Jurnal
Teknologi

Full Paper

DANCE MODELLING, LEARNING AND RECOGNITION SYSTEM OF ACEH TRADITIONAL DANCE BASED ON HIDDEN MARKOV MODEL

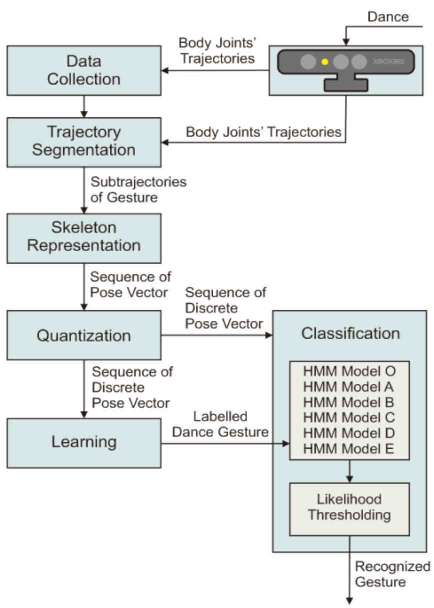
Nurfitri Anbarsanti, Ary S. Prihatmanto*

School of Electrical Engineering and Informatics, Institut Teknologi Bandung, Bandung, Indonesia

Article history
Received
15 June 2015
Received in revised form
1 October 2015
Accepted
13 October 2015

*Corresponding author
anbarsanti@yahoo.com

Graphical abstract



Abstract

The whole dance of Likok Pulo are modeled by hidden markov model. Dance gestures are cast as hidden discrete states and phrase as a sequence of gestures. For robustness under noisy input of Kinect sensor, an angular representation of the skeleton is designed. A pose of dance is defined by this angular skeleton representation which has been quantified based on range of movement. One unique gesture of dance is defined by sequence of pose and learned and classified by HMM model. Six of dance's gesture classes from the phrase "Assalamualaikum" has been trained with hundreds of gesture instances recorded by the Kinect sensor which performed by three of subjects for each gesture class. The classifier system classify the input testing gesture into one of six classes of predefined gesture or one class of undefined gesture. The classifier system has an accuracy of 94.87% for single gesture.

Keywords: Angular skeletal representation, Kinect, dance modelling, dance recognition, gesture recognition, hidden markov model, Likok Pulo dance

1.0 INTRODUCTION

Culture is the identity of a nation. Globalization that hit today's youth, lead the youth generation lose their identity and the value of good wisdom Indonesian culture. If the Indonesian culture is not well maintained, it will disappear in the next few decades. Modeling, learning and classification system of Likok Pulo Dance from Aceh are designed, as one of the initial steps to 'preserve' Indonesian traditional dance into digital form.

Dance is sequence of expressive human body movement and has aesthetic values. Famous Indonesian traditional dance which is known as unique and attractive is Aceh traditional dance such as Likok Pulo. In this study, we choose Likok Pulo dance to communicate with the computer because Likok Pulo dance from Aceh requires synchronous motion among the group of dancers with lined up formation, precision timing of gestures with the rhythm of its music that linearly changed more rapidly. It has several gestures performed repeatedly; and it well accepted in the international environment but still has its strong and decent identity.

Kinect is the Natural User Interface that combines stereoscopic camera and infrared sensors, so it can capture the depth map at a rate about 30 frames per second to estimate the position of the 20 points on the user's skeleton joints. The human body movement such as dance, can be interpreted as a command to communicate with the computer.

Skeleton joint trajectories captured by Kinect sensor is very likely to experience a discontinuity, noise, or instable parameter [7]. Dance movements that involve a lot of body articulation will result in a very large input dimension for the signal trajectories processing systems. So it is necessary to build a representation to reduce the signal entropy and the dimension of data. It must also deal with changes in the dancer's position and orientation relative to the Kinect sensor.

Dance can be defined as sequences of several finite distinct gestures. Gesture transition are not deterministic but probabilistic, due to the unideal dance in real world. Gesture has two aspects of signal characteristics : spatio-temporal variability and segmentation ambiguity [3]. Major approaches for analyzing spatial and temporal patterns include Dynamic Time Warping (DTW), Neural Networks (NNs), dan Hidden Markov Model (HMM) [3]. In this study, HMM-based approach is chosen to model the dance gesture of Likok Pulo dance, because it can be applied to analyzing time-series with spatio-temporal varibilities and can handle undefined patterns [3]. HMM-based dance gesture modelling make us enable to build practical systems that has ability to learn, predict, and classify dance gestures of Likok Pulo Dance from Aceh.

2.0 BACKGROUND

3.1 Related Works

Dance choreography has been captured using various formalization approaces, e.g., Laban notation which is initiated in the early 20 th century. Amy Laviers modelled the motion patterns of ballet as a series or event-driven poses that takes the form of a finite automaton [5]. For a system involving two legs without violating the laws of physics or the rules of ballet, it take the Cartesian composition. Amy Laviers also built automatic generation of Ballet phrases using Linear Temporal Logic and Computation Tree Logic as rich motion specification languages for robots' movements [6]. Yaya Heryadi [10] built a syntactical modeling and classification for performance evaluation of Bali traditional dance, adapting the model of skeleton feature descriptor from Michalis Raptis [7]. Dance's pose is represented by spherical coordinate parameter (θ, φ) from several skeleton joints that is clustered as torso frame, first-degree joints, and second-degree joints.

