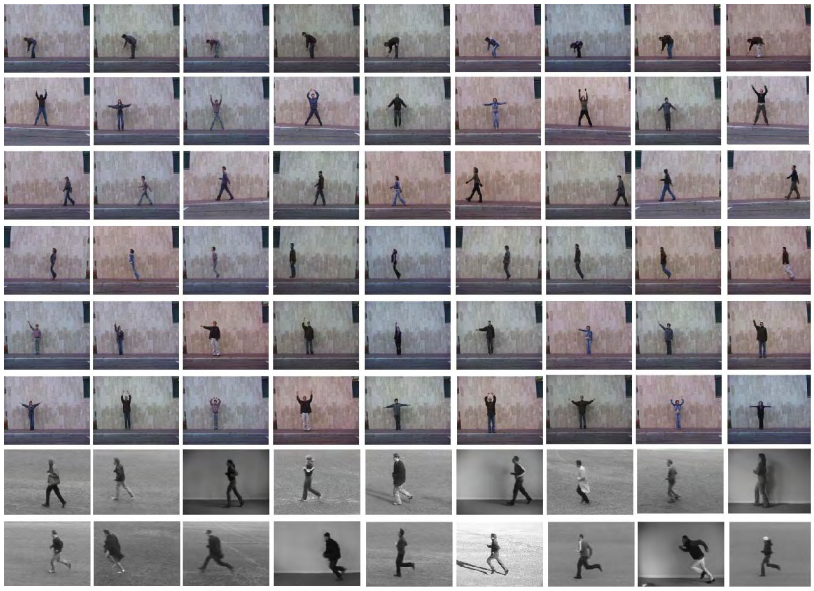


Fig. 2. Sample of the data sets created for the experiments. From top to bottom: bend, jack, walk, jump, wave1, wave2, jog and run, respectively.

A. Data Set

Experiments on two public databases has been conducted: The **KTH Database** [15] and The **Weizmann Database** [2]. Within these, two data sets were created: 1) **S1**: 225 video streams of 3 activities in 3 planar view scenarios from each of the 25 subjects were selected from the KTH Database. The selected activities were walk, run and jog. The aim is to test the efficacy of the QNTs as walking, running and jogging are activities that exhibit very similar movements but have dramatically different meanings. 2) **S2**: 55 video streams from the six activities were selected from the Weizmann Database. The selected activities were bend, walk, jack, jump, one hand waving (wave1) and two hands waving (wave2). The objective here is to test the effectiveness of the proposed approach in distinguishing a wide variety of activities that are performed by different humans. The two data sets are summarised in Table I and samples of the data sets are illustrated in Fig. 2.

TABLE I
TWO DATA SETS FOR HUMAN MOTION ANALYSIS

Data set	Videos	Subjects	Activities
S1	225	25	3
S2	55	9	6

B. Pre-processing and Training

First of all, we defined a five degree of freedoms kinematics structure of human skeleton as illustrated in Fig. 1. All body joints i_s have an axis orientation parallel to the Z-axis in the camera frame. For each created video sequence, the time trajectory of the body joint on the proposed human model were extracted similar to [7]. Then, accordingly to Section III, the GMM of each human activity are trained by a modified EM algorithm with k-means clustering initialisation as in [4]. The number of GMM components in each type of human activity is selected empirically and the results are shown in Tables II and III, respectively. For instance, Fig. 3 shows the Gaussian distribution of modelling the human motions in data set S1.

TABLE II
NUMBER OF GMM COMPONENTS FOR EACH HUMAN MOTION IN DATASET S1

Motion type	Jogging	Walking	Running
No.	4	4	5

TABLE III
NUMBER OF GMM COMPONENTS FOR EACH HUMAN MOTION IN DATASET S2

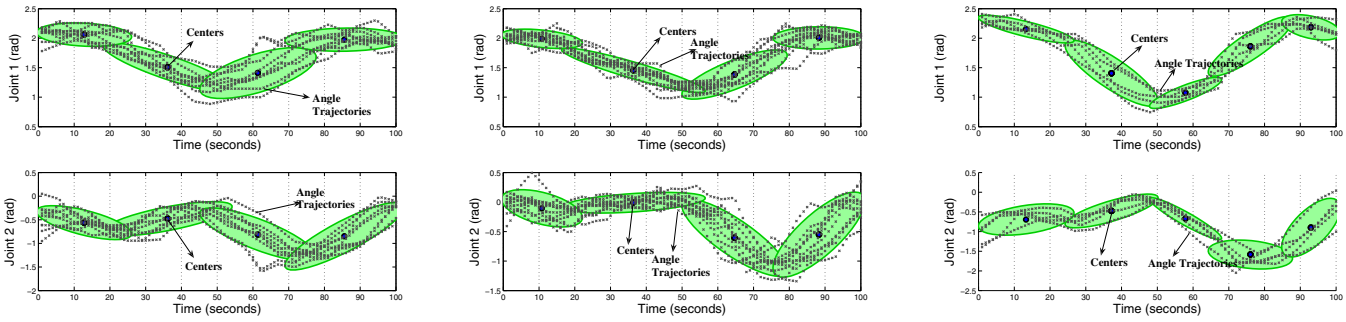
Motion type	Walking	Bending	Jumping	Jacking	Wave1	Wave2
No.	4	2	3	3	6	2

