UTRECHT UNIVERSITY

Faculty of Science

HIERARCHICAL VIDEO BROWSING INTERFACES ON TOUCH-SCREEN MOBILE DEVICES

Author:
Cojocaru Cristian

Supervisor:
Dr. Wolfgang Hürst

Student Number:
3760308

Thesis Number:
ICA-3760308

ABSTRACT

The most widely spread model used in hierarchical video browsing interfaces is the storyboard model, where frames extracted from movies are shown on the screen in an orderly fashion based on the timeline of the film. This research paper evaluates different characteristics of this type of interfaces on mobile devices with touchscreens via a series of detailed user studies. Based on the results and observations from the studies, we propose and implement a new interface design. The usability of this new interface is proven in a final user study.

The results of our experiments show that users prefer simpler methods of interactions in browsing interfaces, even though they are not more efficient or flexible than interfaces using more complex gestures. It was also observed that browsing interfaces focused on a single level of the granularity hierarchy, tends to do better in terms of the number of mistakes and worse in search time than interfaces that show multiple levels on the screen at the same time. Even though the results of the user studies were not all conclusive, they still offer a good basis for the implementation of a combined interface with statistically better results than a simple multi-level browser.

CONTENTS

1. Introduction. 4
1.1 Background 4
1.2 Motivation 5

2. Related Work 6
2.1 Visualization and storyboard design 6
2.2 Interaction methods 7
2.3 3D video browsing interfaces 7

3. Study 1: Mobile interaction gestures for browsing in hierarchical video interfaces 9
3.1 Design Space 9
3.2 Experiment Setup 12
3.3 Results 13
3.3.1 Recorded Data 13
3.3.2 Questionnaire 19
3.4 Conclusion 20

4. Study 2: Hierarchical video browsing models 21
4.1 Design Space 21
4.2 Experiment Setup 23
4.3 Results 25
4.3.1 Recorded Data 25
4.3.2 Questionnaire 30
4.4 Conclusion 31

5. Study 3: Combining the benefits of the tree and grid interfaces 32
5.1 Design Space 32
5.2 Experiment Setup 34
5.3 Results 35
5.3.1 Recorded Data 35
5.3.2 Questionnaire 40
5.4 Conclusion 42

6. Conclusion and Future Work 43
6.1 Conclusion and Summary 43
6.2 Future Work 44

7. Acknowledgments 45

8. Bibliography 46

1. INTRODUCTION

1.1 BACKGROUND

Smartphones are cellular devices that combine the functionality of a phone and a computer in a single product that can be controlled using touch gestures. The first smartphone was introduced to the mass consumer market in 1993 by IBM, but such devices didn't gain much recognition until ten years later (2002), when the BlackBerry 5810 was released. With the introduction of the iPhone in 2007, smart cellular devices registered a boom in popularity, leading to an impressive number of over one billion smartphone users in the present day.

As smartphones are mobile computers, a lot of the problems encountered on regular PCs can be translated to such devices, but must be treated differently due to their mobile nature. Data searching represents one of the biggest research topics in the field of information sciences. Video browsing is an example of an information searching problem, in which the user is actively analyzing video content in order to find a given video document (known item search) or documents related to certain subjects of interest (subject search). The implementation of a video browser presents a set of unique challenges to both interface designers and video content analysis systems.

Figure 1: Storyboard representation of a movie

The most straight-forward solution to this problem is to show video clips in a storyboard - an ordered grid structure that presents time-ordered still images extracted from the video at fixed intervals. Browsing through this type of interfaces can be time consuming, because the user is forced to go over some of the video content that he may skip, in order to examine a certain part of the video. An improvement to the storyboard approach is to display the video data as a tree-like structure – the frames are extracted from the movies at variable time intervals, intervals dependent on the level of the tree currently being examined. This type of organization allows the user to bypass certain segments of the video that are of no interest to him, and to examine in greater detail only the relevant parts.

1.2 MOTIVATION

Smartphones are bounded by a set of restraints imposed by the portable nature of the devices – small screen size, gesture and touch based interaction, etc. Given these restrictions, the solutions that are viable on a desktop computer are not as efficient on mobile phones. The same is true for the hierarchical tree representation of video browsers, which have the deficit of the small size of the screen that restricts the user from viewing many levels of the tree at once. One solution for this problem is to use the capabilities of the device to convey information in a 3D scene instead of just a simple 2D one.

Smartphones are equipped with capacitive multi-touch screens that allow the user to interact with applications using a multitude of gestures deemed more natural than mouse and keyboard interaction, such as pinch, drag or flick. Besides the touch screen interactions, smartphones are also equipped with a variety of sensors like accelerometers and gyroscopes that give information about the spatial positioning of the device, thus allowing the usage of a large number of interaction methods that are unavailable on a stationary desktop computer. One example of such a method of interaction is the so-called "shoebox virtual reality". The shoebox VR is a model of interaction that performs a perspective correction on the 3D space based on the orientation of the device. This change in perspective is used to emulate 3D space on the 2D screen on the device in a more natural and realistic manner.

