AIST DANCE VIDEO DATABASE: MULTI-GENRE, MULTI-DANCER, AND MULTI-CAMERA DATABASE FOR DANCE INFORMATION PROCESSING

Shuhei Tsuchida **Satoru Fukayama** **Masahiro Hamasaki** **Masataka Goto**
National Institute of Advanced Industrial Science and Technology (AIST), Japan
{s-tsuchida, s.fukayama, masahiro.hamasaki, m.goto}@aist.go.jp

ABSTRACT

We describe the *AIST Dance Video Database (AIST Dance DB)*, a shared database containing original street dance videos with copyright-cleared dance music. Although dancing is highly related to dance music and dance information can be considered an important aspect of music information, research on dance information processing has not yet received much attention in the Music Information Retrieval (MIR) community. We therefore developed the AIST Dance DB as the first large-scale shared database focusing on street dances to facilitate research on a variety of tasks related to dancing to music. It consists of 13,939 dance videos covering 10 major dance genres as well as 60 pieces of dance music composed for those genres. The videos were recorded by having 40 professional dancers (25 male and 15 female) dance to those pieces. We carefully designed this database so that it can cover both solo dancing and group dancing as well as both basic choreography moves and advanced moves originally choreographed by each dancer. Moreover, we used multiple cameras surrounding a dancer to simultaneously shoot from various directions. The AIST Dance DB will foster new MIR tasks such as dance-motion genre classification, dancer identification, and dance-technique estimation. We propose a dance-motion genre-classification task and developed four baseline methods of identifying dance genres of videos in this database. We evaluated these methods by extracting dancer body motions and training their classifiers on the basis of long short-term memory (LSTM) recurrent neural network models and support-vector machine (SVM) models.

1. INTRODUCTION

The Music Information Retrieval (MIR) community started with standard music information such as musical audio signals and musical scores, and then extended its scope to other types of music-related multimodal information such

© Shuhei Tsuchida, Satoru Fukayama, Masahiro Hamasaki, Masataka Goto. Licensed under a Creative Commons Attribution 4.0 International License (CC BY 4.0). **Attribution:** Shuhei Tsuchida, Satoru Fukayama, Masahiro Hamasaki, Masataka Goto. "AIST Dance Video Database: Multi-Genre, Multi-Dancer, and Multi-Camera Database for Dance Information Processing", 20th International Society for Music Information Retrieval Conference, Delft, The Netherlands, 2019.

Figure 1. Snapshots of AIST Dance Video Database (AIST Dance DB), which features **multiple genres** (10 major dance genres), **multiple dancers** (solo and group dancing by 40 professional dancers), and **multiple cameras** (at most 9 video cameras surrounding a dancer). These color snapshots are cropped to enlarge dancers in videos.

as images, videos, lyrics, and social-media data. Since dance music – a popular target of MIR research [7, 19, 20, 22, 33, 40, 42, 46, 55, 61, 66, 80] – is originally written for dance and has a strong connection to dance motions, *dance information*, that is, any kind of information related to dance, such as dance music, dance motions, and dancers, can be considered an important aspect of music information that the MIR community should cover. Research on dance motions, however, has not yet received much attention in the community. The goal of this research is to develop a dance information database including original dance videos with copyright-cleared dance music and make it publicly available to the community so that research on dance can attract more attention in the community and new MIR tasks related to dance can emerge in the future.

As a sub-area of MIR research, we propose defining *dance information processing* as meaning various types of processing and research related to dance information.

First, it is necessary for dance information processing to handle dance music. There have been studies on traditional dance tunes [7, 22, 40, 61], electronic dance music [19, 42, 46, 55, 66, 80, 82], and dance-music classification [20, 33]. Second, it is important for dance information processing to advance research related to dance motions. There have been various related studies such as on dance-motion choreography generation driven by music [2,28,29,32,52,53,73], rhythm estimation from dance motions in dance videos [18], music retrieval by dance motions [75], controlling music tempo by dance motions [37], dance-motion identification [60], and dance-motion genre classification [43]. In addition, research on dance motions could have good synergy with research on performer motions made during musical performances [30,47,48,51,67] since both deal with music-related human body movements. This emerging field of dance information processing, however, has lacked a large systematic dance information database that is available to researchers for common use and research purposes.

We therefore built the *AIST Dance Video Database (AIST Dance DB)*, the first large-scale copyright-cleared database focusing on street dances (Figure 1). The AIST Dance DB consists of 13,939 original dance videos covering 10 major street dance genres (break, pop, lock, waack, middle hip-hop, LA-style hip-hop, house, krump, street jazz, and ballet jazz) and 60 original pieces of dance music composed for these genres, each having 6 different musical pieces with different tempi. We had 40 professional dancers (25 male and 15 female), each having more than 5 years of dance experience, dance to those pieces. In addition to 13,890 videos for which each of the 10 genres has 1,380 videos, we added 49 videos that cover 3 typical dancing situations (showcase, cypher, and battle) in which a group of dancers enjoy dancing. The database is carefully designed to cover both solo dancing (12,990 videos) and group dancing (949 videos) as well as both basic choreography moves (10,800 videos) and advanced moves (3,139 videos) originally choreographed by each dancer. We used at most nine video cameras surrounding a dancer to simultaneously shoot from various directions.

To the best of our knowledge, such street dance videos have not been available to researchers, so they will become valuable research materials to be analyzed in diverse ways and used for various machine-learning purposes. For example, the AIST Dance DB will foster new MIR tasks such as dance-motion genre classification, dancer identification, and dance-technique estimation. As a basic task of dance information processing, we propose a dance-motion genre-classification task for street dances. We developed four baseline methods for identifying dance genres of a subset of the dance videos in this database and evaluated them by extracting dancer body motions and training their classifiers on the basis of long short-term memory (LSTM) recurrent neural network models and support vector machine (SVM) models. In our preliminary experiments, we tested the methods on 210 dance videos of 10 genres done with 30 dancers and found that dance genres were identified with a high accuracy of 91.4% when a 32-sec excerpt was given; however, the accuracy dropped to 56.6% when a 0.67-sec excerpt was given. These results can be used as a baseline performance for this task.

2. RELATED WORK

Since the importance of research databases has been widely recognized in various research fields, researchers interested in dance have also spent considerable effort on building dance-related databases [62]. For example, the Martial Arts, Dancing and Sports Dataset [82] has six videos for both the hip-hop and jazz dance genres. These videos were shot from three directions, and the video data includes depth information. However, six videos for each genre is not enough for some tasks, and the videos do not contain enough professional dance motions to effectively express the characteristics of each genre. Several relatively small dance datasets have also been published [14, 17, 63, 79].

