An Automatic Extraction Tool for Ethnic Vietnamese Thai Dances Concepts

Truong-Thanh Ma, Salem Benferhat, Zied Bouraoui, Karim Tabia,
Thanh-Nghi Do, Nguyen-Khang Pham

An Automatic Extraction Tool for Ethnic Vietnamese Thai Dances Concepts (Preprint version)

Truong-Thanh Ma[1], Salem Benferhat[2], Zied Bouraoui[2], Karim Tabia[2], Thanh-Nghi Do[1], Nguyen-Khang Pham[1]

[1] CICT, Can Tho University, Vietnam
Email: truongthanh1511@gmail.com
{dtnghi, nhhoa}@cit.ctu.edu.vn

[2] CRIL, Artois University
CRIL CNRS & Univ Artois, France
Email:{benferhat, bouraoui, tabia}@cril.fr

***Abstract*—In recent year, preservation and promotion of the ICHs are one of the problems of interest. In this paper, we focus on modelling the traditional dance domain, particularly modelling traditional Vietnamese dances. To conserve significant characteristics of dances, we proposed an ontology to represent the significant movements features of Ethnic Vietnamese Thai Dances (EVTDs). Particularly, a detailed description of the movement schemas of EVTDs is presented in this paper. Additionally, we present how to build an automatic extraction tool to collect the fundamental movements data of EVTDs using machine learning. Finally, we represented explicitly how to store those extracted features from raw dance videos into prioritized Ontology-based proposed.**

Keywords: *Vietnamese Traditional Dance, Prioritized Ontology, machine learning, Ethnic Vietnamese Thai Dance*.

I. Introduction

Preservation and promotion of the intangible cultural heritage (ICH) are one of the problems of interest in the world. Therefore, many scientists in computer science domain applying the theoretical and practical knowledge to the field of Intangible Cultural Heritage, particularly in this paper focusing on Southeast Asia countries, because Southeast Asia region is one of the most dynamic regions in the world with a rich cultural heritage. We concentrated primarily on the Vietnam's traditional dances.

Vietnam is a multi-ethnic country existing many different cultures [25] with fifty-four-ethnic groups living in a territory. The traditional dances had became "spiritual foods" of each Vietnamese people, it influences directly the real life from urban to rural. Most of the traditional Vietnamese dance (TVD) is transfered by "word of mouth" [1], the present generation would instruct fundamental movements to the adjacent generation. Additionally, traditional Vietnamese Dances are a steady bridge in educating about human dignity, morality and even historical knowledge. Instead of learning the historical lesson in regular classes as well as participating in the training the course of life skills, the dances has became the digital channel for efficiently educating personality, knowledge and even ethnicity to the generations.

Almost Vietnamese traditional dances built up from the ethnic groups culture, life environments and regions, it contains the large number of the significant characteristics of region-zone. Particularly, the dance movements of the ethnic groups originated from the life activities, each posture is depicted an characteristic action of their life. Therefore, the fundamental movements would be one of the stable foundation as well as being the essential features to determine the different dances. In this paper, we selected a remarkable dance of Thai community in Vietnam to illustrate how to represent the important features. The movements of EVTDs are the combination of fundamental postures presented in [5].

Classifying, detecting, identifying and storing dance videos is a great challenge because most of the Vietnamese dances are stored in raw videos. Therefore, we proposed a methodology to manage the heterogeneous data of EVTDs in order to search, store and query-answer dance videos.

One of the main contributions of this paper is to build an automatic extraction tool using machine learning in order to store the significant features of ethnic Vietnamese Thai dance movements (EVTDMs). These characteristics are put into a lightweight prioritized ontology-based (LPO) attached to recognized probabilistic of each body part in the same frame. Based on the probabilistic of each posture to decide whether the tool put those features into LPO or not.

In the research process, we decomposed our approach into two main stages: the first aspect is to reconstruct a schema of panorama overview of traditional Vietnamese dances; the second aspect concerns a character with the fundamental movements of each EVTD. In this paper, we concentrate primarily on the second stages with respect to detecting, extracting and storing automatically the fundamental movements. Our primary challenge is to determine the principle concepts from EVTD's movements combined with a set of desirable properties to build a useful dance search engine, moreover, because the movements of the performers is quite uncertain and inconsistent regarding the different cases (amateur, speed of music, etc.) as well as the sequential frame of videos is not explicit to recognize those extract movements, our remarkable aim is how to handle the uncertainties and inconsistencies in processing.

The remainder of this paper is structured as follows. In the

next section (section 2) we give recent related works. Section 3 provides a description of EVTD's features and how to encode them into Ontology. How to detect automatically concepts and a methodology to build a LPO are discussed in section 4. Finally, section 5 concludes the paper.

II. Related Works

During the last two decades, people have tried to develop different algorithms for human activity analysis [24][20] for wide applications in the area of surveillance, patient monitoring and many more. Most of the works have been reported on classifying human activity from videos. Recently researchers are trying to classify an activity from a single image [22] [27] [23]. In the video based activity recognition, people have tried with different human activities like walking, jogging, running, boxing, hand waving, hand clapping, pointing, digging and carrying for a single actor [24][4].

Furthermore, with the rapid development of computer science, the large number of applications of management of large and heterogeneous data is built and explored. In the dance domain, several authors using ontology to manage the dance databases, evidences as BalOnSe application [33], a Labanotation based ontology [32]. Moreover, there are a few works on group activities as [16][12]. Due to the increase in multimedia data access through the internet, multimedia data specially video data indexing becomes more and more important. Not only in the retrieval but also for digitization of cultural heritage, this can be an interesting problem. It can be used to analyze a particular dance language.