The goal of the thesis is to present and discuss possible parameters that describe the functionality of hierarchical video browsers, in a series of user studies. We begin by analyzing different methods of interaction within this type of interfaces. Firs we present the most common gesture used – the tap – by acknowledging its strengths, weaknesses and the ways it can be improved. Then we will take a look at multi-touch gestures and their implementations that could solve the specific problems of the tap.

In the second user study, we compare two different approaches to the implementation of a hierarchical vide browsing interface:

- the in-depth view (grid) where the users are presented a large number of extracted frames from a single level of granularity on the whole screen
- an overview (tree) of multiple levels of granularity, but using a smaller number of frames for each level

The effectiveness of each approach is tested and observation about their strengths and weaknesses are made following the user studies.

Using the results gathered during the previous user studies, we implement an interface that combines the benefits of both the grid and tree interfaces using shoebox-like virtual visualization and compare it to a state of the art 2D hierarchical video browsing interface.

2. RELATED WORK

2.1 VISUALIZATION AND STORYBOARD DESIGN

Most mobile devices today come equipped with cameras and permanent access to high speed internet. This means that the available video data to such devices is ever-increasing. This leads to a lot of research in the area with a number of video browsing tools having emerged during the last years. The most common method of presenting video content is by extracting pictures from clips and showing them in a structured manner to the user. One example of video summarization is done using storyboards as presented in the VISTO [1] interface.

A particularly interesting video browsing technique is to show the extracted frames on different levels of a hierarchy based on the time between two consecutive key-frames. This approach was first implemented in 1995, in the work of Zhang et al[2] and Guillemot et al [3] it was shown that a hierarchical interface performed two times better than a classical video player (Real Player). This method of multimedia data presentation is the focus of this thesis, where we will investigate different characteristics of such type of interfaces.

Hürst et al [4] applied the idea of different granularity levels in browsing using navigation bars in the Zoom slider interface. Instead of using just one bar for navigation, the Zoom Slider used several with different navigation speeds (granularities) based on the vertical position at which the navigation gesture occurs. The closer the navigation bar is to the top of the screen, the higher the granularity factor. In our third study, we took this idea and applied it to a keyframe–based interface with different levels of granularity bound together by the movement in the timeline.

An alternative solution for visualizing hierarchical trees was proposed by Jansen et al [6]. Their representation shows frames placed next to their siblings and directly under the parent frame, and shown at a smaller size than the frames in the levels above. As the user goes through the levels of the tree from the root to the edges, the size of individual frames decreases, but the detail (granularity) level of the overall scene increases.

Hürst and Darzentas [7] proposed a hierarchical storyboard browser (HiStory) for mobile devices that takes advantage of the perception and cognition of human visualization. The implementation allows the user to change the granularity of the grid visualization by selecting an anchor image, while maintaining the position of the anchor on both levels. This interaction method was compared to other browsing techniques using storyboards such as page and continuous scrolling.

The AAU video browser [8], [9] presents the data hierarchically in a tree-like structure or allows the parallel exploration of different granularity levels at the same time. The sequential navigation allows fast switching between levels and search paths within the tree, which in turn favors faster browsing times.

In the second study, we decided to test the effectiveness of the two different approaches to video browsing presented in the HiStory [7] browser and the AAU video browser [8], [9]. This study is aimed to analyze the two approaches to hierarchical video browsing interface design and present their advantages and disadvantages based on the data gathered from the users.

2.2 INTERACTION METHODS

Multi-touch interaction methods became the standard for today's mobile devices. This grants an increased importance of the interaction gestures used in the design of mobile apps. Touch interaction systems like Android Touch [12] or Windows Touch [13] offer developers the tools to retrieve information from interactions with the tactile screen, but does not make any attempt to standardize these interactions. One particular question we decided to investigate is whether we can use complex gestures (pinch) in order to interact with

In [14] Lao et al proposed a generalization for touch interactions using an established model. Each touch gesture is classified and defined on three levels: action, motivation and computing. At the action level the available touch types are defined, mapped to the actions carried out by the users at the motivation level, while the technical details of their implementation are discussed at the computing level. One of their general observations is that users are used to gestures that are symmetrical and fewer people are able to perform more complicated gestures.

Furthermore, Kruger et al [15] proposed a formalization of complex touch gestures by defining their features using specialized functions: a pose function describes the blob that is being tracked (one finger, two fingers, one hand…), atomic gestures represent the movement of the tracked blob (line, circle, hold), composition operators describes the temporal progression of multiple gestures (in parallel, successive), while the area constraints define the movements of atomic gestures in relation with each other (converge, spread, intersect).

In one of our studies, we present the most widely used method of interaction used for navigation in hierarchical interfaces: the click gesture. After we analyze its problems, we propose a solution using multi-touch that would tackle the issues of the click.

2.3 3D VIDEO BROWSING INTERFACES

In the third user study, we designed an interface that would combine the benefits of a grid interface showing one level of granularity at one time with the advantages of an interface that presents multiple levels on the screen at once.

In order to show that much data on screen at the same time, the proposed interface would have to take advantage of the 3D space. Numerous research papers study the presentation of storyboards in three dimensional space - Plant and G. Schaefer [18] have shown that picture data presented on the surface of a 3D sphere is preferred by users over traditional two dimensional storyboards.

