Recognition and Re-synthesis of 3D Human Motion with Personalized Variations

Seyed Ali Etemad
Department of Systems and Computer Engineering
Carleton University
Ottawa, Canada
s.a.etemad@gmail.com

Ali Arya
School of Information Technology
Carleton University
Ottawa, Canada
arya@carleton.ca

Abstract— **This paper proposes a 3D view-invariant human action recognition method based on Hidden Markov Models. The natures of the actions, as well as the characteristics of the actors and different performance styles have been successfully recognized. The results have been compared to Nearest Neighbor and Similarity Search based recognition for further evaluation. Also the research addresses the problem of re-synthesis of motion. Transformation of moods, genders, and other characteristics of the actor have successfully been carried out, and the entire action has been re-synthesized for various purposes such as animation.**

Keywords-human action, recognition, hidden Markov models, synthesis, style transformation.

I. INTRODUCTION

Interactive virtual environments are rapidly growing in many applications from arts and entertainment to scientific simulation, education, and the service industry. Very realistic computer games with advanced character and story developments have crossed the boundary between games and movies; virtual social environments such as Second Life now host universities and embassies; and the military training programs use simulations for combat and other situations. Human characters play a significant role in most of these applications, where complex design and powerful multimedia content give the illusion of intelligent computer characters to the users, characters that perform complicated actions, engage in conversation, and display emotions. Historically, character animation is done by the traditional technique of key-framing or recently motion capture for more complicated movements. Regardless of the technique, the majority of animated behaviors in interactive or non-interactive multimedia applications are simply play-back of pre-recorded or pre-animated sequences. This inability of the systems to generate realistic behaviors procedurally limits the applications in terms of their ability to respond to user interactions in real-time by creating appropriate content. Even in non-interactive applications like animated films, there is minimum support from intelligent animation software, and automating the process of character animation will reduce the production cost and effort, and provides animators with a more effective and systematic way to generate content.

The research reported here tries to address the problem of creating personalized animation of human movements with variations such as gender and mood. We divide this problem into two major parts: (1) recognizing basic moves in a series of motion capture data (primary motor themes) and understanding the patterns and parameters that represent personal and style-based variations (secondary motor themes), and (2) reconstructing new behaviors based on desired parameters (personal variations such as gender and mood). The basic tools for the recognition process are Hidden Markov Models. Yet some improvements have been made in the networks to acquire more accurate results. For synthesis we have defined transfer functions corresponding to the variations, and also interpolation between two extreme examples. A series of time warping is critical due to different lengths of the different actions. The training process has been tested with different number of states, as well as iterations, for error minimization. Finally two other methods have been employed for accuracy comparison purpose: Similarity search and Nearest Neighbor. As for synthesis, non-linear time warping, linear interpolation, and some transformation methods were applied. A review of related research can be found in Section II. The following sections will discuss our method and present some experimental results and concluding remarks.

II. RELATED WORK

Many attempts have been made to construct systems, capable of recognizing human actions, regardless of angle of view, gender of the performer, moods and energy levels, clothing, background, and even physical structures of the performer. In [1], Davis et. al. categorizes human action recognition methods into three major fields of: model-free, indirect model, and direct model. Wu et. al. however, in [2], names the three categories as: motion-based, appearance-based, and model-based. While motion-based methods deal with the motion of the body, regardless of the physics of the body, the appearance-based employs the overall form of the body as well as the background, and the respective difference between the two. Finally, the model-based utilizes the appearance as well as the shape, and even other information

978-1-4244-3757-3/09/$25.00 ©2009 IEEE

on the human body and motion [2]. In this study, the model-based method has been used.

HMM has been one of the most employed methods for human action recognition [2, 9, 10, and 11]. In [2], the authors have employed a two-layered HMM system for upper body action recognition. The first layer employs low level features and deals with each arm in separate, while the second layer deals with the relationship of the two arms as a whole. In [9] Li et al have proposed a view invariant 2D action recognition system based on hidden Markov models. Here the key to their view invariant method is the assumption that any action is a linear combination of basis shapes. Wilson et al [10] employ PHMM (Parametric Hidden Markov Models) and train motion recognition models via user-labeled styles. Babu et. al. have proposed a person-independent system based on motion vectors extracted from MPEG video [11].

