CrossMark

Video editing with pen-based technology

Diogo Cabral[1] [ID] · **Nuno Correia**[2]

Abstract The manipulation of video content is still a difficult task due to its complexity and richness. This paper applies pen-based technology to video editing, with the goal to improve such interaction. In this research, digital ink is replaced by video content, aiming to provide a more familiar and creative interaction for video editing and to study how pen gestures can be used on this context as well as what kind of changes are needed in the interface. The concept was implemented in a Tablet PC prototype and evaluated by expert and non-expert users. The user feedback shows that this approach proved to be natural and at the same time to foster user creativity as measured by the Creative Support Index.

Keywords Video editing · Pen-based interfaces · Pen-based video editing

1 Introduction

Video can be considered one of the most complete and complex media [13, 42]. The use of digital platforms not only have changed video editing and visualization processes but also proven to be necessary to improve video content browsing and searching [6, 8, 21, 42]. Nonetheless, manipulating digital video is still a difficult and tedious task [5, 6, 10, 11, 27]. Recent developments in digital video technology, like video sharing platforms (e.g.,

✉ Diogo Cabral
diogo.cabral@staff.uma.pt

Nuno Correia
nmc@fct.unl.pt

1   Madeira-ITI, University of Madeira, Campus da Penteada, 9020-105, Funchal, Portugal

2   NOVA LINCS, Departamento de Informática, Faculdade de Ciências e Tecnologia, Universidade Nova de Lisboa, Campus da Caparica, 2829-516 Caparica, Portugal

YouTube,[1] Vimeo[2]) or the integration of video cameras in mobile devices, have increased the production, distribution and access of video content, making video manipulation not only an important issue for professionals, but also for non-expert users. Therefore, it is crucial to find more natural ways to interact with video content.

In the *Dictionnaire Mondial du Cinèma* [38], it is defined that film editing is the phase of the manufacture of a film in which images and sounds are assembled and arranged. However, digital technology made the editing process equal on film and video [8, 38]. Therefore, it is possible to use the same definition of film editing for video: *editing is the phase of the manufacture of a video in which images and sounds are assembled and arranged.* Editing not only allows to shoot different scenes in a non-linear order, which can be properly selected and arranged on a post-production process, but can also be used for narrative purposes [8, 35], like narrative clarity; dramatic emphasis; subtext and aesthetics. Nonetheless, video editing would benefit with solutions that make the process more efficient [5, 10, 11, 27], easier to learn [6], but at the same time foster the users' creativity [8].

The work featured in this research, applies pen computing to video manipulation, aiming to improve user media interaction. Pen-based technology, a familiar computer input interface for humans [24], can be considered a natural candidate for improving video interaction. It can be used indirectly, through digital ink, or directly, using pen gestures or pressure. Digital ink associated to video content, i.e., pen-based video annotations [2], has been a fairly studied research topic since the 1970's (e.g., the Reiffel's *Telestrator* [30], the GALATEA system [26]) until very recently (e.g., LEAN [27], Creation-Tool [3, 33]). Whereas, how pen gestures or pressure can be used to improve video interaction, particularly on video editing tasks, is still an open issue.

Vogel and Balakrishnan [39], in their study about the difficulties of direct pen interaction with a conventional graphical user interfaces (GUI), alerted for the need of improvement of hardware, base interactions and widgets behavior. The need of interface adaptations for pen-based interactions was also expressed by Marshal [22] in her book about reading and writing on electronic books. Therefore, by applying pen-based technology as an input interface to the manipulation and editing of a time-based visual media like video, three main research questions can be made:

– How to apply pen-based technology to video interaction, making it more natural and at the same time fostering the user's creativity?
– How to improve video manipulation and editing using pen gestures or pressure?
– What kind of changes are needed in a video editor interface, in order to achieve this improvement?

In order to answer these research questions, this work contributes with:

– A set of pen-based interactions for video editing based on the idea of replacing digital ink by video content, which can be painted in a canvas, and their implementation on a Tablet PC prototype exploring pen-based gestures and pressure.
– A preliminary evaluation of the interactions and prototype made by expert and non-expert users. The feedback was made through questionnaires, the Creativity Support Index (CSI) and informal discussions.

---

[1]http://www.youtube.com/

[2]http://vimeo.com/

The main findings of evaluation were that the general idea of using pen-based gestures for painting video content was well accepted by the users, allowing them to explore different outcomes of the final video content; the visual organization and creativity fostered by the concept as well as the easiness to use the pen to manipulate directly video content.

2 Background and related work

In the first motion pictures, dated from the end of the XIX century, there was no editing [8]. The camera recorded an event, an act, or an incident. Many of them were a single short shot. It was with Edwin S. Porter, D.W. Griffith, Sergei Eisenstein and Alfred Hitchcock (with initial sound film editing experiments) in the beginning of the XX century, that film editing took the form that is known today. Since these developments and achievements, film editing is a key process in cinematography.

Davis [9] and Chandler [6] describe three major technological phases of motion picture editing: physical film cutting; electronic videotape editing and digital nonlinear editing. Dancyger [8] defines nonlinear editing as "random-access editing, sourcing shots, scenes, and sounds on an as needed basis". Although film and video editing suffered the revolution of passing from analog- to digital-driven technology [8], still remain time-consuming, frustrating and tedious tasks [5, 10–12, 27]. The difficulty of learning how to use current commercial editing software was also discussed by Chandler [6], and the Jokela et al [17] study showed that the technical complexity of video editing is a barrier for users.

Although non-professional applications (e.g., Movie Maker[3] and iMovie[4]) aim to facilitate the editing task, it is possible to observe that this is mostly achieved by reducing the number of features, when compared with professional and advanced amateur systems (e.g., Avid Media Composer,[5] Adobe Premiere Pro[6] and Final Cut Pro[7]).

Most of the research improvements on digital editing were based on trying to automate the process (e.g., Lienhart's work [18], Silver [5], EWW [40]), a philosophy defended and discussed by Davis [9]. However, Dancyger [8] points out that editing is a creative process, which cannot be made by machines, and most of the developments made on storytelling, interactivity and the relation between the storyteller and the audience were made in the fields of video games and education. By taking these facts in consideration and adding the fact of the increasing popularity of video sharing platforms, mentioned in Section 1, it is possible to conclude that providing more familiar and powerful interfaces that foster easiness and creativity, is a key issue for digital video editing systems. Some research works tried to follow this approach, as described next.

The Video Mosaic [20] and Goldman's work [12] tried to improve the process by combining storyboards with video editing software. In the Hitchcock system [11], the interface presented piles of clustered video clips from the raw data, based on color histogram analysis and visual quality. The user could select the clip from each pile and drag them into the composition panel and generate the new video stream.

The EnhancedMovie [25] explored hand gesture commands for editing tasks. In order to simulate a large-size display in a desk, the system used a camera for gesture detection and a video projection. The EnhancedMovie interface allowed one to select a video clip, from a set of clips, and define the start and the end points of a new video segment, using particular gestures.

The TextableMovie [37] constructed a new video stream while someone was composing a narrative, using text or voice. The system matched the narrative text with the keywords associated to video clips and composed, in real-time, an equivalent video narrative. The TextableMovie was followed by a more tangible and collaborative approach, presented in Movie Pictures [36]. In Movie Pictures, the text narrative was replaced by RFID tokens, each associated with a video clip, and that could be reassembled composing a new video stream. In the Tangible Video Editor (TVE) [43], a tangible interface for video editing based on active tokens was also exploited. The system was composed by a play-controller and sets of clip-holders (Pocket PCs inside plastic cases) and transition connectors, which could be attached to each other. A data stream flowed from the right to the left, traversing all the devices and stopping in the play controller, placed in the beginning of the sequence. The play-controller sent the information sequence to a desktop, which displayed the final movie. In the users' study, some users missed more complete editing features like cut, merge, color correction, cropping or speed control.

In order to provide a more fluid video interaction, different research projects explored pen-based technology for this task. However, most of them applied this type of technology to video browsing and navigation, as presented in the next section.

2.1 Pen-based video interaction and editing

One of the first proposals that used pen computing to control video content was the Marquee [41], which used a pen-based interface to control a VCR device. A user could control the direction and speed of regular VCR controls (play, pause, backward and forward) by drawing a horizontal line using the stylus. The direction and speed of the video content were controlled by the direction and length of the line. In addition, a stylus tap within the control area paused the tape.

The usage of pen-based technology to control or edit video content was also tried in other proposals, such as LEAN [27], Zlider [28], Videotater [10] and the MobileZoom-Slider/ScrollWheel [16].

In the LEAN system [27], novel interfaces, like the TLSlider and PVSlider, were developed to browse video content. The TLSlider departs from the fish-eye frame layout, which focus the user attention in a particular frame, to a sinusoidal frame arrangement, that could be increased or decreased depending on the pen pressure and focusing the user attention in a set of frames. The PVSlider added an extra time bar to video navigation. By changing the PVSlider bar position relatively to the video window it was possible to change the video interval (clip) to watch. In addition, it was also possible to change the playback velocity by dragging the pen along the extra bar in the direction of its endpoints. In addition, LEAN used single-stroke gesture commands, which were parsed using Rubine's feature [31]. Adding to the LEAN project, the same authors have developed the Zlider [28] interface, a timeline with a variable scale which depends on the pen pressure made by the user.

The Videotater [10] used vertical and horizontal pen gestures, on a global timeline, in order to split and join different video segments. In addition, the Videotater presented a poly-focal visualization, allowing to watch the endpoints of a video segment and a pre-defined number of frames in its neighborhood.

