Computer Vision for the Ballet Industry: A Comparative Study of Methods for Pose Recognition

Margaux Fourie[(✉)] and Dustin van der Haar[(✉)]

University of Johannesburg, Kingsway Avenue and University Rds,
Auckland Park, Johannesburg, South Africa
{margauxf,dvanderhaar}@uj.ac.za

Abstract. The presence of computer vision technology is continually expanding into multiple application domains. An industry and an art form that is particularly attractive for the application of computer vision algorithms is ballet. Due to the well-codified poses, along with the challenges that exist within the ballet domain, automation for the ballet environment is a relevant research problem. The paper proposes a model called BaReCo, which allows for ballet poses to be recognised using computer vision methods. The model contains multiple computer vision pipelines which allows for the comparison of approaches that have not been widely explored in the ballet domain. The results have shown that the top-performing pipelines achieved an accuracy rate of 99.375% and an Equal Error Rate (EER) of 0.119% respectively. The study additionally produced a ballet pose dataset, which serves as a contribution to the ballet and computer vision community. By combining suitable computer vision methods, the study demonstrates that successful recognition of ballet poses can be accomplished.

Keywords: Computer vision · Pose recognition · Dance · Ballet industry · Automation

1 Introduction

Movements of the human body can be captured using various types of cameras and sensors. As a result of the increased availability of new capture technologies as well as Graphics Processing Units (GPUs), computer vision technology is increasingly used in order to perform body movement and pose recognition [1,2]. Those involved in human motion and pose-based fields such as medical, sports and performing arts can benefit from additional technological assistance, which contributes to the improvement in field-related tasks. Therefore, new ways of applying computer vision to these fields, in order to enhance and contribute to the lives of those involved, need to be explored [3].

Ballet is a particularly noteworthy form of dance as it has a history reaching back to the 16th century and is, therefore, a well respected, established art

form [4]. It has become foundational for a large majority of other dance forms [5]. The precise structure of ballet is what makes it an attractive art form to investigate and further explore in terms of its relevance to technological fields.

Initially, the scientific domain of technology may seem to be completely separate and unrelated to the artistic domain of ballet. However, this study explored the intricacies of both ballet and technology to reveal how applicable and suitable it may be to bring the two fields together. Artists and scientists have similar aims in their work as both intend to produce work that is novel and original. As a result, the collaboration between an artistic field such as ballet and a scientific field such as computer technology may be valuable to address challenges in both fields [6]. Specific challenges in the ballet environment include individual training correction as well as the accurate documentation of choreographic works.

Ballet pose recognition and correction is an activity executed by ballet dancers and teachers frequently during training sessions. Pose recognition is also relevant to ballet choreography where sequences of postures are involved in the construction of dances. Research has been done on how technology, such as computer vision can be applied to the ballet environment. However, it is still a growing application field with room for the exploration of a range of technological automation approaches. The research problem of the study involves the need for a standardised approach that can recognise multiple ballet poses automatically. This study is, therefore, an initial step towards further research to aid in the tasks of ballet training and choreography.

The paper aims to address the problem with the use of a captured dataset and proposing a model that combines multiple traditional and novel computer vision methods to achieve ballet pose recognition. Background on the problem is provided first, along with current work conducted in related fields. The experiment setup is described next, followed by a discussion on the model. The results are then presented, and the paper ends with a conclusion which highlights key findings and future work.

2   Problem Background

Ballet is a dance form where the dancer's movements are composed out of predefined poses [7]. The clarity with which a dancer performs these poses needs to be enforced to maintain the aesthetics required by the ballet art form. The application of computer vision to the domain of ballet has become a relevant area of research due to growth in the computer vision field over recent years.

Ballet technique is the driving force behind the definite structure of the dance form and is concerned with the creation of strong body lines, poses and fluid movements. Foundational concepts of ballet technique include turnout (outward rotation of the legs), alignment (vertical and horizontal lines of the shoulders and hips) as well as stretched legs and feet [7]. To achieve these classical ideals, it is essential that ballet teachers accurately convey the fundamental principals of technique to students during training, specifically as they relate to different poses.