Finally K-Nearest Neighbors have been employed in [4]. This method requires frontal or lateral views only, and uses the grayscale image sequence of the head for action recognition purpose.

Re-synthesis of motion and stylistic synthesis of movement has not received as much attention as recognition. Matthew Brand et al used entropy minimization methods and Hidden Markov Models for recreation of stylistic motion [5]. They tend to utilize motion capture data and create new styles of previously existing stylistic motion such as dance. In [6], Rose et al exercise linear time warping and basic interpolation and extrapolation for synthesis of existing motion sequences. The key to this approach is the manual selection of key points which determines the span for time warping procedures. Grochow et al propose a novel method called Scaled Gaussian Process Latent Variable Model in [7] which generates most likely poses of a body based on a pre-trained data set. Their system is results in a probability distribution over the span of all probable outcomes. Although this process aims at pose generation, its subsequent function can alternatively result in motion generation. The basic tool of this system is interpolation, with the justification that interpolating different functions does not necessarily result in an interpolation of their styles. In [8] Torresani et al have performed one of the few researches on human motion training and synthesis via Laban principles. Laban movement analysis employs a set of strictly defined perceptual features to describe different styles. The authors utilize motion capture data and interpolate between different movement styles in perceptual space.

The gender, mood and other circumstances of the performers, however has not been subject to research in any of the mentioned literature. Here we aim to tackle this issue as well as recognition of existing captured actions.

III. Data Acquisition and Pre-processing

Motion capture systems provide the data for human movements in 3D space. The measurements that are selected for this research are in the form of angles, meaning each marker is represented by three angles in space with respect to reference point and each frame is represented by a new set of measurements. Since most actions are recognizable from the lower body movement alone, the data which corresponds to that section of the body is extracted from the files. This is done to minimize run time, and to simplify the solutions. Experiment also shows that omitting the upper body does not affect the accuracy of the recognition process.

The rotation values matrix for i^{th} marker for frames 1 to m is as follows, where $\theta_j^{x_i}$ denotes the degree of the x^{th} axis in space, related to i^{th} marker of the j^{th} frame:

$$\overline{\theta}^i = \begin{bmatrix} \theta_1^{x_i}, \theta_1^{y_i}, \theta_1^{z_i} \\ \theta_2^{x_i}, \theta_2^{y_i}, \theta_2^{z_i} \\ \vdots \quad \vdots \quad \vdots \\ \theta_m^{x_i}, \theta_m^{y_i}, \theta_m^{z_i} \end{bmatrix} \tag{1}$$

In (2), the complete angular positioning matrix of m frames and n markers is presented.

$$\Theta = \begin{bmatrix} (\theta_1^{x_1}, \theta_1^{y_1}, \theta_1^{z_1}), (\theta_1^{x_2}, \theta_1^{y_2}, \theta_1^{z_2}), \ldots, (\theta_1^{x_n}, \theta_1^{y_n}, \theta_1^{z_n}) \\ (\theta_2^{x_1}, \theta_2^{y_1}, \theta_2^{z_1}), (\theta_2^{x_2}, \theta_2^{y_2}, \theta_2^{z_2}), \ldots, (\theta_2^{x_n}, \theta_2^{y_n}, \theta_2^{z_n}) \\ \vdots \qquad \vdots \qquad \vdots \qquad \vdots \\ (\theta_m^{x_1}, \theta_m^{y_1}, \theta_m^{z_1}), (\theta_m^{x_2}, \theta_m^{y_2}, \theta_m^{z_2}), \ldots, (\theta_m^{x_n}, \theta_m^{y_n}, \theta_m^{z_n}) \end{bmatrix} \tag{2}$$