Some databases not only have dance videos but also provide different types of sensor data. Stavrakis et al. [70] published the Dance Motion Capture Database, which provides high-quality motion capture data as well as dance videos. This database includes Greek and Cypriot dances, contemporary dances, and many other dances such as flamenco, the belly dance, salsa, and hip-hop. Tang et al. [72] also used a motion capture system to build a dance dataset that contains four dance genres: cha-cha, tango, rumba, and waltz. Essid et al. [23] published a dance dataset that consists of videos of 15 salsa dancers, each performing 2 to 5 fixed choreographies. They were captured using a Kinect camera and five cameras. This dataset is unique since all video data contain inertial sensor (accelerometer + gyroscope + magnemeter) data captured from multiple sensors on the dancers' bodies. All these databases, however, do not handle street dance videos performed by multiple dancers and recorded with multiple camera directions.

Although YouTube-8M [1] and Music Video Dataset [65] were not built for research on dance, the former contains 181,579 dance videos and the latter contains 1,600 music videos including dance-focused videos. It is difficult to use these videos for dance information processing because the videos are not organized from this viewpoint. In comparison, the AIST Dance DB contains videos that are systematically recorded and organized to have attributes such as dance genre names, the names of basic dance moves, labels of dance situations, and information on dancers and musical pieces.

3. AIST DANCE VIDEO DB

3.1 Design policy

We designed the AIST Dance DB by considering the following three important points.

- **Suite of video files of original dance and audio files of copyright-cleared dance music**
  To analyze and investigate the relationships between a musical piece and the dance motions that go along

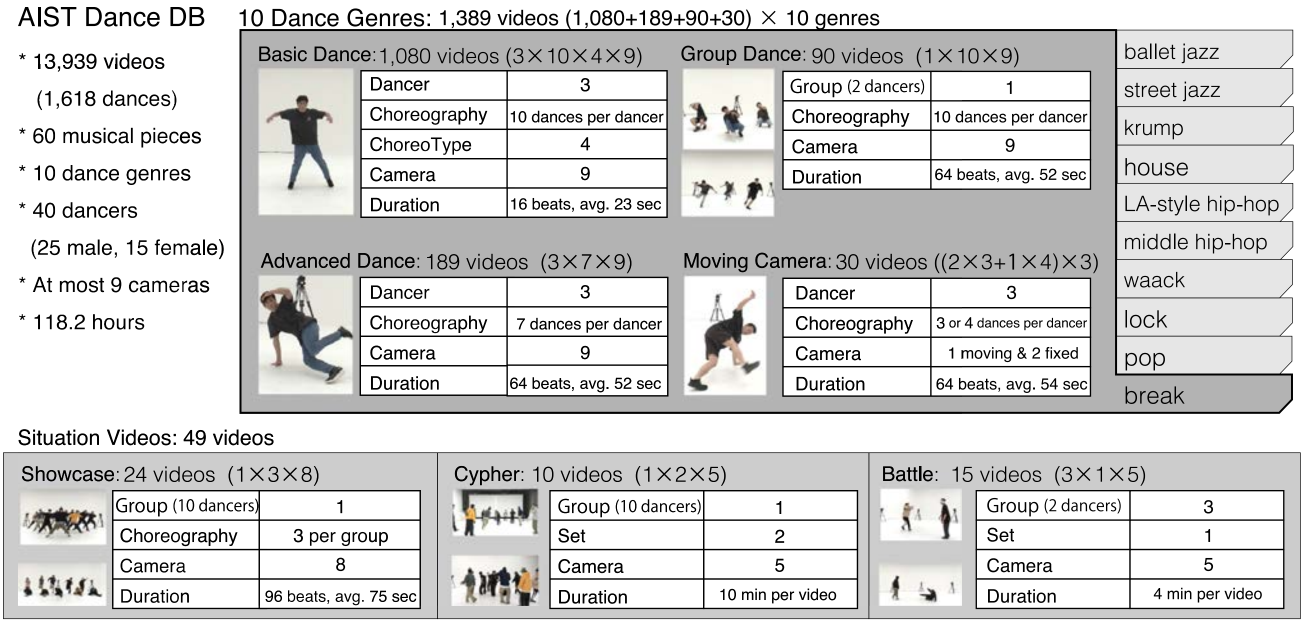

Figure 2. Overview of AIST Dance DB. Each of 10 dance genres has 1,389 videos that consist of 4 categories: **Basic Dance** (120 dances in 1,080 videos for basic genre-specific dance moves with 4 impressions (ChoreoType): intense, loose, hard, and soft), **Advanced Dance** (21 dances in 189 videos for advanced dance moves originally choreographed by each individual dancer), **Group Dance** (10 dances in 90 videos for group dance done with 3 dancers having different original choreographies), and **Moving Camera** (10 dances in 30 dolly-shot videos with moving camera). Other 49 Situation Videos consist of **Showcase** (3 dances in 24 videos assuming stage dance performances done by 10 dancers in front of audiences), **Cypher** (2 dances in 10 videos where 10 dancers line up in a circle and keep dancing in turns), and **Battle** (3 dances in 15 videos where 2 dancers face each other and dance).

with the piece, we first asked professional musicians to create pieces of dance music in different genres for our database, then recorded dance videos in which professional dancers danced while listening to one of the musical pieces. As a result, each video includes not only dance motions but also the musical piece used as background music.

- **Variety of dance genres and choreographies**
  Since processing diverse dance information would require a variety of genres and choreographies, we ensured such a variety by including 10 dance genres, 40 male and female dancers, different numbers of dancers (solo dancing and group dancing), different choreographies, and different levels of difficulty in choreography (basic choreography moves and original advanced moves).
- **Shooting from various directions**
  To analyze the same dance motions from different views and angles, we used multiple video cameras surrounding a dancer so that the cameras can simultaneously shoot from various directions. Even if some body parts cannot be seen from the front camera, they can be seen from the back camera.