The environment in which ballet training typically occurs is a studio class-room where a ballet teacher instructs and corrects about 8–10 students [8]. Expertise and knowledge in ballet are traditionally passed on verbally to future dancers during these coaching sessions [9]. However, a challenge that arises in this environment is that the teacher cannot keep his/her eyes on every student simultaneously. It is therefore often essential for the serious ballet student to receive one-on-one coaching [10]. Furthermore, without proper training and a grasp of technique, ballet dancers risk developing bad habits or injuries [11].

The creative process of constructing dance sequences is known as choreogra-phy, and in ballet, a series of the codified movements are combined to produce the resulting visual expression [12]. The documentation of ballet choreography is important for the preservation and protection of created dances as well as enabling future dancers to study the created works [13]. There is a need for finding new methods to address the challenges in the choreographic domain and this is where the use of technology becomes relevant.

A few studies in current research that address problems in the ballet envi-ronment through the use of technology include wearable technology systems, choreography systems, as well as pose recognition systems. These related works show that pose recognition is an essential first step towards technology-based training and documentation of ballet poses. Ballet-specific research and work in related domains are presented in the subsections that follow.

2.1   Wearable Technology and Other Related Domains

One category of capturing sensors that can be leveraged to collect data on dancers is known as wearable technology. Research by Gupta et al. involves the instruction of beginner adult ballet students through a wearable full-body garment [14]. The garment would allow for the clarification of core movements demonstrated by a teacher wearing the garment which lights up the essential limbs being used. The establishment of the basics and only later moving into the core details of motions is a well-known and effective method to coach ballet dancers regardless of their experience level. Despite the advantages, wearable technologies in ballet have potential restrictions as these garments are typically costly and often require special technical construction [14]. A more practical approach to assist with training would be to make use of vision-based pose data.

A recent relevant vision-based system in the domain of sports was aimed at the creation of a posture analysis tool for basketball free-throw shooting. This basketball free-throw study indicates that it is possible to accurately predict whether or not a basketball throw will be successful based on OpenPose skeleton key-point data as well as correct beginner players [15]. The study demonstrates that it is feasible to make use of the OpenPose [16] library in settings that deal with different body postures.

2.2   Choreography Systems

Dancs et al. investigated the choreographic side of ballet by studying the concept of having a technological tool to help choreographers. The proposed system had the aim to recognise and record a choreographer's movements automatically [17].

Computer vision algorithms proved to be useful for such a system where the consecutive movements of a dancer were to be recorded. The study had the goal to enhance the area of ballet using a Microsoft Kinect to find the joints of the body. The classification algorithms used by the study included, amongst others, the Nearest Neighbor (NN) as well as the Support Vector Machine (SVM) classification algorithms [17]. The approach by Dancs et al. produced promising results with the main algorithms generating accuracy results over 90%.

The limitations of the work proposed by Dancs et al. include the processing speed, which may be affected by any small changes in the system. The authors mentioned that, despite the classification model's recognition being fast, any additions to the model would have to consider how the processing speed would be affected [17]. Another disadvantage of this system is linked with the hardware limitations of the Kinect Sensor. According to a study by Hong et al., the Kinect sensor is unable to detect turnout and crossed feet positions properly [18]. Since turnout and crossed feet occur quite frequently in ballet poses, this challenge is relevant to other related research in the ballet pose recognition domain. Therefore, it may be worthwhile for researchers to explore alternative ways to extract skeleton features from Kinect-captured images without relying on the built-in Kinect skeleton tracking abilities.

2.3   Pose Recognition Systems for Ballet

A known concern in many sport or artistic disciplines that require any level of specialised skill and physical training is that it comes at the cost of being educated in the particular discipline. Saha et al. address this concern with their concept of fuzzy image matching for ballet poses by making the idea of e-learning in ballet easier [19]. The stages of the project involved skin colour segmentation and straight-line approximation in order to reduce the initial image of a dancer to a skeleton and eventually a stick figure representation.

The advantages of pose recognition implementations for ballet are multiple. These types of pose recognition systems lower the level at which ballet students need to rely on the physical presence of a teacher to train. E-learning of ballet through automated pose recognition therefore enables students to practise at any time in any suitable space [20].

One issue pointed out by Banerjee et al. in progressions of the above research was that the proposed algorithm relied on having the dancer wearing only particular colours in order for the skin colour segmentation pre-processing step to be effective [21]. A disadvantage of these systems, therefore, include that they rely heavily on the environmental constraints to be in place for pre-processing to take place effectively.