The related data for the hip positioning marker for frames 1 to m is shown by $\overline{d}$. In (3) d_j^x represents the x axis of the distance of the hip marker with respect to origin. To build and train the recognition system, the hip positioning values must be stacked, and a complete dataset called $\overline{A}$ be constructed as shown in (4).

$$\overline{d} = \begin{bmatrix} d_1^x, d_1^y, d_1^z \\ d_2^x, d_2^y, d_2^z \\ \vdots \quad \vdots \quad \vdots \\ d_m^x, d_m^y, d_m^z \end{bmatrix} \tag{3}$$

$$\overline{A} = \begin{bmatrix} (d_1^x, d_1^y, d_1^z) & (\theta_1^{x_1}, \theta_1^{y_1}, \theta_1^{z_1}, \ldots, \theta_1^{x_n}, \theta_1^{y_n}, \theta_1^{z_n}) \\ (d_2^x, d_2^y, d_2^z) & (\theta_2^{x_1}, \theta_2^{y_1}, \theta_2^{z_1}, \ldots, \theta_2^{x_n}, \theta_2^{y_n}, \theta_2^{z_n}) \\ \vdots & \vdots \\ (d_m^x, d_m^y, d_m^z) & (\theta_m^{x_1}, \theta_m^{y_1}, \theta_m^{z_1}, \ldots, \theta_m^{x_n}, \theta_m^{y_n}, \theta_m^{z_n}) \end{bmatrix} \tag{4}$$

The relative, scalar values were converted to meaningful vectors by subtracting each frame from its previous one, and dividing by the frame rate fr. This is shown through (5) to (9). $\underline{\theta_i}$ and $\underline{d_i}$ represent the movement vectors in rotation and displacement for the i^{th} frame. $\underline{\overline{A}}$ shows the vectorized format of $\overline{A}$. Finally v_{d_i} and θ_{d_i} are the displacement and angular velocities of the i^{th} frame.

These conversions will provides us with the benefit of dealing with vectors which can be interpreted regardless of their previous states. The direction of each vector can be calculated, and the new values are independent of the performer's physical features such as height and weight.

$$\underline{\theta_i} = (\theta_i^{x_1}, \theta_i^{y_1}, \theta_i^{z_1}, \ldots, \theta_i^{x_n}, \theta_i^{y_n}, \theta_i^{z_n}) - (\theta_{i+1}^{x_1}, \theta_{i+1}^{y_1}, \theta_{i+1}^{z_1}, \ldots, \theta_{i+1}^{x_n}, \theta_{i+1}^{y_n}, \theta_{i+1}^{z_n}) \tag{5}$$

$$\underline{d_i} = (d_i^x, d_i^y, d_i^z) - (d_{i+1}^x, d_{i+1}^y, d_{i+1}^z) \tag{6}$$

$$\overline{\underline{A}} = \begin{bmatrix} \underline{d_1} & , & \underline{\theta_1} \\ \underline{d_2} & , & \underline{\theta_2} \\ \vdots & & \vdots \\ \underline{d_{m-1}} & , & \underline{\theta_{m-1}} \end{bmatrix} \quad (7)$$

$$v_{d_i} = fr.\underline{d_i} \quad (8)$$

$$v_{\theta_i} = fr.\underline{\theta_i} \quad (9)$$

To simplify the calculations the values were finally clustered by a radius of three centimetres/s and degrees/s.

The reason that the 11 markers representing the lower body were only selected is as the followed. From (5) we can conclude:

$$\underline{\theta_{i,j}} = \left(\theta_i^{x_j}, \theta_i^{y_j}, \theta_i^{z_j}\right) - \left(\theta_{i+1}^{x_j}, \theta_{i+1}^{y_j}, \theta_{i+1}^{z_j}\right) \quad (10)$$

Therefore the magnitude of $\underline{\theta_{i,j}}$ shown by $\left|\underline{\theta_{i,j}}\right|$ is calculated by:

$$\left|\theta_{i,j}\right| = \sqrt{\left(\theta_i^{x_j} - \theta_{i+1}^{x_j}\right)^2 + \left(\theta_i^{y_j} - \theta_{i+1}^{y_j}\right)^2 + \left(\theta_i^{z_j} - \theta_{i+1}^{z_j}\right)^2} \quad (11)$$

