Modeling Basic Movements of Indonesian Traditional Dance Using Generative Long Short-Term Memory Network

Lukman Zaman, *Member, IAENG,* Surya Sumpeno, Mochamad Hariadi, Yosi Kristian, *Member, IAENG,* Endang Setyati, *Member, IAENG,* Kunio Kondo

***Abstract*—The preservation of traditional dances as an important part of world cultural heritage can be done by recording. While it is convenience to record the dances using video, this medium has limited capability in the reconstruction. On the other hand, recording using a motion capture device gives us the ability to replay them and add alterations in a creative process. In this paper, we propose a method to train traditional dance moves in a generative model using Long Short-Term Memory (LSTM). We use a traditional dance from East Java, Indonesia, that is called Remo Dance as the training data. The dance is recorded with a motion capture device and each basic move is trained into the model. In the sampling process, the trained model reiterates its memory into an unlimited length of dance animation. The generated dance animation has imperfection relative to the training data. This discrepancy gives the intended variations. We use visual assessments, dynamic time warping curves, and a subset of parameters from Laban motion analysis to evaluate the variations. These evaluations show how the variations behave and in what pattern they occur. In general, those variations give slight alterations to the motions that add human-like imperfection and give opportunities for animators and choreographers alike to explore new dances creations.**

***Index Terms*—Long Short-Term Memory, LSTM, generative model, deep learning, Indonesian Traditional dance, Remo dance, cultural heritage.**

I. Introduction

THE preservation of traditional dances could be done in many ways, including recording, performing, and creating new dances based on existing dances. Dances can be recorded using a digital camera or a motion capture device. These digital dance performances then be played back as videos or, in the case of motion capture, as animations. Creating new dances animation using digital media usually involves manually posing the 3D model keyframe by keyframe by the animators. A system that is able to create new dances animations automatically based on recorded existing dances should be useful in supporting said activities.

Currently, there is an increasing number of experiments in artificial intelligence, especially in deep learning; using artificial neural networks, that tackle the arts as the object. The grand aspiration of those researches is the capability of a virtual artist to create artworks which are indistinguishable from human-made ones. There are several fine works in painting domain, such as transferring styles [1] [2], creating doodles [3], and creating pictures from doodles [4]. In music domain, we have music production using generative adversarial network [5]. The creation of dances as a subset of human movements research is also getting more attention. Some works on modeling human motion using neural networks are [6] and [7]. An attempt at creating dance moves using deep learning, specifically for modern dance, has been done in Chor-RNN [8].

Dance as itself is performing art that lives in spatial and time dimension. The digital representation of dance performance is stored in the form of sequential data which also lie on the spatiotemporal dimension. The most common architectures for such data are from the recurrent neural network (RNN) family [9]. Long Short-Term Memory (LSTM) [10] and its variants are types of RNN which have been used to process language, sound, movies, and other sequential data with satisfactory results [11][12][13]. Furthermore, LSTM can act as a classifier or generative model. As a classifier model, LSTM uses the sequence of input data to predict an output value. Whereas as a generative model, LSTM will store the sequence and try to replicate its memory to produce similar sequences [14] [15]. This replication of the output may have an arbitrary length with some variations. Naturally, the ability to produce this kind of output, especially when the objects are from the arts domain encourages researchers as they see the potential to build virtual artists with some level of creativity.

Remo is a traditional dance originating from East Java, Indonesia; among hundreds of other Indonesian traditional dances. It is usually performed as a part of the ceremonial opening of important events[16]. Although Remo represents a male soldier character, it is also common to have either a male or a female dancer to perform this dance as shown in figure 1. Like any other traditional Javanese dances, Remo is composed of repetitions of several basic moves. These repetitions in Remo dance make them a good reason to train those moves into LSTM generative models. The trained LSTM generative models would remember those basic moves and could produce similar movements from their memories.

In this paper, we propose a generative animated model to mimic basic dance moves from a human dancer using Long Short-Term Memory (LSTM) Network. The main contribution of this paper: we build a standard LSTM model which remember basic dance moves. This model can produce animations of similar moves from their memories. The

Manuscript received August 10, 2019; revised April 1, 2020.

Lukman Zaman, Surya Sumpeno and Mochamad Hariadi are with the Department of Electrical Engineering, Faculty of Intelligent Electrical and Informatics Technology, Institut Teknologi Sepuluh Nopember, Indonesia. e-mail: lukman12@mhs.ee.its.ac.id, surya@ee.its.ac.id, mochar@ee.its.ac.id

Lukman Zaman, Yosi Kristian and Endang Setyati are with Department of Informatics, Institut Sains dan Teknologi Terpadu Surabaya, Indonesia. e-mail:lz@stts.edu, yosi@stts.edu, endang@stts.edu

Kunio Kondo is with School of Media Science, Tokyo University of Technology, Japan. e-mail:kondokunio@gmail.com,kondo@stf.teu.ac.jp

Fig. 1: A Dancer Performing Remo Traditional Dance

resulting animations can be repeated for any length of time, and they have slight variations in each of their movements. While we use Remo dance as data in this research, this model also capable of creating other dances that have similar characteristics that is, composed of repetition of several basic moves. With this ability, the model can be seen as a virtual choreographer that simplifies animators tasks in creating dance animations.

Each of the resulting moves from the models has some differences with another. On some level, these discrepancies make they seem more human-like. We can assess the output dance moves visually and by using Dynamic Time Wrapping (DTW) and Laban Motion Analysis (LMA). DTW can match two dance performances and compute the difference between them. LMA, a method that is commonly used in dance communities, could also could give an assessment in terms of human body and its movements.

II. Related Works

The digitalization of several Indonesian traditional dances using motion capture (mocap) device has been done in [17]. Storing human motions, including dance data in Biovision Hierarchy (BVH) format is frequently chosen for its practicality and feasibility. Besides its simplicity for playback purpose, BVH is also suitable for classifying [18], [19], recognition [20], and retrieval of human motions [21].

BVH uses a skeletal model and stores the motions as a chain of sequential poses. Processing the motions with this format requires a certain kind of architecture that is able to process sequential data, such as Hidden Markov Model (HMM). HMM has been used to generate dance motion in [22], [23]. Other architectures which are able to process sequential data is RNN. A fully connected RNN used in [24] is capable to create walking motion without training example.