To reduce the dimension of the data of the skeleton joint signal trajectory from the Kinect sensor before processing it with HMM, there are several methods as follows: mapping joint position in 3D space [11]; grouping joint trajectory relative with K-means clustering [8]; segmenting joint trajectory by the sound of footsteps [13]; using joint angle and joint angular velocity [12]; and using skeleton descriptor [10].

3.2 Hidden Markov Model

Hidden markov model is Markov model with a case where the observation is a probabilistic function of the state. The resulting model (which is called hidden Markov model) is a doubly embedded stochastic process with an underlying stochastic process that is not observable (it is hidden), but can only be observed through another set of stochastic processes that produce the sequence of observations. [2]

A formal characterization of HMM is as follows :

- $S = \{S_1, S_2, S_3, \dots, S_N\}$ -- A set of N states. The state at time t is denoted by q_t.
- $V = \{v_1, v_2, v_3, \cdots v_M\}$ -- A set of Mdistinct observation symbols. The observation at time t is denoted by the variable O_t. The observation symbols correspond to the physical output of the system being modeled.
- $A = \{a_{ij}\}$ -- An $N \times N$ matrix for the state transition probability distribution where a_{ij} is the probability of making a transition from state S_i to S_j :
- $a_{ij} = P[q_t = S_j | q_{t-1} = S_i], \quad 1 \le i, j \le N$
- $B = \{b_j(k)\}$ -- An $N \times M$ matrix for the observation symbol probability distributions where $b_j(k)$ is the probability of emitting v_k at time t in state S_j :
- $b_j(k) = P(O_t = v_k | q_t = S_j), \quad 1 \le j \le N, 1 \le k \le M$
- $\pi = \{\pi_i\}$ -- The initial state distribution where π_i is the probability that the state S_i is the initial state :
- $\pi_i = P[q_1 = S_i], \quad 1 \le i \le N$

Probabilistic notation $A, B,$ and π must satisfy stochastic constraints as follows :

- $\sum_j a_{ij} = 1, \forall i,$ and $a_{ij} \ge 0.$
- $\sum_k b_j(k) = 1, \forall j,$ and $b_j(k) \ge 0.$
- $\sum_i \pi_i = 1,$ and $\pi_i \ge 0.$

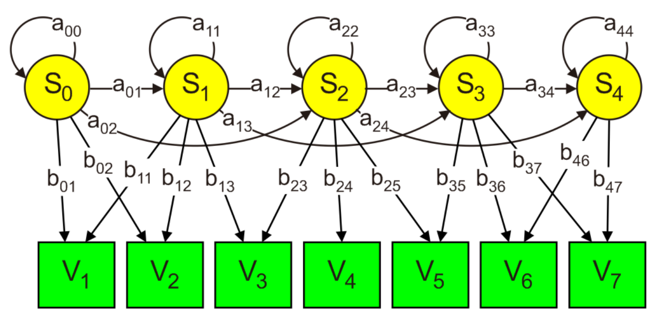


Figure 1 Graphical model of left-right discrete hidden Markov model

An compact notation $\lambda = (A,\ B,\ \pi)$ is used which includes only probabilistic parameters.

The left-right model as shown in Figure 1. It is good for modelling order-constrained time-series whose properties sequentially change over time [3].

3.0 MODELLING THE WHOLE DANCE AND THE DANCE GESTURE

3.1 Modelling The Whole Dance

In this study, some terminologies are used as follows (illustrated in Figure 2):

- Pose – Static configuration of human body, without any movement.
- Gesture – Dynamic movement of human body, which is sequence of poses.
- Phrase – Fragment of choreography which consist of sequence of gestures. The same gestures may be repeated.
- Dance – The whole choreography of a dance from the start to the end, which consist of sequence of phrases.

The whole dance of Likok Pulo dance is modelled as follows :

$$\mathcal{L} = (S, I, P, O, f, e, s_0, S_t)$$

- S -- the finite nonempty set of hidden states. The states correspond to gestures. Its segmentations are determined by the dance expert.
- I -- the finite nonempty set of input.
- P -- the vocabulary of all possible discrete pose of dance.
- O – the finite nonempty set of output, where $O = \{o_1, o_2, \dots, o_T\}$, $o_i \in P^*$, $i \in \{1,2,\dots,T\}$. P^* is the Klenee closure of P, the set consisting of concatenations of arbitrarily many string of element from P (pose). Output O corresponds to gesture trajectories, or its features.
- f – state transition function $f: S \times I \to S$. State transition corresponds to gesture transitions, which for $\forall s \in S$ and $\forall x, y \in I$, satisfies $f(s, xy) = f(f(s, x), y)$ and $f(s, \varepsilon) = s$, where ε is empty transition.

