This manuscript has been accepted for publication in the Proceedings of the 2022 HCI International Conference.
Cite as: Guo, W. et al. (2022). AI-Driven Human Motion Classification and Analysis Using Laban Movement System.
In: Duffy, V.G. (eds) Digital Human Modeling and Applications in Health, Safety, Ergonomics and Risk Management.
Anthropometry, Human Behavior, and Communication. HCII 2022. Lecture Notes in Computer Science, vol 13319. Springer, Cham.

AI-driven Human Motion Classification and Analysis using Laban Movement System

Wenbin Guo[1], Osubi Craig[2], Timothy Difato[1], James Oliverio[1],
Markus Santoso[1], Jill Sonke[3], Angelos Barmpoutis[1]

[1] Digital Worlds Institute, University of Florida, Gainesville, Florida, 32611, USA
[2] Center for the Arts, Migration, and Entrepreneurship, University of Florida, Gainesville, Florida, 32611, USA
[3] Center for the Arts in Medicine and Assist, University of Florida, Gainesville, Florida, 32611, USA
{wenbin, tim,james,markus,angelos}@digitalworlds.ufl.edu,
{ocraig,jsonke}@arts.ufl.edu

Abstract. Human movement classification and analysis are important in the research of health sciences and the arts. Laban movement analysis is an effective method to annotate human movement in dance that describes communication and expression. Technology-supported human movement analysis employs motion sensors, infrared cameras, and other wearable devices to capture critical joints of the human skeleton and facial key points. However, the aforementioned technologies are not mainstream, and the most popular form of motion capture is conventional video recording, usually from a single stationary camera. Such video recordings can be used to evaluate human movement or dance performance. Any methods that can systematically analyze and annotate these raw video footage would be of great importance to this field. Therefore, this research offers an analysis and comparison of AI-based computer vision methods that can annotate the human movement automatically. This study trained and compared four different machine learning algorithms (random forest, K neighbors, neural network, and decision tree) through supervised learning on existing video datasets of dance performances. The developed system was able to automatically produce annotation in the four dimensions (effort, space, shape, body) of Laban movement analysis. The results demonstrate accurately produced annotations in comparison to manually entered ground truth Laban annotation.

Keywords: Artificial intelligence, Human motion classification, Laban movement analysis

1 Introduction

Human movement has been studied in multiple disciplines, including health sciences and the Arts, resulting in a large but disparate assortment of multi-modal datasets, including video, skeletal motion capture, manual annotations, and clinical metadata.

Laban notation is a standardized form of kinetograph annotation [1, 2]. Although it was originally proposed for vector-based choreographic transcription [3], it has been successfully applied to several fields that study human motion [4], such as neuroscience [5], kinesiology [6], human-computer interaction [7], as well as theater and dance [8]. Traditional data collection processes often include Laban movement analysis and annotation that parameterizes observed changes in pre-defined 4-dimensional feature space (effort, space, shape, body). Such analysis requires lengthy manual input from professionals who annotate the recorded data through a time-consuming "watch and pause" process, which is also prone to human errors. The majority of automated Laban analysis focuses on processing 3D point sets (skeletal sequences) captured by specialized motion tracking equipment that may require the installation of markers on the user's body and calibration of the devices [6, 8]. However, these technologies are not mainstream, and until today the most commonly used form of observation and documentation of human motion is conventional video recording, typically done from a single stationary camera discretely placed within the professional setting.

In this project, we proposed using Artificial Intelligence (AI) methods to fully automate the annotation process involved in Laban analysis by training and testing different machine learning algorithms on existing video datasets of human motion, focusing on performative movement. First, we trained four different machine learning algorithms through supervised learning on existing dance video datasets in this study. Second, this study tested feature extraction methods (within and across frames) to improve the annotation accuracy. The trained model was then tested in Laban-annotating existing video sequences and validated using manually produced Laban annotation, which was considered our ground truth. Finally, a software application was developed that can be used by researchers to input raw videos and export automatically produced Laban annotation.