Deep learning has also been used to generate human motion in [25]. This motion synthesis uses auto-encoder to create the abstraction of training data from the captured motion database. For dance creation, several approaches have been taken. Using motion transfer [26] we can create an animation of a person mimicking a dance video by other performer. The Another dance synthesis uses music as the input parameter [27]. In this method, dance motions are cued by varying the music from the training set. Moving on, Chor-RNN is another generative model that uses LSTM to memorize contemporary dance motion [8]. Instead of using skeletal model, this system is using the locations of mocap system markers as dance representation.

III. Remo Dance and Its Moves

Remo dance portrays the moves of a soldier or a knight [28]. This is the most common interpretation of the dance. The other interpretation is based on Nglana dance, the precursor of Remo. Nglana dance is believed as the main inspiration for Remo dance. It portrays the journey of a knight pursuing spiritual maturity. The dance is also inspired by Janaka, a character from Ramayana epics and a symbol for an ideal masculine character.

In Remo dance, said knightly aspects of the inspiration are interpreted into Javanese dance movements. Following is the general characteristic description of Javanese dance moves for a good male character. The performer adopts steady torso movements, with straight back and chest full of breath. Then, the arm motions have stability point at the elbow. Moreover, the dancer performs a series of steady flowing moves and the overall gestures are done in a controlled way. The head and the neck also move in a similar manner, with a calm and sharp gaze from unchanging face expression. These standard Javanese moves are modified in Remo by adding East Java characteristics, such as swift, agile, energetic, as well as quick-tempered-but-quickly-subsided mood. The modifications are rendered into a unique Remo choreography, for instance: several broken moves, quick moves repetitions, and more hand gestures at the front of the torso.

Remo has several variations of movements [29]. Some renowned Remo variations are Remo Trisnawati (Situbondo), Remo Tubi (Surabaya), and Remo Tawi (Jombangan). They are nicknamed after the choreographers' names and their city origins. Those differing qualities of each dance are determined based on the dynamics and the volumes of the moves. For example, Remo Tubi is known to be more dynamic and has more volume, because the dancers change their positions more frequently as compared to those of Remo Tawi. Another variation is the length of the dances. Remo can be performed for as long as 30 minutes, 12 minutes, or—in the case of junior dancers—only 4 minutes.

Currently, there is no formal written documentation of movements in Remo Dance. Although each basic moves has its uniqueness, we can refer to standard traditional Javanese dance moves in [30] to see some connections and similarities. Some moves which are commonly used in traditional Javanese dances are *tanjak* (ready position), *gedrug* (tapping the heel), and *ukel* (spiraling move of the hand).

Moving on, the basic moves are named and defined by tradition. There is so much variation on the complexity of each move. Some moves are very simple, like *tindak* (walking motion) and *kawung* (up and down hands movements). Other moves are quite complex, for instance *iket* (hands and upper body gesture of tying a knot) and *ayam alas* (emulate the movements of a junglefowl). In addition, certain moves concentrate on a single body part, for example *geter kepala* and *pacak gulu* (both are done by rhythmically bobbing the head); whereas other moves require full body movements, for example during the switch between different kind of moves. Moreover, there exist some moves which do not require any movement at all and are only defined by still poses of some

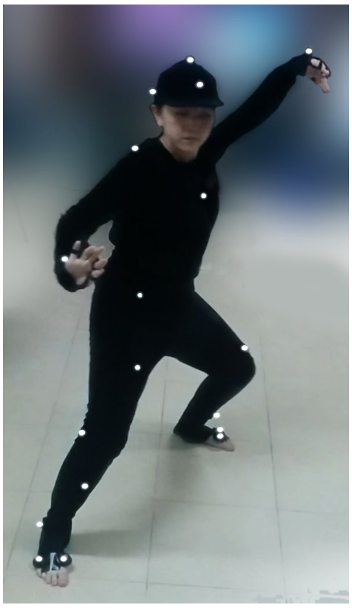


Fig. 2: A Dancer in Motion Capture Session

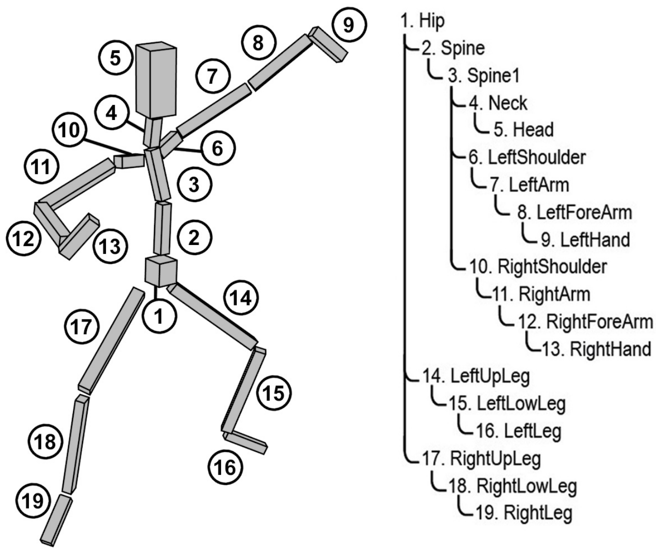


Fig. 3: Skeletal Model Setup for Mocap

body parts. The move *ngendewa* (pose of the arms nocking an arrow) is one of these moves.

If we use the traditional definitions of basic moves, the variations in complexity will create inconsistency in the length of the training features. To counter this problem, we manually select certain basic moves and cut some parts of them as necessary. To illustrate, we select the moves which focus on arms movements, legs movements, or other parts of body movements. By using these selections, we can evaluate our experiments based on certain features of the body parts.

IV. Data Specification

We recorded the short version of the most popular Remo variation. The total length of the dance is 4 minutes. We used Optitrack Motion Capture System with 8 cameras to record the performance as shown in figure 2. The raw data from the motion capture device are in 60 frames per seconds (fps) rate. We down-sampled those data into 30 fps animation and save them in standard Biovision Hierarchy (BVH) format.

BVH format contains two parts of information. The first part is the definition of the hierarchical skeletal model used in animation. Then the second part contains the animation data. In BVH, the animation is defined by a sequence of poses of the skeletal model. These poses are stored in the keyframes, a pose for each keyframe. Each pose is defined by rotation of the bones relative to its joints in the skeletal model. In this format, each pose that is defined by the skeletal model in figure 3 can be written as $p \in \mathbb{R}^{3\times|B|}$ where $|B|$ is the number of the bones. The 3 values in that term are the real numbers for Euler rotations for each bone on the XYZ axis. A motion can be expressed as equation 1.

$$\mathcal{M} : [1 : f] \to \mathcal{P} \subset \mathbb{R}^{3\times|B|} \tag{1}$$

In equation 1, $\mathcal{M}$ is the motion that consists of sequence of poses contained in set $\mathcal{P}$. The number of frames in the motion is f which lies along the time dimension. The usage of the real numbers for expressing rotations of the bones (instead of simple integer symbols) is unique to our model. These real numbers demand more training phases and LSTM layers size but they are required for smooth animation results.