Table 1 Pen-based video interaction systems

	Video browsing	Video editing	Painting with video content
Marquee	Yes	No	No
LEAN	Yes	No	No
Zlider	Yes	No	No
MobileZoomSlide/ ScroolWheel	Yes	No	No
Videotater	Yes	Yes	No
I/O Brush	No	No	Yes
Video as Ink	No	Yes	Yes

The MobileZoomSlider/ScrollWheel [16] interface allowed one to browse video content using pen-based technology to change the timeline scale. The MobileZoomSlider scale is changed by moving the pen, up or down, in the vertical direction, whereas in the ScrollWheel a circular movement is used, in order to change the timeline scale.

The I/O Brush [32], a drawing tool based on a physical brush, equipped with a small video camera, lights and touch sensors, enabled to record color, texture or movement of any physical object and to reproduce it in a digital canvas using the brush. The camera captured a real scene depending on the mode: one frame for the texture, RGB color for the color and 30-frames for movement. The capture was initialized by the touch sensors and illuminated by the lights of the brush. In addition, the coil of a pen tip was embedded in the brush tip with the aim to interact and control a pen-based display used as a digital canvas.

Marquee [41], LEAN [27], Zlider [28] and MobileZoomSlider/ScrollWheel [16] used pen-based interaction only for video browsing and visualization, without changing the content. In Videotater [10], pen-based technology was used to join and split video segments. Nevertheless, the Videotater interface follows the traditional scheme of the global timeline and a separated video window, without taking full advantage of the natural interaction provided by pen-based technology. Even though the I/O Brush [32] was not focused on pen-based video manipulation or control, it shows how digital ink can take different forms, besides imitating regular physical ink, and reproduce different media content.

This research proposes the use of a pen as an input interface for video editing, by "inking" video frames in a canvas and using a video palette formed by video segments. In this canvas, video content can be painted, selected and manipulated, in different directions, fostering visual organization and creativity. In addition, pen pressure is used for zoom features. The different features were implemented in a Tablet PC prototype, called Video as Ink (Table 1).

3 Video as ink: the concept

One advantage of using digital pens, when compared with regular pens, is that they can be used to perform different tasks and digital ink can be replaced by other media or forms. This idea can be found in the research work developed by Ryokay et al. [32], in which the pen (embedded in a physical brush) is used to paint different types of media in a digital canvas, and by Hinkley et al. [15], where the pen takes the form of an x-acto that cuts digital images.

Considering the principle that digital ink is not limited to imitate physical ink, the concept of *videoink* explores a painting metaphor where ink is composed of video content. In the same way a painter places a brush in an ink bucket or a palette and, after, paints in a canvas
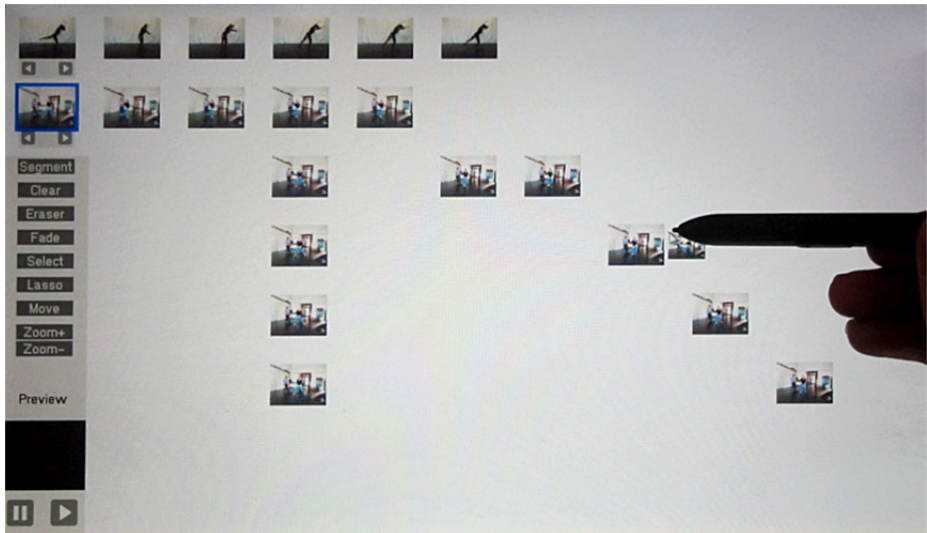
Fig. 1 Inking with video frames: main menu on the right side with two video clips on the *top*, buttons for selecting different features on the middle and a video window to play the generated stream on the *bottom*; canvas with painted frames and video thumbnail as pen's cursor on the *left side*

with the selected ink, one can select a video clip and use a pen to paint on a screen the video frames that belong to that particular clip. Therefore, instead of imitating regular ink, the trail left by the pen is replaced by video content. This change can give the idea that the user is directly manipulating the video content, using a familiar interaction, as well as reduce the number of widgets usually necessary for video manipulation and editing. An implementation of this concept is described next.

3.1 Video as ink: proof-of-concept prototype

A prototype of the *videoink* concept was implemented as part of this research. The prototype was developed for Tablets and exploits pen gestures and pressure for video editing.

In the implementation, the timeline is represented in two dimensions, instead of the more usual 1D horizontal timeline. The two-dimensional timeline allows one to the paint the video content horizontally, vertically or diagonally. The direct use of pen coordinates to add a frame in the canvas would cause successive frame occlusions, causing interesting but not very useful visual effects. In order to avoid such occlusions, the canvas is mapped to a 2D matrix, which is invisible for the user, but where the frames have a pre-reserved space (see Section 3.2). Thus, when the user drags the pen on the screen, the video frames are placed in the correspondent place of the matrix (Fig. 1). The final video stream can be composed by all the video content displayed in the canvas, traversing it from top left corner to bottom right, or by selecting a particular set of frames or segments from the canvas. Since selection can be done horizontally, vertically or diagonally, this second method can be used for non-linear video editing.

The prototype interface is composed of a menu bar, on the left, and a canvas area, on the right. Figure 1 shows two video clips (used as example) on the top left corner, working as video buckets, and a trail of painted frames on the right. The two buttons below each
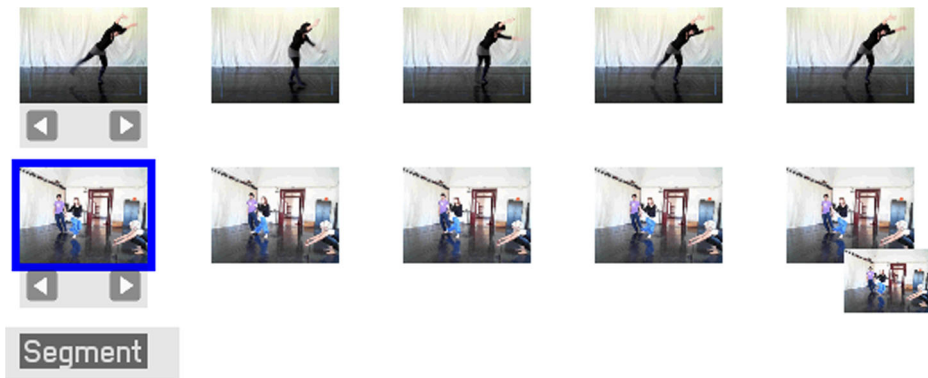
Fig. 2 Frame mode

clip can be used to move from the current frame of the clip to the next or previous one. Different editing features can be chosen in the menu below the clips and, at the bottom of the menu bar, a video window shows a preview of the new video stream. The content is scrollable using up and down buttons at the right side of the canvas. The pen's cursor shows a thumbnail of the next video frame to be "inked".

3.2 The canvas

The prototype includes a canvas where the video content can be painted, selected, moved or erased. This canvas works as a timeline, like in other video editing software, but with the difference that video content can be displayed horizontally, vertically or diagonally. The canvas was defined as horizontally limited, i.e., there is a maximum of frames that can be painted in each row of the canvas, but vertically unlimited. A bidirectionally limited canvas would reduce the working space, whereas an unlimited canvas in both directions could be too confusing. Thus, it was decided that rows should be limited, forcing the user to change to the row below. The implemented approach breaks the traditional horizontal timeline, providing a better visual organization and allowing the exploration of different alternatives much easier.

3.3 Painting video: video frames vs video segments

In the proof-of-concept prototype two basic modes were defined: "Frame" and "Segment". In the "Frame" mode, the user paints a single frame in the canvas, while in the "Segment" mode it is possible to paint a video segment, i.e., a set of consecutive frames, using one single gesture. In the "Frame" mode, all painted frames are displayed (Fig. 2), whereas in the "Segment" mode, only the start and end frames of a segment are shown (Fig. 3). Each segment is represented, horizontally in the canvas, by its start (on the left) and end (on the right) frames connected by a gray box. Even if the frames were painted one by one, the system considers that they compose a video segment if they are horizontally consecutive, i.e., with no holes between them. In this situation the start and end frames are automatically defined by the end points of the frame set. An isolated frame is considered a special segment represented by a single frame, which is simultaneously its start and end frames. Transitions frames placed between two different clips were considered separated segments.
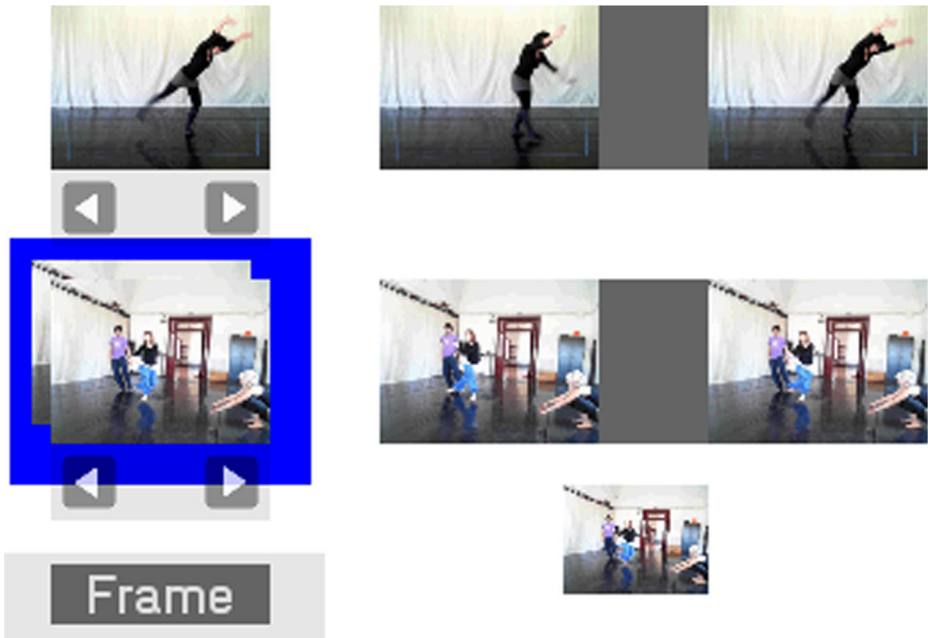
Fig. 3 Segment mode

The visual change of the canvas is synchronized with transformation on the menu. Each mode is represented by a visual change on the selected clip: in the "Frame" mode the clip only shows the current frame, whereas in the "Segment" mode part of the last frame it is displayed behind the current frame (Figs. 2 and 3).