For continuous data, the following equation would old true:

$$\overline{\underline{\theta_{i,j}}} = \frac{1}{N.\Delta t} \int_{i=1}^{m} \left|\theta_{i,j}\right|.dt \quad (12)$$

Where N is the total number of frames and n is the number of markers. But for the discrete data at hand we would have:

$$\overline{\underline{\theta_{i,j}}} = \frac{1}{N.\Delta t} \sum_{i=1}^{m} \left|\theta_{i,j}\right|.\Delta t \quad (13)$$

$$= \frac{1}{N} \sum_{i=1}^{m} \left|\theta_{i,j}\right| \quad (14)$$

Calculating (14) for both the lower half markers, the upper half marker, and the full body markers reveal that despite having less markers in the lower half of the body, roughly 60 percent of the movements of a whole human body is carried out by the lower half, resulting in enough information to construct a system capable of recognizing actions based on the lower half only. Also due to high correlation between left hand and right foot, right hand and left foot, and central hip and upper body, not only omitting the upper body data will not distress the recognition process, but will even stabilize the process by making it less complicated. As another result of working with the lower body data, the HMM will require less training data.

As the last step for construction of data base, the issue of noise reduction is tackled. Through the motion capture sessions, there exist a variety of noise sources such as misplacement and movement of markers during the process, blind spot positioning of a marker for the cameras, initial calibration of the system, and many more. To reduce the impact of the noise, a smoothing filter in the form of a digital Low Pass Filter (LPF) was utilized. Experimental results confirmed a significant improvement following the application of a LPF.

IV. ACTION RECOGNITION BASED ON HIDDEN MARKOV MODELS

To recognize different actions, performed by different performers, Hidden Markov Models were employed. Hidden Markov Models (HMM) are a form of Bayesian networks. There are a number of hidden states in each HMM network. For a network with *n* hidden states $\{s_1, s_2, \ldots, s_n\}$, the transition probability is $P(s_j(t+1) \mid s_i(t))$. Transition property is defined as the probability of the HMM being in state s_i at (t+1) if it has been in state s_j at (t). Another definition in HMM is emission probability, where it is the probability of the HMM producing a certain symbol (observation) at a certain state [3]. Providing the network with some initial probabilities, the probability of observing a sequence of symbols with a length of *m* is:

$$P(Y) = \sum_X P(Y \mid X).P(X) \quad (15)$$

Where $Y = y(0), y(1), \ldots, y(m-1)$ is the desired sequence. Figure 1 shows a basic form of an HMM network.

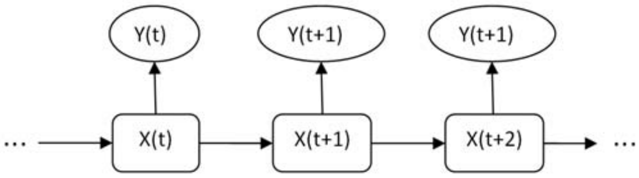

Figure 1. Basic HMM network

For each class of actions and related variations, a different HMM is created which calculates the total emission probability for generating the sequence of feature vectors corresponding to each action. The test data will be employed to the created networks according to the "Recognition Algorithm" presented in the next section, and the HMM resulting in the most likelihood will be determine the class of action that has occurred.

After setting up the HMM networks, the number of hidden states, as well as iterations – in order to reach maximum learning without occurrence of over learning- must be resolved. Figure 2 shows the learning accuracy curve for different number of states in a 50 iteration process. The most accuracy is obtained for 17, 21, 25, 26, and 27 states. Yet the difference between a 17 state and a 27 state hidden Markov model network in this research is insignificant, and therefore to avoid over-learning, 17 states is assigned to the network. In Figure 3 the most efficient number of learning iterations is presented, where 35 iterations show that the log likelihood does not demonstrate significant improvement beyond this point. The completed HMM network is able to recognize the three classes of walking, jumping and running for which it had been trained for, with 100% accuracy.