We use 19 bones in the model as shown in Figure 3. In addition to bone rotation data, there is also the position of the root bone (i.e. the hip bone) which must be stored in each frame. This root position acts as the reference to the relative position of the model in the 3D world. With this bone configuration, each frame of the pose should be defined by (19+1)*3=60 values. The full-length dance animation is a sequence of these poses in a total of 8,200 frames (approximately 4 minutes in 30 fps rate animation). From the full-length dance data, we select several basic moves to be trained into the neural network model. The specification of the selected moves can be seen in Table I.

V. System Architecture and Evaluation Methods

We use a sequence of LSTM modules with Mixture Density Module (MDN) augmented at each output (Figure 4.b). There are several variants in LSTM gating system, but we found those variants have approximately the same performance as standard LSTM (Figure 4.a). This gating system allows an LSTM cell to remember, add, or discard new information based on the previous sequence of inputs. This gating system is required to overcome the vanishing

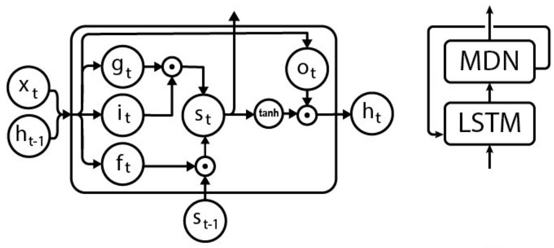


Fig. 4: (a) Standard LSTM, (b) LSTM with Augmented MDN

TABLE I: Training Data Description for Remo Basic Moves

Movement Name	Length (frames)	Desciption
Tindak	95	Walking motion, alternately moving the left and right legs
Gejug kaki, seblak tangan	107	Stomp the ground, throw an arm outward
Tanjak, gedrug kaki	95	Ready position with tapping a heel on the ground
Seblak tangan	145	Throw an arm outward
Lawung kanan-kiri	180	Bringing the left and right arms up and down alternately
Geter kepala	72	Rhythmically moving the head left and right
Seblak, gejug, tanjak	140	Sequence of throwing an arm, stomping the ground, and get to the ready position
Gejug kaki, tindak	137	Stomping the ground dan walking around a circle
Selut	87	Bringing the left and right hands alternatively up and down in front of the torso
Ukel trap jamang	188	Emulating movement of the hands putting up a head gear
Iket (part of)	86	This is a part of a complex motion which shows the movement of the hands tying a knot.

gradient problem—which almost certainly occurs in a deep or long chain of neural network modules [10].

The purpose of the MDN is to add the capability of predicting ambiguous output to the LSTM[31]. In dance movements, ambiguous data commonly occur. In the training process, the MDN are trained to create a collection of Gaussian probability functions. The probability density P as expressed in equation 2, is a combination of m number of Gaussian kernels. It generates vector t from input vector x using α as mixing coefficients.

$$P(t \mid x) = \sum_{i=1}^{m} \alpha_i(x)\phi_i(t \mid x) \quad (2)$$

Each kernel is computed using Gaussian function in equation 3.

$$\phi_i(t \mid x) = \frac{1}{(2\pi)^{c/2}\sigma_i(x)^c} exp(-\frac{\|t-\mu_i(x)\|^2}{2\sigma_i(x)^2}) \quad (3)$$

In equation 3, the center of the i-th kernel is denoted by μ and the variance by σ. All those MDN values are trained alongside the main LSTM module. Without the addition of MDN, the LSTM tends to collapse all probable outputs into a single average value.

Furthermore, the random aspect in MDN could eventually direct the training into an unsuccessful one. The solution to this problem is simply to retry the training process with a different random seed. The visual assessment method is useful for checking whether the model has been trained properly or not. The more detailed evaluation is done by using DTW to compute the distance between the original training data and the generated output. DTW is a common method to compute the similarity between any of two sequential data [32][33][34]. These sequences of data can be asynchronous and differing in lengths.

Besides the aforementioned, we also use LMA to evaluate the generated dances. LMA analyzes dance moves based on four components: body, effort, shape, and space [35]. The body component describes the relationship between different body parts. The effort component describes how energy is used in motions. Then the shape component characterizes how the shape of the body changes over time, and the space component specifies spatial factors covered in a movement. Because LMA evaluates motions both in visual and feeling qualities, there are some challenges to interpret those evaluations into numerical values. In this research, we use a subset from [36] to calculate the numerical values of the LMA components from motion data.

The evaluation using LMA requires explicit positions of body parts. In the skeletal model, the body part positions are defined by the bone joint's locations. Equation 4 converts rotation angles from BVH data into positions of the joints.

$$\mathbb{P} = \begin{bmatrix} X_r \\ Y_r \\ Z_r \end{bmatrix} = M_r \left(M_{r-1} \left(... \left(M_1 \begin{bmatrix} X_0 \\ Y_0 \\ Z_0 \end{bmatrix} \right) \right) \right) \quad (4)$$

In Equation 4, the position of the hip is defined by X_0, Y_0, and Z_0 coordinates. $\mathbb{P}$ is the final position of any joint. If r is the level of the joint in the hierarchical model, its position is computed by merging the transformation matrices from its ancestors' bones up to the root bone. The matrices M are transformation matrices which are composed of translation and rotation matrices.

VI. Training and Sampling Process

The training process involves the feeding of training data consisting of pairs of features and labels into the model. The training data are prepared by cutting the dance data into several segments, which would become the features. These features contain sequences of poses as defined in Equation 1. They are illustrated as the sequence of X_q in Figure 5.a. A single body pose after each segment acts as the label for that particular segment (denoted by X_{q+s} in Figure 5.a). Moreover, the cutting positions of the dance data must cover all possible moves, so that the number of the features is equal to the number of the total frames in the dance data (denoted by n), so an epoch in the training process requires the set of X_q where $q = 1..n$.

For optimal training, we set the parameters for the model based on [37]. The constraint of the training process mainly depends on the GPU memory limit. The capacity of GPU RAM determines how many batches of training data can be processed at the same time. Using 8 gigabyte GPU RAM, on average, each move requires 4 hours of training.