3.2 Contents

An overview of the AIST Dance DB is shown in Figure 2. We built it on the basis of the design policy discussed in Section 3.1. The AIST Dance DB consists of 1,618 street dances in 13,939 videos: 13,890 videos of 10 street dance genres and an additional 49 *Situation Videos* of 3 different situations. The dance genres were decided in consultation with experts on street dances and divided into *Old School* styles (break, pop, lock, and waack), which are dance styles from about the 1970s to 1990s, and *New School* styles (middle hip-hop, LA-style hip-hop, house, krump, street jazz, and ballet jazz), which are dance styles since about the 1990s.

The database also includes 60 musical pieces that are categorized into 10 dance genres. The tempi of the 6 pieces for each genre except for house were set to 80, 90, 100, 110, 120, and 130 beats per minute (BPM); the tempi of the 6 pieces for house were set to 110, 115, 120, 125, 130, and 135 BPM since slow tempi are not fitting for house.

To cover choreographic variations within dance genres, we had 40 professional dancers (25 male and 15 female) participate in video recordings in a professional studio. At least three dancers were assigned to each genre. All dancers had more than 5 years of dance experience. All videos were recorded in full color, although dancers mainly wore monotone clothing when dancing.

For each of the 10 dance genres, we recorded a total of 1,380 videos consisting of 1,080 basic choreography dance videos, 189 advanced choreography dance videos, 90 group dance videos, and 30 dance videos with a moving camera for dolly-in and dolly-out shots. All camera positions were fixed except for the moving camera. As shown in Figure

2, we also recorded 49 *Situation Videos* that consist of 24 videos for showcase, 10 videos for cypher, and 15 videos for battle. Dancers were asked to choreograph their dance to fit the given genre. Basically, each choreography was shot in one take; however, it was taken again when there was a clear mistake. The location of the front camera was designed carefully to capture the full body of the dancer, and the other cameras were located to capture as much of the dancer's body as possible except for group dance. This database is available at `https://aistdancedb.ongaaccel.jp`.

4. DANCE-MOTION GENRE CLASSIFICATION

To illustrate the utility of the AIST Dance DB, we tackled the dance-motion genre-classification task. Since genre classification of music is a popular research topic in the MIR community, genre classification of dance motions could be a good starting point, which would also have applications such as personalized dance-video recommendation.

We developed four baseline methods to provide baseline results. We investigated four research questions the answers of which could contribute to research on dance-motion genre classification: (RQ1) *"Can we classify the 10 genres by using their video frames only?"*, (RQ2) *"How many video frames should be used to train a model?"*, (RQ3) *"Is the ease of classification different by dance genre?"*, and (RQ4) *"Can beat positions help improve classification accuracy?"*.

4.1 Experimental Conditions

As a dataset for our experiments, we created a subset of the advanced choreography dance videos. The dataset consists of 210 dance videos shot from the front camera only and covers the 10 dance genres. Each genre has 21 dance videos by 3 dancers, each of whom uses 7 original choreographies. In total, the dataset covers 210 different choreographies by 30 different dancers. The musical piece in each video has 64 beats (4 beats $\times$ 16 measures in four-four time), where the term "beat" denotes a quarter note.

We split 210 dance videos into a training set (126 videos), a validation set (14 videos), and a test set (70 videos). For each genre, 14 videos by two dancers were used for the training and validation sets, and 7 videos by the remaining dancer was used for the test set. Every dancer and every choreography in the training and validation sets thus does not appear in the test set.

4.2 Methods

An overview of our baseline methods is shown in Figure 3. Each method is trained to classify an excerpt of the input video into one of the 10 dance genres. In the first step of motion-feature extraction, we use the OpenPose library [12] to estimate the dancer's skeleton (body pose and motion) in all video frames (60 frames per second). This can reduce the dependency on the AIST Dance DB since the estimated body pose and motion do not have original RGB pixel information.

Since both pose and motion are important elements that characterize dancing, we obtain dancer poses by calculating 21 joint angles from the skeleton per video frame. Each joint angle is then converted into two dimensional values θ_x and θ_y by calculating the sine and cosine of the angles to make the distance calculation between angular values easier. As a result, we convert the 21-dimensional angular values into a 42-dimensional feature vector. Let us represent this feature vector at the n-th frame of the i-th video as $v_\theta^{(i)}(n)(1 \leq n \leq N^{(i)}\ and\ 1 \leq i \leq I)$, where $N^{(i)}$ is the number of frames in the i-th video and I is the number of videos in the dataset. When some joints in the video have not been detected, we substitute zeros for the values that correspond to the undetected joints. Our methods then represent the body motions by calculating the velocity and acceleration between frames. The velocity $v_{\Delta\theta}^{(i)}(n)$ and acceleration $v_{\Delta^2\theta}^{(i)}(n)$ at the n-th frame are calculated as follows:

$$v_{\Delta\theta}^{(i)}(n) = v_\theta^{(i)}(n) - v_\theta^{(i)}(n-1), \tag{1}$$

$$v_{\Delta^2\theta}^{(i)}(n) = v_{\Delta\theta}^{(i)}(n) - v_{\Delta\theta}^{(i)}(n-1). \tag{2}$$

We then concatenate the above three: $v_\theta^{(i)}(n)$, $v_{\Delta\theta}^{(i)}(n)$, and $v_{\Delta^2\theta}^{(i)}(n)$, into one 126-dimensional vector $v^{(i)}(n)$.

In the second step, we aggregate the 126-dimensional vectors representing body motions within a unit (temporal interval) determined if using beat positions or not. All beat positions are automatically determined by the tempo of each musical piece. Below are the details of the two methods:

- Adaptive method: when using beat positions, the vectors are aggregated within four kinds of tempo-dependent variable length units: we used one, two, three, or four beats as one unit. The length L_a corresponding to a beat ranges from 27 to 45 video frames as the tempo ranges from 80 to 135.
- L-fixed method: the vectors are aggregated within various fixed-length units. The length L of one unit is 20, 40, 60, ..., or 500 video frames.

Each method calculates a unit-level feature vector for every unit in the video. It calculates the mean vector and standard deviation vector (126 dimensions each) among body motions from the n_{start}-th to n_{end}-th video frames of every unit and concatenates those vectors into the 252-dimensional unit-level vector of the k-th unit $\tilde{v}^{(i)}(k)$.

In the third step, each method calculates a window-level feature vector. We use five different window lengths to aggregate unit-level feature vectors into a window-level one to see how many video frames are necessary to identify a dance genre. The window-level feature vector is obtained by concatenating all unit-level vectors $\tilde{v}^{(i)}(k)$ within a window of 2, 4, 8, 16, and 32 units, respectively. The window is then shifted by 1 unit to obtain the next window-level feature vector. We thus obtain five different window-level feature vectors.