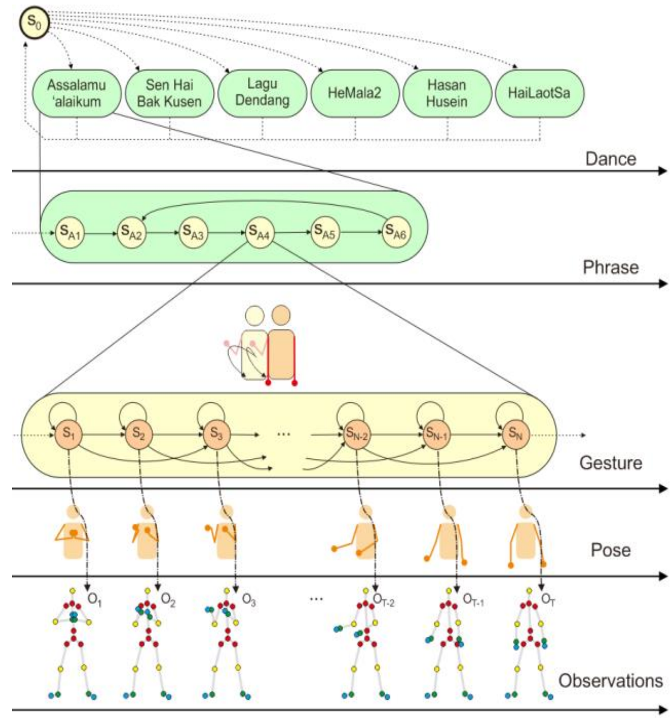


Figure 2 Hierarchy of the Dance

- e – the output map $e : S \times I \to O$.
- s_0 -- initial state, $s_0 \in S$. Initial state corresponds to initial pose or initial gesture of all phrases of Likok Pulo.
- S_t – set of final (or accepting) states, $S_t \subseteq S$. Final states correspond to the end of the phrase.

The model for "Kisah Hasan Husein" phrase and "Assalamualaikum" phrase are illustrated in Figure 3 and Figure 4. Initial states are indicated by using bold circles. Final states are indicated by using double circles. Actually the whole dance has 6-8 phrases.

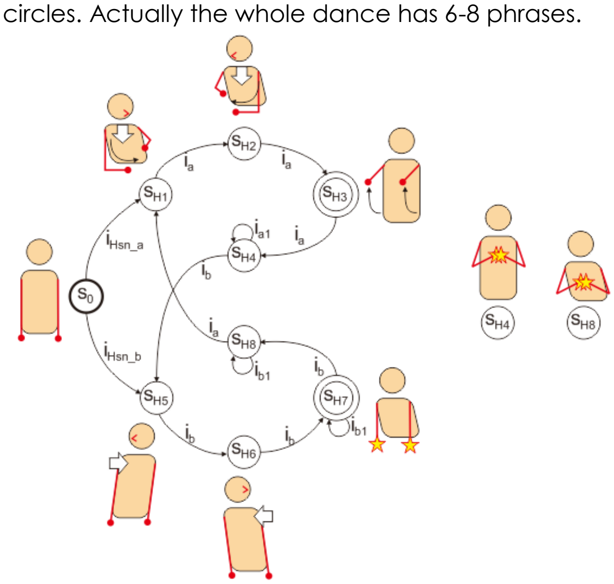


Figure 3 Model for Kisah Hasan Husein Phrase

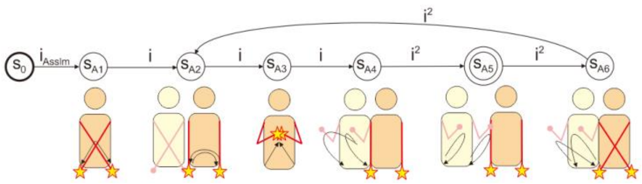


Figure 4 Model for "Assalamualaikum" Phrase

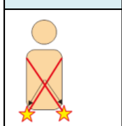
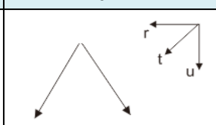
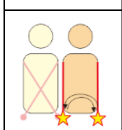
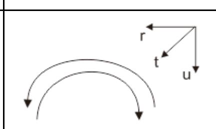
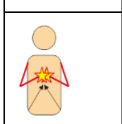
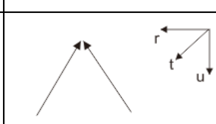
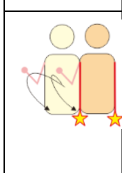
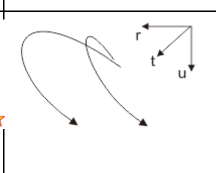
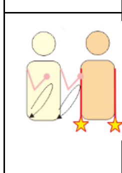
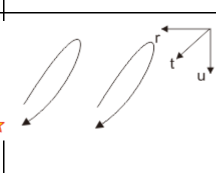
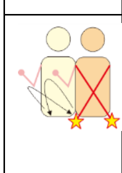
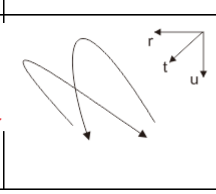
3.2 Modelling The Dance Gestures in HMM

As illustrated in Figure 2, the hidden states $S = \{S_1, S_2, \ldots, S_N\}$ correspond to the pose. The observations symbols $V = \{v_1, v_2, \cdots v_M\}$ correspond to physical output at the system, i.e., the discrete pose vector $P_{u,2}$ (will be explained at chapter IV). Matrix $A = \{a_{ij}\}$ corresponds to transition probability distribution between the gestures S_i. Matrix $B = \{b_j(k)\}$ corresponds to observation symbol probability distribution of discrete vector pose v_i. Matrix $\pi = \{\pi_i\}$, corresponds to initial gesture distribution.

