

Article

# Digitization and Visualization of Folk Dances in Cultural Heritage: A Review

Iris Kico <sup>1,\*</sup>, Nikos Grammalidis <sup>2</sup>, Yiannis Christidis <sup>3</sup> and Fotis Liarokapis <sup>1</sup>

<sup>1</sup> Faculty of Informatics, Masaryk University, Botanicka 68A, 60200 Brno, Czech Republic; liarokap@fi.muni.cz

<sup>2</sup> Information Technologies Institute, Centre for Research and Technology Hellas, 6th km Charilaou-Thermi Res, 57001 Thessaloniki, Greece; ngramm@iti.gr

<sup>3</sup> Department of Communication and Internet Studies, Cyprus University of Technology, Arch. Kyprianou 30, 3036 Limassol, Cyprus; yiannis.christidis@cut.ac.cy

\* Correspondence: iriskico@mail.muni.cz; Tel.: +420-773-652-461

Received: 31 July 2018; Accepted: 12 October 2018; Published: 23 October 2018



**Abstract:** According to UNESCO, cultural heritage does not only include monuments and collections of objects, but also contains traditions or living expressions inherited from our ancestors and passed to our descendants. Folk dances represent part of cultural heritage and their preservation for the next generations appears of major importance. Digitization and visualization of folk dances form an increasingly active research area in computer science. In parallel to the rapidly advancing technologies, new ways for learning folk dances are explored, making the digitization and visualization of assorted folk dances for learning purposes using different equipment possible. Along with challenges and limitations, solutions that can assist the learning process and provide the user with meaningful feedback are proposed. In this paper, an overview of the techniques used for the recording of dance moves is presented. The different ways of visualization and giving the feedback to the user are reviewed as well as ways of performance evaluation. This paper reviews advances in digitization and visualization of folk dances from 2000 to 2018.

**Keywords:** intangible cultural heritage; folk dances; digitization; visualization

## 1. Introduction

Intangible cultural heritage (ICH) refers to oral tradition, presentations, expressions, knowledge, and skills to produce traditional crafts and festive events. This kind of heritage passes from generation to generation and acquires an important role in maintaining quality cultural diversity in growing globalization [1]. ICH in the form of dance, either as an autonomous form of art and expression, or as a part of the music and/or sound culture, has been an object of general interest through the ages. From the wall paintings after the prehistoric age to the contemporary era, humans have been representing themselves dancing and bonding through this animating procedure. Along with hunting or eating and drinking together, it cannot be denied that dancing has been a vital part of humans' life through the ages. Such flows in time can be better pictured by focusing on folk dances, as they are already considered an important part of ICH, directly connected to local culture and ethnic or other type of group identity [2]. These reasons suggest that the preservation of folk dances is more than significant.

The cultural spirit must be passed to the next generation and such a process can be assisted by assorted practices related to folk dances, which are usually taught in person, imitating the teacher's dance moves. Dances can be taught, also, in different ways, such as text documentation, video, and the graphical notation [3]; nevertheless, all these approaches have some limitations. The use of text documentation information about dance and its cultural significance can be presented, but in

such a case, there can be a lack of movements and different dance styles. On the other hand, videos can easily present movements, finding, though, difficulties in successfully presenting additional information about each dance [4]. Preservation includes all actions that help not losing the total of those dance moves that ancestors brought through the time and they are not saved in any savable form. These actions include recordings, digitization, and reconstruction of folk dances. The current state of information technologies has enabled new different ways for the preservation of folk dances that can help overcome the aforementioned issues. It can also help with wider recognition and dissemination of folk dances, the importance of which should be once more stressed; when performed, folk dances incorporate community bonding by default. 'Active' cultural elements that are present in many levels during dancing seem to be essential for the integrity of the peoples' identity, in the vigorous everyday rhythms of societies' development. It is therefore crucial to facilitate the aforementioned recognition and dissemination of the dances, adapting them to the changing dynamics in the evolution of the people. In such a process, the digitization of dancing would be a very significant step. Improving the digitization technology regarding the capturing and modeling of performing arts, especially folk dances, appear to be critical in [5]:

- Promoting cultural diversity,
- making local communities and Indigenous people aware of the richness of their intangible heritage; and
- strengthening cooperation and intercultural dialogue between people, different cultures, and countries.

Many scholars have recognized the significance of the above issues, and, consequently, there are several European projects committed to the preservation of ICH and folk dances.

"Wholodance" ([www.wholodance.eu](http://www.wholodance.eu)) is a European Union (EU) project focusing on developing and applying breakthrough technologies to dance learning, to achieve results with impact on researchers and professionals, but also dance students and the interested public. Among its objectives is the preservation of cultural heritage by creating a proof-of-concept motion capture repository of dance motions and documenting diverse and specialized dance movement practices and learning approaches. A popular learning approach is creating web-based platforms for that purpose. The WebDANCE project ([www.miralab.ch/projects/webdance](http://www.miralab.ch/projects/webdance)) was a pilot project that experimented with the development of a web-based learning environment of traditional dances. The final tool included teaching units and three-dimensional (3D) animation for two dances and demonstrated the potential for teaching folk dances to young people. One more project that committed to passing folk dances for a wide range of users is the Terpsichore project [6]. The focus is to study, analyze, design, research, train, implement, and validate an innovative framework for affordable digitization, modeling, archiving, e-preservation, and presentation of ICH content related to folk dances, in a wide range of users [5].

The need for multimodal ICH datasets and digital platforms for various multimedia digital content has been recognized in different projects [7]. This includes folk dances as well. The European Union (EU) project, i-Treasures ([www.i-treasures.eu](http://www.i-treasures.eu)) [8], was committed to the preservation of ICH. The main objective of the project was to develop an open and extendable platform to provide access to ICH resources. Several folk dances have been recorded and educational game-like applications have been implemented for them [9]. Another project committed to the preservation of the performing arts is the AniAge project (<http://www.euh2020aniage.org/>). This project is committed to the preservation of the performing art related ICHs of Southeast Asia (e.g., local dances that are visually and culturally rich, but are disappearing due to the globalized modernization). Novel techniques and tools to reduce the production costs and improve the level of automation are being developed, without sacrificing the control from the artists. Two areas of technological innovation are targeted, novel algorithms for 3D computer animation, and visual asset management with data analytics and machine learning techniques.

Studies how ICH can become an integral part of future museum practice and policies, supporting practitioners of intangible heritage in safeguarding their cultural heritage, are presented through the IMP (Intangible Cultural Heritage and Museums, <https://www.ichandmuseums.eu/>) project, supported by the European Commission (EC) from the Creative Europe program. Over the course of the project, in co-creation with the participants in its events, practical guidelines, recommendations, and brainstorm exercises will be developed as part of a toolbox.

As it can be seen the area of preservation, ICH and folk dances are very popular and very wide. The need to have a review of the systems used for digitization of dance moves and visualization of folk dances is apparent. To the best of our knowledge, there are not papers that cover both topics at the same time. More about motion capture systems can be found in [10] and about the visualization [11].

This paper presents a review of the digitization and visualization of folk dances and feedbacks for users, as well. The paper is structured as follows: In Section 2, systems and methods used for digitization of dance moves and archiving are explained. In Section 3, the different ways of visualizing approaches are presented and discussed, along with different ways of giving feedback to the user. In Section 4, the evaluation of performance is discussed, and in Section 5, certain conclusions regarding the on-going research are given.

## 2. Dance Digitization and Archival

Digital archives that involve activities to preserve, for future generations, historical and cultural properties through digitization have been undertaken in various places. Such archives do not include just tangible cultural properties, but also intangible cultural assets, as the dance itself [12]. Considering that heritage elements may not always be visible, and given that many of them are likely to disappear, assorted digitization methods offer the possibility to precisely record ICH to preserve it in a visual and digital format [13]. It is important, before advancing to the implementation of any dance moves' digitization, to take the overall dancing environment under consideration, as experienced by the dancer. Music, sounds, smells, the cultural context, dancers' relations, weather conditions, or even the time of the day are factors that are unavoidably part of the experience of folk dancing. Nonetheless, digital technology focuses only on the analysis and visualization of the basic dance moves executed by the bodies, while other factors, such as the aforementioned, should also be documented as additional information.

Folk dance is considered a ritual among people, characteristic of the common residents of a country or region, that is transmitted from generation to generation [14]. Whichever the cause, people gather and perform such rituals for many years, developing bonds among themselves, and connecting with the space they spend, or they used to spend, their everyday life. For such assets to be preserved, dances need to be taught, and recording them can be a flexible tool for this purpose. As the teaching of folk dances using available new technologies is based on recording dance moves and presenting them to the users, the focus should be given to this procedure; the recording of human motion can be a complex process involving different (often multiple) sensors and algorithms comprising motion capture systems. Generally, recording involves not only digitization, but also all the aspects of this digital content management, representation, and reproduction. Digitization represents the first step of the entire recording process and it consists mainly of three phases [15]:

1. Preparation—decision about technique and methodology to be adopted, as well as the place of digitization;
2. digital recording—main digitization process; and
3. data processing and archival—post-processing, modeling, and archival of the digitized dances.

### 2.1. Dance Digitization Systems

Motion capture is applied for digitization to recreate dances in three dimensions and represent them in a three-dimensional (3D) environment [4]. Motion capture is a process of recording moving

objects or people, which would be the dancers in this case. This can be a long and difficult process and it is required that both adequate equipment and software are chosen. There are a lot of systems that can be used for motion capture. These systems can be divided into two main categories: The optical and the non-optical ones.