After a successful training, the trained model remembers the correlations between each movement segment (i.e. a sequence of body poses) and a body pose after that. With these memories, the model could predict the next probable pose after processing a sequence of poses as the feature (Figure 5.b). So, at the beginning of the generation process,

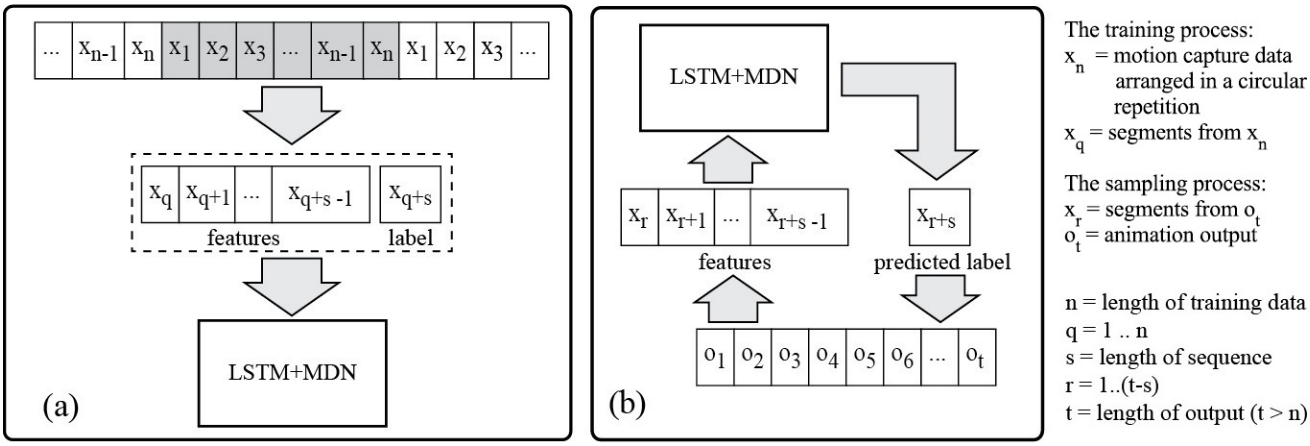


Fig. 5: (a) Training Process, (b) Sampling Process

the trained model is warmed up by giving several known poses. For each segment of the input, it produces an output pose. This output is also augmented to the next input and being fed back to the model. After several cycles, the model will be able to produce consecutive output poses which form an animation that is similar to the trained dance move.

VII. Results and Evaluations

The trained model produces an infinite length of dance. Generally, the outputs are in the form of a sequence of trained moves which are played in loops. We apply the resulting motions onto a 3D human model and visually compares them to assess the dance animation output. We also evaluate these results using DTW [32] and LMA [35] [36]. DTW evaluates the full-length output to find when and how the variances may occur. Then from DTW results, we select two output samples with the least and the most variances. Using LMA method, we analyze further on how those variances occur and describe how the body parts move on the two extremes.

Figure 6 shows three examples from the moves in Table I. These figures are from the rendered 3D model animations of the dance data. The renditions are in simple flat colors and outlines to avoid unnecessary details. The first example in this figure is the *tindak* move. This move only uses legs movement (Figure 6.a). The second example is the *kawung* move, it represents the arms movement with stationary legs (Figure 6.b). Finally, the last example is a transition move from *gejug* to *tindak*. This third example shows the motion of all body parts. In this move, the dancer swirls both his hands as he walks and turns his body sideward, then swivels it back forward (Figure 6.c).

In Figure 6, each move contains three rows of figures. The figures on top rows are the training data. In the middle rows, there are figures for the output motions which are the most similar to the training data, while at the bottom rows are the most different. For clarity purposes, not all frames are included in the figures, but they are down sampled into 1/4 for Figure 6.a and 6.b, and 1/8 for Figure 6.c. The figures show that the variations of the outputs can take forms in the gesture and the position of the dancer. The gesture variations are caused by deviation of the bone rotation angles in the output; whereas the position differences are caused by the shifting of hip bone location. Because this hip bone acts as the root of the model, the difference in its value gives some offset to the overall body.

When we use visual assessment, we can quickly judge the outputs. Then, the analyses using DTW [32] and LMA [35] [36] give more detail about them. In DTW analysis, each repetition in the output is compared with the training data and their distances are calculated. Specifically, the DTW method computes the differences between every single bone in the skeletal model from the output and from the training data. The difference between motions is simply the summation of the differences between all bone pairs. Figure 7 shows the differences over some periods of repetition. The vertical axis shows the DTW distance between the output and the training data; while the horizontal axis shows the repetition number (for example, 40 represents the 40th loop of certain dance move). Each tick in the horizontal axis represents a different length of frames because the length of each motion is varied. In addition, not all DTW bone diagrams are represented in the graph, but only of which are visually important for variation assessment. Usually, the included bones are hip (as the root bone), head, and four body appendages. The graph plot labelled TOTAL is the total DTW distance of all 19 bones in 3D model.

The graphs show that there are consistent behaviors in all of the outputs. They show periodic increasing and decreasing of DTW distance in each loop. Sometimes, the dance moves deviate further from the training data, but they settle back to be more similar after some period. These pattern is repeated for the entire length of the output. Using these peaks and valleys, we can choose the least or most extreme variation from the outputs. For example in Figure 7.a, the output motion for *tindak* moves, the output motion at the 30th repetition has the most similar motion as compared to the training data. This is shown by a low valley in the graph. In contrast, the peak at the 39th point of the horizontal scale shows that the repetition has the most variance. This undulations of the DTW distance occur on all dance moves; although there are differences in how and which different body parts contribute to the total deviations.