Some researchers use space time features to classify the human action. Blank et al. represent the human action as three dimensional shapes included by the silhouettes in the space-time volume [13]. They use space-time features such as local space-time saliency, action dynamics, shape structure and orientation to classify the action. In [2], they recognise human action based on space-time locally adaptive regression kernels and the matrix cosine similarity measure. Klaser et al. localize the action in each frame by obtaining generic spatio temporal human tracks [11]. They have used sliding window classifier to detect specific human actions.

There are several attempts to recognize the movement from multiple videos sources [2][24]. Aggarwal et al. [24] classify the human activity recognition in two classes namely, single-layered approach and hierarchical approach. In single layer approach, activities are recognized directly from videos, while in hierarchical approach, an action is divided into sub-actions [10]. The action is represented by classifying it into sub-actions. Wang et al. [18] have used topic model to model the human activity. They represent a video by Bag-of-Wards representation. Later, they have used a model which is popular in object recognition community, called Hidden Conditional Random Field (HCRF) [19].

III. Ethnic Vietnamese Traditional Dances

A. Ethnic Vietnamese Thai Dances

Thai community in Vietnam is one of the ethnic groups existing the large number of the traditional dances. There are many significant festivals of Thai ethnic group to be held in villages as well as regions during the whole year. In order to understand explicitly with respect to EVTDs, in this section, we present several fundamental movement features to identify the EVTDs in the Vietnam's territory.

In each EVTD, the remarkable characteristics to determine EVTDs is the fundamental movements, in which is the foundation of the creative combination in each motion in order to create the specific dances of Vietnamese Thai people. Correspondingly, the detection of the basis movements is one of the important steps collected automatically the dance dataset for LPO. The following we would present a basis movements schema of EVTDs as well as represent EVTD's LPO based on the schema proposed. Basis movements of EVTDs are divided in five characteristics [5]: Orientation, Arm Posture, Leg Posture, Sitting Posture, Standing Posture as figure 1. They are described in detail as follows:

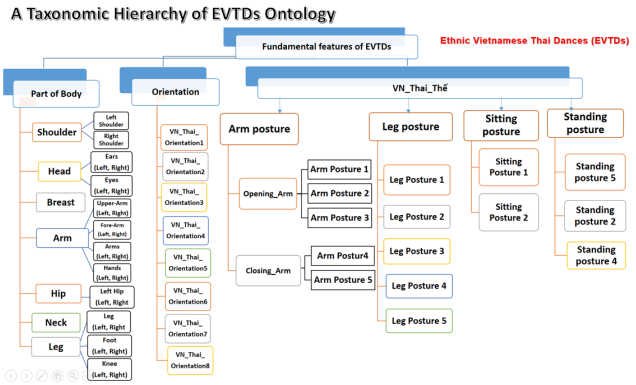

Figure 1: A taxonomic Hierarchy of EVTDs

1) Orientation: Regarding orientation features, it is one of the most significant characteristics because the motions, postures and gestures of Vietnamese traditional dances are always described explicitly through the orientations in almost all documents. They are split in eight orientations as Figure 2, including from orientation 1 to orientation 8. In [5], orientation 1 is the direction of the dancer opposite to spectator (in front of audience), it is also utilized for the first preparation step of performing

2) Arm Posture: Most of arm postures is concentrated on depicting life activities in Thai community, therefore the basis postures is simple and habitual. It is divided in five primary postures as figure 3: VN-Thai-Thế-[i]-Arm (i=1..5). They are grouped into two distinct clusters: open-arm posture and close-arm posture.

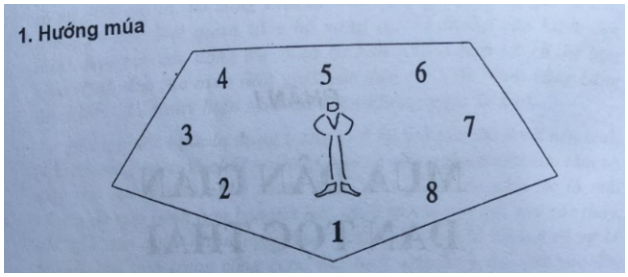

Figure 2: Orientations of EVTDs

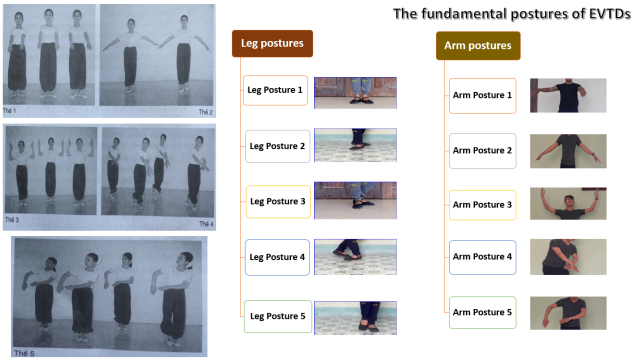

Figure 3: Arm postures and Leg postures of EVTDs

3) Leg Posture: There are five significant leg postures to represent for EVTD movements consisting of VN-Thai-Thế-[j]-Leg (j=1..5) as figure 3.

4) Sitting and standing Posture: Sitting posture is divided in two postures, it consists of VN-Thai-Thế-1-Sitting, VN-Thai-Thế-2-Sitting. There are three standing posture in EVTD as in figure 4 VN-Thai-Thế-5-Standing, VN-Thai-Thế-2-Standing, VN-Thai-Thế-4-Standing.

