HMM-BASED MODEL FOR DANCE MOTIONS WITH POSE REPRESENTATION

Nurfitri Anbarsanti[1], Ary S. Prihatmanto[2]

[1,2] School of Electrical Engineering and Informatics, Institut Teknologi Bandung, Bandung, Indonesia
[1]anbarsanti@yahoo.com, **[2]asetijadi@lskk.ee.itb.ac.id**

Abstract

This paper presents a model for human dance motions based on hidden markov model. The whole dance is defined as sequences of several finite distinct gestures. Dance gestures are cast as hidden discrete states and phrase of dance as a sequence of gestures. In order to map the skeleton motion data to a smaller set of features, an angular skeleton representation of the human pose is also designed, for recognition robustness under noisy input of 3D sensor. A pose of dance is defined by this angular skeleton representation which can be quantified based on range of movement for discrete hidden markov model.

Keywords: **angular skeletal representation, dance modelling, dance recognition, hidden markov model, human motion analysis.**

1. Introduction

Motion analysis and classification is of high interest in a variety of major areas including robotics, computer animation, psychology as well as the film and computer game industries. Dancing is a quintessential form of human body motion which has aesthetic values. In a number of emerging applications, the understanding of human dance motion plays a key role. The applications related to human dance motion range from the production of natural gaits for bipedal humanoid robot; indirect augmented reality during dance performance; to image segmentation and computer animation. The interaction using human dance motion has been studied as an alternative form of human-computer interface by a number of researches.

Real-world processes generally produce observable outputs which can be characterized as signals. Broadly one can dichotomize the types of signal models into the class of deterministic models, and the class of statistical (or stochastic) models [15]. In this paper, human dance motions are modeled by stochastic model due to the dance can be well characterized as a parametric random process, and that the parameters of the stochastic process can be determined in a precise manner.

Dance can be defined as sequences of several finite distinct gestures. Gesture has two aspects of signal characteristics : spatio-temporal variability and segmentation ambiguity [8]. Spatio-temporal variability is fact that the same gesture varies dynamically in size, shape and duration; from the different gesturer or even from the same gesturer. The segmentation ambiguity problem concerns how to determine when a gesture starts and when it ends in continuous signal trajectory. Major approaches for analyzing spatial and temporal patterns include Dynamic Time Warping (DTW), Neural Networks (NNs), dan Hidden Markov Model (HMM) [8]. In this study, HMM-based approach is chosen to model the dance gesture, because it can be applied to analyzing time-series with spatio-temporal varibilities and can handle undefined patterns [8]. HMM-based dance gesture modelling make us enable to build practical systems that has ability to learn, predict, and classify the dance gestures to analyzes the whole dance.

Dance movements that involve a lot of body articulation will result in a very large input dimension for the processing systems. The input dance motion signal such as skeleton joint trajectories captured by 3D sensor is very likely to experience a discontinuity, noise, or instable parameter [16]. In order to analyze the skeleton joint trajectories from 3D sensor, 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 3D sensor.

2. Background

2.1. Related Works

Dance choreography has been captured using various formalization approaches, e.g., Laban notation which is initiated in the early 20 th century. Aristidou, A. Chrysanthou and their colleagues propose a method that can automatically extract motion qualities from dance performances in terms of Laban Movement Analysis (LMA) [3].

Amy Laviers modelled the motion patterns of ballet as a series or event-driven poses that takes the form of a finite automaton [11]. 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 [12]. Yaya Heryadi [6] built a syntactical modeling and

classification for performance evaluation of Bali traditional dance, adapting the model of skeleton feature descriptor from Michalis Raptis [16]. 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.

F. Ofli, C. Canton-Ferrer and their colleagues analyze the relations between the music and the body movements. The body motion synthesis system will take an audio signal as an input and produce a sequence of body motion features, which are correlated with the input audio. The synthesis will be based on the HMM-based audio-body motion correlation model derived from the multimodal analysis [14].

2.2. Markov Model

Consider a system which may be described at any time as being in one of a set of N distinct states, $S_1, S_2, S_3, \ldots, S_N$, as illustrated in Figure 1.

At regularly spaced discrete times $t = 1,2,\ldots$, the system undergoes a change of state (possibly back to the same state) according to a set of probabilities associated with the state.