The selection of each mode (Frame or Segment) follows the same principle when a brush is sank inside an ink bucket, i.e., the more the brush (the pen) is sank, more ink is attached to it and more of it will be painted in the canvas. This idea was implemented using the pressure made by the pen tip against the screen over the selected clip. A pressure threshold was defined and if it is exceeded, the "Segment" mode is triggered. In order to return to the "Frame" mode, is it sufficient to tap on top of one clip with a pressure below this threshold. The threshold was experimentally defined at 99 % of the maximum level of pressure represented by the pen. An alternative (and more traditional) mechanism, based on a switch button (Figs. 2 and 3), was also implemented with the goal of comparing both techniques.

After selecting the mode, the user has to tap or drag in the canvas, in order to paint the video content of the selected video clip (Fig. 4). In the case of the "Frame" mode a single frame is painted each time the user passes with the pen tip over a rectangle of the matrix.

Fig. 4 Painting frames by dragging the pen

Fig. 5 Move a frame by dragging

After a frame is painted, the clip automatically moves to the next frame, following the ink metaphor, i.e., the ink attached to a brush or inside a pen goes into a physical surface, when it is in contact with this surface. If the user selects the "Segment" mode, a video clip, defined by its current and last frames, is painted in the canvas.

In the "Segment" mode, the implementation presents two limitations: 1) the video segments are only represented horizontally, left to right, in the canvas and 2) the dragging gesture is not used for painting segments, i.e., it is sufficient to tap with the pen tip somewhere in the canvas, in order to add a video segment. The vertical representation of video segments can perfectly fit in the concept of video as ink. However, it would introduce an additional level of complexity that was too early to implement, without having the users' feedback about simpler situations, e.g., painting or selecting frame-by-frame in any direction or sequence. Regarding the dragging gesture, since the video segment is defined by the current and last frames of each clip, there is no video content left to be painted. Nonetheless, combining these two limitations it is possible to observe that the dragging gesture could be used to indicate the direction of the video segment to be painted (see Section 5).

3.4 Video editing features

In this implementation, the video editing main features are divided in two main categories: operations and transitions effects. Operations are related with adding, moving, erasing or selecting content and follow the "ink principle", while transitions effects are composed of visual effects that are usually used to link different clips in a smooth or meaningful way. The selection of each operation in the menu changes the behavior of the pen in the canvas. Adding and erasing content work as previously explained but with opposite functions, i.e., by selecting the eraser mode, pen gestures will erase the elements displayed in the canvas. If the user inks in a place where a frame already exists, the older frame is removed and replaced by the new one. Moving frames can be achieved by dragging each frame from its original position to the new place, as shown in Fig. 5. The implementation was limited to move single frames for the same issues presented above, i.e., it would introduce an additional level of complexity without an initial feedback from the users.

When the user hovers the pen over the space between two painted frames, an empty box is displayed (Fig. 6), indicating that the user can add or move other frame between those two. In this situation, the frames on the right or below are shifted and the new frame is added to the canvas.

Regarding the transitions effects, the same input method can be used to add several of them. Therefore, in this implementation only the fade effect was developed, as an example of such features. The fade effect can be selected from the menu and made by tapping the pen in the middle of the two frames used for the transition effect. A new transition frame, resulting from the composition of the adjacent frames (50 % of each one), can be added using the same process of adding a new frame between two already painted. The frames on
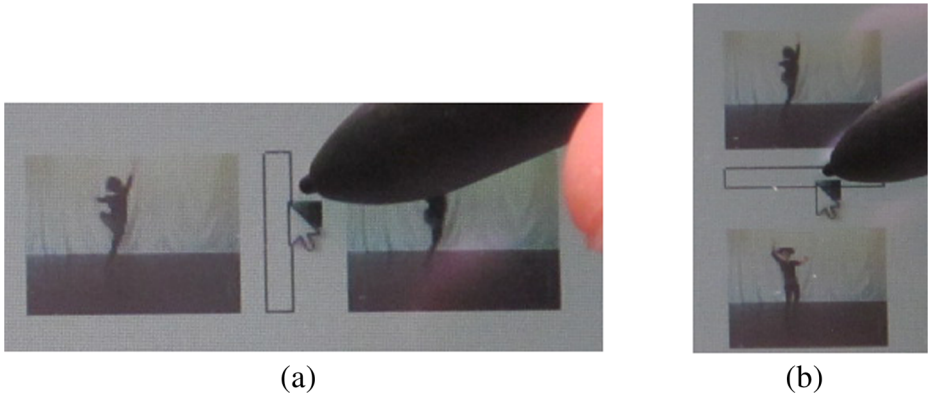
(a)   (b)

Fig. 6 Hit area to add content (hovering the pen). (**a**) On a row. (**b**) On a column

the right or below are shifted and the new frame is added to the canvas. The fade effect can be completed by adding successive blend frames, as shown in Fig. 7. As already mentioned, other transition effects could be developed using the same input method.

3.5 Selecting elements

In order to allow users to select parts of the video two selection methods were developed: paint selection and lasso selection. The paint selection follows the "inking principle". In this mode, the selection is made by pressing or dragging the pen on screen (Figs. 8 and 10a). A frame or segment is selected when the pen passes on top of it. Following
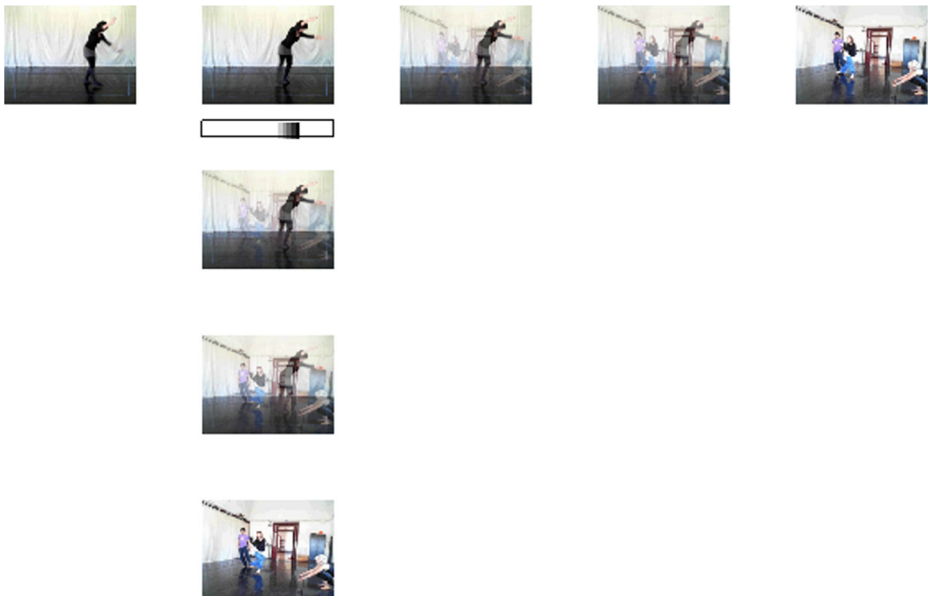
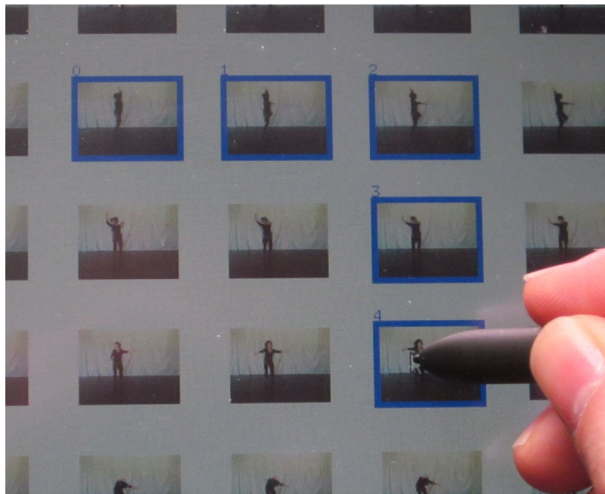**Fig. 7** Transition: fade effect (*horizontal* and *vertical*)

Fig. 8 Paint selection on frames: selected frames present a *blue frame* and a sequence number

the same principle of adding or removing content, paint selection can be made horizontally, vertically or diagonally.