978-1-4244-3757-3/09/$25.00 ©2009 IEEE

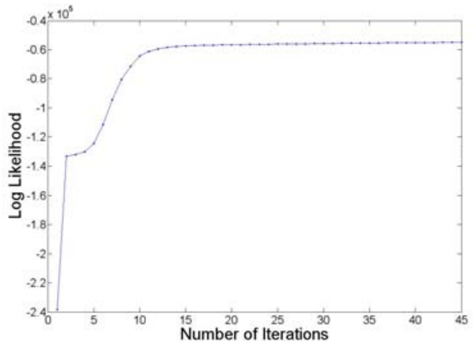

Figure 2. Average Log Likelihood for Different Number of States

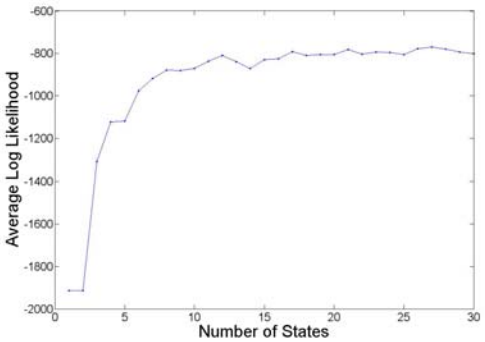

Figure 3. Learning Iterations for 17 states

V. RECOGNITION ALGORITHM

Once a basic HMM network is up and running, the following algorithm is proposed to recognize each movement and its related variation of mood, age and etc. In this algorithm a fine tuning method has been employed to recognize the size of a movement cycle. Not including this section in the algorithm would result in erroneous results, since each HMM is trained using a complete cycle of each action having left intact no extra frames that do not belong to that class of action. The HMMs used in this research show very sensitive to irrelevant and extra frames of data. A hierarchy of Hidden Markov models is presented to classify each class of action and their multiple variations of style, gender, age, and mood. Figure 4 shows the overall appearance and functionality of the recognition process.

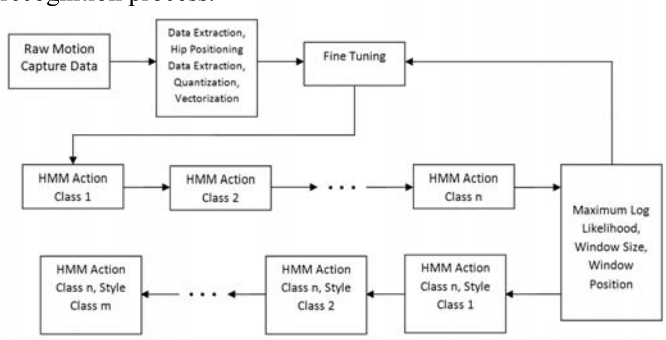

Figure 4. Recognition Process

From the raw motion capture data first the eleven lower body markers are extracted and converted into angular velocities, quantized, and the hip positioning marker data are extracted. All the data are then vectorized in the same fashion that the training data were pre-processed. In this algorithm an HMM network is trained using each class of action data. Sub HMM networks for each style of action is also constructed, i.e. old, young, feminine, masculine, happy, sad, and etc. The test data will provided to all the primary networks, classifying the action into major classes of walking, running, and jumping. Once the major class is distinguished, the data will be fed to classify the respective style of action. The test data is basically an unknown action with an undefined length which is composed of various actions. As mentioned earlier, null frames in the beginning or the end of an action would significantly affect the outcome, therefore a fine-tuning method is proposed. A window with the size of the maximum number of frames used in the training process is placed at the beginning of the test data, decrementing in size and with each decrementing the frames within the window are provided to the first layer of HMM networks. All the resulting log likelihoods and the respective window sizes are stored in an array, for later analysis. The maximum log likelihood, the action class HMM providing the maximum log likelihood, and the respective window size are selected, the window is placed at the end of the previous position of the window, and the same process repeats. Result of this stage is a fractioned matrix, each fraction referring to a class of action. Each new test section is then fed to different style class HMMs of the major class of action, resulting in style, gender, and mood classification.