C. Results and Analysis

For activity classification, a minimum Euclidean distance function (Eq. 9) has been employed. Basically, for each time, the algorithm computes the average distance between the GMM and the QNT. The GMM is classified into its corresponding motion class with the minimum norm distance.

$$\mathrm{Norm}^{(i)} = \left(\sum_{i=1}^{n} |QNT_i - GMM_i|^2 \right)^{1/2} \quad (9)$$

The recognition results for each data set are shown in Tables IV and VI, respectively. From the analysis of the results, the following hypotheses can be made:



(a) Jogging (b) Walking (c) Running

Fig. 3. An example of GMM based human motion from dataset S1

- For all data sets, the percentage of correct classification with the proposed approach is beyond all expectation. The mean of classification accuracy for each data set is about 80%. This has shown that the proposed GMM of human motion can be employed to bridge the gap between low level processing and high level activity understanding.
- The performance of the GMM into classifying human activities is up to a good extent, in particular in S1 where the three activities exhibit very similar movement. In S1, the GMM mis-classified a small number of subjects given by the three tested activities exhibits very similar movement (see Table V). Further analysis on the mis-classified data found that the activity performed is also barely distinguishable from a human perspective.
- In order to test the robustness of the proposed method, we performed a second set of experiments by selecting a wide range of activities performed by different subjects. These activities are deliberately selected to evaluate the proposed solution. As expected, the successful recognition rate of S2 is perfect as for all the chosen activities, the motions differ greatly from each other (see Table VI). For instance, hand waving (wave1 and wave2) are a stationary activity and walking is non-stationary horizontal activity.

TABLE IV
RECOGNITION RATE FOR S1

Activity	Total	Correct	Accuracy
Walking	75	58	77%
Jogging	75	64	85%
Running	75	61	81%
Total			**81%**

TABLE V
CONFUSION MATRIX FOR S1

	Walking	Jogging	Running
Walking	**58**	11	12
Jogging	12	**64**	2
Running	5	0	**61**

TABLE VI
RECOGNITION RATE FOR S2

Activity	Total	Correct	Accuracy
Walking	10	10	100%
Bending	9	9	100%
Jacking	9	9	100%
Jumping	9	9	100%
Wave1	9	9	100%
Wave2	9	9	100%
Total			**100%**

D. Quantitative comparison

A comparison is done with FHMM [12], our previous work [5] and the proposed method in classification task. Basically, HMM are a class of dynamic Bayesian networks where there is a temporal evolution of nodes. In hidden Markov model theory, a HMM model λ is specified by the tuple *(Q,O,A,B,π)* where *Q* is the set of possible state, *O* is the set of observation symbols, *A* is the state transition probability matrix $(a_{ij} = P(q_{t+1} = j|q_t = i))$, *B* is the observation probability distribution $(b_j(k) = P(o_t = k|q_t = j))$ and π is the initial state distribution. Generally, the algorithm used to estimate the HMM model is the well-known Baum-Welch algorithm. It can be regarded as an EM algorithm. However, in FHMM this is replaced by fuzzy C-means clustering. The pre-processing steps of the FHMM was conducted as to [13]. 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 in the data sets. Each model (activity) was trained 50% of randomly selected instances of activities and the best (highest-likelihood) models were kept for comparison as FHMM are known to produce models of varying quality, even when trained repeatedly with the same data. Meanwhile, the QNT is constructed as describe in Section II. 800 particles has been employed to perform the tracking, and the translation, *s* and orientation, *r* in the modified unit circle are chosen as *12* and *16*, respectively.

The comparison of both classification accuracies are provided in Table VII. It is observed that on the two data sets, the performance of the proposed method is much better than the FHMM, however slightly below the QNT. It is worth

pointing out that the QNTs employed in this experiment are constructed from 400 particles with 1% training data while the best HMM were employed for this comparison.

From the analysis of the results, we notice that the effectiveness of FHMM and the proposed method are very much dependant on the accuracy of the training data and the quantity of training data. Meanwhile the QNT is fairly consistent as it is not a statistical learning method thereby does not require large training data. Further looking into the training data achieved, we found out that some of the training data are very sparse. This is one of the possible reason why the QNT performs slightly better than the proposed method. We suggest a better solution such as the active curve axis Gaussian model [18] might alleviate the problem.