The research conducted by Saha et al. progressed over the years as different approaches were tried and tested by the researchers for ballet pose recognition. The results of three current related pose recognition studies, including those by Saha et al. can be seen in Table 1.

Table 1. Similar system results of studies conducted in 2014 and 2015 by Kyan et al. and Saha et al. [20, 22, 23]

No.	Similar system	Algorithm description	Accuracy
1	Posture Recognition in Ballet [20]	Fuzzy discrete membership function for matching	85%
2	Topomorphological Approach to Ballet Posture Recognition [23]	Matching of Radon transforms	91.35%
3	CAVE VR with MS Kinect [22]	L2 norm for similarity matching	92%

The results presented in Table 1 highlight achievements for current pose recognition systems, which indicate there is an opportunity to expand and improve on previous work. Along with the limited availability of ballet-datasets, this opportunity warrants the creation of a sufficiently sized ballet pose dataset as well as the implementation of multiple computer vision methods. These aspects of dataset creation as well as method implementations are addressed by this study and described in the following section.

3   Experiment Setup

To obtain consistent and scientific results, this study required certain constraints to be in place during the data capturing phase. The first constraint is that capturing should take place within a ballet studio with the appropriate floor surface for the safe execution of ballet poses. In the case of this study, the floor surface consists of ballet dance mats. Another environmental constraint includes that no mirrors or clutter should be present in the background of the capture space. Furthermore, the lighting conditions for capturing should be at a suitable level, not being too bright or too dim. In terms of role constraints, the study requires participants to be advanced level dancers that could execute the determined poses clearly. Another constraint involves the clothing that is necessary for a capture session. Standard black ballet attire has to be worn to maintain consistency and provide the assurance that there are no unnecessary variations or noise in the captured data.

A total of eight ballet poses are selected for the study, and a primary dataset was created with thirty real ballet dancers performing each respective pose in a studio. During data collection, the participants were instructed to perform the eight different ballet poses of varying difficulty. The poses include Demi-Plié, Second Position, Tendu, Sussous, Retiré, Développé, Arabesque and Penché. In order to capture the data for this study, a Microsoft Kinect sensor and a GoPro camera were used. These sensors allow for the collection of image, depth and video data.

The Microsoft Kinect images of the poses are all captured at a resolution of 640 by 480 pixels to build the dataset for this study. Once all the pose data of each dancer is gathered, it is necessary to arrange the dataset in such a way that it would be easy to feed the data to the relevant computer vision methods. The data is split into a training and testing set with 80% of the data used for training, and 20% used for testing. The dataset contains 7198 images in total, with an even distribution of images amongst different classes for both the training and testing sets. The gathered depth data was not used in this study due to the initial focus being on achieving ballet pose recognition through various methods on normal image data. The depth data will, therefore, be utilised in future work. A link to sample images in the dataset is available at: http://bit.ly/2vMZ1gq.

Once the data has been captured, and the necessary constraints are in place for the study, the relevant computer vision methods for different pipeline phases can be considered. The captured dataset is, therefore, the starting point for each of the model's pipeline implementations which will be presented next.

4   Model

For ballet pose recognition to take place using computer vision methods, this paper proposes the BaReCo model. This model consists of three broad pipeline categories, namely traditional, OpenPose and Artificial Neural Network pipelines. The variations for each of the broader categories are presented in this section, ultimately introducing eight individual computer vision pipelines. The methods involved in these pipelines were chosen based on their presence in related research [16,17,23] as well as their general relevance to computer vision pose-based problems. Furthermore, this study contributes by applying computer vision methods to the ballet environment that have not yet been explored by related research. Images from the captured dataset were used as input for each of the pipelines.