### 2.1.1. Optical Marker-Based Systems

The use of these systems requires the dancer(s) to wear a specially designed suit, covered with reflectors that are placed in their main articulations. Special cameras are strategically positioned to perform the tracking of the reflectors during the dancer's movement. Each camera generates two-dimensional (2D) coordinates for each reflector. Using the set of the 2D data captured by all cameras, 3D coordinates of the reflectors are generated. An important advantage of these systems is the very high sample rate that enables the capturing of fast movements. Of course, as dances include sophisticated moves by default, systems with a high degree of precision should be chosen. Another advantage of the optical systems is the freedom of the dancer's movement, as there are no cables that can limit the movements. The disadvantage, though, is the possible occlusion of some markers in some of the cameras. This problem can compromise the entire recording process if occluded data is unrecoverable. Another disadvantage is that users typically wear suits, to which markers are attached. Furthermore, a characteristic of these systems is the lack of interactivity, since the obtained data must be processed before they become usable [10,16].

#### Active Markers

Motion capture systems with active markers use LED's that emit their own light. Stavrakis et al. [17] used the Phasespace Impulse X2 motion capture system with active LEDs. This system uses eight cameras that can capture 3D motion using modulated LEDs. The dancer wears a special suite with 38 markers and active LEDs, as shown in Figure 1. As it has been mentioned, it applies here that active markers require additional wiring and may limit the freedom of the dancer's moves.



Figure 1. Optical motion capture system with active markers [18].

#### Passive Markers

Optical systems with passive markers are in use for motion capture [4,7]. These systems consist of cameras that can capture markers placed on various positions on the dancer's body, as shown in Figure 2. Markers are coated with a reflective material to reflect light that is produced near the cameras' lens. Before usage, cameras need to be calibrated. The number of cameras and markers can vary. Systems with passive markers can have problems with marker identification.



**Figure 2.** Optical motion capture system with passive markers [19].

A passive optical motion capture system is also used by Mustafa et al. [20] and Hegarini et al. [21]. The camera's threshold can be adjusted so only the bright reflective markers will be sampled, ignoring the skin and fabric. Even though the accuracy of optical systems is limited by the number of markers available, they still provide the highest accuracy and shortest response time.

### 2.1.2. Marker-Less Motion Capture Systems

Marker-less capture methods based on computer vision technology [22–24] can overcome the limitations of passive optical motion capture systems and can provide movement freedom for dancers. However, these systems are susceptible to error approximation, do not fully exploit global spatiotemporal consistency constraints, and are generally less precise than systems with markers. These systems do not require any additional equipment for tracking the dancer's movement. The movements are recorded in one or multiple video streams and computer vision algorithms analyze these streams. The motion capture process is completely software-based [10]. Next, the paper will first describe marker-less motion cameras based on popular depth sensors, then, will examine modern techniques for 2D/3D pose estimation based on a single RGB (red, green, blue) camera, and, finally, examine some multiview camera setups.

#### Depth Sensors

According to the technologies used, the most popular depth (range) sensors can be categorized as follows: Structured light, time-of-flight (ToF), and embedded stereo. The structured light approach is an active stereovision technique, where a sequence of known (usually infra-red (IR)) patterns is sequentially projected onto an object and is deformed by the geometric shape of the object. The object is then observed from a standard RGB camera (RGB—red, green, and blue light are added together in various ways to reproduce a broad array of colors), and depth information can be extracted by analyzing the distortion of the observed pattern, i.e., the disparity from the original projected pattern. The ToF approach is based on measuring the time that the light emitted by an illumination unit requires to travel to an object and back to the sensor array. In the continuous wave (CW) intensity modulation approach, which is commonly used, the scene is actively illuminated using near infrared (NIR) intensity-modulated, periodic light and shifting of the phase of the returning light is detected. In the embedded stereo approach, the depth of each pixel is determined from data acquired using a stereo or multiple-camera setup system, based on triangulation. Using state-of-the-art sensors (e.g.,

Zed camera, high-resolution and high frame-rate 3D video capture) depth perception for indoors and outdoors applications at up to 20 m can be achieved.

The concept of depth cameras is not new, but Microsoft Kinect has made such sensors accessible to all. The first Kinect camera used a structured light technique to generate real-time depth maps containing discrete range measurements of the physical scene, while the second version achieved improved performance based on a time-of-flight approach [25]. Microsoft discontinued all Kinect products starting from October 2017, however, they recently announced a new “Project Kinect for Azure” product, planned to be released in 2019. In [26], an algorithm is presented that fuses depth data streamed from a moving Kinect sensor into a single global implicit surface model of the observed scene in real-time. An extension of this technique is the DynamicFusion approach [27], which reconstructs scene geometry whilst simultaneously estimating a dense volumetric 6D motion field that warps the estimated geometry into a live frame. In [28], an algorithm using no temporal information is presented, which is used by the Kinect sensor to quickly and accurately predict 3D positions of body joints from a single depth image. Many scholars used the Kinect sensor as a low-cost sensor for motion capture [2,29,30], as it provides real-time 3D skeleton tracking in dark and bright indoor areas (since it uses infra-red). However, it is almost useless in sunlight, because the IR structured lighting pattern gets completely lost in ambient IR. A limitation of the sensor is that it can only record the front side of the body, and the movement area is limited. Also, the Kinect depth data are inherently noisy. Depth measurements often fluctuate, and depth maps contain numerous holes, where no readings have been obtained [26].

## 2D and 3D Pose Estimation Based on a Single RGB Camera

Pose estimation and action recognition are two crucial tasks for understanding human motion. Pose estimation refers to the process of estimating the configuration of the underlying kinematic or skeletal articulation structure of a person [31]. Estimating human pose from video input is an increasingly active research area in computer vision that could give rise to numerous real-world applications, including dance analysis. Traditional methods for pose estimation model structures of body parts are mainly based on handcrafted features. However, such methods may not perform well in many cases, especially when dealing with occlusions on body parts.

Recently, great technological advances were made in 2D human pose estimation from simple RGB images, mainly due to the efficiency of deep learning techniques, and particularly the convolutional neural networks (CNN), a class of deep neural networks mostly applied to analyzing visual imagery. A new benchmark dataset is introduced by Andriluka et al. [32], followed by a detailed analysis of leading human pose estimation approaches, providing insights for the success and failures of each method. Some very effective open source packages have become increasingly popular, such as OpenPose [33], a real-time method to estimate multiple human poses that was efficiently developed at Robotics Institute of Carnegie Mellon University. OpenPose represents a real-time system to jointly detect a human body, hand, and facial keypoints (130 keypoints in total) on single images, based on convolutional neural networks (CNN). More specifically, OpenPose extends the “convolutional pose” approach proposed in [34] and estimates 2D joint locations in three steps: (a) By detecting confidence maps for each human body part, (b) by detecting part affinity fields that encode part-to-part associations, and (c) by using a greedy parsing algorithm to produce the final body poses. In addition, the system’s computational performance on body key point estimation is invariant to the number of people detected in the image [33,35].

In [36], a weakly-supervised transfer learning method is proposed for 3D human pose estimation in the wild. It uses mixed 2D and 3D labels in a unified deep neural network that has a two-stage cascaded structure. The module combines (a) a 2D pose estimation module, namely the hourglass network architecture [37], producing low-resolution heat-maps for each joint, and (b) a depth regression module, estimating a depth value for each joint. An obvious advantage from combining these modules in a unified architecture is that training is end-to-end and fully exploits the correlation between the

2D pose and depth estimation sub-tasks. Furthermore, in [38], a real-time method is presented to capture the full global 3D skeletal pose of a human using a single RGB camera. The method combines a CNN-based pose regressor with a real-time kinematic skeleton fitting method, using the CNN output to yield temporally stable 3D global pose reconstructions based on a coherent kinematic skeleton. The authors claim that their approach has comparable (and, in some cases, better) performance with Kinect and is more broadly applicable than RGB-depth (RGB-D) solutions (e.g., in outdoor scenes or when using low-quality cameras). RGB-D (red, green, blue plus depth) cameras provide per-pixel depth information aligned with image pixels from a standard camera. In [39], a fully feedforward CNN-based approach is proposed for monocular 3D human pose estimation from a single image taken in an uncontrolled environment. The authors use transfer learning to leverage the highly relevant mid- and high-level features learned on the readily available in-the-wild 2D pose datasets in conjunction with the existing annotated 3D pose datasets. Furthermore, a new dataset of real humans with ground truth 3D annotations from a state-of-the-art marker-less motion capture system is produced.

A promising recent advancement is the recovery of parameterized 3D human body surface models, instead of simple skeleton models. This paves the way for a broad range of new applications, such as foreground and part segmentation, avatar animation, virtual reality (VR) applications, and many more. In [40], dense human pose estimation is performed by mapping all human pixels of an RGB image to a surface-based representation of the human body. The work is inspired by the DenseReg framework [41], where CNNs were trained to establish dense correspondences between a 3D model and images “in the wild” (mainly for human faces). The approach is combined with the state-of-the-art Mask-RCNN (Region-CNN) system [42], resulting in a trained model that can efficiently recover highly accurate correspondence fields for complex scenes involving tens of persons with moderate computational complexity. In [43], a “Human Mesh Recovery” framework is presented for reconstructing a full 3D mesh of a human body from a single RGB image. Specifically, a generative human body model, SMPL (Skinned Multi-Person Linear model) [44], is used, which parameterizes the mesh by 3D joint angles and a low-dimensional linear shape space. The method is trained using large-scale 2D key point annotations of in-the-wild images. Convolutional features of each image are sent to an iterative 3D regression module, whose objective is to infer the 3D human body and the camera in a way that its 3D joints project onto the annotated 2D joints. To deal with ambiguities, the estimated parameters are sent to a discriminator network, whose task is to determine if the 3D parameters correspond to bodies of real humans or not. The method runs in real-time performance, given a bounding box containing the person. Additional information and reviews of the progress in the field can be found in the recent literature [45–47].

