Bachelor Thesis

Development of a User Interface Concept for Viewing Temporal Video Annotations

Hochschule Harz, University of Applied Sciences
Department of Automation and Computer Science
Degree Program: Computer Science in Media (B.Sc.)

Date of Completion: September 7, 2010

Written By

Stefanie Müller
Computer Science in Media, Term 8
Matriculation Number: 12699

Fürst-Franz-Josef-Strasse 15, A201,
9490 Vaduz, Liechtenstein
s.mueller@public-files.de

First Supervisor

Prof. Holger Reckter
Hochschule Harz

Second Supervisor

Dr. Gregor Miller
University of British Columbia

Abstract

Video is a complex information space that requires advanced navigational aids for effective browsing. The increasing number of temporal video annotations offers new opportunities to provide video navigation according to a user's needs. We present a novel video browsing interface called TAV (Temporal Annotation Viewing) that provides the user with a visual overview of temporal video annotations. TAV enables the user to quickly determine the general content of a video, the location of scenes of interest and the type of annotations that are displayed while watching the video.

The visual overview of TAV consists of an enhanced video timeline with icons representing the location and content of each temporal video annotation. Icons provide more semantic information for navigation than other visual cues, such as colour stripes which are abstract and difficult to distinguish. However, icons also require more space on the video timeline. This soon becomes a problem with the increasing density of temporal video annotations throughout the video. Additionally, the emerging trend for consuming video on devices with small screens, such as mobile phones, motivates our research. We present a novel solution called SCADE (Scaling with Deformation) that adjusts the appearance of icons according to available screen real-estate while keeping the advantage of incorporating detailed semantic information.

In our user study, we evaluate the trade-off between detailed semantic information and limited screen real-estate by comparing the finding time of video scenes when icons or colour stripes are used as visual cues on the video timeline. Our results show that icons provide the user with more information to reason about the content of a video scene while having a finding time similar to colour stripes. However, with an increasing density on the video timeline the search with icons is still slower than with colour stripes.

Future work will include further development of TAV and SCADE to integrate the results from participants feedback, as well as an extended user study to evaluate the improvements.

Acknowledgements

First of all, I would like to thank Prof. Dr. Sidney Fels and Prof. Holger Reckter who one year ago set up a collaboration to exchange students between the University of British Columbia (UBC) in Vancouver, Canada, and the Hochschule Harz (HH) in Wernigerode, Germany. Looking back on the experience I made during my research term at UBC, I can say that this thesis would not have been possible in the same quality without the knowledge I gained in the past seven months.

It was a pleasure to work in such an inspiring environment and I would like to show my gratitude to my supervisor Gregor Miller who oversaw my work on this thesis and made available his support in a number of ways. His advice helped me a lot to find my way through the research process and I am proud to say that I learned many important things while working with him.

With the completion of this thesis, I finish my bachelor degree in Computer Science in Media at HH. It was an honor for me to be part of this extraordinary degree with its unique corporate identity and I owe my deepest gratitude to Prof. Martin Kreyßig, Prof. Jürgen Singer (Ph.D), Prof. Holger Reckter and Prof. Eberhard Högerle for all the energy they invest to further the development of their students.

I am also grateful that I had the chance to meet Nicolas d'Alessandro during my stay at UBC and to build this wonderful friendship with him. I would like to thank him for his unlimited confidence in my skills and competences and his availability during the difficult times I have had when writing this thesis.

I warmly thank my parents Siglinde and Jürgen Müller for their unconditional trust in my personal and professional development and their support throughout the last twentytwo years of my life. Without their love this thesis would not have been possible.

Finally, I want to thank the German Academic Exchange Service (DAAD) for providing financial support for my stay at UBC in the form of a scholarship.

Contents

Abbreviations

BREB	Behavioural Research Ethics Board
HCT	Human Communication Technology Laboratory
HH	Hochschule Harz
NTSC	National Television Systems Committee
OSD	On Screen Display
PAL	Phase-Alternation-Line
SCADE	Scaling With Deformation
TAV	Temporal Annotation Viewing
TCPS	Tri-Council Policy Statement: Ethical Conduct for Research Involving Humans
UBC	University of British Columbia
UIST	ACM Symposium on User Interface Software and Technology

List of Figures

Chapter 1

Introduction

Source: Google Maps [28]

1. Introduction

This thesis is the result of a research internship at the Human Communication Technology Laboratory of the University of British Columbia. In this introductory chapter, we give background information on the research internship, followed by a description of the research project and the user interface concept covered in this thesis. Finally, we explain the research process for developing the remaining part of this thesis.

1.1 Research Internship

Source: HH [30], UBC [40]

Figure 1.1: Research internship facilitation agreement made effective on August 11, 2009, between the Hochschule Harz in Germany and the University of British Columbia in Canada.

The work presented in this thesis is the result of a collaboration between the University of British Columbia (UBC) in Vancouver, Canada, and the Hochschule Harz (HH) in Wernigerode, Germany (Figure 1.1). Between both universities a research internship facilitation agreement was made effective on August 11, 2009. In this agreement, both universities ensure to provide exchange students with the opportunity to gain practical research experience in a research internship.

This thesis is the result of a first exchange between the degree program Computer Science in Media (B.Sc.) of HH and the Human Communication Technology Laboratory (HCT) of UBC which took place from February 2 to August 23, 2010 (Figure 1.2). The purpose of this exchange was to enhance the practical knowledge acquired in Computer Science in Media with the advanced research skills imparted at the HCT Laboratory.

Source: CS [26], HCT [29]

Figure 1.2: First exchange between the degree program Computer Science in Media („Medieninformatik“) of HH and the HCT Laboratory of UBC.

The degree program Computer Science in Media at the Department for Automation and Computer Science of HH educates undergraduate students in concept, design, realization and marketing of digital content for multimodal applications. The degree is divided into two equally weighted core areas. The first core area focuses on technical courses, such as mathematics, computer graphics and coding, whereas the second core area aims at imparting knowledge in creative courses, such as graphical design, photography, film, audio, animation and interface design.

The HCT Laboratory at the Department of Electrical and Computer Engineering of UBC researches issues concerning the use of information technology for realizing effective communication of human experience. Based on this, the laboratory tries to explore, create and develop new technologies that support communication in novel ways.

1.2 Research Project MyView

During the stay at the HCT Laboratory, we worked on the MyView project. Within the last years more than fourty researchers participated in the MyView project and developed the framework used for on-going research.

MyView is a system for the creation and playback of multiple streams of context aware[1] video data (Figure 1.3). The novel techniques developed in the MyView project are currently applied to an ice hockey broadcast event. Instead of watching the linear video sequence predefined by a broadcast director, all camera angles located around the hockey field are directly delivered to the viewer. Based on this, the viewer gains the ability to, for instance, switch between camera angles according to his personal interest and to select ice hockey players to retrieve additional information, such as player statistics.

SOURCE: MYVIEW [34]

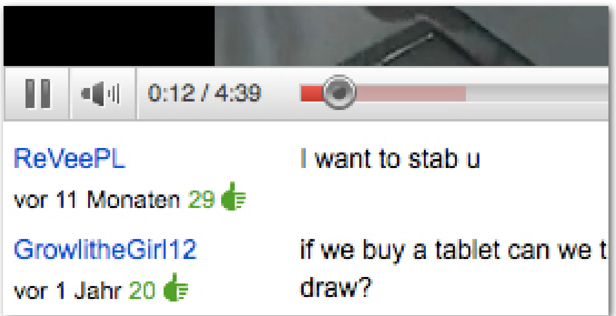
Figure 1.3: In the MyView project, we explore novel techniques for creating and interacting with video.

To provide the viewer with this interactive enriched video content, the MyView project contains three components: a multi-camera capture, both online and offline processing services for human tracking and identification and an interface for browsing in the resulting multi-video space[2].

In this thesis, we focus on the third component: the interface for browsing the context aware multi-video space which is called the VideoDiver. The VideoDiver allows the user to personalize his video experience by interacting with the video content. Within the last years, we developed three user interface concepts for the VideoDiver which are explained in detail in chapter 2. The goal of this thesis is to extend the current collection of user interface concepts by focussing on how temporal video annotation can support video navigation and video search according to a user's interests. In the following section, we introduce our work on temporal video annotation and motivate our research.

1.3 THESIS CONTEXT: TEMPORAL ANNOTATION VIEWING

Although video annotation is a growing phenomenon, the current approach for viewing and finding video annotations is still at a basic level. Even though many annotations refer to a specific subset (or scene) of the video and are therefore temporal in nature, the established approach does not take this characteristic into account. On popular video platforms, such as YouTube, video comments (a widely used type of annotation) are displayed in a single entry list that does not change during playback (Figure 1.4). However, we believe that this static approach is not appropriate for a time-based medium, such as video. We propose that video annotations are displayed simultaneously with the scene to which they refer.



SOURCE: YOUTUBE [42]

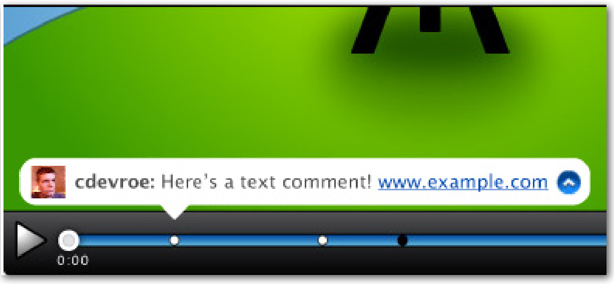
Figure 1.4: Static video comment list on YouTube that does not change during video playback.

Furthermore, enabling the user to add temporal video annotations provides new opportunities for browsing video. The static approach for visualizing video annotations cannot be easily used to provide navigational cues. However, with a temporal approach video anno-

1 A context aware system in the video domain acquires its context by collecting information from the video frames, by abstracting and understanding the content of the video frames and by applying behaviour based on the context in the video frames [47]. For instance, by counting how many scores a player made in a hockey game until a certain scene, the system knows that when the player is selected in this scene, information about player scores until this scene are requested.
2 A multi-video space is a collection of videos that cover a common context, such as the playing field of a hockey game.

tations can be used to support the user with additional information, such as the scene content and locations of interest. We believe that providing the user with a visual overview of temporal video annotations improves video navigation speed.

In addition to the problem of non-temporal video annotation, there is also a lack of filtering mechanisms. The increasing number of annotations requires an interface that enables the user to define which annotations are relevant. For example, only the annotations made by a specific person (such as a friend) or the annotations related to a specific event in the video (such as a goal in a sporting event).



Source: Viddler [15]

Figure 1.5: Colour cues showing the location of temporal video annotations.

In this thesis, we explore how visual cues positioned on a video timeline to represent the location and content of temporal video annotations support video navigation and video search. We believe that visual cues in the form of icons provide more effective navigational aids than the widely used colour stripes. We suggest to use icons due to the fact that their visual appearance provides more detailed semantic information than the commonly used colour stripes which are abstract and difficult to distinguish (Figure 1.5).

However, icons require more space on the video timeline which quickly becomes a problem with the increasing density of temporal video annotation throughout the video. Furthermore, the emerging trend for consuming video on devices with small screens, such as mobile phones, creates a need for a novel solution addressing the issue of limited screen real-estate.

In this thesis, we present a novel solution that scales icons with deformation to adjust their appearance according to the available screen real-estate while remaining the advantage of incorporating detailed semantic information.

1.4 Research Process For Thesis Development

To develop our novel solution, we first study existing work on the MyView project to gain a better understanding of recent issues in the field of personalized video browsing (chapter 2). Afterwards, we define our research field and review related literature to subsequently define our research question (chapter 3). Based on this, we derive our novel solution and describe our prototype implementation (chapter 4). We then explain how we prepare and conduct a user study to evaluate our novel solution (chapter 5). Finally, we state our conclusion and give an overview of future work (chapter 6).

Chapter 2

MyView Project

Source: MyView [34]

2. MyView Project

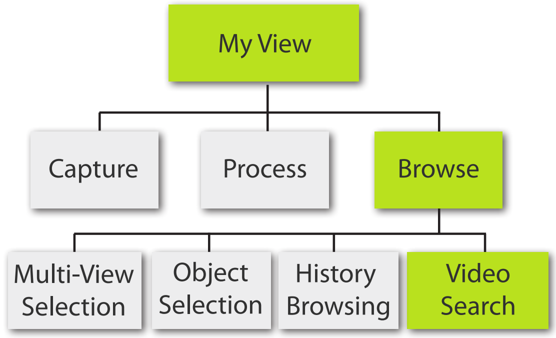
Our work is part of the MyView project. In the following sections, we list the components of the MyView system and motivate the project by giving reasons why personalized video experience is an emerging research topic. Finally, we describe the MyView video browsing interface called VideoDiver due to the fact that we set the focus of our thesis on this component.

2.1 Components of the MyView Project

MyView is a system for the creation and playback of multiple streams of context aware video data. In the MyView project we develop three components in parallel: the multi-camera capture, both online and offline processing services for human tracking and identification and an interface for browsing in the resulting multi-video space (Figure 2.1).

In this thesis, we focus on the third component: the interface for browsing the context aware multi-video space which is called the VideoDiver. The VideoDiver allows the user to personalize his video experience by interacting with the video content. Within the last years, we developed three user interface concepts for the VideoDiver: the multi-view navigation for switching between camera angles, the object selection for accessing metadata[3] and the history browsing for reproducing and sharing the personal video experience. The goal of this thesis is to extend the current collection of user interface concepts by focussing on how enhanced video timelines can be used as a navigational aid for personalized video browsing.

The following sections cover our motivation to create the VideoDiver and describe its three existing components to lay out the foundation for further discussion of the novel user interface concept we develop in this thesis.



Source: Own Figure

Figure 2.1: MyView project overview. In this thesis we focus on the browsing interface and develop a new interface concept for video search.

2.2 Motivation: Personalized Video Experience

Since television became popular in the early 50's, watching video has been mostly a passive experience. People settled down in front of a TV screen to consume video in the way it was produced by the director. Video was watched as a linear playback from the start to the end without enabling the user to interact with the video content.

Although the technical progress in computers and the availability of high speed internet access provide new opportunities for interacting with video, the viewing experience still follows the established television paradigms. When watching video on popular sites, such as YouTube, the user is still only provided with traditional video transport controls, such as fast-forward, rewind and pause/play, but additional functionality to customize the viewing experience is missing.

We believe that video browsing is a highly personal experience and that new interface concepts for browsing video are required. For instance, when watching an ice hockey game a user might be interested in a certain player and wants to see only the camera angle that focuses on the player. However, in the established approach the decision which camera angle is shown in the video is made by the

3 Metadata provides information about the content of a video. For example, a single video frame may include metadata that describes what is seen in the video frame or may include technical data, such as the frame resolution and colour depth. Metadata is especially useful in video since information about the content of a video frame is not easily accessible by a computer.

director who chooses the best view for the entire audience. Enabling the user to choose the camera angle allows him to browse the video more efficiently according to his wants. Additionally, the user might be interested in accessing information that is related to the video content. For instance, the user wants to see how many goals a certain player scored until the current playback time of the video. Enabling the user to access this information by interacting with the video content will enrich the user's viewing experience.

Creating new interaction paradigms for accessing video data in a personalized way will take the user's viewing experience to the next level. Video will no longer be watched as a linear sequence of images, but will be navigated through in multiple ways, enabling each user to have his own view on the video data.

2.3 VideoDiver: A Multi-Video Browsing Interface

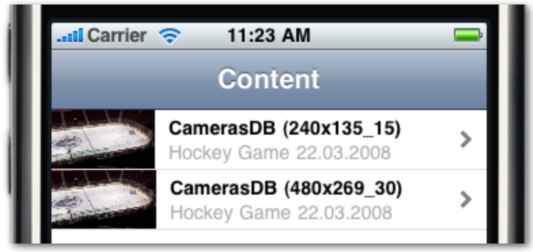
To address the lack of interaction with video content, we develop the VideoDiver, a novel browsing interface that enables the user to navigate a context aware multi-video space according to his interests.

We build a proof-of-concept-implementation of the VideoDiver on the iPhone by reason of an emerging trend for consuming video on mobile platforms. However, the multi-touch control of current mobile devices has significant differences compared to the traditional keyboard/mouse input as known from personal computers.

While developing the VideoDiver, we focus on the use of predefined gestures and built-in elements of the iPhone because users are already aware of their functionality. We adapt them for the VideoDiver whenever possible and invent new interaction paradigms only where no available gestures or built-in-elements are appropriate.

In the following sections, we describe the initial video content and camera selection when launching the VideoDiver and explain the user interface.

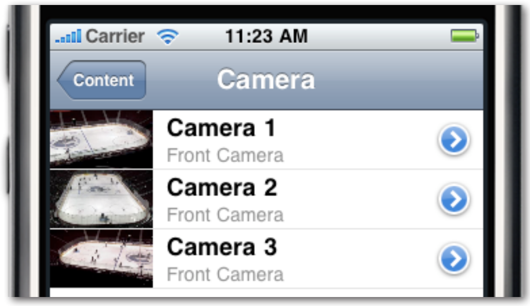
2.3.1 Selecting Video Content and Camera



Source: MyView [34]

Figure 2.2: Selection of a video content.

After launching the VideoDiver, the user is presented with an interface that displays the available video content (Figure 2.2). For test purposes, we install video content from an ice hockey game in high and low resolution.



Source: MyView [34]

Figure 2.3: Selection of a camera angle.

After selecting a content, the user enters a second screen that shows which camera angles are available for his previous content choice (Figure 2.3).

Currently, the user has to select the camera angle manually. However, we plan to integrate a search algorithm that automatically provides the user with the best camera angle, for instance, on a certain hockey player.

Figure 2.4: The VideoDiver loading screen after selecting a content and a camera.

After selecting the camera angle, the VideoDiver starts the content loading as described in figure 2.4 and displays the video browsing interface.

2.3.2 Enhanced Traditional Video Player

Figure 2.5: VideoDiver Interface: 1) Video timeline, 2) VideoGrid, 3) Rewind, 4) Play/Pause, 5) Fast-Forward, 6) VideoFlow, 7) Volume Slider

The VideoDiver user interface is built on top of a traditional video player. We use the iPhone built-in video player as a template to add the functionality for browsing a context-aware multi-video space (Figure 2.6). The iPhone built-in video player is used for watching non-interactive video. However, for our advanced video browsing interface, we can adapt several user interface elements and the look and feel to provide the user with a common user interface that facilitates orientation.

Figure 2.6: The iPhone built-in video player is used as a template for the VideoDiver.

Like the built-in video player, the VideoDiver displays the video content in fullscreen to completely immerse the user in the viewing experience. By tapping the screen once, an on-screen display (OSD) is shown that contains a scrubber bar with time labels on top of the screen and the traditional video transport controls play/pause, fast-forward and rewind in the middle of the screen (Figure 2.5). However, this functionality is not sufficient for interactive video in which a user can choose his preferred camera angle and can access metadata embedded in the video content. The VideoDiver enhances the OSD with new interactive elements for the multi-view selection (Figure 2.5: VideoGrid, VideoFlow) and enables the user to directly interact with the video content on the screen, for instance, by selecting objects to view additional information.

In the following sections, we explain the three components of the VideoDiver that allow the user to browse the context-aware multi-video space and demonstrate how they contribute to a personalized video experience.

2.3.3 Multi-View Selection

In multi-camera broadcast events, the broadcast manager usually selects the most appropriate camera angle for the entire audience. However, the director's choice becomes inappropriate when a user is interested in specific parts of the video and thus, the video content selected for the entire audience does not fit his personal interests. We believe that for these cases the role of the broadcast manager needs to be extended to the single user, enabling him to select the best view according to his own preferences.

Giving the user access to a multi-video space requires new interaction paradigms to effectively browse the collection of related videos. All camera angles in a multi-video space cover a shared physical space, for instance, the playing field of a hockey game, but have different spatial areas and temporal boundaries. Displaying the spatial and temporal boundaries of camera angles in an interactive overview provides the user with the most effective choice according to his personal interests.

In the following sections, we present our two multi-view selection interfaces called VideoFlow and VideoGrid. Futhermore, we present VideoSwiping as an interaction method for switching between adjacent camera angles and show how automatical view selection via search criteria can support the user in his personalized video experience.

VideoFlow

Source: MyView [34]

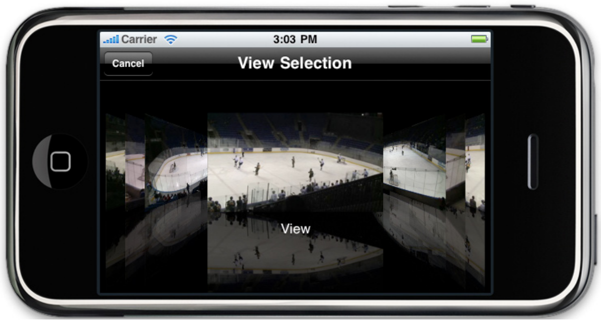


Figure 2.7: VideoFlow displays camera angles according to their horizontal spatial relations.

In VideoFlow, a spatial overview of all camera locations is presented to the user (Figure 2.7). When activated, the currently watched camera angle is centered on the screen. The adjacent cameras are arranged either on the left or the right side according to their spatial positions to the currently active camera. This allows the user to understand the spatial arrangement of multiple camera angles covering a shared physical space.

VideoFlow is based on the CoverFlow functionality of the iPhone built-in iPod application and adapts its gestures (Figure 2.8). Dragging with the finger on the screen moves the overview to the left or right, a single tap on one of the videos activates it and causes the VideoDiver to return to the video player. We adjust the static representation of images as known from CoverFlow to animated video representations for the VideoFlow to fit the time-based nature of video.

Source: CoverFlow [25]

Figure 2.8: CoverFlow from the iPhone built-in application iPod is adapted for the VideoFlow functionality of the VideoDiver.

While VideoFlow provides the user with a common interface for browsing through a collection of content representations, it has several limitations when adapted for spatial representation purposes. Currently, VideoFlow only works for camera angles that are horizontally aligned while a concept for browsing the camera angles vertically is in progress. Furthermore, VideoFlow only shows which camera angles are adjacent, but does not provide the user with information on the spatial distance between camera angles. As known from CoverFlow, VideoFlow only shows a small number of camera angles and their relations at once. However, it might be necessary for the user to make his choice based on a complete overview of all spatial relations between camera angles. Finally, the user has to skim through VideoFlow to see all available camera angles which is not appropriate for a high number of cameras angle.

VideoFlow is activated by tapping the outer right button on the OSD (Figure 2.5). The activation of VideoFlow results in leaving the context of the video player due to limited screen real-estate on the iPhone. However, the active camera angle continues playing in the VideoFlow interface.

VideoGrid

Source: MyView [34]

Figure 2.9: VideoGrid displays camera angles according to their horizontal and vertical spatial relations.

VideoGrid presents all available camera angles in a matrix display mode (Figure 2.9). Camera angles are sorted in rows and columns enabling the user to make his choice based on the overview. VideoGrid is used to display the horizontal relations between camera angles by arranging them in rows and the vertical relations by arranging them in columns. However, VideoGrid currently does not work when camera angles are not equally distributed among the physical space they share. Furthermore, camera representations, as seen in figure 2.9, have to have a certain size in order to provide the user with information on the video content. Due to the required size of camera representations and the limited screen real-estate, a high number of cameras angles causes scrolling of the VideoGrid. Future work will examine filter and search mechanisms to show only the camera angles that correspond to a user's interest.

VideoGrid is activated by tapping the outer left button on the OSD (Figure 2.5). As seen in VideoFlow, the camera overview of VideoGrid is displayed on a seperate screen due to limited screen real-estate on the iPhone.

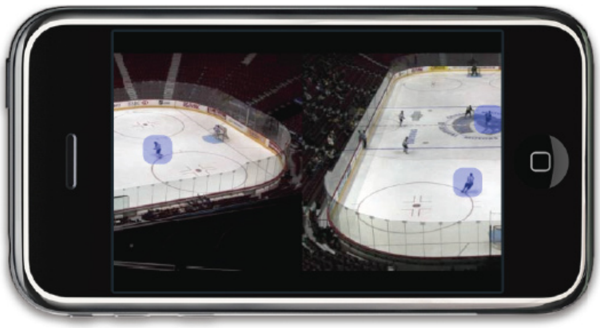
Source: Safari [39]

Figure 2.10: Safari user interface for selecting favorite bookmarks is adapted for the VideoGrid functionality of the VideoDiver.

VideoGrid is based on the bookmarks overview functionality of the Safari webbrowser (Figure 2.10). Dragging with the finger on the VideoGrid moves the overview to the left and right, respectively up and down. A single tap on one of the videos representations activates the corresponding video and causes the VideoDiver to return to the video player.

VideoSwiping

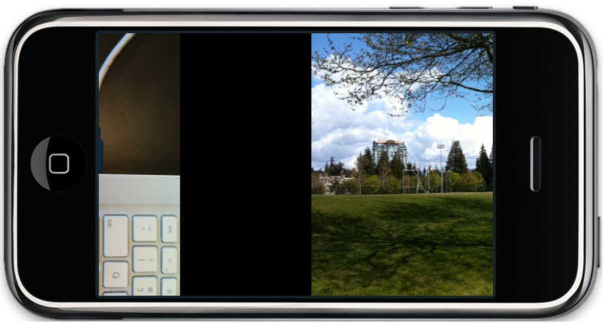
Beside the option to choose one camera angle from the set of all available cameras, we also enable the user to switch between adjacent camera angles without leaving the video player.



Source: MyView [34]

Figure 2.11: The swipe gesture allows the user to switch between adjacent camera angles.

VideoSwiping (Figure 2.11) is based on the swiping gesture of the iPhone built-in Photo application (Figure 2.12). Dragging to the left on the screen pushes the current camera angle out of the video window and displays the camera angle adjacent to the right. Dragging to the right results in displaying the camera angle adjacent to the left. This allows the user to step through each camera angle while remaining spatial and temporal context.



Source: MyView [34]

Figure 2.12: The swipe gesture is adapted from the iPhone built-in application Photos.

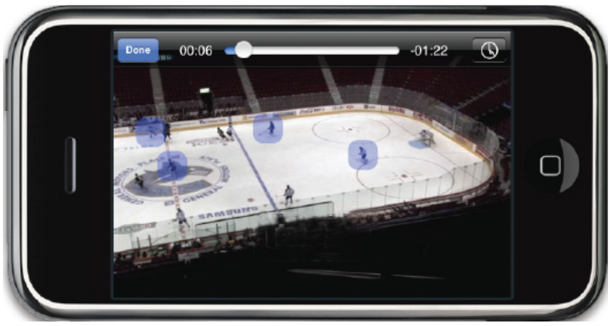
View Selection Via Search

Beside these options to directly determine which camera angle is shown in the video player, there is also a need for an indirect approach. Users might be interested in, for instance, a certain player or a specific event during an ice hockey game. The effort of selecting the best camera angle manually can be avoided by applying a search depending on the user's interest. This becomes especially helpful with an increasing set of available camera angles which is difficult to explore manually.