VI. EXPERIMENT AND RESULTS

The system was constructed based on figure 4. A data set consisting of 7 samples for each act of running, jumping, feminine walking, masculine walking, happy walk, and sad walk were captured. The goal is to determine whether a test action is in the form of walking, jumping or running. Then, if the action were walking, which one of the mentioned characteristics is it more likely to possess. Seven training samples per each class is relatively undersized and while the results were very significantly accurate, it is expected that the results improve as larger data sets be used for training. A test set of 6 actions were used for testing the system. Table 1 shows the results of classifications using this system. Five out of six test samples were classified correctly, resulting in 83.3% accuracy, while happy walk was confused with feminine walk.

TABLE I. TEST SAMPLES AND RESULTS FOR HMM NETWORK

Test Sample	Test Action	Test Style	Class Results	Style Results
1	Run	-	Run	-
2	Jump	-	Jump	-
3	Walk	Feminine	Walk	Feminine
4	Walk	Masculine	Walk	Masculine
5	Walk	Happy	Walk	Feminine
6	Walk	Sad	Walk	Sad

978-1-4244-3757-3/09/$25.00 ©2009 IEEE

Two other methods were also used for comparison to the functionality of the proposed HMM method: Nearest Neighbor and Similarity Search. For both methods to operate properly, the action length of all the members of the training set as well as the test sets must be equal. A basic shrink or stretch was performed to deal with this issue.

The similarity search, as expected, provided inaccurate results, as it does not take into account different styles. The similarity search algorithm basically searches the data base for similar files based on minimizing a difference function and attributes the class and style of the most similar training data as the result. The problem with this method is that first of all, a very huge data set for accurate results, and second, it is not an intelligent method. Table 2 demonstrates the results for the same test set performed on the same training set as the HMM network.

TABLE II. TEST SAMPLES AND RESULTS FOR SIMILARITY SEARCH

Test Sample	Test Action	Test Style	Class Results	Style Results
1	Run	-	Walk	Happy
2	Jump	-	Jump	-
3	Walk	Feminine	Walk	Feminine
4	Walk	Masculine	Run	-
5	Walk	Happy	Run	-
6	Walk	Sad	Walk	Sad

TABLE III. TEST SAMPLES AND RESULTS FOR NEAREST NEIGHBOR

Test Sample	Test Action	Test Style	Class Results	Style Results
1	Run	-	Walk	Happy
2	Jump	-	Jump	-
3	Walk	Feminine	Walk	Feminine
4	Walk	Masculine	Walk	Sad
5	Walk	Happy	Walk	Sad
6	Walk	Sad	Walk	Sad

The deficiency of similarity search especially for style classification is clear according to Table 2. The basic movement classification however, is more reliable than style classification.

The nearest neighbor method is used in some literature for recognizing movements. In [4] for instance, Madabhushi et al have employed this method on extracted feature vectors of the head movement for recognition purpose. They have employed grayscale images of successive frames and were able to recognize up to 12 actions successfully. The range of accuracy varies from 67 to 100 percent. The gender, mood, or style of the actor however, has not been subject to research via this method. Although this method is suitable for small databases [12], the reason for lack of interest in style classification using this method is that different styles of one action are very similar in terms of feature space representation and therefore the calculated Euclidean distances would be very similar. As a result, the style classification accuracy is insignificant. Here we performed this method on the mentioned data base and the results are provided in Table 3.

VII. STYLICTIC ACYION SYNTHESIS

Once the recognition process is successfully complete, the next subject is to synthesize and transform existing captured actions to desired styles. A simple example would be to convert a male walking actor into a female actor, or a happy actor to sad, and etc. Prediction and synthesis of such transformations would be vastly valuable for animation, game, and educational purposes.