### Multiview RGB-D Systems

A number of scholars have used a multiple Kinect sensor approach for motion capture. A multiple RGB-depth (RGB-D) capturing system, along with a novel sensor’s calibration method, is presented in [48]. A robust, fast reconstruction method from multiple RGB-D streams is also proposed, based on an enhanced variation of the volumetric Fourier transform-based method, and accompanied by an appropriate texture mapping algorithm. Furthermore, generic, multiple depth stream-based methods for accurate real-time human skeleton tracking is proposed, extending previous work [49,50]. In [9,51], a motion capture approach using three Kinect sensors is used for dance motion capture for a game application for dance learning and performance evaluation. Dance digitization is done in two ways for different types of performances in [52]. Both ways are marker-less motion capture without disturbing the dancer’s moves using additional equipment. Solo and trio performances are captured using three camcorders, all facing the stage, but placed into different positions. Duo performance is captured using two Kinects and five 2D/HD camcorders. The first Kinect is used for one dancer and the second one for the other dancer while two 2D camcorders were used for HD close-ups of both. The other camcorders were used to capture sequences of the whole stage. A model-based method to accurately reconstruct human performances captured outdoors in a multi-camera setup is presented in [53]. The proposed

approach deforms a template of the actor model in a way that it accurately reproduces the performance filmed with a calibrated and synchronized multi-view video. The fit is achieved in two stages: First, the coarse skeletal pose is estimated, and, subsequently, the non-rigid surface shape and body pose are jointly refined.

### 2.1.3. Non-Optical Marker-Based Systems

Non-optical systems' marker-based systems for motion capture can be categorized as follows, with respect to the technology used [10]:

- Acoustic systems;
- mechanical systems;
- magnetic systems; and
- inertial systems.

In acoustic systems, a set of sound transmitters are placed on the dancer's main articulations, while three receptors are positioned in the capture site. The emitters are sequentially activated, producing a characteristic set of (typically ultrasonic) frequencies that receptors pick up and use to calculate the emitter's position in 3D space. The number of transmitters that can be used is limited [10]. An advantage of these systems is their stability, even if obstructions between the dancer and the receptor or metallic object interference issues emerge. On the other hand, problems are a restriction of movements, due to cables and possible external sound sources, which might affect the capture process [16]. One more downside is the difficulty in obtaining a correct description of the data in a certain instant.

Mechanical systems are made of potentiometers and sliders that are put in the desired articulations and enable the display of their positions. Motion capture is done using an exoskeleton. Every joint is connected to an angular encoder. The value of the movement of each encoder is recorded by computers. Knowing the relative position of every joint, it is possible to reconstruct movements. These systems have some advantages that make them attractive: They are not affected by magnetic fields or unwanted reflections and do not need a recalibration process, which makes their use easy [10]. They also offer high precision, but the accuracy depends on the position of the encoders. The downside of mechanical systems is that they are generally significantly obstructive. The exoskeleton uses wired connections to connect encoders and the computer. This makes freedom of movements limited. It is quite complicated to measure the interaction between several exoskeletons, making the recording of more people at the same time difficult to implement. Figure 3 illustrates an actor wearing a mechanical motion capture suit.



Figure 3. Mechanical motion capture suit [54].

Magnetic systems use a set of receptors placed on the dancer's articulations, which measure the 3D position and orientation in relation to the emitter antenna. Magnetic systems are used for real-time application due to its quick setup capabilities. For instance, it is likely that no calibration is needed [4]. These systems are cheap compared to other motion capture systems. Disadvantages of these systems include a large number of cables that reduces freedom of the dancer's movements and high-power consumption [55]. An alternative system that eliminates this drawback is proposed in [56]. Interference in the magnetic field caused by various metallic objects is possible and it represents one more disadvantage of these systems.

Inertial systems use inertial sensors distributed on the dancer's body. An advantage of these systems is portability; no spatial setting is needed and cost are lower when compared to optical systems. An inertial motion capture system is used in [11]. Each sensor in this system measures rotational rates. The system live streams the dancer's motion to an avatar. Inertial systems have a limitation in the interpretation of feet in relation to a reference surface in movements, such as jumping and sitting. Also, these systems during the time can produce large error between the real motion and the captured data, and due to the inaccuracies of used sensors, error is accumulated [55]. In inertial systems, positional drift can compound over time. Figure 4 shows the inertial motion capture suit.



Figure 4. Inertial motion capture suit [54].

#### 2.1.4. Comparison of Motion Capture Technologies

In Table 1, an overview of the previously described systems is given.

**Table 1.** Motion capture systems.

System	Advantages	Disadvantages	Data Captured/Data Analysis/Real Time (or Not)	References
Optical marker-based systems	<ul style="list-style-type: none"> <li>- high sample rate</li> <li>- no limitation for the number of reflectors</li> <li>- light weight</li> </ul>	<ul style="list-style-type: none"> <li>- possible marker occlusion</li> <li>- possible difficulty in marker identification (for passive markers)</li> <li>- wiring (for active markers) restricts movements</li> <li>- lack of interactivity as post-processing is needed</li> <li>- expensive (\$100,000–\$250,000)</li> <li>- at least two pairs of tracker-sensor relationships for valid triangulation</li> </ul>	<ul style="list-style-type: none"> <li>- 3D marker positions and orientations using triangulation</li> <li>- Often non-real-time as denoising and post-processing may be required (esp. for passive markers)</li> </ul>	[4,7,17,20,21]
Marker-less systems	<ul style="list-style-type: none"> <li>- no additional equipment</li> <li>- freedom of movements</li> <li>- low cost (Microsoft Kinect from \$100)</li> <li>- work in bright and dark areas</li> <li>- real-time tracking</li> </ul>	<ul style="list-style-type: none"> <li>- limited movement area</li> <li>- less precise than optical systems</li> </ul>	<ul style="list-style-type: none"> <li>- RGB image (for optical systems) or RGB image + depth (for depth sensors)</li> <li>- joint positions and orientations (after body part identification and human pose recognition) in hardware (real-time) or software (near real time using GPUs)</li> </ul>	[2,7,9,29,30,51]
Acoustic systems	<ul style="list-style-type: none"> <li>- no obstruction issues</li> <li>- no metallic interference issues</li> </ul>	<ul style="list-style-type: none"> <li>- obtrusive</li> <li>- external noise and sound reflection issues</li> <li>- reduced accuracy</li> </ul>	<ul style="list-style-type: none"> <li>- 3D joint positions through either time-of-flight of the acoustic waves and triangulation or phase coherence</li> <li>- real-time</li> </ul>	[10,16]
Mechanical systems	<ul style="list-style-type: none"> <li>- not affected by magnetic fields</li> <li>- no recalibration needed</li> <li>- no unwanted reflections</li> <li>- no occlusion</li> <li>- low cost (\$5000–\$10,000)</li> </ul>	<ul style="list-style-type: none"> <li>- obtrusive</li> <li>- restricted movements</li> <li>- fixed configuration of sensors</li> <li>- no global translation</li> </ul>	<ul style="list-style-type: none"> <li>- the relative position of each joint based on angular encoders attached to an exoskeleton</li> <li>- real-time</li> </ul>	[10,16]
Magnetic systems	<ul style="list-style-type: none"> <li>- no calibration needed</li> <li>- real-time</li> <li>- cheaper than optical systems (\$5000–\$150,000)</li> </ul>	<ul style="list-style-type: none"> <li>- obtrusive</li> <li>- interference with magnetic fields</li> <li>- high power consumption</li> </ul>	<ul style="list-style-type: none"> <li>- 3D position and orientation of each joint in relation to emitter antenna</li> <li>- real-time</li> </ul>	[10,55,56]
Inertial systems	<ul style="list-style-type: none"> <li>- portability</li> <li>- cheaper than optical systems (price range \$1000–\$80,000)</li> <li>- smaller latency than optical systems</li> <li>- much higher sampling rate compared to optical systems</li> <li>- no occlusion</li> </ul>	<ul style="list-style-type: none"> <li>- measurements drift over time periods</li> <li>- battery packs and wires on the performers' body</li> <li>- smaller capture area compared to optical</li> </ul>	<ul style="list-style-type: none"> <li>- velocity, orientation, and acceleration of each sensor, with respect to a base station</li> <li>- 3D position and/or orientation of each joint by post-processing (integration)</li> <li>- some are real-time</li> </ul>	[10,11,16,55]

Based on the above table, we can conclude that different motion capture sensors/techniques should be used, depending on the unique needs of each application. Parameters that should be considered when selecting an appropriate motion capture technique for a particular application include:

- Cost;
- required accuracy;
- requirements for interactivity/real-time performance;
- required easy calibration/self-calibration;
- number of joints to be tracked;
- weight/size of markers;
- level of restriction to (dancer) movements; and
- environmental constraints (e.g., existence of metallic objects or other noise sources affecting specific techniques).

One of the applications of motion capture lies in animation and special effects. Motion capture is an important source of motion data for computer animation, education, sports, the film industry, video-based games, medicine, ICH education and dances, and the military. More about specific applications can be found in [57].

Some advantages of using motion capture for the mentioned purposes are that it can accurately capture difficult-to-model physical movement, can provide virtual reality (VR) or augmented reality (AR), and that it takes fewer hours of work to animate the character. The downsides are that the motion capture requires special programs and data processing, and the price of motion capture equipment is high.