Moving on, by using DTW diagrams, we can explain how different parts of the body take part in contribute to the difference in motions. For example, the graph show that

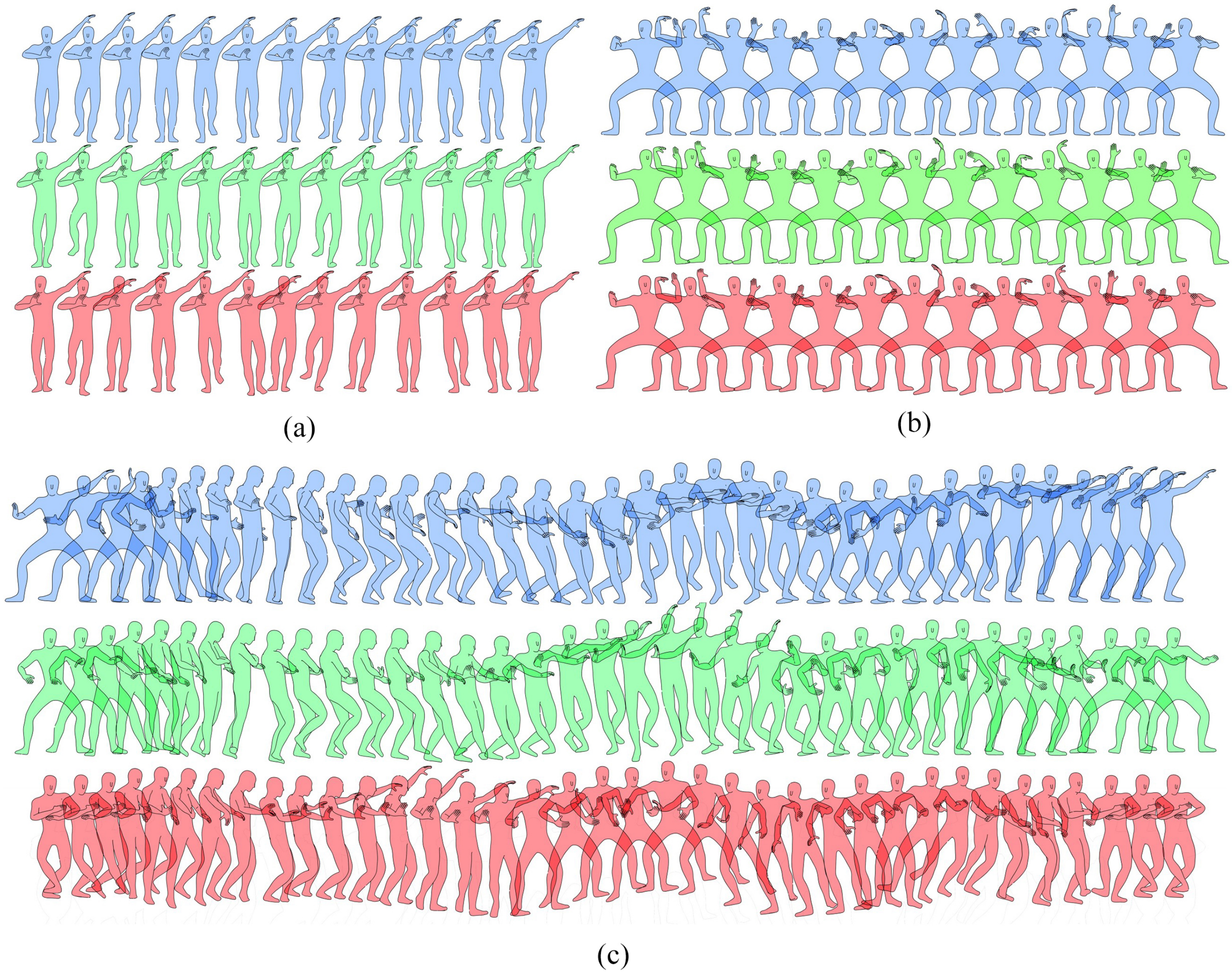


Fig. 6: The Generated Moves of (a) *Tindak* shows the variation in the legs movement, (b) *Kawung* for the arms, and (c) *Gejuk-Tindak* transition for showing the variation in a full body movement. The top rows are the training data, the middle rows are the most similar, and the bottom rows are the most different output

difference in legs motion have a prominent effect on the total difference. Besides, the different scales in the vertical axis must also be carefully observed. Some moves like *kawung* (Figure 7.b) have steep peaks and valleys in the graph plot. Nevertheless, because the distance is relatively small, the difference is hardly noticeable in visual representation. Logically, motions involving full body parts like transition move in Figure 7.c, generally have a large DTW distance as more bones contribute their own differences to the total distance.

Note that the comparisons using DTW only calculate the distance between the rotation angles of the bones excluding hip bone. As the root bone, the location of the hip bone would affect the positions of other bones. However, although the hip bone causes location displacement which is visually perceptible in 3D models, it does not affect the DTW distances of bones other than itself. Hence, it would be a different case when we use LMA features to assess the movements. Unlike DTW that allows many kinds of data, LMA features are strictly computed from various locations of body parts. By converting rotation data into position of the joints using Equation 4, it means that the position of the hip bone would likely affect the LMA greatly.

LMA has a full range of features which can be used to analyze various aspects of human motion. In this paper, we choose a subset from those features to help us get further insight into the resulting variations of the generated dance moves. For instance, for *tindak* move which relies heavily on legs movement, we use feet velocity as well as right and left volume features (Figure 6.a). Next, we use hands velocity to examine hands-intensive movement in *kawung* (Figure 6.b).

Each LMA diagram in Figure 8 uses three graph plots to show how the most similar and the most different output vary from the training data. The diagrams show how the variations occur in terms of several LMA parameters. In velocity diagrams (Figure 8.a), the peaks show when the body parts move quickly and the valleys show when they momentarily pause (e.g. at the moment before changing the direction). In volume diagrams, the peaks show when the body taking most spaces, usually when the dancer spread her arms or feet (Figure 8.b and 8.c). The other LMA features are used accordingly.

We can notice that even though the shapes of the output graphs are similar or at least they have some resemblances to