Manske[10] introduced a method to present hierarchical set of key-frames in a conic tree-like 3D visualization. The hierarchical tree is computed based on the information provided by the histogram on a set of parameters such as color, amount of motion or number of objects that are in a scene inside each frame. Schoeffmann et al [11] also combined the advantages of hierarchical video browsing with 3D projection to provide an intuitive way to navigate within a video using a 3D carousel.

Klaus Schoeffmann et all [19] performed a similar study by comparing diverse 3D shaped interfaces with classical storyboards, arguing that interfaces presented in 3D space can show a large number

of images on the screen. This assumption was confirmed by David Ahlstrom [20], as his results shown that 3D interfaces got a 12% improvement in trial completion time over their 2D counterparts.

A novel method of interaction with 3D space on smartphones takes advantage of the gravitational sensors (gyroscope and accelerometers) of the device to change the perspective over the virtual space as the user tilts his phone. This method is known as the "shoebox virtual reality" and its characteristics were studied by Martin de Jong [21].

Mathijs T. Lagerberg [22] performed a series of user studies where he compared a classical interface that uses swiping as their method of interaction, with four interfaces enhanced by the shoebox effect: plane, stacks, hollow cylinder and regular cylinder. The results have shown that the 2D interface scored significantly worse than the shoebox-enriched interfaces.

Steven Wijden [23] also tested the usefulness of the shoebox method inside different 3D shapes - sphere, hollow cylinder, hollow box, tunnel and pipe – by measuring how much time users would take to find certain pictures inside each interface. For the purpose of our study, we selected to place a detailed view of a certain level of granularity – accessible by tilting using the shoebox effect - beneath the main interface, on the same plane. However, different positions in the 3D space may be tested further, in order to find the ideal position of the detailed view.

3. STUDY 1: MOBILE INTERACTION GESTURES FOR BROWSING IN HIERARCHICAL VIDEO INTERFACES

3.1 DESIGN SPACE

Modern smart-phones are equipped with sensitive surfaces capable of registering the contact and movement of the user's fingers on multiple points at the same time. Using this information, such devices can be programmed to recognize a number of pre-defined gestures used to control the applications present on the phone. This presents the developers with unique opportunities, as well as challenges when designing the interaction of a program – a gesture that is un-intuitive to the user can severely lessen their experience.

Modern media browsers for mobile devices already take full advantage of such methods of interaction – Marco Hudelist, Klaus Schoeffmann and David Ahlström [1] use the tap gesture to select a picture and present it in greater detail, while Steven Wijden and Wolfgang Hürst [2] use the swipe gesture as a method of navigation inside a 3D grid filled with images.

Browsing through the content of a movie is usually done by splitting it in key-frames and showing them in an orderly fashion to the user inside a 2D storyboard. Showing all possible frames to the user would be pointless as the redundant information shown on the screen would slow down browsing speed. In order to avoid this problem, the key-frames presented to the user are selected at regular time intervals or based on a content detection algorithm. However, in the case of long movies, browsing through the entire content still takes a long time.

A solution to this problem is to present the data in a hierarchical manner based on the time interval between two adjacent images, called levels of granularity. Normally, switching between different levels is done using the tap gesture, but this interaction presents two problems to app designers:

- the tap gesture generally overlaps with other actions – play the movie starting with the taped frame, present the selected image in more detail…
- this interaction doesn't provide a counterpart method for getting back to the previous level, so it must be done using other forms of interaction - such as the smart-phone's back button, thus breaking the natural flow of the user's interaction

An interesting research question generated by these problems is to identify what other gestures can be used for switching between different levels of granularity and test their effectiveness. In order to determine such a method of interaction, we must look at the different types of applications that are used for similar purposes and try to integrate them into our design. One example of an application that uses hierarchical browsing on multiple levels is Google Maps. The gesture used by Google is the pinch gesture – the user places his fingers on the screen and then narrows or widens the distance between his thumb and index finger. If we look at other programs, we can see that this interaction is consistently used for zooming.

This interaction method seems like a good way to deal with the issues presented by using the tap gesture – as we can both increase or decrease the space between the fingers, we have two opposed gestures at our disposal that can be used for both switching to a higher or lower level of granularity.

Despite the fact that the pinch is a solution to the problem posed by using tap, it doesn't present a straightforward way of integration within a video browser interface, as in this case the levels are differentiated in time and not space.

As we will further analyze the possibility of using this gesture for our interface, let's define the action of approaching the fingers as "close pinch" and the opposite action as "open pinch".

In order to address the problems specific to the tap gesture – as described above - we propose the following method of interaction using the pinch gesture:

1a. Initially, the user places the first finger over one image and the second over another in order to select them.

1b. At this point, doing an open pinch gesture would result in getting to a new level whose granularity is computed by dividing the time between the two pictures to the number of images on the screen.