In the fourth step, we prepare four baseline methods by combining adaptive or L-fixed methods with LSTM-based or SVM-based models. Each method classifies every

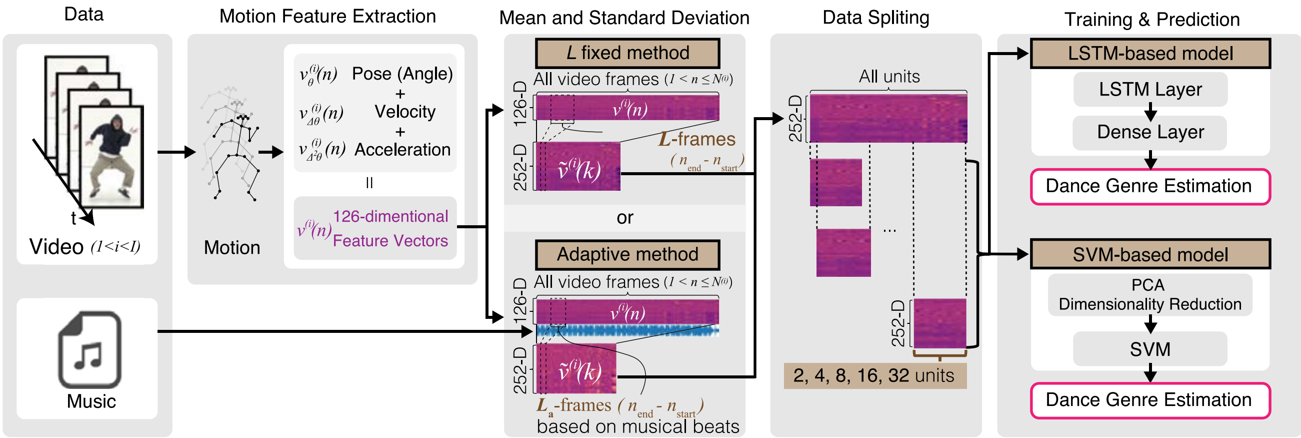

Figure 3. Overview of four baseline methods for dance-motion genre-classification task: (1) L-fixed method with LSTM-based model, (2) L-fixed method with SVM-based model, (3) adaptive method with LSTM-based model, and (4) adaptive method with SVM-based model.

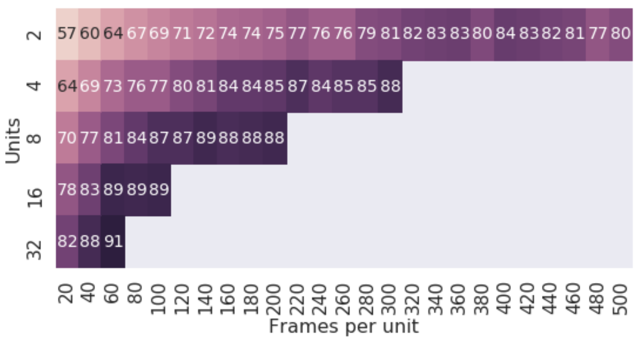

Figure 4. Comparison of the genre-classification accuracy of the L-fixed method with the LSTM-based model with regard to different combinations of the number of frames per unit and the number of units. Blank indicates that combined size of frames exceeds the length of video.

window-level feature vector into the 10 dance genres. For the LSTM-based model, we use a bi-directional recurrent neural network (RNN) with one layer of LSTM [39] cell. The network outputs a 10-dimensional one-hot vector representing dance genres. A rectified linear unit activation function is applied to the output of the LSTM. Batch normalization is applied to the output layer of the dense layers. We use cross entropy as the loss function and a batch size of 10. We train this model with a learning rate of $5e-4$ through 100 epochs and record the trained model at the minimum validation loss. This model is implemented in PyTorch [57]. For the SVM-based model, we first obtain 200-dimensional vectors by using principal component analysis to reduce the dimension of the training data, then train the SVM model. Finally, we estimate the dance genre of every window-level feature vector in a video by using these two models.

5. RESULTS

To answer RQ1 in Section 4, we investigated the accuracy of dance-motion genre classification using three-fold cross-validation (each fold used a different dancer for the test set). We first calculated the ratio of the correct estimation for every dance genre, then averaged over genres to obtain the genre-classification accuracy. The best genre-classification accuracy was 91.4% when we used the L-fixed method with the LSTM-based model where the number of frames was 60 and the number of units was 32. In this dataset, we found that dance genres can be estimated with relatively high accuracy. In the case of the L-fixed method with the SVM-based model, however, the best accuracy was dropped to 84.0%.

To answer RQ2 in Section 4, we analyzed the genre-classification accuracy when the number of frames per unit and the number of units were changed as shown in Figure 4. We found that dance-motion genre classification can be executed with an accuracy of 56.6% by using only 0.67 sec corresponding to 40 frames (20 frames per unit $\times$ 2 units) of a video. This was much shorter than we expected.

To answer RQ3 in Section 4, we conducted an analysis by creating a confusion matrix for the L-fixed method with the LSTM-based model and found that krump is relatively easy to estimate and house is relatively difficult to estimate. We also found that the estimation performance depends on the number of frames per unit and the number of units to calculate the input to the classifiers. With a small number of frames and units, street jazz and ballet jazz were easily confused by the classifier and the estimation accuracy of house dropped. This can be understood from the fact that street jazz and ballet jazz contain similar poses and house contains many movements that are commonly found in other dance genres, such as simple lateral movements.

To answer RQ4 in Section 4, we confirmed that the highest accuracy of the adaptive method using musical beats was 83.4% and that of the L-fixed method was 91.4%, both with the LSTM-based model. In the case of the adaptive method with the SVM-based model, the best accuracy was further dropped to 80.7%. In this way, the preliminary answer to RQ4 was not positive. Since there would be much room for improvement when using beat positions, we leave this for future research.

6. DISCUSSION

6.1 Dance information processing

As shown in Figure 5, dance information processing can be classified into four categories: (a) dance-motion analysis, (b) dance-motion generation, (c) dance-music analysis, and (d) dance-music generation. The goal of (a) dance-motion analysis is to automatically analyze every aspect of dance motions including dance-motion genre classification, dancer identification, dance-technique estimation, structural analysis of choreographies, and dance-mood analysis. A typical research topic of (b) dance-motion generation is to automatically generate motions for dance robots and computer-graphics dancers so that their motions can be natural and indistinguishable from human motions. As described at the beginning of this paper, research related to dance music, including (c) dance-music analysis and (d) dance-music generation, has been popular in the MIR community. There is still room for improvements regarding analyzing and classifying dance music in more depth and generating dance music in various styles and for various purposes.