Two basic methods were selected for this part of our research: Interpolation and Transform Function Implementation. The problem of unequal lengths of action, as well as different speeds associated with the personal preference and style of an actor, however, requires a comprehensive solution. To tackle this problem, we decided to use nonlinear time warping. This technique would eliminate the element of different actor speeds and it would convert all the captured data into equal-length actions. Each marker was warped separately along its respective markers from other actions.

As explained in previous sections, the data is in the form of angles. A data in the form of Θ shown in (2) is used. In each marker sequence an instance where the vectorized velocity format of Θ is equal to zero, indicates a significant event. This event would mean that a specific marker has changed rotation direction on a specific axis. This event may happen more than once, yet where the largest angle is observed in Θ is the most critical. Figure 5 shows three walking cycles where "X" marks the occurrence time of this event.

The non-linear time warping here would be aimed at transforming the three signals such that the three "X" s would be aligned, and the three signals would be the same length. The result is shown in figure 6.

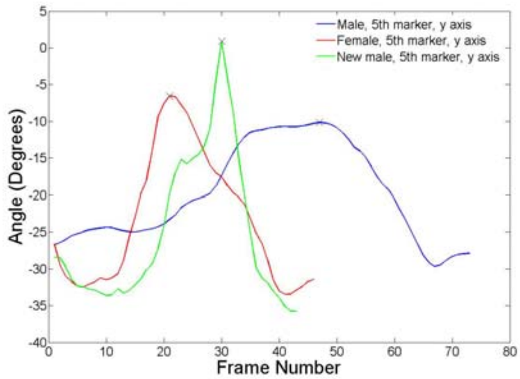

Figure 5. Three different actors walking, X marks the instance where angular direction alters

The resulted signals are suitable for the two mentioned methods. Interpolation between two existing actions results in a third action which is different from both the original actions. An interpolation between a male walking actor and a female walking actor attributed some of the feminine characteristics of the female actor to the male actor. The weight of each initial action could be tuned to create more feminine or more masculine actions. The resulting action contained some noise. This noise is from the same nature of that explained in section 4, therefore a similar filter for noise elimination was applied.

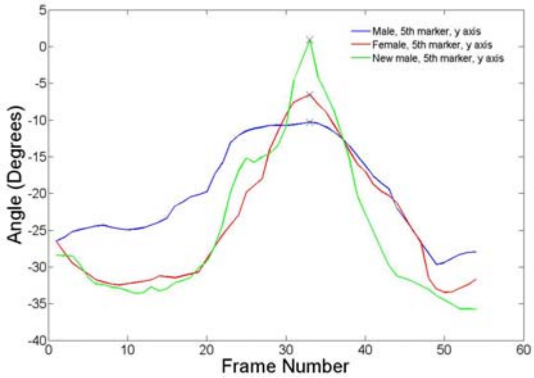

Figure 6. Nonlinear time warped signals

The second technique – Transform Function Implementation was also performed successfully. In this method the transform function $H_{i,j}(m)$ of action i to action j was formed. Action i can be any action such as a male walk, where action j can be another action such as a female walk. The resulting transform function was filtered for noise elimination. The outcome was then applied to action k which is a new action. The expected result for k being a new male walk is a prediction of k walking if he were a female, which was met with a respectively high accuracy. Figure 7 shows the diagram for this method.

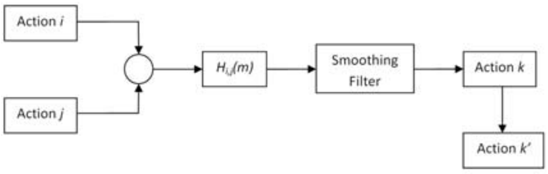

Figure 7. Using transform function for style synthesis

The original masculine walk is presented in figure 8. Figure 9 shows the converted male-to-female action using interpolation while figure 10 shows the same action recreated using the transfer function method. Both visual and correlation analysis of the output confirms the significance of the two methods.