The indirect search automatically provides the user with the best camera angle according to his wants by self-directed switching, for instance, to the camera angle that has the best quality of view on a certain ice hockey player.

We believe that the combination of direct and indirect selection of camera angles provides the user with new opportunities for accessing a multi-video space.

2.3.4 Metadata Access Through Object Selection



Source: MyView [34]

Figure 2.13: Object selection with bounding boxes, for instance, ice hockey players on the playing field.

Metadata contains additional information about the underlying video content, such as player statistics and hyperlinks to related websites or videos. In the following sections, we describe how metadata can be accessed through the selection of objects within a video.

Definition of Objects

Objects are recognizable named regions within a video scene, such as hockey players, referees, coaches, the puck or advertisement. They are displayed as interactive areas which enable the user to define his selection. The selection of objects can be used to access additional information, for instance, selecting a player results in the player's statistics being displayed right next to him. Selected objects can also be used to build up a hyperlink structure between camera angles. For example, if a selected hockey player leaves the current camera angle, the video player automatically switches to the next camera angle with a better view point.

Unique Object Identifier

To identify an object in each video frame and across multiple camera angles, a unique identifier for each object is required. The identifier enables the object to serve as an active hyperlink between videos and as a criteria for search.

Object Selection

Objects in a video can be static or moving by nature. Static objects, such as advertisement banners on the sport field sideline, can be easily selected due to the fact that they maintain their spatial location in the video frame. However, moving objects, such as hockey players, are difficult to select because their movements are non-predictable and often fast. Furthermore, hockey players tend to overlap each other when trying to reach the puck which makes effective selection difficult. The current approach to select a moving target, for instance, as seen in computer games, requires the user to chase a target while it is moving on the screen.

However, this requires the user to predict the movement of the object and to click while it is in motion. We develop a novel interaction technique that enables the user to temporarily pause the video and to make his selection afterwards while the object is no longer in motion. Following this, the user can restart the video.

Results of our user study show that the common chase technique is faster with large and slow objects, whereas the pause approach is advantageous for small and fast moving targets.

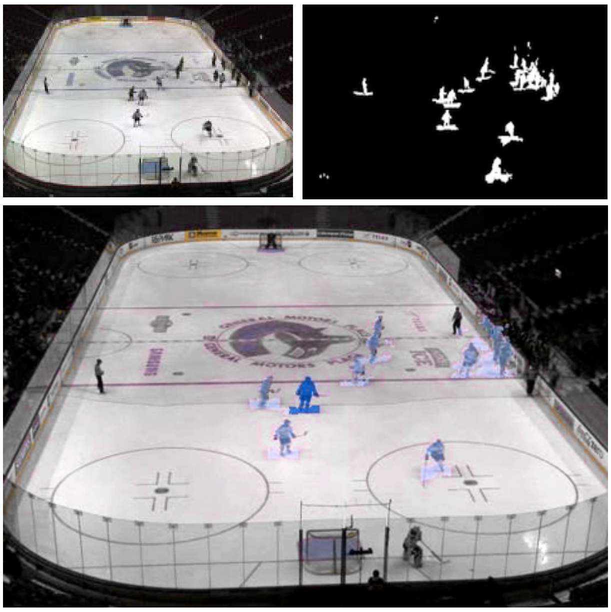
Resulting Action From Object Selection

There are several open issues for object selection that we address in our work. First, it is not obvious what the resulting action from selecting an object is. For instance, a selection can imply that the user wants to see additional information for the object or it can express that the camera angle should adjust automatically according to the best view for this object. We explore ways how the user can be informed about the available options and how he can make an effective selection. For this, we need to consider that the video might continue playing while the user makes his selection and that every interface element on the screen hides the underlying video content.

Displaying Object Selection State

Furthermore, we examine how the selection state of objects can be displayed. A common approach to enable object selection is to use bounding boxes that are drawn as rectangular shapes around an object (Figure 2.13). However, bounding boxes do not fit the shape of a selected object, such as an ice hockey player. When the distance between objects decreases, the imprecise object outline caused by a bounding box results in an increased error rate due to overlapping effects.

In the MyView project, we use background subtraction masks for each frame that enable us to display the active selection of a hockey player as described in figure 2.14. This approach avoids overlapping areas that appear when using bounding boxes. However, colouring a player's body results in a modified visual appearance of him on the screen. We are currently exploring other ways to display selection states of objects.



Source: MyView [34]

Figure 2.14: Regular video frame (top left), object subtraction mask (top right), coloured player shapes after background subtraction (below).

In conclusion, we believe that enabling the user to access metadata in a multi-video space through selectable content creates new ways for browsing video and enriches the user's viewing experience.

2.3.5 History browsing

Rewinding a video to a point of interest is an essential functionality of a video player. With non-interactive video, rewinding means to go back along the video content the broadcast director provided for the viewer. However, when video is no longer watched in a linear way it is less clear what the definition of rewinding the video is. In the following sections, we show how a history record helps the user to keep track of his video experience and how sharing history records creates a new motivation for social exchange.

The Experio

When video is no longer watched in a linear way, a record of the pathway a user takes is required to represent his visual experience. Every time a user interacts with the video, for instance, by switching camera angles or by selecting objects on the screen, the interaction is added to a history record in order to make the viewing experience traceable. Based on this, we call the record of a user's browsing structure through a video space an Experio.

Spatial-Temporal Record

Each Experio is a spatio-temporal record similar to those found in video editing software, such as Avid Express Pro [45] or Final Cut Pro [48] (Figure 2.15).

Source: Final Cut Pro [27]

Figure 2.15: Video timelines in the video editing software Final Cut Pro.

While browsing a video space with the VideoDiver the user acts as a video editor choosing camera angles from a list of sources and the temporal intervals in which they are displayed on the screen. The Experio is therefore a non-linear edit decision list containing multiple video sequences that are defined by start time and duration. The representation of a video sequence in the history record is called a video clip. Each video clip is colour-coded according to the corresponding camera angle and contains a thumbnail of the start frame.

Time-based and Node-based Display Mode

Source: MyView [34]

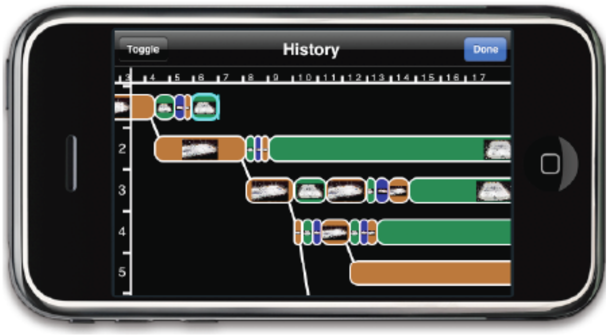


Figure 2.16: History Browser in time-mode. The video clip width is determined by the temporal duration a camera angle is watched.

We develop two display modes for the history record which are called the time-mode and the node-mode. In the time-mode (Figure 2.16) the horizontal axis corresponds to the playback time of the video. Each video clip width represents the duration the user watched the corresponding camera. However, when camera angles are changed often, the video clip width is very small. Although the user is provided with zooming functionality, it is difficult for him to recognize the thumbnail that represents the underlying video content.

Source: MyView [34]

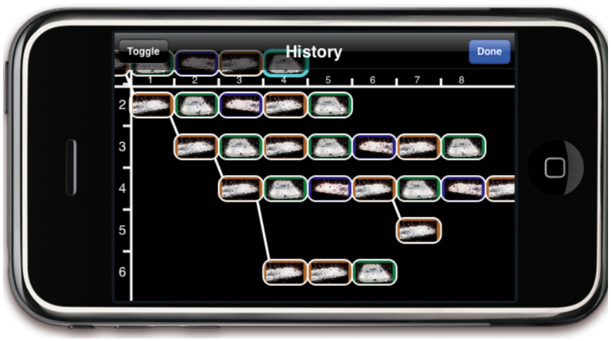


Figure 2.17: History Browser in node-mode. All video clips have an equal width.

Therefore, we develop a second display mode called node-mode (Figure 2.17). In the node-mode all video clips have the same width and thus, thumbnails can be easily viewed even when video clips are watched only for a short time. However, in this mode the user cannot determine the duration a video clip was watched.

There are several open issues with the history record that we address in our future work for the MyView project. First, the use of one thumbnail per video clip is not appropriate to represent a video sequence. Currently, we use the first frame of the video sequence as a representation of the underlying video content. However, the part of the video sequence for which the user is looking for, such as a goal in a hockey game, is often not represented appropriately by the first frame due to the fact that it happened later in the video clip.

For future work, we suggest to use the thumbnail as a mini video player which shows the video sequence according to the position of the playhead in the history record. Furthermore, thumbnails of one camera angle can look very similar throughout the video, for instance, a camera showing the playing field of a hockey game from a long shot. For these camera angles it might be more appropriate to focus on the part of the video frame that is changing, in this example the moving hockey players.

Recording Multiple Experios

Every time a user rewinds an existing Experio to a certain point and starts another interaction, a new branch is added to the history record. The additional branch provides space for recording a new Experio while maintaining the existing ones. When the video is watched several times while interacting differently, the history record gets the shape of a tree with each branch containing one Experio. However, when new branches are added, the history tree becomes more and more difficult to access. When the number of rows in the history tree grows, filter and search functionality is needed in order to maintain fast access of the history record.

Reproducing and Sharing

Derived from a user's interactions, the history record enables the user to reproduce his viewing experience. It allows the user to rewind his interaction path through the video to, for instance, find a more appropriate camera angle according to his interests. Furthermore, the user can share the history record on social networking sites by making his personal viewing experience available to his connections. When several users interact with the same video, their Experios, created while watching the video, can be grouped together. These collections of Experios allow users to see how the video was watched by others and which people have the same interests.

2.3.6 Video Search With Temporal Annotation

We have shown how the MyView project contributes to a personalized video experience by exploring ways for creating and browsing context-aware multi-video spaces. We presented the VideoDiver, a novel browsing interface that is built as a proof-of-concept implementation on the iPhone. We demonstrated how multi-view selection, object selection and history browsing enable the user to switch between different camera angles, to access metadata and to reproduce and share his personal video experience. In this thesis, we explore a fourth way of personalizing a user's video browsing experience by focusing on how temporal video annotation and enhanced video timelines provide navigational aids for video search and navigation.

Chapter 3

Research Question

Source: Schoeffmann et al. [13]

3. Research Question

In addition to the three existing components of the VideoDiver described in the previous chapter, we aim at developing a user interface concept for video search and navigation. Our concept allows the user to quickly locate scenes according to his interests. In the following sections, we describe our research field, discuss related work and state the research question.

3.1 Research Field Video Navigation and Search

Source: iPod Player [32]

Figure 3.1: Traditional video player without any search functionality. The user is dependent on the scrubber bar to locate a scene of interest.

Searching for a particular scene within a video is a common activity. However, traditional video players do not provide the user with supporting functionality to find a scene of interest. In traditional video players, users rely on the scrubber bar for rewinding and fast-forwarding the video to a scene of interest, but the scrubber bar only contains information about the current playback time of the video and its overall duration (Figure 3.1). When a user cannot memorize the temporal location of the scene of interest, searching with a scrubber bar quickly becomes a frustrating and time-consuming task.

Novel video players include functionality to address the search and navigation problem. Most of them focus on the textual approach by using keyword-input into a search engine [6, 7, 8]. While this is useful in many cases, the textual approach is not appropriate when the user's vocabulary for the search request does not match the keywords that are attached to the searched scene. Furthermore, the textual approach requires the user to first make an input into a search engine to be afterwards provided with navigational cues for a scene location.

In their paper about recent advances and challenges of semantic video search, Chang et al. [3] emphasize that users need interfaces to solve search tasks visually beside the textual approach. They highlight that a lot of work has been done to successfully enhance the keyword based search, but that there is still a lack of effective visualization of video data. According to Chang et al., visualization interfaces are a critical component for task completion and task efficiency when searching video scenes and therefore, one of the main challenges for future work in the field of video search and navigation is to enhance visual cues.

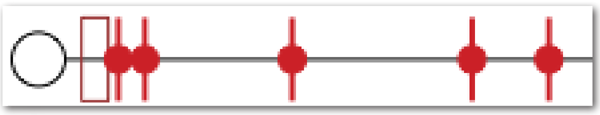
Work in the field of visual video navigation has either enhanced existing user interface elements, such as the video timeline and the video window, or invented new user interface concepts. In this thesis, we focus on the existing user interface element video timeline by reason that users are already familiar with this video player component and novel enhancements can be easier understood in the context of the existing functionality. In this chapter, we discuss work from the field of visual video navigation with enhanced timelines, list remaining open issues and state our research question.

3.2 Related Work on Enhanced Video Timelines

In the following section, we discuss existing work from the field of visual video navigation with enhanced timelines. We begin with user interface concepts using single-colour cues and show their limits in representing detailed semantic information[4]. Following this, we explain how multi-coloured representations have helped to overcome this problem by mapping semantic information onto a colour range. Finally, we have a look on how icons can be used as a more intuitive way of representing the underlying video content. We conclude with an evaluation of visually enhanced timelines and show how filtering visual cues supports the user in navigating video according to his interests.

3.2.1 Single-Colour Semantics

eduKEN [20] and Viddler.com [15] are two examples how video timelines with single-colour semantic cues[5] can provide the user with information on the location of scenes of interest.

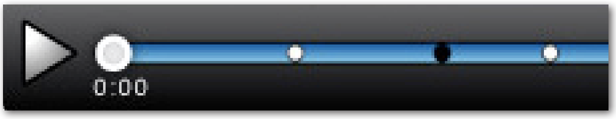


Source: eduKEN [20]

Figure 3.2: eduKEN video timeline with single-colour marks shows the temporal location of video comments.

eduKEN is a video player that displays the position of temporal video comments as single-colour marks in an enriched timeline (Figure 3.2). A user can easily locate scenes of interest with temporal video comments by navigating along the red marks on the video player timeline.

However, single-colour semantic cues only show where video comments are made, but do not provide any information on the underlying video content. A comment may refer to a specific event in a video, such as a goal in a sport game, or may refer to a specific object, such as a hockey player.



Source: Viddler [15]

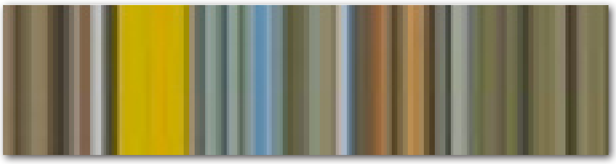
Figure 3.3: The video timeline on Viddler.com shows the location of user comments (white dots) and the location of tags (black dots).

On Viddler.com different colour marks are used to differentiate between types of annotations. As described in figure 3.3, white dots in the timeline are associated with user comments, black dots represent tags[6]. As seen with eduKEN, these single-colour semantic cues do not help the user to identify an annotation by its underlying video content.

3.2.2 Multi-Colour Semantics: Low-Level Features

Most work that visualized the underlying video content in more detail to provide navigational cues comes from the area of low-level feature extraction, such as extracting dominant colour [1], sound [1] and motion [13, 14].

Dominant Colour



Source: Barbieri et al. [1]

Figure 3.4: The ColorBrowser video timeline provides the user with simple navigational cues based on the dominant colour of the video frame.

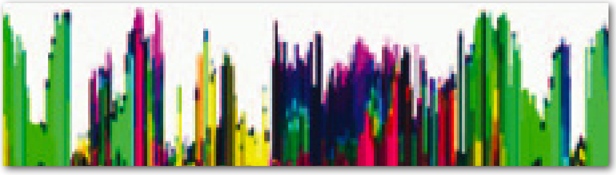
Barbieri et al. [1] extract the dominant colour and sound volume from a video and map the data onto a colour range. Afterwards, Barbieri et al. embedd the colour stripes for each video scene in a video timeline to provide the user with additional information while browsing the video (Figure 3.4).

By conducting experiments with the ColorBrowser, Barbieri et al. show that even simple multi-colour semantic cues, such as the do-

4 In this thesis, semantic information is understood as the information a human gains by interpreting the content of a video, such as recognizing a certain hockey player instead of defining him as a collection of coloured pixels.
5 A semantic cue is a visual mark that represents the semantic information of a video scene.
6 In the context of video, a tag is a keyword or term added to a video scene [52], such as the tag „goal" assigned to a video scene where a hockey team scores. A tag is a kind of metadata.

minant colour of a video frame, can provide useful navigational aids for video browsing when embedded in a timeline. For instance, during a soccer match the dominant colour on the screen is normally green, but during the breaks the colour changes by reason that interviews are conducted and the current scores are discussed.

Motion



Source: Schoeffmann et al. [13]

Figure 3.5: Schoeffman et. al. introduce a video timeline showing motion direction and motion intensity based on the underlying video content.

Schoeffmann et al. [13, 14] use motion to visualize the underlying video content. They extract motion direction and motion intensity from the video data and use colour and brightness to display the information in a timeline (Figure 3.5).

While it is difficult to identify a single video scene based on the motion of its objects, the visualization of motion helps to identify scenes that are similar to a prototype scene. According to Schoeffmann et al., it is easy for users to find similar scenes because they have the same colour representation in the visually enhanced timeline. For instance, in ski flying the athletes are the fastest moving video objects. Comparing the colour representation of one ski-jump enables the user to identify ski-jumps of competiting athletes with similar speed.

Furthermore, users are able to identify potential scenes of interest, for instance, by knowing that scenes with fast motion are usually thrilling and that these are associated with a red colour in the timeline. With this approach, Schoeffmann et al. try to link the video high-level semantics[7] with the low-level feature extraction of motion.

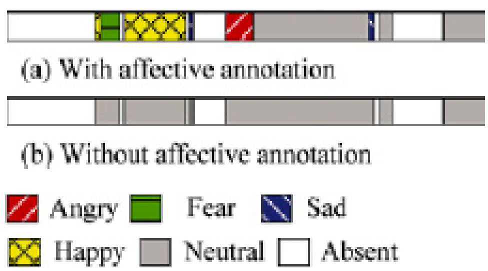
Limits of Low-Level Feature-Extraction

According to Chang et al. [3], one of the main challenges in the field of video navigation is to close the semantic gap between the video data and human's interpretation. Low-level feature extraction based on, for instance, dominant colour, sound and motion is useful for finding similar scenes, but it is in most cases not suitable for representing the underlying video content as it is understood by the viewer, for instance, in the form of characters and events in a story. In the next section, we discuss how high-level feature extraction based on interpreting the underlying video content can be used to provide more meaningful navigational cues for the user.

3.2.3 Multi-Colour Semantics: High Level Features

High-level feature extraction, such as detecting faces to extract human emotion, is still not at a level to create reliable data from the video bit stream. Therefore, work that focuses on interpreting the video content presented in the following section makes either use of manual annotation or focuses on textual sources, such as temporal user comments, to extract information about the video content.

Emotion



Source: Chen et al. [2]

Figure 3.6: The EmoPlayer video timeline displays character emotions by mapping them onto a colour range.

7 High level semantic information is created by interpreting the video content from a human perspective, for instance, by associating facial expressions of people with emotions. Low level semantic information only contains the information a computer gained by analyzing the video frame, for instance, that a persons facial expression changes, but without further interpretation of this gesture.

Chen et al. [2] develop EmoPlayer, a video player that displays character emotions by using a multi-colour timeline. The user chooses a character from a drop down box and views his emotions along the timeline while watching the video.

As described in figure 3.6, different colours in the timeline represent different types of emotions of the selected character. The absent of colour in the timeline indicates that the selected character is not present in the scene. At time, EmoPlayer uses manual annotation due to the lack of techniques to automatically extract character emotions from the video data. However, Chen et al. describe how computer vision[8], speech processing and natural language processing will allow to take over the manual effort in future.

In a user study, Chen et al. compare their interface with affective annotation to an interface that only displays a character's appearance in the video. Their results show that providing more semantic information significantly increases video navigation speed, especially when subjects are not familiar with the video content.

Trustworthiness

Source: Nakamua et al. [12]

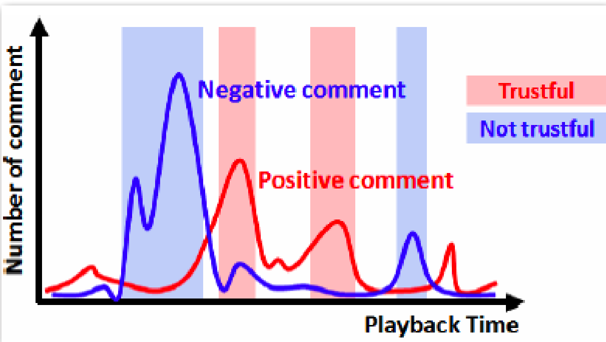


Figure 3.7: Nakamua et al. introduce a video timeline that displays the trustworthiness of video scenes based on extracting positive and negative words from temporal user comments.

A second example how navigational cues can be provided using the underlying video content is based on user comments from a social video site. Nakamua et al. [12] analyse temporal video comments to provide the user with a visual overview of the trustworthiness of video scenes. They extract keywords from the user comments and compare them to a dictionary that assigns words into either a positive (trustworthy) or negative (untrustworthy) category.

In their prototype, Nakamua et al. use an enhanced timeline with two display modes as described in figure 3.7. First, the timeline contains graphs for each positive and negative changes in user comments. Second, these two graphs are evaluated by the system and summarized in a bar chart that shows if a video scene is rated overall trustworthy or not.

The graph interface is one example how timelines cannot only be used to provide qualitative navigational cues, but can also support quantitative search. A graph based interface enables the user to see both the type of information related to a video scene and the level of strength of this information. For instance, if we apply this technique to the previously discussed EmoPlayer, the timeline could not only show that a character is sad by displaying a blue colour stripe, but also show the level of sadness by adjusting the height of the blue colour stripe. However, this approach raises the question how to measure the levels of information in such complex contexts like emotions.

The quantitative overview of trustworthiness displayed by using two graphs also influences the user's viewing behaviour. Users are able to see which scenes of a video are rated least trustworthy and therefore, access the video scene to examine the cause.

8 „A branch of artificial intelligence and image processing concerned with computer processing of images from the real world. Computer vision typically requires a combination of low level image processing to enhance the image quality (e.g. remove noise, increase contrast) and higher level pattern recognition and image understanding to recognise features present in the image." [46]

3.2.4 Iconic Semantics

All of the work described previously focuses on using colours to enhance the video timeline. However, colours are abstract representations of the underlying video content and users usually do not intuitively associate a video scene with a specific colour.

To close the semantic gap between colour cues and the underlying video content, Moraveji et al. [11] use a tool-tip with icon and text descriptor that is displayed when the cursor is placed over the timeline (Figure 3.8).

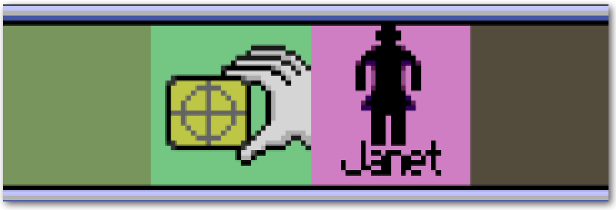
Source: Moraveji et al. [11]

Figure 3.8: A tool-tip with icon and text descriptor closes the semantic gap between the video content and its abstract colour representation.

The tool-tip provides information on the meaning of the colour stripe, but requires additional effort to access this information. Furthermore, the tool-tip only shows the explanation for a single colour stripe at once.

Davis et al. [5] develop a user interface concept with icon based semantic cues in order to create more meaningful representations for video content. Their iconic language contains over 3500 iconic primitives for object and action description. The icon primitives can be combined into icon sets and later be placed onto a video timeline.

In order to close the gap between the general abstract information that an icon delivers, Davis et al. enable the user to add titles to the icons to be more specific about their meanings. For instance, an icon that shows the abstract shape of a person can be enriched with a textual caption giving the name of the person (Figure 3.9). The additional information provided by the text descriptor enables the user to differentiate between visually similar things, but also reduces the effort necessary to create the icons and keeps the icon library at an affordable size.



Source: Davis et al. [5]

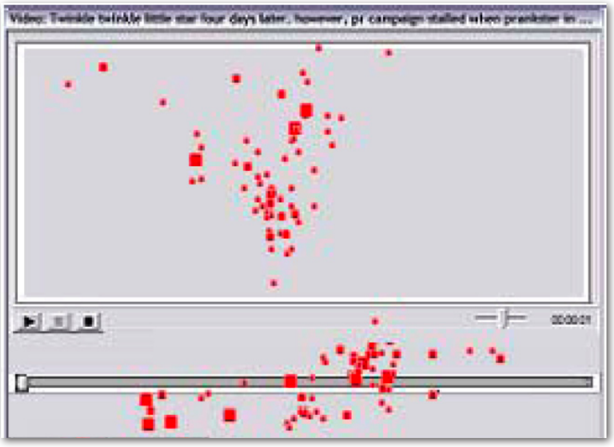
Figure 3.9: Adding text-based information to icons allows to be more specific about their meanings.

However, Davis et al. state that it is generally difficult to find a descriptive language for a temporal media since it mainly contains actions that are hard to describe with a single icon. As a solution, Davis et al. suggest to use animated icons to describe actions. However, this makes reading the video representation more complicated since the user can no longer have a brief look on the iconic description, but has to wait until the icon animation has finished.

3.2.5 Evaluating Visually Enhanced Video Timelines

The usefulness of enhanced video timelines is proven by Moraveji et al. [11] who conduct an eye-tracking experiment. For their eye-tracking experiment, Moraveji et al. use eye fixation as a unit of measurement.

The results from the eye-tracking experiment show that users pay more attention to the timeline when it embedds visual cues. While this might be seen as an indicator for distraction, the results of measuring the average number of eye fixations needed to solve a task provide evidence that task completeness and task efficiency increases with the visually enhanced timeline.



Source: Moraveji et al. [11]

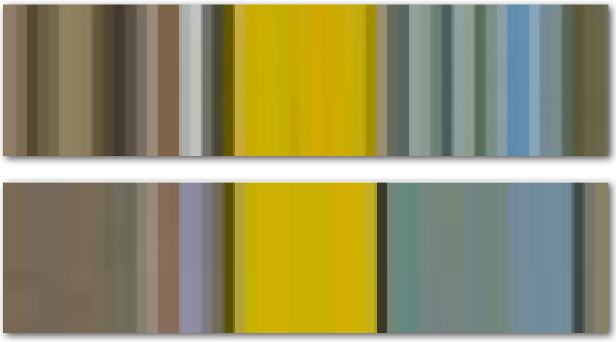
Figure 3.10: Results of an eye-tracking experiment by measuring eye-fixation to evaluate the usefulness of enhanced video timelines.

However, the measurements show that task efficiency is directly related to the distribution of visual cues along the timeline. When visual cues are unique, task efficiency increases significantly more than with common visual cues that appear in regular distances along the timeline.