Furthermore, our research community could investigate various interactions between those categories as well as advance each of the four categories. For example, we could combine (a) dance-motion analysis and (c) dance-music analysis. Analyzed dance motions would be helpful for structural analysis of musical pieces in dance music videos. Analyzed music structure could be used to analyze dance motions in a context-dependent manner. Another interesting topic of research is to find musical pieces suitable for dance motions, which will be useful for developing automatic DJ systems that recommend musical pieces suitable for various dancers on dance floors and at dance events. Analyzing the relationships between dance motions and music in existing dance videos is also important to develop systems for assisting people in creating and editing more attractive music-synchronized videos. We could also combine (a) dance-motion analysis and (b) dance-motion generation. If three-dimensional dance motions can be accurately extracted from a large collection of existing dance videos, they could be useful for automatic dance-motion-generation systems based on machine learning. If dance styles of dancers can be analyzed and modeled, it could become possible to transfer those styles to artificial computer-graphic dancers or robot dancers.

Dance information processing will naturally use multimodal dance information for different research topics. There have been many related studies such as on dance-practice support [3, 8, 11, 15, 16, 24, 36, 49, 68, 71], choreography-creation support [25, 59, 81], dance-performance augmentation [6,9,10,13,26,27,31,38,45,56, 76,78], dance-group support and analysis [35,50,54,69,77], dance archive [44, 58, 64, 83], dance performance alignment [21, 34], dance video editing [5, 41, 74], and dance-style transfer [4]. We look forward to advances in this emerging research field of dance information processing.

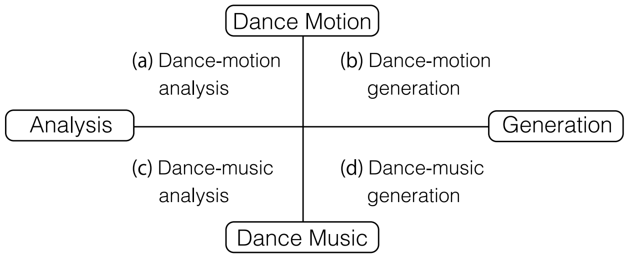

Figure 5. Overview of dance information processing.

6.2 AIST Dance Video Database

We believe that the AIST Dance DB will contribute to the advancement of dance information processing as other research databases have also contributed to the advancement of their related research. In particular, this database will advance research on dance-motion analysis (Figure 5). In addition to the dance-motion genre-classification task we discussed in this paper, it is possible to develop systems that can classify dance videos into advanced dance, basic dance, moving camera, and group dance, as shown in Figure 2. Since various dancers dance to the same musical pieces in this database, their individual differences can be analyzed in depth, and such analyzed results could be useful in developing dancer-identification systems. The basic dance videos could be useful for analyzing subtle individual differences of motions since all the dancers have the same basic motions. By using videos recorded from different directions, systems that can recognize dance motions from any direction could also be developed. As we illustrated in Section 4.2, using image-processing technologies, such as OpenPose, makes it possible to extract dance motions from dance videos for use in machine learning for various purposes including automatic dance-motion generation.

From the viewpoint of the MIR community, it is essential for the AIST Dance DB to include 60 pieces of dance music in synchronization with dance motions. This will lead to various research topics such as dance-music classification with or without using dance motions, dance-motion classification with or without using dance music, and detailed multimodal analysis of the correlation between dance motion and music. Furthermore, since this database is publicly available, it can be used for designing benchmarks of evaluating technologies.

7. CONCLUSION

The main contributions of this work are threefold: 1) we built the first large-scale shared database containing original street dance videos with copyright-cleared dance music, 2) we proposed and discussed a new research area *dance information processing*, and 3) we proposed a dance-motion genre-classification task and developed four baseline methods. We hope that the AIST Dance DB will help researchers develop various types of dance-information-processing technologies to give academic, cultural, and social impact.

8. ACKNOWLEDGMENTS

This work was supported in part by JST ACCEL Grant Number JPMJAC1602, Japan.

9. REFERENCES

[1] S. Abu-El-Haija, N. Kothari, J. Lee, P. Natsev, G. Toderici, B. Varadarajan, and S. Vijayanarasimhan. Youtube-8m: A large-scale video classification benchmark. *arXiv preprint arXiv:1609.08675*, 2016.

[2] O. Alemi, J. Françoise, and P. Pasquier. Groovenet: Real-time music-driven dance movement generation using artificial neural networks. In *Workshop on Machine Learning for Creativity, 23rd ACM SIGKDD Conference on Knowledge Discovery and Data Mining, SIGKDD 2017*, volume 8, page 26, 2017.

[3] D. S. Alexiadis, P. Kelly, P. Daras, N. E. O'Connor, T. Boubekeur, and M. B. Moussa. Evaluating a dancer's performance using kinect-based skeleton tracking. In *Proceedings of the 19th ACM International Conference on Multimedia, MM 2011*, pages 659–662, 2011.

[4] A. Aristidou, Q. Zeng, E. Stavrakis, K. Yin, D. Cohen-Or, Y. Chrysanthou, and B. Chen. Emotion control of unstructured dance movements. In *Proceedings of the ACM SIGGRAPH / Eurographics Symposium on Computer Animation, SCA 2017*, pages 9:1–9:10, 2017.

[5] E. Bakala, Y. Zhang, and P. Pasquier. MAVi: A movement aesthetic visualization tool for dance video making and prototyping. In *Proceedings of the 5th International Conference on Movement and Computing, MOCO 2018*, pages 11:1–11:5, 2018.

[6] Louise Barkhuus, Arvid Engström, and Goranka Zoric. Watching the footwork: Second screen interaction at a dance and music performance. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI 2014*, pages 1305–1314, 2014.

[7] P. Beauguitte, B. Duggan, and J. D. Kelleher. A corpus of annotated irish traditional dance music recordings: Design and benchmark evaluations. In *Proceedings of the 17th International Society for Music Information Retrieval Conference, ISMIR 2016*, pages 53–59, 2016.