2 Related Works

Laban movement analysis was originally developed by Rudolf Laban [2]. Laban system has been widely used for human movement analysis and annotation, and its reliability has been extensively studied and evaluated [9]. Figure 1 shows the detailed hierarchy of the Laban movement analysis in four dimensions. Shape quality has six elements: opening, enclosing, rising, sinking, advancing, and retreating. Shape quality can be described by observing the positioning of the body along the vertical, horizontal, and sagittal axes. Effort quality has eight elements: light, strong, free, bound, sustained, quick, indirect, and direct. The majority of the elements in effort quality can be related to motion features such as velocity and acceleration. The space zone has six elements: side-open, side-across, up, down, forward, and backward. These elements describe the trajectory of human motions. Body phrasing has three elements: impulsive, swing, and impactive. These elements explain the kinematic chains and global locomotion.

With the use of motion tracking technologies, there have been numerous applications of Laban movement analysis in dance emotion recognition [10], folk dance evaluation [8], parameterizing interpersonal behavior [6], and signal interpretation [11], among others. Several of these examples were focused on behavior patterns recognition and human-computer interaction using Laban movement analysis, and the majority of them have been restricted to the effort and shape dimensions of the 4-dimensional Laban annotation space.

More recently, Microsoft Kinect sensors have been employed for Laban-based motion tracking and analysis [12–14]. More specifically, Ajili [13] compared the effectiveness of machine learning methods to classify human actions using Kinect sensors. The results indicated robust classification across the Laban dimensions. Similarly, Kim [14] extracted motion features such as velocity and acceleration of joints to analyze the Laban movement. Although motion sensors such as Kinect have been consistently becoming more affordable, they are still not in mainstream use by consumers. On the contrary, the popularity of video sharing and streaming platforms, ranging from teleconferencing tools to social media, have established conventional video recording (without additional sensors) as the dominant form of recording and documenting the human activity. In professional settings (such as clinical, performative, etc.), the use of a single stationary video camera has been the current standard for observing and recording human movement.

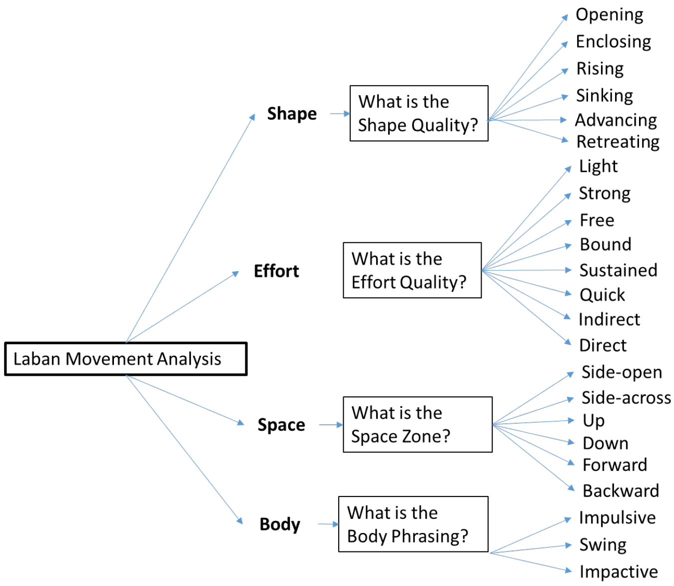

Fig. 1. Laban movement annotation hierarchy

Our method involved the processing of raw video footage of human movement using a sequence of artificial intelligence algorithms. First, human skeleton detection algorithms extracted key skeletal features from the input videos. Then feature vectors were calculated from in-frame and across-frames values and were used to train and test the effectiveness of four machine learning algorithms (random forest, K neighbors, neural network, and decision tree) in automatic Laban classification across the four dimensions of the Laban analysis space. Random forest is an estimator that fits many decision tree classifiers on various sub-samples of the dataset and uses averaging to improve the predictive accuracy and control over-fitting. K neighbors is a classifier that captures the similarity, such as calculating the distance between points on a graph. The neural network is a multi-layer perceptron classifier to optimize the log-loss function. The feedforward network generates a set of outputs from a set of inputs in the neurons.