## 2.2. Post-Processing

Motion capture is mainly used for reproducing human animation. Motion capture data should be an accurate reflection of the real performance. Therefore, sensor information is transformed into an animated human figure [58]. The output of every system is similar—a set of the 3D positions in space is captured every frame. These data are usually transferred to some software and translated to the movement of the animation character. Motion capture data require cleaning because of the inaccuracy and unreliability of the data due to marker occlusion that can make the data noisy and incomplete. Regarding the data acquisition type, motion capture systems are classified into two categories [10]:

- Direct acquisition; and
- indirect acquisition.

Direct acquisition systems do not require any type of post-processing. Direct acquisition is a good solution since the recorded signals are coupled with discrete gestures uniquely. Each sensor captures a specific physical variable of the gesture. This category includes magnetic, mechanical, and acoustic systems, as they have been analyzed above. These systems are more obtrusive and offer a lower sampling rate. Indirect acquisition systems include optical systems. These systems enable more freedom for the user and a higher sampling rate. Data captured using these kinds of systems are processed by dedicated software. Most optical motion capture systems require human intervention. Identifying markers can be done through labeling, finding missing markers due to occlusion, and correcting possible errors detected in a rigidity test [59].

Modification of the pre-captured motions is still an open question. A lot of estimation and smoothing techniques are used for data post-processing, e.g., linear interpolation, Kalman filtering, a priori knowledge about rigid bodies [4]. Despite data post-processing being not required for some systems, all of them require the data to be cleaned, filtered, and mapped to a skeleton. In [4], post-processing involves two parts: Trajectory reconstruction and labeling of markers/trajectories. With this double stage completed, it is possible to visualize the technical skeleton. The subject skeleton and its animation are derived from that.

### 2.3. Archiving and Data Retrieval

Designing digital dance archives is an important step in the data storage process. The archives should be scalable, so new data and metadata can be added. Currently, there is no standardized method of dance recording and archiving, but several datasets are available. The archive available in [60] includes a textual description about dance types, video recordings, and motion capture data of individual performances, metadata of dancers appearing in performances, and the locations where these dances are performed. The largest publicly available motion capture database is [61], which contains movements associated with a variety of activities, including dances. Data are available in different formats, e.g., C3D, ASF/AMC. As few data of a two-subject interaction exist, the HDM12 database [62] provides Argentine Tango dance sequences, recorded of 11 different dance couples. More information regarding databases that contain locomotion, exercise, and every-day movements can be found in [63,64].

To fully use and exploit databases, efficient retrieval and browsing methods are needed. This is a difficult task due to complex spatio-temporal variances in human motions. Reusing existing data is much more time- and cost- efficient than capturing the whole motion from scratch. Motion retrieval systems can perform queries based on the different input, e.g., text, motion clips or key frames. For large datasets, efficient methods exist, such as techniques based on the query-by-example paradigm. This paradigm is based on the retrieval of all documents from a database containing parts similar to a given data fragment. For example, in [65], a motion retrieval system is presented that allows efficient retrieval of logically related motions based on the above-mentioned paradigm. Logically related motions do not need to be numerically similar. That means that even though motions are different considering timing, intensity, and execution style, they can describe the same motion. A key frame-based human motion capture data retrieval system, which uses a wooden doll as the input device, is described in [66]. After the user finishes inputting the key frames, the motion sequences are retrieved from the database and ranked based on the similarities to the key frames.

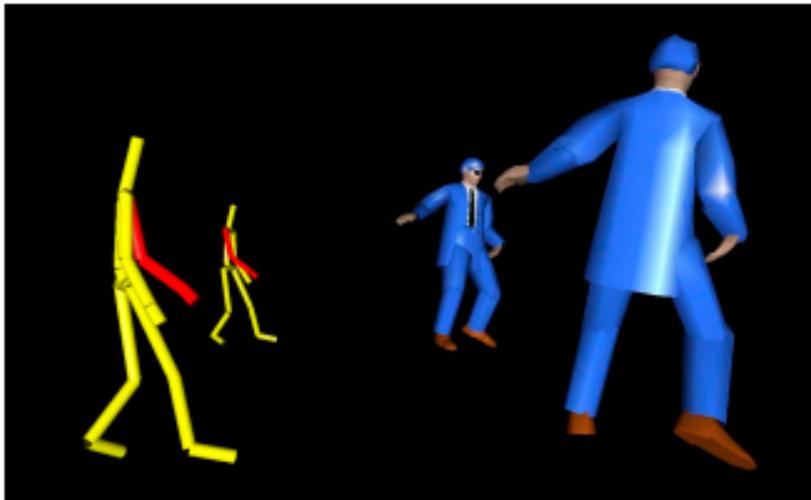
In [67], the human character is divided into three parts to reduce the spatial complexity. The temporal similarity of each part is measured by a self-organizing map and Smith-Waterman algorithm. The overall similarity between two motion clips is achieved by integrating the similarities of the separate parts. Muller et al. [68] proposed a system where a query consists of a short motion clip and a query-dependent specification of a motion aspect that determines the desired notion of similarity. More about motion data retrieval can be found in the literature [69–71].

## 3. Visualization

### 3.1. Types of Visualization and Feedback

After recording the dance, it is often useful to visualize the collected data. Different ways of visualizing dances and presenting them to the users for learning purposes can be found in the relevant literature. The interface of a learning application for visualization should be interesting, simple, and intuitive for the user. Users should be orientated to the dance learning process and not spend too much time on learning how to use any application. Another factor of major importance is the direct feedback to the users, depending on their performance, so that they are aware of their success in completing their “task”.

Video is an efficient way of preserving dances, but it suffers from a lack of feedback. In [4], a 3D viewer for dance learning was developed, and several functionalities were integrated. The user can watch the dance, choose the point of view and zoom level, and control the speed of the 3D animation. A VR training application combined with motion capture was proposed in [72]. The demonstration of the dance moves is done by rendering the 3D animation with OpenGL and the user can change the speed and point of view. The user is recorded during the learning process and several types of feedback are provided. The first type of feedback is illustrated in Figure 5. The color of a cylinder indicates whether the position of the body segment is correct.



**Figure 5.** Immediate feedback for the user [72].

The second type of feedback is a scoring mechanism, i.e., the user is shown a report about performance. The third type is a slow-motion replay, allowing the user to realize his/her mistakes. For making the learning process intuitive and motivating, a platform for visualizing dance events in the 3D virtual environment was developed [73]. For the user, it is possible to manipulate the 3D dancer through functionalities of start, stop, zoom, and focus, and to change the camera position.

In [29], users can learn by observing a teacher's dance moves. Performance of the teacher is recorded, and the position of key joints is extracted and stored. They can choose to watch the teacher's performance and imitate moves. They are also recorded, and an extraction of features is done. Feedback is given in two ways. Either by being evaluated by experts, or by the learning system matching the features of the teacher and the user. Moreover, the concept of the user observing the teacher's dance moves and repeating them is presented in [74]. The user can select the dance and the 3D avatar of the teacher. After watching the teacher's performance, the user is recorded during the performance of dance moves. The user's moves are compared to the motion template and an evaluation of the performance is given to the user.

The combination of gaming and learning introduced a new area in the educational domain. The popularity of games, especially among young people, makes them ideal for educational purposes. Serious games have the potential in teaching because they can promote training, knowledge acquisition, and skill development through interactive, engaging, or even immersive activities [9]. Game-like applications can be found in the literature for the preservation of folk dances. The process of adding games or game-like elements to encourage participation is known as gamification. Nowadays, gamification has become a popular way to encourage specific behaviors and increase motivation and engagement [75].

Creating a virtual 3D gaming environment where users can see their dance from any orientation is proposed in [76]. In this environment, users can also step forward/backward, pause, or continuously play back at a decreased framerate. Feedback is given by scoring a user's motion against a teacher's motion. More information regarding calculating scores can be found in [76].

Furthermore, in [30], a game interface is implemented, shown in Figure 6. The avatar of the teacher is shown in the corner, and the user, whose avatar is shown in the middle of the scene, must imitate dance moves. Real data from the second version of the Microsoft Kinect and a high-precision motion capture system, Qualisys, are used. The user's moves are captured using the Microsoft Kinect and sent to the game framework. The framework for dance learning runs on Unity. Feedback is given in the form of a score value with a comment. If the score is higher than 50%, then the next exercise is presented. Otherwise, the user must repeat the same exercise from the beginning.



Figure 6. Game interface [30].

A similar game-like application is presented in [9]. The user can watch the tutorial, where it is explained how to play the game. 3D animations, video recordings, and dance music are presented to the user. The learning activity consists of several exercises, while each exercise contains several dance moves, which are presented to the user one by one. To proceed to the next exercise, the user must perform the current exercise correctly three to five times and feedback is presented with an appropriate comment. The Kinect depth sensor is used for motion capture. Motion data are transmitted to the 3D game module for the visualization process.

A game for learning purposes using the Unity 3D game engine and Microsoft Kinect was also proposed [17]. The virtual dancer performs dance moves and the user is asked to repeat them. The user's dance moves are captured using a Microsoft Kinect sensor. Both avatars are displayed at the same time. Feedback is given through hints and advice on what the user should improve.

In another approach, the cave automatic virtual environment (CAVE) was proposed [77]. The CAVE is an immersive virtual environment where the projector's screen is placed between three to six walls of the room-sized cube. The user can watch the virtual teacher and repeat dance moves. Afterward, the user and the teacher perform the dance movements using the CAVE system. Three types of feedback and two types of playback are provided. The first type of feedback is side by side. Virtual models of the user and the teacher are shown on the screen, side by side. The user can watch performances and compare them. The second type is an overlay. Figures of the teacher and the student are overlaid, and it is possible to see the difference in moves. The third type is a score graph. The user's performance is recorded and presented in the form of a number or a trace.