4.0 HMM-BASED DANCE GESTURE LEARNING AND CLASSIFICATION SYSTEM

4.1 Definition of the Dance Gesture Classes

Table 1 Dance Gestures used in this study

Gesture	Wrist Trajectories	Class	Description
		O	"Clapping the hand in front of the chest" to "crossing the hand over the thigh".
		A	"Crossing hand over the thigh" to "straightening the hand over the thigh".
		B	"Straightening the hand over the thigh" to "clapping the hand in front of the chest".
		C	"Clapping the hand in front of the chest" to "swinging the hand to the rightside" to "straightening the hand over the thigh".
		D	"Straightening the hand over the thigh" to "swinging the hand to the leftside" to "straightening the hand over the thigh".
		E	"Straightening the hand over the thigh" to "swinging the hand to the rightside" to "crossing the hand over the thigh".

The gesture classes used in this study are 6 dance gesture elements of Likok Pulo traditional dance on "Assalamualaikum" phrase (Table 1). $f(s_0, i) = O$, $f(O, i) = A$, $f(A, i) = B$, $f(B, i) = C$, $f(C, i^2) = D$, $f(D, i^2) = E$, $f(E, i^2) = A$. Sequence of dance gestures is recognized as "Assalamualaikum" phrase if it is $\{s_0 O(ABCDE)^7 ABCD\}$.

4.2 Sensing System Environment

XBOX Kinect sensor is used to capture the skeleton joints at about 30 fps. The system is implemented in MATLAB and Simulink Environment. The code to trigger the Kinect and its interface are written in C++ and compiled using mex compiler.

4.3 Skeleton Representation

The skeleton representation must satisfy these objectives [7]: (1) Robust coordinate system based on human body orientation, so that the skeleton representation does not depend to the position of the Kinect sensor. (2) Continuity and stability of the signal. (3) Reduce the dimension of the signal while maintaining the character of the motion.

Torso PCA Frame

The joints of the human torso (defined by red skeletal nodes in Figure 5) rarely exhibit strong independent motion with large angle. Due to the strong noise in the depth sensing system, individual torso points, in particular shoulder and hips, may exhibit unrealistic motion that it would like to be limited. Therefore, the torso can be considered as a rigid body which provides 3D orthonormal basis will be used as reference frame for the remaining joints.

Its principal components as follows : $\vec{u}$, the vector with the direction out of the upper to the lower (in most dancing, the player's torso will never stand upside-down relative to the sensor); $\vec{r}$, the vector with the direction out of the right body to the left side of the body; $\vec{t}$, is the cross product of two principal components, $\vec{t} = \vec{u} \times \vec{r}$.

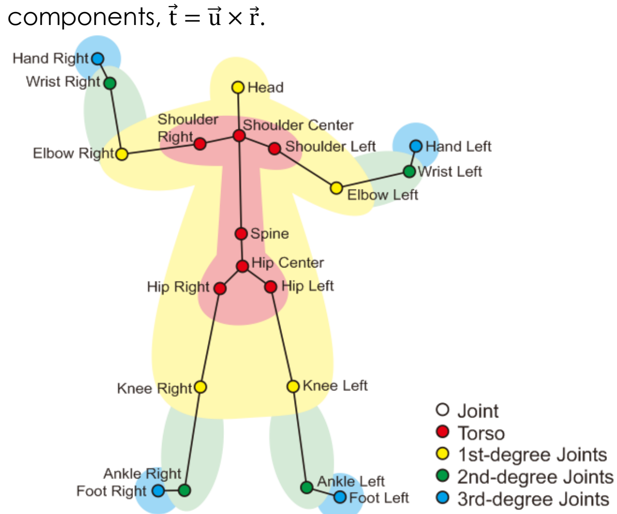


Figure 3 Hierarchy of skeleton joints

First-Degree Joints

First-degree joints (defined by yellow skeletal nodes in Figure 5) are represented relative to the adjacent joint in the torso in a coordinate system derived from torso PCA frame as illustrated in Figure 6(a). The torso PCA frame is translated to RS (right shoulder) and construct spherical coordinate system such that the origin is RS, its azimuth axis is $\vec{u}$ and its zenith axis is $\vec{r}$.

Azimuth φ is the angle between $\vec{u}$ and $\overrightarrow{(\mathrm{RS}, \mathrm{RE_P})}$ where $\mathrm{RE_P}$ is the projection of RE onto the plane whose normal is $\vec{r}$. Elevation θ is the angle between $\overrightarrow{(\mathrm{RS}, \mathrm{RE_P})}$ and $\overrightarrow{(\mathrm{RS}, \mathrm{RE})}$. Then each first-degree joint is represented with two angles (θ, φ). Angular representation for RS is $\{\mathrm{RS_\theta}, \mathrm{RS_\varphi}\}$.

Second-Degree Joints

Second-Degree joints (defined by green skeletal nodes in Figure 5) are represented relative to the adjacent joint in the first-degree joints in a coordinate system $\{\mathrm{u_p}, \mathrm{r_p}, \mathrm{t_p}\}$ which is derived from rotationed torso PCA frame $\{\vec{u}, \vec{r}, \vec{t}\}$ by angle $\{\mathrm{RS_\theta}, \mathrm{RS_\varphi}\}$ as illustrated in Figure 6(b). The vector $\overrightarrow{\mathrm{u_p}}$ protuding out of the vector $\overrightarrow{(\mathrm{RS}, \mathrm{RE})}$. The vector $\overrightarrow{\mathrm{u_p}}$ be a zenith axis of the spherical coordinate system with origin RE. The azimuth axis is $\overrightarrow{\mathrm{r_p}}$ and the zenith axis is $\overrightarrow{\mathrm{u_p}}$. Each second-degree joint is represented with two angles (θ, φ). Angular representation for RE is $\{\mathrm{RE_\theta}, \mathrm{RE_\varphi}\}$. Knee joint is represented by one angle θ.