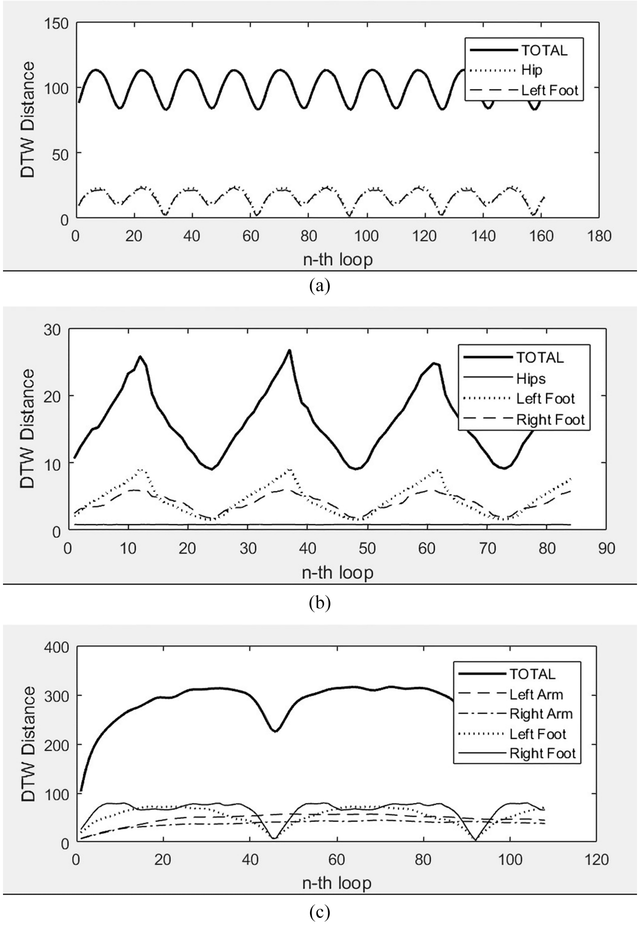


Fig. 7: The DTW distance between output moves and training moves, (a) *tindak*, (b) *kawung*, and (c) *gejuk-tindak* transition

the training data, they are generally more jagged. This shows that there are inherent noises in the output. In the active movements of the body parts, big motions usually overcome the noises, thus making them invisible. The noises are more visible when they occur in stationary body parts. The noises can also exist in the time domain. The jerk diagram in Figure 8.d depict the values from the first derivative of acceleration of the hip. They show erratic values, signifying the noises which occur in the time domain. However, these values are relatively small and their effects are not visible.

VIII. Discussion and Future Works

Along with the aforementioned examples, the evaluation of the outputs from the trained models for other Remo dance moves shows similar behaviors. As Figure 6 shows, the shifting in the hip bone position can be a problem when the variation on the hip position gives undesirable sliding legs animation or makes the legs to float over or be buried underground. Applying an inverse kinematics mechanism [38] on the legs can fix this problem. Other than that, a more complex solution can be applied by adding physics simulation which is able to compute the hip location based on the pacing of the legs[24].

As compared to that of hip location, variations of the bone angles are smaller and even desirable. The slight difference in bone rotations gives a human imperfection aspect which adds a natural looks to the animation. The more challenging problem is when the variations are caused by time shifting. For most dances, including Remo, the accompaniment music gives the cue for the dance beat. Specifically, the Remo dancers follow the sound of Javanese hand drum called *kendhang* for the timing. Naturally, these discrepancies in time produce invalid moves for standard Remo dance. From another perspective, however, when these variations can sway a motion into an invalid move, a choreographer can take advantage of them. Choreography process always involves observation, emotional response, improvising, and evaluation [39][40]. Using these generated variations should help the choreographer at least at the observation, improvisation, and evaluation phases.

IX. Conclusion

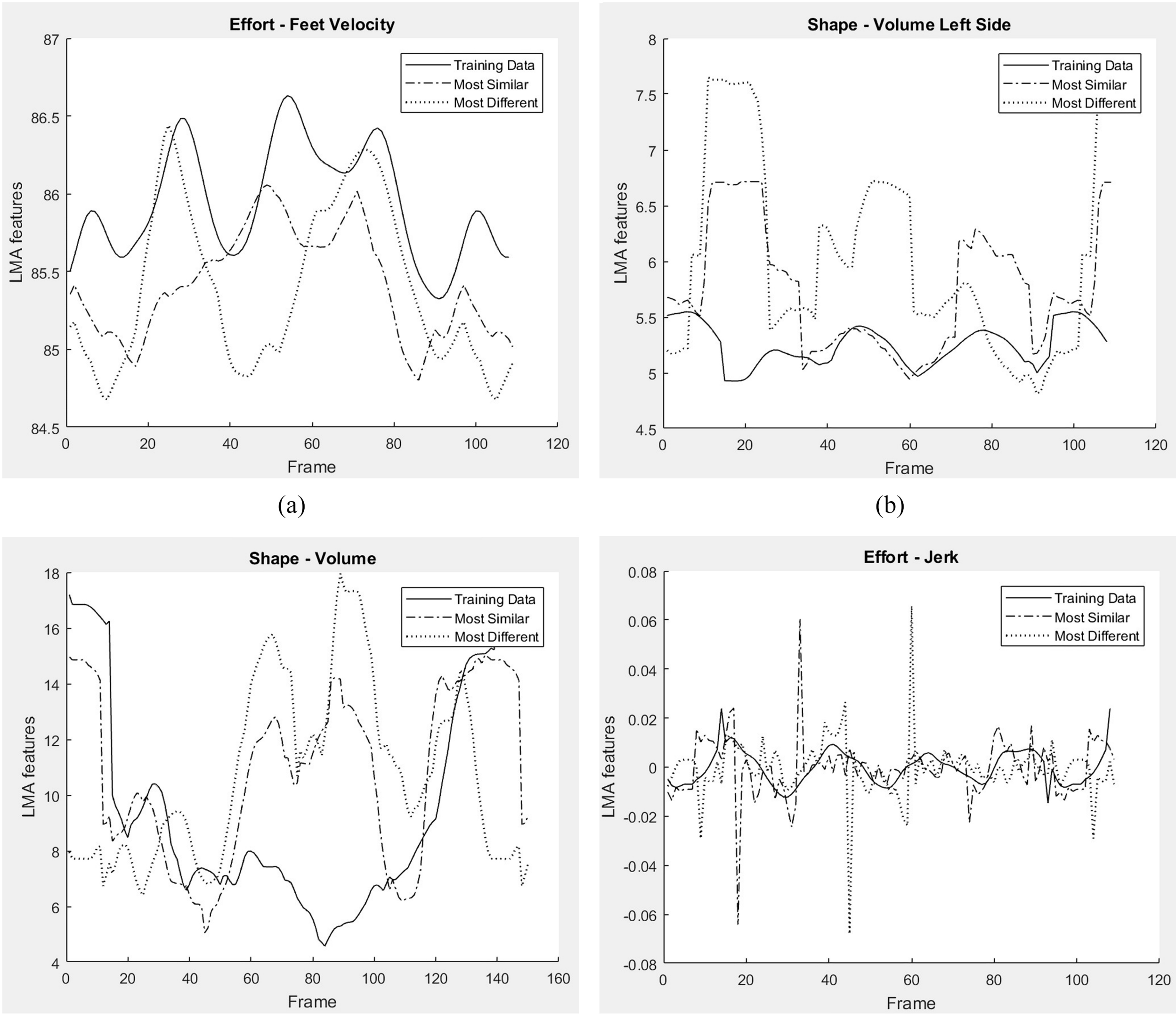
The trained generative models of Remo dance basic moves using LSTM can help animators to create unlimited variations of said basic moves. These variations in the output animation can be both in the spatial and temporal dimension. The generated variations also have a potential ability to help the choreographers explore new dance moves. The current limitation in using this method lies on the computation process requirement—capped by the availability of the GPU memory. When this limitation is overcome in the next development of GPU technology, theoretically, the limit of the length of the dance moves that can be trained into the models also increases. In that condition, it is feasible to build virtual choreographers which are able to combine and vary the trained dance moves.