3 Methods

As in the majority of AI-based methods, our process included the following four milestones: 1) Data preparation, 2) Feature extraction, 3) Training and testing, 4) Automated annotation.

In this pilot study, we used performative video datasets from the Digital Worlds Institute at the University of Florida and videos from the Dance Motion Capture Database of the University of Cyprus [15]. A selection of four videos was manually annotated using Laban movement analysis. In order to identify the skeleton and body joints in the raw input videos, we used Facebook Detectron2 [16], which is an open-source platform for human motion analysis.

Features extracted from single frames (in-frame features) were initially used as the baseline to check the accuracy of four traditional machine learning algorithms. After finding the best algorithms, features within and across frames were extracted to improve the annotation accuracy. Figure 2 shows the total of 17 key points extracted from the video, including nose, left eye, right eye, left ear, right ear, left shoulder, right shoulder, left elbow, right elbow, left wrist, right wrist, left hip, right hip, left knee, right knee, left ankle, and right ankle.

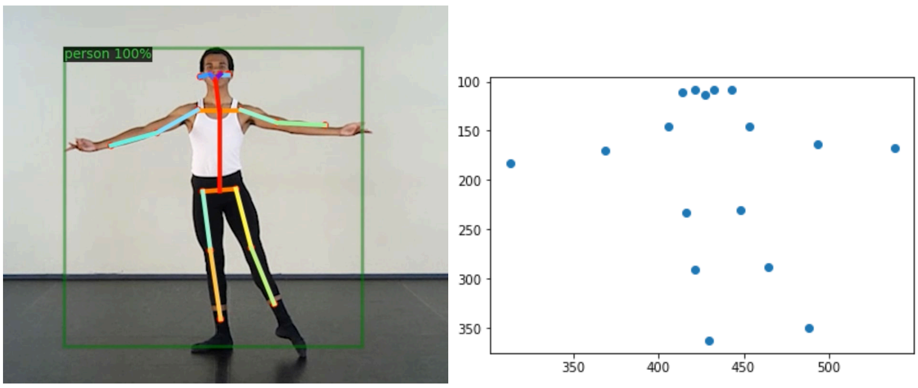

Fig. 2. Skeleton detection from input video (left), Extracted body key points (right)

The x, y coordinates of each key point were estimated for each frame of the video, generating 34 time-series/signals for our analysis. Figure 3 shows the x and y coordinates as signals in the left and right plots, respectively. By visually inspecting these plots, it is evident that patterns can be observed from the corresponding motion activity. Subsequently, these features and other quantities derived from these features, such as distance, velocity, and acceleration, can be correlated with the four dimensions of the Laban analysis system.

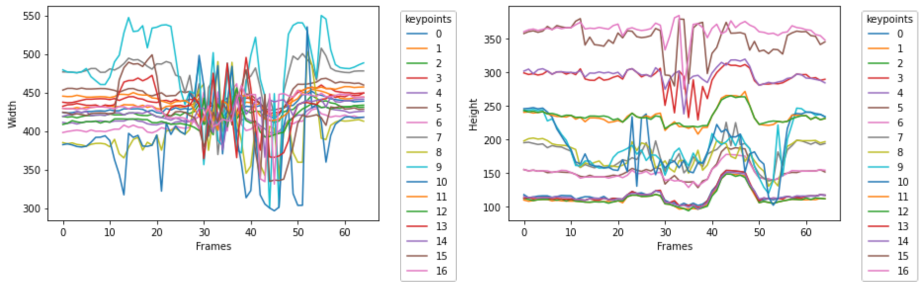

Fig. 3. Plots of the human joint movements. Left: x-coordinate, right: y-coordinate.

After the feature extraction step, we trained four machine learning models to receive a video sequence as input and recognize the body motion changes across the 4-dimensional domain used in Laban annotation, including effort, space, shape, and body. Table 1 shows an example of Laban movement annotation in four dimensions. Each dimension has two elements, and this study created four binary classifiers to annotate the human movement.