[8] V. Boer, J. Jansen, A. T. Pauw, and F. Nack. Interactive dance choreography assistance. In *Proceedings of the 14th International Conference on Advances in Computer Entertainment Technology, ACE 2017*, pages 637–652, 2017.

[9] C. Brown. Machine tango: An interactive tango dance performance. In *Proceedings of the Thirteenth International Conference on Tangible, Embedded, and Embodied Interaction, TEI 2019*, pages 565–569, 2019.

[10] A. Camurri. Interactive dance/music systems. In *Proceedings of the 21th International Computer Music Conference, ICMC 1995*, pages 245–252, 1995.

[11] A. Camurri, R. K. El, O. Even-Zohar, Y. Ioannidis, A. Markatzi, J. Matos, E. Morley-Fletcher, P. Palacio, M. Romero, A. Sarti, S. D. Pietro, V. Viro, and S. Whatley. WhoLoDancE: Towards a methodology for selecting motion capture data across different dance learning practice. In *Proceedings of the 3rd International Symposium on Movement and Computing, MOCO 2016*, pages 43:1–43:2, 2016.

[12] Z. Cao, T. Simon, S. E. Wei, and Y. Sheikh. Real-time multi-person 2D pose estimation using part affinity fields. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017*, pages 7291–7299, 2017.

[13] F. Caputo, V. Mc Gowen, J. Geigel, S. Cerqueira, Q. Williams, M. Schweppe, Z. Fa, A. Pembrook, and H. Roffe. Farewell to dawn: A mixed reality dance performance in a virtual space. In *Proceedings of the ACM SIGGRAPH 2016 Posters, SIGGRAPH 2016*, pages 49:1–49:2, 2016.

[14] D. Castro, S. Hickson, P. Sangkloy, B. Mittal, S. Dai, J. Hays, and I. A. Essa. Let's dance: Learning from online dance videos. *arXiv preprint arXiv:1801.07388*, 2018.

[15] J. C. P. Chan, H. Leung, J. K. T. Tang, and T. Komura. A virtual reality dance training system using motion capture technology. *Journal of IEEE Transactions on Learning Technologies*, 4(2):187–195, 2011.

[16] E. Charbonneau, A. Miller, and J. J. LaViola. Teach me to dance: Exploring player experience and performance in full body dance games. In *Proceedings of the 8th International Conference on Advances in Computer Entertainment Technology, ACE 2011*, pages 43:1–43:8, 2011.

[17] W. Chu and S. Tsai. Rhythm of motion extraction and rhythm-based cross-media alignment for dance videos. *Journal of IEEE Transactions on Multimedia*, 14(1):129–141, 2012.

[18] W. T. Chu and S. Y. Tsai. Rhythm of motion extraction and rhythm-based cross-media alignment for dance videos. *Journal of IEEE Transactions on Multimedia*, 14(1):129–141, 2012.

[19] N. Collins. Influence in early electronic dance music: An audio content analysis investigation. In *Proceedings of the 13th International Society for Music Information Retrieval Conference, ISMIR 2012*, pages 1–6, 2012.

[20] S. Dixon, E. Pampalk, and G. Widmer. Classification of dance music by periodicity patterns. In *Proceedings of the 4th International Conference on Music Information Retrieval,ISMIR 2003*, pages 159–165, 2003.

[21] A. Dremeau and S. Essid. Probabilistic dance performance alignment by fusion of multimodal features. In *Proceedings of the IEEE International Conference on*

Acoustics, Speech and Signal Processing, ICASSP 2013, pages 3642–3646, 2013.

[22] B. Duggan, B. O'Shea, M. Gainza, and P. Cunningham. Machine annotation of sets of traditional irish dance tunes. In *Proceedings of the 9th International Society for Music Information Retrieval Conference, ISMIR 2016*, pages 401–406, 2008.

[23] S. Essid, X. Lin, M. Gowing, G. Kordelas, A. Aksay, P. Kelly, T. Fillon, Q. Zhang, A. Dielmann, V. Kitanovski, R. Tournemenne, A. Masurelle, E. Izquierdo N. E. O'Connor, P. Daras, and G. Richard. A multimodal dance corpus for research into interaction between humans in virtual environments. *Journal on Multimodal User Interfaces*, 7(1–2):157–170, 2013.

[24] A. Z. M. Faridee, S. R. Ramamurthy, H. M. S. Hossain, and N. Roy. Happyfeet: Recognizing and assessing dance on the floor. In *Proceedings of the 19th International Workshop on Mobile Computing Systems & Applications, HotMobile 2018*, pages 49–54, 2018.

[25] M. C. Felice, S. F. Alaoui, and W. E. Mackay. Knotation: Exploring and documenting choreographic processes. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems, CHI 2018*, pages 448:1–448:12, 2018.

[26] M. Fujimoto, N. Fujita, Y. Takegawa, T. Terada, and M. Tsukamoto. A motion recognition method for a wearable dancing musical instrument. In *Proceedings of the International Symposium on Wearable Computers, ISWC 2009*, pages 11–18, 2009.

[27] M. Fujimoto, N. Fujita, T. Terada, and M. Tsukamoto. Lighting choreographer: An LED control system for dance performances. In *Proceedings of the 13th ACM International Conference on Ubiquitous Computing, UbiComp 2011*, pages 613–614, 2011.

[28] S. Fukayama and M. Goto. Music content driven automated choreography with beat-wise motion connectivity constraints. In *Proceedings of the 12th Sound and Music Computing Conference, SMC 2015*, pages 177–183, 2015.

[29] A. Gkiokas and V. Katsouros. Convolutional neural networks for real-time beat tracking: A dancing robot application. In *Proceedings of the 18th International Society for Music Information Retrieval Conference, ISMIR 2017*, pages 286–293, 2017.

[30] R. I. Godøy and A. R. Jensenius. Body movement in music information retrieval. In *Proceedings of the 10th International Society for Music Information Retrieval Conference, ISMIR 2009*, pages 45–50, 2009.

[31] D. P. Golz and A. Shaw. Augmenting live performance dance through mobile technology. In *Proceedings of the 28th International BCS Human Computer Interaction Conference on HCI 2014 - Sand, Sea and Sky - Holiday HCI, BCS-HCI 2014*, pages 311–316, 2014.

[32] M. Goto and Y. Muraoka. A beat tracking system for acoustic signals of music. In *Proceedings of the ACM International Conference on Multimedia, MM 1994*, pages 365–372, 1994.