Acknowledgment

The first author is grateful for the doctoral program funded by General Directorate of Higher Education, Ministry of Research and Education, Indonesia for financial support. We also give thanks to Human-Centric System Laboratory, ITS, Surabaya for their motion capture facility and Stefanie Tanujaya as the Remo dancer.

References

[1] L. A. Gatys, "Image Style Transfer Using Convolutional Neural Networks," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 2414–2423.
[2] B. He, F. Gao, D. Ma, B. Shi, and L.-y. Duan, "ChipGAN : A Generative Adversarial Network for Chinese Ink Wash Painting Style Transfer," in *2018 ACM Multimedia Conference on Multimedia Conference*, 2018, pp. 1172–1180.
[3] T. Zhou, C. Fang, Z. Wang, J. Yang, B. Kim, Z. Chen, J. Brandt, and D. Terzopoulos, "Learning to Sketch with Deep Q Networks and Demonstrated Strokes," *ArXiv: 1810.05977*, 2018. [Online]. Available: http://arxiv.org/abs/1810.05977
[4] P. Isola, J.-y. Zhu, T. Zhou, and A. A. Efros, "Image-to-Image Translation with Conditional Adversarial Networks." *arXiv Computer Vision and Pattern Recognition*, 2018.
[5] L.-C. Yang, S.-Y. Chou, and Y.-H. Yang, "MidiNet: A Convolutional Generative Adversarial Network for Symbolic-domain Music Generation," *arXiv:1703.10847 [cs.SD]*, 2017.
[6] J. Martinez, M. J. Black, and J. Romero, "On human motion prediction using recurrent neural networks," 2017.
[7] D. Pavllo, C. Feichtenhofer, M. Auli, D. Grangier, and E. Zurich, "Modeling Human Motion with Quaternion-based Neural Networks," 2019.
[8] L. Crnkovic-Friis and L. Crnkovic-Friis, "Generative Choreography using Deep Learning," *7th International Conference on Computational Creativity*, no. June, pp. 1–6, 2016.
[9] J. L. Elman, "Finding Structure in Time," *Cognitive Science*, vol. 14, no. 2, pp. 179–211, 1990.



(a) (b)

(c) (d)

Fig. 8: Several examples of LMA Diagram for the moves, (a) feet velocities in *tindak*, (b) left arm moves in *kawung*, (c) all body volume, and (d) jerk motion in *gejuk-tindak* transition

[10] S. Hochreiter and J. Schmidhuber, "Long Short-Term Memory," *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997. [Online]. Available: http://www.mitpressjournals.org/doi/10.1162/neco.1997.9.8.1735

[11] J. A. Ramírez-Quintana, M. I. Chacon-Murguia, and J. F. Chacon-Hinojos, "Artificial neural image processing applications: A survey." *Engineering Letters*, vol. 20, no. 1, pp. 68–80, 2012.

[12] W. Cao, A. Song, and J. Hu, "Stacked residual recurrent neural network with word weight for text classification." *IAENG International Journal of Computer Science*, vol. 44, no. 3, pp. 277–284, 2017.

[13] Q. Zhuge, L. Xu, and G. Zhang, "Lstm neural network with emotional analysis for prediction of stock price." *Engineering Letters*, vol. 25, no. 2, pp. 167–175, 2017.

[14] A. Karpathy, "The Unreasonable Effectiveness of Recurrent Neural Networks," 2015. [Online]. Available: http://karpathy.github.io/2015/05/21/rnn-effectiveness/

[15] A. Graves, "Generating sequences with recurrent neural networks," *arXiv preprint arXiv:1308.0850*, pp. 1–43, 2013.

[16] T. Wibisono, *Tari Remo di Surabaya: Dari Terob, Tobong, Menuju Kelas*, 2016.

[17] E. Hegarini, Dharmayanti, and A. Syakur, "Indonesian traditional dance motion capture documentation," in *International Conference on Science and Technology-Computer (ICST)*, 2016.

[18] A. Kleinsmith, P. R. De Silva, and N. Bianchi-Berthouze, "Cross-cultural differences in recognizing affect from body posture," *Interacting with Computers*, vol. 18, no. 6, pp. 1371–1389, 2006.

[19] A. Kleinsmith and N. Berthouze, "Recognizing Affective Dimensions from Body Posture," *Affective Computing and Intelligent Interaction*, vol. LNCS-4738, pp. 48–58, 2007.

[20] H. Liu, Z. Ju, X. Ji, C. S. Chan, and M. Khoury, *Human Motion Sensing and Recognition*. Springer-Verlag, 2017, vol. 675. [Online]. Available: http://link.springer.com/10.1007/978-3-662-53692-6

[21] X. Wei, B. Xiao, and Q. Zhang, "A retrieval method for human Mocap data based on biomimetic pattern recognition," *Computer Science and Information Systems*, vol. 7, no. 1, pp. 99–110, 2010.

[22] M. Brand and A. Hertzmann, "Style machines," *SIGGRAPH*, pp. 183–192, 2000.

[23] F. Ofli, E. Erzin, Y. Yemez, and A. M. Tekalp, "Learn2Dance: Learning statistical music-to-dance mappings for choreography synthesis," *IEEE Transactions on Multimedia*, vol. 14, no. 3 PART 2, pp. 747–759, 2012.

[24] T. Reil and P. Husbands, "Evolution of central pattern generators for bipedal walking in a real-time physics environment," *IEEE Transactions on Evolutionary Computation*, vol. 6, no. 2, pp. 159–168, 2002.

[25] D. Holden, J. Saito, and T. Komura, "A deep learning framework for character motion synthesis and editing," *ACM Transactions on Graphics*, vol. 35, no. 4, pp. 1–11, 2016.

[26] C. Chan, S. Ginosar, T. Zhou, and A. A. Efros, "Everybody Dance Now," *arXiv:1808.07371*, vol. 1, no. 1, 2018.