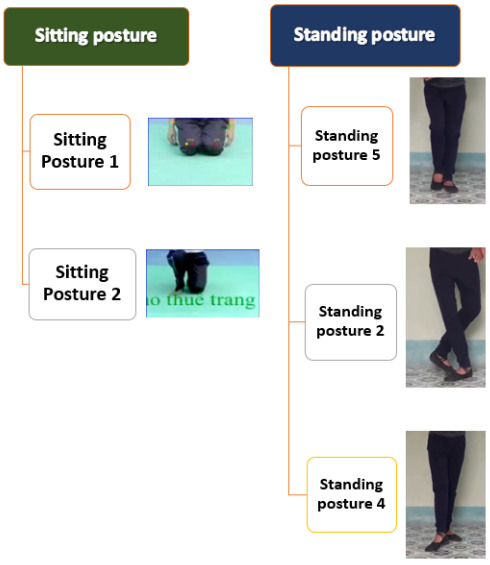

Figure 4: Sitting postures and standing postures of EVTDs

B. Encoding EVTD Movements into Prioritized Ontology

This section proposes implementing a modelling of EVTD using lightweight ontology by DL-Lite. One of the main advantage of DL-Lite is that query-answering is down in an efficient way. Realizing the significance of DL-Lite in representing ontology-based model as well as perceiving further what is the motivation behind using DL-Lite in order to build an application *(semantic Web)* for EVTDs through OWL2-QL language, therefore, we proposed to select DL-Lite to represent implementing ontology-based modelling of EVTDs in this paper.

1) Presentation of DL/DL-Lite: Description Logics (DLs) provide the formal foundation for ontologies, and the tasks related to the use of ontologies in various application domains are posing new and challenging requirements with respect to the trade-off between the expressive pover of a DL and the efficiency of reasoning over knowledge bases (KBs) expressed in the DL. Description Logics [**?**] are also the logical frameworks underlying the ontology language. A description logic knowledge base is formed by a terminological base, called TBox, and an assertional base, called ABox. The TBox contains inclusion axiom concepts and rules regarding knowledge of the application domain whereas the ABox stores data (individuals and constants) DL-Lite is a family of tractable DLs specifically altered for applications that utilize the huge number of data, for which query answering is the significant reason assignment.

2) DL-Lite syntax: Let [**?**] N_C, N_R and N_I be three pairwise disjoint sets of atomic concepts, atomic roles and individuals respectively. Let $A \in N_C, P \in N_R$, three connectors $'\neg'$, $'\exists'$, $'-'$ are used to define complex concepts and roles. We only present $DL-Lite_{core}$ the core fragment of all the $DL-Lite$ family and we would utilize $DL-Lite_R$ instead of $DL-Lite_{core}$. Nevertheless, we concentrate on two important members of the $DL-Lite$ family in this paper, including $DL-Lite_R$. Considering basic concepts (*resp.* roles) B (*resp.* *R*), complex concepts (*resp.* roles) C (*resp.* *E*) are defined in DL-Lite as follows:

$$B \rightarrow A \mid \exists\, R \qquad C \rightarrow B \mid \neg B$$
$$B \rightarrow A \mid P^- \qquad E \rightarrow R \mid \neg R$$

where P^- represents the inverse of P and A is an atomic concept, P is an atomic role.

A DL-Lite knowledge base (KB) is a pair $\mathcal{K} = \langle \mathcal{T}, \mathcal{A} \rangle$ where $\mathcal{T}$ is the TBox and $\mathcal{A}$ is the ABox. The TBox $\mathcal{T}$ includes a finite set of inclusion axioms on concepts and on roles respectively of the form: $B \sqsubseteq C$ and $R \sqsubseteq E$. The ABox contains a finite set of assertions (facts) of the form A(a) and P(a,b) where $A \in N_C, P \in N_R$ and $a, b \in N_I$.

3) Implementing EVTD movements ontology using DL-Lite: We built a prioritized ontology of EVTD movements based on the schema proposed as figure 1. We present how to encode EVTD movements into LPO using DL-Lite, as follows:

TBox: T_{EVTD}
$Orientation \sqsubseteq EVTD - Movements$
$ArmPosture \sqsubseteq EVTD - Movements$
$LegPosture \sqsubseteq EVTD - Movements$

$SittingPosture \sqsubseteq EVTD - Movements$
$StandingPosture \sqsubseteq EVTD - Movements$
$CaUocMovement \sqsubseteq Movements$
$VN - Orientation - 1 \sqsubseteq Orientation$
$VN - Orientation - 3 \sqsubseteq Orientation$
$ArmPosture - 1 \sqsubseteq ArmPosture$
$ArmPosture - 2 \sqsubseteq ArmPosture$
...
$LegPosture - 1 \sqsubseteq LegPosture$
$LegPosture - 2 \sqsubseteq LegPosture$
...
$XeKhanDance \sqsubseteq EthnicVietnameseThaiDance$
$XeMaHinhDance \sqsubseteq EthnicVietnameseThaiDance$
$CaUocMovement \sqsubseteq XeKhanDance$
$ChauPoMovement \sqsubseteq XeKhanDance$
$QuatBoHeoMovement \sqsubseteq XeKhanDance$
...
$\exists hasOrientation \sqsubseteq Part - Of - Body$
$\exists hasOrientation^{-} \sqsubseteq Orientation$
$\exists hasPose \sqsubseteq EthnicVietnameseThaiDance$
$\exists hasPose^{-} \sqsubseteq ArmPosture$
$\exists hasArmPosture \sqsubseteq EthnicVietnameseThaiDance$
$\exists hasArmPosture^{-} \sqsubseteq ArmPosture$
$\exists hasLegDance \sqsubseteq EthnicVietnameseThaiDance$
$\exists hasLegPosture^{-} \sqsubseteq LegPosture$
$\exists hasConfidence \sqsubseteq EVTD_{M}ovement$
$\exists hasConfidence^{-} \sqsubseteq IntegerType$
...