Table 1. Laban movement annotation

#	Shape	Effort	Space	Body
1	Enclosing	Light	Side-open	Impulsive
2	Opening	Strong	Side-across	Swing

For our training dataset, we used dance videos from the Graphics Lab at the University of Cyprus. Additional video datasets from the University of Florida Digital Worlds Institute were used for evaluation and testing. We randomly selected 70% of the compiled dataset for supervised training, and the remaining 30% was used for testing. The annotation system was implemented in Python using existing libraries (Scikit-learn) to expedite the project's prototyping phase. The system was developed and tested in UF High-Performance Computing Facility (HiPerGator). Facebook Detectron2 was used to detect the human key points, estimating a total of seventeen key points for each human skeleton. The feature vector that we used to detect the Laban annotation consisted of the velocity and acceleration of each detected key point as well as distances between specific key points, as listed in Table 2.

More specifically, for the experiments in this pilot study, we calculated the distance, velocity, and acceleration between ankle and hip, velocity and acceleration of ankle, hip, and wrist [8]. Each dimension in Laban movement analysis contained multiple elements, and not all the videos covered every element of four dimensions.

Therefore, two elements in each dimension were used in this study, and multilabel classification was generated.

Table 2. Features for Laban annotation

Distance	Velocity	Acceleration
Ankle and hip	ankle	ankle
Wrist and nose	Hip	Hip
Wrist and hip	Wrist	Wrist

In our experiments, we also compared distances between adjacent joints in order to study the suitability of each feature in our feature vector (such as the distance from hip to knee, elbow to wrist, elbow to shoulder, and shoulder to shoulder). Hip to knee distance had a minimum standard deviation, and it was regarded as a normalization factor. Based on the average thigh length of humans, this study normalized the data based on the hip to knee distance.

Performance evaluation was based on the precision, recall, and F1 scores. Precision is the ability of the classifier not to label as positive a sample that is negative. The recall is the ability of the classifier to find all the positive samples, and its value ranges from 0-1, with 1 being the best. F1 can be interpreted as a harmonic mean of the precision and recall, and its value also ranges from 0-1. The relative contribution of precision and recall to the F1 score are equal. The formula for the F1 score is:

$$F1 = 2 * (\text{precision} * \text{recall}) / (\text{precision} + \text{recall}) \quad (1)$$

Finally, the automated annotation was visualized by Matplotlib, which was a comprehensive library for creating animated and interactive visualization.

4 Results

Multilabel classification treats each label independently whereas multilabel classifiers treat the multiple classes simultaneously. The multilabel approach taken in this pilot study is to break the Laban movement analysis problem into four binary classification tasks and one for each Laban dimension. Each binary classifier decides body (impulsive or swing), shape (enclosing or opening), effort (light or strong), and space (side-open or side-across), respectively. Table 3 shows the results of Laban annotation using four machine learning algorithms, including random forest, K neighbors, neural network, and decision tree. It should be noted that all four algorithms were used in the multilabel classification framework.

The data indicated that Random forest had the best performance compared to the other machine learning algorithms. Additional experiments were performed with larger feature vectors that were generated by adding cross-frame features. The added features were the average, maximum, minimum, and standard deviation of the original frames for a duration of two seconds. These additional features increased the accuracy of Laban annotation, especially the space dimension, side-across, and side-open, which are highly related to the content of the previous frames, as reported in Table 4.

This pilot study also explored the accuracy based on different frame intervals. The frame interval experiments were varied with increments of 10 frames at a time because the manually produced labels were based on ten frames interval. Cumulative mean, standard deviation, maximum, minimum for 10-60 frames interval were calculated as the input data for training. Based on the results, the frame interval that produces the best results was 60 frames.