The lasso selection is a more traditional method for graphical selection, usually presented on image editing tools, where the user has to draw a lasso around the frames or segments that should be selected (Figs. 9 and 10b), allowing to select a set of frames or segments with a singular gesture. The selection is made by traversing the frames or segments displayed in the canvas from left to right, top to bottom. The point that represents the center of mass of each frame, in the case of the "Frame" mode, or the point in middle of start and end frames of a segment, in the case of the "Segment" mode, is verified if it is inside or outside the
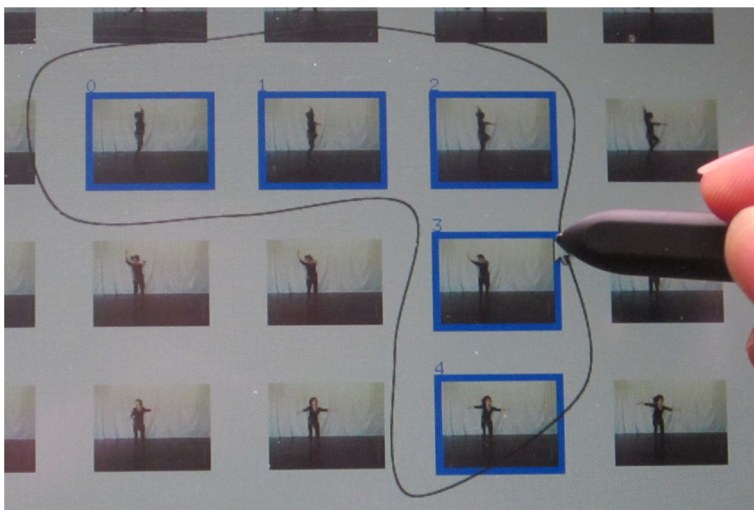

Fig. 9 Lasso selection on frames: selected frames present a *blue frame* and a sequence number
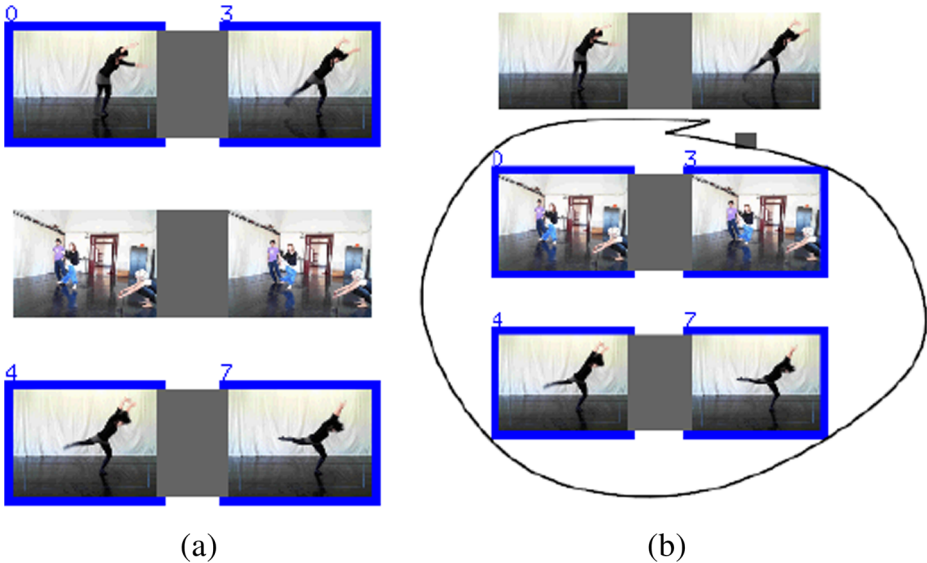
(a)											(b)

Fig. 10 Selecting segments: selected segments present a *blue frame* and a sequence number). - (**a**) Paint Selection. (**b**) Lasso Selection

lasso. This verification is achieved by using a ray-casting algorithm, i.e., counting how many times a ray, starting from that point and going to any fixed direction, intersects the edges of the polygon (the lasso). If the number of intersections is even, the point is outside, if it is odd is inside. The prototype uses an implementation of the ray-casting algorithm made by Alexander Motrichuk.[8] In this implementation the rays are calculated using a horizontal left cross over direction approach and the programmer has to choose if the points that are in the boundary of the polygon can be considered inside or outside. Since it could be awkward to draw a lasso line over a point and this point (and the corresponding frame or segment) will not be included in the selection, it was considered that boundary points are inside the lasso. In both methods, paint and lasso selection, the selected frames present a number on top of each one, defining the order of the frame selection. This ordering is considered for creating a new video stream.

The two selection methods are limited to create a new video stream. However, they could be combined with other video editing features, in order to apply an action to a set of frames, e.g., move a selected set of frames or segments.

3.6 Pressure-based zoom

A pressure-based zoom mechanism was developed as part of the proof-of-concept prototype, aiming to provide a fluid interaction for zooming the canvas. It is composed of two buttons, "Zoom+" and "Zoom-", and the pressure made by the pen tip on top of each button causes a proportional scale of the canvas (Fig. 11). The pressure levels are proportionally scaled into an interval of values between 0 (lowest pressure level) and 1 (maximum pressure level), following the expression (1). These values are used directly for scaling the
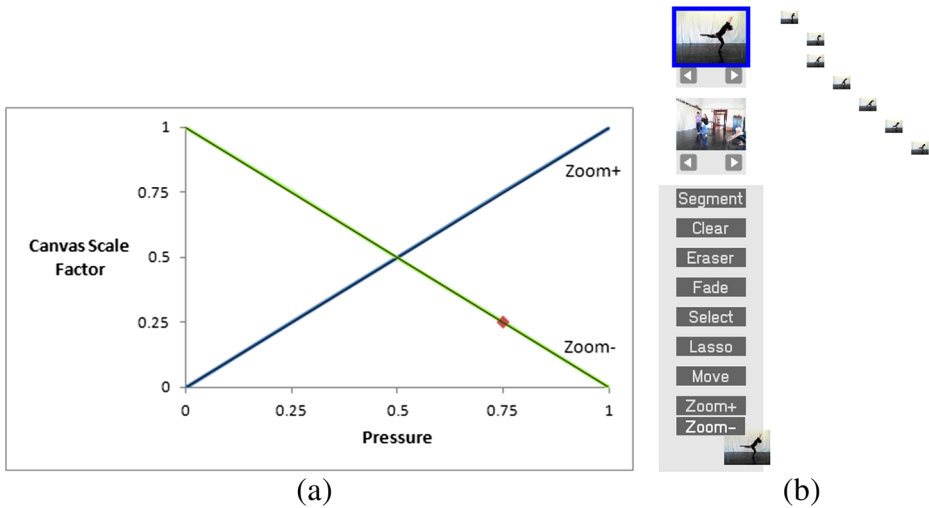
---

[8]http://paulbourke.net/geometry/polygonmesh/InsidePolygonWithBounds.cpp

Fig. 11 Pressure-based zoom. (**a**) Relation between pressure and scale factor. Example: *Red dot* - zooming out with a pressure of 0.75 causes a scaling factor of 0.25. (**b**) Canvas zoomed out (e.g., scaling factor of 0.25)

content displayed in the canvas, where $CanvasScaleFactor_{Zoom+} = ScaledValue$ and $CanvasScaleFactor_{Zoom-} = 1 - ScaledValue$.

$$ScaledValue = \frac{Current PressureValue}{Maximum PressureValue} \qquad (1)$$

3.7 Preliminary evaluation

A preliminary and qualitative study of the prototype was made providing an initial user feedback about the usage of pen for video editing as well as if it could foster their creativity. The study involved 12 participants and was composed of a set of basic tasks, a questionnaire, which included questions from the Creative Support Index (CSI) [4, 7], and an informal discussion.

3.7.1 Participants and tasks

The study involved 12 participants, 8 non-experts and 4 experts. All the experts worked (at that time or in the past) in video or film productions, one accumulate his professional work with video-jockey (VJ) activities and another is a professional designer. 3 of the non-expert participants record video content for fun and work purposes and also use professional software.

The participants were mostly male (66.67 %) and the mean of ages was $\bar{x} = 33.67$ ($\sigma = 6.89$). More than half (7) had a Master degree, 2 had a Bachelor degree, 2 had a Bachelor degree complemented with post-graduate studies and 1 held a PhD. All of them usually record video content, most (83.3 %) use a video or photo camera, 66.67 % use their mobile phones and 8.33 %, corresponding to one user, record video using a webcam. 75 % of the participants edit their videos. From those that do not edit one reported the difficulty and boredom of the task, a second reported the lack of time to do it (the same issue was reported

by one of the users that edit) and the third does not care about editing his own videos. Most
of the participants (75 %) had occasionally experimented pen-based technology, 16.67 %
never had experimented this technology before and 8.33 % (one of the experts) frequently
use it.

Before the start of each test, the video as ink concept was briefly introduced to the partici-
pants. After, they were asked to experiment the different tool features, which took around 15
minutes, and to answer the questionnaire. The tasks included adding frames and segments
on the canvas; selecting a video segment on the menu using pen pressure as well as with a
regular switch button; moving and deleting frames on a canvas; adding an additional frame
and a fade effect between content already painted; selecting content (painted on the canvas)
by inking and using the lasso tool (and play the selection in the video window); and zoom-
ing in and out using pen pressure. Questions related to particular features were answered
immediately after each feature was experimented and more generic questions were left to
the end of the test.

The test was made on a Windows Tablet PC, a Lenovo X220, with the rotated screen
blocking the use of physical keyboard and touchpad. It was asked to the participants to only
use the tablet's pen, during the test. The tablet's pen detects 1024 pressures levels.