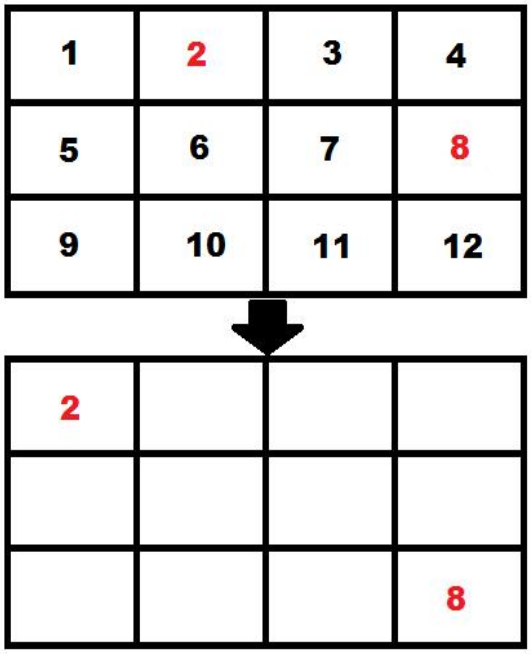


Figure 2: Moving to a lower granularity level using the pinch:
The user performs the pinch gesture starting with his fingers above images in red (2 and 8) and switches to a granularity level defined by the two selected pictures

2. In order to get back to the previous level of granularity the user just needs to perform a close pinch.

The pinch method also comes with another potential advantage – the users have more freedom in selecting their own granularity as opposed to navigating through the pre-determined levels offered when using the tap. The downside is that this gesture is more complicated to perform in comparison to the simple tap.

In order to prove the effectiveness of this approach to hierarchical browsing we must put it to the test in a practical experiment whose aim is to meet the following goals:

1. Test how the pinch gesture compares to the tap in the context of known-item video search tasks
2. Test whether the users prefer the pinch closed gesture as a method to getting to the previous level in favor of the back button
3. See what is the jump in the granularity factor the users prefer

When designing the experiment, there are a number of other parameters that can have an influence over the results, so must be taken into account:

- **Image size**

The size of the image represents the width of the selected inside the storyboard. In the work of Wolfgang Hürst et al [16], it is shown that a size of 70 pixels is enough the users understand the content of the images. As we are using screens of limited size, the number of images is directly influenced by their size – the larger their size, the smaller number of pictures we can fit onto the screen.

When selecting image size, we must also keep in mind that granularity levels in the pinch implementation are also dependent on the number of images (time between selected pictures/total number). We must also take into account that, as the images get smaller, the harder it becomes to interact with the interface because the user has to place his fingers over the pictures in order to switch between granularity levels.

Given the above information, we decided to select a size of 120 pixels as the width of the image, resulting in a total of 9 frames on the screen.

- **Granularity levels**

The starting granularity level must be selected for both methods of interaction. In order to not deviate from the point of the experiment, we don't want to give the user options other than hierarchical browsing, so we selected a granularity factor of 3 minutes in order to cover the whole video on the initial screen, without the need for swiping.

As described previously, in the case of the pinch gesture, granularity levels can be freely selected by the user, but for the tap, we must pre-define the jump between different levels. We have selected a granularity jump of 4x between two adjacent levels – the time between two frames in the next level is four times smaller than the current one. This factor gives the user enough information such that he does not lose perspective of his current place in the timeline, but also without getting too much redundant information from the previous level.

Another important element of decision in the case of the tap gesture is where the selected image should be placed inside the next level of granularity. Inspired by the work done by Wolfgang Hürst and Dimitri Darzentas [4], we decided to place the frame on exactly the same spot it was in the previous level. This would take advantage of the user's spatial memory and allow them to find the images easier. However, this approach also poses problems when selecting images close to the beginning or the end of the grid, because the number of new pictures shown before or after the selected frame is small.

User feedback suggested that this particular problem manifested a great deal in our implementation (due to the small number of pictures on the screen at one time – 9) and that is why we decided to change the placement for the second half of the experiment. Such, the selected frame would be placed on the middle of the grid when changing to a new level in the second version of the interface, which we call "click v2".

3.2 EXPERIMENT SETUP

The purpose of this test is to determine if we can successfully use the pinch gesture as a method of interaction for navigating through hierarchical video browsers.

- **Participants**

The test was conducted with 29 participants, all students enrolled in the "Multimodal Interaction" course at Utrecht University, 2012 – 2013. As described, all the test subjects have a background in Computer Science and are accustomed to the technology and methods used. The participants consist of 1 female and 28 males, with ages ranging between 21 to 28 years old.

- **Apparatus**

All the 29 user tests were performed on a GT-S7500 device with a resolution of 320x480 pixels shown on a 62x114 mm screen. For each of the tasks performed by the users, the task correctness, the completion time and number of gestures are recorded.

- **Data Set**

In order to simulate a KIS task, the data was selected from a well known movie series – The Lord of the Rings trilogy. This type of tasks presumes that the users are familiar with the data, hence our selection. Out of the 29 people participating in the experiment, only 2 were unfamiliar with the movies. Frames from the movies were extracted at regular intervals, resulting in 130 total images captured.

- **Procedure**

- **Hands-on Task**

Prior to beginning the tests, an overall explanation about the interfaces and methods of interaction is given to the participants, followed by a try-out task where they are free to test with the given interface. When they get used to the interaction method, the recorded tests begin.

- **Tasks**

When starting a new task, a picture is selected pseudo-randomly from the frames and shown on screen as the target frame. The pseudo-random algorithm selects images in such a way that the users have to "zoom" in to the smallest level of granularity in order to find the target (thus make use of the tap or pinch gestures). When a user sees the target image on screen, he confirms that he found it by tapping the options button. Each user has to complete a total of 10 tasks using each gesture.