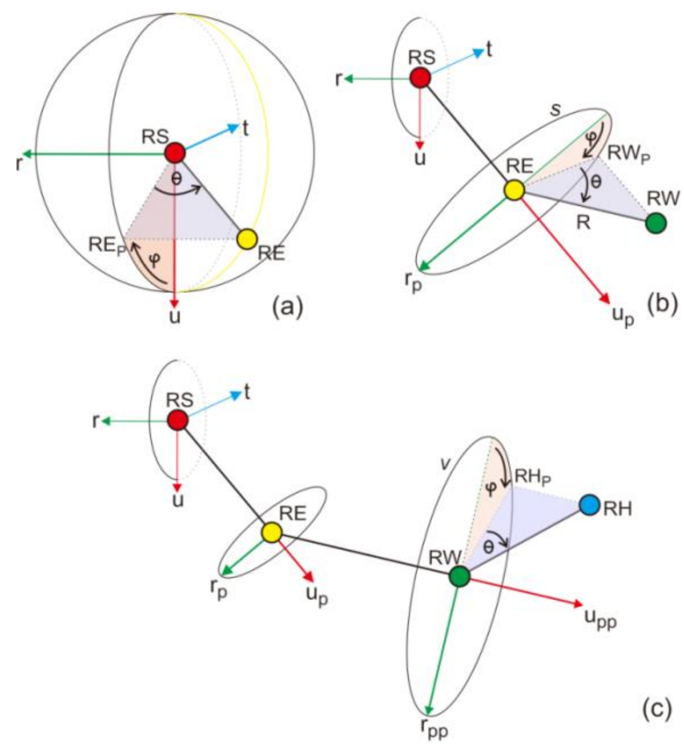


Figure 4 Spherical coordinate system for (a) first-degree joints, (b) second-degree joints, (c) third-degree joints

Third-Degree Joints

Third-Degree joints (defined by blue skeletal nodes n Figure 5) are represented relative to the adjacent joint in the second-degree joints in a coordinate system $\{\mathrm{u_{pp}}, \mathrm{r_{pp}}, \mathrm{t_{pp}}\}$ which is derived from rotationed frame $\{\mathrm{u_p}, \mathrm{r_p}, \mathrm{t_p}\}$ by angle $\{\mathrm{RE_\theta}, \mathrm{RE_\varphi}\}$ as illustrated in Figure 6(c). The vector $\overrightarrow{\mathrm{u_{pp}}}$ protuding out of the vector $\overrightarrow{(\mathrm{RE}, \mathrm{RW})}$. The vector $\overrightarrow{\mathrm{u_{pp}}}$ be a zenith axis of the spherical coordinate system with origin RW. The azimuth axis is $\overrightarrow{\mathrm{r_{pp}}}$ and the zenith axis is $\overrightarrow{\mathrm{u_{pp}}}$. Each third-degree joint is represented with two angles (θ, φ). Angular representation for RW is $\{\mathrm{RW_\theta}, \mathrm{RW_\varphi}\}$.

It is needed to use Wearable Inertial Measurement Units (WIMU) [9] to obtain accurate angles at third-degree joints because Kinect sensor can not detect third-degree joints orientation and position accurately.

Human Pose Representation

For the scope of body poses which involves up to second-degree joints,

- Upper body poses are represented by an 8-tuple $\mathrm{P_{u,2}} = (\mathrm{LE_\varphi}, \mathrm{LE_\theta}, \mathrm{LS_\varphi}, \mathrm{LS_\theta}, \mathrm{RS_\theta}, \mathrm{RS_\varphi}, \mathrm{RE_\theta}, \mathrm{RE_\varphi})$.
- Lower body poses are represented by the 6-tuple $\mathrm{P_{l,2}} = (\mathrm{LK_\theta}, \mathrm{LH_\varphi}, \mathrm{LH_\theta}, \mathrm{RH_\theta}, \mathrm{RH_\varphi}, \mathrm{RK_\theta})$.

For the scope of body poses which involves up to third-degree joints,

- Upper body poses are represented by an 12-tuple $\mathrm{P_{u,3}} = \begin{pmatrix} \mathrm{LW_\varphi}, \mathrm{LW_\theta}, \mathrm{LE_\varphi}, \mathrm{LE_\theta}, \mathrm{LS_\varphi}, \mathrm{LS_\theta}, \ldots \\ \mathrm{RS_\theta}, \mathrm{RS_\varphi}, \mathrm{RE_\theta}, \mathrm{RE_\varphi}, \mathrm{RW_\theta}, \mathrm{RW_\varphi} \end{pmatrix}$
- Lower body poses are represented by the 12-tuple $\mathrm{P_{l,3}} = \begin{pmatrix} \mathrm{LA_\phi}, \mathrm{LA_\theta}, \mathrm{LK_\theta}, \mathrm{LH_\phi}, \mathrm{LH_\varphi}, \mathrm{LH_\theta}, \ldots \\ \mathrm{RH_\theta}, \mathrm{RH_\varphi}, \mathrm{RH_\phi}, \mathrm{RK_\theta}, \mathrm{RA_\theta}, \mathrm{RA_\phi} \end{pmatrix}$
- Head poses are represented by the 3-tuple $\mathrm{H} = (\mathrm{H_\varphi}, \mathrm{H_\theta}, \mathrm{H_\phi})$

$\mathrm{P_{u,2}}$ will be used for implementing HMM-based dance learning and classification.