For each raw video, we represented each feature in each frame, particularly, for an example of Frame10 presented in ABox as follows:

ABox: T_{EVTD}
...
$hasConfidence(VN - Orientation - 1 - 94, 94)$
$hasOrientation(RightShoulder, VN - Orientation - 1 - 94)$
$hasOrientation(LeftShoulder, VN - Orientation - 2 - 74)$
...
1. $hasArmPosture(Frame10, VN - Thai - ArmPosture - 3 - 67)$
2. $hasLegPosture(Frame10, VN - Thai - LegPosture - 1 - 89)$
3. $hasPose(Frame10, RightShoulder - Orientation - 5 - 73)$
4. $hasPose(Frame10, LeftShoulder - Orientation - 1 - 94)$
5. $hasPose(Frame10, LeftHip - Orientation - 1 - 55)$
6. $hasPose(Frame10, RightHip - Orientation - 6 - 34)$
...

The explanation of (1) that the accuracy of ArmPosture-3 is 67% in Frame 10. At the same frame 10, left shoulder with orientation-1 obtained 94% of accuracies represented in (4) and so on. Our primary aim is to store each motion of each body part of performer/dancer in the same frame into prioritized ontology-based where collect the dance movement dataset to serve for searching, reasoning and query-answering.

The reason why we built prioritized ontology because a dance video is sequential frames where we are difficult to determine explicitly how dance postures and movements are right. Therefore, the probability and accuracy of each motion based on machine learning algorithms is quite expected. Example: considering a dance video from frame 1 to frame 100, just exiting frame 20 satisfied the essential requirements of those motions. Based on the accuracies classified of machine learning which we identified whether the features are set into LPO or not (above 80% to be selected).

IV. On the Automatic Detection of Concepts

In order to detect the concepts automatically, we had utilized TF-Openpose to identify skeletons of performers. We also aggregated Deep Convolutional Neural Networks to recognize primary postures in each frame. The principle purpose of the automatic detection is to enrich the ontology by assertions representing concepts of LPO presented in each frame. In this section, a description of how to detect automatically significant features through machine learning will be discussed, particularly using human pose estimation and DCNN architectures to build a detection tool.

A. Human Pose Estimation

We utilize TF-Openpose (written in python using Tensorflow library instead of Caffe library) for estimating the positions of human joints and articulated pose estimation in order to support for depicting each movements in EVTD. Moreover, we ameliorated and improved TF-Openpose through algorithms of input image processing and modified several essential arguments of CNNs.

The primary purpose of using HPE for EVTD movements is to determine concretely parts of body in raw dance videos to aim at extracted and represent the motions in each dance movement. The architecture simultaneously predicts detection confidence maps and affinity fields that encode part-to-part association as in Figure5. The network is split into two branches: Branch 1 is responsible for predicting confidence maps, and Branch 2 is to predict the affinity fields. TF-Openpose takes a 2D color image as input and produces the 2D location of anatomical key-points for each person. The (x,y) coordinates of the final pose data array could be normalized to the range depending on the key-point scale. It can be estimated 18 key-point body pose from COCO 2016 dataset.

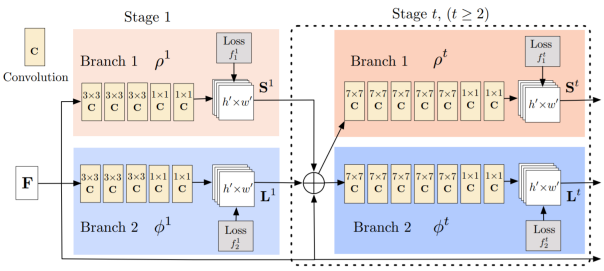

Figure 5: The architecture using CNN in Openpose

Realizing the requirements of a high configuration regarding GPU for Openpose handled, we proposed to utilize TF-Openpose [1] instead of orginal Openpose version. It is a human pose estimation library developed based upon the foundation of the Openpose library using Tensorflow and OpenCV. It also provides several variants that have made the changes to the network structure for real-time processing on the CPU or low-power embedded devices. We concentrated on two variations of models to find optimized network architecture: CMU [29] and Mobile-Net [34]. (1) With respect to CMU, it is the model based VGG pre-trained network which described in the Openpose's original paper using COCO dataset for training, it

[1] https://github.com/ildoonet/tf-pose-estimation

is converted from Caffe format to utilize in Tensorflow; (2) Based on the Mobile-Net paper [34], with 12 convolutional layers are used as feature-extraction layers. The experimental result as in figure 6.

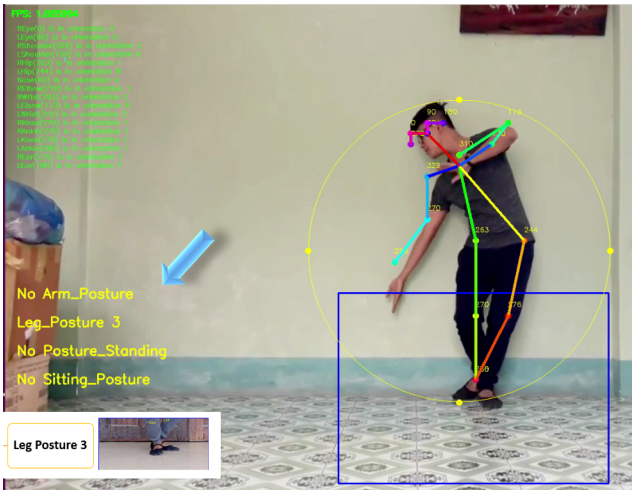

Figure 6: Using HPE and DCNN to detect the basis features

B. Using Deep Convolutional Neural Networks

In order to detect and recognize the significant postures (particularly in Leg, Standing and Sitting Posture), we proposed to utilize deep convolutional neural networks (DCNNs) which have been applied to visual tasks since the late 1980s.