- **Interface version**

Due to the problems caused by selecting images close to the end or the beginning of the grid, the first implementation of the slick interface obtained subpar results. Based on user feedback, we implemented a second click interface, used in the second part of our experimentation.

The first half of the tester base (14 users) tested the pinch interface and the first version of the click, while the second half (15 users) was given the pinch and the second version of click as test interfaces. As a consequence, we decided to treat the results of each half of the experimentation separately.

➢ Questionnaire

Following the practical tests, the users answered a series of questions about which interaction method they prefer and why.

3.3 RESULTS

3.3.1 RECORDED DATA

In the next section the data recorded from the user tasks – time and number of mistakes - are shown and discussed.

➢ **Total Time** – (in milliseconds) the duration the users took to finish all the tasks using one of the interaction methods

User	Total Time	
	Click v1	Pinch
1	396559	145175
2	852141	256586
3	1003659	410824
4	184661	296802
5	671609	319033
6	732375	285109
7	438053	425477
8	242646	156092
9	178807	169864
10	230248	208301
11	206186	130401
12	296777	279855
13	243858	183460
14	112037	141988
Average	**413544**	**243497,6**
Standard deviation	**274690,3**	**93942,8**

Table 1: Total time results – the interface using the pinch gesture outperforms the one using the click v1

As we can see, the first version of the click interaction method performs worse than the pinch interaction in terms of task completion time. This is caused by the fact the beginning and ending parts of the video are very hard to explore because of the low number of new pictures shown when switching to a new level of granularity.

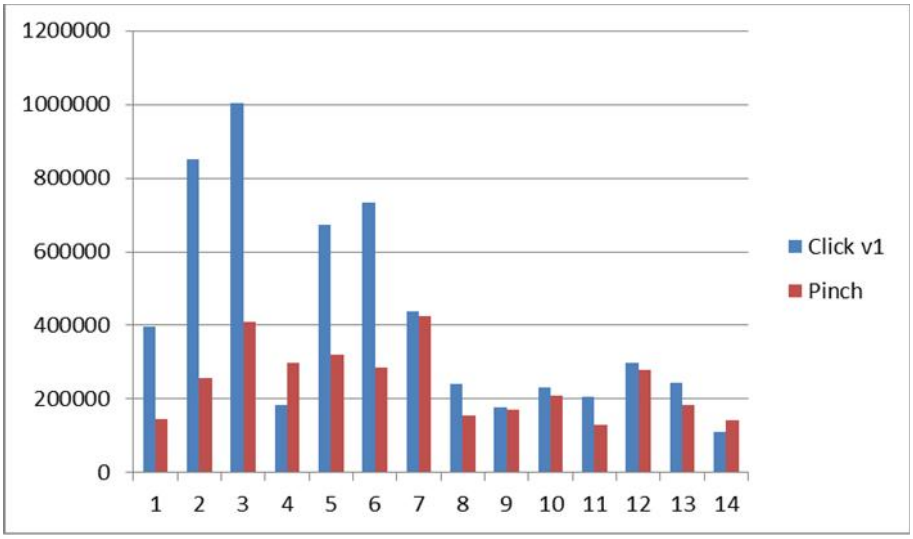


Figure 3: The overall task completion time of
both the click and pinch interfaces for each user

From the graphic showing the task completion time for each user we can observe that the click interface performs particularly worse during the first half of the tests where it was the first interface shown to the users. On the second half of the tests, where the pinch is shown first, the overall task completion time begins to even out for each user.

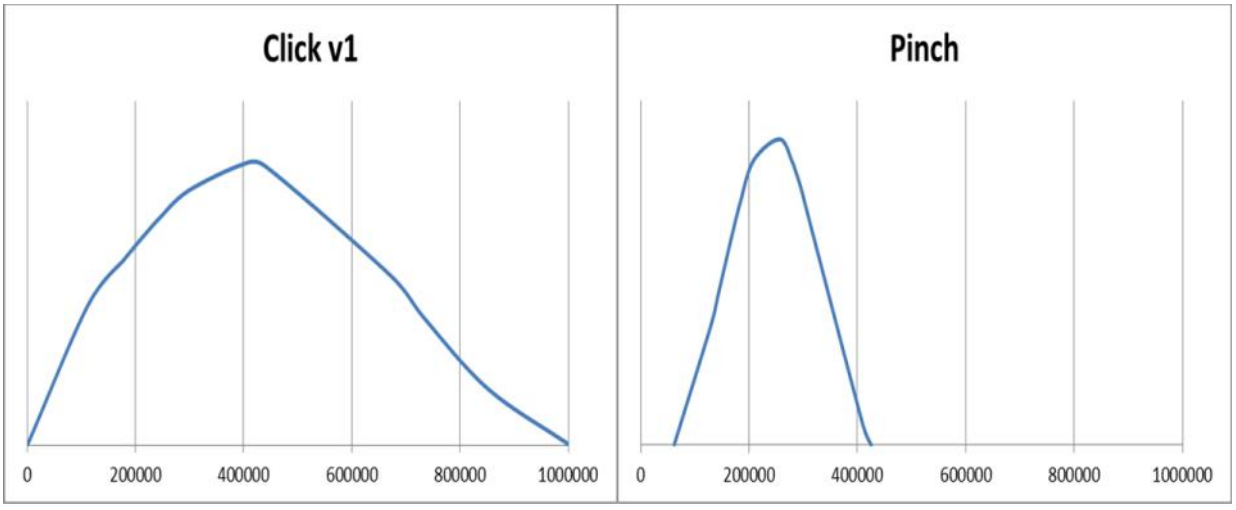


Figure 4: Normal distribution of the recorded time
for both the pinch and click interfaces