4.4 Data Collection and Segmentation

Each of three subjects performing 'Assalamualaikum' dance phrase multiple times. It has been collected 2169 isolated dance gestures data from three subjects which are classified to 6 sets of data for each of dance gesture classes (Table 2). Total isolated data are partitioned into 80% training data and 20% test data.

Segmentation between gestures in a continuous joint trajectory signal is done in two ways : (1) By detecting hand's clap by clap sensors in the smartgloves and sent by bluetooth module. (2) By using a time window. The system inform the performing subject to do the gesture with limited time. The border between the gestures are identified through timestamps.

Table 2 Isolated dance gesture pattern for the HMM Training and Testing

Gesture	Only Kinect	W/Clap Sensors	Total Data	80% Training Data	20% Test Data
O	387	0	387	310	77
A	394	71	465	372	93
B	415	92	507	406	101
C	488	127	615	492	123
D	415	131	546	437	109
E	386	113	499	400	99

4.5 The System Block Diagram

Figure 7 shows the block diagram with major component designed for this study. The system acquire the dance gesture from Kinect sensor, collect the body joints' trajectories, segment and fetch the principal component that is human pose vector. The system quantize the human pose vectors and the HMM model learn from training data and recognize from testing data the dance gesture from quantized pose vector.

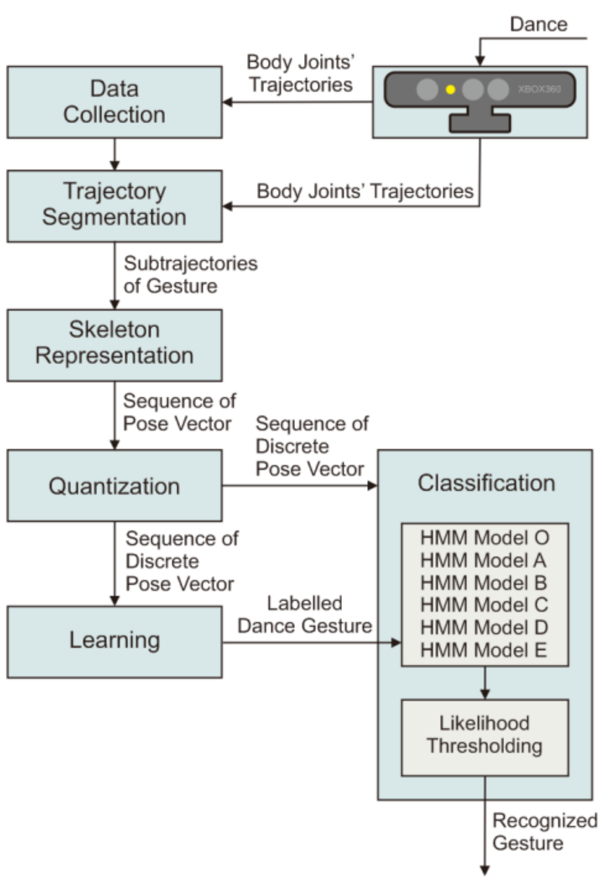
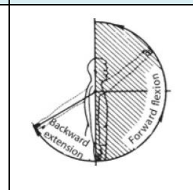
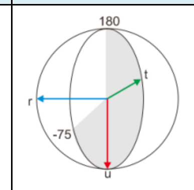
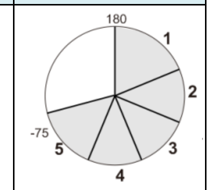
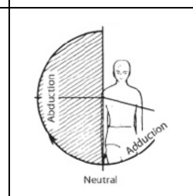
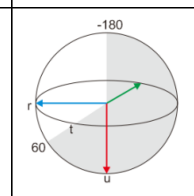
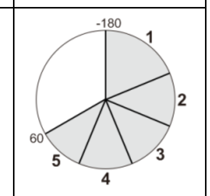
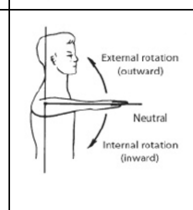
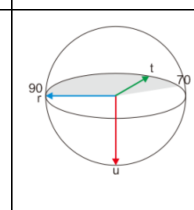
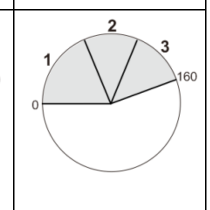
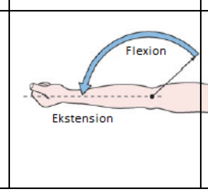
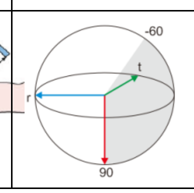
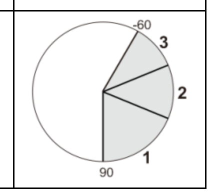


Figure 5 System block diagram with major components

4.6 Pose Vector Quantization based on Range of Movement

Table 3 Directional codewords for each joint angle based on Range of Movement (ROM)

Joint Angle	Physical Representation	Range of Joint Angle	Directional Codewords
RS_φ			
RS_θ			
RE_φ			
RE_θ			

Isolated (segmented) joint trajectory signals are pre-processed using skeleton representation and gives $P_{u,2} = (RS_\varphi, RS_\theta, RE_\varphi, RE_\theta, LS_\varphi, LS_\theta, LE_\varphi, LE_\theta)$, that is dance's pose vector. The combination of all possible values of the pose vector elements is infinite. For discrete HMM-based approach, each element of pose vector should be converted to one of the 3 or 5 directional codewords, based on range of movement [1], [4] as in Table 3. All possible pose's configuration involving up to second-degree joints on one arm are 225 poses. All possible pose's configuration on two arms are $225 \times 225 = 50625$ poses.