As we had known, there are three main types of layers used to build Deep CNN architectures: convolutional layer, pooling layer and fully connected layer. Most of the CNN architectures is obtained by stacking the number of these layers. Deep convolutional neural networks, trained on large datasets, achieve convincing results and are currently the state-of-the-art approach for this task illustrated in figure7.

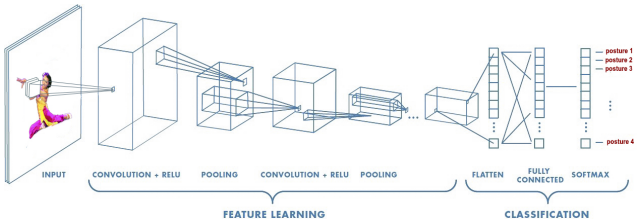

Figure 7: Neural network with many convolutional layers

Because the limited number of the postures features to represent EVTDs, we selected DCNNs to detect automatically those postures on the image frames of video. We utilized an open source neural network library written in Python called Keras [2] in which integrated many architectures being compatible with all the backends (TensorFlow, Theano, and CNTK).

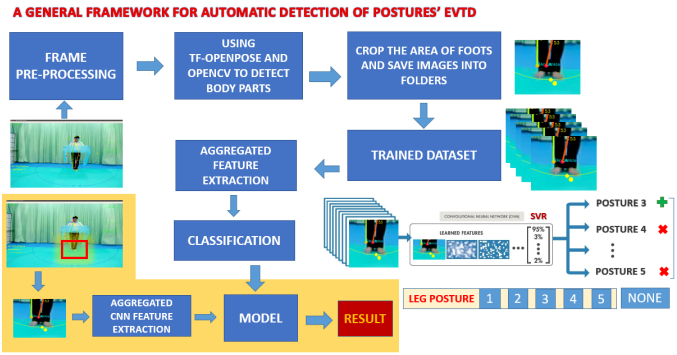

Figure 8: Framework to detect automatically dance postures

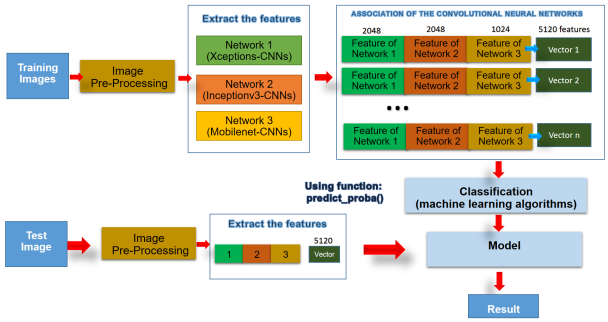

Figure 9: The aggregation of three DCNNs architectures to extract features

In this research, we implemented a general framework for automatic detection and classification of EVTD's fundamental postures as in figure 8. We introduce a classification model aggregating consequently the deep CNN architectures to extract the features (including Xceptions model, Inceptionv3 model and MobileNet model as in Figure 9). Realizing that each CNN architecture deals with the different cases of the particular dataset as well as combining the features to represent a vector is quite expected because it will be boosted strongly the discrimination between the classifications. In addition, the dataset of ETVDs focus principally on the high resolution image frames of dance videos, specially, the resolution and size of images will be grown rapidly in the future. For these reasons, we selected the methodology using deep CNNs to extract the significant features and ML algorithms to classify. After having sets of the collected images, we used several algorithms to advance the quality of images (image pre-processing). In each image frame, we extracted the features from three CNN architectures, particularlly, including Xceptions [35] (2048 features), InceptionV3 [36] (2048 features), Mobilenet [34] (1024 features). The next step, we aggregated consequently (respectively) the extracted features to have a feature vector with 5120 dimensions. After that, the comparison among several ML algorithms to have a best selection for classification is fully essential presented in section V.

The main idea of this model is a classification tool to update

[2]https://keras.io/models/model/

flexibly the novel architectures which will be published in the future aimed at advancing the accuracy in classification as soon as the dataset accelerated. In addition, we are also able to extract the features in parallel with each CNN architecture, nevertheless, we do not experiment the parallel models in this paper.

C. An automatic detection tool

In order to store and to detect the fundamental features into LPO, we implemented an automatic extraction tool of EVTD's movements (poses) by python language illustrated in figure 12. We utilized DCNN's architecture models proposed to classify the postures (leg, sitting and standing postures). Additionally, we also used HPE to describe each body parts of performer/dancer. Furthermore, we selected Owlready2 library (using python language) with HermiT reasoner to build our lightweight prioritized ontology. Our tool allowing users is able to extract automatically or extract manually (in each frame) features to put into ontology. As presented in above sections, each feature we inserted the probability of each movements to serve for selecting the suitable features in many sequential frames.

On the other hand, because the contrast between light and shade is different as well as the resolution of each video is also distinct, our tool implemented a simple adjustable set of image processing manually to support detection and classification.

