Virtual Reality Annotator: A Tool to Annotate Dancers in a Virtual Environment

Claudia Ribeiro(✉), Rafael Kuffner, and Carla Fernandes

Universidade NOVA de Lisboa, Faculdade de Ciências Sociais e Humanas,
BlackBox Project, Lisbon, Portugal
`claudia.ribeiro@fcsh.unl.pt`

Abstract. In this paper we describe the Virtual Reality Annotator, a visualization tool that allows users to dynamically annotate dancers in a virtual reality environment. The current annotation types supported are sketches, speech-to-text, and highlighting. Each type of annotation can be applied to a bone or a set of bones of the skeleton data or a set of points of the point-cloud data. Using a wireless mouse, users can interact with the 3D objects of the virtual reality space, specifically a 3D menu that allows to choose the type of annotation and desired color as well as the skeleton and point cloud data. Examples of usage of this system using data from expert dancers are presented and discussed as well as suggestions for future work.

Keywords: Virtual reality · Annotations · Contemporary dance
Visualization · Interaction

1 Introduction

After Merce Cunningham has pioneered the use of the LifeForms software in the late 1980s, choreographer William Forsythe was probably one of the strongest enthusiast of the idea of exploring the huge potentiality of computation and design to enhance the transmission, learning and creative processes of contemporary dance works. His first project connecting contemporary dance to digital media technology was developed in the 1990s in a CD-ROM (which has quickly become rather emblematic until today) with the title "Improvisation Technologies", and it clearly established a trendy basis for the use of computational design to communicate ideas behind dance composition. This project was the inspiration for several other contemporary choreographers and dancers

(a.o. Wayne MacGregor, Emio Greco PC (DS/DM Project)[1], Siobhan Davies (RePlay archive)[2], Rui Horta (TKB project)[3] and João Fiadeiro (BlackBox project)[4] towards the documentation and "archiving" of dance compositional methodologies, on the one hand, and towards the progress of more innovative Arts and Science studies leading to a fresher vision over the importance of deeply analyzing intangible and ephemeral art forms such as contemporary dance.

Performing arts such as dance were traditionally taught either by example or by following conventional scores on paper. In the specific field of contemporary dance, with different body movements emerging and the impossibility of creating a controlled vocabulary of movements in order to compose a score, watching videos of previous performances or of rehearsals is often the most effective way to learn a specific choreography. A common video, though, is not sufficient to communicate what is envisioned by the choreographer [9].

Video annotation systems have been used in this field to shorten the knowledge gap between choreographers and dancers, especially in the transmission process of very detailed bodily information. Nevertheless, current video annotators support a limited set of 2D annotation types, such as text, audio, marks, hyperlinks, and pen annotations. Some of them offer animation functionalities, which are only applied to classical choreographies, including ballet, for which numerous notation systems already exist, allowing to represent a wide spectrum of movement complexes [14]. This is not the case for contemporary dance, where movements are unpredictable and can change with every execution, either during rehearsals or live performances.

Previous works [16] have further developed this field by capturing dance performances in 3D using depth cameras, and extending the 2D annotations to this space. By doing so, new types of contextualized annotations are introduced, where information can be attached to individual performers and accompanying them during the time where the annotation is visible. However, annotating in a 2D environment and transposing the data to 3D presents several limitations. The mapping between views is not a straightforward task, and the 2D input during annotation limits the possibilities for other type of annotations that can make use of 3D annotations.

To overcome the limitations of transposing 2D annotations to a 3D environment, we have developed the Virtual Annotator that allows users to annotate specific body parts and movement sequences in a three-dimensional virtual reality space. The system was implemented in Unity3D[5] integrated with the Oculus Rift V2 development kit, and the interaction with the point cloud and skeleton data is provided by a wireless mouse.

Section 2 describes background work related to video annotators and their uses in the context of dance. Section 3 presents the Virtual Reality Annotator where is described data capture and visualization, the software architecture and interaction. Section 4 describes and discusses the obtained results. We finalize with conclusions and future work (Sect. 5).

2   Related Work

Several video annotation systems have been proposed to shorten the knowledge gap between the choreographer and dancers. The "Creation-Tool" [3] (labelled as DancePro in the framework of the recently concluded EuropeanaSpace project), the "Choreographer's notebook" prototype [18], Dance Designer, ReEnact, Danceforms, and more recently Motion Bank's Piecemaker2GO, are relevant examples, all of them serving quite specific purposes. DancePro is to our knowledge the most efficient video annotator to assist choreographers directly while rehearsing in situ, as it allows taking annotations over the captured videos in real-time.

Both non-specific systems [11,23] and specific ones [6,7,10,14,20] have been used with varying features or limitations. Dance targeted software will typically allow a choreographer to choose from a series of poses to create scores and visualize it in different manners, which can be limitative in the expressiveness of the movements.

The sketch based system from Moghaddam et al. [14] allows the user to compose a digital choreography by individually sketching each individual dance pose, which overcomes the previous limitation. These sketches are used to estimate poses that will be applied to a 3D avatar. However, this system does not include annotations or scoring as the other dance-specific software.

These systems have either the limitation of having a single viewpoint where one annotates, or a limited subset of movements to be used, or lack of expressiveness. Previous research has tackled this problem [16] by extending a 2D annotator to a 3D environment. This was performed by translating the 2D annotations on a video to a reconstructed point cloud in the 3D environment. However, the expressiveness of the 3D annotations was compromised by the translation process, since users do not directly annotate in the 3D environment.

Wearable technology and motion tracking have developed to the point where Virtual Reality (VR) and Augmented Reality (AR) are usable in a dance context with minimal disruption to its traditional practice. The article from Gould [8] discusses "AR art and performance", and how the mixture with technology creates a different type of art, putting the *"body at the heart of its becoming"*. The displayed content can now depend heavily on the perspective, gestures, and positions of the body of the one whos visualizing the work. One given example is the "Eyebeam" project where a dancer is partnered with an AR avatar which can only be seen through a device. A similar goal is shared by the WhoLoDance [4], which already uses head mounted display technology (HMD).

Both approaches show the importance of embodied systems and presence [17] in the context of dance. This has been used as a different approach for

teaching dance. Kyan et al. [12] developed a system to train ballet poses in a cave environment, similarly to previous work from Chan et al. [5]. However, by not displaying the virtual world through an HMD, the sense of presence and body ownership is considerably lower.

Annotating through an HMD has been mainly targeted at AR scenarios, where real world problems can be observed and tagged for later inspection. The survey paper from [22] reviews these types of annotations in great depth. Virtual reality has not been thoroughly used for video annotation, due to the fact that free-viewpoint videos and point cloud based data that register real world events still not being commonplace. Different techniques have been proposed to annotate static point clouds in an non immersive scenario [2,21], but have limitations when translating them to HMD, where one cannot resort to using hand-held devices with an auxiliary screen, or other peripherals such as keyboard for input. Static inspection [1] and annotation of point clouds [13] has been done using VR, with a focus on architectural problems and rich environments. Using the advantages of embodied experiences in dance to annotate captured point cloud videos through a HMD is a problem that has yet to be addressed.

3   Virtual Reality Annotator

3.1   Data Capture and Point Cloud Visualization

A wide-baseline setup was used in the present study to capture point cloud and skeleton data, where each view was captured by a Kinect sensor. Kinect sensors positioned triangularly about two meters apart to optimize the capturing of point cloud and skeleton data and at the same time allowing dancers enough space to dance.

Regarding data synchronization, a network-based synchronization program was developed allowing triggering the capture remotely and simultaneously on each computer. The calibration of extrinsic and intrinsic parameters was performed using OpenCV and manual inputs from the developers since the process was performed in a controlled scenario.

Point clouds are generated using depth information, and all the streams were integrated in a single point cloud based on the calibration data (position and rotation) of each viewpoint. The amount of data produced capturing at 30 fps supports brief stretches of each performance to be viewed. We used Unity3D as a platform for rendering the recorded datasets, where the user could freely navigate the camera around the performance scene.

3.2   Architecture

The Virtual Reality Annotator is a network-based application (see Fig. 1) that integrates several software modules that support tracking, creating and managing annotations, point cloud and skeleton visualization and finally an interaction module that processes user inputs.
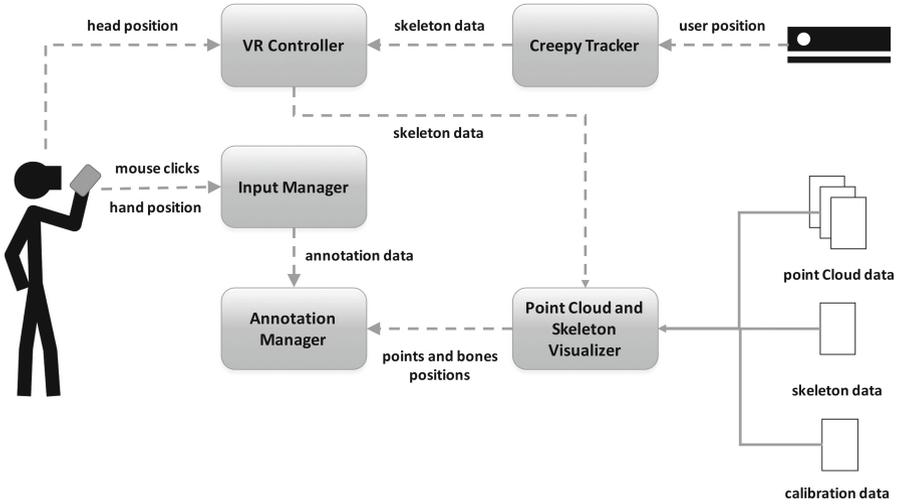
Fig. 1. Virtual Reality Annotator modular architect diagram

Users have an abstract body representation created by the skeleton provided by the "Creepy Tracker" module [19], since their real body is covered by the HMD. This system works by combining skeleton information from various Kinect cameras, allowing users to move freely in the capture space. Skeleton data is received by the application through the network, with users identifying themselves at the startup of the application by raising their hand.

The annotation manager module is an aggregated class containing a set of modules responsible for storing and managing annotation types in the Virtual Reality environment. Each annotation class has an associated mono behavior responsible for processing user input and managing the 3D objects related to the annotation. In this manner extending the current software with new annotation types is straightforward and involves minimal changes to the code.

The position where the annotation should be created in the virtual world is provided by the VR Controller module, which receives and processes the skeleton data given by the Creepy Tracker and the head orientation given by the Oculus Rift. This data is also used to update the user skeleton data in the VR environment.

The Input Manager receives the mouse clicks (see Fig. 2) and hand position. Based on this data, it toggles the mono behavior of the annotation type that is currently active. Moreover, it is responsible for visualizing and hiding the 3D menu and drawing the contextual heads-up display attached to the users' hand.

To interact with the menu, a raycast is drawn starting from the users' hand position and, when a collision is detected, the appropriate method is executed enabling or disabling the appropriate menu option. The 3D menu has five options: highlight points, speech-to-text, 3D drawing, change color, and delete (see Fig. 2).
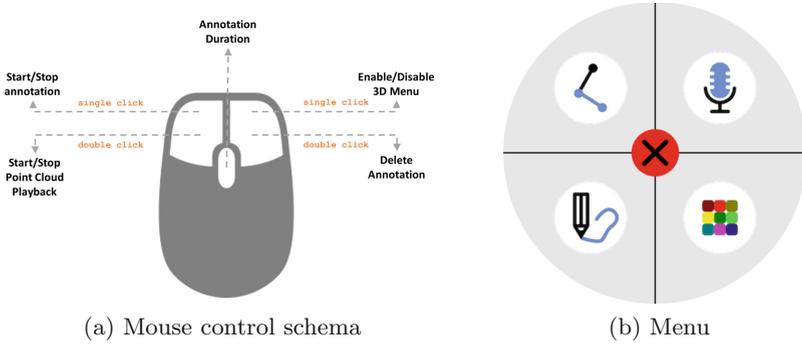
(a) Mouse control schema   (b) Menu

Fig. 2. 2 Figures side by side

3.3   User Interaction Paradigm

Interaction with the system is performed by free navigation in the environment, and mostly based on the position of the users' dominant hand. Figure 2(a) displays the input commands of the wireless mouse. Menu interaction is performed by right clicking to open a menu in the looking direction, and selection of the current function by pointing and left clicking.

Color selection (Fig. 4b) is performed by pointing at the desired color on a RGB spectrum and left clicking. Annotations are created by holding the left mouse button and performing the desired action. For 3D drawing (Fig. 3a) the hand position is used as a virtual brush. For cloud highlighting (Fig. 3b) the same metaphor is used, except the user chooses paints the desired points instead of a general area. Finally, text annotations are created at the users' hand location.
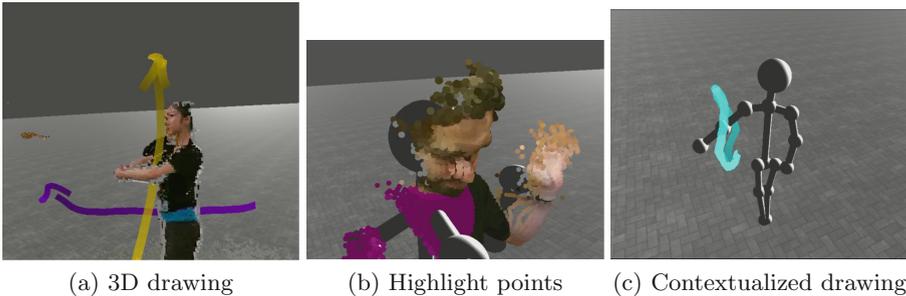

(a) 3D drawing   (b) Highlight points   (c) Contextualized drawing

Fig. 3. Implemented annotations/functions in our system

The duration of an annotation can be adjusted by placing the dominant hand near the center of the annotation, clicking and scrolling the mouse wheel until the desired duration is reached. This is then confirmed via a second wheel click. An annotation can be deleted by a double right-click near the desired annotation.
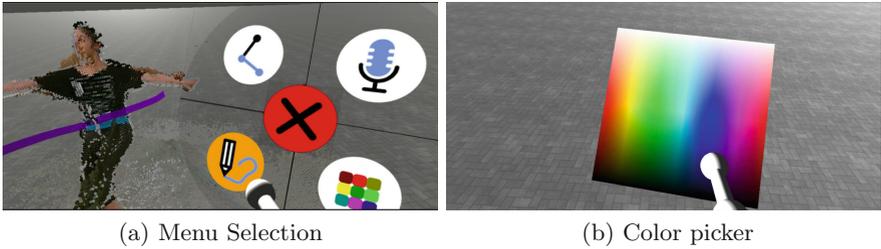
(a) Menu Selection   (b) Color picker

Fig. 4. Examples of interaction

4   Results and Discussion

The implemented system allows users to visualize and annotate temporal point-cloud data that can be associated with skeletal information. This is particularly important in the case of highlighting 3D points of a time-sequenced point cloud.

Given that there is no temporal relation between 3D points in different frames, we associate the annotations to the skeleton joints, which are updated every frame and passed to the shader that will affect the points in the vicinity of that joint. This is optional for 3D drawing and text annotations.

Some uses for this are to attach a text annotation, like a name of the person, to follow a certain subject in the video, or to create a note related to a series of actions. The same applies to drawing, which might be related to a certain body part, or simply markings on the floor. If annotations are created closer than a certain threshold to a specific skeleton joint in the frame, they are associated to that body part.

This is one of the advantages of the proposed system over previous attempts, where we are able to have both static and dynamic annotations, contextual or not, and affecting unstructured data (point cloud). Moreover, such 3D contextual annotations are only possible in the immersive VR environment, due to the inexistent depth component in more traditional inputs.

As opposed to more complex input solutions, our system has the advantage of using known metaphors for users (painting with your hand, pointing at something you want). Also, such an embodied solution allows users to more accurately perceive and interact with space such as it is perceived in a dance studio. Allowing one to freely navigate around the data overcomes the stated limitations of video-based teaching approaches [9].

The described system has been applied to both documenting and archiving dance choreographers' work [15] in the context of contemporary dance.

Instead of adding annotations targeted at teaching, scholars could highlight characteristic movements or tendencies of a certain choreographer. This is a crucial application for enhancing digital cultural heritage archives, where common videos combined with text-based descriptions are not able to efficiently represent certain temporal and spatial aspects embedded in an immediate context of the documented works. An added benefit of this type of annotated data is that it

can be used to develop comparative studies on different genres, styles, or periods of time in the work of choreographers and dancers.

Some of our current limitations are shared by other VR-based systems. Some users may experience motion sickness in prolonged interactions with the system, due to the fact that the markerless tracking solution applied is subject to some noise and delay as mentioned in [19]. Also, the tracked space by current VR systems is limited, which is a problem if the captured area is larger than the VR tracked space, in that it restricts the users possibility to reach the position where they want to annotate. For a more detailed view of the interaction, please see the video attached to this publication.

5   Conclusion and Future Work

In this paper we described the Virtual Reality Annotator, which is a software tool that allows a user to visualize and annotate dynamically point cloud and skeleton data. Currently, is supports three types of annotations, namely highlight points, speech-to-text and 3D Drawing. Our tool supports dynamic annotations which is an improvement on previous existing systems.

One future development in the area of digital dance teaching is allowing students to overlay themselves with annotated data, which could also be visualized as a mirror. A scoring mechanism could be defined and used by instructors to help perfecting certain captured movements. Specific types of annotations for this matter would be created, so an instructor could create a virtual class that a student user could follow. Finally, using the perceived advantages of embodied VR applications in the context of dance, novel applications can be developed in the same direction to support other aspects of this domain, such as digital composing, drafting a dance piece, or editing existing data to better express an idea or concept underlying a specific movement.

References

1. Addison, A.C., Gaiani, M.: Virtualized architectural heritage: new tools and techniques. IEEE MultiMedia **7**(2), 26–31 (2000)
2. Bacim, F., Nabiyouni, M., Bowman, D.A.: Slice-n-swipe: a free-hand gesture user interface for 3D point cloud annotation. In: 2014 IEEE Symposium on 3D User Interfaces (3DUI), pp. 185–186. IEEE (2014)
3. Cabral, D., Valente, J.G., Aragão, U., Fernandes, C., Correia, N.: Evaluation of a multimodal video annotator for contemporary dance. In: Proceedings of the International Working Conference on Advanced Visual Interfaces, AVI 2012, pp. 572–579. ACM, New York (2012)

4. Camurri, A., El Raheb, K., Even-Zohar, O., Ioannidis, Y., Markatzi, A., Matos, J.-M., Morley-Fletcher, E., Palacio, P., Romero, M., Sarti, A., Di Pietro, S., Viro, V., Whatley, S.: WhoLoDancE: towards a methodology for selecting motion capture data across different dance learning practice. In: Proceedings of the 3rd International Symposium on Movement and Computing, MOCO 2016, pp. 43:1–43:2. ACM, New York (2016)

5. Chan, J.C.P., Leung, H., Tang, J.K.T., Komura, T.: A virtual reality dance training system using motion capture technology. IEEE Trans. Learn. Technol. **4**(2), 187–195 (2011)

6. ChoreoPro. Dance designer (2014). https://www.choreopro.com

7. Credo Interactive. Danceforms (2014). http://charactermotion.com/products/danceforms/

8. Gould, A.S.: Invisible visualities: augmented reality art and the contemporary media ecology. Convergence **20**(1), 25–32 (2014)

9. Guest, A.H.: Dance Notation: The Process of Recording Movement on Paper. Dance Horizons, New York (1984)

10. James, S., Fonseca, M.J., Collomosse, J.: ReEnact: sketch based choreographic design from archival dance footage. In: Proceedings of International Conference on Multimedia Retrieval, ICMR 2014, pp. 313:313–313:320. ACM, New York (2014)

11. Kipp, M.: ANVIL: the video annotation research tool. In: Handbook of Corpus Phonology. Oxford University Press, Oxford (2010)

12. 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. (TIST) **6**(2), 23 (2015)

13. Lubos, P., Beimler, R., Lammers, M., Steinicke, F.: Touching the cloud: bimanual annotation of immersive point clouds. In: 2014 IEEE Symposium on 3D User Interfaces (3DUI), pp. 191–192, March 2014

14. Moghaddam, E.R., Sadeghi, J., Samavati, F.F.: Sketch-based dance choreography. In: 2014 International Conference on Cyberworlds (CW), pp. 253–260, October 2014

15. Ribeiro, C., dos Anjos, R.K., Fernandes, C.: Capturing and documenting creative processes in contemporary dance. In: Proceedings of the 4th International Conference on Movement Computing, MOCO 2017, pp. 7:1–7:7. ACM, New York (2017)

16. Ribeiro, C., Kuffner, R., Fernandes, C., Pereira, J.: 3D annotation in contemporary dance: enhancing the creation-tool video annotator. In: Proceedings of the 3rd International Symposium on Movement and Computing, MOCO 2016, pp. 41:1–41:4. ACM, New York (2016)

17. Sanchez-Vives, M.V., Slater, M.: From presence to consciousness through virtual reality. Nat. Rev. Neurosci. **6**(4), 332–339 (2005)

18. Singh, V., Latulipe, C., Carroll, E., Lottridge, D.: The choreographer's notebook: a video annotation system for dancers and choreographers. In: Proceedings of the 8th ACM Conference on Creativity and Cognition, C&C 2011, pp. 197–206. ACM, New York (2011)

19. Sousa, M., Mendes, D., Dos Anjos, R.K., Medeiros, D., Ferreira, A., Raposo, A., Pereira, J.M., Jorge, J.: Creepy tracker toolkit for context-aware interfaces. In: Proceedings of the 2017 ACM on Interactive Surfaces and Spaces, ISS 2017. ACM, New York (2017)

20. Umino, B., Soga, A., Hirayama, M.: Feasibility study for contemporary dance e-learning: an interactive creation support system using 3D motion data. In: 2014 International Conference on Cyberworlds (CW), pp. 71–76. IEEE (2014)

21. Veit, M., Capobianco, A.: Go'then'tag: a 3-D point cloud annotation technique. In: 2014 IEEE Symposium on 3D User Interfaces (3DUI), pp. 193–194. IEEE (2014)
22. Wither, J., DiVerdi, S., Höllerer, T.: Annotation in outdoor augmented reality. Comput. Graph. **33**(6), 679–689 (2009)
23. Wittenburg, P., Brugman, H., Russel, A., Klassmann, A., Sloetjes, H.: ELAN: a professional framework for multimodality research. In: Proceedings of LREC, vol. 2006, p. 5 (2006)