3.7.2 Results

The questionnaire was composed of seven questions with semantic differential numerical
scale answers, four (Q1, Q3, Q4, Q6) about perceived difficulty (1 for Difficult - 5 for Easy)
and three (Q2, Q5, Q7) about mode usage rate (1 for Rarely - 5 for Frequently); a set of
questions (Q8) and pairwise factor rankings (Q9) that define the Creative Support Index
(CSI) [4]; one (Q10) based on Microsoft "Product Reaction Cards" classification [1] and a
last open question (Q11) for comments and suggestions.

In order to compare the mode preferences and the perceived difficulty, Friedman tests
(Q1, Q4 and Q7) as well as Wilcoxon Matched-Pairs Signed-Ranks tests (Q2, Q3, Q5 and
Q6), were conducted based on the null hypothesis (H0), i.e, there was not a significant
difference between answers, and on a alpha level of 0.05.

In Q1, the participants were asked to rate the perceived difficulty for adding a video
frame ($\tilde{x} = 5.00$), a video segment ($\tilde{x} = 5.00$) and a transition (the fade effect) ($\tilde{x} = 5.00$)
(Fig. 12). The Friedman test did not present a significant difference ($\chi^2 = 0.7, df = 2, p > 0.05$) between the different features.

Afterwards they were asked to rate the usage (Q2) and the perceived difficulty (Q3) of
the two methods for switching between "Frame" and "Segment" modes: by pressing on
top of the selected clip or tapping on the switch button. Regarding the usage rating, the
Wilcoxon Matched-Pairs Signed-Ranks test ($W = 31, Z = -0.20, p > 0.05$) did not
present a significant difference between using the pressure mechanism ($\tilde{x} = 4.00$) and the
switch button ($\tilde{x} = 4.50$) (Fig. 13).

However, regarding the perceived difficulty, the Wilcoxon Matched-Pairs Signed-Ranks
test ($W = 4.5, Z = -2.34, p < 0.05, r = 0.48$) showed a significant difference between
using the pressure mechanism ($\tilde{x} = 4.00$) and the switch button ($\tilde{x} = 5.00$) (Fig. 14). The
results of Q2 and Q3 show that there is no preference between pressing on top of the selected
clip or tapping on the switch button but the switch button is perceived as the easiest method.

The perceived difficulty of the different ways of generating a new video stream, with no
selection ($\tilde{x} = 5.00$), using the paint selection ($\tilde{x} = 5.00$) and using the lasso selection
($\tilde{x} = 4.00$) was studied in Q4 (Fig. 15). The Friedman test showed a significant differ-
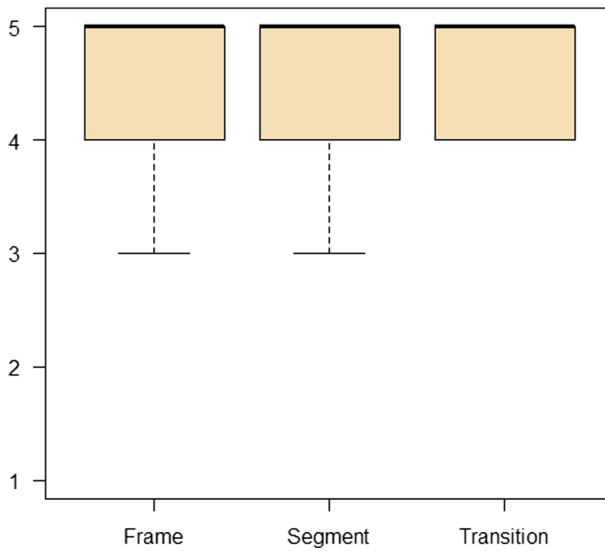ence ($\chi^2 = 6.41, df = 2, p < 0.05$). However, the pairwise comparison, using Wilcoxon

Fig. 12 Results for the perceived difficulty for adding video frame, a video segment and a transition (the *fade effect*) - Median scores

Matched-Pairs Signed-Ranks tests, did not show a significant difference between the different possible comparisons, i.e., between no selection and lasso selection ($W = 7.5$, $Z = -1.82$, $p > 0.05$), no selection and paint selection ($W = 2.5$, $Z = -0.52$, $p > 0.05$) and
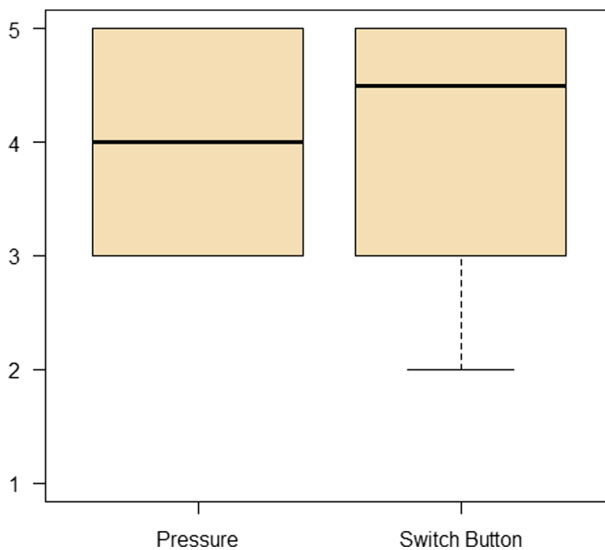

Fig. 13 Results for the usage of the two methods for mode switching: pressing on top of the selected clip or tapping on the switch button - Median scores
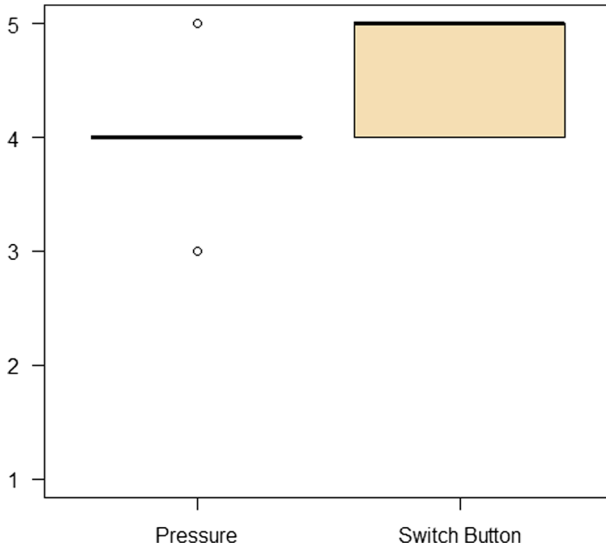
Fig. 14 Results for the perceived difficulty for two methods for mode switching - Median scores

paint selection and lasso selection ($W = 11.5$, $Z = −1.49$, $p > 0.05$). Therefore, it is possible to conclude that there is no significant difference between the perceived difficulty of the different ways of generating a new video stream.

In Q5, it was asked to the participants to rate the usage of the selection modes: paint selection and lasso selection. The Wilcoxon Matched-Pairs Signed-Ranks test ($W = 11.5$, $Z = −2.20$, $p < 0.05$, $r = 0.45$) showed a significant difference between paint selection
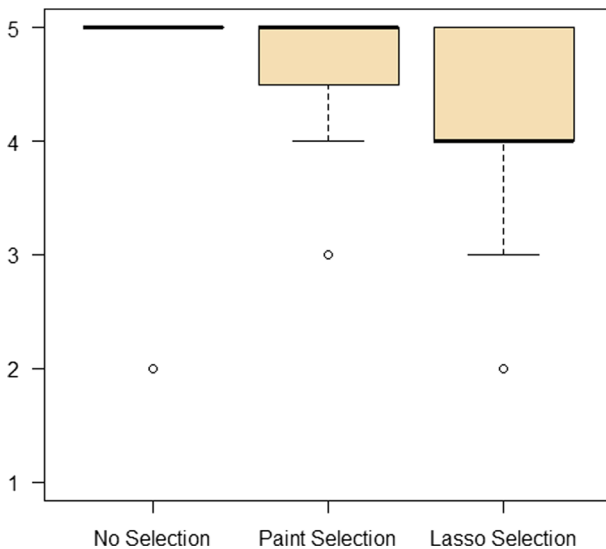

Fig. 15 Results for the perceived difficulty for different ways of generating a new video stream: no selection, paint selection and lasso selection - Median scores
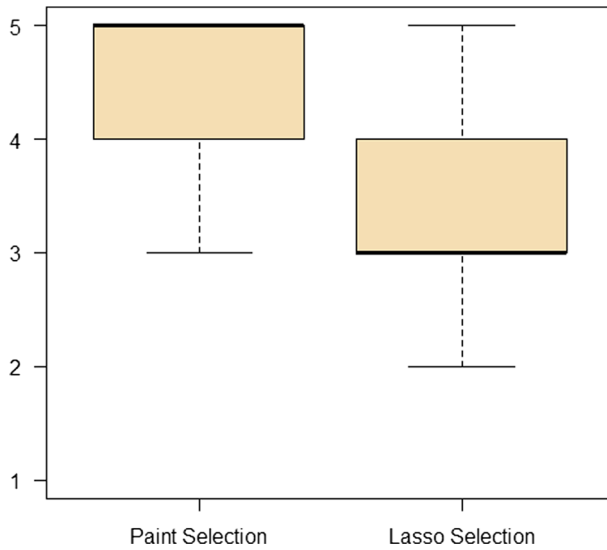
Fig. 16 Results for the usage of the selection modes: paint selection and lasso selection. - Median scores

$(\tilde{x} = 5.00)$ and lasso selection $(\tilde{x} = 3.00)$ (Fig. 16). Therefore, Q5 shows a preference for using the paint selection mode.