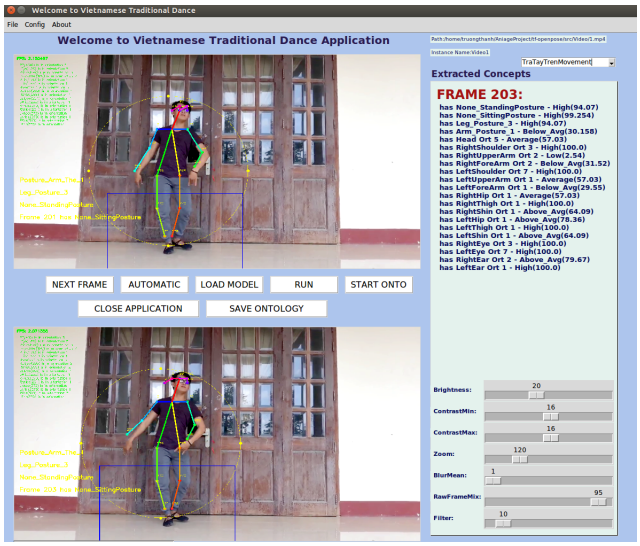

Figure 10: An automatic extraction tool of EVTDs

In order to be compatible for extracting the features of EVTDs where existing many fundamental movements separately. We also designed a simple function to change flexibly the classes/concepts for storing easily the different movements. In the next section, we will present experimental results and how to represent the extracted features into LPO with the probability.

Table I: Datasets of the fundamental posture of EVTDs

	Postures (Pos)						
	P1	**P2**	**P3**	**P4**	**P5**	**None Pos**	**Total**
Leg Postures	1260	685	1185	494	673	1407	5704
Sitting Posture	1252	798	none	none	none	2163	4213
Standing Posture	none	685	none	494	673	2084	3936

V. Experimental Result

We implemented the propositional tool on the computer supporting graphical card (NVIDIA GeForce GTX 950M with total memory is 8107 MB) to run TensorFlow on multiple GPUs. In this section, we would present two parts: (1) the experimental result of classifying the postures, (2) presenting LPO implemented and how to put the features into LPO by the tool implemented.

A. Classification of significant postures

We used python programming language and Keras libary to implement the EVTD postures automatic detection framework on each frame. We randomly split the dataset into two different training(2/3) and test (1/3) sets *(with the posture image datasets collected as in Table I, collected from 15 videos)*. In addition, we also implemented scikit-learn library [3] to use several ML algorithms including logic regression, Support vector machine (SVM - C=10000, Gama=0.002), Random Forest (200 decision trees), Stochastic Gradient Descent classifier (SGD), K-Nearest Neighbors (KNN - K=5), Naïve Bayes classifier. The experimental results of the propositional CNN model are in Table II for Rank-1 accuracy). The result achieved the high accuracy (F1) using Logic Regression as follows: 98.88% of Leg postures (for 06 classes), 99.84% of Sitting Posture (for 03 classes), 99.37% of Standing Posture (for 04 classes). In general, most of the classification results achieved the high accuracy around above 90%.

The result of this classification will be one of the preliminary of dataset collected. In implementation process, we realized that this collected dataset is not able to represent and reflect most of the different angle in dance. It exists some difficulties in classification including the difference between the postures of a professional dancer and an amateur person as well as the distinction from the distinct directions to look. Therefore, the necessary to collect a huge dataset is absolute expected in the future.

B. Implementation of EVTD's Lightweight Prioritized Ontology-based

Our LPO has 105 classes, 09 object properties, 09 data properties, 321 individuals and 2322 axiom including orientations, postures and body parts. It is implemented by

[3]http://scikit-learn.org/stable/

Table II: Comparisons of Rank-1 accuracy of algorithms

Algorithm	Postures		
	Leg Posture	Sitting Posture	Standing Posture
Logic Regression	98,88	99.84	99,37
SVM (Linear)	98,84	99.80	99,27
SGD Classifier	92,18	94,23	93,78
Random Forest	96,47	97,63	98,02
K-Nearest Neighbors	94,11	96,15	95,07
Naïve Bayes	96,30	97,06	97,34

owlready2 library [4], to store fundamental features (including the possibility of each movements characteristics). In this sub-section, we present how to manage significant features of EVTD postures into LPO.

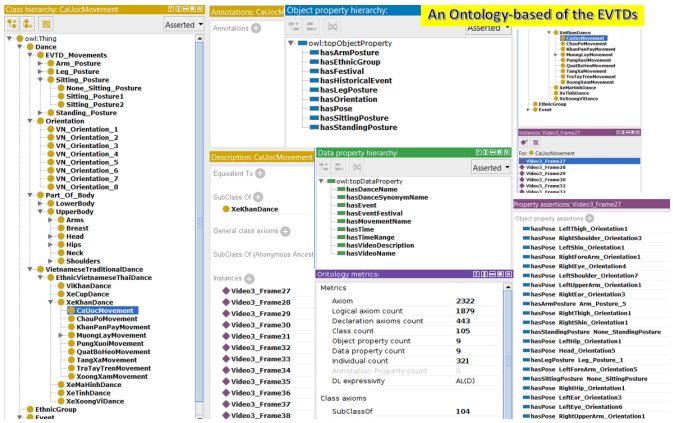

Figure 11: An Ontology-based Modelling of EVTDs

We proposed two methodologies to structure prioritized ABox as follows:(1) storing each significant feature with the specific probability as the left side of figure 12 (from 0% to 100%), (2) storing each essential characteristic with the probability range as right side of 12 that be classify including the levels: High (80%-100%), Above Average (60%-80%), Average (40%-60%), Below Average (20%-40%), Low (0%-20%) illustrated as top part of figure12.