4.1   Traditional Pipeline

The traditional pipeline implementation for this study involves five stages, namely capturing, pre-processing, localisation, feature extraction and classification as illustrated in Fig. 1. For the pre-processing stage of this pipeline grayscaling and histogram equalisation was chosen. The histogram of oriented gradients (HOG) method was used for localisation to isolate the dancer's body in each

frame. This method was also used for the feature extraction phase to gather key features for classification. The classification stage of the pipeline introduces three options namely a Support Vector Machine (SVM), a Random Forest (RF) and a Gradient Boosted Tree (GBT).
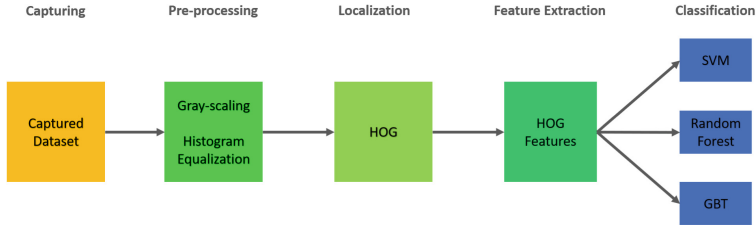
Fig. 1. Traditional pipeline architecture
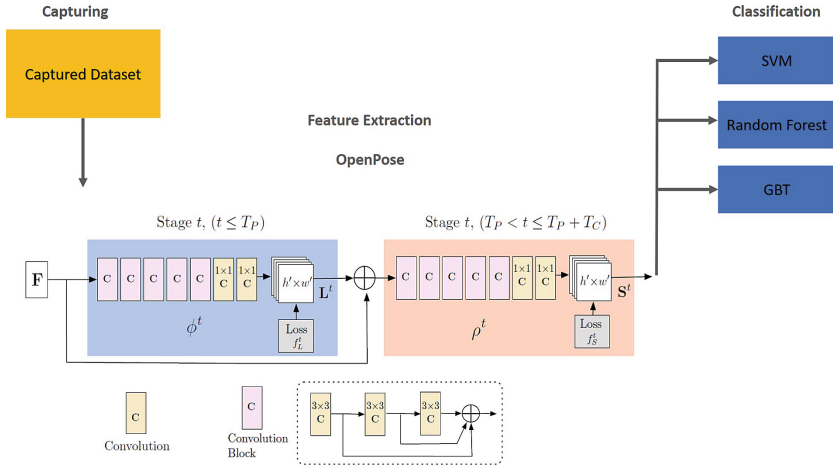
4.2   OpenPose Pipeline

The OpenPose pipeline for this study consists of three main phases, including capturing, feature extraction and classification, which is illustrated in Fig. 2. OpenPose is a recent and useful approach which uses a multi-stage Convolutional Neural Network (CNN) for extracting human skeleton key-point data from an image [16]. The classification phase of this pipeline uses the same three classifier variations that the previously described traditional pipeline used.
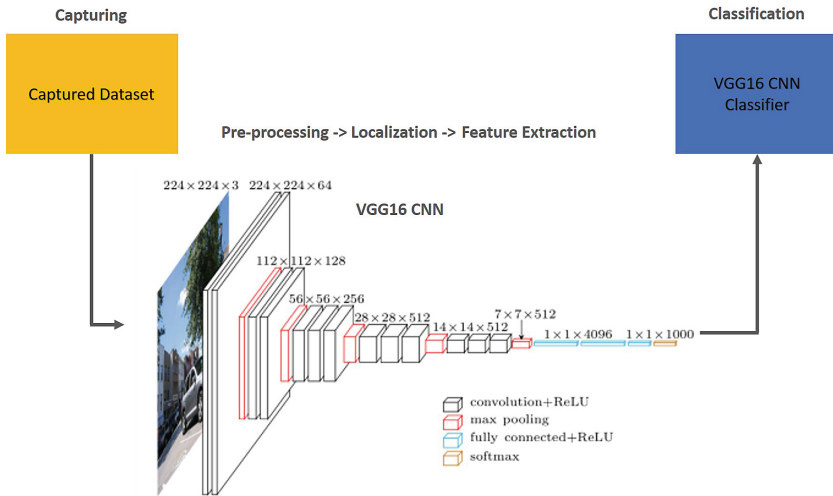
4.3   VGG16

The third category of pipelines the BaReCo model made use of is a deep learning approach which is concerned with Deep Neural Networks. The VGG16 Convolutional Neural Network (CNN) is a deep learning algorithm utilised for the model which is illustrated in Fig. 3. Since accurate results have been achieved using CNNs for various computer vision problems [24], it is suitable to utilise the approach for the implementation of this study.