Following the normal distribution of both interfaces, we can see that the results of the Click interaction are more spread out (standard deviation is 275690) around the average than in the case of the Pinch interaction (standard deviation is 93942). It can also be observed that the average value of the interface using pinch is smaller than the one using click as a method of interaction.

To measure the significance of the results, we compare them using a Paired t-test with 13 **degrees of freedom,** for a significance level of **95%**. We conduct the test one-paired in order to test the hypothesis if the click has a higher task completion time than the pinch (H_a: $D_{click\text{-}pinch}>0$) as opposed to the null hypothesis that the two interfaces would score the same (H_0: $D_{click\text{-}pinch}=0$). Following the calculations, we obtain a **T-Value** of 2.80 and a **P-Value** of 0.9925. The results show that we can say with 99% confidence that the pinch will score a lower task completion time than the click. As the probability we obtained is higher than the set significance level of 95%, we can accept the alternate Hypothesis as being true.

User	Total Time	
	Click v2	Pinch
1	154892	338341
2	330592	324880
3	322622	211991
4	632053	390725
5	491330	768764
6	628924	321461
7	219403	236272
8	638846	616485
9	372702	479985
10	265818	374055
11	340811	434164
12	452933	934140
13	281882	219469
14	196933	265359
15	415375	368986
Average	**383007,7**	**419005,1**
Standard deviation	**153033,7**	**200085,5**

Table 2: The total task completion time of the pinch and the second version of click

The results from the second interface show that the second version of the click interface scores a bit better than the first version, with an improvement of 30537 miliseconds on average. The most interesting observation is that the second version of the click outperforms the pinch, but mostly due to the worse performance of the pinch, which lost 175508 miliseconds on average when compared to the results of the first 14 participants.

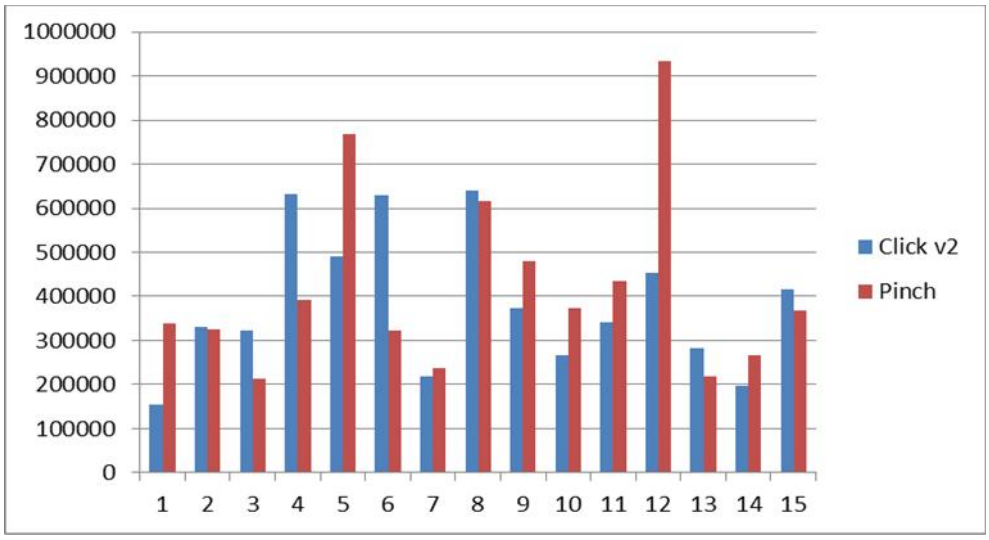


Figure 5: The task completion time of each user
for both the click and the pinch interfaces