[33] F. Gouyon and S. Dixon. Dance music classification: A tempo-based approach. In *Proceedings of the 5th International Conference on Music Information Retrieval, ISMIR 2004*, 2004.

[34] M. Gowing, P. Kell, N. E. O'Connor, C. Concolato, S. Essid, J. Lefeuvre, R. Tournemenne, E. Izquierdo, V. Kitanovski, X. Lin, and Q. Zhang. Enhanced visualisation of dance performance from automatically synchronised multimodal recordings. In *Proceedings of the 19th ACM International Conference on Multimedia, MM 2011*, pages 667–670, 2011.

[35] C. F. Griggio and M. Romero. Canvas dance: An interactive dance visualization for large-group interaction. In *Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems, CHI EA 2015*, pages 379–382, 2015.

[36] T. Großhauser, B. Bläsing, C. Spieth, and T. Hermann. Wearable sensor-based real-time sonification of motion and foot pressure in dance teaching and training. *Journal of the Audio Engineering Society*, 60(7/8):580–589, 2012.

[37] C. Guedes. Controlling musical tempo from dance movement in real-time: A possible approach. In *Proceedings of the 29th International Computer Music Conference, ICMC 2003*, 2003.

[38] N. Hieda. Mobile brain-computer interface for dance and somatic practice. In *Proceedings of the 30th Annual ACM Symposium on User Interface Software and Technology, UIST 2017*, pages 25–26, 2017.

[39] S. Hochreiter and J. Schmidhuber. Long short-term memory. *Journal of Neural Computation*, 9(8):1735–1780, 1997.

[40] A. Holzapfel and E. Benetos. The sousta corpus: Beat-informed automatic transcription of traditional dance tunes. In *Proceedings of the 17th International Society for Music Information Retrieval Conference, ISMIR 2016*, pages 531–537, 2016.

[41] T. Jehan, M. Lew, and C. Vaucelle. Cati dance: Self-edited, self-synchronized music video. In *Proceedings of the ACM SIGGRAPH 2003 Sketches &Amp; Applications, SIGGRAPH 2003*, pages 1–1, 2003.

[42] T. Kell and G. Tzanetakis. Empirical analysis of track selection and ordering in electronic dance music using audio feature extraction. In *Proceedings of the 14th International Society for Music Information Retrieval Conference, ISMIR 2013*, pages 505–510, 2013.

[43] D. Kim, D. H. Kim, and K. C. Kwak. Classification of K-Pop dance movements based on skeleton information obtained by a kinect sensor. *Sensors*, 17(6):1261, 2017.

[44] E. S. Kim. Choreosave: A digital dance preservation system prototype. In *Proceedings of the American Society for Information Science and Technology*, pages 48:1–48:10, 2011.

[45] H. Kim and J. A. Landay. Aeroquake: Drone augmented dance. In *Proceedings of the ACM Designing Interactive Systems Conference, DIS 2018*, pages 691–701, 2018.

[46] P. Knees, Á. Faraldo, P. Herrera, R. Vogl, S. Böck, F. Hörschläger, and M. L. Goff. Two data sets for tempo estimation and key detection in electronic dance music annotated from user corrections. In *Proceedings of the 16th International Society for Music Information Retrieval Conference, ISMIR 2015*, pages 364–370, 2015.

[47] B. Li, A. Maezawa, and Z. Duan. Skeleton plays piano: Online generation of pianist body movements from MIDI performance. In *Proceedings of the 19th International Society for Music Information Retrieval Conference, ISMIR 2018*, pages 218–224, 2018.

[48] J. MacRitchie, B. Buck, and N. J. Bailey. Visualising musical structure through performance gesture. In *Proceedings of the 10th International Society for Music Information Retrieval Conference, ISMIR 2009*, pages 237–242, 2009.

[49] C. Mousas. Performance-driven dance motion control of a virtual partner character. In *Proceedings of the IEEE Conference on Virtual Reality and 3D User Interfaces, VR 2018*, pages 57–64, 2018.

[50] A. Nakamura, S. Tabata, T. Ueda, S. Kiyofuji, and Y. Kuno. Multimodal presentation method for a dance training system. In *Proceedings of the SIGCHI Conference Extended Abstracts on Human Factors in Computing Systems, CHI EA 2005*, pages 1685–1688, 2005.

[51] K. Nymoen, R. I. Godøy, A. R. Jensenius, and J. Tørresen. Analyzing correspondence between sound objects and body motion. *Journal of ACM Transactions on Applied Perception*, 10(2):9:1–9:22, 2013.

[52] F. Ofli, E. Erzin, Y. Yemez, and A. M. Tekalp. Multimodal analysis of dance performances for music-driven choreography synthesis. In *Proceedings of the IEEE Conference on Acoustics, Speech, and Signal Processing, ICASSP 2010*, pages 2466–2469, 2010.

[53] F. Ofli, E. Erzin, Y. Yemez, and A. M. Tekalp. Learn2Dance: Learning statistical music-to-dance mappings for choreography synthesis. *Journal of IEEE Transactions on Multimedia*, 14(3-2):747–759, 2012.

[54] K. Özcimder, B. Dey, R. J. Lazier, D. Trueman, and N. E. Leonard. Investigating group behavior in dance: an evolutionary dynamics approach. In *Proceedings of the American Control Conference, ACC 2016*, pages 6465–6470, 2016.

[55] M. Panteli, N. Bogaards, and A. K. Honingh. Modeling rhythm similarity for electronic dance music. In *Proceedings of the 15th International Society for Music Information Retrieval Conference, ISMIR 2014*, pages 537–542, 2014.

[56] J. A. Paradiso, K. Y. Hsiao, and E. Hu. Interactive music for instrumented dancing shoes. In *Proceedings of the 25th International Computer Music Conference, ICMC 1999*, pages 453–456, 1999.

[57] A. Paszke, S. Gross, S. Chintala, G. Chanan, E. Yang, Z. DeVito, Z. Lin, A. Desmaison, L. Antiga, and A. Lerer. Automatic differentiation in pytorch. In *Proceedings of the 31st Conference on Neural Information Processing Systems, NIPS 2017*, 2017.

[58] K. E. Raheb and Y. Ioannidis. Modeling abstractions for dance digital libraries. In *Proceedings of the 14th ACM/IEEE-CS Joint Conference on Digital Libraries, JCDL 2014*, pages 431–432, 2014.