Different kinds of feedback are proposed in [78]. There are two types of feedback: The vibrotactile and the acoustic one. Vibrotactile feedback is given with vibrational devices placed on each ankle, indicating the direction to which the dancer should move. The idea for acoustic feedback is to choose sounds for correct and incorrect steps. Both feedbacks are given to the dancer during the performance. A brief overview of the types of visualization is given in Table 2.

Table 2. Types of visualization.

Type of Visualization	Advantages	Disadvantages
Video	<ul style="list-style-type: none"> <li>- addresses movement/audio aspect</li> <li>- the most efficient way of preserving dance</li> </ul>	<ul style="list-style-type: none"> <li>- lack of interaction with the user</li> <li>- lack of wider context, meaning, and significance of the dance</li> <li>- lack of feedback</li> </ul>
Virtual reality (VR) environment	<ul style="list-style-type: none"> <li>- real-time feedback</li> <li>- ability to identify which part of the body moves incorrectly</li> <li>- reconstruction of real environments</li> <li>- realistic avatars</li> <li>- increased user engagement</li> </ul>	<ul style="list-style-type: none"> <li>- the distraction of the user may require experience to identify mistakes</li> <li>- immersion can cause motion sickness</li> </ul>
Game-like application (3D game environment)	<ul style="list-style-type: none"> <li>- improved interaction</li> <li>- more realistic environment and avatars</li> <li>- improved visualization of information</li> <li>- increased user engagement</li> </ul>	<ul style="list-style-type: none"> <li>- may be hard to identify mistakes</li> </ul>

### 3.2. Movements Recognition

Human activity recognition is an important area of computer vision research. Analysis, processing of motion capture datasets, and their reuse for the synthesis of novel motions is still a problem that needs to be solved in a better way. Motion, in general, consists of different actions, like dance moves, but also stylistic variations of moves. A challenging task for motion analysis and synthesis algorithms is generating plausible dance motions. The motion analysis framework in [79] is based on Laban movement analysis (LMA). LMA takes into consideration stylistic variations of the movement, which is very important for dances. This was implemented in the context of motion graphs, and used for elimination of potentially problematic transitions and synthetization of style-coherent animations without prior labeling of the data. Extracting relevant spatio-temporal features from dance movements of known emotions following the LMA was proposed in the framework in [80]. A set of effective and consistent features for emotion characterization was identified. These features were used to map a new input motion to their emotion coordinates on the Russell's circumplex model (RCM) of affect. The two-way mapping between the motion features and emotion coordinates through the radial basis function (RBF) regression and interpolation was implemented and can stylize freestyle highly dynamic dance movements at interactive rates.

The recognition of salsa dance steps is proposed in [81]. Using principal component analysis (PCA), motion features are extracted from 3D sub-trajectories of dancers' body-joints. The classification of dance gestures is done using a hidden Markov model (HMM).

The automatic extraction of choreographic patterns from the motion capture data is proposed in [82]. Choreographic patterns can provide an abstract representation of the dance semantics and encode the overall dance storytelling. The key-frame extraction method implements a hierarchical scheme that exploits spatio-temporal variations of dance features. An introduced spatio-temporal summarization algorithm considers 3D motion captured data represented by 3D joints that model the human skeleton. The global holistic descriptors are extracted to localize the key choreographic steps derived from the 3D human joints. Each segment is further decomposed into more detailed

sub-segments. The abstraction scheme uses the concept of a sparse modeling representative selection (SMRS) modified to enable spatio-temporal modeling of the dance sequences through a hierarchical decomposition algorithm. This approach was evaluated on thirty folkloric sequences.

The multilinear motion model for analysis and synthesis of personalized stylistic human motion was presented in [83]. Using this model, it is possible to adjust the parameters that control the “identity” and “style” variations of the action. Also, it is possible to interactively adjust the attribute parameters to match the constraints specified by the user. With this approach, the power and flexibility of multilinear motion models were demonstrated.

A system that recognizes the actions from skeleton data is presented in [84]. For each frame, features are extracted based on the relative position of joints, temporal differences, and normalized trajectories of motion. These features are used to train deep neural network-hybrid multi-layer perceptron, which simultaneously classifies and reconstructs input data.

A comprehensive comparative study of classifiers and data sampling schemes for dance pose identification based on motion capture data is presented in [85]. In this work, the effectiveness of several classifiers for dance recognition from skeleton data was tested. Classifiers that are used for dance pose identifications include k nearest neighbors (kNN), naïve Bayes, discriminant analysis, classification trees, and support vector machines. These are well-known classifiers used for recognition. The feature extraction process involved subtraction between successive frames and PCA for dimensionality reduction.

Motion-capture-based human identification, as a pattern of recognition discipline, can be optimized using a machine learning approach. The concept of learning motion features directly from raw joint coordinates by a modification of the Fisher’s linear discriminant analysis with maximum margin criterion (MMC), which was introduced in [86]. The point of interest is to find an optimal feature space where a template is close to those from the same person and different from those of different persons. To evaluate this technique, a large number of samples were extracted from the Carnegie Mellon University (CMU) database [61], and a number of submotions were extracted and filtered out. The final database and the evaluation framework are publicly available in [87]. A similar approach for extracting robust features from raw data using a modification of linear discriminant analysis with maximum margin criterion is presented in [88,89].

Human movements can be considered as a set of trajectories and can be used for the extraction of distance-time dependency signals (DTDS). In [90], several functions are proposed that compare various combinations of extracted signals. Signals were normalized and used as input parameters for the computation of the similarity of patterns. The Manhattan distance, the Euclidean distance, and the dynamic time warping (DTW) were used to measure the similarity of two DTDSs. The DTW-like comparison led to the 96% effectiveness.

#### 4. Performances Evaluation

The next step, after visualization, is to evaluate a dance performance. A typical approach to evaluating the performance is to use ground truth data [91]. Typically, for learning folk dances, ground truth data are provided by professionals. The ground truth data and the data obtained from the user are compared using different metrics and algorithms. The difficulty here is that different dancers have different dancing styles. It is possible for two dancers to dance the same dance in a different, but correct, way.

For measuring the motion similarity between the user and the teacher, two metrics were proposed in [51], using the knee-distance and ankle-distance for each frame. A specific normalization process was used to ensure the invariance of these metrics. Calculating maximum correlation coefficients between the user’s normalized distances and the teacher’s normalized distances, two motion accuracy scores were introduced. Furthermore, a choreography score was derived as the precision of the correct detection of motion patterns. The coefficients and the choreography score were then fed as input to a two-level fuzzy inference system (FIS) that outputs the final performance score.

Gesture recognition algorithms are often used to recognize a specific pose. These algorithms are very popular for hand gesture recognition, but they are becoming more popular in dance move recognition. A hidden Markov model-based system for real-time gesture recognition and performance evaluation was presented in [30]. Performed gestures were decoded using the system to provide, at the end of the recognized gesture, a likelihood value as a score that was used for the evaluation.

The comparison of three different measures for evaluating dance performance was proposed in [72]. The system computes the distance between the teacher's and learner's sequence using dynamic time wrapping (DTW) and the Euclidian distance between joint angles, joint positions, or joint velocities. To check if there is a significant difference between the different values, a T-test is used.

An automatic dance analysis tool for the evaluation of a learner's performance was proposed in [92]. The first metric evaluates the quality of movements and the accuracy the learner achieves. The second metric uses "timing" to assess the dancer's ability to keep in step with the teacher. With respect of both scores, the learner's performance is evaluated.

Wei et al. [93] proposed a three-part scheme for the evaluation of a dance sequence. The first part is related to motion correctness estimation, the second part is for rhythm management estimation, and the third part is for a comprehensive evaluation of the performance. By calculating the matching cost between the testing data and the standard model trained by the teacher, a motion correctness estimation was determined. Rhythm mistakes were detected using the second part of the scheme and a comprehensive grading level was provided for the user using the third part of the scheme.

Even though different metrics and algorithms are in use for performance evaluation there are still many open questions. As was already mentioned, ground truth data are needed. In the case of folk dances, ground truth data are collected using different motion capture systems. However, there is no perfect motion capture system, so capturing the "true" dance moves is elusive in a way. Furthermore, capturing the performance of the learner requires a real- (or near real-) time motion capture system, as performance evaluation, visualization, and provision of feedback to the user need to be supported. Hence, most of these systems use low-cost depth and/or optical sensors, such as Kinect. Studies [94–96] have shown that the performance of human motion estimation is highly dependent on the quality of the motion capture data and on the algorithms used. In addition, the selection of appropriate metrics for dance performance evaluation is not an easy task. For instance, there is a need to compensate for differences between the skeletal models (e.g., lengths of body parts) of the teacher and learner or differences in their motion styles.

## 5. Conclusions

In a struggle to preserve ICH through the ages, new technologies are currently used for digitization of folk dancing. The goal is to enhance and, hence, safeguard this significant element of peoples' identity. As seen, many tools currently exist to transform ICH, and specifically folk dances, into digital information, suitable for various purposes and applications. In all processing stages, i.e., preparation, recording, data processing, and archival, different systems have been used, with different advantages and disadvantages. Surprisingly enough, body sensors, cameras, and other high-end pieces of hardware are used for the precise recording of a procedure, which is normally considered a technology-free expression of dancing groups. Using these technologies, accurate folk dancing representations can be produced that will later make the transmission of knowledge easier.

The result of such processes is continuously proven to provide new ways for dance teachers, choreographers, game developers, and more to communicate detailed dancing elements to learners, dancers, and gamers, respectively. Interactivity, the flexibility of the system, or feedback for the users, referring to either a hardware- or software-based platform, are important factors for the successful digitization of folk dances. As it has been demonstrated through this research paper, the study of these parameters defines the success of the evolution of the current systems and contributes significantly to the knowledge for the development of new, advanced systems for recording, digitizing, and visualizing folk dances, as well as other forms of ICH.