Table 3. Laban annotation using four machine learning algorithms

	Random Forest			**K neighbors**			**Neural Network**			**Decision Tree**		
	P	R	F1	P	R	F1	P	R	F1	P	R	F1
enclosing	1.00	0.97	0.98	0.93	0.93	0.93	0.85	0.97	0.90	0.80	0.97	0.88
impulsive	1.00	0.45	0.62	0.50	0.09	0.15	0.56	0.43	0.49	0.38	0.35	0.36
light	0.98	1.00	0.99	0.97	0.99	0.98	0.97	0.98	0.98	0.97	0.97	0.97
opening	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.99	0.99	1.00	0.98	0.99
side-across	0.81	0.76	0.79	0.79	0.67	0.72	0.79	0.79	0.79	0.69	0.73	0.71
side-open	0.80	0.84	0.82	0.81	0.73	0.77	0.81	0.81	0.81	0.74	0.70	0.72
strong	1.00	0.45	0.62	0.50	0.10	0.17	0.50	0.40	0.44	0.40	0.40	0.40
swing	0.98	1.00	0.99	0.96	0.99	0.98	0.97	0.98	0.98	0.97	0.97	0.97

Table 4. Laban annotation using cumulative frames

	Random Forest (Single frame)			**Random Forest (Cross frame)**		
	P	R	F1	P	R	F1
enclosing	1.00	0.97	0.98	0.88	0.78	0.82
impulsive	1.00	0.45	0.62	0.94	0.61	0.74
light	0.98	1.00	0.99	0.98	1.00	0.99
opening	1.00	1.00	1.00	0.99	0.99	0.99
side-across	0.81	0.76	0.79	0.95	0.91	0.93
side-open	0.80	0.84	0.82	0.92	0.96	0.94
strong	1.00	0.45	0.62	0.94	0.61	0.74
swing	0.97	1.00	0.99	0.98	1.00	0.99

Figure 4 describes the confusion matrix of Laban annotation using random forest (Cross frame). The results showed the high accuracy of the detection in Laban movement four dimensions including eight elements.

Figure 5 shows an example of our results from a single frame of one of the input videos. The automatically detected human skeleton and the corresponding key points are superimposed on the image. For this frame, the automated Laban annotation in four dimensions was: Shape=opening, Effort=light, Space=side-open, Body=swing.

Our developed software was able to successfully produce Laban annotation for each frame of an input video. Our classifier had two elements for each Laban dimension: effort (light, strong), shape (enclosing, opening), space (side-open, side-across), and body (swing, impulsive). The multilabel binary classifier could produce the annotation such as “light, opening, side-open, swing” for each frame of the input video. The labels reflected the human movement in each frame.