In order to represent body parts, we implemented TF-Openpose to detect skeleton. From each body part extracted, we classified into eight orientations (O_i (i=1..8)) as in Figure 13. From each image flat and each body part, we identified the angle between x-axis with the body part orientation to classify, for example right-shoulder has the 93-degree angle, being O_5. In this demonstration, we only experimented with one dancer to represent and to store movements of EVTDs.

[4]https://pypi.org/project/Owlready2/

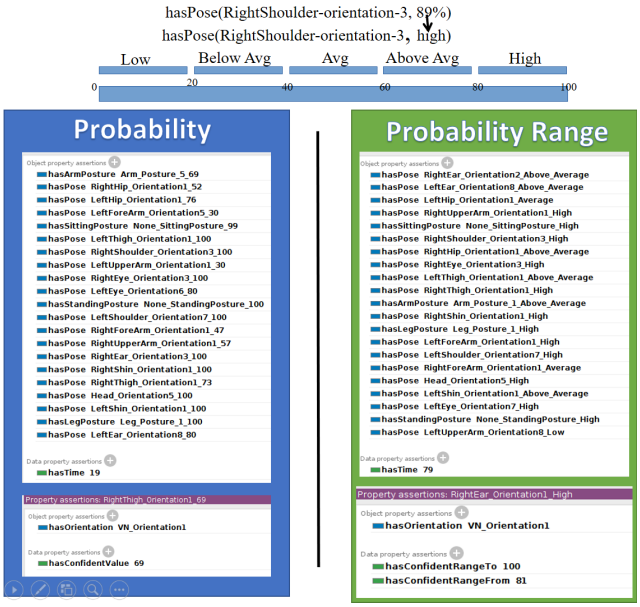

Figure 12: Range of Probability

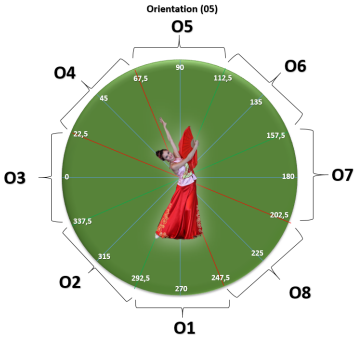

Figure 13: An orientation circle

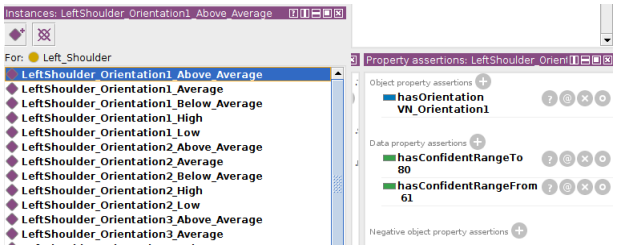

Figure 14: An example of left-shoulder part represented eight orientations with case of the probability range

Regarding the case of postures that did not belongs to the fundamental knowledge in [5], "non-posture" classification is utilized to represent. As discussed above, based on the probability of postures and poses that we will decide the selection of the essential frame to store into ontology.

VI. Conclusion and Future Works

With the aim of the preservation and promotion in the intangible cultural heritage in general as well as developing an application to store the Vietnamese traditional dances. In particular, we presented a methodology to identify automatically the significant concepts of EVTDs to build an intelligent repository. Using the machine learning algorithms combines with the CNN architectures and HPE to detect the important features are discussed in this paper. On the basis of the propositional tool, we collected and managed the heterogeneous dance data of EVTDs (from raw videos with low revolution).

The work presented in this paper is one of the important first steps regarding preserving and promotion of EVTDs based on the background of artificial intelligent. These initial steps would be the foundation for creating universal traditional dance repository to aim to support for advanced heterogeneous digital storage, indexing, classification, reasoning and searching dance videos. Based on the concepts extracted automatically and stored into prioritized ABox, the next step will classify, annotate and query-answering the dance videos of EVTDs. Our next plans include the collection of the dance dataset, improve how to detect the fundamental movements which is more compatible and build an universal lightweight ontology-based for EVTDs.

VII. Acknowledgements

This work has received support from the European Project H2020 Marie Sklodowska-Curie Actions (MSCA), Research and Innovation Staff Exchange (RISE): Aniage project (High Dimensional Heterogeneous Data based Animation Techniques for Southeast Asian Intangible Cultural Heritage Digital Content), project number 691215.

REFERENCES

[1] L.T.Loc, "Mua dan gian cac dan toc Viet Nam", in *Thoi-dai Publishing house*, 1994.

[2] H. Seo and P. Milanfar. Action recognition from one exam-ple.IEEE Trans. on Pattern Analysis and Machine Intelli-gence, 33(5):867–882, 2011

[3] V.Hoc, "Nghe thuat mua Viet Nam, thoang cam nhan", in *Nation Publishing House*, 2001.

[4] I. Laptev and T. Lindeberg. Velocity adaptation of space-time interest points. InICPR, pages 52–56, September 2004

[5] T.V. Son, Đ. T. Hoàn, N. T. M. Hương, "Mua dan gian mot so dan toc vung Tay Bac", in *Culture and Nation Publishing House*, 2003.

[6] Kyan, G. Sun, H. Li, L. Zhong, P. Muneesawang, N. Dong, B. Elder, and L. Guan, "An approach to ballet dance training through ms kinect and visualization in a cave virtual reality environment," In ACM Transactions on Intelligent Systems and Technology, vol. 6, no. 2, p. 23, 2015.

[7] [2] G. Chantas, A. Kitsikidis, S. Nikolopoulos, K. Dimitropoulos, S. Douka, I. Kompatsiaris, and N. Grammalidis, "Multi-entity bayesian networks for knowledge-driven analysis of ich content", in Proc. 1st International Workshop on Computer vision and Ontology Applied Cross-disciplinary Technologies in conj. with ECCV, pp. 355–369, Springer, 2014.