Even though the existing technology can help with the preservation of folk dances and ICH, there are still some open questions awaiting further exploration. Motion capture systems that have been presented are used for recordings, but all of them demonstrate certain disadvantages. Furthermore, different motion capture sensors, algorithms, and parameterizations are required, depending on the particular needs of each application. Improving the existing systems, or finding new (e.g., hybrid) solutions in the future, could lead to improved estimation of human pose and motion, as well as better ground truth data and performance evaluation. Music acquires a huge role in dances. To learn a dance correctly, dance moves should be synchronized with music in the right way. This needs to be considered during the digitization and visualization of folk dances. A lot of effort has been put into micro-parameters for folk dancing visualization, but still, there is space for further improvements and explorations. For instance, to make a visual representation of the dancer more realistic, proper dance clothes can be added. Simulation of moving the clothes during the dance performance can help with this. Screens are also traditionally used for visualization and presenting the dance and feedback for the users. The users may not always face the screen during the learning process. It is needed to provide more screens, so users can track the performance. Visualization using virtual reality has open questions on how to visualize the body of the user so that the user can track their movements. Also, moving in VR can cause motion sickness. Video representation has a lot of disadvantages, but it is still a very good way of teaching dances and appears able to compete with the existing 3D applications.

Different ways of performance evaluation and giving feedback have been described, and specific benefits for certain groups have been shown. For this model to work as flawlessly as possible, and to visualize a future system that applies recognition of dance movements at a more sophisticated level, there is a crucial path to be followed: Applying different algorithms, especially machine learning and deep learning algorithms, can automate the tasks of feature extraction and pattern recognition, and so create an advanced system that will easily be used by more groups at a much wider level.

**Author Contributions:** Conceptualization, I.K. and F.L.; Data curation, I.K., N.G., Y.C. and F.L.; Funding acquisition, F.L.; Investigation, I.K., N.G. and Y.C.; Methodology, I.K.; Project administration, F.L.; Resources, I.K., N.G. and Y.C.; Supervision, F.L.; Validation, N.G., Y.C. and F.L.; Visualization, I.K.; Writing—Original draft, I.K.; Writing—Review & editing, F.L.

**Funding:** This work is supported under the H2020 European Union funded project TERPSICHORE: Transforming Intangible Folkloric Performing Arts into Tangible Choreographic Digital Objects, under the grant agreement 691218.

**Conflicts of Interest:** The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, and in the decision to publish the results.

## References

1. UNESCO. What Is Intangible Cultural Heritage. Available online: <https://ich.unesco.org/en/what-is-intangible-heritage-00003> (accessed on 5 June 2018).
2. Protopapadakis, E.; Grammatikopoulou, A.; Doulamis, A.; Grammalidis, N. Folk Dance Pattern Recognition over Depth Images Acquired via Kinect Sensor. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2017**, *XLIII-2/W3*, 587–593. [CrossRef]
3. Hachimura, K.; Kato, H.; Tamura, H. A Prototype Dance Training Support System with Motion Capture and Mixed Reality Technologies. In Proceedings of the 2004 IEEE International Workshop on Robot and Human Interactive Communication Kurashiki, Okayama, Japan, 20–22 September 2004; pp. 217–222.
4. Magnenat Thalmann, N.; Protosaltou, D.; Kavakli, E. Learning How to Dance Using a Web 3D Platform. In Proceedings of the 6th International Conference Edinburgh, Revised Papers, UK, 15–17 August 2007; Leung, H., Li, F., Lau, R., Li, Q., Eds.; Springer: Berlin/Heidelberg, Germany, 2008; pp. 1–12.
5. Doulamis, A.; Voulodimos, A.; Doulamis, N.; Soile, S.; Lampropoulos, A. Transforming Intangible Folkloric Performing Arts into Tangible Choreographic Digital Objects: The Terpsichore Approach. In Proceedings of the 12th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISIGRAPP 2017), Porto, Portugal, 27 February 2017–1 March 2017; Volume 5, pp. 451–460.

6. Transforming Intangible Folkloric Performing Arts into Tangible Choreographic Digital Objects. Available online: <http://terpsichore-project.eu/> (accessed on 25 June 2018).
7. Grammalidis, N.; Dimitropoulos, K.; Tsalakanidou, F.; Kitsikidis, A.; Roussel, P.; Denby, B.; Chawah, P.; Buchman, L.; Dupont, S.; Laraba, S.; et al. The i-Treasures Intangible Cultural Heritage dataset. In Proceedings of the 3rd International Symposium on Movement and Computing (MOCO'16), Thessaloniki, Greece, 5–6 July 2016; ISBN 978-1-4503-4307-7. [[CrossRef](#)]
8. Dimitropoulos, K.; Manitsaris, S.; Tsalakanidou, F.; Denby, B.; Crevier-Buchman, L.; Dupont, S.; Nikolopoulos, S.; Kompatsiaris, Y.; Charisis, V.; Hadjileontiadis, L.; et al. A Multimodal Approach for the Safeguarding and Transmission of Intangible Cultural Heritage: The Case of i-Treasures. *IEEE Intell. Syst.* **2018**. [[CrossRef](#)]
9. Kitsikidis, A.; Dimitropoulos, K.; Ugurca, D.; Baycay, C.; Yilmaz, E.; Tsalakanidou, F.; Douka, S.; Grammalidis, N. A Game-like Application for Dance Learning Using a Natural Human Computer Interface. In Part of HCI International, Proceedings of the 9th International Conference (UAHCI 2015), Los Angeles, CA, USA, 2–7 August 2015; Antona, M., Stephanidis, C., Eds.; Springer International Publishing: Basel, Switzerland, 2015; pp. 472–482.
10. Nogueira, P. Motion Capture Fundamentals—A Critical and Comparative Analysis on Real World Applications. In Proceedings of the 7th Doctoral Symposium in Informatics Engineering, Porto, Portugal, 26–27 January 2012; Oliveira, E., David, G., Sousa, A.A., Eds.; Faculdade de Engenharia da Universidade do Porto: Porto, Portugal, 2012; pp. 303–331.
11. Tsampounaris, G.; El Raheb, K.; Katifori, V.; Ioannidis, Y. Exploring Visualizations in Real-time Motion Capture for Dance Education. In Proceedings of the 20th Pan-Hellenic Conference on Informatics (PCI'16), Patras, Greece, 10–12 November 2016; ACM: New York, NY, USA, 2016.
12. Hachimura, K. Digital Archiving on Dancing. *Rev. Natl. Cent. Digit.* **2006**, *8*, 51–60.
13. Hong, Y. The Pros and Cons about the Digital Recording of Intangible Cultural Heritage and Some Strategies. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2015**, *XL-5/W7*, 461–464. [[CrossRef](#)]
14. Giannoulakis, S.; Tsapatsoulis, N.; Grammalidis, N. Metadata for Intangible Cultural Heritage—The Case of Folk Dances. In Proceedings of the 13th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications, Funchal, Madeira, 27–29 January 2018; pp. 534–545.
15. Pavlidis, G.; Koutsoudis, A.; Arnaoutoglou, F.; Tsioukas, V.; Chamzas, C. Methods for 3D digitization of Cultural Heritage. *J. Cult. Herit.* **2007**, *8*, 93–98. [[CrossRef](#)]
16. Sementille, A.C.; Lourenco, L.E.; Brega, J.R.F.; Rodello, I. A Motion Capture System Using Passive Markers. In Proceedings of the 2004 ACM SIGGRAPH International Conference on Virtual Reality Continuum and Its Applications in Industry (VRCAI'04), Singapore, 16–18 June 2004; ACM: New York, NY, USA, 2004; pp. 440–447.
17. Stavarakis, E.; Aristidou, A.; Savva, M.; Loizidou Himona, S.; Chrysanthou, Y. Digitization of Cypriot Folk Dances. In Proceedings of the 4th International Conference (EuroMed 2012), Limassol, Cyprus, 29 October–3 November 2012; Ioannides, M., Fritsch, D., Leissner, J., Davies, R., Remondino, F., Caffo, R., Eds.; Springer: Berlin/Heidelberg, Germany, 2012; pp. 404–413.
18. Johnson, L.M. Redundancy Reduction in Motor Control. Ph.D. Thesis, The University of Texas at Austin, Austin, TX, USA, December 2015.
19. Matus, H.; Kico, I.; Dolezal, M.; Chmelik, J.; Doulamis, A.; Liarokapis, F. Digitization and Visualization of Movements of Slovak Folk Dances. In Proceedings of the International Conference on Interactive Collaborative Learning (ICL), Kos Island, Greece, 25–28 September 2018.
20. Mustafa, N.; Idris, M.Z. Accessing Accuracy of Structural Performance on Basic Steps in Recording Malay Zapin Dance Movement Using Motion Capture. *J. Appl. Environ. Biol. Sci.* **2017**, *7*, 165–173.
21. Hegarini, E.; Syakur, A. Indonesian Traditional Dance Motion Capture Documentation. In Proceedings of the 2nd International Conference on Science and Technology-Computer (ICST), Yogyakarta, Indonesia, 27–28 October 2016.
22. Pons, J.P.; Keriven, R. Multi-View Stereo Reconstruction and Scene Flow Estimation with a Global Image-Based Matching Score. *Int. J. Comput. Vis.* **2007**, *72*, 179–193. [[CrossRef](#)]
23. Li, R.; Sclaroff, S. Multi-scale 3D Scene Flow from Binocular Stereo Sequences. *Comput. Vis. Image Underst.* **2008**, *110*, 75–90. [[CrossRef](#)]