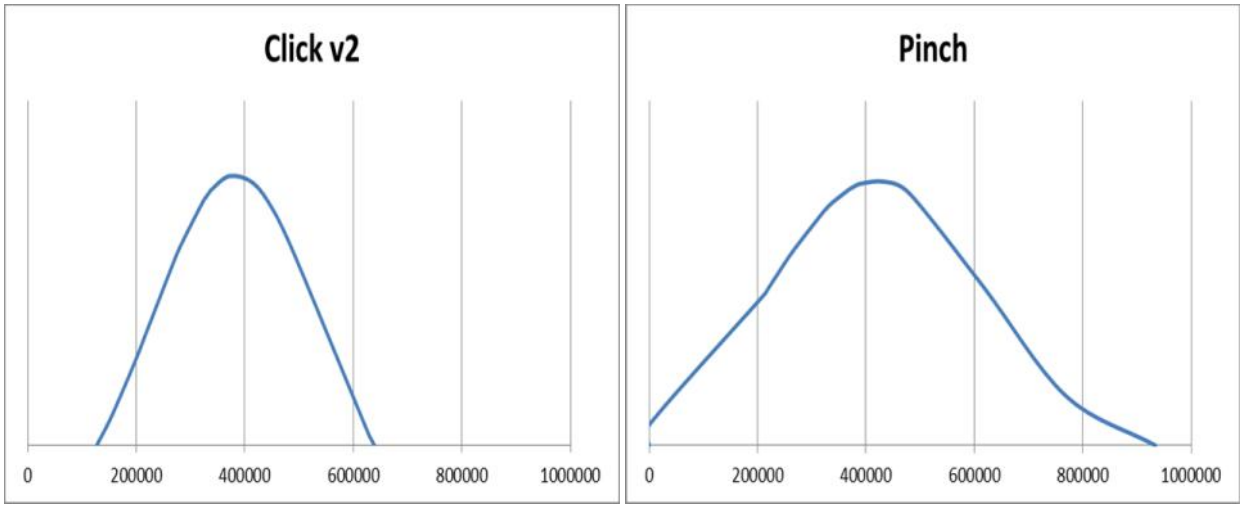


Figure 6: The normal distribution of the total task completion time
in the second part of the study

As the overall score of the click was better than the score of the pinch, we will perform a test on the hypothesis with an expected significance level of 95%. The alternate hypothesis states that the pinch scores are higher than the click (H_a: $D_{pinch-click}>0$) as opposed to the null hypothesis (H_0: $D_{click-pinch}=0$). We obtain a **T-Value** of 0.74 for 14 **degrees of freedom**, with the corresponding P-Value of 0.76. The probability of the alternate hypothesis to be true (76%) is lower than the expected 95% significance level, thus we reject it and accept the null hypothesis as true - there is no statistical difference between the task completion time of the two interfaces.

- **Errors** – the number of tasks where the test subjects selected a picture other than the target

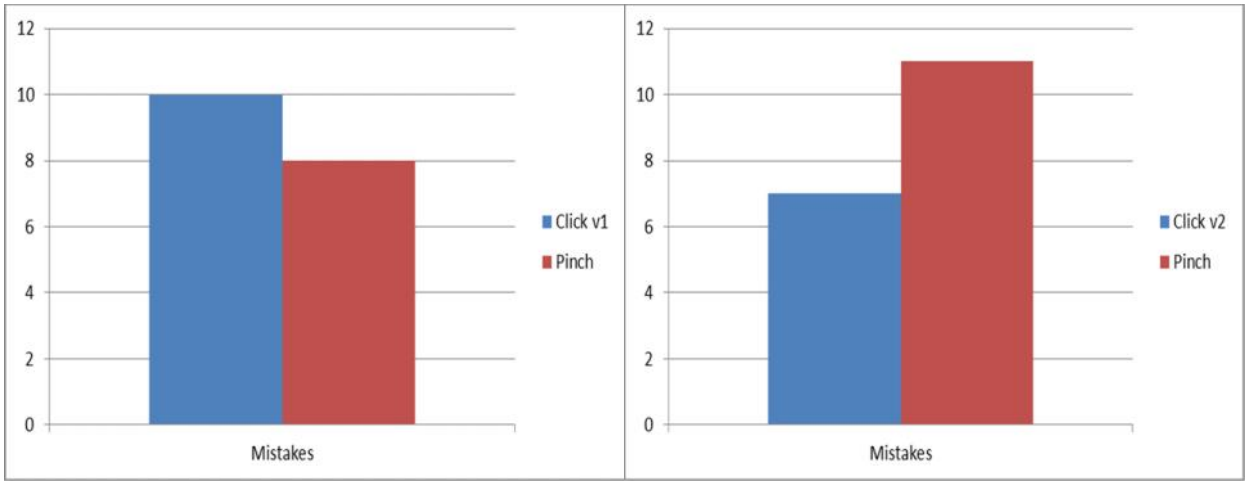


Figure 7: The total number of mistakes for the first part (left)
and the second part of the study (right)

The number of errors registered, follow the same pattern as the overall task time – the first version of click performs worse than the pinch, while the second version performs better. However, due to the small number of mistakes done – 36 errors in 464 tasks – the difference between the interfaces is insignificant.

- **Gestures** – number of actions required for a user to finish a task

Because the pinch gesture provides alternative means to go back to previous levels of granularity – and potential solutions to one of the problems of the click based interfaces - we regard with great interest the question if the users would still have used the back button once they had the close pinch gesture as an alternative.