4.4   Faster Region-Based Convolutional Neural Network

A family of algorithms that are extremely effective for performing object detection and localisation tasks are known as Region-Based Convolutional Neural Networks (R-CNNs). One of these algorithms is known as the Faster R-CNN algorithm which is more efficient than its predecessors [26,27] because it makes use of Region Proposal Networks (RPNs) for determining regions of interest within an image. The Faster Region-Based CNN approach is considered to be an end-to-end deep learning object detection pipeline. Accurate object detection effectively assists in visual recognition tasks, which makes the Faster R-CNN an

Fig. 2. OpenPose pipeline architecture with the multi-stage CNN architecture of Open-Pose [16]

Fig. 3. VGG16 pipeline architecture [25]

appropriate method choice for the pose recognition problem of this study. A visual illustration of the Faster R-CNN pipeline for the study is presented in Fig. 4.
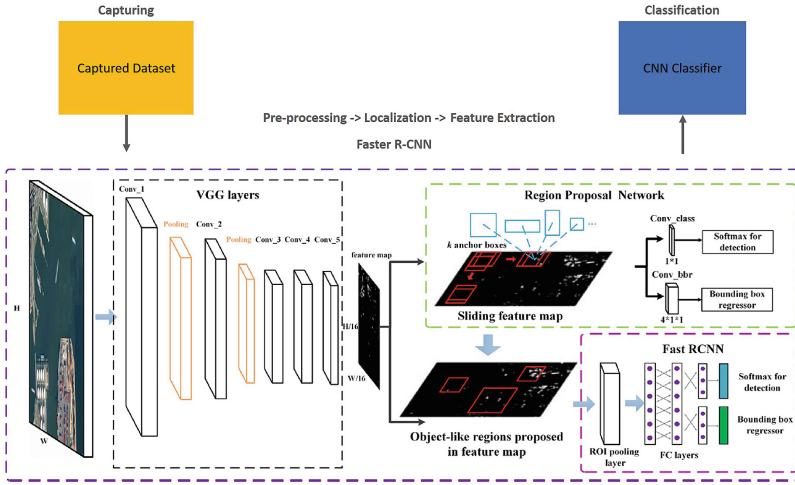
Fig. 4. Faster R-CNN pipeline architecture [27]

5   Results

Table 2 presents the results of this study and highlights the top-performing pipelines. It can be observed that the OpenPose and Random Forest Approach achieves the best results out of all the pipelines with the best scores for the considered metrics. This suggests that the OpenPose data had a positive effect on classification and that the Random Forest classifier is an appropriate model for pose recognition. Another favourable result is that of the VGG16 CNN pipeline, which has the lowest EER of 0.119% and a high accuracy score of 98.472%.

Table 2. Summary of the results obtained by this study as percentages.

Pipeline variation	Accuracy	EER	Recall	Precision	F1	Specificity
OpenPose+RF	**99.375**	0.268	**99.375**	**99.399**	**99.377**	**99.911**
VGG16 CNN	98.472	**0.119**	98.472	98.639	98.466	99.782
OpenPose+GBT	99.305	1.261	99.306	99.320	99.307	99.901
Traditional+HOG+SVM	99.199	0.139	99.178	99.016	99.096	99.887
OpenPose+SVM	99.097	0.308	99.097	99.125	99.098	99.871
Traditional+HOG+RF	98.958	0.486	99.048	98.135	98.523	99.860
Traditional+HOG+GBT	98.379	0.757	98.254	97.781	97.991	99.776
Faster R-CNN	96.049	0.237	95.381	96.577	95.561	99.433

The metrics gathered for the pipelines of this study were promising with accuracy scores above 95% for all the classifiers, which indicates that correct classifications were made for the majority of the involved poses. One interesting