The perceived difficulty of the pressure-based zoom mechanism was rated in Q6. A Wilcoxon Matched-Pairs Signed-Ranks test $(W = 14, Z = -0.33, p > 0.05)$ did not show a significant difference between pressure-based zoom mechanism $(\tilde{x} = 4.00)$ and
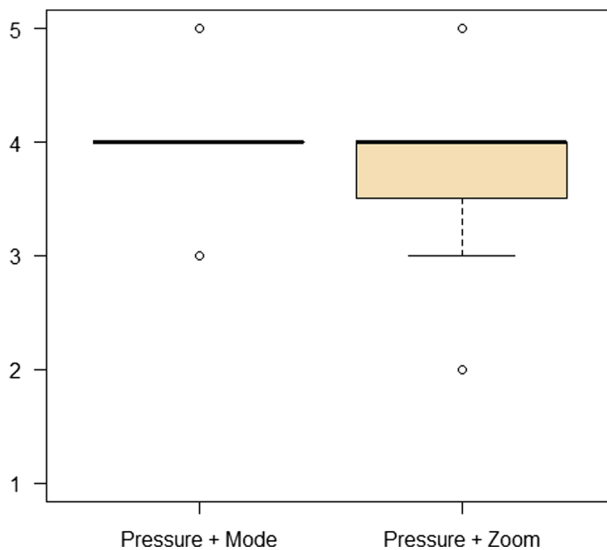

Fig. 17 Results for the perceived difficulty of the two pressure-based mechanisms: mode switch and zoom - Median scores
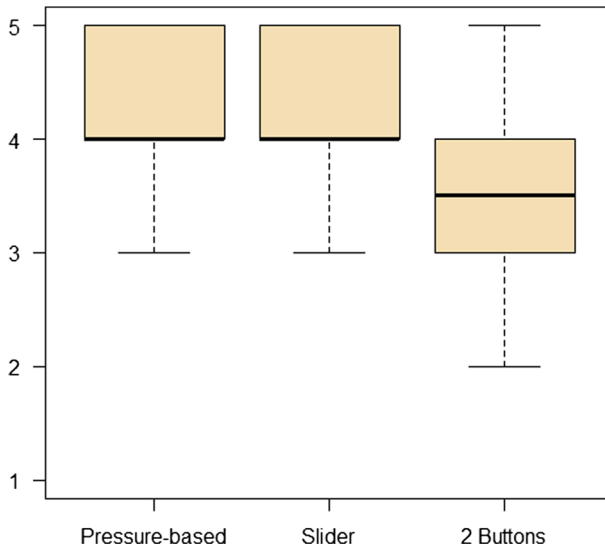
Fig. 18 Results for the usage of the different zoom interfaces: pressure-based, slider and two buttons - Median scores

the pressure mode switch "Frame"/"Segment" mechanism ($\tilde{x} = 4.00$) classified in Q3 (Fig. 17).

In Q7 (Fig. 18), it was asked to the participants to rate the usage between the pressure-based zoom ($\tilde{x} = 4.00$) and other two commonly used zooming mechanisms: a slider that goes to the right or left ($\tilde{x} = 4.00$), increasing or decreasing the zoom, and two simple buttons ($\tilde{x} = 3.50$), one that zooms in and other that zooms out. The Friedman test did also not show a significant difference ($\chi^2 = 3.80, df = 2, p > 0.05$) between the usage of the different zooming mechanisms. It is possible to say there is not a clear preference between the different zooming methods but more studies are needed to confirm this hypothesis. However, these results show a lower preference for the two buttons interface for zooming.

The Creative Support Index (CSI) [4, 7] was studied in questions Q8 and Q9. The CSI is a measurement tool for evaluating creativity support. The CSI is composed of six factors: *exploration*, *expressiveness*, *enjoyment*, *immersion*, *collaboration* and *results worth effort*. The survey metric generates an index between 0 and 100 of the creativity support afforded by a system, tool or interface.

In CSI, the participants have to answer a set of questions, each related to one of the factors, and compare each factor against the other five, assessing the relative importance of these factors (Table 2). The questions are scored in a scale from 0 (Highly Disagree) to 10 (Highly Agree) and the factors are pairwise ranked from 0 to 5. Each answer is multiplied by its associated ranked factor and the overall score for the CSI is calculated by summing all the weighted answers and the result of the sum is divided by 1.5, resulting in a value between 0 and 100.

The original work [4] includes a set of six questions but the software[9] developed by the authors considers two sets of six questions. The calculus of the CSI for these two sets considers the average between the two answers associated to each factor.

───────────────────

[9]http://www.erincarroll.net/csi.html

Table 2 CSI questions and factors

Questions	Factor
Set 1	
I was satisfied with what I got out the system or tool.	Results Worth Effort
It was easy for me to explore many different ideas, outcomes, and possibilities.	Exploration
I would be happy to use this system or tool on a regular basis.	Enjoyment
I was able to be very creative while doing this activity.	Expressiveness
My attention was fully tuned to the activity, and I forgot about the system or tool that I was using.	Immersion
It was really easy to share ideas and designs with other people inside this tool.	Collaboration (not used)
Set 2	
The system was helpful in allowing me to track different ideas, outcomes, or possibilities.	Exploration
I enjoyed this system or tool.	Enjoyment
What I was able to produce was worth the effort I had to exert to produce it.	Results Worth Effort
The system or tool allowed me to be very expressive.	Expressiveness
I became so absorbed in the activity that I forgot about the system or tool that I was using.	Immersion
The system or tool offered support for multiple users.	Collaboration (not used)

As reported by Carrol et al. [4], for non-collaborative tools (as the one being evaluated), questions on the collaboration factor can confuse the users. The authors argument that, in these situations, the users tend to give low scores in the questions and pairwise comparisons related to collaboration. In addition, the authors of this metric also report that the 15 pairwise factor comparisons can be too tedious for the participants. Considering these aspects, it was decided to remove the two questions about collaboration (one per set), and to eliminate the collaboration factor from the pairwise comparisons. This decision reduced the numbers of questions to five per set, ten in total, and the pairwise comparisons to ten. In order to compute the overall CSI score, the lack of answers related to the collaboration factor was compensated by adding one unit to the other factors and setting the collaboration factor to

Table 3 CSI Factor Results

Factors	Avg. factor counts (σ)	Avg. factor score (σ)	Avg. weighted factor score (σ)
Results worth effort	2.58 (1.31)	16.33 (2.53)	41.75 (19.77)
Exploration	4.33 (0.78)	17.00 (1.70)	73.75 (15.39)
Immersion	2.42 (1.16)	13.08 (5.60)	31.08 (22.03)
Expressiveness	3.25 (1.22)	17.17 (2.37)	56.75 (25.60)
Enjoyment	2.42 (1.16)	17.00 (2.30)	41.25 (21.43)
Collaboration	0 (–)	0 (–)	0 (–)

The Avg. Factor Counts is a outcome of the pair wise comparisons; the Avg. Factor Score is a outcome of the two sets of questions and the Avg. Weighted Factor Score is the product of the previous two
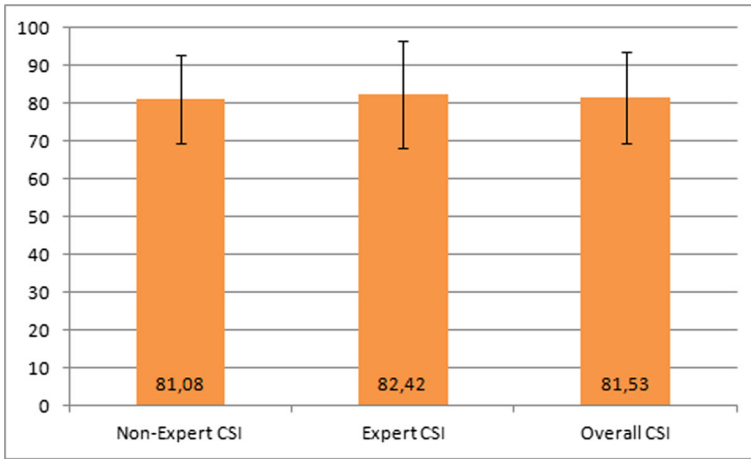
Fig. 19 CSI means: non-expert, expert and overall. *Error bars* represent the standard deviation

zero. This approach assumes that if the users had to choose between collaboration and other factor in the pairwise comparisons, they would always choose the other factor.

Table 3 presents the means for the CSI factors used for the index calculation and Fig. 19 shows the different result of the CSI index. It is possible to observe that the means for non-experts ($\bar{x} = 81.08$, $\sigma = 11.67$) and for experts ($\bar{x} = 82.42$, $\sigma = 14.35$) are very close. The overall CSI mean is 81.53 ($\sigma = 11.97$), which can be considered a high value in the CSI scale. During this research the comparison with other tools was difficult, since only few studies use this metric. In [23] the CSI was computed for an interactive quadruped animation tool (CAT) and in more recent studies [7] for Adobe Photoshop, AutoDesk SketchBook Express and for a Bimanual Color Exploration Plugin (BiCEP). Table 4 compares the different CSI values for the different applications.

In Q10, the users were asked to classify the tool with 28 words of the Microsoft "Product Reaction Cards" [1]. Figure 20 presents the percentage for each word. The most selected (>= 50 %) words were: "easy to use", attractive", "creative", inspiring", and "novel". The comments and suggestions, asked in Q11, were sometimes replaced by an informal discussion. Most of the written comments were about GUI features, like having tooltips or having some visual feedback in the canvas synchronized with the preview window.