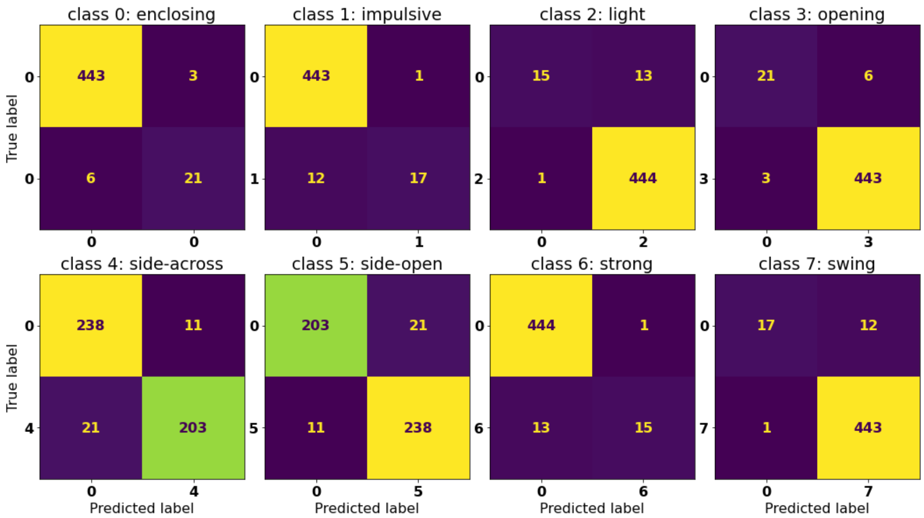

Fig. 4. Confusion matrix of each Laban movement dimension

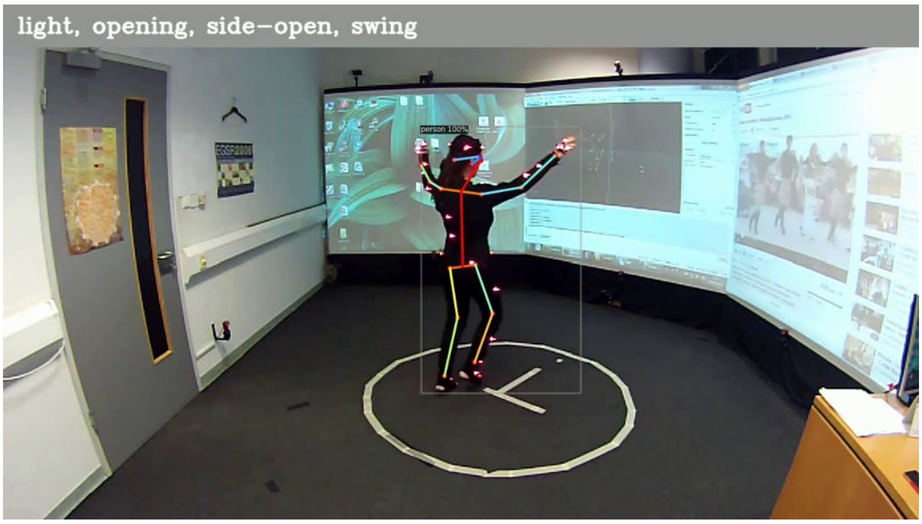

Fig. 5. Sample demonstration of our results. Detected skeletal features are shown in the frame. Automated Laban classification result for this frame: Effort=light, Shape=opening, Space=side-open, Body=swing.

5 Conclusion

In this pilot study, we developed a method for recognizing Laban movement dimensions from input videos using machine learning algorithms. The study trained and compared four different machine learning algorithms (random forest, K neighbors, neural network, and decision tree) through supervised learning on existing human motion video datasets. Our results obtained high accuracy in three dimensions: shape (enclosing, opening), effort (light, strong), and body (swing, impulsive) for the single-frame analysis. The annotation accuracy was increased in space (side-open, side-across) when cross-frame features were included in our feature vector. These results demonstrated that Laban movement analysis could be fully automated using machine learning algorithms that operate on raw input videos and that AI-driven standardized human motion evaluation systems can be developed.

Our results indicated that such AI-driven movement classification could enable fully automated reporting in commonly used Laban annotation using only a single stationary camera. Furthermore, an AI-based solution to this problem may lead to standardization of data processing and optimize professionals' time and decrease human errors. In the future, this can significantly impact the health science and art areas that use Laban analysis, by bringing AI to new domains and applications such as atlas construction of choreographic data and clinical assessment in dance therapy.

6 Limitations and Future Work

Although AI-based Laban annotation has already shown the improvement of efficiency, this study only trained a limited number of four videos, and more datasets need to be trained in the future. Only a specific genre of dances was included in the dataset; inclusion of more genres may lead to larger variability of cross frames features based on the rhythms and motion patterns. Our binary classifier only detected two elements in each Laban dimension, and all the elements in the four dimensions need to be tested in the Laban annotation in the future.

Facebook Detectron2 only contains x, y coordinates which may bias the annotation to patterns that are visible in the 2-dimensional plane of the video frame. Newer pose estimation models with 3D coordinate detection need to be tested in the future. The inclusion of the z-dimension (depth) of the joints may help us detect more elements in the Laban movement analysis.

In the future, more algorithms can be tested using Tensorflow, such as Convolutional Neural Network (CNN), through supervised learning on the time-series of the features extracted from the video datasets.

Acknowledgment

This project was supported by the 2020 UF AI Catalyst grant from the UF Office of Research. The authors acknowledge the University of Florida Research Computing for providing computational resources and support that have contributed to the research results reported in this publication.

