3D Annotation in Contemporary Dance: Enhancing the Creation-Tool Video Annotator

Claudia Ribeiro
CLUNL, Faculdade de
Ciências Sociais e Humanas,
FCSH, University of Lisbon
Lisbon, Portugal
claudia.ribeiro@fcsh.unl.pt

Rafael Kuffner
CLUNL, Faculdade de
Ciências Sociais e Humanas,
FCSH, University of Lisbon
Lisbon, Portugal
rafael.kuffner@fcsh.unl.pt

Carla Fernandes
CLUNL, Faculdade de
Ciências Sociais e Humanas,
FCSH, University of Lisbon
Lisbon, Portugal
carla.fernandes@fcsh.unl.pt

João Pereira
INESC-ID, Instituto Superior
Técnico, University of Lisbon
Lisbon, Portugal
jap@inesc-id.pt

ABSTRACT
Annotated videos have been used in the context of dance performance not only as a way to record and share compositions and knowledge between different choreographers, but also as a powerful learning tool. Restraining the viewpoint of the user to the recorded point of view can be an obstacle in several scenarios. Alternatives that introduce the concept of a three-dimensional space have been developed, but coming short either on the freedom of concepts that the user is able to introduce, or on resorting to a non-natural representation. This article describes a follow-up work on the previously developed Creation-Tool [2] extending the existing functionality to tackle this problem. The developed system places the 2D annotations onto a three-dimensional point cloud, captured by depth sensors coupled with cameras around the performance, thus enabling the user to freely visualize the annotated performance three-dimensionally at an arbitrary point of view.

Author Keywords
Annotations; Video Annotations; Contemporary Dance; 3D visualization; Image Segmentation;

ACM Classification Keywords
H.5. Information Interfaces and Presentation: H.5.1Multimedia Information Systems-*Video*; H.5.2User Interfaces-*Prototyping*;

MOCO'16, July 05-06, 2016, Thessaloniki, GA, Greece

DOI: http://dx.doi.org/10.1145/2948910.2948961

INTRODUCTION
The global performing arts community is urging the need for innovative systems which:

(a) document, transmit and preserve the unexplored knowledge contained in performance composition processes, (b) assist artists with tools to facilitate their choreographic or dramaturgic practices, preferably on a collaborative basis. At present, existing digital archives for performing arts mostly function as linear e-libraries, not allowing higher degrees of interactivity or active user intervention. One of the reasons for this limitation is related to the reduced abilities offered by available video annotators. Specifically, current video annotators support a very 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 Ballet choreographies where numerous notation systems exist that allow to represent a very high number of movements and combinations of movements [8]. This is not the case for contemporary dance, where the movement is unpredictable and can change with every execution, either during rehearsals or in live performances.

Previously to the BlackBox project, the Creation-Tool video annotator has been developed which can be used by choreographers to analyse and improve their work, by recording and annotating a rehearsal or a live performance, for later review or for sharing his/her notes with the performers [2]. Although this tool has provided significant advance, it still presents the limitations previously described.

The BlackBox project endeavours to fill this gap and create a new paradigm for the documentation of performance composition by augmenting a 2D video annotator with 3D visualizations of the resulting annotations. In this paper a system is described that couples a 2D multi-view performance captured with video and a 3D multi-view captured with the Kinect sensor, translating the 2D annotations taken on the Creation-Tool to the captured point clouds by using feature matching and image segmentation. These annotations can be visualized later on in a moving point cloud using an arbitrary viewpoint. Therefore, the contributions of this paper are the following: (a) a novel system that combines Computer Vision and Visualization techniques to enhance video annotation with a 3D representation; and (b) its application to a live contemporary dance improvisation, demonstrating the possibilities of exploring this novel form as a means to document and preserve the implicit knowledge contained in performance composition processes.

AUGMENTING THE CREATION-TOOL WITH 3D ANNOTATIONS

Traditionally, performing arts such as dance are taught either by example or by looking at conventional scores on paper. With different dance 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 way to learn a specific dance. A common video, though, is not sufficient to communicate what is envisioned by the choreographer [6].

Video annotation systems have been used in this field in order to shorten the knowledge gap between the choreographer and dancers. Software tools such as Dance designer [3], Reenact [7], Danceforms [4], and the Sketch-based system of Moghaddam et. al [8] try different ways of tackling this problem: by introducing three-dimensional information, either by synthetic characters representing the performers [3], by a different layer of annotation with 3D information augmenting the 2D [4] or by sketch based interactions [7, 8]. The compromise made to have 3D representations of the movements is either to define a subset of movements to be displayed on a virtual avatar, or to have an unnatural visualization of the result with sketches or text annotation on synthetic representations of the performers, which strays away from the classical experience of watching a performance in order to learn dance.

Our work tries to combine the best of both approaches. We argue that the best alternative is to have a full three-dimensional reconstruction of the performance where the user can watch it as a free viewpoint video, with merged annotations from each one of the recording streams. Capturing three-dimensional data using depth sensors of a live performance and enhancing it up with annotations is our solution for contemporary dance choreographers to document and share their compositions (see Fig. 1).

We use a wide-baseline setup with each view being captured by a color camera and the depth information by a Kinect sensor, coupled closely to the used camera in order not to introduce a big stereo disparity on the annotation conversion process. Each view is positioned in opposite sides of the room in order to minimize loss of data due to occlusion. Adding more viewpoints decreases the probability of its occurrence. Initially a program was developed which captured the required information using only the Kinect. Nevertheless, initial tests showed that capturing both depth maps, color images and video, simultaneously with the Kinect, resulted in high frame

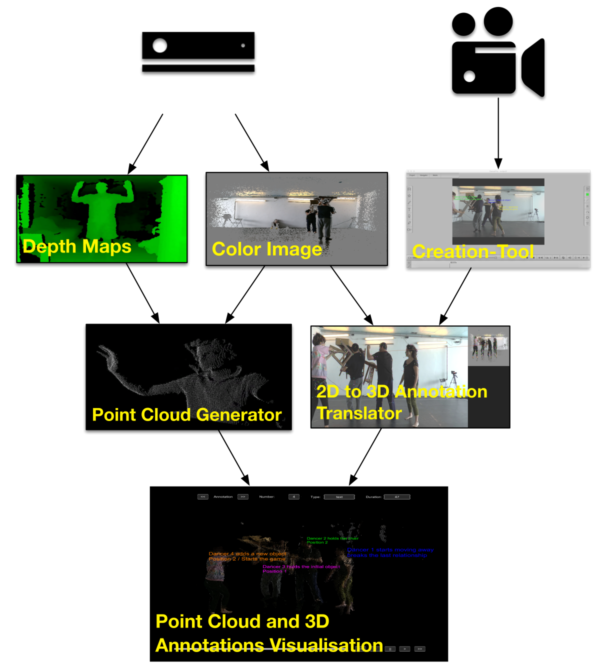

Figure 1. Schematic representation of the developped system. At the top, the data input devices: Microsoft Kinect and video camera. In the middle, two modules used to generate the point clouds (left) and convert the 2D annotations in 3D annotations (right). At the bottom, the 3D Annotation Visualizer build on top of Unity3D.

losses, which would compromise the quality of the 3D data visualization. In order to solve this limitation, a video camera was attached and aligned to each Kinect used in the final setup (see Fig. 2).

Annotations are made on each one of the 2D videos using the Creation-Tool, which encodes them on an XML file with a pixel 2D position and starting times. An external developed tool processes the XML file to assign 3D coordinates to these so that they can be visualized in the point cloud video. This process is made using SURF feature detectors [1] on the video frame and on the color frame from the Kinect sensor. The closest feature to the annotation position is selected, and its corresponding match on the Kinect frame is used for depth information in order to calculate the three-dimensional coordinates.

The low resolution and high noise frequency of the kinect streams reduces the quality of the matches to a lower level than expected. We found that performing background subtraction greatly improves the quality of the feature matching process, with the downside of losing the capability of annotating the static elements of the scene. In the context of our case-study this has not been considered as a problem, since our main focus is on annotating dancers in movement. Point clouds are generated using the depth information and all the streams are integrated in a single point cloud based on the calibration data (position and rotation) of each viewpoint. We have used Unity3D as a platform for rendering the recorded datasets, allowing the user to navigate the camera

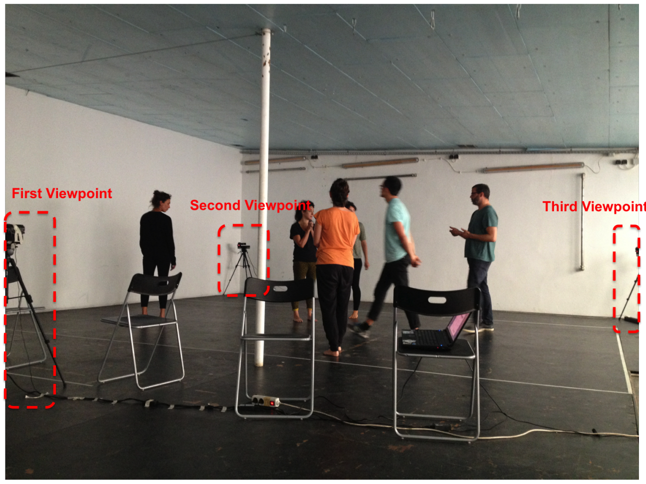

Figure 2. Representation of the setup used with three kinects and video cameras during the case-study done with João Fiadeiro and a group of his contemporary dancers.

freely around the performance, as well as to control the playback of the video segment. Annotations are displayed on the three-dimensional position that best matches the point clicked on the original video.

CASE STUDY - REAL TIME COMPOSITION

Within the context of the BlackBox project, we have been working with the Portuguese choreographer João Fiadeiro, who has done extensive work to developed a method for improvisation in contemporary dance, the Real-Time Composition (RTC) method [5]. To understand and analyse his method, two separate sessions of RTC involving João Fiadeiro and seven dancers from his company were recorded using both video cameras as well as the Microsoft Kinect.

Figure 3. First example of the Creation-Tool: each dancer has a text annotation of the *first* and *second position* of the RTC method.

Three different point of views were captured during the improvisation sessions. The space available for the improvisation was delimited on the floor with duck tape to prevent the dancers from getting out of the space captured by the cameras and the Kinects. Before each session started, the alignment between video cameras and kinect was tested. Each improvisation session lasted around three hours including breaks and feedback discussions related to what was being composed by the group during each set (improvisation sub-session).

The collected data were used to generate the corresponding point clouds which were analysed together with the videos. The main criteria used to choose the sketches for annotation was related to the richness of the information provided by the improvisation and related to the choreographer's work. The corresponding video was annotated using the Creation-Tool [2]. The main annotation types used were text and hyperlink annotations, thus enhancing some concepts underlying the RTC method (e.g Position and Relationship).

Figure 4. Result of the conversion of the 2D annotation of Dancer 2 (see Figure 3) into a 3D annotation using the SURF feature detector.

Finally the video was segmented in order to determine the 3D position of each of the annotations (see Figure 4). The segmentation process generated a XML file which contains detailed information of each of the annotations, as well as the newly calculated z coordinates. The newly generated XML file and the corresponding point clouds were then loaded into the 3D Annotation Visualizer. 5, Figure 6 and Figure 7, which represent present the same scene from the tree viewpoints captured with the video cameras and kinects.

Figure 5. Result of the 3D Annotator Visualizer for the annotations depicted in Figure 3. (viewpoint 1)

As shown in the resulting images, the 3D environment provides a richer setting to analyse and understand contemporary dance, since it allows to observe a particular scene from multiple points of view and therefore solving the problems of occlusion in 2D video. Moreover, this representation presents less ambiguity for future researchers studying a particular performance or dancer on their own, since it preserves the correct locations of the annotations. And as it allows to capture different points of view, it can potentially be extended to

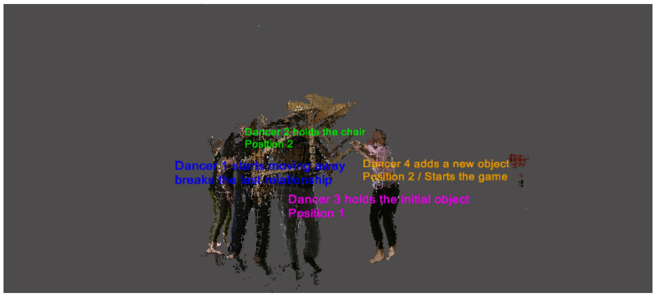

Figure 6. Result of the 3D Annotator Visualizer for the annotations depicted in Figure 3. (viewpoint 2)

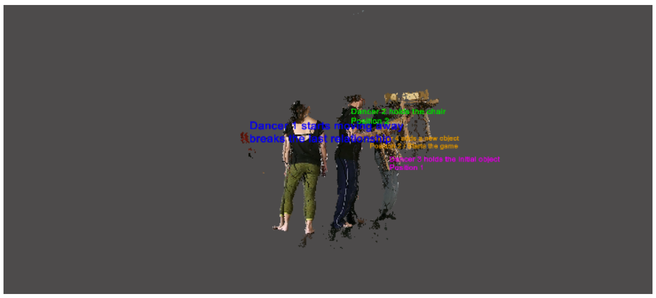

Figure 7. Result of the 3D Annotator Visualizer for the annotations depicted in Figure 3. (viewpoint 3)

support collaboration as well, either in real-time or in post-processing.

CONCLUSION AND FUTURE WORK

We have presented a case-study were video and point cloud data was captured during several improvisation sessions involving a choreographer and several dancers. This data was annotated later on and then used to visualize how text annotations would be represented in a 3D environment.

The development of this work has allowed us to more deeply understand the challenges involved when working with this kind of data (point clouds) and these types of media (video, kinects). In particular, point clouds can correctly represent the three-dimensional reality of a performance, but have clear drawbacks when it comes to visualization, such as background/foreground confusion, and loss of definition in close-up planes. Therefore, alternative visualization techniques need to be explored using this data format in order to circumvent these obstacles and provide a more realistic experience. Moreover, in order to generate point cloud data, both color images and depth data have to be used, which results in a very high volume of data. This represents a challenge, both in terms of the limits imposed by computer memory when processing this data and when archiving it.

Another challenge is related to the annotations. Specifically, not all types of 2D annotations can be easily transcribed to 3D, especially freehand sketches, which do not provide enough information regarding the objects they relate to. This hardens the decision of knowing which is the appropriate depth for them. Therefore, further research needs to be carried out in order to understand which type of annotations would make sense in a 3D environment for this type of application area. Also, annotating directly on video may not be the most appropriate way of harnessing this kind of data. A combined system, which incorporates both video and a kinect stream, could potentially solve the limitations related to the determination of the depth of the different types of annotations.

Finally, having a 3D representation of the scene can be further exploited to enhance how the annotations are related to the dancer(s) and/or to the sequence of movements of a particular scene. For example, point cloud segmentation algorithms can be used to segment the dancers into cluster and use effects such as changes in colors of particular parts of the dancers bodies, in order to visually relate annotations, dancers and movements.

ACKNOWLEDGMENTS

This work was supported by the European Research Council under the project (Ref. 336200) - ”BlackBox: A collaborative platform to document performance composition: from conceptual structures in the backstage to customizable visualizations in the front-end”. We extend our acknowledgments to João Fiadeiro and the dancers who have participated in this case-study, as well as to the team at Re.Al for their technical support and collaboration.

REFERENCES

1. Bay, H., Ess, A., Tuytelaars, T., and Gool, L. V. Speeded-up robust features (surf). *Computer Vision and Image Understanding 110*, 3 (2008), 346 – 359. Similarity Matching in Computer Vision and Multimedia.
2. Cabral, D., Valente, J. a. G., Aragão, U., Fernandes, C., and Correia, N. Evaluation of a multimodal video annotator for contemporary dance. In *Proceedings of the International Working Conference on Advanced Visual Interfaces*, AVI ’12, ACM (New York, NY, USA, 2012), 572–579.
3. ChoreoPro. Dance designer. web. `https://www.choreopro.com`, 2014.
4. Credo Interactive. Danceforms. web. `http://charactermotion.com/products/danceforms/`, 2014.
5. Fiadeiro, J. *Wissen in Bewegung. Perspektiven der künstlerischen und wissenschaftlichen Forschung im Tanz.* transcript Verlag, Bielefeld, 2007, ch. Wenn Du das nicht weißt, warum fragst Du dann?, 103–112.
6. Guest, A. H. *Dance notation: The process of recording movement on paper*. New York: Dance Horizons, 1984.
7. James, S., Fonseca, M. J., and Collomosse, J. Reenact: Sketch based choreographic design from archival dance footage. In *Proceedings of International Conference on Multimedia Retrieval*, ICMR ’14, ACM (New York, NY, USA, 2014), 313:313–313:320.
8. Moghaddam, E., Sadeghi, J., and Samavati, F. Sketch-based dance choreography. In *Cyberworlds (CW), 2014 International Conference on* (Oct 2014), 253–260.