[8] A. Kitsikidis, K. Dimitropoulos, E. Yilmaz, S. Douka, and N. Grammalidis, "Multi-sensor technology and fuzzy logic for dancer motion analysis and performance evaluation within a 3d virtual environment", in Universal Access in Human-Computer Interaction. Design and Development Methods for Universal Access, pp. 379–390, Springer, 2014.

[9] A. Masurelle, S. Essid, and G. Richard, "Multimodal classification of dance movements using body joint trajectories and step sounds", in Proc. 14th International Workshop on Image Analysis for Multimedia Interactive Services, pp. 1–4, IEEE, 2013.

[10] A. Gupta, P. Srinivasan, J. Shi, and L. S. Davis. Understand-ing videos, constructing plots: Learning a visually groundedstoryline model from annotated videos. InCVPR, 2009.

[11] A. Klaser, M. Marszałek, C. Schmid, and A. Zisserman. Hu-man focused action localization in video. InInternationalWorkshop on Sign, Gesture, Activity, 2010

[12] M. S. Ryoo and J. K. Aggarwal. Recognition of compositehuman activities through context-free grammar based repre-sentation. InCVPR, pages 1709 – 1718, October 2006.

[13] M. Blank, L. Gorelick, E. Shechtman, M. Irani, and R. Basri.Actions as space-time shapes. InICCV, volume 2, pages1395–1402, October 2005.

[14] L.N. Canh, "Nghệ thuật múa Hà Nội truyền thống và hiện đại", *in Hà Nội Publishing House*, 2011.

[15] Karavarsamis S., Ververidis D., Chantas G., Nikolopoulos S., Kompatsiaris Y. Classifying salsa dance steps from skeletal poses; Proceedings of the 2016 14th International Workshop on Content-Based Multimedia Indexing (CBMI); Bucharest, Romania. 15–17 June 2016; pp. 1–6.

[16] T. Lan, Y. Wang, W. Yang, and G. Mori. Beyond actions:Discriminative models for contextual group activities. InNIPS, 2010.

[17] L.N. Canh, "Nghệ thuật múa truyền thống Khmer Nam bộ", *in Vietnamese Dance College, Cultural and Nation publishing house*, 2013.

[18] Y. Wang and G. Mori. Human action recognition by semi-latent topic models.IEEE Trans. on Pattern Analysis andMachine Intelligence Special Issue on Probabilistic Graphi-cal Models in Computer Vision, 31(10):1762–1774, 2009.

[19] Y. Wang and G. Mori. Hidden part models for human ac-tion recognition: Probabilistic vs. max-margin.IEEE Trans.on Pattern Analysis and Machine Intelligence, 33(7):1310–1323, 2011

[20] N. Nayak, R. Sethi, B. Song, and A. Roy-Chowdhury. Mo-tion pattern analysis for modeling and recognition of com-plex human activities.Visual Analysis of Humans: Lookingat People, Springer, 2011.

[21] L.N.Canh. "Múa tín ngưỡng dân gian Việt Nam", in *Social Science Publishing house*, 1998.

[22] W. Yang, Y. Wang, and G. Mori. Recognizing human actions from still images with latent poses. In CVPR, pages 2030–2037, June 2010.

[23] B. Yao, A. Khosla, and L. Fei-Fei. Classifying actions and measuring action similarity by modeling the mutual context of objects and human poses. InICML, Bellevue, USA, June 2011.

[24] J. K. Aggarwal and M. S. Ryoo. Human activity analysis: Areview.ACM Computing Surveys(To appear), 2011

[25] L.N.Canh. "Đại cương nghệ thuật múa", in *Culture and information publishing house*, 2003.

[26] D.Lin and P.Paten. Induction of semantic classes from natural language text. In proceedings of SIGKDD-01. pp.317-322. 2001.

[27] S. Maji, L. Bourdev, and J. Malik. Action recognition from adistributed representation of pose and appearance. InCVPR,pages 3177–3184, June 2011.

[28] A.G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, H. Adam, "MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications", arXiv preprint, arXiv:1704.04861, 2017.

[29] Z. Cao, T. Simon, S. Wei, Y. Sheikh, "Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields", arXiv preprint, arXiv:1611.08050, 2016.

[30] F. Chollet. Xception: Deep learning with depthwise separable convolutions. arXiv preprint arXiv:1610.02357, 2016.

[31] X. Xia, C.Xu, Inception-v3 Flower Classification, 2017 2nd International Conference on Image, Vision and Computing, 978-1-5090-6238-6/17/ ©2017 IEEE

[32] K. El Raheb, Y.Ioannidis, "A Labanotation based Ontology for Representing Dance Movement", in Proceedings of Gesture Workshop, Athens, 2011.

[33] El Raheb, K., Papapetrou, N., Katifori, V., Ioannidis, Y. BalOnSe:Ballet Ontology for Annotating and Searching Video performances. In Proceed-

ings of the 3rd International Symposium on Movement and Computing (p. 5). ACM, 2016.
[34] A.G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, H. Adam, "MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications", arXiv preprint, arXiv:1704.04861, 2017.
[35] F. Chollet. Xception: Deep learning with depthwise separable convolutions. arXiv preprint arXiv:1610.02357, 2016.
[36] X. Xia, C.Xu, Inception-v3 Flower Classification, 2017 2nd International Conference on Image, Vision and Computing, 978-1-5090-6238-6/17/ ©2017 IEEE