We denote the actual state at time t as q_t. A full probabilistic description of the above system would, in general, require specification of the current state as well as all the predecessor state. For the special case of a discrete, first order, Markov chain, this probabilistic description is truncated to just the current and the predecessor state, i.e,

$$P\,[q_t = S_j \mid q_{t-1} = S_i,\ q_{t-2} = S_k,\ \ldots\ ] = P\,[q_t = S_j \mid q_{t-1} = S_i] \quad \ldots (1)$$

Furthermore we only consider those process in which the right-hand side of (1) is independent of time, thereby leading to the set of state transition probabilites a_{ij}.

2.3. 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 (so we called it as hidden), but can only be observed through another set of stochastic processes that produce the sequence of observations [15].

A formal characterization of HMM is as follows :

a. $S = \{S_1, S_2, S_3, \ldots, S_N\}$, a set of N states. The state at time t is denoted by q_t .

b. $V = \{V_1, V_2, V_3, \ldots, V_M\}$, a set of M distinct 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.

c. $A = \{a_{ij}\}$, 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 :
$a_{ij} = P\,[q_t = S_j \mid q_{t-1} = S_i],\ 1 \le i, j \le N$

d. $B = \{b_j(k)\}$, an N x 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 :
$b_j(k) = P\,(O_t = v_k \mid q_t = S_j),\ 1 \le j \le N,\ 1 \le k \le M$.

e. $\pi = \{\pi_i\}$, the initial state distribution where π_i is the probability that the state S_i is the initial state :
$\pi_i = P[\,q_1 = S_i\,],\ 1 \le i \le N$

An compact notation $\lambda = (A, B, \pi)$ is used which includes only probabilistic parameters. Probabilistic notation A, B, and π must satisfy stochastic constraints as follows :

a. $\sum_j a_{ij} = 1, \forall\, i$, and $a_{ij} \ge 0$.

b. $\sum_k b_j(k) = 1, \forall\, j$, and $b_j(k) \ge 0$.

c. $\sum_i \pi_i = 1$, and $\pi_i \ge 0$.

The left-right model as shown in Fig. 2. It is good for modelling order-constrained time-series whose properties sequentially change over time [8]. Since the left-right model has no backward path, the state index either increases or stays the same as time increases. In other words, the state proceeds from left to right or stays where it was. On the other hand, every state in the ergodic or fully connected model can reach every other state in a single transition.

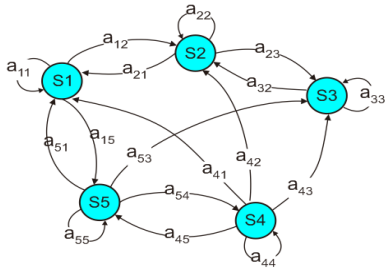

Figure 1. A Markov Chain with 5 State and with Selected State Transition

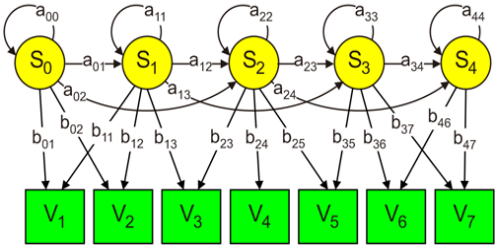

Figure 2. Graphical Model of Left-Right Discrete Hidden Markov Model

3. 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 3) :

a. Pose, is static configuration of human body, without any movement.
b. Gesture, is dynamic movement of human body, which is sequence of poses.
c. Phrase, is fragment of choreography which consist of sequence of gestures. The same gestures may be repeated.
d. 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 :

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

a. S, the finite nonempty set of hidden states. The states corresspond to gestures. Its segmentations are determined by the dance expert.
b. I, the finite nonempty set of input.
c. P, the vocabulary of all possible discrete pose of dance.
d. O, the finite nonempty set of output, where $O = \{o_1, o_2, o_3, \ldots o_T\}$, $o_i \in P^*$, $i \in \{1, 2, \ldots, T\}$. 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.
e. f, state transition function $f : S \times I \rightarrow S$. 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.
f. e, the output map $e : S \times I \rightarrow O$.
g. s_0, initial state, $s_0 \in S$. Initial state corresponds to initial pose or initial gesture of all phrases of Likok Pulo.
h. S_t, set of final (or accepting) states, $S_t \subseteq S$. Final states correspond to the end of the phrase.

For Likok Pulo Dance as case study [1], the model for each phrase are illustrated in Figure 4, Figure 5, Figure 6, Figure 7, Figure 8, and Figure 9. Initial states are indicated by using bold circles. Final states are indicated by using double circles. Actually the Likok Pulo dance has 6-8 phrases.