References

1. Hutchinson, A., Guest, A. H., & Hutchinson, W. A. (1977). *Labanotation: or, kinetography Laban: the system of analyzing and recording movement*. Taylor & Francis.
2. Laban, R. (1966). Choreutics: Annotated and Ed. *London: Macdonald & Evans*.
3. Longstaff, J. S. (2001). Translating "vector symbols" from Laban's choreographie. *Twenty-Second Biennial Conference of the International Council of Kinetography Laban, ICKL, 26 July-2 August, Ohio State University, Columbus, Ohio. USA*, *4*(1926), 70–86. Retrieved from http://www.ickl.org/conf01_proceedings/ICKL01_070_086.pdf
4. Davies, E. (2007). *Beyond dance: Laban's legacy of movement analysis*.
5. Foroud, A., & Whishaw, I. Q. (2006). Changes in the kinematic structure and non-kinematic features of movements during skilled reaching after stroke: A Laban Movement Analysis in two case studies. *Journal of Neuroscience Methods*, *158*(1), 137–149. https://doi.org/10.1016/j.jneumeth.2006.05.007
6. Roudposhti, K. K., Santos, L., Aliakbarpour, H., & Dias, J. (2012). Parameterizing interpersonal behaviour with Laban movement analysis - A Bayesian approach. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, 7–13. https://doi.org/10.1109/CVPRW.2012.6239349
7. Rett, J., Dias, J., & Ahuactzin, J. M. (2010). Bayesian reasoning for Laban Movement Analysis used in human-machine interaction. *International Journal of Reasoning-based Intelligent Systems*, *2*(1), 13–35. https://doi.org/10.1504/IJRIS.2010.029812
8. Aristidou, A., Stavrakis, E., Charalambous, P., Chrysanthou, Y., & Himona, S. L. (2015). Folk dance evaluation using laban movement analysis. *Journal on Computing and Cultural Heritage*, *8*(4), 1–19. https://doi.org/10.1145/2755566
9. Bernardet, U., Alaoui, S. F., Studd, K., Bradley, K., Pasquier, P., & Schiphorst, T. (2019). Assessing the reliability of the Laban Movement Analysis system. *PLoS ONE*, *14*(6), 1–23. https://doi.org/10.1371/journal.pone.0218179
10. Wang, S., Li, J., Cao, T., Wang, H., Tu, P., & Li, Y. (2020). Dance Emotion Recognition Based on Laban Motion Analysis Using Convolutional Neural Network and Long Short-Term Memory. *IEEE Access*, *8*, 124928–124938. https://doi.org/10.1109/ACCESS.2020.3007956
11. Santos, L., & Dias, J. (2011). Motion patterns: Signal interpretation towards the Laban Movement Analysis semantics. *IFIP Advances in Information and Communication Technology*, *349 AICT*, 333–340. https://doi.org/10.1007/978-3-642-19170-1_36
12. Bernstein, R., Shafir, T., Tsachor, R., Studd, K., & Schuster, A. (2015). Laban Movement Analysis Using Kinect. *International Journal of Computer, Electrical, Automation, Control and Information Engineering*, *9*(6), 1574–1578.
13. Ajili, I., Mallem, M., & Didier, J. Y. (2017). Robust human action recognition system using Laban Movement Analysis. *Procedia Computer Science*, *112*, 554–563. https://doi.org/10.1016/j.procs.2017.08.168
14. Kim, W. H., Park, J. W., Lee, W. H., Chung, M. J., & Lee, H. S. (2013). LMA based

emotional motion representation using RGB-D camera. *ACM/IEEE International Conference on Human-Robot Interaction*, 163–164. https://doi.org/10.1109/HRI.2013.6483552

15. University of Cyprus. (2018). Dance Motion Capture Database. Retrieved December 18, 2021, from http://dancedb.cs.ucy.ac.cy/main/performances
16. Wu, Y., Kirillov, A., Massa, F., Lo, W.-Y., & Girshick, R. (2019). Detectron2. Retrieved December 20, 2021, from https://github.com/facebookresearch/detectron2