Table 4 CSI Values: Comparision betwen (CAT) [23]; Adobe Photoshop, AutoDesk SketchBook Express and BiCEP [7]; and VideoInk

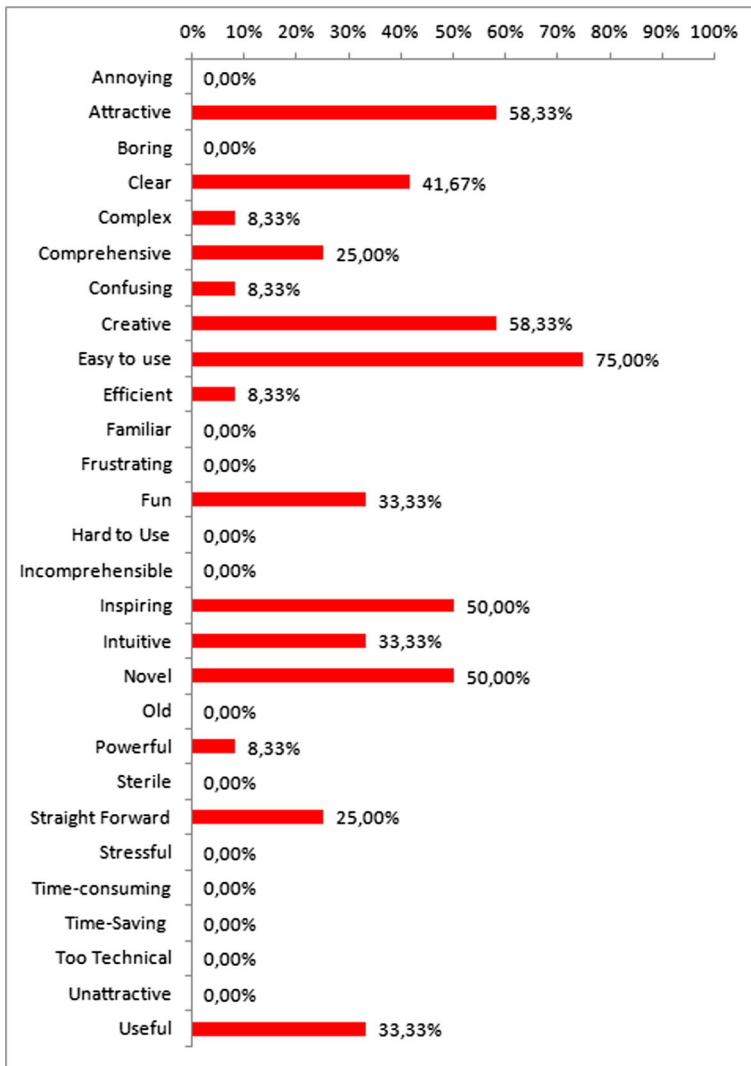|	N=	CSI (σ)
CAT	10	79.20 (17.80)
Photoshop	5	84.20 (18.84)
SketchBook express	11	64.79 (17.06)
BiCEP	16	76.52 (16.25)
VideoInk	12	81.53 (11.97)

Fig. 20 Classification with Microsoft "Product Reaction Cards"

3.7.3 Informal discussions with participants

During or after each test, there was an informal discussion with the participants, both expert and non-expert. However, there was an intentional focus on this type of feedback in the tests with the experts. The general feedback from the experts was positive. They particularly enjoy the possibility to explore different outcomes using the two-dimensional canvas as well as to use the pen, when compared with the mouse. Three of them pointed out the visual organization of the video material in the canvas as an advantage of the tool. When they were asked if they considered the canvas and paint mechanism confusing or creative, all them

answered as being more creative. Two of them reported the advantage of having pen and touch interactions in such approach.

In order to do multiple operations, three of them would also like to have the possibility to use multiple lasso selections, i.e., select different video blocks for preview or move them around the canvas, and one of them would like to have the possibility to change the selection order using this tool, e.g., use the path of lasso to order them. One participant would like to have the ability to zoom in a frame and edit it, as in an animation tool, and have some visual feedback on the frames about the segment that they belong to. In addition, other participant referred that using this tool, the hand gestures easily followed his thoughts but would like to have more time to experiment the tool.

Five of the participants (3 experts and 2 non-experts) referred the need of moving the menu bar to other places of the canvas, particularly to the right side of the screen. The reason of this observation was that the right-handed users had to pass their hand over the canvas, so they could reach the menu bar, causing uncomfortable gestures. Regarding the "Segment" mode, one non-expert participant referred that he/she would like to have the ability to define the start and end frames of a video clip, independently of how they are placed in a frame set.

4 Discussion

The general aspects of the video as ink concept were well received by the participants. Adding content to the canvas was generally well perceived by the participants, although, it was observed that the majority of them did not realize at first, the rectangles between the two frames, indicating that additional content can be added between them. Regarding these rectangles, it was also observed that sometimes it was difficult to hit them with the pen. Larger hit areas with additional visual tips could help on this task. It was also observed, during the tests, that the usage of a small frame (the next frame to be painted) as a cursor could be confusing and some users, while performing moving tasks, tried to pick and drop the frame, instead of dragging it on canvas.

The participants could not perceive which mechanism for mode switching, pressing on the clip or tapping on the switch button, would use more. However, the switch button was perceived as the easiest one. In addition, the participant that frequently uses the pen tested the two modes a couple of times. It was observed that the pressing technique reduces one step, i.e., using a single gesture one selects the clip and the segment mode, but the amount of time needed to achieve the pressure threshold was longer than just tapping on the switch button.

Regarding the construction of a new video stream, there was not a significant difference between the different methods: no selection, paint selection and lasso selection. Nonetheless, when comparing the two modes of selection, the participants preferred the paint selection. Considering the comments about the usage of multiple lasso selection, this improvement could change the users's feedback on the usage of this selection tool.

The pressure-based zoom mechanism was confused, by half of participants, with a time-based mechanism, i.e., the zoom was proportional to the time that the user pressed on one of the two buttons. After the participants discover that the zoom was dependent on the pressure made on the buttons, they have tried to control it. All participants considered hard to control the pressure, due to the high sensibility of the pen, specially with low pressure levels. The same control problem was reported in Ramos et al study [29] about pressure widgets. In addition, three participants would like to have a scale, e.g., 0 % to 100 %, in order to perceive the maximum and minimum zoom levels. Even though the tests did

not show a significant difference between the pressure-based mechanism presented in the prototype and the traditional zoom methods, slider and two buttons, this last one was the less considered for usage by the users. Therefore, it would be interesting to use an interface similar to the Zlider [28] in the *videoink* prototype.

The CSI score presents a high value in the scale but the few studies using this metric makes the comparison with other tools more difficult. Regarding the usage of this metric and due to the short experience with the tool, some users reported some difficulty to answer questions related with *results worth effort* and *immersion* factors. The most ranked factors of CSI metric were *expressiveness* and *exploration*. The combination of these results with the words chosen by the majority of the users to classify their experience with the prototype ("easy to use", attractive", "creative", "inspiring","novel") and with the feedback given during the informal discussions shows that the *videoink* concept fosters creativity using natural interactions.

5 Conclusions and future work

The main principles for a pen-based approach to video editing were presented in this work. The described approach uses video content as digital ink, which can be painted in a canvas that works as a two-dimensional canvas. In the context of this research, a Tablet PC implementation of the concept combined with different video editing features was carried out. The concept was well received by the users, allowing them to explore different outcomes of the final video content. Users pointed out the visual organization and creativity fostered by the *videoink* concept. The easiness to use the pen to manipulate directly video content, when compared with regular software tools that are mainly developed for the mouse or touchpad, was also noted by the users. Nonetheless, a set of limitations and future research directions were detected.

The recent developments of multi-touch pens unfold new modes of video interaction using pen-based technology. The work developed by Song et al. [34] and by Liu and Guimbretière [19] use specific sensors on the pen's barrel, in order to detect grips and gestures made by users' fingers. The pen prototype developed by Hinkley et al. [14], uses accelerometer, gyro and magnetometer sensors, in order to capture motion and context sensing techniques. How these finger grips and gestures can be used on video interaction still needs an answer.

As mentioned in Section 3.3, the proof-of-concept of the *videoink* prototype presents two major limitations regarding video segments: 1) the video segments are represented horizontally, left to right, in the canvas and 2) the dragging gesture is not used for painting segments. The positive feedback from the users on painting and selecting video content on a canvas as a way to explore different outcomes of the final video stream, indicates that the development of bidirectional video segments could be worth the effort of implementing such complex feature.

Finally, this research only considered the visual dimension of video editing. However, sound is also an important factor on video and film editing. Pen-based interactions for sound editing should also be considered on future research work, e.g., for moving sound segments and tracks in the canvas.