[27] J. Françoise, P. Pasquier, O. Alemi, J. Françoise, and P. Pasquier, "GrooveNet : Real-Time Music-Driven Dance Movement Generation using Artificial Neural Networks," *In Proceedings of SIGKDD 2017Workshop on Machine Learning for Creativity*, no. July, p. 6, 2017.

[28] Wahyudianto, “Karakteristik Ragam Gerak dan Tatarias-Busana Tari Ngremo sebagai Wujud Presentasi Simbolis Sosio Kultural,” *IMAJI*, vol. 4, no. 2, 2006.

[29] A. T. Rukmana Sari and W. Wahyudi, “Rekonstruksi Gerak Pada Tari Remo Tawi Jombang,” *Joged*, vol. 10, no. 2, p. 577, 2018.

[30] C. Brakel-Papenhuijzen, *Classical Javanese Dance: The Surakarta Tradition and Its Terminology*. KITLV Press, 1995.

[31] C. M. Bishop, “Mixture Density Networks,” *Neural Computing Research Group*, 1994.

[32] M. Müller, *Information Retrieval for Music and Motion*. Springer-Verlag Berlin Heidelberg, 2007.

[33] K. A. Darabkh, A. F. Khalifeh, I. F. Jafar, B. A. Bathech, and S. W. Sabah, “A yet efficient communication system with hearing-impaired people based on isolated words of arabic language,” *IAENG International Journal of Computer Science*, vol. 40, no. 3, pp. 183–192, 2013.

[34] K. Tamura and T. Ichimura, “Classifying of time series using local sequence alignment and its performance evaluation,” *IAENG International Journal of Computer Science*, vol. 44, no. 4, pp. 462–470, 2017.

[35] E. Davies, *Beyond Dance: Laban’s Legacy of Movement Analysis*. Routledge, 2007.

[36] A. Aristidou, E. Stavrakis, P. Charalambous, Y. Chrysanthou, and S. L. Himona, “Folk Dance Evaluation Using Laban Movement Analysis,” *Journal on Computing and Cultural Heritage*, vol. 8, no. 4, pp. 1–19, 2015.

[37] L. Zaman, S. Sumpeno, and M. Hariadi, “Training Strategies for Remo Dance on Long Short- Term Memory Generative Model,” in *International Conference on Computer Engineering, Network, and Intelligent Multimedia*, Surabaya, 2018.

[38] S. Tejomurtula and S. Kak, “Inverse kinematics in robotics using neural networks,” *Information sciences*, vol. 116, no. 2, pp. 147–164, 1999.

[39] S. C. Minton, *Choreography: a basic approach using improvisation*. Human Kinetics, 2017.

[40] J. M. Smith-Autard, *Dance composition: A practical guide to creative success in dance making*. Routledge, 2014.

Lukman Zaman receive his bachelor degree in computer science from Sekolah Tinggi Teknik Surabaya (STTS), Surabaya, Indonesia, in 1998. He receive his master degree in Informatics from Institut Teknologi Sepuluh Nopember (ITS), Surabaya, Indonesia, in 2003 He is currently pursuing a Ph.D degree in Institut Teknologi Sepuluh Nopember (ITS), Surabaya, Indonesia. He joined as a faculty member in STTS since 1999. His current research interests include interactive media, artificial itelligence, and computer graphics. He is an IAENG member.

Surya Sumpeno earned his bachelor degree in Electrical Engineering from Institut Teknologi Sepuluh Nopember, Surabaya, in 1996, and MSc degree from the Graduate School of Information Science, Tohoku University, Japan in 2007. He earned doctor degree in Electrical Engineering from Institut Teknologi Sepuluh Nopember, Surabaya, in 2011. His research interests include natural language processing, human computer interaction and artificial intelligence. He is an IAENG and IEEE member.

Mochamad Hariadi received the B.E. degree in Electrical Engineering Department of Institut Teknologi Sepuluh Nopember (ITS), Surabaya, Indonesia, in 1995. He received both M.Sc. and Ph. D. degrees in Graduate School of Information Science Tohoku University Japan, in 2003 and 2006 respectively. Currently, he is the staff of Electrical Engineering Deparment of Institut Teknologi Sepuluh Nopember, Surabaya, Indonesia. He is the project leader in joint research with PREDICT JICA project Japan and WINDS project Japan. His research interest is in Video and Image Processing, Data Mining and Intelligent System. He is a member of IEEE, and member of IEICE

Yosi Kristian receive his bachelor degree in computer science and master degree in information technology from Sekolah Tinggi Teknik Surabaya (STTS), Surabaya, Indonesia, in 2004 and 2008 respectively. He receive his Ph.D degree on Electrical Engineering from Institut Teknologi Sepuluh Nopember (ITS), Surabaya, Indonesia. He joined as a faculty member in STTS since 2004, and currently as an Associate Profesor in Department of Computer Science. His current research interests include Machine Learning, Intelligent System, and Computer Vision. He is an IAENG and IEEE member.

Endang Setyati earned her bachelor degree in Mathematics from Institut Teknologi Bandung, Indonesia, in 1992 and master degree in Informatics from Institut Teknologi Sepuluh Nopember (ITS), Surabaya, Indonesia, in 2000. She receive her doctorate degree at the Graduate School of Electrical Engineering, ITS, in 2017. She is an associate profesor at the IT Department, Sekolah Tinggi Teknik Surabaya. Her research interests are mathematical logic, digital image processing and human computer interaction. She is an IAENG member.

Kunio Kondo is a Professor at the School of Media Science, Tokyo University of Technology. He earned his Bachelor degree from Nagoya Institute of Technology in 1978 and Dr. Eng from the University of Tokyo in 1988. He is the former Associate Professor at the Department of Information and Computer Sciences, Saitama University in 1989-2007, faculty member at Tokyo Polytechnic University in 1988-1989, technical staff at Nagoya University in 1973-1988, part-time teacher at Tokyo University in 1991-2007, Aichi Prefectural University of Fine Arts and Music in 1989-1999, and Kyushu Institute of Design in 2002-2010. His research interests are computer graphics, animation, game, and interactive modelling. He won the IPSJ Anniversary Best Paper Award in 1985, JSGS Research Award in 1985, and JSGS Best Paper Award in 2011. He is the President of The Institute of Image Electronics Engineers of Japan, former President of The Society for Art and Sceince, former Vice President of Japan Society of Graphic Science, and Chair of SIG on Computer Graphics and CAD of Information Processing Society of Japan, Board member of Asia Digital Art and Design Association.