4.7 Dance Gesture Learning

The parameters of each HMM models estimated using the Baum Welch algorithm iteratively. Training likelihood curves generally appeared to be stable after 25 cycles, but has not really converge until approximately 70 cycles. The training stops after 100 cycles. The number of states in gesture models ranges from three to five, depending on the complexity of the gesture shape. Increasing the number of hidden states may lower down the recognition rate.

4.8 Dance Gesture Classification

The trained HMM models is used to classify the test data which is different from the training data. Test data which is obtained from 20% of the total data collection are used to test each of HMM models. Classification is done by using a score value P(O | λ) to assess the likelihood (degree of match) between the input test gesture and gesture models. Score computation is done with the forward probabilities:

$$\text{Scoring} = P(O|\lambda) = \sum_{i=1}^{N} \alpha_T(i)$$

$\alpha_t(i)$ = Forward probabilities.

One input test datum of dance gesture tested by six trained HMM model to find the one model that reflects the highest likelihood. If the value of its maximum likelihood pass the predefined threshold value for that model, then the test datum is classified as that model. If the log-likelihood is minus infinity (has no likelihood at all), then the gesture is not classified to any gestures. Threshold of each model is the minimum of the scores of tested training data. The results of classification system is shown in Table 4 and Figure 8.

Table 4 The classification results of dance gesture

Gesture	Detected gesture pattern by the model (%)						Undetected
	O	A	B	C	D	E	
O	96.64	1.29	0.00	0.00	0.00	0.00	2.07
A	0.22	95.27	0.22	0.00	0.43	0.00	3.87
B	0.79	0.00	93.69	1.18	0.99	0.20	3.16
C	0.00	0.16	0.00	93.33	0.00	0.81	5.69
D	0.00	0.00	0.00	0.00	96.52	0.18	3.30
E	0.00	0.20	0.00	0.00	0.00	93.79	6.01

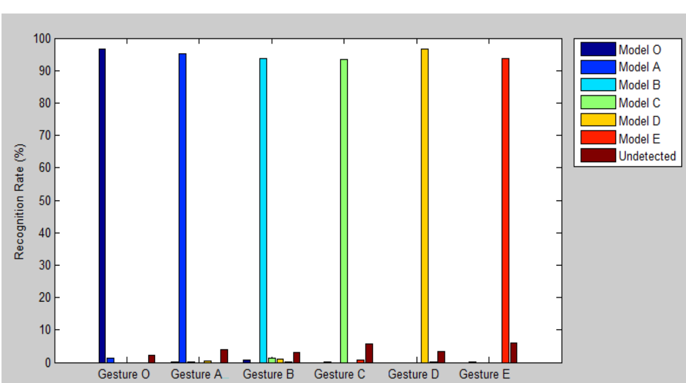


Figure 6 Bar chart of classification results of dance gesture

The system was implemented using the Matlab and Simulink programming package. The implementation of Likok Pulo Dance's Gesture Recognition on Matlab Environment is shown in Figure 9 (when gesture detected) and Figure 10 (when gesture undetected).

The using of angular skeleton sepresentation can reduce data dimension from 33-tuple data to 8-tuple data. The accuration has been also increased : the gestures that is detected as false are decreased from 10% to 0.22%; the dance gestures that is detected as true are increased from 80.7% to 94.87%. This result is shown in Table 5.

Figure 7 Realtime implementation, gesture O detected

Figure 8 Realtime implementation, gesture O undetected

Table 5 Effect of Using Skeleton Representation and Maximum Likelihood

	Before	After
Data Dimension	33-tuple	8-tuple
Detected as False	10 %	0.22 %
Detected as True	80.7 %	94.87 %

4.9 Dance Phrase Recognition

Twenty of complete recording data of "Assalamualaikum" dance phrase are collected. Each gesture of the phrase is segmented using time window given by the system. The complete phrases are tested with each trained HMM models. The results are shown in Table 6. Results of 20 complete phrases compared to ideal phrase are shown in Figure 11.

Table 6 Testing results of 20 complete phrases

Phrase	True Gesture	False Gesture	Gesture Recognition Rate (%)	Complete Phrase Detected
1	37	3	92.50	No
2	39	1	97.50	No
3	39	1	97.50	No
4	40	0	100.00	Yes
5	40	0	100.00	Yes
6	39	1	97.50	No
7	40	0	100.00	Yes
8	39	1	97.50	No
9	40	0	100.00	Yes
10	40	0	100.00	Yes
11	39	1	97.50	No
12	40	0	100.00	Yes
13	40	0	100.00	Yes
14	40	0	100.00	Yes
15	37	3	92.50	No
16	40	0	100.00	Yes
17	39	1	97.50	No
18	40	0	100.00	Yes
19	40	0	100.00	Yes
20	37	3	92.50	No
Mean			98.125 %	55 %