References

1. Benedek J, Miner T (2002) Measuring desirability: new methods for evaluating desirability in a usability lab setting. In: Proceedings of UPA usability professional association conference. Microsoft Corporation
2. Cabral D, Correia N (2009) Pen-based video annotations: a proposal and a prototype for tablet pcs. In: Proceedings of the 12th IFIP TC13 Human-computer interaction international conference, part II, lecture notes in computer science, INTERACT'09, vol 5727. Springer, Berlin, pp 17–20
3. Cabral D, Valente JG, Arago U, Fernandes C, Correia N (2012) Evaluation of a multimodal video annotator for contemporary dance. In: Proceedings of the 11th international working conference on advanced visual interfaces, AVI'12. ACM, New York, pp 572–579
4. Carroll EA, Latulipe C, Fung R, Terry M (2009) Creativity factor evaluation: towards a standardized survey metric for creativity support. In: Proceedings of the seventh ACM conference on creativity and cognition, C&C '09. ACM, New York, pp 127–136. doi:10.1145/1640233.1640255
5. Casares J, Long AC, Myers BA, Bhatnagar R, Stevens SM, Dabbish L, Yocum D, Corbett A (2002) Simplifying video editing using metadata. In: Proceedings of the 4th conference on designing interactive systems, DIS '02. ACM, New York, pp 157–166. doi:10.1145/778712.778737
6. Chandler G (2012) Cut by cut: editing your film or video, 2nd edn. Michael Wiese Produtions, Studio City
7. Cherry E, Latulipe C (2014) Quantifying the creativity support of digital tools through the creativity support index. ACM Trans Comput-Hum Interact 21(4):21:1–21:25. doi:10.1145/2617588
8. Dancyger K (2011) The technique of film and video editing: history, theory, and practice, 5th edn. Focal Press
9. Davis M (2003) Editing out video editing. MultiMedia IEEE 10(2):54–64. doi:10.1109/MMUL.2003.1195161
10. Diakopoulos N, Essa I (2006) Videotater: an approach for pen-based digital video segmentation and tagging. In: Proceedings of the 19th annual ACM symposium on User interface software and technology, UIST '06. ACM, New York, pp 221–224. doi:10.1145/1166253.1166287
11. Girgensohn A, Boreczky J, Chiu P, Doherty J, Foote J, Golovchinsky G, Uchihashi S, Wilcox L (2000) A semi-automatic approach to home video editing. In: Proceedings of the 13th annual ACM symposium on user interface software and technology, UIST '00. ACM, New York, pp 81–89. doi:10.1145/354401.354415
12. Goldman DB, Curless B, Salesin D, Seitz SM (2006) Schematic storyboarding for video visualization and editing. ACM Trans Graph 25(3):862–871. doi:10.1145/1141911.1141967
13. Goldman DR (2007) A framework for video annotation, visualization, and interaction. Ph.D. thesis, University of Washington
14. Hinckley K, Chen XA, Benko H (2013) Motion and context sensing techniques for pen computing. In: Proceedings of the 2013 graphics interface conference, GI '13. Canadian Information Processing Society, Toronto, pp 71–78. http://dl.acm.org/citation.cfm?id=2532129.2532143
15. Hinckley K, Yatani K, Pahud M, Coddington N, Rodenhouse J, Wilson A, Benko H, Buxton B (2010) Pen + touch = new tools. In: Proceedings of the 23nd annual ACM symposium on user interface software and technology, UIST '10. ACM, New York, pp 27–36. doi:10.1145/1866029.1866036
16. Hürst W, Götz G (2008) Interface designs for pen-based mobile video browsing. In: Proceedings of the 7th ACM conference on designing interactive systems, DIS '08. ACM, New York, pp 395–404. doi:10.1145/1394445.1394488
17. Jokela T, Mäkelä K, Karukka M (2007) Empirical observations on video editing in the mobile context. In: Proceedings of the 4th international conference on mobile technology, applications, and systems and the 1st international symposium on Computer human interaction in mobile technology, Mobility '07. ACM, New York, pp 482–489. doi:10.1145/1378063.1378140
18. Lienhart R (1999) Abstracting home video automatically. In: Proceedings of the seventh ACM international conference on multimedia (Part 2), MULTIMEDIA '99. ACM, New York, pp 37–40. doi:10.1145/319878.319888
19. Liu S, Guimbretière F (2012) Flexaura: a flexible near-surface range sensor. In: Proceedings of the 25th annual ACM symposium on user interface software and technology, UIST '12. ACM, New York, pp 327–330. doi:10.1145/2380116.2380158
20. Mackay W, Pagani D (1994) Video mosaic: laying out time in a physical space. In: Proceedings of the second ACM international conference on multimedia, MULTIMEDIA '94. ACM, New York, pp 165–172. doi:10.1145/192593.192646

21. Mackay WE, Davenport G (1989) Virtual video editing in interactive multimedia applications. Commun ACM 32(7):802–810. doi:10.1145/65445.65447
22. Marshall CC (2010) Reading and writing the electronic book. Morgan and Claypool Publishers
23. Martin T (2012) Interactive quadruped animation. Master's thesis, University of, California, Santa Barbara
24. Meyer A (1995) Pen computing: a technology overview and a vision. ACM SIGCHI Bull 27(3):46–90. doi:10.1145/221296.221308
25. Nakanishi Y, Ishii Y, Koike H, Oka K, Sato Y (2004) Enhancedmovie: movie editing on an augmented desk as a large-sized display. In: Adjunct proceedings of the 17th annual ACM symposium on user interface software and technology, UIST Adjunct Proceedings '04. ACM. UIST ARCHIVE
26. Potel MJ, Sayre RE (1976) Interacting with the galatea film analysis system. ACM SIGGRAPH Comput Graph 10(2):52–59. doi:10.1145/965143.563285
27. Ramos G, Balakrishnan R (2003) Fluid interaction techniques for the control and annotation of digital video. In: Proceedings of the 16th annual ACM symposium on user interface software and technology, UIST '03. ACM, New York, pp 105–114. doi:10.1145/964696.964708
28. Ramos G, Balakrishnan R (2005) Zliding: fluid zooming and sliding for high precision parameter manipulation. In: Proceedings of the 18th annual ACM symposium on User interface software and technology, UIST '05. ACM, New York, pp 143–152. doi:10.1145/1095034.1095059
29. Ramos G, Boulos M, Balakrishnan R (2004) Pressure widgets. In: Proceedings of the SIGCHI conference on human factors in computing systems, CHI '04. ACM, New York, pp 487–494. doi:10.1145/985692.985754
30. Reiffel L (1971) Superimposed dynamic television display system, us patent 3617630
31. Rubine D (1991) Specifying gestures by example. ACM SIGGRAPH Comput Graph 25(4):329–337. doi:10.1145/127719.122753
32. Ryokai K, Marti S, Ishii H (2004) I/o brush: drawing with everyday objects as ink. In: Proceedings of the SIGCHI conference on Human factors in computing systems, CHI '04. ACM, New York, pp 303–310. doi:10.1145/985692.985731
33. Silva J, Cabral D, Fernandes C, Correia N (2012) Real-time annotation of video objects on tablet computers. In: Proceedings of the 11th international conference on mobile and ubiquitous multimedia, MUM '12. ACM, New York, pp 19:1–19:9. doi:10.1145/2406367.2406391
34. Song H, Benko H, Guimbretiere F, Izadi S, Cao X, Hinckley K (2011) Grips and gestures on a multi-touch pen. In: Proceedings of the SIGCHI conference on human factors in computing systems, CHI '11. ACM, New York, pp 1323–1332. doi:10.1145/1978942.1979138
35. Thompson R, Bowen CJ (2009) Grammar of the edit, 2nd edn. Focal Press, Burlington
36. Vaucelle C, Africano D, Davenport G, Wiberg M, Fjellstrom O (2005) Moving pictures: looking out/looking in. In: ACM SIGGRAPH 2005 educators program, SIGGRAPH '05. ACM, New York, pp 27:1–27:7. doi:10.1145/1187358.1187391
37. Vaucelle C, Davenport G (2004) A system to compose movies for cross-cultural storytelling: textable movie. In: Proceedings of the second international conference on technologies for interactive digital storytelling and entertainment, lecture notes in computer science, TIDSE 2004. Springer-Verlag, Berlin, pp 126–131
38. Viviani C, Baptiste M, Gili JA, Logette L, Sauvaget D et al (2011) Dictionnaire mondial du Cinèma. Larousse, Paris
39. Vogel D, Balakrishnan R (2010) Direct pen interaction with a conventional graphical user interface. Human-Comput Interact 25(4):324–388. doi:10.1080/07370024.2010.499839
40. Weda H, Campanella M (2007) Use study on a home video editing system. In: Proceedings of the 21st British HCI group annual conference on people and computers: HCI...but not as we know it - volume 2, BCS-HCI '07. British Computer Society, Swinton, pp 123–126. http://dl.acm.org/citation.cfm?id=1531407.1531439
41. Weher K, Poon A (1994) Marquee: a tool for real-time video logging. In: Proceedings of the SIGCHI conference on Human factors in computing systems: celebrating interdependence, CHI '94. ACM, New York, pp 58–64. doi:10.1145/191666.191697
42. Yeo BL, Yeung MM (1997) Retrieving and visualizing video. Commun ACM 40(12):43–52. doi:10.1145/265563.265571
43. Zigelbaum J, Horn MS, Shaer O, Jacob RJK (2007) The tangible video editor: collaborative video editing with active tokens. In: Proceedings of the 1st international conference on tangible and embedded interaction, TEI '07. ACM, New York, pp 43–46. doi:10.1145/1226969.1226978

Diogo Cabral is an Assistant Professor at University of Madeira and Researcher at Madeira-ITI/LARSyS, Portugal. His research work crosses the fields of Multimedia and Human-Computer Interaction. Previously, he worked as Postdoctoral Researcher at Helsinki Institute for Information Technology HIIT, University of Helsinki, Finland, where he lectured the Human-Computer Interaction course and collaborated in different research projects. He received the Ph.D. degree in Computer Science from Universidade Nova de Lisboa, Portugal, in 2014.

Nuno Correia is a Professor at Faculdade de Ciências e Tecnologia, Universidade Nova de Lisboa, Portugal, where he teaches and does research work in the areas of Multimedia and Interaction. He participated in the creation of a national Digital Media Ph.D. program that he coordinates at FCT/UNL and currently he also coordinates the Ph.D. program in Computer Science. He is national representative in IFIP TC14 (Entertainment Computing) since 2006. Nuno Correia participated in several national and European projects. He was a researcher at Interval Research Corporation, Palo Alto, California, in 1996/97. Previously, he was a researcher at INESC, Lisbon, Portugal. Nuno Correia supervised 6 doctoral theses and about 30 master theses already completed and currently supervises several doctoral and master dissertations. He is author or co-author of more than 90 publications in journals, conferences and books. He participated in the organization and was chair or co-chair of several international conferences (e.g., ACE, MobileHCI).