These results are in agreement with Chen et al.'s statements after conducting user studies with the EmoPlayer. Frequency and distribution of one type of emotion on the timeline have significant impact on how fast a scene is located. Thus, with increasing appearance and density the video navigation speed decreases.

For this reason, the number of different types of visual cues in the enhanced timeline is an important key to video scene navigation speed. A coarse-grained range of visual cues results in the same colour stripe appearing often along the timeline which makes it difficult for the user to identify the searched scene. A fine-grained range of visual cues provides unique semantic identifiers, but results in lots of similar colour stripes among the timeline due to a limited range of distinct colours.

In the previously discussed ColorBrowser, Barbieri et al. smoothed their visual cues to reduce the number of different types to provide the user with visually distinct colours as described in figure 3.11.



Source: Barbieri et al. [1]

Figure 3.11: Colour representation with fine-grained colour range (top) and after smoothing (below) to provide the user with visually distinct colour stripes.

However, calculating the average always results in a loss of detailed information. It is therefore necessary to define how many different visual cues the timeline includes to provide unique semantic information on each video scene while remaining to be easily identifiable due to the distinct visual appearance.

3.2.6 Filtering Visual Cues

According to Costa et al. [4], users have different interests when watching a video. Some users might be interested in characters of a movie and others in technical features. It is therefore not necessary to display all available visual cues, but necessary to provide the user with an interface that adjusts according to the user's specific requirements.

This coincides with Chen et al. findings while conducting a user study with the EmoPlayer. The results show that the retrieval time of video scenes is influenced by two points: in the first step, whether a target could be related to the general content type of a timeline (a character) and in the second step whether the target could be related to a specific visual cue on the timeline (an emotion). When the user is not able to associate the video scene with its visual representation on the timeline, the visual cues are not helpful to access the video content. It is therefore essential to define the content type of an enhanced timeline according to a user's wants when navigating video.

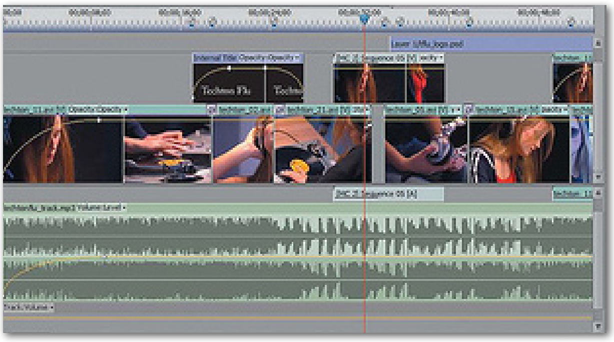
In the EmoPlayer user study, participants are provided with a drop down box to choose the visual cues presented on the video timeline. However, adjusting the content of a single timeline by using a drop down box requires additional effort. According to Chen et al., it is of more value to display different kinds of visual cues in multiple timelines to give the user quick access to the information. Furthermore, having multiple timelines allows the user to understand relations between multiple types of visual cues. For instance, displaying the emotions of each character on a separate timelines enables the user to compare the emotional changes among characters and helps the user to reason coherence which is important to comprehend the structure of a video clip.

Costa et al. design a user interface with multiple timelines to address the issue of adjusting the visual cues according to a user's personal interests. Costa et al.'s video player, shown in figure 3.12, allows the user to have multiple perspectives over the same video data. The interface consists of multiple timelines each corresponding to a specific view on the video, for instance, one showing the transcript of the video and another showing the script of the director. While watching the video, the user determines which additional information is displayed alongside with the audio visual signal of the video scene.

Source: Costa et al. [4]

Figure 3.12: Users select one or multiple timelines (bottom) to watch additional information displayed in small screens (top) according to their interests.

The interface concept of Costa et al. refers to the timeline model of common video editing tools, such as Adobe Premiere [44] or Adobe After Effects [43] (Figure 3.13).

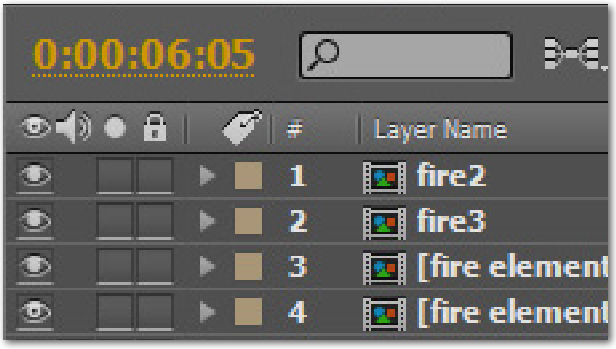


Source: Adobe Premiere [23]

Figure 3.13: Multiple video timelines in the video editing software Adobe Premiere.

In those video editing tools, the content of the video is determined by a composition of multiple video timelines each adding data to the video. For instance, one video timeline contains the video data and another one the text captions that are displayed alongside with the video. Seeing both video timelines vertically aligned enables the editor to determine the relation between the appearance of the text captions and the video data. For instance, when interviewing athletes in the break of a sport game, every time a new interview starts the name of the athlete is displayed.

The editor defines which video timelines influence the final video composition by activating and deactivating the video timelines (Figure 3.14). For instance, deactivating the video timeline with the text captions results in displaying only the audio visual signal of the video.



Source: Adobe After Effects [22]

Figure 3.14: In the video editing software Adobe After Effects, the editor determines which video content is displayed by activating and deactivating the video timelines via the eye-icon on the left side.

However, these techniques are currently only available in video-editing software and are not included in video players on popular social video sites, such as YouTube.

3.2.7 Video Semantics From User Comments

To broaden our knowledge on how enhanced video timelines can be used with temporal video annotations, we also refer to recent literature that contains information on how temporal video annotation, especially in the form of user comments, is organized in social video sharing systems [9, 19]. To maintain the narrow focus of this thesis we will not explain the literature in detail, but recommend to read how video comments can be used to extract video semantics [18] for video categorization [10], video ranking [17], video search [21] and video scene extraction [16].

3.3 Research Problem
Semantic Information vs. LimitedScreen Space

Most of the work described in the previous section has focused on using colour stripes as a representational means to provide the user with visual cues on the underlying video content. However, we believe that there are several disadvantages when using colour-based semantic cues.

First of all, there is a semantic gap between the video content and its colour representation. Users usually do not associate colour with a specific moment of a video. They may remember a goal in a sport video, but they do not intuitively associate this information with a colour. Based on this, the choice which colours are mapped onto which semantic information when creating the visual cues remains subjective and not generally valid. To address this problem, a legend with additional textual information is provided in most cases to explain the meaning of each colour to the user. However, users may want to access the information of visual cues without the additional effort of referring to a second source, such as a legend.

By reason that colour stripes are abstract representations of the underlying video content, users need to learn the meanings of the colours when using the interface for the first time. However, when the interface is not used on a regular basis users may not be able to memorize the meaning of the colours and thus increased error rates resulting from missing or wrong associations may appear. Therefore, a more intuitive visual representation of the underlying video content is required.

Furthermore, there is only a limited range of distinct colours available. Since each semantic cue has to have a unique representation, it quickly becomes difficult to differentiate between colours when the number of types of semantic cues increases. It is therefore necessary to provide the user with more enhanced visual cues that enable him to quickly identify different types of semantic information.

Since colour stripes are abstract, non-intuitive representations of the underlying video content that quickly come to their limits when the the number of types of semantic information increases, we suggest to enhance video timelines with icons. We believe that it is more valuable to provide the user with icons embedded in the timeline since they provide the user with more detailed semantic information on the underlying video content. For instance, instead of using a black colour stripe representing a goal in a hockey game, we suggest to use a puck icon as described in figure 3.15. The user is no longer required to learn the meaning of the colour, but can immediately access the semantic information of the underlying video content by seeing the icon on the timeline.

Source: Own Figure

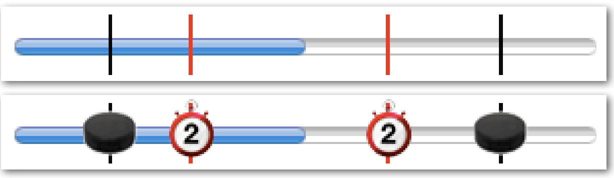
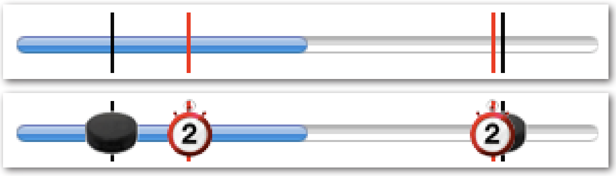


Figure 3.15: Enhanced video timelines with colour stripes (top) are less intuitive and provide less semantic information on the underlying video content then those with icons (bottom).

However, providing the user with more detailed semantic information in the form of icons also requires more space on the timeline. Icons have to have a certain size to be recognizable which quickly becomes a problem when adjacent icons are placed in small distances and start to overlap (Figure 3.16).



Source: Own Figure

Figure 3.16: When visual cues appear in small temporal distances, the semantic information of colour stripes remains accessible while icons start to overlap due to limited screen real-estate.

The problem of limited screen real-estate becomes even more significant due to an emerging trend for consuming video on mobile devices. The screen resolution of mobile devices, such as the iPhone 3G, is limited to 320px to 480px. On such a low resolution screen, space is an important source that when not used effectively makes accessing the detailed semantic information of icons difficult.

In this thesis, we investigate the trade-off between visual cues showing more detailed semantic information and the limited screen real-estate. While colour stripes give only little semantic information about the underlying video content, they also need just a single pixel in width to be recognizable by the user. On the other hand, icons give more detailed semantic information on the underlying video content, but also need more space on the video timeline. In our work, we examine how the enhanced timeline can provide the user with detailed semantic information on the underlying video content even when there is only little space available.

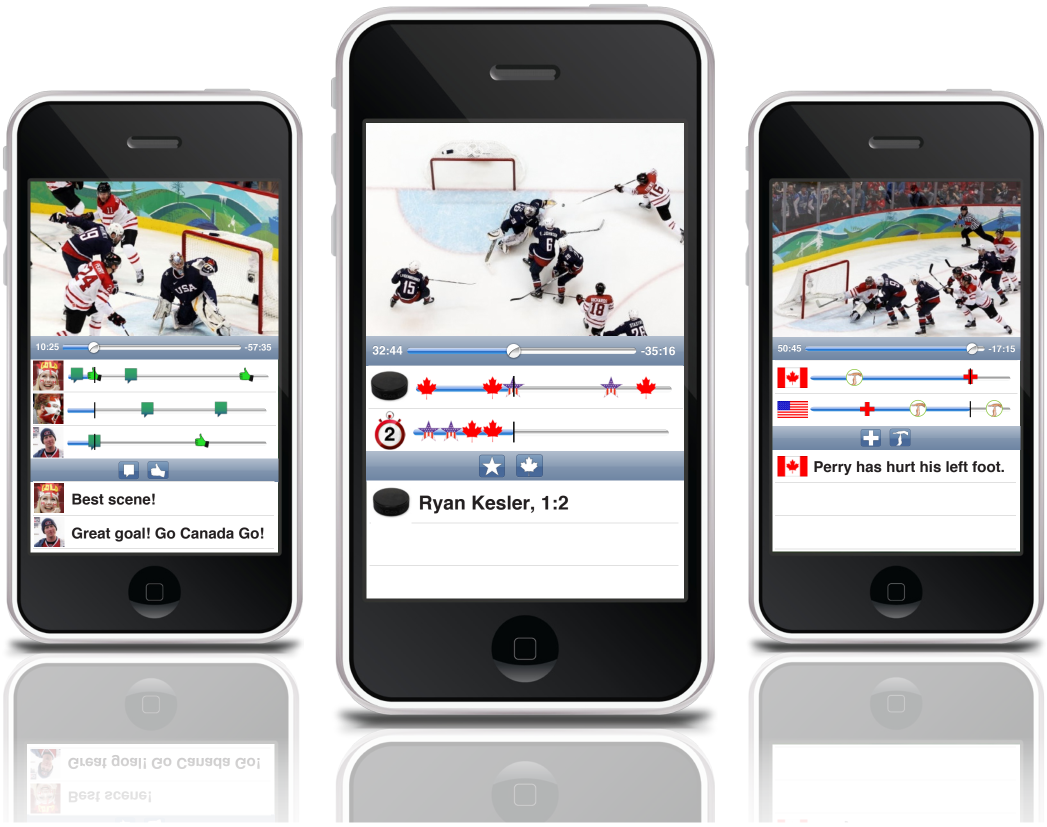
The results of our work will contribute to the field of visual video navigation with enhanced timelines by providing a solution for video scene retrieval with detailed semantic cues on small screens. Our work is especially relevant due to the on-going behavioral change of consuming video.

With the technical development of mobile devices and the progress in providing fast wireless internet access, users can nowadays consume video whereever and whenever they want. Mobile devices are more and more used as on the way video players, but due to external disruptions the time spent for one video session is short. Our work will contribute to access video scenes of interest faster by providing the user with detailed semantic cues even when the screen real-estate is very limited.

The usage of video has changed from a mainly passive video experience, using only traditional video elements to control the video playback, to more enhanced interactions, such as annotating video scenes with temporal tags and comments. Every year more users annotate video scenes while watching a video on social video sites, such as YouTube. Since these temporal video annotations refer to the underlying video scene, they can be used as detailed semantic cues for navigating the video. However, current approaches either do not visualize the temporal video annotation in an enhanced timeline or provide only single-colour semantic cues. Our work creates value of the increasing number of temporal video annotation by using them to provide visual cues for navigating video. In the following chapter, we discuss our solution for detailed semantic cues on small screens with limited screen real-estate and present our prototype.

Chapter 4

Solution and Prototype



Source: MyView [34]

4. Solution and Prototype

In the following section, we present our novel solution for the research problem and describe our prototype implementation that is later used for evaluating the solution by conducting a user study. First, we derive our novel solution by comparing and contrasting different approaches for displaying semantic visual cues on a limited screen real-estate. Afterwards, we show how we integrate our solution into the VideoDiver.

4.1 Solution: Adjustable Icon Representations

Iconic visual cues contain more detailed semantic information on the underlying video content than colour stripes. However, when the distance between icons is smaller than their width and overlapping appears, an effective approach is needed to determine how the visual information of icons can remain accessible for the user. In the following sections, we present three different approaches and describe their advantages and disadvantages for navigating video to find scenes of interest.

4.1.1 Scaling Icons Without Deformation

Scaling icons without deformation is an effective approach to use the available space on the video timeline according to the distribution of visual cues. As described in figure 4.1, when adjacent icons are placed close to each other their size is scaled down according to the available space. A video timeline with few icons that are equally distributed results in displaying the icons in a large size. If icons are distributed among a certain scene of the video and only little space is available, the smaller icons are used to provide the user with non-overlapping visual cues.

Figure 4.1: Large icons are used when enough screen space is available (top), however when icons start to overlap (middle), icons are scaled down to avoid overlapping (bottom).

Distinct Versions For Each Icon Size

However, it is not appropriate to only reduce the resolution of an icon when its size is decreased. Instead, distinct versions for every icon size are required in order to provide the user with the most detailed semantic information according to the available screen real-estate (Figure 4.2). Shrinking the rich, detailed high resolution icon to a smaller size is not a suitable approach since the visual elements of the icon are often no longer recognizable. On the other hand, if the low-resolution version with little semantic details is created first and afterwards scaled up, the opportunity to provide the user with more fine-grained semantic information is lost.

This approach of providing the user with differently designed icons is applied, for instance, to the application icons on the iPhone. To increase the quality and usability of applications, developers are required to provide versions of the application icon in various sizes. For instance, the application icon for the app store[9] requires a high resolution icon with rich semantic details to attract the user (512x512px), but a version with lower resolution is used for the home screen on the device (114x114px) and for the search results (58x58px).

9 The app store is Apple's sales platform for distributing the applications for their mobile devices. Each application has its own app store page that contains a high resolution image for advertisement purposes and information about the application.

Figure 4.2: Different versions of icons are used according to the available screen space. Larger icons are more detailed and richer in texture.

However, displaying detailed semantic information does no longer work when the size of icons is strongly reduced due to the fact that simple visual representations replace the complex shape and textural appearance of the original icons. When small temporal distances between the visual cues on the video timeline appear, the information becomes difficult to access for the user (Figure 4.3).

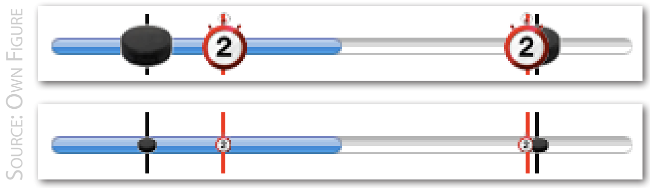


Figure 4.3: When the size of the icons is scaled from large icons (top) to small icons (bottom) the semantic information becomes difficult to access.

Local and Global Scaling

We consider two approaches called local and global scaling that describe how a group of icons on the video timeline adjusts their size according to the available screen real-estate (Figure 4.4). With local scaling, the icon size is adjusted according to the available space of adjacent icons. For small distances adjacent icons are scaled down to avoid overlapping, for larger distances adjacent icons keep their size. Based on this, the accessibility of semantic information represented by icons is related to the available screen real-estate which makes local scaling an effective approach.

However, local scaling of icons creates the impression that video scenes with larger icons are more important than others. An approach that adresses this problem is global scaling which scales all icons according to the smallest distance between two adjacent icons on the entire video timeline (Figure 4.4). However, global scaling results in a reduction of the level of detail of icons on the entire timeline even when more screen space is available.

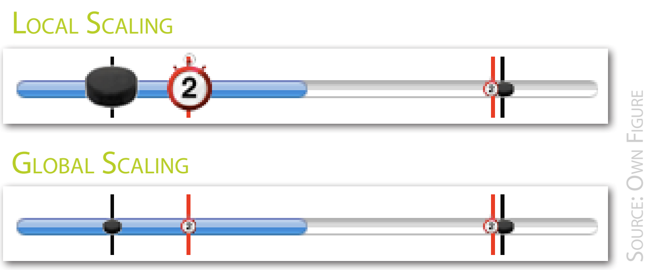


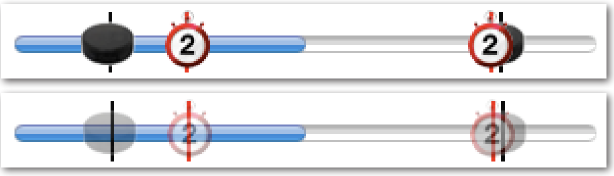
Figure 4.4: The size of icons can be defined by the space available between adjacent icons (local scaling, top) or by the smallest distance on the entire timeline (global scaling, bottom).

In conclusion, scaling icons without deformation in either local or global mode is useful to provide the user with the most detailed semantic information according to the available space between adjacent icons on the timeline. While this approach is suitable for most levels of distribution along a timeline, it is not appropriate when the space between adjacent icons is strongly reduced and icons become very small. Even though this problem can be solved by providing the user with zooming functionality, it requires additional effort by the user. In our work, we aim at developing a user interface concept for video timelines in which the detailed semantic information of visual cues is accessible at first sight.

4.1.2 Overlapping Icons With Transparency

When icons start to overlap, a part of their visual information is no longer accessible by the user. Overlapping reduces the amount of semantic information that can be represented by an icon. Based on this, users are no longer able to interprete the semantic information of an icon when a part of the icon is hidden since in some cases icons are only recognizable when seen as a whole.

The problem of hiding the visual information of an icon when not enough screen space is available can be addressed by adding transparency. When applied, the video timeline shows the visual appearance of both adjacent icons layered at the overlapping section as described in figure 4.5.



Source: Own Figure

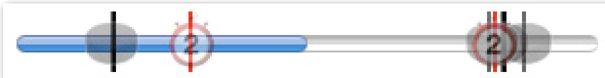
Figure 4.5: When icons overlap, parts of the icons are no longer visible (top). Adding transparency makes the hidden part visible (bottom).

Accessing Information

While the transparency approach is helpful to partwise regain the previously hidden information, it also has several disadvantages regarding the access of semantic information. First of all, layering the visual appearance of multiple icons results in both icons being no longer completely accessible. For instance, when layered colours or textural patterns of icons complement in an unexpected way, the transparency can either destroy existing shapes or create new ones. This becomes especially significant when multiple icons are placed in small distances (Figure 4.6).

Interaction On The Timeline

The colour and texture merging that appears when multiple icons are layered cannot only visually confuse the user, but also has impact on the interaction with the video timeline. The visual appearance of two layered icons implies that both are equally close to the foreground which raises the question which icon is selected when an interaction with the layered section is made.



Source: Own Figure

Figure 4.6: Multiple icons layered with 40% transparency. The combination of colours and textures of different icons can cause visual confusion.

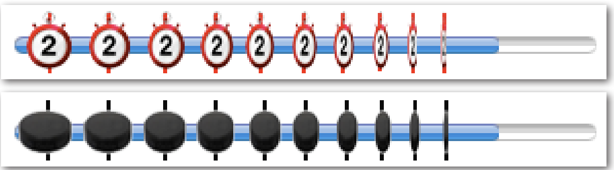
In conclusion, adding transparency to otherwise overlapping icons is helpful to partwise regain the hidden information. With the transparency approach, icons can keep their original size and scaling does not need to be applied. However, layering visual information of icons can easily cause confusion, especially when multiple icons are placed in small distances.

4.1.3 Scaling Icons With Deformation (SCADE)

We believe that the solution for the trade-off between detailed semantic information and limited screen real-estate has to take into account the characteristics of both existing approaches colour stripes and icons. The previously discussed solution statements for scaling without deformation and overlapping with transparency are not suitable due to the fact that they reinvent new display modes instead of merging the advantages of both colour stripes and icons. In the following section, we present a novel solution called SCADE (Scaling With Deformation) that combines the small width of colour stripes with the detailed semantic information of icons according to the available screen real-estate.

Deforming Icons to Colour Stripes

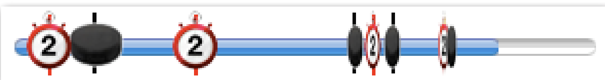
With SCADE the visual appearance of icons is adjusted according to the available space on the video timeline. When icons are placed in large distances to each other, they are displayed in full size providing the user with the most detailed semantic information. However, when adjacent icons are placed in small distances, their aspect ratio changes by scaling the width of the icon and keeping the height as described in figure 4.7.



Source: Own Figure

Figure 4.7: Different display states of an icon from the original size (left) to a single pixel in width mimicing a colour stripe (right).

Scaling the icons in their width while maintaining their height draws them near to the visual appearance of the colour stripes. While the shape of the icon gets lost when being deformed, a part of the textural pattern and the colour remains. Based on this, the deformed icon always displays the most detailed semantic information according to the available screen space (Figure 4.8).



Source: Own Figure

Figure 4.8: A video timeline using SCADE for displaying icons based on the available screen space.

Maintaining The Semantic Context

Furthermore, the different display states of the deformed icon enable the user to maintain the context of the semantic information when screen real-estate decreases. We believe that the user is able to draw conclusions from the original shape and colour of the icon to the deformed versions. Following this, scaling the icons with deformation allows the user to access detailed semantic information even when the icons do not have their original size.

While the space-saving colour stripes cannot be extended to show more detailed semantic information when more screen space is available, we found a solution to use the more detailed semantic information of space-consuming icons for every degree of available screen space. Based on this, we believe that our novel approach SCADE is advantageous in any video browsing situation compared to the colour stripes.

4.2 Prototype: Temporal Annotation Viewing

To evaluate SCADE as an effective approach for adjusting icons according to available screen space, we develop a user interface concept for temporal annotation viewing that integrates SCADE on its video timeline. TAV (Temporal Annotation Viewing) displays the location and content of temporal video annotations by using visual cues on multiple video timelines. The TAV user interface, an example of which is shown in figure 4.9, consists of a traditional video player (1), multiple video timelines (2) with visual cues (3), a filter toolbar (4) and an annotation viewing interface (5). In the following sections, we describe each of these components and explain how they contribute to a personalized video experience.

4.2.1 Video Player

The TAV user interface consists of a traditional video player vertically adjacent to our novel browsing interface for visual cues. Based on the built-in video player of the iPhone, the traditional video transport controls, such as play, pause, rewind and fast-forward are placed on an OSD. Selecting the screen fades the OSD in to make the transport controls accessible for the user. After a few seconds of not being used, the OSD fades out to provide the user with the best view on the video content. Furthermore, the OSD contains a scrubber bar with time labels to show the current temporal location in the video and the remaining duration.

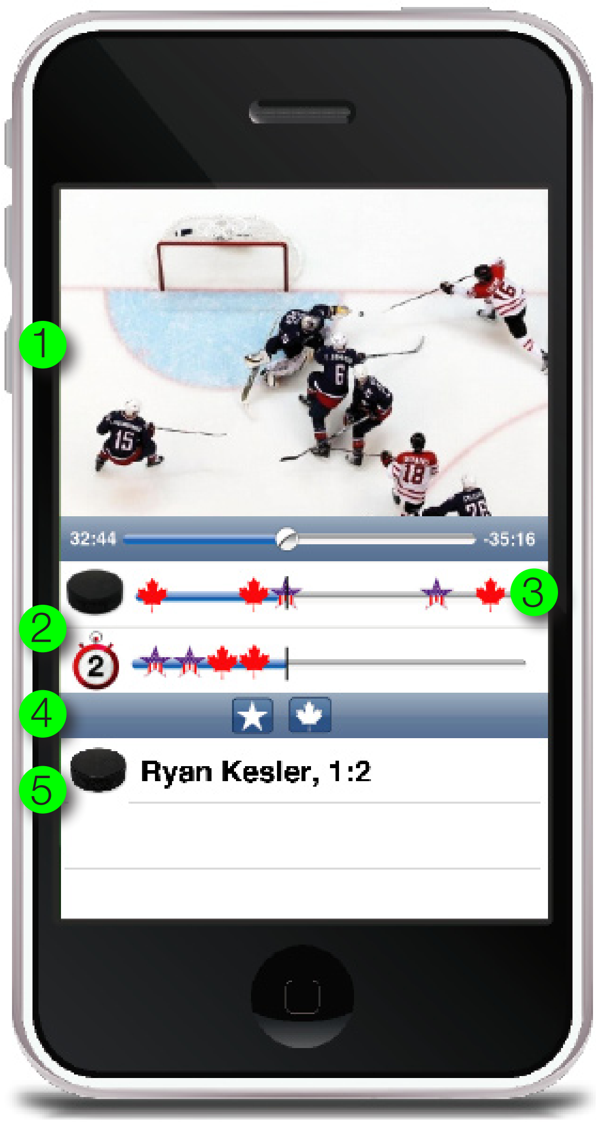


Figure 4.9: TAV User Interface, 1) Video Player, 2) Video Timelines, 3) Annotations/Visual Cues, 4) Annotation Filter, 5) Annotation Viewing

4.2.2 Annotation Timelines

Below the traditional video player, multiple timelines are displayed. Each timeline refers to a specific type of video annotation. In the case shown in figure 4.9, the types of video annotation are goals and two minute penalties in an ice hockey game. Each timeline consists of a visual identifier on the left side followed by a scrubber bar with multiple icons. Annotations are located according to their temporal appearance in the video and can be of different types. Our example shows icons representing Canada and USA in a hockey game.