However, it is worth pointing out that the proposed method has two advantages: (1) it bridges the low-level vision signal data and high-level vision understanding in human motion analysis systems. (2) a mixture regressor and interpolation method can be employed to convert Gaussian symbols into numerical values. This will provide a two-way connections for human motion representation used for both numerical and symbolic human motion analysis systems.

TABLE VII
COMPARISON WITH QNT [5] AND FHMM

	Proposed Method	QNT [5]	FHMM
S1	81%	85%	78%
S2	100%	100%	100%

V. CONCLUDING REMARKS

There are always two essential parts in human motion recognition: the low-level vision processing and the high-level vision understanding that is based on it. In this paper, we proposed a framework based on Gaussian mixture model-based clustering algorithm and qualitative normalised templates to bridge the low-level vision signal data and high-level vision understanding in human motion analysis systems. Empirically, we have demonstrated that the proposed method produces a very encouraging recognition rate on two public databases. A comparison with the FHMM also shows that our proposed approach is significant in a variety of aspects. Our future work aim at developing a compositional model which uses predominantly knowledge-based techniques to translate between high-level human motion scenarios and the proposed Gaussian mixture models of human motions in this paper.

REFERENCES

[1] D. Ayers and M. Shah. Monitoring human behavior from video taken in office environment. *Image and Vision Computing*, 19(12):833–846, October 2001.

[2] M. Blank, L. Gorelick, E. Shechtman, M. Irani, and R. Basri. Action as space-time shapes. In *Proceedings of IEEE International Conference on Computer Vision*, volume 2, pages 1395–1402, Beijing, China, 2005.

[3] M. Brand, N. Oliver, and A. Pentland. Coupled hidden markov models for complex action recognition. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 994–999, 1997.

[4] S. Calinon, F. Guenter, and A. Billard. On learning, representing and generalizing a task in a humanoid robot. *IEEE Transactions on Systems, Man and Cybernetics, Part B. Special issue on robot learning by observation, demonstration and imitation*, 36(5):1–12, 2006.

[5] C. S. Chan, H. Liu, D. Brown, and N. Kubota. A fuzzy qualitative approach human motion recognition. In *Proceedings of the IEEE International Conference on Fuzzy Systems*, pages 1242–1249, 2008.

[6] J. Gonzàlez, X. Varona, F. Roca, and J. Villanueva. aspaces: Action spaces for recognition and synthesis of human actions. In *Second International Workshop on Articulated Motion and Deformable Objects*, volume 2492 of *Lecture Notes in Computer Science*, pages 942–946. Springer, 2002.

[7] M. Isard and A. Blake. Condensation: Conditional density propagation for visual tracking. *International Journal of Computeter Vision*, 29(1):5–28, 1998.

[8] Y. Ivanov and A. Bobick. Recognition of visual activities and interactions by stochastic parsing. *IEEE Transaction on Pattern Recognition and Machine Intelligence*, 22(8):852–872, 2000.

[9] D. G. Kendall. A survey of the statistical theory of shape. *Statistical Science*, 4(2):87–120, 1989.

[10] H. Liu, D. Brown, and G. Coghill. Fuzzy qualitative robot kinematics. *IEEE Transactions on Fuzzy Systems*, 16(3):808–822, 2007.

[11] H. Liu and G. Coghill. Fuzzy qualitative trigonometry. In *Proceedings of the IEEE International Conference on Systems, Man and Cybernetics*, volume II, pages 1291–1296, Hawaii, USA, 2005.

[12] M. Mohamed and P. Gader. Generalized hidden markov modles. i. theoretical frameworks. *IEEE Transactions on Fuzzy Systems*, 8(1):67–81, 2000.

[13] M. Mohamed and P. Gader. Generalized hidden markov modles. ii. application to handwritten word recognition. *IEEE Transactions on Fuzzy Systems*, 8(1):82–94, 2000.

[14] N. Oliver, B. Rosario, and A. Pentland. A bayesian computer system for modeling human interactions. *IEEE Transactions on Pattern Analysis and Machine Intellingence*, 22(8):831–843, 2000.

[15] C. Schuldt, I. Laptev, and B. Caputo. Recognizing human actions: A local svm approach. In *Proceedings of the International Conference on Pattern Recognition*, volume 3, pages 32–36, Hong Kong, 2004.

[16] N. Vaswani, A. Roy Chowdhury, and R. Chellappa. Activity recognition using the dynamics of the configuration of interacting objects. In *Proceedings of the 2003 Computer Vision and Pattern Recognition*, volume 2, pages 633–640, Madison, Wisconsin, USA, 2003.

[17] A. Wilson and A. Bobick. Parametric hidden markov models for gesture recognition. *IEEE Transaction on Pattern Analysis and Machine Intelligence*, 21(9):884–900, 1999.

[18] B. Zhang, C. Zhang, and X. Yi. Active curve axis gaussian mixture models. *Pattern Recognition*, 38(12):2351–2362, 2005.