24. Chun, C.W.; Jenkins, O.C.; Mataric, M.J. Markerless Kinematic Model and Motion Capture from Volume Sequences. In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Madison, WI, USA, 18–20 June 2003.
25. Sell, J.; O'Connor, P. The Xbox One System on a Chip and Kinect Sensor. *IEEE Micro* **2014**, *34*, 44–53. [[CrossRef](#)]
26. Izadi, S.; Kim, D.; Hilliges, O.; Molyneaux, D.; Newcombe, R.; Kohli, P.; Shotton, J.; Hodges, S.; Freeman, D.; Davison, A.; et al. KinectFusion: Real-time 3D Reconstruction and Interaction Using a Moving Depth Camera. In Proceedings of the 24th Annual ACM Symposium on User Interface Software and Technology (UIST'11), Santa Barbara, CA, USA, 16–19 October 2011; ACM: New York, NY, USA, 2011; pp. 559–568.
27. Newcombe, R.A.; Fox, D.; Seitz, S.M. DynamicFusion: Reconstruction and tracking of non-rigid scenes in real-time. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, 7–12 June 2015.
28. Shotton, J.; Fitzgibbon, A.; Cook, M.; Sharp, T.; Finocchio, M.; Moore, R.; Blake, A. Real-time human pose recognition in parts from single depth images. In Proceedings of the Computer Vision and Pattern Recognition (CVPR) 2011, Colorado Springs, CO, USA, 20–25 June 2011.
29. Kanawong, R.; Kanwaratraton, A. Human Motion Matching for Assisting Standard Thai Folk Dance Learning. *GSTF J. Comput.* **2018**, *5*, 1–5. [[CrossRef](#)]
30. Laraba, S.; Tilmanne, J. Dance performance evaluation using hidden Markov models. *Comput. Animat. Virtual Worlds* **2016**, *27*, 321–329. [[CrossRef](#)]
31. Moeslund, T.B.; Hilton, A.; Kruger, V. A survey of advances in vision-based human motion capture and analysis. *Comput. Vis. Image Underst.* **2006**, *104*, 90–126. [[CrossRef](#)]
32. Andriluka, M.; Pishchulin, L.; Gehler, P.; Schiele, B. 2D human pose estimation: New benchmark and state of the art analysis. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Columbus, OH, USA, 23–28 June 2014.
33. Cao, Z.; Simon, T.; Wei, S.E.; Sheikh, Y. Realtime multi-person 2D pose estimation using part affinity fields. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2011.
34. Wei, S.E.; Ramakrishna, V.; Kanade, T.; Sheikh, Y. Convolutional pose machines. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 4724–4732.
35. Simon, T.; Joo, H.; Matthews, I.; Sheikh, Y. Hand keypoint detection in single images using multiview bootstrapping. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; Volume 2.
36. Zhou, X.; Huang, Q.; Sun, X.; Xue, X.; Wei, Y. Towards 3D human pose estimation in the wild: A weakly-supervised approach. In Proceedings of the 2017 IEEE International Conference on Computer Vision (ICCV), Venice, Italy, 22–29 October 2017.
37. Newell, A.; Yang, K.; Deng, J. Stacked hourglass networks for human pose estimation. In Proceedings of the 14th European Conference Computer Vision (ECCV) 2016, Amsterdam, The Netherlands, 11–14 October 2016; Liebe, B., Matas, J., Sebe, N., Welling, M., Eds.; Springer: Cham, Switzerland, 2016; pp. 483–499.
38. Mehta, D.; Sridhar, S.; Sotnychenko, O.; Rhodin, H.; Shafiei, M.; Seidel, H.P.; Theobalt, C. VNect: Real-time 3D human pose estimation with a single RGB camera. *ACM Trans. Gr.* **2017**, *36*, 44. [[CrossRef](#)]
39. Mehta, D.; Rhodin, H.; Casas, D.; Fua, P.; Sotnychenko, O.; Xu, W.; Theobalt, C. Monocular 3D human pose estimation in the wild using improved CNN supervision. In Proceedings of the 2017 International Conference on 3D Vision (3DV), Qingdao, China, 10–12 October 2017; pp. 506–516.
40. Güler, R.A.; Neverova, N.; Kokkinos, I. DensePose: Dense human pose estimation in the wild. In Proceedings of the CVPR, Salt Lake, UT, USA, 18–22 June 2018.
41. Güler, R.A.; Trigeorgis, G.; Antonakos, E.; Snape, P.; Zafeiriou, S.; Kokkinos, I. DenseReg: Fully convolutional dense shape regression in-the-wild. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017.
42. He, K.; Gkioxari, G.; Dollar, P.; Girshick, R. Mask R-CNN. In Proceedings of the IEEE International Conference on Computer Vision (ICCV), Venice, Italy, 22–29 October 2017.
43. Kanazawa, A.; Black, M.J.; Jacobs, D.W.; Malik, J. End-to-end recovery of human shape and pose. In Proceedings of the Computer Vision and Pattern Recognition (CVPR), Salt Lake, UT, USA, 18–22 June 2018.

44. Loper, M.; Mahmood, N.; Romero, J.; Pons-Moll, G.; Black, M.J. SMPL: A skinned multi-person linear model. *ACM Trans. Gr.* **2015**, *34*, 248. [[CrossRef](#)]
45. Gong, W.; Zhang, X.; Gonzalez, J.; Sobral, A.; Bouwmans, T.; Tu, C.; Zahzah, E. Human Pose Estimation from Monocular Images: A Comprehensive Survey. *Sensors* **2016**, *16*, 1996. [[CrossRef](#)] [[PubMed](#)]
46. Ke, S.; Thuc, H.L.U.; Lee, Y.J.; Hwang, J.N.; Yoo, J.H.; Choi, K.H. A Review on Video-Based Human Activity Recognition. *Computers* **2013**, *2*, 88–131. [[CrossRef](#)]
47. Neverova, N. Deep Learning for Human Motion Analysis. Ph.D. Thesis, Universite de Lyon, Lyon, France, 2016. [[CrossRef](#)]
48. Alexiadis, D.S.; Chatzitofis, A.; Zioulis, N.; Zoidi, O.; Louizis, G.; Zarpalas, D.; Daras, P. An integrated platform for live 3D human reconstruction and motion capturing. *IEEE Trans. Circuits Syst. Video Technol.* **2017**, *27*, 798–813. [[CrossRef](#)]
49. Alexiadis, D.S.; Zarpalas, D.; Daras, P. Real-time, full 3-D reconstruction of moving foreground objects from multiple consumer depth cameras. *IEEE Trans. Multimed.* **2013**, *15*, 339–358. [[CrossRef](#)]
50. Alexiadis, D.S.; Zarpalas, D.; Daras, P. Real-time, realistic full body 3D reconstruction and texture mapping from multiple Kinects. In Proceedings of the IVMS 2013, Seoul, Korea, 10–12 June 2013.
51. Kitsikidis, A.; Dimitropoulos, K.; Yilmaz, E.; Douka, S.; Grammalidis, N. Multi-sensor technology and fuzzy logic for dancer's motion analysis and performance evaluation within a 3D virtual environment. In *Part of HCI International 2014, Proceedings of the 8th International Conference (UAHCI 2014), Heraklion, Crete, Greece, 22–27 June 2014*; Stephanidis, C., Antona, M., Eds.; Springer International Publishing: Basel, Switzerland, 2014; pp. 379–390.
52. Kahn, S.; Keil, J.; Muller, B.; Bockholt, U.; Fellner, D.W. Capturing of Contemporary Dance for Preservation and Presentation of Choreographies in Online Scores. In Proceedings of the 2013 Digital Heritage International Congress, Marseille, France, 28 October–1 November 2013.
53. Robertini, N.; Casas, D.; Rhodin, H.; Seidel, H.P.; Theobalt, C. Model-based outdoor performance capture. In Proceedings of the 2016 Fourth International Conference on 3D Vision (3DV), Stanford, CA, USA, 25–28 October 2016.
54. Meta Motion. Available online: <http://metamotion.com/> (accessed on 10 September 2018).
55. Vlastic, D.; Adelsberger, R.; Vannucci, G.; Barnwell, J.; Gross, M.; Matusik, W.; Popovic, J. Practical Motion Capture in Everyday Surroundings. *ACM Trans. Gr.* **2007**, *26*. [[CrossRef](#)]
56. Yabukami, S.; Yamaguchi, M.; Arai, K.I.; Takahashi, K.; Itagaki, A.; Wako, N. Motion Capture System of Magnetic Markers Using Three-Axial Magnetic Field Sensor. *IEEE Trans. Magn.* **2000**, *36*, 3646–3648. [[CrossRef](#)]
57. Sharma, A.; Agarwal, M.; Sharma, A.; Dhuria, P. Motion Capture Process, Techniques and Applications. *Int. J. Recent Innov. Trends Comput. Commun.* **2013**, *1*, 251–257.
58. Bodenheimer, B.; Rose, C.; Rosenthal, S.; Pella, J. The Process of Motion Capture: Dealing with the Data. In Proceedings of the Eurographics Workshop, Budapest, Hungary, 2–3 September 1997; Thalmann, D., van de Panne, M., Eds.; Springer: Vienna, Austria, 1997; pp. 3–18.
59. Gutemberg, B.G. Optical Motion Capture: Theory and Implementation. *J. Theor. Appl. Inform.* **2005**, *12*, 61–89.
60. University of Cyprus. Dance Motion Capture Database. Available online: <http://www.dancedb.eu/> (accessed on 28 June 2018).
61. Carnegie Mellon University Graphics Lab: Motion Capture Database. Available online: <http://mocap.cs.cmu.edu> (accessed on 25 June 2018).
62. Voge, A.; Kruger, B. *HDM12 Dance—Documentation on a Data Base of Tango Motion Capture*; Technical Report, No. CG-2016-1; Universitat Bonn: Bonn, Germany, 2016; ISSN 1610-8892.
63. Muller, M.; Roder, T.; Clausen, M.; Eberhardt, B.; Kruger, B.; Weber, A. *Documentation Mocap Database HDM05*; Computer Graphics Technical Reports, No. CG-2007-2; Universitat Bonn: Bonn, Germany, 2007; ISSN 1610-8892.
64. ICS Action Database. Available online: <http://www.miubiq.cs.titech.ac.jp/action/> (accessed on 25 June 2018).
65. Demuth, B.; Roder, T.; Muller, M.; Eberhardt, B. An Information Retrieval System for Motion Capture Data. In Proceedings of the 28th European Conference on Advances in Information Retrieval (ECIR'06), London, UK, 10–12 April 2006; Springer: Berlin/Heidelberg, Germany, 2006; pp. 373–384.