[59] K. E. Raheb, G. Tsampounaris, A. Katifori, and Y. Ioannidis. Choreomorphy: A whole-body interaction experience for dance improvisation and visual experimentation. In *Proceedings of the 2018 International Conference on Advanced Visual Interfaces, AVI 2018*, pages 27:1–27:9, 2018.

[60] M. Raptis, D. Kirovski, and H. Hoppe. Real-time classification of dance gestures from skeleton animation. In *Proceedings of the 10th ACM SIGGRAPH/Eurographics Symposium on Computer Animation, SCA 2011*, pages 147–156, 2011.

[61] L. Risk, L. Mok, A. Hankinson, and J. Cumming. Melodic similarity in traditional french-canadian instrumental dance tunes. In *Proceedings of the 16th International Society for Music Information Retrieval Conference, ISMIR 2015*, pages 93–99, 2015.

[62] A. Rizzo, K. E. Raheb, S. Whatley, R. M. Cisneros, M. Zanoni, A. Camurri, V. Viro, J.-M. Matos, S. Piana, and M. Buccoli. Wholodance: Whole-body interaction learning for dance education. In *Proceedings of the Workshop on Cultural Informatics, co-located with the EUROMED International Conference on Digital Heritage, EUROMED 2018*, pages 41–50, 2018.

[63] M. Saini, S. P. Venkatagiri, W. T. Ooi, and M. C. Chan. The jiku mobile video dataset. In *Proceedings of the 4th ACM Multimedia Systems Conference, MMSys 2013*, pages 108–113, 2013.

[64] M. Sakashita, K. Suzuki, K. Kawahara, K. Takazawa, and Y. Ochiai. Materialization of motions: Tangible representation of dance movements for learning and archiving. In *Proceedings of the ACM SIGGRAPH 2017 Studio, SIGGRAPH 2017*, pages 7:1–7:2, 2017.

[65] A. Schindler and A. Rauber. An audio-visual approach to music genre classification through affective color

features. In *Proceedings of the European Conference on Information Retrieval, ECIR 2015*, pages 61–67, 2015.

[66] H. Schreiber and M. Müller. A crowdsourced experiment for tempo estimation of electronic dance music. In *Proceedings of the 19th International Society for Music Information Retrieval Conference, ISMIR 2018*, pages 409–415, 2018.

[67] R. A. Seger, M. M. Wanderley, and A. L. Koerich. Automatic detection of musicians' ancillary gestures based on video analysis. *Journal of Expert Systems with Applications*, 41(4):2098–2106, 2014.

[68] A. Soga, Y. Yazaki, B. Umino, and M. Hirayama. Body-part motion synthesis system for contemporary dance creation. In *Proceedings of the ACM SIGGRAPH 2016 Posters, SIGGRAPH 2016*, pages 29:1–29:2, 2016.

[69] A. Soga and I. Yoshida. Interactive control of dance groups using kinect. In *Proceedings of the 10th International Conference on Computer Graphics Theory and Applications, GRAPP 2015*, pages 362–365, 2015.

[70] E. Stavrakis, A. Aristidou, M. Savva, S. L. Himona, and Y. Chrysanthou. Digitization of cypriot folk dances. In *Proceedings of the 4th International Conference on Progress in Cultural Heritage Preservation, EuroMed 2012*, pages 404–413, 2012.

[71] J. K. T. Tang, J. C. P. Chan, and H. Leung. Interactive dancing game with real-time recognition of continuous dance moves from 3D human motion capture. In *Proceedings of the 5th International Conference on Ubiquitous Information Management and Communication, ICUIMC 2011*, pages 50:1–50:9, 2011.

[72] T. Tang, J. Jia, and M. Hanyang. Dance with melody: An LSTM-autoencoder approach to music-oriented dance synthesis. In *Proceedings of the 26th ACM International Conference on Multimedia, MM 2018*, pages 1598–1606, 2018.

[73] T. Tang, J. Jia, and H. Mao. Dance with melody: An LSTM-autoencoder approach to music-oriented dance synthesis. In *Proceedings of the 26th ACM International Conference on Multimedia, MM 2018*, pages 1598–1606, 2018.

[74] S. Tsuchida, S. Fukayama, and M. Goto. Automatic system for editing dance videos recorded using multiple cameras. In *Proceedings of the 14th International Conference on Advances in Computer Entertainment Technology, ACE 2017*, pages 671–688, 2017.

[75] S. Tsuchida, S. Fukayama, and M. Goto. Query-by-dancing: A dance music retrieval system based on body-motion similarity. In *Proceedings of the 24th International Conference on MultiMedia Modeling, MMM 2019*, pages 251–263, 2019.

[76] S. Tsuchida, T. Takemori, T. Terada, and M. Tsukamoto. Mimebot: spherical robot visually imitating a rolling sphere. *International Journal of Pervasive Computing and Communications*, 13(1):92–111, 2017.

[77] S. Tsuchida, T. Terada, and M. Tsukamoto. A system for practicing formations in dance performance supported by self-propelled screen. In *Proceedings of the 4th Augmented Human International Conference, AH 2013*, pages 178–185, 2013.

[78] S. Tsuchida, T. Terada, and M. Tsukamoto. A dance performance environment in which performers dance with multiple robotic balls. In *Proceedings of the 7th Augmented Human International Conference, AH 2016*, pages 12:1–12:8, 2016.

[79] P. Wang and J. M. Rehg. A modular approach to the analysis and evaluation of particle filters for figure tracking. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2006*, pages 790–797, 2006.

[80] K. Yadati, M. Larson, C. C. S. Liem, and A. Hanjalic. Detecting drops in electronic dance music: Content based approaches to a socially significant music event. In *Proceedings of the 15th International Society for Music Information Retrieval Conference, ISMIR 2014*, pages 143–148, 2014.

[81] T. Yamaguchi and H. Kadone. Bodily expression support for creative dance education by grasping-type musical interface with embedded motion and grasp sensors. *Sensors*, 17(5):1171, 2017.

[82] W. Zhang, Z. Liu, L. Zhou, H. Leung, and A. B. Chan. Martial arts, dancing and sports dataset. *Journal of Image Vision Computing*, 61(C):22–39, 2017.

[83] X. Zhang, T. Dekel, T. Xue, A. Owens, Q. He, J. Wu, S. Mueller, and W. T. Freeman. MoSculp: Interactive visualization of shape and time. In *Proceedings of the 30th Annual ACM Symposium on User Interface Software and Technology, UIST 2018*, pages 275–285, 2018.