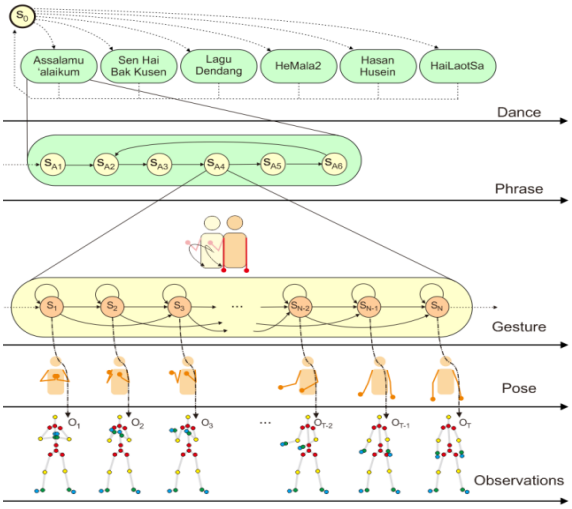

Figure 3. Hierarchy of The Whole Dance

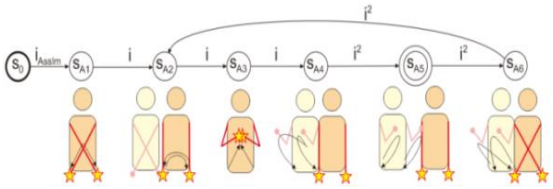

Figure 4. Model for "Assalamualaikum" Phrase

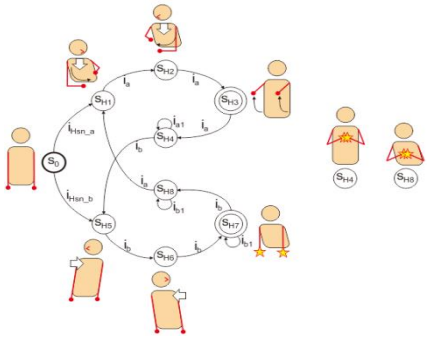

Figure 5. Model for Kisah Hasan Husein Phrase

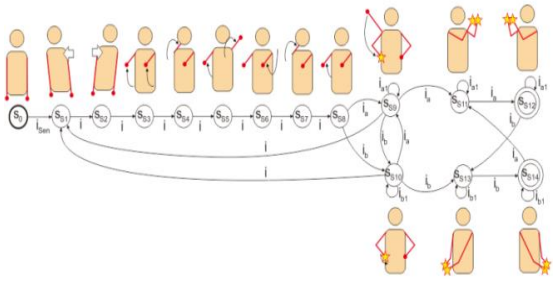

Figure 6. Model for Hai Aneuk Sen Phrase

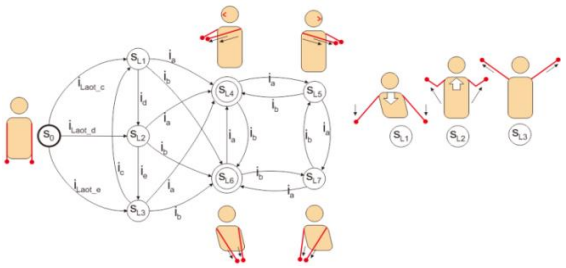

Figure 7. Model for Hai Laot Sa Phrase

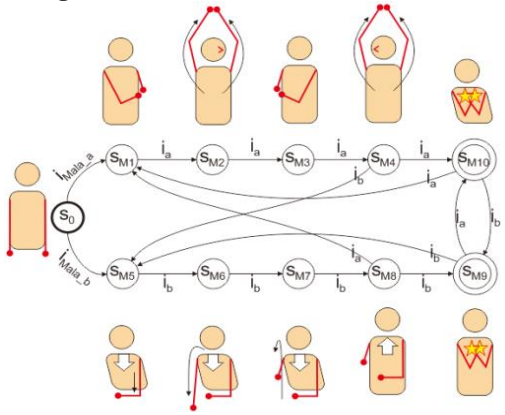

Figure 8. Model for He Mala Mala Phrase

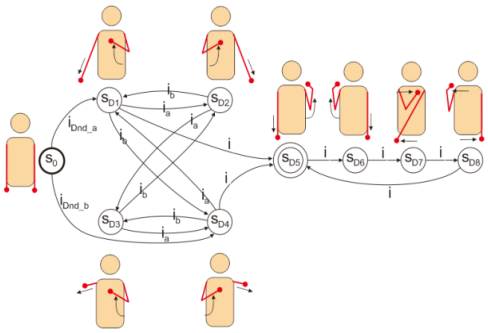

Figure 9. Model for Lagu Dendang Phrase

If the model are implemented in practical systems, input I can corresponds to music rhythm or pre-defined input. Output O can correspond to, e.g., indirect augmented reality during dance performance, or automatic generation of the dance for dancing humanoid robot.