Based on the usage of multiple video timelines with annotations sorted by type, more detailed information for video navigation is provided. The user cannot only see where goals occurred in the video, but also which team scored the goal.

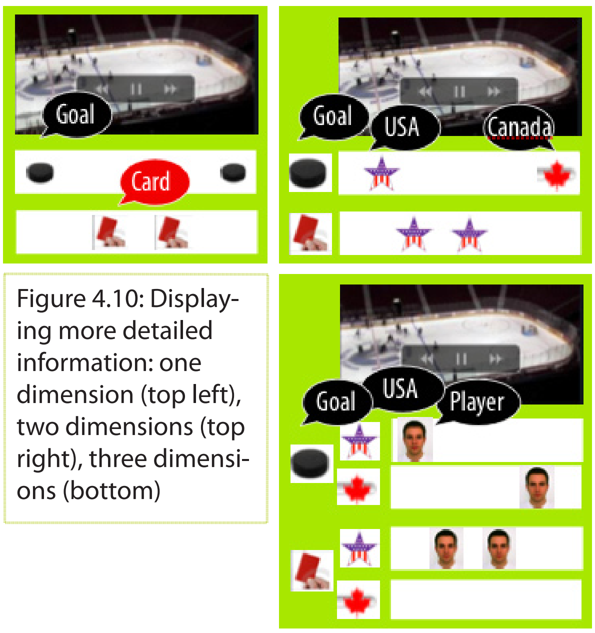


Figure 4.10: Displaying more detailed information: one dimension (top left), two dimensions (top right), three dimensions (bottom)

We define the level of detailed information provided on a video timeline in dimensions (Figure 4.10). Showing at which time goals are scored in a video is a simple one-dimensional information. Displaying on a video timeline at which time goals are scored in a video and which team made the goal is called a two-dimensional information. Adding more detailed information, such as which player scored the goal of a certain team, increases the number of dimensions to three. However, each additional level of information displayed as a visual cue requires screen space which quickly becomes a problem, especially on the small screens of mobile devices.

4.2.3 Annotation Filtering

When the number of visual cues on the video timeline increases, effective filter mechanisms are necessary to provide the user with only the information that is of interest for him. In order to provide the user with the ability to adjust the visual overview according to his wants, we integrate two basic filter mechanisms.

A filter toolbar (Figure 4.9) enables the user to determine which types of annotations are displayed on the video timeline. For instance, if a user is only interested in goals and penalties referring to Canada, he can deactivate USA's annotations by selecting the related filter button.

A second filter method can be applied by deactivating a timeline. For instance, if a user is only interested in goals but not in penalties, he can deactivate the second timeline by selecting it on the left side.

The two filter mechanisms described above are still at a basic level and allow the user only to choose between a limited range of available options. As soon as the number of video timelines and the number of different types of annotations increases, the user is no longer able to see all available options at once due to limited screen real-estate.

Further development of the prototype includes more advanced filter mechanisms such as a keyword based search which provides the user with only the corresponding video timelines and visual cues for his personalized video browsing experience.

4.2.4 Annotation Viewing

While the video is playing, the video annotations are shown in the viewing table (Figure 4.9). This enables the user to watch annotations simultaneously with the video scene they refer to.

Annotations are added to and removed from the viewing table according to the playhead position on the video timeline. Other work [20] has focused on using a list of all temporal annotations of the video that scrolls vertically according to the playhead position on the video timeline. However, we believe that temporal video annotation is only suitable in the context of its corresponding video scene. Therefore, we display only the video annotations in the viewing table that refer to the video scene currently shown in the video player.

4.2.5 Interaction

There are three ways for the user to navigate through the video by using the visual cues. First, the user can drag the playhead of the scrubber bar to forward or rewind the video to the location of a certain visual cue. The playhead position on the scrubber bar is synchronized to the position of a small playhead on each video timeline (Figure 4.11).

When the user moves the playhead of the scrubber bar, the playheads on the video timelines follow to provide the user with a precise navigational aid. When a playhead hits an icon and activates it, the icon is highlighted by using a brighter version of its graphical representation. When the playhead moves on and the icon is no longer active, it returns to its former graphical representation.

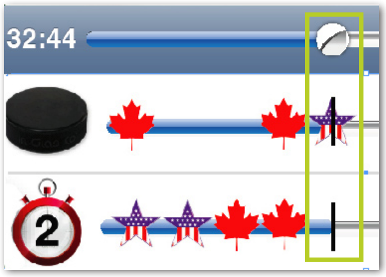


Figure 4.11: The playhead of the scubber bar is synchronized with the playheads on the video timelines.

To go to a scene of interest, the user can also directly interact with the icon by selecting it on the video timeline. The direct selection allows the user to go to a specific scene of interest without forwarding or rewinding the video between the current temporal position and the scene he is looking for.

Compared to the scrubber bar approach in which the user has to drag the playhead along the entire timeline to go to a certain position, the direct selection enables the user to quickly reach a target. This becomes especially useful when several icons are of interest but positioned in large distances to each other on the video timeline.

However, when many icons are located on a video timeline and SCADE is applied, selecting a target with, for instance, a finger on the iPhone touch screen becomes difficult. In those cases where the user is no longer able to directly select an icon on the video timeline, the scrubber bar approach becomes more appropriate.

TAV contains a third method for navigating icons based on existing methods of video editing and postproduction software, such as Adobe After Effects. Due to the need to quickly forward and rewind the video to the next or previous scene of interest, we added one button on each the left and right side of the video timeline (Figure 4.12). The buttons allow the user to move from the current position in the video to the following or previous icon on the video timeline. For instance, if the user is only interested in the video scenes with goals, he can easily use the skip buttons to go to the next scene with an icon representing a goal. This method can be used when the user is interested in a specific video scene with icon, but the video timeline does not allow direct selection due to a large number of visual cues.

However, when using the skip approach the context of the adjacent video scenes is not given due to its nature of excluding video scenes when jumping to the next icon.

Figure 4.12: Skip-buttons to forward to the next or previous icon on a video timeline.

4.2.6 Implementation

We develop our prototype in the integrated development environment XCode 3.1.4. with the iPhone software development kit 3.1.3. We test our prototype in the iPhone Simulator as well as on a device, such as the iPod 2G 8GB. In the following section, we describe three main challenges when building the prototype: the OpenGL rendering for displaying the video frames, the loading and caching of video frames and the xml file for providing the visual cues according to the active video content.

OpenGL for Video Player

Since the implementation of our novel user interface concept TAV is embedded in the VideoDiver of the MyView project, we follow the existing guidelines for prototype development. On the long term, the MyView project aims at streaming video content in real-time from a live broadcasting event instead of using existing local video content. When streaming video content from a set of cameras, the video material is delivered as a permanently growing number of single video frames. Based on this, we cannot make use of the built-in video player of the iPhone which has only very limited functionality. Furthermore, the current application programming interface does not provide developers with free access to the video player components and prevents enhancements of the built-in video player.

For our own video player, we use OpenGL to render the video frames on the screen. Due to the memory limitations of the device, we apply OpenGL subtexturing to keep the computing time at a minimum. In OpenGL subtexturing a part of the texture is replaced by a subrectangle that is drawn on top of it.

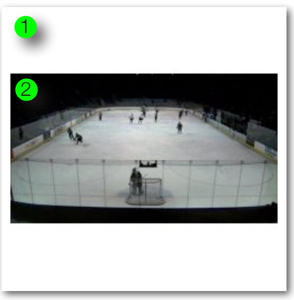


Figure 4.13: OpenGL subtexturing, 1) plain texture with a size bigger than the video frame and with a height and a width that are a pow of two, 2) the video frame as a subtexture

Subtexturing is not only the fastest way to update the image of an existing texture, it is also an effective method to display images with a width and height being not a pow of two and thus, do not fit the OpenGL texture requirements. Since we use video frames for our prototype, the texture requirements of OpenGL soon become a problem because none of the common video formats, such as PAL[10] and NTSC[11], fit those dimensions.

To solve this problem, we use the size of the video frame to generate a plain texture that is larger than the video frame and has a height and width that are a pow of two (Figure 4.13). Afterwards, we subtexture the video frame onto the plain texture and update it every time the render method is called. Based on this, the memory usage is optimized and video frames can be displayed without further modification.

Caching and Loading Content on the Device

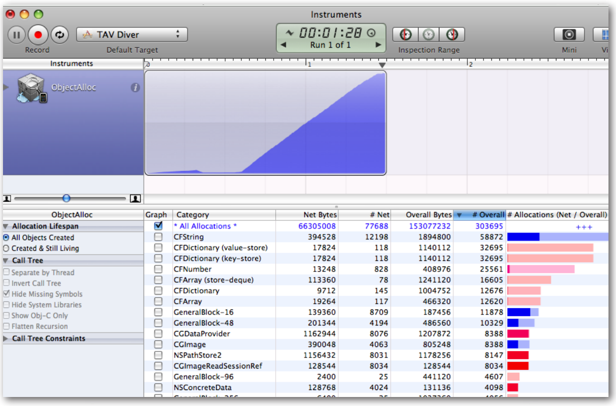
At present, we are working with local content which is a set of video frames of existing video material stored on a hard drive. While the high number of video frames is not a problem when working with the iPhone Simulator on a desktop computer, loading the large number of single files on a mobile device, such as the iPod 2G 8GB, is inappropppriately slow. Due to the fact that the local video content does not change during several launches of the application, we use caching to reduce the amount of time necessary to start the application.

Caching allows us to provide the video frames faster for each usage of the prototype application since they are copied to the device only once in an initial launch. To cache the video frames, we first transfer them with the ressources of the application bundle to the device. Afterwards, we copy the video frames from the application bundle to the cache directory of the application. Finally, we remove the vi-

10 PAL (Phase Alternate Line) is a television standard with a resolution of 720 x 576px and a framerate of 25fps [50].
11 NTSC (introduced by the National Television Systems Committee) is a television standard with a resolution of 720 x 480px and a framerate of 30fps [49].

deo frames from the application bundle when building the application for the next time and change the path to the video content to point to the cache directory.

Although we apply caching and optimize the OpenGL rendering, there are still limitations for the number of video frames we can load. We use the performance measuring tool Instruments [31] to improve our memory management where possible. However, we need to load all video frames into the application to enable rewinding and fast-forwarding of the video which is a problem due to the size of the working memory available on the iPod 2G 8GB (Figure 4.14).



Source: Instruments [31]

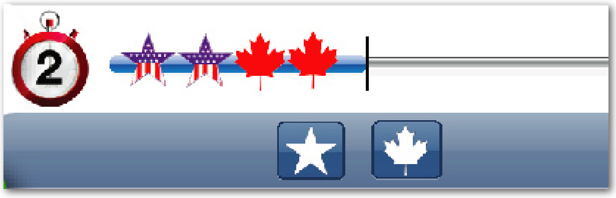
Figure 4.14: Performance of the TAV prototype when measuring object allocation with Instruments to optimize memory usage.

At the moment, we are able to load around three minutes of video content with a framerate of 25fps, a resolution of 240px to 135px and a file size of 8KB. While this is sufficient to run a user study and answer the research question of this thesis, we will work on the optimization of the file management during future prototyping. Additionally, we will focus on streaming network content from the capture system of the MyView project which provides us with new challenges and opportunities for using the prototype at hand.

XML for Visual Cues

We load the data for the video timeline with visual cues from an xml file. The interface adjusts automatically according to the content created from parsing the xml file. For instance, the vertical position of the filter toolbar and the viewing table changes according to the number of video timelines with visual cues.

The xml file structure is based on the arrangement of the interface components of the video timelines with visual cues and the filter toolbar. The interface consists of any number of timelines. Each timeline has an identifier image which is shown at the left side of the timeline and several visual cues. Each visual cue has a frame number for its location on the timeline and a specific type, for instance, USA or Canada as seen in figure 4.15.



Source: Own Figure

Figure 4.15: The xml structure is based on the interface structure of timelines, visual cues and filters.

The type of a visual cue is related to an element of the filter toolbar. Since filters are applied globally to all video timelines that contain the annotation type, they are positioned independent from the timeline structure in the xml file. Based on this, the xml file first contains the filter for each visual cue type and then lists each video timeline with its visual cues representing the temporal video annotation. In the following section, we present the basic structure of our xml file.

```
<Filters>
        <Filter type=„„ image=„„/>
        [...]
</Filters>
```

```
<Timelines>
  <Timeline image="">
    <Annotation frame=„ filtertype=„>
        Content of the Annotation
        displayed in the Viewing Table
    </Annotation>
    [...]
  </Timeline>
  [...]
</Timelines>
```

Loading the timeline data from an xml file allows us to quickly change between different video contents and the corresponding visual cues. Due to the characteristics of xml as a highly portable language, our data sets can also be used with other applications that explore the field of video navigation with visual cues.

At present, the xml file has to be created manually which is sufficient to create the number of data sets required to run the user study for evaluating our novel solution SCADE. However, the prototype will be further developed to enable the user to create his own visual cues while watching a video. For instance, if the user points an interesting scene, he can create a visual cue to easily retrieve the video scene later. Future work will cover the development of a user interface concept that provides the user with the functionality to create the visual cues as seen in the current prototype.

4.2.7 Conclusion

In this chapter, we presented different solutions for the research problem and discussed why SCADE is the most effective approach. Furthermore, we described our user interface concept called TAV that consists of a traditional video player, multiple video timelines with visual cues, a filter toolbar and an annotation viewing interface. Finally, we explained the most challenging parts of the implementation of the prototype. In the next chapter, we discuss the preparation, conduction and results of our user study that we run based on our prototype implementation of TAV and SCADE.

Chapter 5

User Study

Source: MyView [34]

5. User Study

Our user study is designed to evaluate if icons that are adjusted with our novel approach SCADE result in faster video navigation compared to colour stripes. In the user study, participants are given a video player containing a video timeline with visual cues and are asked to find video scenes by searching for visual cues on the timeline. In the following sections, we describe how we designed, prepared, conducted and evaluated the user study.

5.1 Design of the User Study

In this section, we give an overview of our user study design containing the hypothesis, the dependent and independent variables, the mixed experimental design and the ethics approval for conducting the user study.

5.1.1 Hypothesis and Null Hypothesis

The hypothesis and null hypothesis according to our research question are:

H-1: Finding a video scene with icons is faster than with colour stripes even when the number of visual cues on the video timeline increases.

H-0: Finding a video scene with icons is not faster than with colour stripes when the number of visual cues on the video timeline increases.

To define a user study design according to our hypothesis, we have to consider which dependent and independent variables are included in the setup.

5.1.2 Dependent Variables

In our user study, we measure two dependent variables which are called finding time and error rate that help us to evaluate the influence of the independent variables.

Finding Time

The finding time is the time a user needs to find a video scene by using the visual cues on the video timeline. The finding time is measured in milliseconds and provides us with data to evaluate the influence of different visual cues on the video timeline.

Error Rate

The error rate is the number of tasks with errors divided by the number of total tasks. An error is measured each time a user confirms that he found the correct video scene but made a mistake.

5.1.3 Independent Variables

In our user study, we examine the influence of two independent variables on the finding time and error rate. The first independent variable is called display mode of visual cues and has two levels which are colour stripes and icons. The second independent variable is called distribution of visual cues and has five levels.

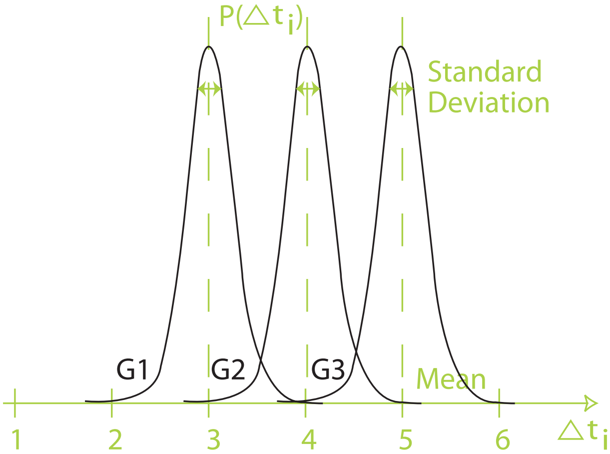
Display Mode of Visual Cues

Based on the hypothesis, we differentiate between two levels of display modes which are colour stripes and icons. While colour stripes consist of only a plain colour, icons provide more semantic details based on their shape and textural pattern.

Distribution of Visual Cues on the Timeline

For controlling the distribution of visual cues on the video timeline, we consider the uniform distribution, the Gaussian distribution and the Poisson distribution as potential methods to position the visual cues. In the following section, we discuss the practicability of each method in the context of our user study and give reasons why we focus on the Poisson distribution.

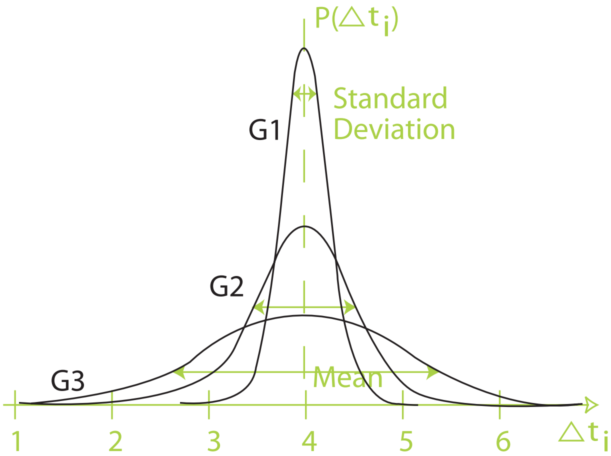
In the uniform distribution, the probability of all values is equal which does not correspond to the characteristics of a video timeline. It is likely that visual cues on the video timeline group around interesting scenes and that other parts of the video timeline do not have any visual cues. Using the uniform distribution is therefore not appropriate to model a video timeline as required for our user study.



Source: Own Figure

Figure 5.1: Gaussian distribution. Keeping the standard deviation constant and changing the mean results in a larger or smaller average temporal distance between visual cues.

The Gaussian distribution provides us with two variables the mean and the standard deviation for creating different levels of distribution. For instance, if we keep the standard deviation constant and adjust the mean, we are able to control the average temporal distance between visual cues (Figure 5.1). Based on this, a low mean corresponds to a small average temporal distance between visual cues whereas a high mean corresponds to larger temporal distances in average.



Source: Own Figure

Figure 5.2: Gaussian distribution when keeping the mean constant. A low standard deviation results in visual cues that are nearly equally distributed on the video timeline.

Another way of creating different distributions is to keep the mean constant and adjust the standard deviation (Figure 5.2). For instance, a low standard deviation corresponds to visual cues that are placed in nearly equal temporal distances, whereas a high standard deviation corresponds to a high variance in temporal distances. When visual cues are distributed unequal along the timeline, it is more likely that there is a high density of visual cues among certain video scenes.

However, the Gaussian distribution is mainly applied to physical attributes of materials, such as errors in a mass production or the average size of people. Our aim to distribute visual cues on a video timeline is a time-based problem for which applying the Poisson distribution is more appropriate.

The Poisson distribution covers time-based distributions, such as people arriving in a shopping mall or cars passing by on a highway. The Poisson distribution is based on a rate, such as people per hour or cars per minute, and visualizes the deviation from this rate. For instance, if people entering a shopping mall are counted, the result might be that there are fifty people per hour. However, repeating the observation several hours creates a slightly different rate each time, for instance, fourtyfour people per hour or sixty people per hour. Entering these rates into a graph results in a Poisson distribution with one value being the most likely rate which is called lambda λ. If lambda is small only few people enter the shopping mall in one hour, if lambda is large many people enter the shopping mall in one hour.

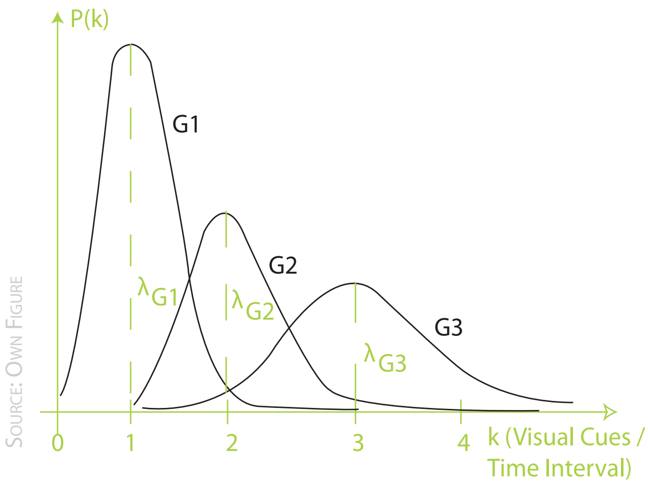


Figure 5.3: Poisson distribution with different levels of λ. When λ increases visual cues on the video timeline are distributed more densely.

Since our problem of distributing visual cues on a video timeline is a time-based problem, we focus on the use of the Poisson distribution based on a rate of visual cues per time interval (Figure 5.3). For instance, if we take a one minute video and aim for a distribution with few visual cues, we define a lambda of one visual cue per ten seconds. According to the Poisson distribution, most time intervals of ten seconds will have one visual cue, but some others will have two or three visual cues or none. If we take the same video and apply a lambda of five, most time intervals will have five visual cues and the remaining time intervals have less or more visual cues.

With the Poisson distribution, we have a standardized way of calculating the distribution of visual cues along the timeline. With regard to the limited length of the user study, we select five levels for lambda representing different dense distributions of visual cues along the timeline.

5.1.4 Mixed Design

We choose a mixed design with the independent variable display mode being setup as a between-subjects experimental design and the independent variable distribution being setup as a within-subjects experimental design.

Participants are randomly assigned to the two levels of display mode (colour stripes, icons) and with the assigned display mode participants serve in all levels of distribution (λ1, λ2, λ3, λ4, λ5) as described in figure 5.4.

Distribution

Display Mode		λ1	λ2	λ3	λ4	λ5
	Colour Stripes	P1	P1	P1	P1	P1
	Icons	P2	P2	P2	P2	P2

P1: Participant 1
P2: Participant 2

Source: Own Figure

Figure 5.4: Mixed design. Participants are assigned to either the colour stripe or the icon group and use this display mode for all levels of distribution.

The mixed design combines the advantages of both the between-subjects and the within-subjects design. The between-subjects design for the display mode reduces the fatique effect of participants and keeps the time for the user study at an appropriate length since participants are not exposed to all conditions. Furthermore, the between-subjects design allows us to use a larger task set for a single condition since the number of tested conditions per participants is only half of the number of conditions the user study has in total.

The within-subjects design for the distribution allows us to keep the number of participants at an appropriate size by reason that participants are exposed to all levels of distribution. Based on the combination of the between-subjects and within-subjects design, we reduce carry-over and learning effects while keeping the number of participants at an affordable size and having enough statistical power.

5.1.5 Ethics Approval

In Canada, research ethics boards are responsible to ensure that research involving human subjects follows ethical principles. Before a user study is conducted, the researcher has to submit a human research compliance protocol to one of the human ethics research boards. In our user study, we are conducting behavioural research and therefore applied for a certification of the UBC Behavioural Research Ethics Board (BREB). The application consists of a detailed user study description and all documents used during the user study, such as the consent form and protocols.

In order to submit an application to the BREB, researchers are required to complete a tutorial for the „Tri-Council Policy Statement: Ethical Conduct for Research Involving Humans" (TCPS). The research ethics' introductory tutorial consists of multiple sections containing information about the principles, standards and procedures for conducting research involving human subjects. Topics covered by the TCPS are, for instance, the free and informed consent of all research subjects and the privacy and confidentiality during the user study and for storing data. After passing a multiple choice test, the tutorial finishes with a „Certificate of Completion". The date of completion has to be entered into the application to the BREB and the certificate itself must be available in a printed form while conducting the user study.

5.2 Preparation of the User Study

In the following section, we describe how we prepare the material for the user study by choosing, selecting and editing a set of hockey videos and by adding the visual cues according to the Poisson distribution.

5.2.1 Selecting the Video Material

To gain video material for the user study, we search for ice hockey videos showing games with the Vancouver Canucks due to the fact that our lab is located in Vancouver, Canada. For our search, we use the social video site YouTube and select summaries of hockey games called highlights to increase the density of events happening during a short time period of the video.

5.2.2 Listing the Positions of Events

After downloading the three to eight minute long videos, we make notes on the events happening in the video to find potential temporal locations for visual cues. While noting the events in the video, we focus on identifying common events that appear among all ice hockey videos to later use the same visual cues in each of the videos.

5.2.3 Limiting the Number of Different Events

We focus on using the same visual cues in each video by reason that different visual cues can lead to additional influencing factors when measuring the finding time. For instance, when the icons of one video are easier recognizable than the icons of another video, the user's finding time is not only influenced by the distribution on the timeline, but also by the recognizability of the icons.

Source: YouTube [42]

Figure 5.5 Overview of events in an ice hockey game. 1) referee, 2) celebrating the score, 3) coach, 4) fight, 5) audience, 6) score

Therefore, we define seven events that happen regularly in an ice hockey game which are: score when a hockey team makes a goal, score was blocked when a hockey team tried to make a goal but was stopped by the other team, celebrating the score when the hockey players cheer after they made a goal, the coach appears, the audience is shown, the referee interferes and a fight between two players happens (Figure 5.5).

5.2.4 Editing the Video Sequences

After making notes on all events in the hockey videos, we take out six one minute video sequences that have a high event frequency. The six video sequences correspond to the five levels of lambda for the Poisson distribution plus one training video. We keep the length of the videos limited to one minute due to the memory limitations on the iPod. Subsequently we have chosen the video sequences, we render the material into jpg sequences with a resolution of 240px to 135px and a file size of 8kb to fit the requirements for video material on our prototype.

5.2.5 Defining The Positions of Visual Cues

Following the preparation of video material, we create the visual cues and their locations on the timeline for each video. We implement a noise generator based on existing code of the Poisson and the uniform distribution [51] and expand the code to calculate the distribution of visual cues on the video timeline.