observation that can be seen in the confusion matrices in Fig. 5 is that a common pose misclassification that occurred was the incorrect prediction of a Tendu as a Développé as well as an Arabesque as a Penché. A reason for this may be due to the fact that these poses are closely related to similar body orientations and arm lines. The Receiver Operating Characteristic (ROC) curves of three implemented pipelines are presented in Fig. 6 which also indicate that recognition of the Développé and Tendu poses were in some cases not as accurate as other poses. However, the ROC curves generally show that the classifiers performed well in recognising the relevant ballet poses.
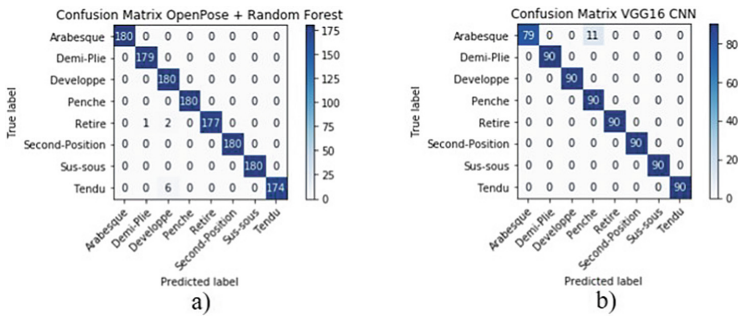

Fig. 5. Confusion matrices of the OpenPose + Random Forest and VGG16 pipelines
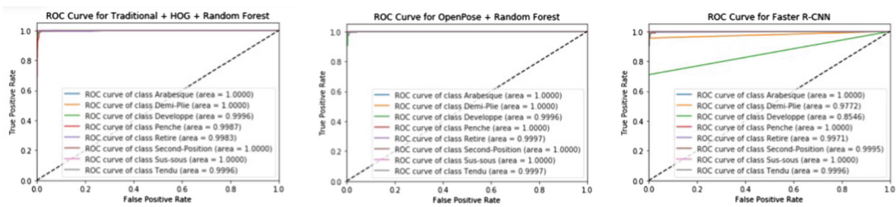

Fig. 6. The Receiver Operating Characteristic (ROC) curves of three implemented pipelines

6   Conclusion and Future Work

This study has validated the use of computer vision in the ballet domain to achieve pose recognition successfully. Some key findings of the study indicate that closely related poses are the potential cause for errors in recognition. The results also reveal that the use of novel deep learning techniques such as OpenPose and artificial neural networks, along with traditional classification approaches, yield promising results which could drive automation forward in the ballet industry.

The most accurate pipeline of this study made use of OpenPose key-point data. This study has, therefore, successfully utilised the OpenPose approach for the domain of ballet poses. The successful results of the OpenPose pipelines in

this study indicate that the combination of deep learning for feature extraction with traditional classifiers is a feasible approach for ballet pose recognition. The results attained by this study is largely due to the quality of the created dataset, which contributes to research in the ballet pose recognition domain as no similar dataset is openly available.

Future work for the study includes making improvements to the current implementation, as well as the application of other computer vision approaches on image as well as video data. Future research may, therefore, expand from static pose recognition to movement-based recognition and tracking. For ballet training there is also value in building upon the work of this study by looking at deviations from the standard ballet poses in order to correct dancer poses. Further computer vision approaches that may be of interest in future work include different Convolutional Neural Network architectures, N-shot learning as well as Recurrent Neural Networks.

The developed solution ultimately has the potential to affect the realm of ballet training and choreography in a technologically focused world. In a broader sense, this study indicates that exciting explorations can be made in artistic industries as their potential is found within the field of computer vision.

References

1. Nishani, E., Çiço, B.: Computer vision approaches based on deep learning and neural networks: deep neural networks for video analysis of human pose estimation. In: 2017 6th Mediterranean Conference on Embedded Computing (MECO), pp. 1–4. IEEE (2017)
2. Yao, B., Hagras, H., Alhaddad, M.J., Alghazzawi, D.: A fuzzy logic-based system for the automation of human behavior recognition using machine vision in intelligent environments. Soft Comput. **19**(2), 499–506 (2015). https://doi.org/10.1007/s00500-014-1270-4
3. Kale, G.V., Patil, V.H.: A study of vision based human motion recognition and analysis. Int. J. Ambient Comput. Intell. (IJACI) **7**(2), 75–92 (2016)
4. Di Orio, L.: Ballet: Method to Method. Dance Informa American Edition (2013)
5. New York Film Academy: Ballet and modern dance: using ballet as the basis for other dance techniques (2014)
6. Clay, A., Domenger, G., Conan, J., Domenger, A., Couture, N.: Integrating augmented reality to enhance expression, interaction & collaboration in live performances: a ballet dance case study. In: 2014 IEEE International Symposium on Mixed and Augmented Reality-Media, Art, Social Science, Humanities and Design, ISMAR-MASH'D, pp. 21–29. IEEE (2014)
7. Royal Academy of Dancing: The Foundations of Classical Ballet Technique. Royal Academy of Dancing (1997)
8. Kassing, G., Jay, D.M.: Beginning Ballet Technique. Human Kinetics (1998)
9. Trajkova, M., Cafaro, F.: E-ballet: designing for remote ballet learning. In: Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing Adjunct - UbiComp 2016, pp. 213–216 (2016)