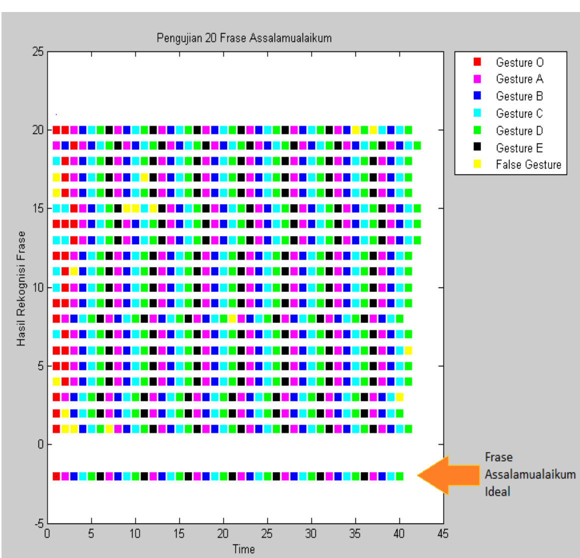


Figure 9 Testing results of 20 complete phrases

5.0 CONCLUSION

Hidden Markov model can be used to model the whole dance; dance gestures cast as hidden discrete states and phrase as a sequence of gestures.

Skeleton representation that is quantized based on range of movement can effectively handle noisy joint trajectory data, reduce the data dimension, and handle the change of position and orientation of user relative to the Kinect sensor.

HMM are an effective and efficient method of both learning and classifying dance gestures involving several joints.

These are suggestion for future work : (1) Observation of the dance can be expanded up to lower body, and/or expanded to third-degree joints. It is required additional inertial sensors for capturing position and orientation of third-degree joints (palm hands and foots) due to Kinect sensor can't detect it. (2) Skeleton representation can be deepened to consider the dynamic aspects of the human body; (3) Skeleton representation can be deepened to also consider the dynamic aspects of the human body; (4) For wider dance movement, the segmentation process must be independent from wearable sensors. Segmentation process can use gesture spotting method based on Viterbi [3]; (5) It may be desirable to further optimize the HMM training and expectation-maximization algorithms to allow the system to give near instant feedback. This would make it truly viable for an human-computer interaction application.

Acknowledgement

The author greatly appreciate the contributions of Mr. Ary Setijadi Prihatmanto for the guidance and the teaching; and also the help of Sayid Tarmizi and Sundari Mega related to Likok Pulo dance.

References

[1] Jahn, D. C. W. 1979. Musculoskeletal Examination–Range Of Motion. *The Journal of the CCA*. 23(2): 51-60.

[2] Rabiner, L. R. 1989. A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition. *Proceedings of the IEEE*. 77(2): 257-286.

[3] Lee, H. K., Kim, J. H. 1999. An HMM-Based Threshold Model Approach for Gesture Recognition. *IEEE Transactions on Pattern Analysis and machine Intelligence*. 21(10): 961-973.

[4] Apley, A. G. 2010. *Apley's Sistem of Orthopaedics and Fractures.* Ninth edition. London: Hodder Arnold.

[5] LaViers, A. & Egerstedt, M. 2011. The Ballet Automaton : A Formal Model for Human Motion. *American Control Conference (ACC), 2011*. San Francisco, CA, USA. 29 Jun-01 Jul 2011. 3837-3842.

[6] LaViers, A., Chen, Y., Belta, C., Egerstedt, M. 2011. Automatic Generation of Balletic Motions. *Proceedings of the 2011 IEEE/ACM Second International Conference on*

Cyber-Physical Systems. Chicago, IL, USA. 12-14 April 2011. 13-21.

[7] Raptis, M., Kirovski, D., Hoppe, H. 2011. Real-Time Classification of Dance Gestures from Skeleton Animation. *ACM SIGGRAPH Symposium on Computer Animation*. Vancouver, BC, Canada. 5-6 August 2011.

[8] Hall, J.C. 2011. How to Do Gesture Recognition with Kinect Using Hidden Markov Models (HMMs). [Online]. From : http://www.creativedistraction.com/demos/gesture-recognition-kinect-with-hidden-markov-models-hmms/ [Accessed on 12 March 2014].

[9] Gowing, M., Concolato, C., Izquierdo, E. 2011. Enhanced Visualisation of Dance Performance from Automatically Synchronised Multimodal Recordings. *Proceedings of the 19th ACM international conference on Multimedia.* New York, USA. November 28-December 01, 2011. 667-670.

[10] Heryadi, Y., Fanany, M. I., Arymurthy, A. M. 2012. A Syntactical Modeling and Classification for Performance Evaluation of Bali Traditional Dance. *International Conference on Advanced Computer Science and Informations (ICACSIS).* Indonesia. 1-2 December 2012.

[11] Huang, J., Lee, C., Ma, J. 2012. *Gesture Recognition and Classification using the Microsoft Kinect*. Stanford: Stanford University.

[12] Zhang, H., Du, W. X., Li, H. 2012. *Kinect Gesture Recognition for Interactive System.* Stanford: Stanford University.

[13] Masurelle, A., Essid, S., Richard, G. 2013. Multimodal Classification of Dance Movements using Body Joint Trajectories and Step Sounds. *14th International Workshop on Image Analysis for Multimedia Interactive Services (WIAMIS). Paris, France. 3-5 July 2013.* 1-4.