Our noise generator returns the Poisson distribution after the number of intervals on the timeline and the lambda value according to low or high density of distribution are inserted. We split each video in six time intervals of ten seconds and use values for lambda beginning from one visual cue per time interval up to five visual cues per time interval.

```
1 annotations in the 1. interval [-0.00,-300.00].
1. annotation position at frame: 125.
--------------
4 annotations in the 2. interval [-300.00,-600.00].
1. annotation position at frame: 506.
2. annotation position at frame: 477.
3. annotation position at frame: 579.
4. annotation position at frame: 554.
--------------
1 annotations in the 3. interval [-600.00,-900.00].
1. annotation position at frame: 758.
--------------
0 annotations in the 4. interval [-900.00,-1200.00].
--------------
2 annotations in the 5. interval [-1200.00,-1500.00].
1. annotation position at frame: 1228.
2. annotation position at frame: 1396.
```

Source: Own Figure

Figure 5.6: Results from our noise generator using a Poisson and a uniform distribution to define the temporal location of visual cues.

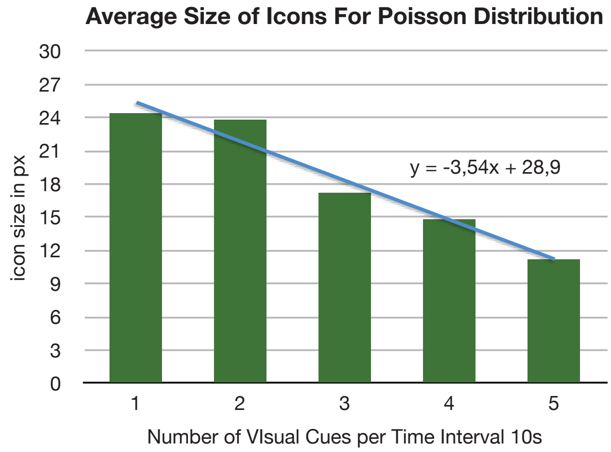
However, the Poisson distribution only returns the number of visual cues in each time interval, but not the exact frame number of each visual cue on the timeline. We use a uniform distribution to randomly assign the visual cues to a certain frame number in the given Poisson time interval (Figure 5.6).

Based on the list of all temporal positions for five different distributions, we apply one distribution to each video according to the frequency of events happening in the video. For instance, the video with the lowest event frequency is assigned to the smallest value of lambda.

Afterwards, we compare the list of events with the temporal location of visual cues to decide which visual cue corresponds to one of our predefined events, such as coach, audience, referee or fight. Since each video has five finding tasks assigned, we need at least five meaningful visual cues for which we can ask the participant in the user study. Following this, we randomly assign three other visual cue types to fill the remaining temporal locations as defined by the Poisson distribution. In the user study, we do not ask for these random visual cues since their only purpose is to control the density of distribution on the video timeline.

5.2.6 Using SCADE to Adjust Icons

While colour stripes have a fixed width of seven pixels during the user study, icons are adjusted with our novel approach SCADE to correspond to available screen space. When the number of visual cues increases on the timeline, SCADE decreases the width of the icons from their maximum size of twentyfive pixels up to a minimum of seven pixels. Therefore, increasing the lambda in the Poisson distribution directly influences the appearance of icons on the video timeline (Figure 5.7).



Source: Own Figure

Figure 5.7: When the number of visual cues on the timeline increases, the size of the icons decreases when our novel approach SCADE is applied.

5.2.7 Choosing the Graphical Representations

We choose the graphical representations of visual cues by ourselves, but evaluate our choice in a pilot study. While there are many results from search engines that demonstrate a common sense of icon design (Figure 5.8), it is difficult to define the meanings of the colour stripes. We aim at using the colours in the way participants know them from their daily life, such as the colour green representing a success (score) and the colour red representing a failure (scoring was blocked). For other visual cues, we refer to the visual appearance of the related video scene, such as using the colour black for the referee based on his tricot colour.

Source: Referee [35], [36], [37], [38]

Figure 5.8: Search results from search engines, for instance, when looked for referee icons demonstrate a common sense in icon design.

After defining the graphical representations of all visual cues, we edit the icons to have seven levels of deformation and prepare the xml files with the location of visual cues for each video.

5.2.8 Simplifying the User Interface of TAV

For the user study, we simplify the user interface concept of TAV since we aim at evaluating the trade-off between detailed semantic information and limited screen real-estate, but do not focus on how multiple timelines, annotation viewing and filter mechanisms support video search and navigation. Therefore, our user study prototype only consists of a video player, a video timeline with visual cues and a scrubber bar for fast-forwarding and rewinding the video.

5.3 Conduction of the User Study

In the following section, we describe how we gain participants for the user study and give an overview of the user study conduction by including information about the consent form, the training session, the task session and the questionnaire.

5.3.1 Contacting Participants

We invite members of the Media and Graphics Interdisciplinary Centre and members of the Human Communication Technology Laboratory of the University of British Columbia to participate in the user study. Team members of the MyView project are excluded from the invitation to ensure that no previous knowledge of the project influences the results. The experimenter uses the mailing list of both labs to contact potential participants. Additonally, the experimenter invites potential participants in a conversation by handing over invitation cards (Figure 5.9).

Source: MyView [34]

Figure 5.9: User study invitation card handed over to potential participants in a conversation about the project.

During the invitation, participants are informed that the user study is about video browsing on a mobile device. Furthermore, the experimenter states that the user study takes roughly fourtyfive minutes and that each participant receives a 10$ compensation for his participation.

5.3.2 Scheduling the Experiment

For organizing the user study, each participant chooses dates from an online schedule in which he marks his times of availability. Afterwards, the experimenter selects one of the suggested times and sends the participant an email of confirmation. Participants who are invited in a conversation receive an invitation card with the user study time entered in the plain field (Figure 5.9).

The user study is conducted from 9am to 4pm on August 19 and August 20, 2010. Each user study session starts on the hour, takes fourtyfive minutes and contains fifteen minutes for the experimenter to reset the user study prototype and to make a short break.

5.3.3 Participant Statistics

Overall twelve participants subscribed for the user study, six participants on each of the two conduction days. Eight out of twelve participants are males (67%), four are females (33%) as described in figure 5.10.

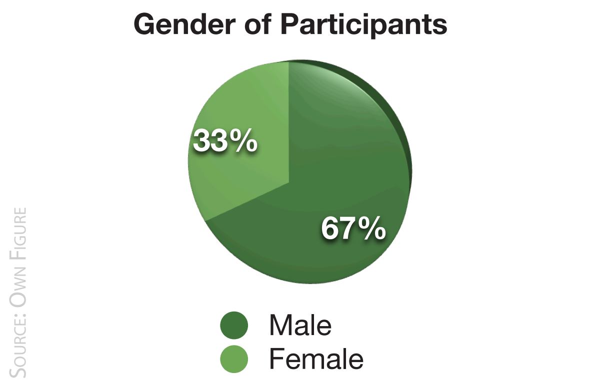


Figure 5.10: Eight out of twelve participants are males (67%), four are females (33%).

Participants are between 23 and 53 years old (Figure 5.11). The average age of participants is 30.25 years.

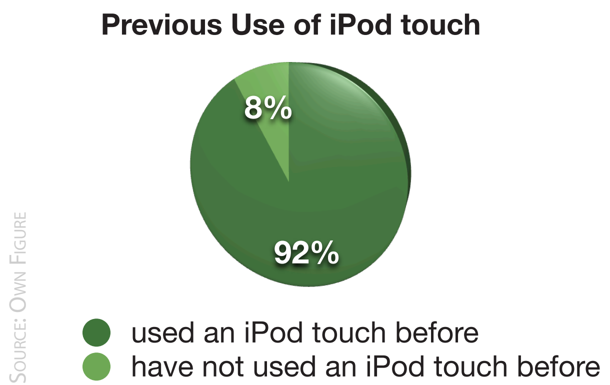


Figure 5.12: Eleven out of twelve participants have used an iPod touch or an iPhone before they participated in the user study.

Eleven out of twelve participants have used an iPod touch or an iPhone before and are familiar with the user interface elements (Figure 5.12).

Participants are distributed randomly to one of the display modes either icons or colour stripes. To ensure that both groups come with the same statistical power, both have an equal number of participants assigned (Figure 5.13). Six participants use the icons during the user study and the other six participants use the colour stripes.

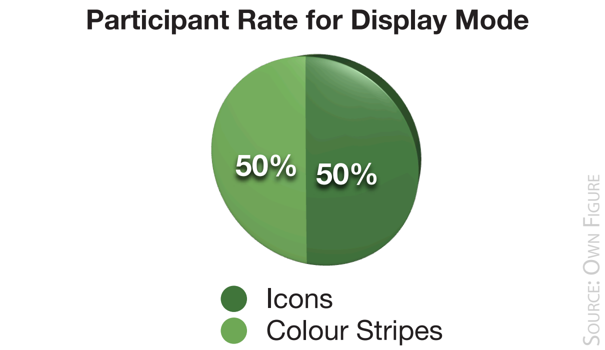


Figure 5.13: Participants are randomly assigned to one of the user study groups. However, both of the groups have six participants (50%).

5.3.4 Sequence of Steps During The Conduction

Participants either come to the user study room by themselves or the experimenter leads them to the right location. A door sign is used

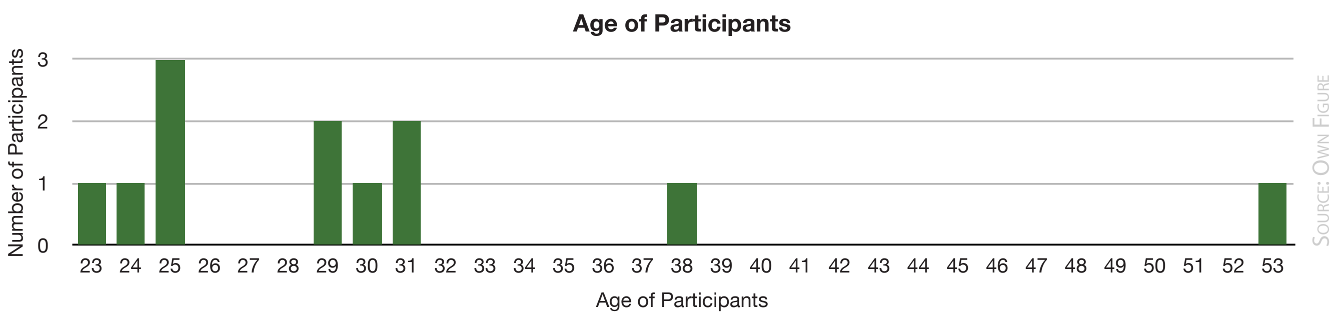


Figure 5.11: Participants are between 23 and 53 years old. The average age of all participants is 30.25 years.

to mark the user study room for participants. Additionally, the door sign informs by-passing people that a user study is under conduction and disturbance is not allowed.

Introducing the Participant to the User Study

After entering the room, participants are assigned to a seat. The experimenter starts the user study by thanking the participant for his will to participate. Following this, the experimenter repeats the information from the invitation by summing up the general topic of the user study. The experimenter explains that the user study is about browsing video on a mobile device and that the participant will use the iPod touch. The experimenter adds that the participant will be given a set of hockey videos and that he will be asked to find specific video scenes within these hockey videos.

Signing the Consent Form

Following this general information, the experimenter informs the participant that he has to sign the consent form to confirm that he agrees with all experimental conditions. A printed version of the consent form is given to the participant and each passage is explained by the experimenter. The experimenter briefs the participant that there is only minimal risk involved in the study, that his data is kept confidental and that he has the right to withdraw from the user study at any time. The participant is encouraged to take as much time as he needs to go over the consent form and he is asked to sign the consent form on the last page if he agrees with the conditions. The experimenter also emphasizes that the participant can ask questions at any time during the user study.

Overview of the Task Session

After the participant has signed the consent form, the experimenter gives a detailed explanation about the remaining part of the user study. The experimenter states that the user study consists of thirty tasks that are divided into six sessions each containing five tasks. After each of the six sessions a short break is made to avoid fatigue effects of the participant. The participant is informed that each task will asked him to find a specific video scene and that tasks are phrased in the manner of, for instance, „Find the video scene with the audience“ or „Find the video scene with the referee“. Thereafter, the experimenter asks the participant if he has any questions about this part of the user study.

Explaining the User Interface of the Prototype

To explain the user interface of the prototype application, a paper sheet showing the main screens of the application is used. Depending on the level of display mode the participant is assigned to, the paper sheet shows either the user interface for icons or the user interface for colour stripes. The participant is informed that the first screen is the task introduction screen which contains the current task number and the total number of tasks in the active task session. The experimenter states that the participant can take as much time as he likes when seeing this screen and should press the „Start Task“ button as soon as he feels ready (Figure 5.14).

Source: MyView [34]

Figure 5.14: Task introduction screen showing the number of the current task and the number of total tasks in the task session.

After the task is started, the participant is introduced to a loading screen that shows the progress of loading the next video which takes approximately ten seconds (Figure 5.15).

Source: MyView [34]

Figure 5.15: Loading screen displaying the progress of loading the hockey video for the current task.

As soon as the loading is finished the participant is presented to the video browsing screen (Figure 5.16).

Source: MyView [34]

Source: MyView [34]

Figure 5.16: Video browsing screen for both display modes icons (top) and colour stripes (bottom) containing a done-button (1), a task text field (2), a skip-button (3), a video player (4), a video timeline with visual cues (5) and a scrubber bar (6).

The experimenter explains that the participant should first read the text of the task which is positioned on top of the screen. Following this, the experimenter introduces the participant to the other user interface elements. First he goes over the elements known from traditional video players that are a video window showing the hockey video in the middle of the screen and a scrubber bar used to fast-forward or to rewind the video which is located at the bottom of the screen.

Afterwards, the experimenter informs the participant that the user study examines how visual cues can help to locate specific video scenes. Depending on the level of display mode the participant is assigned to, the experimenter either uses the word „icons" or „colour stripes" instead of „visual cues". Based on this explanation, the experimenter introduces the participant to the video timeline that contains different visual cues. The experimenter states that each visual cue represents the meaning of a specific video scene. The participant is told that each task in the user study requires him to identify a video scene that corresponds to one of the visual cues seen on the video timeline.

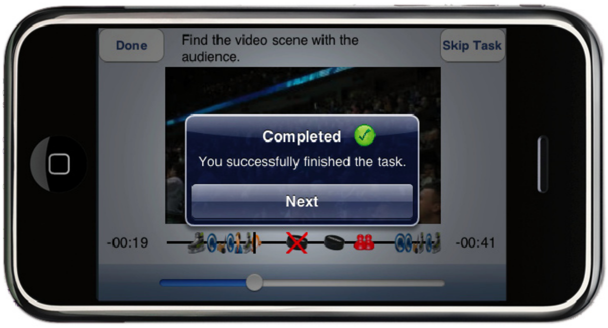
As an example, the task „Find the video scene with the audience" is used. The experimenter gives the participant a second paper sheet that contains a legend showing the meaning of the different visual cues. The participant is asked to use the legend to find the corresponding visual cue for the audience. After the participant identified the visual cue for the audience, the experimenter shows him how he can use the scrubber bar to move to the corresponding video scene. A vertical black line on the video timeline moves according to the position of the playhead on the scrubber bar and gives additional feedback when a visual cue is reached.

Furthermore, each visual cue highlights when the playhead reaches its position.

The participant is informed that after he reached the correct visual cue on the timeline, he has to press the done button in the upper left corner to confirm his choice. The experimenter suggests to use the right index finger for dragging on the scrubber bar and the left index finger for pressing the done button.

The experimenter explains that after hitting the done button, the participant is either presented to a task completed screen (Figure 5.17) or to a task failed screen (Figure 5.18).

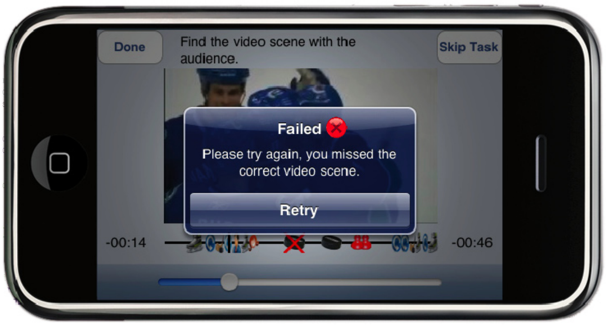
When seeing the task completed screen, the participant successfully found the video scene and finished the task. By pressing the next task button, the participant moves to the task introduction screen to start the next task.



Source: MyView [34]

Figure 5.17: The task completed screen is shown after the participant has found the correct video scene and pressed the done-button.

When seeing the task failed screen, the participant missed the correct video scene and is asked to retry the search. The experimenter informs the participant that he can try finding the correct video scene as often as he likes. Additionally, the experimenter emphasizes that it is more valuable for the user study if the participant is searching until he successfully completed the task. For the case of not being able to finish a task, the participant is informed about the skip-task button in the top right corner of the screen which stops the current task and moves to the task introduction screen which presents the next task.

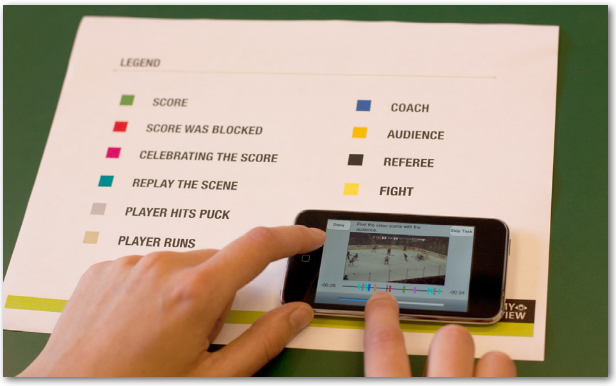


Source: MyView [34]

Figure 5.18: The task failed screen is shown when the participant pressed the done-button but missed the correct video scene.

Explaining the Legend of Visual Cues

After explaining the user interface, the experimenter goes over the different meanings of the visual cues on the legend paper sheet. The meaning of each visual cue is read aloud and a short explanation is given. The experimenter advices the participant to put the device on the free space of the legend sheet in the lower right corner to always have the best view on the meanings of the visual cues (Figure 5.19).



Source: MyView [34]

Figure 5.19: User study setup showing a participant that is doing the task session with colour stripes.

Furthermore, the experimenter informs the participant that tasks referring to hockey teams label the teams by the colour of their tricots. For instance, instead of asking for a goal of the Vancouver Canucks, the task asks for the goal of the white-blue team. The experimenter gives reason for this by explaining that the participation in the experiment does not require any previous hockey knowledge.

Following the overview of the user interface and the legend, the experimenter emphasizes that the participant should try to find the video scenes as fast as possible and with as few errors as possible. Afterwards, the participant is asked if he has any questions concerning the user interface and the legend and when denied is introduced to the training session.

Conducting the Training Session

The training session is used to allow the participant to familiarize himself with the user interface. Based on this, the training session minimizes learning effects during the user study. The experimenter states that the first five out of the thirty tasks are training tasks and that the participant can take as much time as he likes for the completion by reason that the data is not used for evaluation. The experimenter starts the prototype application on the device and the participant conducts the training session. At the end of the training session the participant is asked if he has any questions. Furthermore, he is encouraged to repeat the training session if necessary or to progress to the task session of the user study.

Conduction the Task Session of the User Study

After the participant emphasized that he feels familiar with the user interface, the experimenter informs the participant that he will now start the experiment with the remaining twentyfive tasks. The experimenter repeats the goal for the participant to be as fast as possible and to make as few errors as possible when searching for the video scenes. The experimenter then starts the tasks as described previously with a break after each task session.

Filling in a Questionnaire

Subsequently the participant completed all tasks, he is asked to fill in a questionnaire. The questionnaire contains four rating questions that require the participant to evaluate how often he is searching for video scenes in familiar and unfamiliar video sources and if he thinks the visual cues helped him to locate scenes of interest and to reason about the underlying video content.

Following the rating questions, the partcipant is asked to give his opinion on several topics by writing into a plain text field or by drawing sketches for suggestions that are easier communicated visually. The questions ask the participant, for instance, if he can imagine other types of visual cues that can be used on a video timeline, where he sees advantages in using the visual cues presented to him during the user study and where he sees room for enhancement concerning the visual cues he has seen on the video timeline.

Like the user interface and the legend paper sheet, the questionnaire depends on the level of display mode participants are assigned to. The questionnaire for a participant in the colour stripe group asks him about advantages and enhancements concerning colour stripes and if he can imagine other visual cues than colour stripes on the video timeline, whereas a participant in the icon group is asked the same question but concerning icons.

After the participant finished giving his opinion and suggestions, he is asked to fold his questionnaire once to ensure confidentiality of his data. Later, the experimenter gives each consent form and questionnaire a number for reference purposes. With handing over the questionnaire, the participant completes the user study and the experimenter thanks him again for his participation.

5.4 User Study Results

Based on our user study conduction, we gain twelve data sets for each of the twentyfive search tasks which makes 300 data sets in total (150 colour stripes, 150 icons).

In our user study, five tasks are assigned to each of the five levels of distribution which results in 60 data sets per distribution.

From these 60 data sets half provide insights into the finding time and error rate when colour stripes are used on the video timeline. The other half corresponds to icons that are adjusted with our novel approach SCADE.

According to this, we have an appropriate data collection to evaluate how the finding time is influenced by the display mode and the density of distribution of visual cues.

5.4.1 Cleaning the Data Set

Before we evaluate the data with regard to a comparison of finding time between colour stripes and icons, we clean the data to remove values that either contain errors or have a finding time above a certain treshold.

Removing Data With Errors

First, we remove the data sets that contain one or more errors made by the participant while searching for the correct video scene. We delete these data sets by reason that some participants stated that they missed the video scene by a few seconds because the visual cues are too small to select properly.

Since we did not measure the frame number where the error occured, we cannot reason which data set with errors resulted from this cause. However, tasks with errors appeared only in 8% of the icon and colour stripe data sets (Figure 5.20).

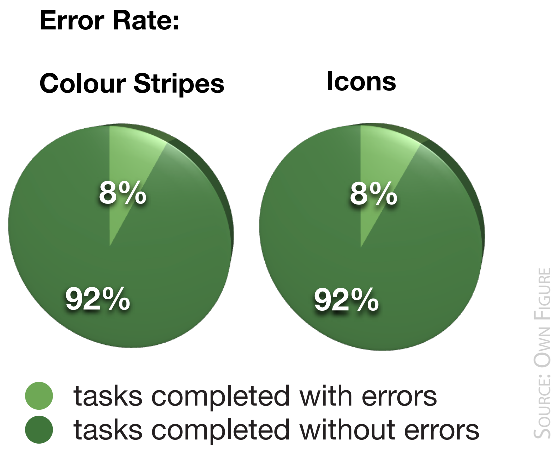


Figure 5.20: In 8% of all data sets participants made errors while searching for the correct video scene.

Removing Data With Time Over Threshold

Furthermore, we remove all data sets that have a finding time over the treshold of 15s. Only 4% of all data sets in both display modes have a finding time above 15s and those are often far beyond this value (Figure 5.21).

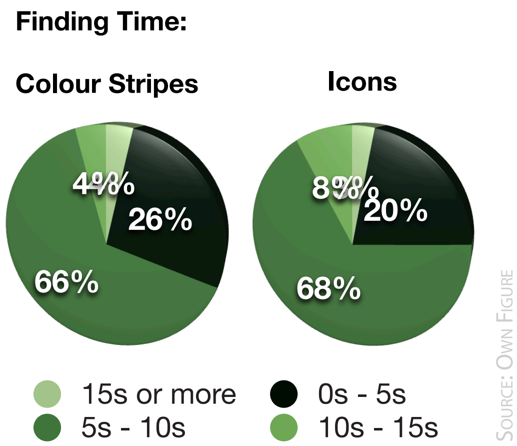


Figure 5.21: Only 4% of data sets in both display modes have a finding time above 15s. We remove these data sets from our evaluation.

5.4.2 Comparing the Average Finding Time

After cleaning the data sets, we calculate the average finding time for each task (Figure 5.22). Our calculation shows that the difference of finding time between the colour stripes and the icons is small and often less than one second. Evaluated over the whole task session, colour stripes are faster in 60% of all tasks, whereas icons are faster in 40% of all tasks (Figure 5.23).

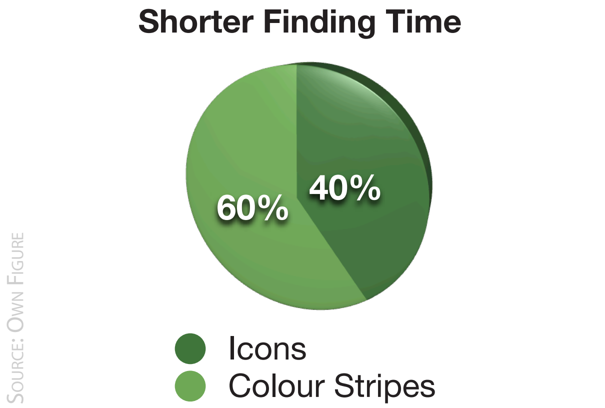


Figure 5.23: When the finding time of colour stripes and icons is compared for each task, colour stripes are faster in 60% and icons in 40% of all cases.

However, when the finding time is sorted according to the five different levels of distribution, we can state that with increasing density the finding time with icons increases faster than with colour stripes (Figure 5.24).

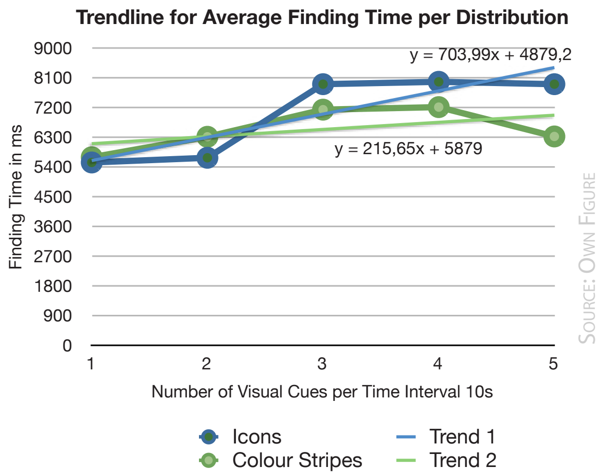


Figure 5.24: Finding time according to the five levels of distribution. The trend line shows that with increasing density, the finding time of icons increases faster than the finding time of colour stripes

5.4.3 Questionnaire Results

The questionnaire we use for the user study collects both quantitative and qualitative data.