3.2. Modelling the Dance Gesture in HMM

The left-right HMM model with two degrees is used because it is good for modelling order-constrained time-series whose properties sequentially change over time. As illustrated in Figure 2, the hidden states $S = \{S_1, S_2, S_3, \ldots, S_N\}$ correspond to the pose. The observations symbols $V = \{V_1, V_2, V_3, \ldots, V_M\}$ correspond to physical output at the system, i.e., the discrete pose vector $P_{u,2}$ (will be explained at chapter 4). 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. Human Pose Representation

The human pose representation must satisfy these objectives as follows [16]: (a) Robust coordinate system based on human body orientation, so that the skeleton representation does not depend to the position of the Kinect sensor. (b) Continuity and stability of the signal. (c) Reduce the dimension of the signal while maintaining the character of the motion.

There are several method to reduce the redundant dimension of the skeleton joint trajectories, such as [5,6,7,9,13]. Angular skeleton representation is chosen as the most appropriate representation of the pose of dance.

4.1. Torso PCA Frame

The joints of the human torso (defined by red skeletal nodes in Figure 10 and Figure 11) 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, $\vec{t} = \vec{u} \times \vec{r}$.

4.2. First-Degree Joints

First-degree joints (defined by yellow skeletal nodes in Figure 10) are represented relative to the adjacent joint in the torso in a coordinate system derived from torso PCA frame as illustrated in Figure 11 (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 $\vec{r}$. Then *RE* (right elbow)'s position is described by :

a. Its radius *R*, is the distance of *RE* from the *RS*. Since the length of the bones is constant, *R* can be ignored.

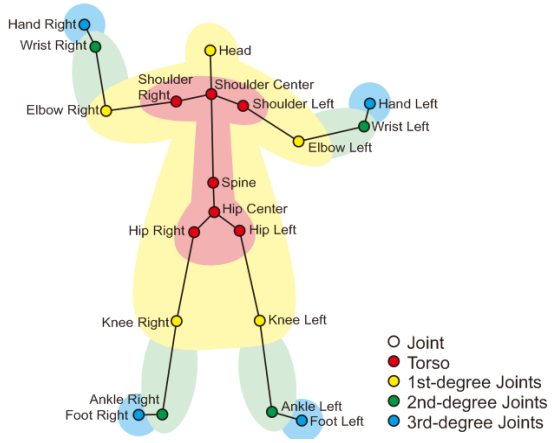

Figure 10. Hierarchy of Skeleton Joints.

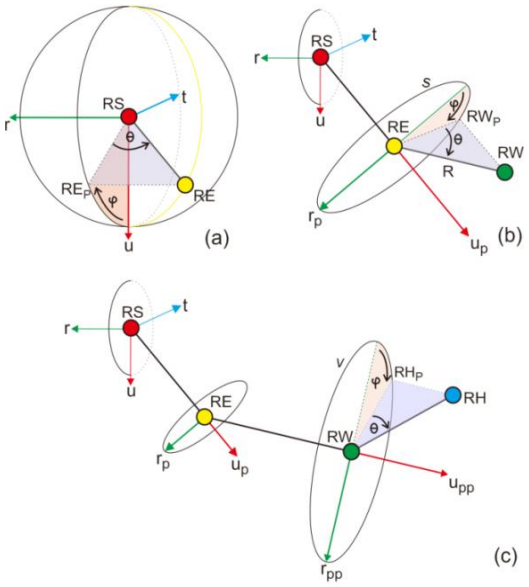

Figure 11. Spherical Coordinate System for (A) First-Degree Joints, (B) Second-Degree Joints, (C) Third-Degree Joints.

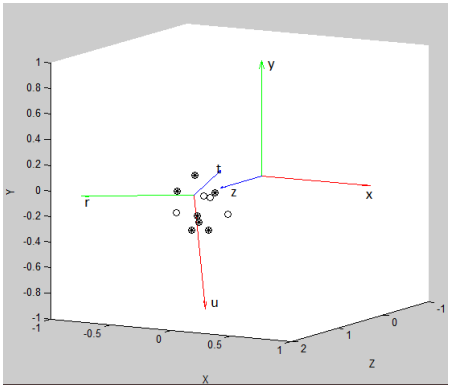

Figure 12. Results of Applying PCA to the Matrix Torso

b. Its azimuth φ, is the angle between $\vec{u}$ and vector $(\overrightarrow{RS, RE_P})$ where RE_P is the projection of RE onto the plane whose normal is $\vec{r}$.
c. Its elevation θ, is the angle between $(\overrightarrow{RS, RE_P})$ and $(\overrightarrow{RS, RE})$.