66. Feng, T.C.; Gunwardane, P.; Davis, J.; Jiang, B. Motion Capture Data Retrieval Using an Artist's Doll. In Proceedings of the 2008 19th International Conference on Pattern Recognition, Tampa, FL, USA, 8–11 December 2008.
67. Wu, S.; Wang, Z.; Xia, S. Indexing and Retrieval of Human Motion Data by a Hierarchical Tree. In Proceedings of the 16th ACM Symposium on Virtual Reality Software and Technology (VRST'09), Kyoto, Japan, 18–20 November 2009; ACM: New York, NY, USA, 2009.
68. Muller, M.; Roder, T.; Clausen, M. Efficient Content-Based Retrieval of Motion Capture Data. *ACM Trans. Gr.* **2005**, *24*, 677–685. [[CrossRef](#)]
69. Muller, M.; Roder, T. Motion Templates for Automatic Classification and Retrieval of Motion Capture Data. In Proceedings of the 2006 ACM SIGGRAPH/Eurographics Symposium on Computer Animation, Vienna, Austria, 2–4 September 2006; pp. 137–146.
70. Ren, C.; Lei, X.; Zhang, G. Motion Data Retrieval from Very Large Motion Databases. In Proceedings of the 2011 International Conference on Virtual Reality and Visualization, Beijing, China, 4–5 November 2011.
71. Muller, M. *Information Retrieval for Music and Motion*, 1st ed.; Springer: Berlin/Heidelberg, Germany, 2007; ISBN 978-3-540-74048-3.
72. Chan, C.P.J.; Leung, H.; Tang, K.T.J.; Komura, T. A Virtual Reality Dance Training System Using Motion Capture Technology. *IEEE Trans. Learn. Technol.* **2011**, *4*, 187–195. [[CrossRef](#)]
73. Bakogianni, S.; Kavakli, E.; Karkou, V.; Tsakogianni, M. Teaching Traditional Dance using E-learning tools: Experience from the WebDANCE project. In Proceedings of the 21st World Congress on Dance Research, Athens, Greece, 5–9 September 2007; International Dance Council CID-UNESCO: Paris, France, 2007.
74. Aristidou, A.; Stavrakis, E.; Charalambous, P.; Chrysanthou, Y.; Loizidou Himona, S. Folk Dance Evaluation Using Laban Movement Analysis. *ACM J. Comput. Cult. Herit.* **2015**, *8*. [[CrossRef](#)]
75. Hamari, J.; Koivisto, J.; Sarsa, H. Does Gamification Work—A Literature Review of Empirical Studies on Gamification. In Proceedings of the 2014 47th Hawaii International Conference on System Science, Waikoloa, HI, USA, 6–9 January 2014.
76. Alexiadis, D.; Daras, P.; Kelly, P.; O'Connor, N.E.; Boubekeur, T.; Moussa, M.B. Evaluating a Dancer's Performance using Kinect-based Skeleton Tracking. In Proceedings of the 19th ACM International Conference on Multimedia (MM'11), Scottsdale, AZ, USA, 28 November–1 December 2011; ACM: New York, NY, USA, 2011; pp. 659–662.
77. Kyan, M.; Sun, G.; Li, H.; Zhong, L.; Muneesawang, P.; Dong, N.; Elder, B.; Guan, L. An Approach to Ballet Dance Training through MS Kinect and Visualization in a CAVE Virtual Reality Environment. *ACM Trans. Intell. Syst. Technol.* **2015**, *6*. [[CrossRef](#)]
78. Drobny, D.; Borchers, J. Learning Basic Dance Choreographies with Different Augmented Feedback Modalities. In Proceedings of the Extended Abstracts on Human Factors in Computing Systems (CHI '10), Atlanta, GA, USA, 14–15 April 2010; ACM: New York, NY, USA, 2010; pp. 3793–3798.
79. Aristidou, A.; Stavrakis, E.; Papaefthimiou, M.; Papagiannakis, G.; Chrysanthou, Y. Style-based Motion Analysis for Dance Composition. *Int. J. Comput. Games* **2018**, *34*, 1–13. [[CrossRef](#)]
80. Aristidou, A.; Zeng, Q.; Stavrakis, E.; Yin, K.; Cohen-Or, D.; Chrysanthou, Y.; Chen, B. Emotion Control of Unstructured Dance Movements. In Proceedings of the ACM SIGGRAPH/Eurographics Symposium on Computer Animation (SCA'17), Los Angeles, CA, USA, 28–30 July 2017; ACM: New York, NY, USA, 2017.
81. Masurelle, A.; Essid, S.; Richard, G. Multimodal Classification of Dance Movements Using Body Joint Trajectories and Step Sounds. In Proceedings of the 2013 14th International Workshop on Image Analysis for Multimedia Interactive Services (WIAMIS), Paris, France, 3–5 July 2013.
82. Rallis, I.; Doulamis, N.; Doulamis, A.; Voulodimos, A.; Vescoukis, V. Spatio-temporal summarization of dance choreographies. *Comput. Gr.* **2018**, *73*, 88–101. [[CrossRef](#)]
83. Min, J.; Liu, H.; Chai, J. Synthesis and Editing of Personalized Stylistic Human Motion. In Proceedings of the 2010 ACM SIGGRAPH Symposium on Interactive 3D Graphics and Games (I3D'10), Washington, DC, USA, 19–21 February 2010; ACM: New York, NY, USA, 2010.
84. Cho, K.; Chen, X. Classifying and Visualizing Motion Capture Sequences using Deep Neural Networks. In Proceedings of the 9th International Conference on Computer Vision Theory and Applications (VISAPP 2014), Lisbon, Portugal, 5–8 January 2014.

85. Protopapadakis, E.; Voulodimos, A.; Doulamis, A.; Camarinopoulos, S.; Doulamis, N.; Miaoulis, G. Dance Pose Identification from Motion Capture Data: A Comparison of Classifiers. *Technologies* **2018**, *6*, 31. [CrossRef]
86. Balazia, M.; Sojka, P. Walker-Independent Features for Gait Recognition from Motion Capture Data. In *Structural, Syntactic, and Statistical Pattern Recognition*; Robles-Kelly, A., Loog, M., Biggio, B., Escolano, F., Wilson, R., Eds.; Lecture Notes in Computer Science; Springer: Cham, Switzerland, 2016; Volume 10029.
87. Gait Recognition from Motion Capture Data. Available online: <https://gait.fi.muni.cz/> (accessed on 8 September 2018).
88. Balazia, M.; Sojka, P. Gait Recognition from Motion Capture Data. *ACM Trans. Multimed. Comput. Commun. Appl.* **2018**, *14*. [CrossRef]
89. Balazia, M.; Sojka, P. Learning Robust Features for Gait Recognition by Maximum Margin Criterion. In Proceedings of the 23rd IEEE/IAPR International Conference on Pattern Recognition (ICPR 2016), Cancun, Mexico, 4–8 September 2016.
90. Sedmidubsky, J.; Valcik, J.; Balazia, M.; Zezula, P. Gait Recognition Based on Normalized Walk Cycles. In *Advances in Visual Computing*; Bebis, G., Ed.; Lecture Notes in Computer Science; Springer: Berlin/Heidelberg, Germany, 2012; Volume 7432.
91. Black, J.; Ellis, T.; Rosin, P.L. A Novel Method for Video Tracking Performance Evaluation. In Proceedings of the IEEE International Workshop on Visual Surveillance and Performance Evaluation of Tracking and Surveillance (VS-PETS), Nice, France, 11–12 October 2003.
92. Essid, S.; Alexiadis, D.; Tournemene, R.; Gowing, M.; Kelly, P.; Monaghan, D.; Daras, P.; Dremeau, A.; O'Connor, E.N. An Advanced Virtual Dance Performance Evaluator. In Proceedings of the 2012 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Kyoto, Japan, 25–30 March 2012.
93. Wei, Y.; Yan, H.; Bie, R.; Wang, S.; Sun, L. Performance monitoring and evaluation in dance teaching with mobile sensing technology. *Pers. Ubiquitous Comput.* **2014**, *18*, 1929–1939. [CrossRef]
94. Wang, Y.; Baciú, G. Human motion estimation from monocular image sequence based on cross-entropy regularization. *Pattern Recognit. Lett.* **2003**, *24*, 315–325. [CrossRef]
95. Tong, M.; Liu, Y.; Huang, T.S. 3D human model and joint parameter estimation from monocular image. *Pattern Recognit. Lett.* **2007**, *28*, 797–805. [CrossRef]
96. Luo, W.; Yamasaki, T.; Aizawa, K. Cooperative estimation of human motion and surfaces using multiview videos. *Comput. Vis. Image Underst.* **2013**, *117*, 1560–1574. [CrossRef]



© 2018 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).