Search in Familiar and Unfamiliar Videos

The first question asks participants how often they search for video scenes in familiar video sources, such as a favorite tv series or an old family video, and how often they search for video scenes in unfamiliar video sources, such as a video tutorial or a sportsgame that they

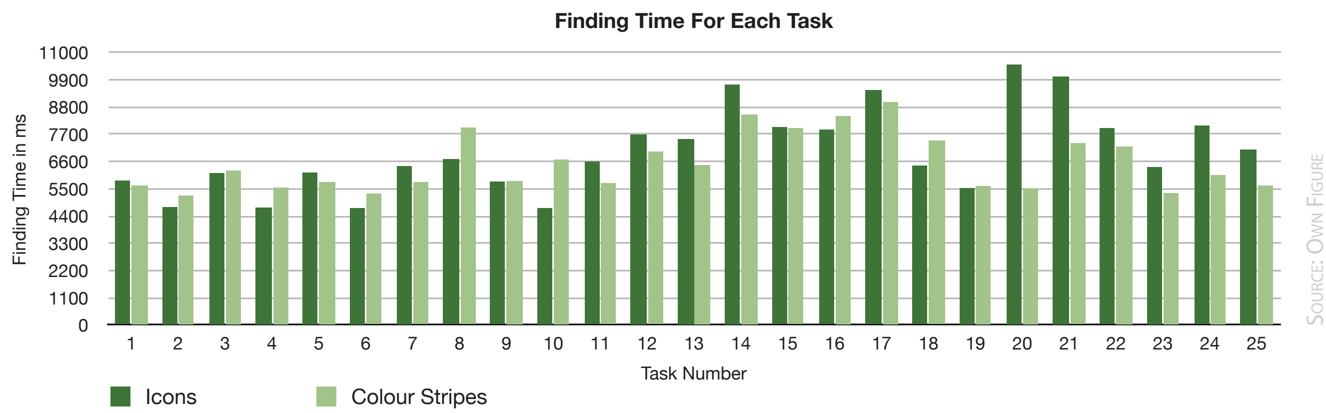


Figure 5.22: Finding time for each task when either colour stripes or icons are used as visual cues on the timeline.

missed. Participants have four options which are labeled: very often, often, occassionally and never. The results show that on average participants search more often in familiar video sources than in unfamiliar video sources as described in figure 5.25.

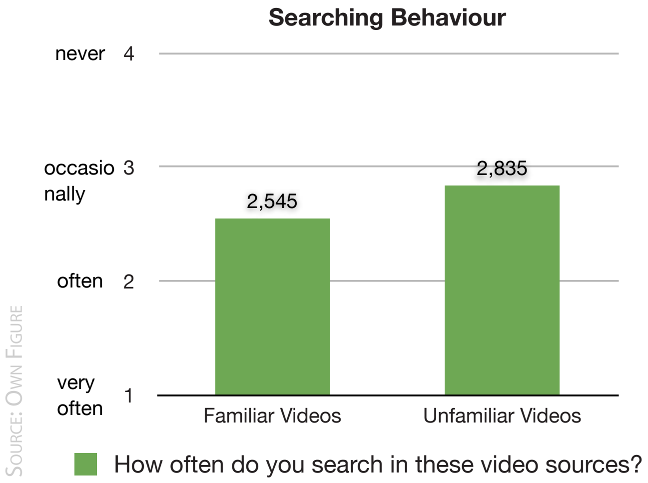


Figure 5.25: Participants search more often in familiar video sources such as a favorite tv series or a family video than in unfamiliar video sources.

Location of Scenes of Interest and Reasoning About Underlying Video Content

In the second question, participants are required to rate if the visual cues used in the user study helped to see where scenes of interest are located and if the visual cues helped to reason about the content of the underlying video scene.

Participants in both groups the colour stripes and the icons agree strongly that the visual cues are helpful to see where scenes of interest are located. However, when asked if the visual cues helped to reason about the underlying video content, participants in the icon group agree stronger than participants in the colour stripe group as described in figure 5.26.

After these two quantitative questions, participants are requested to insert their opinions and suggestions into a plain text field that can be either used for written statements or can be used for drawing sketches of visual enhancements.

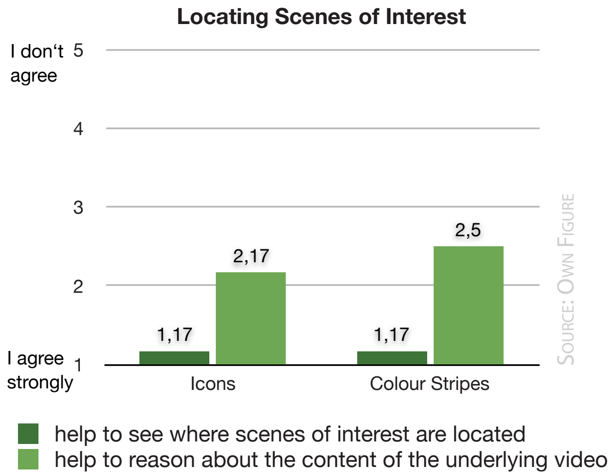


Figure 5.26: Participants opinions when asked how helpful the visual cues are for locating scenes of interest and reasoning about the video content.

Other Types of Visual Cues

The third question asks participants which other visual cues they can imagine to be positioned on a timeline to represent the content of a video scene. Four out of six participants in the colour stripe group mention that symbols or icons are better visual cues. One participant notes that he prefers colour stripes with small letters such as an „R" for referee or colour stripes that have small symbols assigned. Another participant suggests to use colour stripes for regular activities and markers such as „!" or „*" for specific events. Finally, one participant explains that speech bubbles which appear while browsing through the video would be more effective.

When the same question is asked to participants in the icon group, they mention more enhanced visual cues, such as using stills of the video scene, animations and words or abbreviations. However, none of the participants in the icon group states that colour stripes are a better means to represent the content of a video scene.

Advantages of Colour Stripes/Icons

The fourth question requires participants to state which advantages they see concerning the visual cues used on the video timeline during the user study. Participants in both groups agree that the visual cues support the finding of video scenes, that they provide an easier navigation, that they save time when searching and that they support knowing the interval and number of scenes of interest throughout the video. One participant notes that the visual cues are especially helpful for longer videos.

Additionally, participants in the colour stripe group mention that colour stripes require only little space, that they are distinguishable, easy to remember and not ambiguous.

Participants in the icon group mention that icons are easily recognizable (especially when having different shapes and colours), that they are very good scalable and that they become familiar when used more often.

Enhancements for Colour Stripes/Icons

The last question asks participants where they see room for enhancements concerning the visual cues used on the video timeline during the user study. Participants in both groups suggest that the visual cues scale to a larger size when the playhead reaches their location. This user interface concept is known, for instance, from Apple's operation system Mac OS X (Figure 5.27).

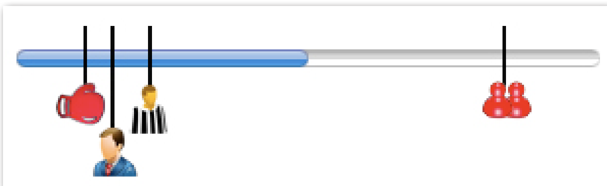
One participant proposes that the visual cues are placed in the video window and not on the timeline because he felt distracted by referring to the video timeline for searching. Other participants state that they prefer placing the visual cues directly on the scrubber bar and not on a second video timeline. Finally, one participant notes that he would like to select the visual cues directly.

Source: Mac Dock [33]

Figure 5.27: Scaling according to the mouse position as known from Mac OS X.

Participants in the colour group mainly state that they had problems to select the colour stripes due to their small size. Their suggestions include to either make all colour stripes thicker or to adjust the width according to the duration of the corresponding video scene. Two participants respond that the large set of different colours is confusing and that they would prefer using fewer colours for broader categories.

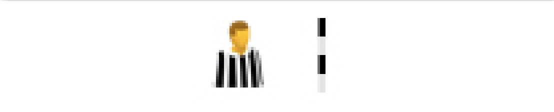
Participants in the icon group mention that some icons are not suitable for scaling. For instance, the goal icon showing a hockey puck looks like a small black dot on the video timeline when adjusted with SCADE. One participant notes that the aspect-ratio of icons has to be constant because otherwise the icons are difficult to recognize. This participant suggests to position the icons with a vertical offset to solve the space problem (Figure 5.28).



Source: Own Figure

Figure 5.28: Positioning icons with a vertical offset to gain space for displaying the icons in full size.

Another participant proposes not to change the aspect-ratio but to provide a distinct icon for each width. For instance, when the icon is thin and has the shape of a colour stripe, it still can contain textural patterns representing the original icon (Figure 5.29).



Source: Own Figure

Figure 5.29: Providing a distinct icon for each width of the visual cue by referring to textual patterns.

5.5 Interpretation Of Experiment Results

As described in the previous section, the difference between the finding time with colour stripes and with icons is very small according to our results. Additionally, colour stripes are faster than icons in the majority of cases when the finding time of each task is evaluated. Finally, our results show that with an increasing density of distribution, the finding time with icons increases more than with colours.

Based on this, adjusting the icons with our novel approach SCADE does not support the user in finding video scenes faster than with the colour stripes. However, we believe that our novel approach SCADE is advantageous compared to existing approaches, such as overlapping the icons or layering the icons with transparency, but not yet at a level of development to provide the user with faster video navigation than colour stripes.

5.5.1 Limitation of Video Length

Furthermore, our user study design was limited due to the resources available for writing this thesis. First of all, our prototype is, at present, not advanced enough to run a user study with a complex setup. Especially the memory limitations on the device caused several problems and finally we were only able to use video sequences with a length of one minute. However, when videos are short, the event frequency has to be high to be able to place enough meaningful visual cues on the video timeline for the user study tasks.

5.5.2 Uniqueness of Visual Cues

Following this, the visual cues we used on the video timeline were mostly unique since most events in the one minute video happened only once. For instance, if a task required the participant to find „the man in the audience with the yellow poster" there was only a single audience icon on the timeline due to the fact that the audience was only shown once in the video.

During the user study, several participants stated that they often paid no attention to the video window and used only the video timeline with visual cues to find the correct video scene since solving the task did not require to monitor the video window. For few tasks, we were able to integrate several visual cues of the same type into a single video sequence. Based on this, participants were required to check if the visual cue corresponds to the video scene the task was asking for. For instance, in one video sequence, we used two audience cues: one corresponding to two girls and one corresponding to a man. When participants were asked for one of these scenes, they were required to review each audience cue and its corresponding video scene to make sure they found the correct video scene.

Furthermore, when the length of video sequences is limited to one minute, the finding time of video scenes is very short. Following this, differences concerning the finding time between icons and colour stripes are less obvious than with longer video sequences.

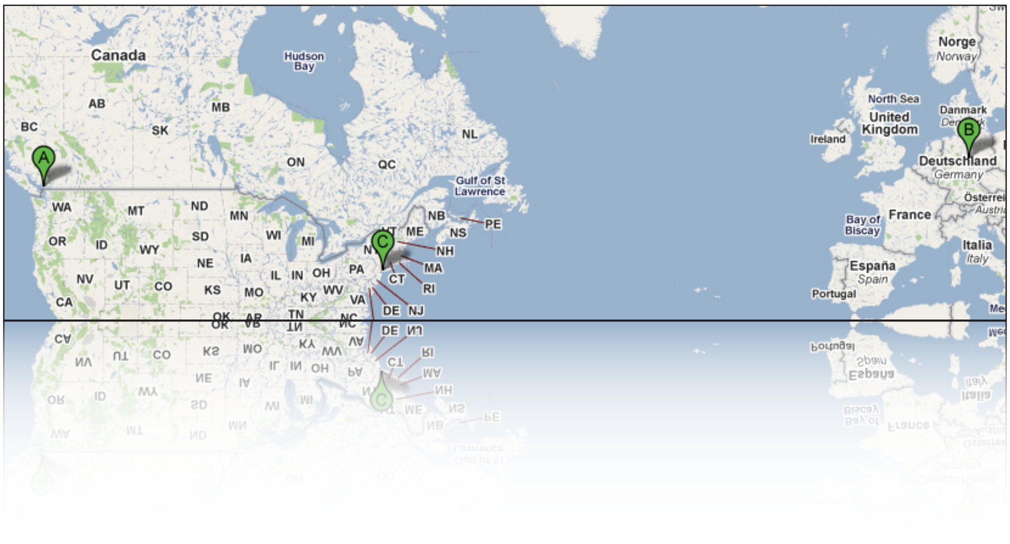
5.5.3 Task Complexity

Additionally, the legend paper sheet made the completion of tasks simple since participants were only required to find a certain visual cue, but not to memorize its meaning. Basically, using a legend paper sheet is an appropriate approach due to the fact that we do not want to measure how well participants memorize the meanings of colour stripes and icons, but aim at measuring how quickly participants can locate visual cues on a video timeline. However, if participants can refer to the meanings of visual cues at each moment, search tasks have to be more complex than those we used in our user study. Task complexity can be increased, for instance, by using longer videos, by raising the number of visual cues of the same type and by distributing the visual cues more densely on the video timeline.

With regard to the insights we gained in our user study, we believe that future evaluation will deliver more meaningful results concerning the usefulness of our novel approach SCADE. Future work will cover further development of our prototype to run a more complex user study setup and to expand our knowledge on video browsing with visual cues.

Chapter 6

Conclusion and Future Work



Source: Google Maps [28]

6. Conclusion and Future Work

In the following section, we review the work covered in this thesis, make suggestions for future work and give information about a publication we submitted to a conference on user interface software and technology.

6.1 Overview of the Research Process

In this thesis, we presented a novel user interface concept called TAV that is developed as a part of the MyView project to further personalize video browsing. Due to the massive growth of video content online and offline, advanced user interfaces are needed to provide the user with effective means for browsing video collections according to his personal interests. Recent literature contains many examples how large video collections can be searched effectively. However, we believe that the search in a single video source has to be explored in more detail.

People often access video to look for a specific scene, for instance, to find a part of a favorite tv series or to search in a sports video for the highlights of the game. Existing literature on video search and navigation has mainly focused on the textual approach, such as using keyword input into a search engine. However, only little work has examined effective visualization approaches. In our work, we focused on the usage of visual cues that are positioned on a video timeline. Each visual cue represents the content and temporal location of a video annotation. Based on this, our novel user interface concept TAV enables the user to view the temporal video annotations simultaneously with their related video scene. Furthermore, TAV allows the user to apply filter mechanisms to the visual cues to only display the temporal video annotations that are of interest for him.

Most existing approaches for visual video navigation with timelines make use of colour stripes to show where interesting video scenes are located. However, colour stripes are abstract and difficult to distinguish due to a limited range of distinct colours. We suggested to use icons since they incorporate more semantic information and are better recognizable.

However, while colour stripes only need a single pixel in width to represent a video scene, icons need more space on the video timeline which soon becomes a problem with regard to the limited screen real-estate, especially on mobile devices. To address this problem, we developed a novel approach called SCADE that adjusts icons according to the available screen space.

We conducted a user study based on TAV and Scade to examine the trade-off between detailed semantic information and limited screen real-estate. In our user study, we measured the finding time of video scenes when colour stripes or icons are used on the video timeline. However, due to our limited user study setup we were not able to find enough evidence that icons adjusted with SCADE provide faster video navigation. We believe that the knowledge we gained during this first user study will help to further the development of our prototype. Following this, we will be able to run a more complex user study setup to find more evidence for the usefulness of SCADE.

6.2 Categorization: Colour and Pattern

While it was appropriate to evaluate only colour stripes and icons due to the limits of this thesis, we think that a third category of visual cues has to be taken into account to examine the trade-off between detailed semantic information and limited screen real-estate. Beside the evaluation of colour stripes (colour, no pattern) and icons (colour and pattern), we also have to consider the usefulness of black and white icons (no colour, but pattern). Following this, we will be able to reason in more detail how colour and pattern contribute to faster video navigation.

6.3 Integration of Tav in the MyView Project

TAV is one of the four user interface concepts for personalized video browsing developed in the MyView project. In this thesis, we focused on developing a first version of TAV. Future work will cover the integration of TAV into the existing user interface concepts to make possible new ways of interacting with video. For instance, the combination of the object selection with the filter mechanisms of TAV provides the user with a novel approach to define for which object in the scene he wants to see visual cues on the timeline. If a user selects a hockey player in the scene, TAV shows only the visual cues concerning the selected player, such as visual cues showing where the player scored or where the player got a penalty.

6.4 Uist Poster Submission

On June 30, 2010, we submitted a poster to the ACM Symposium on User Interface Software and Technology called UIST which is the premier forum for innovations in the software and technology of human-computer interfaces. The poster represents our on-going work on temporal video annotation and will be presented from October 3 to 6, 2010, in New York City as part of UIST 2010. Attached to the poster is a two pages abstract that emphasizes the importance of temporal video annotation as a recent research topic and states our novel solutions for the recent challenges in navigating time-based media, such as video. The feedback of the reviewer that oversaw our poster provided us with additional sources of related work and included advices for our on-going research. For future development of our novel interface concept TAV, we hope to gain new insights during the poster presentation and to meet researchers that work on similar problems.

6.5 Research Term at Hct

Our work at the HCT Laboratory was an experience that helped us to improve our research skills throughout the whole research process. We are confident that the knowledge we gained in the last seven months will be of important value for further research. For the future, we hope that the collaboration between the HCT Laboratory of UBC and the degree program Computer Science in Media of HH will enable other students to progress on their research skills.

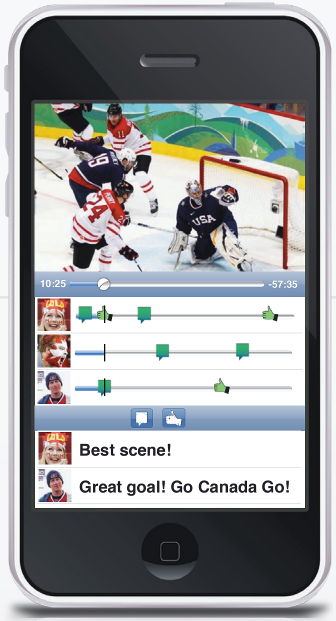
Attachments

Using Temporal Video Annotation as a Navigational Aid for Video Browsing

Stefanie Müller | Gregor Miller | Sidney Fels | Human Communication Technology Laboratories (HCT) | University of British Columbia | {smueller, gregor, ssfels}@ece.ca

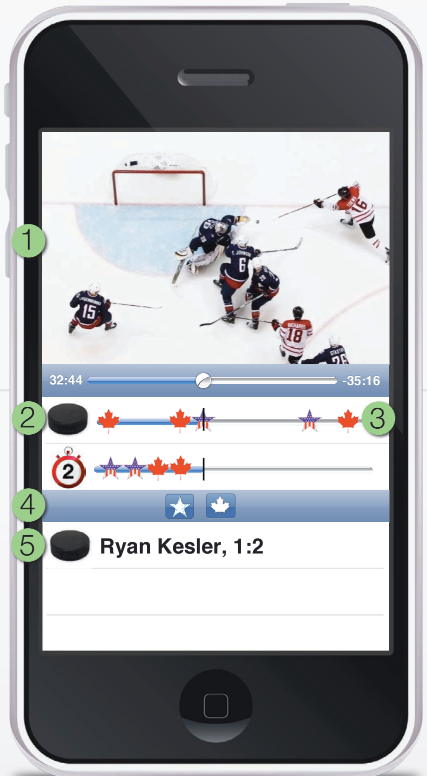
Introduction

- video is a complex information space
- requires advanced navigational aids
- temporal video annotation offers new opportunities for navigating video
- novel video browsing interface: a visual overview of temporal video annotations

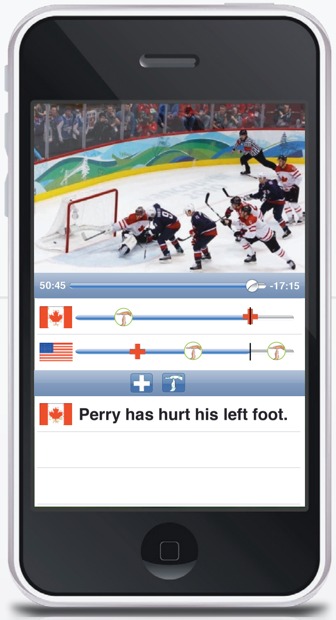


User Interface

- one annotation layer for each annotation type
- a layer contains multiple icons showing the temporal position and type of annotations
- annotations can be filtered
- contents of annotations are displayed according to video playback



1. Video
2. Annotation Layer
3. Annotation Mark
4. Annotation Filter
5. Annotation Viewing



User Benefits

Enables the user to
- determine the general content of a video
- locate scenes of interest
- watch annotations simultanously with the video
- define which annotations are displayed while watching

Conclusion

- a visual overview of temporal video annotations provides the user with additional navigational cues for browsing video
- future work: user study to evaluate novel user interface concept

Using Temporal Video Annotation as a Navigational Aid for Video Browsing

Stefanie Müller, Gregor Miller and Sidney Fels
Dept. of Electrical and Computer Engineering
University of British Columbia
{smueller, gregor, ssfels}@ece.ubc.ca

ABSTRACT
Video is a complex information space that requires advanced navigational aids for effective browsing. The increasing number of temporal video annotations offers new opportunities to provide video navigation according to a user's needs. We present a novel video browsing interface called TAV (Temporal Annotation Viewing) that provides the user with a visual overview of temporal video annotations. TAV enables the user to quickly determine the general content of a video, the location of scenes of interest and the type of annotations that are displayed while watching the video. An ongoing user study will evaluate our novel approach.

ACM Classification: H5.2 [Information interfaces and presentation]: User Interfaces. - Graphical user interfaces.

General terms: Design, Human Factors

Keywords: Video Browsing, Video Navigation, Video Annotation, Video Search

INTRODUCTION

Although video annotation is a growing phenomenon, the current approach for viewing and finding video annotations is still at a basic level. Even though many annotations refer to a specific subset (or scene) of the video and are therefore temporal in nature, the established approach does not take this characteristic into account. On popular video platforms such as YouTube, video comments (a widely used type of annotation) are displayed in a single entry list that does not change during playback. However, we believe that this static approach is not appropriate for a time-based medium such as video. We propose that video annotations be displayed simultaneously with the scene to which they refer.

Furthermore, enabling the user to add temporal video annotations also provides new opportunities for browsing video. The static approach for visualizing video annotations cannot be easily used to provide navigational cues. However, with a temporal approach video annotations can be used to support the user with additional information such as the scene's content and locations of interest. We believe that providing the user with a visual overview of temporal video annotations improves video navigation speed.

In addition to the problem of non-temporal video annotation there is also a lack of filtering mechanisms. The increasing number of annotations requires an interface that enables the user to define which annotations are relevant. For example, only the annotations made by a specific person (such as a friend) or the annotations related to a specific event in the video (a goal in a sporting event).

In the following sections we present related work in the field of video browsing with visual cues and introduce our novel approach called TAV.

RELATED WORK

The existing literature contains many examples of research on visualizing the underlying video content either by extracting image-based features such as dominant color, sound volume or motion [3] or by interpreting the video's content, for example in the form of visualizing character's emotions [1]. However, none of this work has focused on using temporal video annotation as a means to provide navigational aids based on a video's content. We believe that the increasing number of temporal video annotations is a valuable source for navigational cues while browsing through a video.

Furthermore, current visualization approaches do not support the user with effective filter mechanisms. Costa et al. [2] emphasized how temporal annotations can provide multiple perspectives on a video. We believe that enabling the user to choose the visual cues according to his needs will result in a more efficient browsing experience.

NAVIGATING VIDEO WITH VISUAL CUES

We have developed TAV, a novel video browsing interface that provides the user with a temporal overview of video annotations including visual cues that represent the underlying video content.

This poster was submitted to UIST 2010 on June 30, 2010, as a presentation of on-going work. The poster was accepted on July 23, 2010, and minor changes according to the comittee were made until the camera-ready deadline on August 13, 2010.

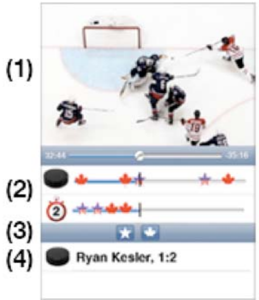


Figure 1: TAV User Interface, (1) traditional video player, (2) annotation layer, (3) annotation filter, (4) annotation viewing

Figure 2: A thumbnail's size below the playhead indicates the number of video annotations in a scene to identify scenes of interest.

Searching Video by Annotations

TAV's user interface, an example of which is shown in Figure 1, consists of a traditional video player (1) vertically adjacent to our novel browsing interface. Below the video player multiple timelines (2) are displayed. Each timeline refers to a specific type of video annotation, in this case goals or penalties in a hockey game. Each timeline consists of a visual identifier on the left side followed by a scrub bar with multiple annotation icons. Annotations are located according to their temporal appearance in the video and can be of different types – our example shows icons representing Canada and USA in a hockey game. Based on this, the user cannot only see where goals occurred in the video but also which team made the goal. The user can drag the playhead to advance to a scene of interest.

Viewing Video with Annotations

While the video is playing the video annotations are shown in the viewing table (4). This enables the user to watch annotations simultaneously with the video scene they refer to. Annotations are added to and removed from the viewing table according to the playhead's position on the video timeline.

Filtering Annotations

In order to provide the user with the ability to adjust the visual overview according to their needs we integrated two filter mechanisms. A filter toolbar (3) enables the user to determine which types of annotations are displayed on the timelines. For instance, if a user is only interested in goals and penalties referring to Canada, he can deactivate USA's annotations by selecting the related filter button. A second filter method can be applied by deactivating a timeline. For instance, if a user is only interested in goals but not in penalties, he can deactivate the second timeline by selecting it on the left side.

Using Annotations to Identify Interesting Scenes

Besides the qualitative use of video annotation as a visual cue for representing the underlying video content, we also provide quantitative navigational aids. A high number of annotations added to a video scene usually indicates that many people found this scene interesting. TAV visualizes the quantitative distribution of video annotations by placing different scaled thumbnails below the video timeline (Figure 2). This enables the user to quickly determine which scenes got the most attention by other viewers and therefore might be of interest. We addressed the limited space problem by providing the user with zooming functionality that allows him to define the level of detail according to his needs.

CONCLUSION

We presented a novel video browsing interface called TAV that enables the user to navigate through video by using a visual overview of temporal video annotations. We emphasized the potential of temporal video annotations as a valuable source for navigational cues and demonstrated how simple filter mechanisms can help the user to focus on information relevant to him while browsing through a video. Future work will include the evaluation of our novel user interface concept by conducting a user study.

REFERENCES

1. L. Chen, G.C. Chen, C.Z. Xu, J. March, and S. Benford. EmoPlayer: A media player for video clips with affective annotations. *Interacting with Computers*, vol. 20, no. 1, pp. 17-28, 2008
2. M. Costa,, N. Correia, N. Guimaraes. Annotations as Multiple Perspectives of Video Content. In *Proceedings of the Tenth ACM International Conference on Multimedia*, pp. 283-286, 2002
3. K. Schoeffmann, M. Lux, M. Taschwer, and L. Boeszoermenyi. Visualization of Video Motion in Context of Video Browsing. In *Proceedings of the IEEE International Conference on Multimedia and Expo*, 2009

Certificate of Completion

This is to certify that

Stefanie Mueller

has completed the Interagency Advisory Panel on Research Ethics'
Introductory Tutorial for the
Tri-Council Policy Statement: Ethical Conduct for Research Involving Humans (TCPS)

Issued On: June 19, 2010

TCPS Certificate of Completion.