Each first-degree joint is represented with two angles (θ, φ). At joint RS, its angular representation is $\{RS_\theta, RS_\varphi\}$.

4.3. Second-Degree Joints

Second-Degree joints of the human skeleton (right and left wrist, right and left ankle) are defined by green skeletal nodes in Figure 11. These joints are represented relative to the adjacent joint in the

first-degree joints in a coordinate system $\{u_p, r_p, t_p\}$ which is derived from rotationed torso PCA frame $\{\vec{u}, \vec{r}, \vec{t}\}$ by angle $\{RS_\theta, RS_\varphi\}$ as illustrated in Figure 11 (b). The vector $\vec{u}_P$ protuding out of the vector $(\overrightarrow{RS, RE})$. The vector $\vec{u}_P$ be a zenith axis of the spherical coordinate system with origin RE. The azimuth axis is $\vec{r}_P$ and the zenith axis is $\vec{u}_P$. Then RW (right wrist)'s position is described by :

a. Its radius R, is the distance of RW from the origin RE, can be ignored.
b. Its azimuth φ, is the angle between $\vec{r}_P$ and $(\overrightarrow{RE, RW_P})$. Vector $\vec{r}_P$, is the result of rotationing vector $\vec{r}$ into plane S whose normal is $\vec{u}_P$. RW_P is the projection of RW onto the plane S.
c. Its elevation θ is the angle between $(\overrightarrow{RE, RW_P})$ and $(\overrightarrow{RE, RW})$.

Each second-degree joint is represented with two angles (θ, φ). Angular representation for RE is $\{RS_\theta, RS_\varphi\}$. Knee joint is represented by one angle θ.

4.4. Third-Degree Joints

Third-Degree joints of the human skeleton (right and left hand, right and left food) are defined by blue skeletal nodes in Figure 10 and Figure 11. These joints are represented relative to the adjacent joint in the second-degree joints in a coordinate system $\{u_{pp}, r_{pp}, t_{pp}\}$ which is derived from rotationed frame $\{u_p, r_p, t_p\}$ by angle $\{RE_\theta, RE_\varphi\}$ as illustrated in Figure 11(c). The vector $\vec{u}_{pp}$ protuding out of the vector $(\overrightarrow{RE, RW})$. The vector $\vec{u}_{pp}$ be a zenith axis of the spherical coordinate system with origin RW. The azimuth axis is $\vec{r}_{pp}$ and the zenith axis is $\vec{u}_{pp}$. Then RH (right hand)'s position is described by :

a. Its radius R, is the distance of RH from the origin RW, can be ignored.
b. Its azimuth φ is the angle between $\vec{r}_{pp}$ and $(\overrightarrow{RW, RH_P})$. Vector $\vec{r}_{pp}$ is the result of rotationing vector $\vec{r}$ into plane V whose

normal is $\vec{u}_{pp}$. RH_P is the projection of RH onto the plane V.

c. Its elevation θ is the angle between $\overrightarrow{RW, RH_P}$) and $\overrightarrow{RW, RH}$).

Each third-degree joint is represented with two angles (θ, φ). Angular representation for RW is $\{RW_\theta, RW_\varphi\}$.

It is needed to use Wearable Inertial Measurement Units (WIMU) [4] to obtain accurate angles at third-degree joints because Kinect sensor can not detect third-degree joints orientation and position accurately.

4.5. Human Pose Vector

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

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

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

a. Upper body poses are represented by an 12-tuple $P_{u,3} = (LW_\varphi, LW_\theta, LE_\varphi, LE_\theta, LS_\varphi, LS_\theta, RS_\theta, RS_\varphi, RE_\theta, RE_\varphi, RW_\theta, RW_\varphi)$.
b. Lower body poses are represented by the 12-tuple $P_{l,2} = (LA_\phi, LA_\theta, LK_\theta, LH_\phi, LH_\varphi, LH_\theta, RH_\theta, RH_\varphi, RH_\phi, RK_\theta, RA_\theta, RA_\phi)$.
c. Head poses are represented by the 3-tuple $H = (H_\varphi, H_\theta, H_\phi)$.