10. Dance Spirit: Working one-on-one: what to expect from private lessons (2014)
11. Grieg, V.: Inside Ballet Technique. Dance Books (1994)
12. Speck, S., Cisneros, E.: Ballet for Dummies. Wiley, Hoboken (2003)
13. Snyder, A.F.: Securing our dance heritage: issues in the documentation and preservation of danceatio (1999)
14. Gupta, M., Hallam, J., Keen, E., Lee, C., McKenna, A.: Ballet hero: building a garment for memetic embodiment in dance learning. In: Proceedings of the 2014 ACM International Symposium on Wearable Computers Adjunct Program - ISWC 2014 Adjunct, pp. 49–54 (2014)
15. Nakai, M., Tsunoda, Y., Hayashi, H., Murakoshi, H.: Prediction of basketball free throw shooting by OpenPose. In: Kojima, K., Sakamoto, M., Mineshima, K., Satoh, K. (eds.) JSAI-isAI 2018. LNCS (LNAI), vol. 11717, pp. 435–446. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-31605-1_31
16. Cao, Z., Hidalgo, G., Simon, T., Wei, S.-E., Sheikh, Y.: OpenPose: real-time multi-person 2D pose estimation using part affinity fields. arXiv preprint arXiv:1812.08008 (2018)
17. Dancs, J., Sivalingam, R., Somasundaram, G., Morellas, V., Papanikolopoulos, N.: Recognition of ballet micro-movements for use in choreography. In: 2013 IEEE/RSJ International Conference on Intelligent Robots and Systems, pp. 1162–1167 (2013)
18. Hong, G.-S., Park, S.-W., Park, S.-H., Nasridinov, A., Park, Y.-H.: A ballet posture education using IT techniques: a comparative study. In: Proceedings of the Sixth International Conference on Emerging Databases: Technologies, Applications, and Theory, pp. 114–116. ACM (2016)
19. Saha, S., Banerjee, A., Basu, S., Konar, A., Nagar, A.K.: Fuzzy image matching for posture recognition in ballet dance. In: 2013 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE) (2013)
20. Saha, S., Konar, A., Janarthanan, R.: Posture recognition in ballet dance a case study on fuzzy uniform discrete membership function. In: Proceedings of the 2014 International Conference on Control, Instrumentation, Energy and Communication (CIEC), pp. 708–711 (2014)
21. Banerjee, A., Saha, S., Basu, S., Konar, A., Janarthanan, R.: A novel approach to posture recognition of ballet dance. In: 2014 IEEE International Conference on Electronics, Computing and Communication Technologies (CONECCT) (2014)
22. Kyan, M., et al.: An approach to ballet dance training through ms kinect and visualization in a cave virtual reality environment. ACM Trans. Intell. Syst. Technol. (TIST) 6(2), 23 (2015)
23. Saha, S., Konar, A.: Topomorphological approach to automatic posture recognition in ballet dance. IET Image Process. 9(11), 1002–1011 (2015)
24. Géron, A.: Hands-On Machine Learning with Scikit-Learn and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems. O'Reilly Media Inc., Newton (2017)
25. Frossard, D.: VGG in TensorFlow (2016)
26. Ren, S., He, K., Girshick, R., Sun, J.: Faster R-CNN: towards real-time object detection with region proposal networks. In: Advances in Neural Information Processing Systems, pp. 91–99 (2015)
27. Deng, Z., Sun, H., Zhou, S., Zhao, J., Lei, L., Zou, H.: Multi-scale object detection in remote sensing imagery with convolutional neural networks. ISPRS J. Photogram. Remote Sens. 145, 3–22 (2018)