THE UNIVERSITY OF BRITISH COLUMBIA

Media and Graphics Interdisciplinary Centre
Forest Sciences Centre Building
FSC 3640 - 2424 Main Mall
Vancouver, BC V6T 1Z4
Tel: 1-604-822-8990 Lab: 1-604-822-9248

Consent Form

On the Usage of Temporal Video Annotation as a Navigational Aid for Video Browsing

Principal Investigators

Dr. Sidney Fels, Associate Professor, Department of Electrical and Computer Engineering, University of British Columbia (604) 822-5338

Student Investigators

Stefanie Mueller, Bachelor student, Human Communication Technologies Laboratories (HCT), University of British Columbia (604) 822-4583

INVITATION TO PARTICIPATE

You are being invited to take part in this research study because you are a healthy adult, between the ages of 19 and 50 years and you have either normal vision or wear corrected lenses.

OVERVIEW

This study is designed to examine the usefulness of temporal video annotation as a navigational aid for browsing video. As part of our continuing research program, our interest is to consider the personal video browsing experience thoroughly. We are conducting investigations to extend existing user interface guidelines for customized browsing of complex video spaces. The purpose of this study is to create a user interface concept for browsing a video by using temporal video annotation as a navigational aid.

PROCEDURES

In this experiment, you will be asked to find a specific video scene within a video by using the provided interface for browsing. The video will be presented on a mobile device with touchscreen (iPhone) and you will use your fingers to navigate through the video. Your primary goal will be to respond as fast as possible.
A camera will monitor your responses. The data resulting from the tasks will be used to study the usage of the interface in detail. This information will help us to understand how using temporal video annotation as a navigational aid supports video browsing.
At the end of the experiment you will be asked to fill in a short questionnaire concerning the different user interfaces that you used during the experiment.

RISKS

The risks involved in participating in this experiment will be minimal. That is, the risks are no greater than the risks in everyday life. You might experience some slight fatigue, as you will be asked to maintain focused attention and concentration throughout the experiment.

CONFIDENTIALITY

Any information resulting from this research study will be kept confidential. All documents will be identified only by a code number and kept in a locked filing cabinet in the principal investigators research office. You will not be identified by name in any reports or scientific publications of the completed study. All backup computer files will be kept in a locked filling cabinet, and any data files that reside on the data analysis computer in the Media and Graphics Interdisciplinary Centre (room 3641, Forestry Center, UBC), will be number coded. Only Dr. Fels and his research assistants will have password access to these files. The data will be used for publication purposes only, and will be retained for a minimum of five years post-publication or completion of the study, after which time raw data will be destroyed.

Your confidentiality will be respected. No information that discloses your identity will be released or published. Research records identifying you may be inspected in the presence of the Investigator by representatives or the Natural Sciences and Engineering Research Concil of Canada, or the UBC Research Ethics Board for the purpose of monitoring the research. However, no records which identify you by name or initials will be allowed to leave the Investigators' offices.

RIGHTS TO WITHDRAW FROM THE RESEARCH

Your participation is voluntary and you may withdraw at any time without consequences. Data collected up to the point of your withdrawal from the study will be kept for data analysis purposes under strict provisions of confidentiality.

COMPENSATION

You will receive a $10 compensation for your participation in this project.

Contact Information About the Project

If you have any questions or require further information about the project you may contact Sidney Fels at (604) 822-5338 or Stefanie Mueller at (604) 822-4583.

Contact For Information About The Rights Of Research Subjects

If you have any concerns about your treatment or rights as a research subject, you may contact the Research Subject Information Line in the UBC Office of Research Services at 604-822-8598.

CONSENT

Your signature below indicates that you have received a copy of this consent form for your own records.

Your signature indicates that you consent to participate in this project. You do not waive any legal rights by signing this consent form.

I, ________________________________, agree to participate in the project as outlined above. My participation in this project is voluntary and I understand that I may withdraw at any time.

____

Participant's Signature Date

____

Student Investigator's Signature Date

	The University of British Columbia *Office of Research Services* ***Behavioural Research Ethics Board*** *Suite 102, 6190 Agronomy Road* *Vancouver, BC V6T 1Z3*

H10-01894 **Moving Hyperlinks - Temporal Annotation Viewing (Version 1.0)**

Principal Investigator: Sidney S. Fels

1. Principal Investigator & Study Team - Human Ethics Application [View Form]

1.1. Principal Investigator Please select the Principal Investigator (PI) for the study. Once you hit Select, you can enter the PI's name, or enter the first few letters of his or her name and hit Go. You can sort the returned list alphabetically by First name, Last name, or Organization by clicking the appropriate heading.	Last Name / First Name / Employer.Name / Email Fels / Sidney S. / Electrical and Computer Engineering / ssfels@ece.ubc.ca
Enter Principal Investigator Primary Department and also the primary location of the PI's Institution:	
1.2. Primary Contact Provide the name of ONE primary contact person in addition to the PI who will receive ALL correspondence, certificates of approval and notifications from the REB for this study. This primary contact will have online access to read, amend, and track the application.	Last Name / First Name / Rank Miller / Gregor / Research Fellow
1.3. Co-Investigators List all the Co-Investigators of the study. These members WILL have online access which will allow them to read, amend and track the application. These members will be listed on the certificate of approval (except BC Cancer Agency Research Ethics Board certificates). If this research application is for a graduate degree, enter the graduate student's name in this section.	Last Name / First Name / Institution/Department / Rank Mueller / Stefanie / UBC/Applied Science/Electrical and Computer Engineering / Visitor
1.4. Additional Study Team Members - Online Access List the additional study team members who WILL have online access to read, amend, and track the application but WILL NOT be listed on the certificate of approval.	Last Name / First Name / Institution/Department / Rank
1.5. Additional Study Team Members - No Online Access Click Add to list study team members who WILL NOT have online access to the application and will NOT be listed on the certificate of approval.	Last Name / First Name / Institution / Department / Rank / Job Title / Email Address
1.6.1. All undergraduate and graduate students and medical residents are expected to complete the TCPS Tutorial before submission. It is strongly recommended that the Principal Investigator and all Co-Investigators are familiar with the TCPS. Indicate completion of the TCPS tutorial below: All Undergraduate/Graduate Students:	Yes

Ethics approval form for the BREB.

<table>
<tr><td>1.6.2. All Medical Residents:</td><td>N/A (no medical residents participating in this study)</td></tr>
<tr><td>Comments:</td><td></td></tr>
<tr><td>1.7. Project Title Enter the title of this research study as it will appear on the certificate. If applicable, include the protocol number in brackets at the end of the title.</td><td>Temporal Annotation Viewing</td></tr>
<tr><td>1.8. Project Nickname Enter a nickname for this study. What would you like this study to be known as to the Principal Investigator and study team?</td><td>Moving Hyperlinks - Temporal Annotation Viewing</td></tr>
<tr><td>NOTE, if this application was converted to RISe from our previous database, ORSIL, here is the previous ORSIL application number for your information.</td><td></td></tr>
<tr><td colspan="2">2 Study Dates and Funding Information - Human Ethics Application [View Form]</td></tr>
<tr><td>2.1. A. Start date:</td><td>August 1, 2010</td></tr>
<tr><td>2.1. B. End date:</td><td>August 31, 2012</td></tr>
<tr><td>2.2. Types of Funds Please select the applicable box(es) below to indicate the type(s) of funding you are receiving to conduct this research. You must then complete section 2.3 and/or section 2.4 to enter the name of the source of the funds to be listed on the certificate of approval.</td><td>Grant</td></tr>
<tr><td>If you selected Other, specify the type of funding below.</td><td></td></tr>
<tr><td>2.3. Research Funding Application/Award Associated with the Study Submitted to the UBC Office of Research Services Please click Add to identify the research funding application/award associated with this study. Selecting Add will list the sources of all research funding applications that have been submitted by the PI (and the person completing this application if different from the PI). If the research funding application/award associated with this study is not listed below, please enter those details in question 2.4.</td><td>
<table>
<tr><th>UBC Number</th><th>Title</th><th>Sponsor</th></tr>
<tr><td>F09-03200</td><td>Diving experiences: wayfinding and sharing experiences with large, semantically tagged video</td><td>Bell University Laboratories</td></tr>
<tr><td>05-3714</td><td>Using a parallel distributed camera array for on-demand user-selected video</td><td>Bell University Laboratories</td></tr>
<tr><td>F09-03200</td><td>Diving experiences: wayfinding and sharing experiences with large, semantically tagged video</td><td>Natural Sciences and Engineering Research Council of Canada (NSERC)</td></tr>
<tr><td>F08-04994</td><td>Future Digital Lifestyle Research</td><td>Natural Sciences and Engineering Research Council of Canada (NSERC)</td></tr>
<tr><td>05-3714</td><td>Using a parallel distributed camera array for on-demand user-selected video</td><td>Natural Sciences and Engineering Research Council of Canada (NSERC)</td></tr>
<tr><td>05-2780</td><td>MyView: Using a Parallel Distributed Camera Array for On-Demand User-Selected Video</td><td>Bell Canada</td></tr>
</table>
</td></tr>
<tr><td>2.3.1. Is this a DHHS grant?</td><td></td></tr>
</table>

2.3.2. If yes, please select the appropriate DHHS funding agency from the selection box, and attach the grant application.	DHHS Sponsor List:
Attach DHHS Grant Application for each sponsor listed above	
2.4. Research Funding Application/Award Associated with the Study not listed in question 2.3. Please click Add to enter the details for the research funding application/award associated with this study that is not listed in question 2.3.	UBC Number Title Sponsor
2.4.1. Is this a DHHS grant?	
2.4.2. If yes, please select the appropriate DHHS funding agency from the selection box, and attach the grant application.	DHHS Sponsor List:
Attach DHHS Grant Application for each sponsor listed above	
2.5. Conflict of Interest Do any of the following statements apply to the Principal Investigator, Co-Investigators and/or their partners/immediate family members? Receive personal benefits in connection with this study over and above the direct cost of conducting this study. For example, being paid by the funder for consulting. (Reminder; receiving a finders fee for each subject enrolled is not allowed). Have a non-financial relationship with the sponsor (such as unpaid consultant, advisor, board member or other non-financial interest). Have direct financial involvement with the sponsor (source of funds) via ownership of stock, stock options, or membership on a Board. Hold patent rights or intellectual property rights linked in any way to this study or its sponsor (source of funds).	no
4. Study Review Type - Human Ethics Application **[View Form]**	
4.1. UBC Research Ethics Board Indicate which UBC Research Ethics Board you are applying to and the type of study you are applying for:	UBC Behavioural Research Ethics Board - Behavioural
4.2. Institutions and Sites for Study A. Enter the locations for the institutions and sites where the research will be carried out under this Research Ethics Board approval (including specimens processed by pathology, special radiological procedures, specimens obtained in the operating room, or tissue requested from pathology). Click Add and enter the appropriate letter to see the locations for the institutions and sites where the research will be carried out under this Research Ethics Board approval: B for BC Cancer Agency C for Children's and Women's Health Centre of BC P for Providence Health Care U for UBC Campus V for Vancouver Coastal Health (VCHRI/VCHA). If you are NOT using any of these sites select N/A from the list.	Institution Site UBC Vancouver (excludes UBC Hospital)

B. Please enter any other locations where the research will be conducted under this Research Ethics Approval (e.g. private physician's office, community centre, school, classroom, subject's home, in the field - provide details).	
4.3. A. If this proposal is closely linked to any other proposal previously/simultaneously submitted, enter the Research Ethics Board number of that proposal.	H09-00066
B. If applicable, please describe the relationship between this proposal and the previously/simultaneously submitted proposal listed above.	It is nearly identical except for a change in interface elements that the study is investigating. Also, a different student is running the experiment.
C. Have you received any information or are you aware of any rejection of this study by any Research Ethics Board? If yes, please provide known details and attach any available relevant documentation in question 9.8.	
4.4. If this research proposal has received any independent scientific/methodological peer review, please include the names of committees or individuals involved in the review. State whether the peer review process is ongoing or completed. A. External peer review details:	NSERC reviewed
B. Internal (UBC or hospital) peer review details:	none
C. If this research proposal has NOT received any independent scientific/methodological peer review, explain why no review has taken place.	it has been reviewed
4.5. After reviewing the minimal risk criteria on the right, does your application fall under minimal risk (and therefore is eligible to be considered for Delegated Review, executive review or review by an Undergraduate Research Review Committee)?	yes
4.6.A. Pandemic Research Does this study involve research concerning H1N1 or any other urgent public health event such that it requires urgent review and approval? [if no, move on to 5, if yes, answer 4.6B]	no
4.6.B. Does this pandemic study require review and approval by multiple Canadian Research Boards (i.e. more than those covered under the certificate of approval for this application) [If no, move on to 5, if yes, answer 4.6.C]	
4.6.C. Are you the Lead Investigator for this pandemic study? (i.e. the pandemic study involves numerous co-investigators from various sites external to UBC and you have been selected as the lead investigator for the entire project) [If YES, move on to 5, if NO move on to 4.7]	
4.7. Pandemic Research Lead PI REB Please review the guidance note on the right and then answer the following question: If the study has	

NOT been approved by the Lead PI's REB, UBC's REBs will not proceed to review the study independently. They will be participating in the Lead REB approval process and accordingly, your application is premature. Please discontinue this application and submit a new application as soon as the study approval by the Lead PI REB has been obtained. If the study HAS been approved by the Lead PI's REB, UBC's REBs will make every effort to review your study as quickly as possible. In order to ensure that the required documentation is incorporated into the RISe system, you will be directed to respond to Question 9. For more information please see the accompanying guidance note. Has this study been reviewed and approved by the Lead Principal Investigator's REB?	
4A. Study Review Type - Undergraduate Behavioural Research [View Form]	
4. A1. Has this study been approved by another Canadian Research Ethics Board?	no
If Yes, provide the name of the Research Ethics Board (REB) and the REB contact information below and proceed to the next page. Attach all relevant documentation in Section 9 of the form, including all documents submitted to the other Canadian REB. The application and correspondence between the researcher and the REB must be attached in Question 9.8. If No complete question 4. A2.	
4. A2. Does this study involve individual, honours thesis or course based research by UNDERGRADUATE students that is being conducted as part of an undergraduate course offered by The University, that is NOT PART OF A FACULTY MEMBER'S research program	no
If Yes, please select the applicable Undergraduate Student Research Review Committee from the list of established committees below. NOTE: There are currently no committees established, so please select No Research Committees Available:	
5. Summary of Study and Recruitment - Human Ethics Application [View Form]	
5.1.A Provide a short summary of the project written in lay language suitable for non-scientific REB members. DO NOT exceed 100 words and do not cut and paste directly from the study protocol.	
5.1.B Summarize the research proposal:	Video is a complex temporal information space that requires advanced navigational aids for browsing. A lot of work has been done to explore how textual search (e.g. keywords) can support finding information in a video. However, only little work has focused on solving the problem visually. Work in the field of visual video navigation has either enhanced existing user interface elements or invented new user interface concepts. For this experiment, we focus on the existing user interface element video timeline by reason that

	users are already familiar with this video player component. Existing literature on enhanced video timelines has mostly focused on using colour as a means to provide navigational cues. However, colour provides only little semantic information on the underlying video content. Other work has concentrated on the use of icons as a more intuitive means of representing the semantic information of the underlying video content. However, icons need more space when placed on a video timeline and when the number of navigational cues on the timeline increases the information of icons may no longer be accessible for the user. The problem of limited screen real-estate becomes even more significant due to an emerging trend for consuming video on mobile platforms with small screens. In this experiment, we evaluate our solution for the conflict between displaying enhanced semantic information with icons and the limited screen real-estate. Subjects will be required to perform simple tasks of finding a specific video scene and pointing on a video timeline. For study purposes, we will use video material from an ice hockey game as broadcasted regularly on TV. The video material will not be considered as disturbing, upsetting, shocking or unpleasant by subjects. We compare the existing user interface concept with colours to our novel interface concept that displays and scales icons according to the available screen real-estate. An application will be developed to provide a video player with a visually enhanced timeline to record video scene retrieval speed. A new model for displaying icons as visual cues in enhanced timelines will be developed to extend previous research on visual video navigation. We are developing the software and user interface ourselves in order to explore these principles and determine under which circumstances our user interface is most beneficial or most preferred.
5.2. Inclusion Criteria. Describe the subjects being selected for this study, and list the criteria for their inclusion. For research involving human pluripotent stem cells, provide a detailed description of the stem cells being used in the research.	Subjects will be adult males or females aged 19 years or older and must have good vision (corrective lenses are acceptable) and the motor skills necessary to operate a touchscreen to acquire targets on a mobile device.
5.3. Exclusion Criteria. Describe which subjects will be excluded from participation, and list the criteria for their exclusion.	The visually impaired or subjects deficient in motor skills.
5.4. Provide a detailed description of the method of recruitment. For example, describe who will contact prospective subjects and by what means this will be done. Ensure that any letters of initial contact or other recruitment materials are attached to this submission on Page 9.	The contact information will be collected by the co-investigator and bachelor student Stefanie Mueller, who will also make the initial contact with the subjects. Participants will be contacted by email to inquire about possible participation. We are going to recruit people from the lab as well as people from the street.
5.5. Describe how prospective normal/control subjects will be identified, contacted, and recruited, if the method differs from the above.	

5.6 If existing records (e.g. health records, clinical lists or other records/databases) will be used to IDENTIFY potential subjects, please describe how permission to access this information, and to collect and use this information will be obtained.	
5.7. Summary of Procedures	Subjects will be asked to do several tasks involving the acquisition of targets on a mobile device screen using alternate display modes, a colour mode and an icon mode. We will measure how well they perform on these tasks (speed and accuracy) and determine their personal preferences.
6. Subject Information and Consent Process - Human Ethics Application [View Form]	
6.1. How much time will a subject be asked to dedicate to the project beyond that needed for normal care?	1 hour
6.2. If applicable, how much time will a normal/control volunteer be asked to dedicate to the project?	
6.3. Describe what is known about the risks (harms) of the proposed research.	None. The task is similar to a person working on a mobile device.
6.4. Describe any potential benefits to the subject that could arise from his or her participation in the proposed research.	Testing hand-eye coordination and determining strategies for efficient target acquisition.
6.5. Describe any reimbursement for expenses (e.g. meals, parking, medications) or payments/gifts-in-kind (e.g. honoraria, gifts, prizes, credits) to be offered to the subjects. Provide full details of the amounts, payment schedules, and value of gifts-in-kind.	At the end of the experiment, subjects receive a $10 compensation for their time, whichever they completed or not the requested tasks.
6.6. Specify who will explain the consent form and invite the subject to participate. Include details of where the consent will be obtained, and under what circumstances.	Student researcher when subject contacts him/her to be a subject. The consent form will be available when subject arrives at the experiment room. The experiment and the form will be explained to them and the interactive video player used for the testing will be shown to them.
6.6.A. If you are asking for a waiver or an alteration of the requirement for subject informed consent please justify the waiver or alteration and confirm that the study meets the criteria on the right.	
6.7. How long after receiving the consent form will the subject have to decide whether or not to participate? If this will be less than twenty-four hours, provide an explanation.	Subjects will receive the consent form via email at least 24 hours before the actual experiment.
6.8. Will every subject be competent to give fully informed consent on his/her own behalf? Please click Select to complete the question and view further details.	Will subject be competent to give fully informed consent? | Details of the nature of the incompetence | If not, who will consent on his/her behalf? | If not, will he/she be able to give assent to participate? | If Yes, explain how assent will be sought. Yes [Details]
6.9. Describe any situation in which the renewal of consent for this research might be appropriate, and how this would take place.	

6.10. What provisions are planned for subjects, or those consenting on a subject's behalf, to have special assistance, if needed, during the consent process (e.g. consent forms in Braille, or in languages other than English).	
6.11. Describe any restrictions regarding the disclosure of information to research subjects (during or at the end of the study) that the sponsor has placed on investigators, including those related to the publication of results.	
7. Number of Subjects - Human Ethics Application for Behavioural Study **[View Form]**	
7.1. Indicate external approvals below: A. Other Institutions:	no
B. Please select Add to enter the name of the institution and if you have already received approval attach the approval letter.	Name of Institution
C. Other Jurisdiction or Country:	no
D. Please select Add to enter the name of the jurisdiction or country and if you have already received approval attach the approval letter.	Name of Jurisdiction or Country
E. Has a Request for Ethics Approval been submitted to the institution or responsible authority in the other jurisdiction or country? (Send a copy to the Research Ethics Office when approval is obtained).	no
F. If a Request for Approval has not been submitted, provide the reasons below:	
G. Does this research involve aboriginal communities or organizations; or aboriginals as an identified subject category?	
If YES, ensure that you are familiar with the guidance documents linked on the right. Also attach a copy of the research agreement with the community (if available) in Question 9.8. Please describe the community consent process. If no community consent is being sought, please justify.	
7.2. A. How many subjects (including controls) will be enrolled in the entire study? (i.e. the entire study, world-wide)	20
B. How many subjects (including controls) will be enrolled at institutions covered by this Research Ethics Approval? (i.e. only at the institutions covered by this approval)	
Of these, how many are controls?	
7.3. Are any of the following procedures or methods involved in this study? Check all that apply.	None of these Methods
7.4. Who will actually conduct the study and what are their qualifications to conduct this kind of research? (e.g., describe relevant training, experience, degrees, and/or courses).	Stefanie Mueller, Bachelor Student will conduct the study. Stefanie has taken relevant courses in Interface Design, Human Computer Interface and Interface Laboratory in the Hochschule Harz, University of Applied Sciences (Germany). She completed the TCPS on June 19 2010. Stefanie will also benefit from the expertise of primary advisor Dr. Sidney Fels for conducting this HCI experiment.

8. Confidentiality - Human Ethics Application for Behavioural Study [View Form]	
8.1. Security of Data during the course of the study How will data be stored? (e.g. computerized files, hard copy, videotape, audio recordings, PDA, other) How will security of the data be maintained? (For example, study documents must be kept in a secure locked location and computer files should be password protected and encrypted, data should not be stored or downloaded onto an unsecured computer or portable lap-top, backup files should be stored appropriately). If any data or images are to be kept onthe Web, what precautions have taken to prevent it being copied?	Any information resulting from this research study will be kept confidential. All documents will be identified only by a code number and kept in a locked filing cabinet in the principal investigators research office. All backup computer files will be kept in a locked filling cabinet, and any data files that reside on the data analysis computer in the Human Communication Technologies Laboratories (HCT) (room 509, ICIS, UBC), will be number coded.
8.2. Access to Data Who will have access to the data? (For example, co-investigators or students). How will all of those who have access to the data be made aware of his or her responsibilities concerning privacy and confidentiality issues?	Only Dr. Fels and the study co-investigators will have password access to the data. Subjects will have access to published results of the research. Those with access will be briefed verbally and/or by electronic memo by one or more co-investigator(s) on privacy and confidentiality issues with data.
8.3. Protection of Personal Information Describe how the identity of research subjects will be protected both during and after the research study, including how subjects will be identified on data collection forms	Participant names will not be used in written research. Subjects will be identified with code numbers.
8.4. Transfer of Data Will any data that identify individuals be transfered (available) to persons or agencies outside of the University?	no
If YES, describe in detail what identifiable information is released, to whom, how the data will be transferred, how and where it will be stored and what safeguards will be used to protect the identity of subjects and the privacy of their data. Attach the data transfer agreement if applicable.	
8.5. Retention and Destruction of data UBC policy requires that data be kept for at least 5 years within a UBC facility. If you intend to destroy the data at the end of the storage period describe how this will be done to ensure confidentiality (e.g. tapes should be demagnetized, paper copies shredded). UBC has no explicit requirement for shredding of data at the end of this period; however, destruction of the data is the best way of ensuring that confidentiality will not be breached. Please note that the responsibility for the security of the data rests with the Principal Investigator.	Data will be stored by Fels in a locked cabinet within a private office for at least the required 5 years. Data will be shredded or erased if it is no longer to be kept after that time.
8.6. Future use of data Are there any plans for future use of either data or audio/video recordings? Provide details, including who will have access and for what purposes, below.	There are no future plans at this time.

8.7. Feedback to subjects Are there any plans for feedback on the findings or results of the research to the subject? Provide details below.	publications will be available publicly in conference and journal proceedings as well as available for download on the internet
9. Documentation - Human Ethics Application [View Form]	
9.1.A. Protocol Examples of types of protocols are listed on the right. Click Add to enter the required information and attach the documents.	Name Version Date Password (if applicable)
9.1.B. Health Canada regulatory approval (receipt will be acknowledged)	Name Version Date Password (if applicable)
9.1.C. FDA IND or IDE letters (receipt will be acknowledged)	Name Version Date Password (if applicable)
9.2. Consent Forms Examples of types of consent forms are listed on the right. Click Add to enter the required information and attach the forms.	Name Version Date Password (if applicable) Consent Form TAV July 21, 2010 [View]
9.3. Assent Forms Examples of types of assent forms are listed on the right. Click Add to enter the required information and attach the forms.	Name Version Date Password (if applicable)
9.4. Investigator Brochures/Product Monographs (Clinical applications only) Please click Add to enter the required information and attach the documents.	Name Version Date Password (if applicable)
9.5. Advertisement to recruit subjects Examples are listed on the right. Click Add to enter the required information and attach the documents.	Name Version Date Password (if applicable)
9.6. Questionnaire, questionnaire cover letter, tests, interview scripts, etc. Please click Add to enter the required information and attach the documents.	Name Version Date Password (if applicable)
9.7. Letter of initial contact Please click Add to enter the required information and attach the forms.	Name Version Date Password (if applicable)
9.8. A. Other documents: Examples of other types of documents are listed on the right. Click Add to enter the required information and attach the documents.	Name Version Date Password (if applicable)
B. If a Web site is part of this study, enter the URL below. Since URL's may change over time or become non-existent, you must also attach a copy of the documentation contained on the web site to one of the sections above or provide an explanation.	
10. Fee for Service - Human Ethics Application for Behavioural Study [View Form]	
Mechanism for Submitting Fee. Please indicate which of the following method of payment will be used for this application:	
Contact information regarding where to send the invoice.	

This user study invitation card is given to participants after the experimenter asked them for their participation. The experimenter enters the date of participation in the empty field and hands it over to the participant as a reminder.

	August 2010													
	Thu 19										Fri 20			
	10:00 AM	11:00 AM	12:00 PM	1:00 PM	2:00 PM	3:00 PM	5:00 PM	6:00 PM	7:00 PM	8:00 PM	8:00 AM	9:00 AM	10:00 AM	11:00 AM
Cam	OK													
Johnty		OK												
Gordon Chang													OK	
Andrew			OK											
Krusty the clown				OK										
ivy						OK								
Laurence					OK									
Salima												OK		
Daesik														
Nima														OK
martin														
vincent														OK

An online schedule is used to organize the user study. Participants are invited to subscribe to the schedule and afterwards are contacted by the experimenter for further information.

