Dance Modelling, Learning and Recognition System of Aceh Traditional Dance based on Hidden Markov Model

Nurfitri Anbarsanti[1], Ary S. Prihatmanto[2]
School of Electrical Engineering and Informatics, Institut Teknologi Bandung, Bandung, Indonesia
E-mail: [1]anbarsanti@yahoo.com, [2]asetijadi@lskk.ee.itb.ac.id

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. The system was implemented using the Matlab and Simulink programming package. Six of dance's gesture classes from the phrase "Assalamualaikum" has been trained with hundreds of gesture instances recorded by the XBOX 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.

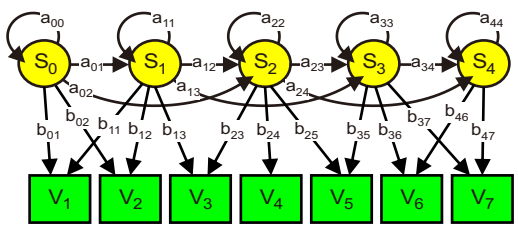
Index Terms—angular skeletal representation, Kinect sensor, dance modelling, dance recognition, gesture recognition, hidden markov model, Likok Pulo dance.

Aim

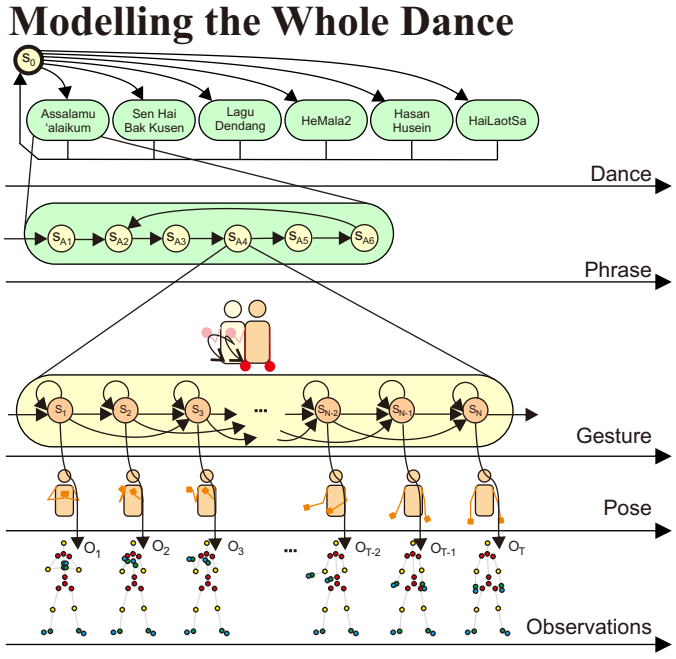
1. Model the whole dance of Likok Pulo Dance and its hierarchy.
2. Build the skeleton representation to handle discontinuities, instability, and dimension redundancy of Kinect's data trajectories.
3. Build the learning and classification system of Likok Pulo's gestures.

Hidden Markov Model

Markov model with a case where the observation is a probabilistic function of the state. The resulting 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.



Modelling the Whole Dance



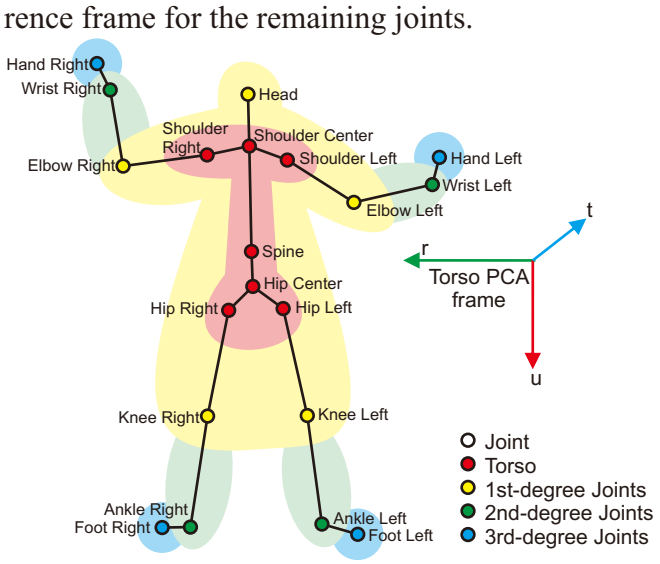
The whole Likok Pulo dance is modelled as follows:

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

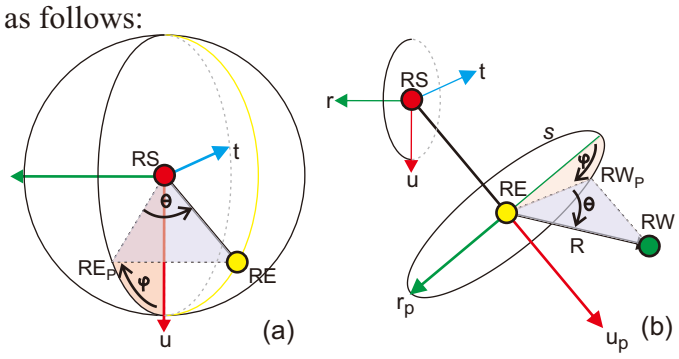
S, the finite nonempty set of hidden states. The states correspond to gestures.
I, the finite nonempty set of input.
P, the vocabolary of all possible discrete pose.
O, the finite nonempty set of output, where $O = \{o_1, o_2, ..., o_T\}$, $o_i \in P^*$, $i \in \{1,2,...,T\}$. P^* is the Klenee closure of P, the set of consisting of concatenations of arbitrarily many string of element from P (pose). Output O corresponds to gesture trajectories, or its features.
f, state transition $f : S \times I \rightarrow S$. 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.
e, the output map $e : S \times I \rightarrow O$.
s_0, initial state, $s_0 \in S$. Initial state corresponds to initial pose or initial gesture of all phrases.
S_t, set of final (or accepting) states, $S_t \in S$. Final states correspond to the end of the phrase.

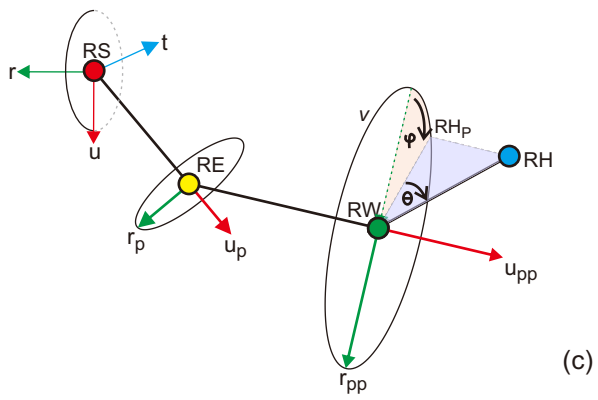
Skeleton Representation

The torso can be considere as a rigid body which provides 3D orthonormal basis will be used as reference frame for the remaining joints.



Each first-degree joint is represented with two angle which is derived from spherical coordinates system as follows:



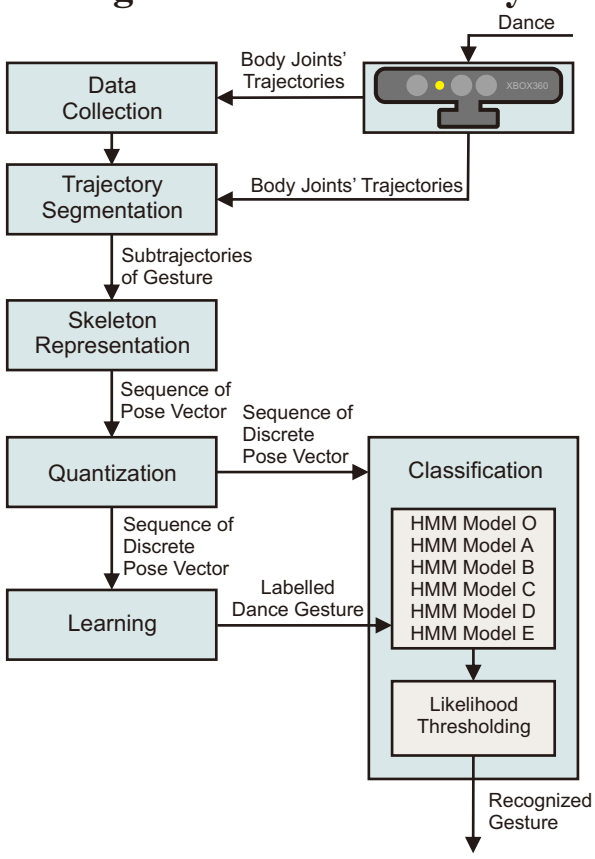


Spherical coordinate system for (a) first-degree joints (b) second-degree joints, (c) third-degree joints.

Conclusion

- Hidden Markov model used to model the dance.
- 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.

Learning and Classification System



It has been collected 2169 isolated dance gestures data which are classified to 6 dance gesture classes. Total isolated data are partitioned into 80% training data and 20% test data. Each element of pose vector are converted to one of the 3 or 5 directional codewords, based on ROM. 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 are225x225=50625poses. Classification results and realtime implementation as follows: