
Multimodal Video Annotation for Contemporary Dance Creation

Diogo Cabral¹

diogocabral@di.fct.unl.pt

Urândia Carvalho¹

urandiacarvalho@gmail.com

João Silva¹

jms19199@fct.unl.pt

João Valente¹

p110356@fct.unl.pt

Carla Fernandes²

carla.fernandes@fctsh.unl.pt

Nuno Correia¹

nmc@di.fct.unl.pt

¹CITI, Departamento de Informática,
Faculdade de Ciências e Tecnologia, FCT,
Universidade Nova de Lisboa,
Quinta da Torre, 2829-516
Caparica, Portugal

²CLUNL,
Faculdade de Ciências Sociais e Humanas, FCTSH,
Universidade Nova de Lisboa,
Avenida de Berna, 26 C
1069-061 Lisboa, Portugal

Copyright is held by the author/owner(s).

CHI 2011, May 7–12, 2011, Vancouver, BC, Canada.

ACM 978-1-4503-0268-5/11/05.

Abstract

This paper presents a video annotator that supports multimodal annotation and is applied to contemporary dance as a creation tool. The prototype, developed for Tablet PCs, explores bimanual interaction, using pen and touch input interfaces. This combination can be more natural and familiar than the traditional input interfaces (keyboard or mouse). Contemporary dance is a domain where this type of flexible interaction with video material is relevant in order to augment and improve the rehearsal and creative processes. Motion tracking is used to define the dynamic behavior of the annotations and voice input complements the other modalities. The paper describes the design decisions done by the multidisciplinary development team and the current status of the tool.

Keywords

Annotations, Video Annotations, Pen-based Video Annotations, Tablet PCs, Contemporary Dance

ACM Classification Keywords

H.5. Information interfaces and presentation: H.5.1 Multimedia Information Systems-*Video*; H.5.2 User Interfaces- *Input devices and strategies, Interaction styles*

General Terms

Human Factors

Introduction

Creative processes of choreographers, or other authors in the performative arts, involve several rehearsal iterations where video based annotation can significantly enhance the process. The TKB (Transmedia Knowledge-Base for contemporary dance) project has the aim to provide a research space for rigorous, critical exploration of the relationship between linguistics, dance studies, new digital media and thought/consciousness. The Creation Tool developed in the scope of TKB was conceived and designed to assist these creative processes. It enables to record and augment video materials with personal annotations on composition or creative video editing during the rehearsal periods of a particular work.

Currently, there is a large growth in popularity of tactile and multi-touch devices and therefore it is increasingly necessary to develop applications and interfaces that explore this type of interaction. This type of direct input device is often called natural and familiar when compared to the traditional mouse and keyboard. Pen and touch as input methods offer the opportunity to craft new user experiences that are uniquely well suited to how people naturally work with pen and paper [5,8]. This interaction, using pen and touch, is superior in terms of speed, accuracy and user preference, than other existing combinations [2]. Additionally, pen-based technology combined with digital video allows the association between handwritten annotations, or sketches, with video segments, called pen-based video annotations [3].

This paper presents a tool that supports the capture and multimodal annotation of video content. The

annotations associated to video content can be text, ink strokes or audio. The prototype was developed for Tablet PCs exploring bimanual pen and touch interaction. Presently, due to their design and computational power, Tablet PCs are a very interesting approach to electronic paper [3]. Their affordances create a human-computer interaction similar to human-pen-paper interaction [3, 7].

Contemporary Dance - A Case Study

In this paper, we present a video annotator, which can be used in different environments. In our case study, the system is used as a creation tool for contemporary dance. One choreographer can use the system to analyze and improve his work, by recording and annotating a rehearsal, or a live performance, for a later review or for sharing his notes with the performers.

The system can be used for after event analysis, using a pre-recorded video stream, or in a live event, using a live video stream. This second mode requires a familiar interface and setup, so that the usage of such a tool does not interfere with an already existing creation process. The usage of Tablet PCs + Video Camera (webcam or regular) and pen + touch input imitates the human interaction with pen and paper on electronic devices. This technology can be applied to any environment without interfering with existing creative processes and work methodologies.

Motion is an essential element in dance and graphical video annotations, particularly those made using a stylus, can require motion tracking methods. Graphical notes should follow the associated video elements [3]. Therefore, this case study also has the aim to explore

pen-based video annotations combined with motion tracking.

The creation tool is one of the components under development in the framework of the TKB project, an international transdisciplinary research project working in the intersection of cognitive linguistics, dance studies and new media. Two choreographers, invited by the TKB project, are following the development of the tool and will use it in their creative process, contributing for future usability tests.

Implementation

The prototype is being developed for Tablet PCs and implemented in C++, using the OpenCV 2.1¹, openFrameworks v0.061² and Qt GUI 4.7.0³ frameworks [4]. The developed interface was designed exploring two forms of bimanual inputs, using direct touch and through a stylus. Presently, the touch input is used for mode and tool switching, exploring the non-preferred hand mode switching [7]. In addition, the tool allows video annotations, using the keyboard, digital ink or voice as input. Motion tracking methods are being applied to video annotations, in order to maintain their context. The video content associated with the annotations is tracked and the annotations move accordingly. This feature will be generalized, allowing the selection and association of any element or video region to an annotation, using pen gestures.

¹ <http://opencv.willowgarage.com/wiki/>

² <http://www.openframeworks.cc/>

³ <http://qt.nokia.com/>



figure 1. Video annotation made with the Creation Tool

Interface Design

The interface design started from the analysis of the user needs, technical specifications and required content and features. Usability, user-friendly design and the ability to categorize the information into a coherent structure were the main concerns during the interface creation process. The navigation structure was developed based on the two levels of human-computer-interaction, hardware (physical interface-Tablet PC) vs. software (virtual interface-contents).

The size of Tablet PCs fosters bimanual interaction, allowing easy changes between pen and touch modes, and faster navigation through interface contents.

The video annotator interface is composed by a video display area, presenting a live or pre-recorded stream, in which it is possible to augment the content with annotations, using a stylus or a virtual keyboard. It is

also possible to annotate in the area around the video window, avoiding the occlusion of video elements by the user's notes. Annotations, in the form of sketches, provide to the choreographer more freedom in the creative process. In addition, annotations can be recorded using a microphone, thus creating voice annotations. This type of notes combined with a wireless microphone, allow the choreographer to have a remote annotation input. All annotations are shown in a timeline, as well as the correspondent frames of video.

The contents were organized and distributed in five sections: in the upper left side of the display, there is the *system menu* with the standard system functionalities (file, open, save, edit, help) and, in the left side, the *tool bar* (selection, draw, text, mark, link and motion tracking tools). In the upper right side, the *project menu* manages all the content and files generated through the annotation process and, in the right side, *the toolsets* (related with toolbar tools, e.g., line color, line width, eraser) are included. Additionally, in the bottom of the display, the *timeline* is used for video and notes navigation (see figure 1).

During the choreographer's creation process a huge amount of documents, such as annotations, videos, references, links to external files, are generated, compiled and organized in one support, the creation tool. The design attempted to create an interface with the least graphical noise possible, neutral colors (gray scale-which avoids eyestrain) and a simple approach in order to separate graphic interface elements from all the visual content information. To improve user's feedback with pen and touch interaction, all buttons (type/icons) have an immediate feedback in response to button presses. Sizes and spacing between them are

strictly related with the touch screen specifications in order to have a better performance [6].

Video Annotations

Aiming to correspond to the needs of the choreographers two types of annotations were defined: regular annotations and bookmarks. The bookmarks are concepts, defined by the choreographer, represented by a keyword, such as "light" or "stage", and have a correspondent icon, in contrast to regular annotations that do not have a pre-defined structure.

Annotations can be made using three different inputs: keyboard (touch keyboard), pen and voice. The pen-based and voice inputs allow two distinct types of interactions not supported by the keyboard. The voice annotations, combined with a wireless microphone, provide an instantaneous and location-free method of annotation. Pen-based video annotations provide to the user, in this case the choreographer, the ability to sketch over a video stream, allowing more freedom in the creative process. Pen-based annotations can also be used to sketch the icons of bookmark annotations. The annotations made using a stylus can require motion tracking, with the aim to follow a particular video element.

VOICE ANNOTATIONS

The system permits the recording of two different audio streams, allowing the usage of two audio recording devices, e.g., video camera and a microphone. This feature is achieved by using the Libsndfile⁴ and RtAudio⁵ frameworks, which will allow the

⁴ <http://www.mega-nerd.com/libsndfile/>

⁵ <http://www.music.mcgill.ca/~gary/rtaudio/>

simultaneous but separated recording of audio annotations and the audio of the video stream. A voice recognition method will be applied in the future, with the aim to transform the voice annotations into text.

PEN-BASED VIDEO ANNOTATIONS

The system implements and explores the concept of pen-based video annotations presented in [3]. In order to maintain context, pen-based video annotations require motion tracking. The development of the prototype includes motion tracking methods that can be applied to the environment of the case study.

Motion Tracking

With the goal of tracking dancers in the scene, an object tracking library is under development. A basic design is already in place and OpenCV's implementation of the CAMSHIFT tracker [1] has been adapted to it. Due to the availability of the tracker this allowed for a quick way to test the object tracking framework. Despite being a useful way to find and overcome some shortcomings in the framework's design, this tracker has proven rather limited for use in actual dance videos, especially whenever the dancers and parts of the background have similar color histograms (see figure 2).

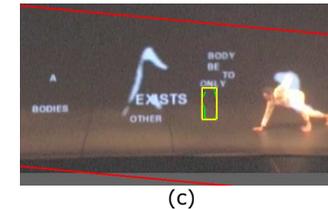


figure 2. Pen-based Video Annotation with CAMSHIFT tracking: a) Pen-Based Video Annotation in Green b) Motion tracking with CAMSHIFT starts. The yellow rectangle is the annotation's bounding box, and the red rectangle is CAMSHIFT's output. c) Motion tracking with CAMSHIFT loses performer.

The work currently being done consists of implementing a more robust tracker, using this framework, for use in the video annotation tool. To this end, the background subtraction algorithm presented in [9] will be adapted to the framework, since it seems to be robust enough to handle lighting and camera position changes, and it has a readily available implementation.

Conclusions and Future Work

In this paper, a video annotator for Tablet PCs, which was mainly developed for bimanual pen and touch interaction, was described. The system also allows keyboard (touch keyboard), pen and voice based annotations.

The video annotator is applied to contemporary dance as a creation tool. It was conceived to assist the creative processes of choreographers, working as a digital notebook for personal annotations, during their rehearsals. While used by choreographers invited by the TKB project, it can be complemented with other

detailed annotations on the finished pieces. In this case, it allows the creators to use the tool not only for their own individual purposes, but also to export their notes to a general web-based archive to be launched further on. If used as a single tool, we believe it can be of use to several other purposes, such as in social and anthropological field work, creative video and educational settings.

The system is still under development and some important steps have yet to be performed, such as the integration of the main interface with the motion tracking framework and usability testing. Nevertheless, some conclusions can already be reported. The user interaction for video annotations is highly dependent of how precise a video element can be defined and tracked by the system. The selection of specific elements depends on the accuracy of the methods used to define and track these elements. Annotations will also include hypermedia notes, allowing the user to annotate video content with different media, such as images, web pages or other videos.

Acknowledgements

This work is partially funded by the project “TKB - A Transmedia Knowledge Base for contemporary dance” (PTDC/EAT/AVP/098220/2008 funded by FCT/MCTES), the UTAustin-Portugal, Digital Media, Program (SFRH/BD/42662/2007) and by CITI/DI/FCT/UNL. We would like to thank to the choreographers Rui Horta and Stephan Jürgens for their suggestions and help.

References

[1] Bradski, G. R. Computer Vision Face Tracking For Use in a Perceptual User Interface. *Intel Technology Journal* 2, 2 (1998), 12–21.

[2] Brandl, P., Forlines, C., Wigdor, D., Haller, M. and Shen, C. Combining and measuring the benefits of bimanual pen and direct-touch interaction on horizontal interfaces. In *Proc. AVI 2008*. ACM Press (2008), 154-161.

[3] Cabral, D. and Correia, N. Pen-Based Video Annotations: A Proposal and a Prototype for Tablet PCs. In *Proc. INTERACT 2009*, LNCS 5727, Springer-Verlag (2009), 17-20.

[4] Cabral, D. and Valente, J. Programmer’s Guide for QT Gui + openFrameworks (OF) in C++ (Visual Studio 2008). Technical Report. CITI and DI, FCT/UNL (2011).

[5] Hinckley, K., Yatani, K., Pahud, M., Coddington, N., Rodenhouse, J., Wilson, A., Benko, H. and Buxton, B. Pen + touch = new tools. In *Proc. UIST 2010*. ACM Press (2010), 7-36.

[6] iPhone Human Interface Guidelines for Web Applications: User Experience. Technical Report. Apple Inc. (2010).

[7] Li, Y., Hinckley, K., Guan, Z. and Landay, J. A. Experimental analysis of mode switching techniques in pen-based user interfaces. In *Proc. CHI 2005*. ACM Press (2005), 461-470.

[8] Sellen, A. J. and Harper, R. H.R. *The Myth of the Paperless Office*. MIT Press, Cambridge, MA, USA, 2003.

[9] Zivkovic, Z., and van der Heijden, F. Efficient Adaptive Density Estimation per Image Pixel for the Task of Background Subtraction. *Pattern Recognition Letters* 27, 7 (2006), 773-780.