This angular pose representation can be directly used for the humanoid robot which has kinematics model equivalent to 24-DOF NAO [10] or 28-DOF Aaron. $P_{u,2}$ will be used to be observation symbol of HMM-based dance learning and classification [1].

4.6. Range of Pose Vector based on Range of Movement

Flexibility of a joint is defined as the range of movement (ROM) allowed at a joint, measured by the number of degrees using geniometer from the starting position of a segment until the final position of a full range of motion. The range of upper body movement for $P_{u,2}$, obtained from [2] is Tabel 1

4.7. Testing the Torso PCA Frame

Skeleton joints position are captured by Kinect sensor. Applying PCA to the matrix torso is performed using Matlab, and the result is in Figure 12. Torso joints are indicated by using dark circles. Elbow and wrist joints are indicated by using white circles.

To process second-degree joints at right arm, torso PCA frame $\{\vec{u}, \vec{r}, \vec{t}\}$ is rotationed by angle $\{RS_\theta, RS_\varphi\}$ and gives $\{u_p, r_p, t_p\}$. Then the frame $\{u_p, r_p, t_p\}$ is translated to joint RE as the origin. The calculations are performed by using Matlab and the result is in Figure 13.

5. 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.

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 3D sensor can not detect it. Skeleton representation can be deepened to also consider the dynamic aspects of the human body.

Acknowledgement

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

Table 1. Range of Movement for Upper Body Pose Vector

Left Arm	Right Arm
$-60 \le LS_\varphi \le 180$	$-60 \le RS_\varphi \le 180$
$-75 \le LS_\theta \le 180$	$-180 \le RS_\theta \le 75$
$20 \le LE_\varphi \le 180$	$0 \le RE_\varphi \le 160$
$-60 \le RE_\theta \le 90$	$-60 \le RE_\theta \le 90$

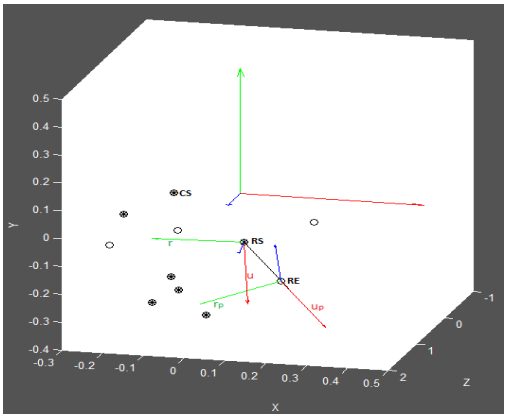

Figure 13. Transforming Torso PCA Frame for Right Elbow Calculations

References

[1] Anbarsanti, N., Setijadi, A., *"Dance Learning and Recognition System based on Hidden Markov Model. A Case Study : Aceh Traditional Dance"*, 4th International Conference on System Engineering and Technology, Bandung, Indonesia, November 24-25, 2014.

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

[3] Aristidou, A., Chrysanthou, *"Feature Extraction for Human Motion Indexing of Acted Dance Performances"*, International Conference on Computer Graphics Theory and Applications, (GRAPP), Lisbon, Portugal, 5-8 January, 2014.

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

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

[6] Heryadi, Y., Fanany, M. I., and Arymurthy, A. M., *"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.

[7] Hao Zhang, Wen Xiao Du, and Haoran Li, "*Kinect Gesture Recognition for Interactive System"*, Stanford, Stanford University, 2012.

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

[9] Justin Huang, Chun-wei Lee and Junji Ma, "*Gesture Recognition and Classification using the Microsoft Kinect"*, Stanford, Stanford University, 2012.

[10] Kofinas, N., Orfanoudakis, E., Lagoudakis, M. G., *"Complete Analytical Inverse Kinematics for NAO"*, 2013 13th International Conference on Autonomous Robot Systems (Robotica), Lisbon, 24 April 2013.

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

[12] LaViers, A., Yushan Chen, Belta, C., and Egerstedt, Magnus., "*Automatic Generation of Balletic Motions"*. Proceedings of the 2011 IEEE/ACM Second International Conference on Cyber-Physical Systems (pp 13-21), Chicago, IL, USA, 12-14 April 2011.

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

[14] Oflu, F., Canton-Ferrer, C., Demir, Y., *"Audio-Driven Human Body Motion Analysis and Synthesis"*, eNTERFACE'07 Workshop on Multimodal Interface, Istanbul, Turkey, July 16-August 10, 2007.

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

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