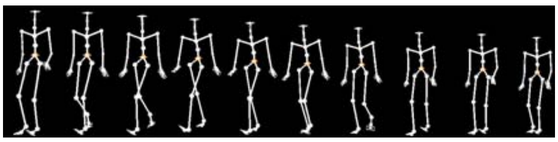

Figure 8. Original male walk

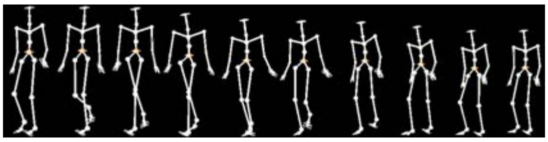

Figure 9. Female walk re-created using interpolation

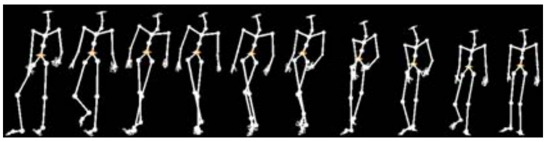

Figure 10. Female walk re-created using transfer function method

VIII. Conclusion

In this paper, we proposed a recognition method for human actions, based on HMM. The method proved to be significantly more accurate when compared to NN and similarity search. The input data extracted from motion capture files were altered and manipulated to construct the input data. The classes of action, as well as the style class were both subject to recognition.

As the second section of this research, the aim of altering the style of various actions was performed successfully. The key to this transformation lay in the non-linear time warping technique developed here. We were able to convert different styles such as feminine, masculine, happy, sad and other style to on another prior to time warping and using two different methods of interpolating between styles, as well as constructing the necessary transform functions.

References

[1] A. F. Bobick and J. W. Davis, "The recognition of human movement using temporal templates," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 23, Issue 3, pp. 257–267, March 2001.

[2] Y. C. Wu, H. S. Chen, W. J. Tsai, S. Y. Lee, and J. Y. Yu, "Human action recognition based on layered-HMM," 2008 IEEE International Conference on Multimedia and Expo, pp. 1453–1456, June 2008.

[3] L. J. Luotsinen, H. Fernlund, and L. Boloni, "Teamwork recognition of embodied agents with hidden Markov models," 2007 IEEE International Conference on Intelligent Computer Communication and Processing, pp. 33–40, September 2007.

[4] A. Madabhushi and J. K. Aggarwal, "Using head movement to recognize activity," Proceedings of the 15th International Conference Pattern Recognition, vol. 4, pp. 698–701, September 2000.

[5] M. Brand and A. Hertzmann, "Style machines," Proceedings of the 27th Annual Conference on Computer Graphics and Interactive Techniques, pp. 183–192, 2000.

[6] C. Rose, M. F. Cohen, and B. Bodenheimer, "Verbs and Adverbs: Multidimensional Motion Interpolation," 1998 IEEE conference on Computer Graphics and Applications, vol. 18, Issue 5, pp. 32–40, September 1998.

[7] K. Grochow, S. L. Martin, A. Hertzmann, and Z. Popović, "Style-based inverse kinematics," Special Issue: Proceedings of the 2004 SIGGRAPH Conference,vol. 23, Issue 3, pp. 522–531, August 2004.

[8] L. Torresani, P. Hackney, and C. Bregler, "Learning Motion Style Synthesis from Perceptual Observations," Advances in Neural Information Processing Systems, no. 19, pp. 1393–1400, 2006.

[9] X. Li and K. Fukui, "View Invariant Human Action Recognition Based on Factorization and HMMs," IEICE Transactions on Information and Systems, pp. 1848–1854, July 2008.

[10] A. D. Wilson and A. F. Bobick, "Parametric hidden Markov models for gesture recognition," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 21, Issue 9, pp. 884–900, September 1999.

[11] R. Venkatesh Babu, B. Anantharaman, K. R. Ramakrishnan, and S. H. Srinivasan, "Compressed domain action classification using HMM," 2001 IEEE Workshop on Content-Based Access of Image and Video Libraries, pp. 44–49, December 2001.