MYVIEW

USER STUDY

AUGUST 19 - 20
(THURSDAY - FRIDAY)

PLEASE DO NOT DISTURB.

MY VIEW

This paper sheet is used on the door of the user study room to inform by-passing people about the on-going user study and to help participants to find the location of the user study.

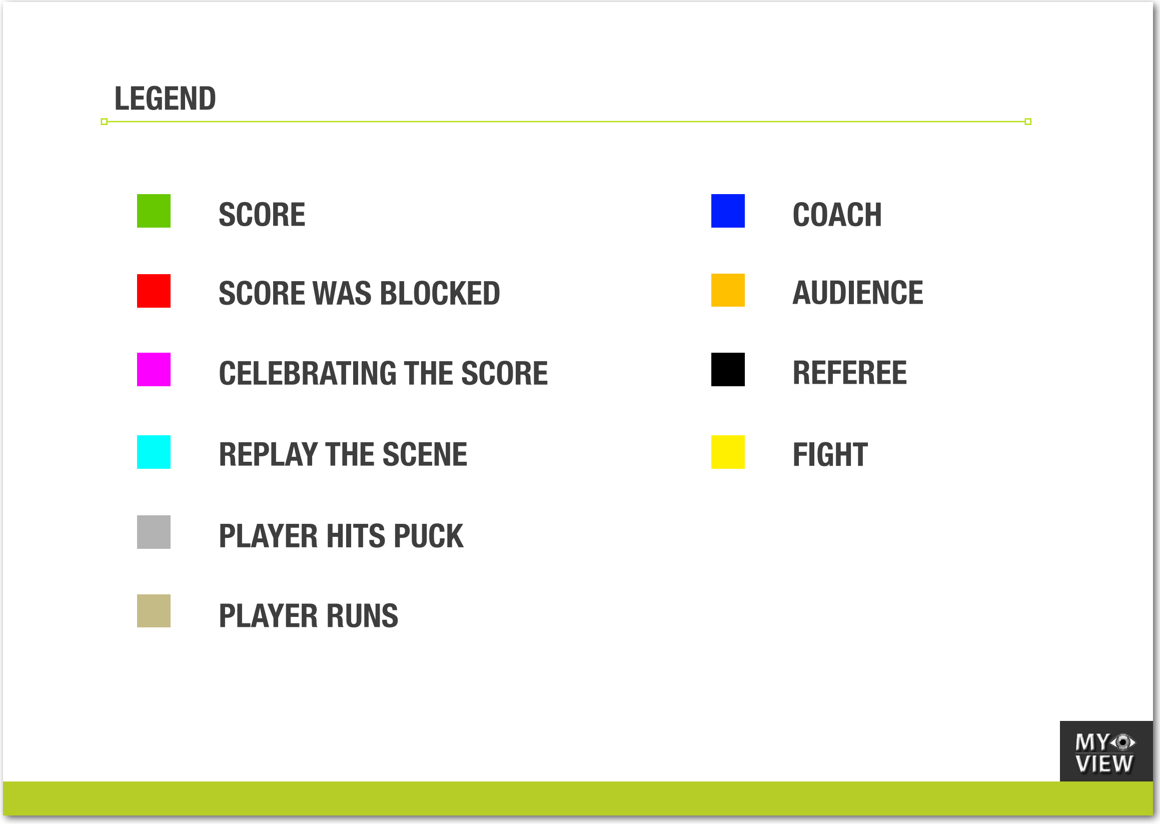
USER INTERFACE

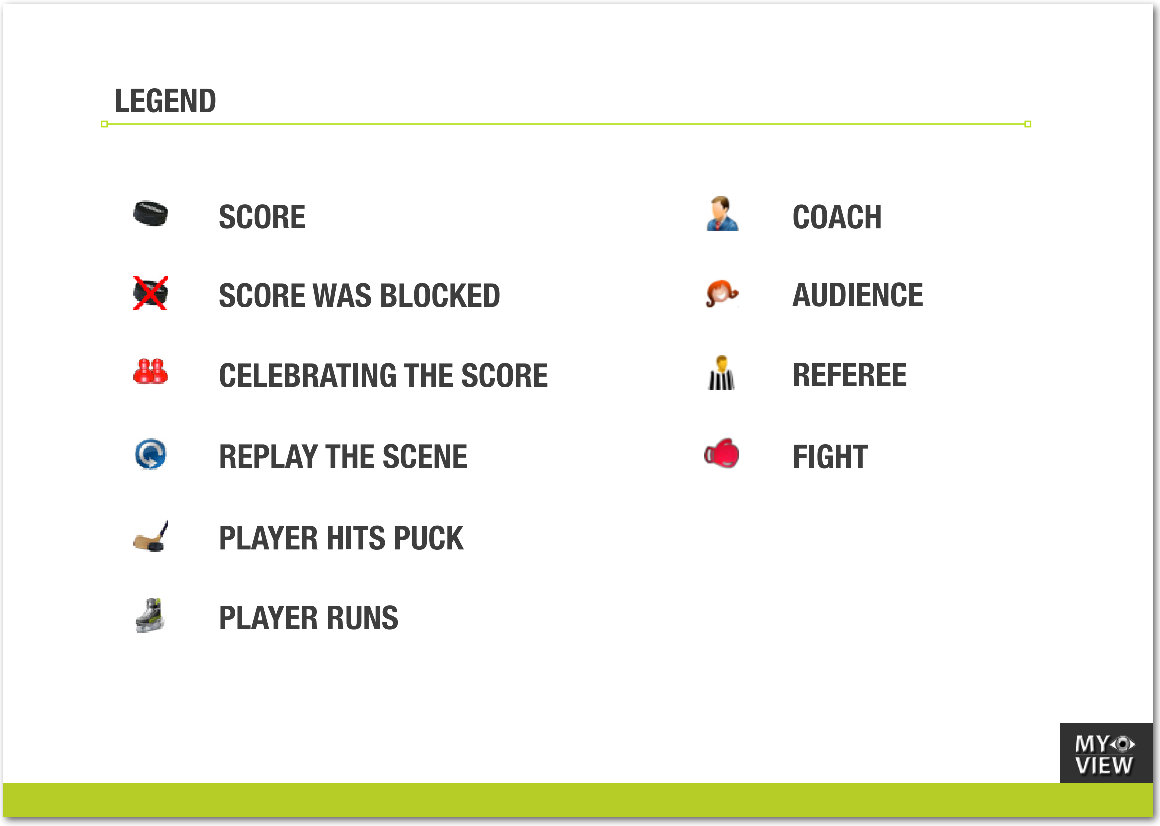
This paper sheet is used during the user study for the colour stripe participant group to explain the user interface. After the particpant confirmed that he understands the user interface, the experimenter starts the training session on the device.

USER INTERFACE

MY VIEW

This paper sheet is used during the user study for the icon participant group to explain the user interface. After the particpant confirmed that he understands the user interface, the experimenter starts the training session on the device.





These legends explain the meanings of the different visual cues used during the user study. Legend 1 (top) is used for the colour stripe participant group, whereas legend 2 (bottom) is used for the icon participant group.

User Study

Questionaire

1. How often do you search for scenes:

- in a familar video source (e.g. a favorite tv series, an old family video)

  very often *often* *occasionally* *never*

- in an unfamiliar video source (e.g. a video tutorial, a sportsgame that you missed)

  very often *often* *occasionally* *never*

2. Please rate:

- the colour stripes helped me to see where scenes of interest are located

  1 (I agree strongly) *2* *3* *4* *5 (I don't agree at all)*

- the colour stripes helped me to reason about the content of the underlying video scene

  1 (I agree strongly) *2* *3* *4* *5 (I don't agree at all)*

3. Can you think of other visual cues than colour stripes to represent the content of a video scene?

4. Which advantages do you see in using colour stripes on the video timeline?

5. Where do you see room for enhancements concerning the use of colour stripes on the video timeline?

MY VIEW

This questionnaire is used during the user study for the colour stripe participant group after they finished the task section. The questionnaire is provided in a printed form to give the participant the opportunity to draw sketches of user interface suggestions beside the regular verbal statements.

USER STUDY

QUESTIONAIRE

1. How often do you search for scenes:

- in a familar video source (e.g. a favorite tv series, an old family video)

  very often   *often*   *occasionally*   *never*

- in an unfamiliar video source (e.g. a video tutorial, a sportsgame that you missed)

  very often   *often*   *occasionally*   *never*

2. Please rate:

- the icons helped me to see where scenes of interest are located

  1 (I agree strongly)   *2*   *3*   *4*   *5 (I don't agree at all)*

- the icons helped me to reason about the content of the underlying video scene

  1 (I agree strongly)   *2*   *3*   *4*   *5 (I don't agree at all)*

3. Can you think of other visual cues than icons to represent the content of a video scene?

4. Which advantages do you see in using icons on the video timeline?

5. Where do you see room for enhancements concerning the use of icons on the video timeline?

MY VIEW

This questionnaire is used during the user study for the icon participant group after they finished the task section. The questionnaire is provided in a printed form to give the participant the opportunity to draw sketches of user interface suggestions beside the regular verbal statements.

	Participant 1	Participant 2	Participant 3	Participant 4	Participant 5	Participant 6
DisplayMode	icons	colour stripes	icons	colour stripes	icons	colour stripes
Gender	male	male	male	male	female	female
Age	31	25	30	29	31	24
Used iPod before	yes	yes	yes	yes	yes	yes

	Participant 7	Participant 8	Participant 9	Participant 10	Participant 11	Participant 12
DisplayMode	icons	colour stripes	icons	colour stripes	icons	colour stripes
Gender	female	male	male	male	female	male
Age	23	25	25	38	53	29
Used iPod before	yes	yes	yes	yes	no	yes

DisplayMode	50%	colour stripes/icons
Gender	66,67% male, 33,33% female	
Age	30,25	years in average
Used iPod before	91,67%	have used iPod touch before

Data collected about the participants: the display mode to which they are randomly assigned, the gender of the participant, the age of the participant and if they have used an iPod touch before.

Finding Time Icons	Participant 1	Participant 3	Participant 5	Participant 7	Participant 9	Participant 11		Icon Size	Lambda
Task 26	8268	6228	5030		3779		5826,25	25	1
Task 27	5817	4967	4317	4412	4290		4760,6	25	1
Task 28	5520	6084	8130	6800	3512	6807	6142,167	25	1
Task 29	4861	3736	7540	3143	4456		4747,2	22	1
Task 30	7275	5995	6357	5984	5166		6155,4	25	1
							5526,323	24,4	
Task 6	5184	4486	5456			3756	4720,5	22	2
Task 7	6981	6571	4870	4959	7056	8058	6415,833	22	2
Task 8	6227	8300	9630	4007	5691	6300	6692,5	25	2
Task 9	5493	4769	5996	8234	4922	5288	5783,667	25	2
Task 10	4554	4148	3992	4684	5238	5761	4729,5	25	2
							5668,4	23,8	
Task 11	10607	4468	6261	7155	4481		6594,4	10	3
Task 12	7546	7654	4847	12760	5619		7685,2	25	3
Task 13	7374	5838	9430	6279	7428	8701	7508,333	22	3
Task 14	11974	7340	10921	6805	14269	6940	9708,167	19	3
Task 15	7841	6519	7194	7624	7220	11524	7987	10	3
							7896,62	17,2	
Task 16	7964	6170	7224	10190			7887	7	4
Task 17	14502	8972	8057	9159	7142	9031	9477,167	25	4
Task 18	7936	5072	8566	4727	5844		6429	19	4
Task 19	6076	4725	5140	4534		7173	5529,6	16	4
Task 20	9787		9472	12583		10229	10517,75	7	4
							7968,103	14,8	
Task 21	11494	5573	5243	15000	7907	15000	10036,17	13	5
Task 22	7969	7184	8098	7998	5436	10982	7944,5	13	5
Task 23	6056	5207	6820	6228		7558	6373,8	10	5
Task 24	8222	5782	6910	6484	5880	15000	8046,333	10	5
Task 25	8029	6002	6342	7669	7364		7081,2	10	5
							7896,4	11,2	
	Application crashed								
	Participants made errors								
	Values above threshold								
	Finding time smaller than other display mode								
	Average								

Overview of measured experiment data in the icon group: finding time for all tasks, errors participants made, time values over the threshold of 15s, the icon size of each task and the corresponding lambda of the Poisson distribution.

Finding Time Colours	Participant 2	Participant 4	Participant 6	Participant 8	Participant 10	Participant 12		Colour Size	Lambda
Task 26	7093	5051	7458	3775	6606	3794	5629,5	7	1
Task 27	4965	4895	5559	5891	4484	5578	5228,67	7	1
Task 28	6375	6874	7443	4684	4532	7507	6235,83	7	1
Task 29	4700	6095	5618	4982	5441	6489	5554,17	7	1
Task 30	5121	7031	6018		4746	5964	5776	7	1
							5684,83	7	
Task 6	4285	4664	7521		5148	4859	5295,4	7	2
Task 7	4573	4072	11018	4373	4458	6080	5762,33	7	2
Task 8	9284	7912	9643	7546	8753	4638	7962,67	7	2
Task 9	6049	3674	5750	5109	6689	7623	5815,67	7	2
Task 10	4701	6115	8915	4047	7348	8964	6681,67	7	2
							6303,55	7	
Task 11		3659	5932			7585	5725,33	7	3
Task 12	5987		6527	6600	10097	5761	6994,4	7	3
Task 13	6427	5205	7308	4721	7624	7478	6460,5	7	3
Task 14	7809	6725	15000	6614	6366		8502,8	7	3
Task 15	8387	4878	8765	7127	9615	8887	7943,17	7	3
							7125,24	7	
Task 16	7925	4590	15000		6960	7713	8437,6	7	4
Task 17	4601		15000	6582	11282	7524	8997,8	7	4
Task 18	6580	10490	10254	5294	6326	5710	7442,33	7	4
Task 19	4490	4177	8232	4511	5686	6611	5617,83	7	4
Task 20	5649		6238		5727	4449	5515,75	7	4
							7202,26	7	
Task 21	7995	11922	8239	5672	6055	4152	7339,17	7	5
Task 22	5868		15000	5498	4215	5463	7208,8	7	5
Task 23	5212	5273	6820	4478	6272	3928	5330,5	7	5
Task 24	5166	4473	9467	5680	5915	5638	6056,5	7	5
Task 25	5339	5361	7014	3616	7755	4717	5633,67	7	5
							6313,73	7	
	Application crashed								
	Participants made errors								
	Values above threshold								
	Finding time smaller than other display mode								
	Average								

Overview of measured experiment data in the colour group: finding time for all tasks, errors participants made, time values over the threshold of 15s, the colour stripe size of each task and the corresponding lambda of the Poisson distribution.

Bibliography (Paper)

[1] M. Barbieri, G. Mekenkamp, M. Ceccarelly, J. Nesvadba, The Color Browser: A Content Driven Linear Video Browsing Tool, Multimedia and Expo, ICME 2001, IEEE International Conference on, pp. 627-630, 2001

[2] L. Chen, G.C. Chen, C.Z. Xu, J. March, S. Benford, EmoPlayer: A media player for video clips with affective annotations, In: Interacting with Computers, vol. 20, no. 1, pp. 17-28, 2008

[3] S. Chang, W. Ma, A. Smeulders, Recent Advances and Challenges of Semantic Image/Video Search, In: Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing, 2007

[4] M. Costa, N. Correia, N. Guimaraes, Annotations as Multiple Perspectives of Video Content, In: Proceedings of the Tenth ACM International Conference on Multimedia, pp. 283-286, 2002

[5] M. Davis, Media Streams: An Iconic Visual Language for Video Representation, In: Readings in Human-Computer Interaction: Toward the Year 2000, pp. 854-866, 1995

[6] A. Haubold, P. Dutta, J.R. Kender, Evaluation of Video Browser Features and User Interaction with VAST MM, In: Proceedings of the 16th ACM international conference on Multimedia, pp. 449-458, 2008

[7] A. Janin, L. Gottlieb, G. Friedland, Joke-o-mat HD: Browsing Sitcoms with Human Derived Transcripts, ACM Multimedia 2010, 2010

[8] W.H. Hsu, L.S. Kennedy, S.F. Chang, Video search reranking through random walk over document-level context graph, In: Proceedings of the 15th international conference on Multimedia, 2007

[9] D. Kitayama, N. Oda, K. Sumiya, Organizing User Comments in a Social Video Sharing System by Temporal Duration and Pointing Region, In: Proceedings of the 2008 International Work shop on Information-Explosion and Next Generation Search, pp. 55-58, 2008

[10] J.K.W. Leung, C.H. Li, T.K. Ip, Commentary-based Video Categorization and Concept Discovery, In: Proceedings of the 2nd ACM workshop on Social web search and mining, pp. 49-56, 2009

[11] N. Moraveji, Improving Video Browsing with an Eye-Tracking Evaluation of Feature-Based Color Bars, Digital Libraries, In: Proceedings of the 2004 Joint ACM/IEEE Conference on, pp. 49-50, 2004

[12] S. Nakamua, M. Shimizu, K. Tanaka, Can Social Annotation Support Users in Evaluating the Trustworthiness of Video Clips?, In: Proceeding of the 2nd ACM workshop on Information credibility on the web, pp. 59-62, 2008

[13] K. Schoeffmann, M. Lux, M. Taschwer, L. Boeszoermenyi, Visualization of Video Motion in Context of Video Browsing, In: Proceedings of the IEEE International Conference on Multimedia and Expo, 2009

[14] K. Schoeffmann, M. Taschwer, L. Boeszoermenyi, The Video Explorer - A Tool for Navigation and Searching within a Single Video based on Fast Content Analysis, In: Proceedings of the first annual ACM SIGMM conference on Multimedia systems, pp. 247-258, 2010

[15] Viddler.com, Accessed August 2010, www.viddler.com

[16] S. Wakamiya, D. Kitayama, K. Sumiya, Scene Extraction for Video Clips based on the Relation of Text, Pointing Region and Temporal Duration of User Comments, Database and Expert Systems Application, 20th International Workshop on, pp. 289-294, IEEE, 2009

[17] L. Wiyartanti, H. Yo-Sub, K. Laehyun, A Ranking Algorithm for User-generated Video Contents based on Social Activities, Digital Information Management, Third International Conference on, pp. 260-265, 2008

[18] R. Wu, J. Sun, J. Chen, H. Wu, Video Semantics discovery from the video captions and comments, Electronic Imaging and Multimedia Technology V, 2007

[19] D. Yamamoto, T. Masuda, S. Ohira, and K. Nagao, Collaborative Video Scene Annotation Based on Tag Cloud, Lecture Notes in Computer Science, Advances in Multimedia Information Processing, Springer Berlin, 2008

[20] W.G. Yee, A. Yates, O. Frieder, and A. Moehrle, eduKEN: A Tool for Fine-grained Video Comment Collection and Analysis, In: Proceedings of the 19th International Conference on World Wide Web, pp. 1325-1328, 2010

[21] W.G. Yee, A. Yates, S. Liu, O. Frieder, Are Web User Comments Useful for Search, 7th Workshop on Large-Scale Distributed Systems for Information Retrieval, 2009

Bibliography (Figures)

[22] Adobe After Effects, Accessed August 27, 2010,
http://img193.imageshack.us/i/firez.jpg/

[23] Adobe Premiere, Accessed August 22, 2010,
http://www.flickr.com/photos/amit-agarwal/85590368/

[24] Apple Icons, Accessed August 5, 2010,
http://developer.apple.com/iphone/library/documentation/userexperience/conceptual/mobilehig/IconsImages/IconsImages.html

[25] CoverFlow, Accessed August 31, 2010,
http://images.anandtech.com/reviews/gadgets/2008roundup/CoverFlow.jpg

[26] CS (Computer Science in Media), Accessed August 8, 2010,
http://www.medieninformatik.de/templates/medieninformatik/images/logos/logo_mi.jpg

[27] Final Cut Pro, Accessed August 31, 2010,
http://www.bethlehemmedia.net/eng/components/com_virtuemart/shop_image/product/62db9991d56f66d0a64e0f0bc52f32c0.jpg

[28] Google Maps, Accessed August 31, 2010,
http://www.googlemaps.com

[29] HCT (Human Communication Technology Laboratoy), Accessed August 8, 2010,
http://www.ece.ubc.ca/~hct/images/hct_logo1.jpg

[30] HH (Hochschule Harz), Accessed August 8, 2010,
http://upload.wikimedia.org/wikipedia/de/thumb/7/72/Logo_HochschuleHarz.svg/800px-Logo_HochschuleHarz.svg.png

[31] Instruments, Screenshot of Instruments, Made on September 4, 2010,
http://developer.apple.com/technologies/tools/

[32] iPod Player, Accessed August 10, 2010,
http://www.ploem.be/blog/wp-content/uploads/020.PNG

[33] Mac Dock, Screenshot of the Dock on Mac OS X 10.5.8, Made on September 4, 2010,
http://www.apple.com/de/macosx/

[34] MyView, all figures referenced with this source are part of the internal sources of the MyView project and can be accessed by contacting Gregor Miller (gregor@ece.ubc.ca)

[35] Referee 1, Accessed September 4, 2010, http://www.table-soccer.org/news/images/icon_referee06.gif

[36] Referee 2, Accessed September 4, 2010, http://www.iconarchive.com/icons/afterglow/forum-faces-2/128/The-Referee-icon.png

[37] Referee 3, Accessed September 4, 2010, http://www.marquette.edu/bea/write/graphics/referee-2.jpg.jpeg

[38] Referee 4, Accessed September 4, 2010, http://www.bradfitzpatrick.com/store/images/products/pc010-cartoon-referee.jpg

[39] Safari, Screenshot of Safari 5.0.1 Bookmarks, Made on September 4, 2010, http://www.apple.com/de/safari/

[40] UBC, Accessed August 8, 2010, http://www.cs.ubc.ca/events/acmicpc2009/ACM_info_files/UBC.jpg

[41] Video Player, Accessed August 24, 2010, http://www.iphonespies.com/wp-content/uploads/2009/07/iphone-3gs-video-player.jpg

[42] YouTube, Accessed August 24, 2010, http://www.youtube.com

Bibliography (Other)

[43] Adobe After Effects, Accessed September 4, 2010,
http://www.adobe.com/de/products/aftereffects/

[44] Adobe Premiere, Accessed September 4, 2010,
http://www.adobe.com/de/products/premiere/

[45] Avid Xpress Pro, Accessed September 4, 2010,
http://www.avid.de/de/products/xpress-pro/

[46] Computer Vision, Accessed September 4, 2010,
http://dictionary.die.net/computer%20vision

[47] Context Aware Video Data, Accessed September 4, 2010,
Schmidt, Albrecht (2002). „Ubiquitous Computing - Computing in Context".
PhD dissertation, Lancaster University.. http://www.comp.lancs.ac.uk/~albrecht/phd/

[48] Final Cut Pro, Accessed September 4, 2010,
http://www.apple.com/de/finalcutstudio/finalcutpro/

[49] NTSC, Accessed September 4, 2010,
http://www.yourdictionary.com/computer/ntsc

[50] PAL, Accessed September 4, 2010,
http://www.yourdictionary.com/computer/pal

[51] Poisson Code, Accessed July 8, 2010,
http://www.mathtools.net/files/net/poisson.zip

[52] TAG, Accessed September 4, 2010,
http://en.wikipedia.org/wiki/Tag_%28metadata%29

Summarization of Thesis Content in German Language

Konzeptentwicklung einer Benutzeroberfläche zum Ansehen zeitbasierter Videokommentare

Video ist ein komplexer Informationsraum, der erweiterte Navigationshilfen für effektives Explorieren benötigt. Die ansteigende Zahl zeitbasierter Videokommentare eröffnet neue Möglichkeiten für eine Videonavigation, die sich nach den Interessen des Benutzers richtet. In der vorliegenden Bachelorarbeit präsentieren wir einen neuartigen Video-Player genannt TAV (Temporal Annotation Viewing), der dem Benutzer eine visuelle Übersicht über zeitbasierte Videokommentare bereitstellt. Benutzer können anhand der Übersicht den allgemeinen Inhalt eines Videos erfassen, interessante Videoszenen lokalisieren und bestimmen, welche Videokommentare während des Ansehens angezeigt werden.

Die Übersicht zeitbasierter Videokommentare in TAV besteht aus einer Zeitleiste mit Piktogrammen, welche die Position und den Inhalt der Videokommentare darstellen. Während die visuelle Gestaltung von Piktogrammen mehr semantische Informationen für die Navigation bereitstellt als andere visuelle Repräsentationsformen wie Farbstreifen, die abstrakt und nur schwer zu unterscheiden sind, so ist es als Nachteil anzusehen, dass Piktogramme mehr Platz auf der Zeitleiste benötigen. Diese Eigenschaft wird schnell zum Problem, wenn die Anzahl von zeitbasierten Videokommentaren im Video zunimmt und somit auch die Dichte der Piktogramme auf der Zeitleiste ansteigt. Zusätzlich spielt auch der zunehmende Trend Videos auf mobilen Endgeräten mit kleinen Bildschirmen wie z.B. Handys anzusehen eine entscheidende Rolle. In unserer Arbeit präsentieren wir eine neuartige Lösung genannt SCADE (Scaling with Deformation), welche die visuelle Erscheinung von Piktogrammen an den verfügbaren Bildschirmplatz anpasst und dabei die semantische Information vermittelt durch die visuelle Gestaltung der Piktogramme beibehält.

In unserer Benutzerstudie untersuchen wir den Konflikt zwischen detaillierter semantischer Information und begrenztem Bildschirmplatz. Dabei vergleichen wir die Zeit, die ein Benutzer benötigt um eine bestimmte Videoszene zu finden, wenn Piktogramme oder Farbstreifen auf der Zeitleiste verwendet werden. Unsere Ergebnisse zeigen, dass die visuelle Gestaltung von Piktogrammen dem Benutzer mehr Information bereitstellt um auf den Inhalt einer Videoszene zu schlussfolgern und dabei eine ähnliche Suchzeit aufweist wie Farbstreifen. Steigt die Dichte der zeitbasierten Videokommentare auf der Zeitleiste jedoch an, so sind Piktogramme auch bei der Verwendung von SCADE langsamer als Farbstreifen. Zukünftige Forschung wird die Weiterentwicklung von TAV und SCADE umfassen, um die Hinweise der Teilnehmer unserer Benutzerstudie zu integrieren. Anschließend werden wir weitere umfassende Benutzerstudien durchführen, um die Verbesserungen zu evaluieren.

Declaration of Authorship

I certify that the work presented here is, to the best of my knowledge and belief, original and the result of my own investigations, except as acknowledged, and has not been submitted, either in part or whole, for a degree at this or any other University.

Location, Date

Signature