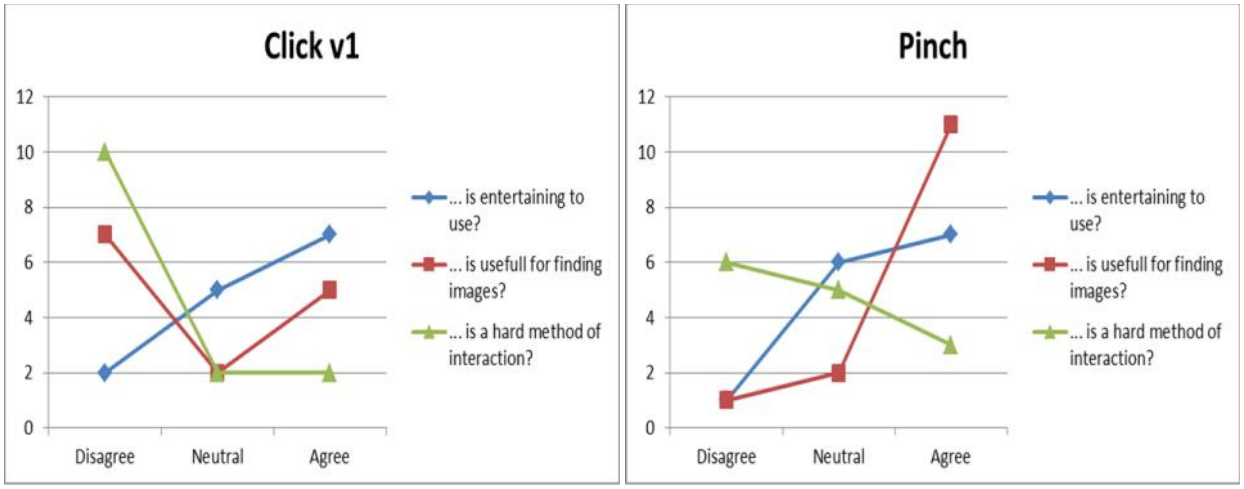
User	Close Pinch	Back Button
1	6	0
2	35	0
3	27	0
4	35	0
5	21	0
6	39	0
7	65	2
8	15	0
9	10	0
10	37	1
11	7	0
12	62	0
13	19	0
14	10	0
15	78	0
16	13	0
17	29	6
18	75	0
19	98	0
20	46	0
21	14	0
22	260	0
23	73	2
24	55	0
25	42	1
26	245	0
27	34	0
28	30	0
29	59	0
Total	**1548**	**12**

Table 3: The number of gestures performed by the users to get to a previous level of granularity, either by clicking the back button or using the close pinch gesture

In the above table, it is clearly shown that users stopped using the back button once they got an alternative that did not interrupt the natural flow of their actions – 12 uses of the back button compared to 1548 uses of the close pinch. This data confirms our assumption that the pinch can solve one of the problems presented by the tap gesture.

3.3.2 QUESTIONNAIRE

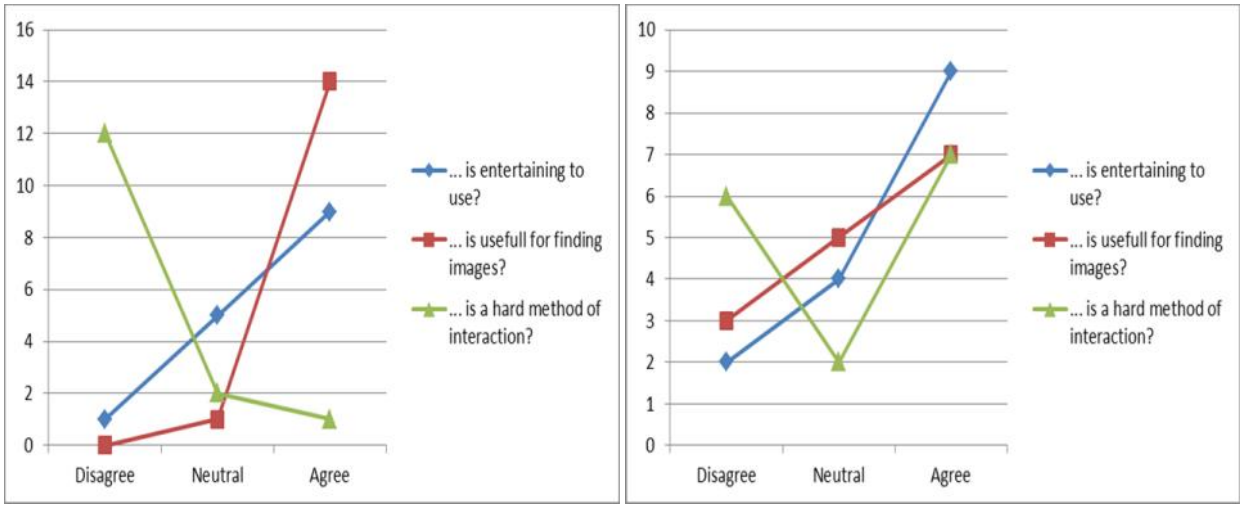
This section will describe the data resulted from the answers to the questions asked after the completion of the practical tasks.



Click v1 Pinch

Figure 8: The user responses on how difficult, helpful and entertaining is the interface using the first version of the click (left) or the pinch (right)

The data gathered from the questionnaires shows that the first version of click got mixed results in difficulty and usefulness, while the pinch scored a lot better – the majority of users considered it useful and easier to use than Click v1. However, in terms of fun the click scored better than the pinch.



Click v2 Pinch

Figure 9: The user responses on how difficult, helpful and entertaining is the interface using the second version of the click (left) or the pinch (right)

When compared to the second implementation of the tap gesture, the pinch scored worse in all three categories – difficult to use, usefulness and fun. This comes from the fact that the users found the pinch a lot more difficult to use when compared to the simplicity of the tap.

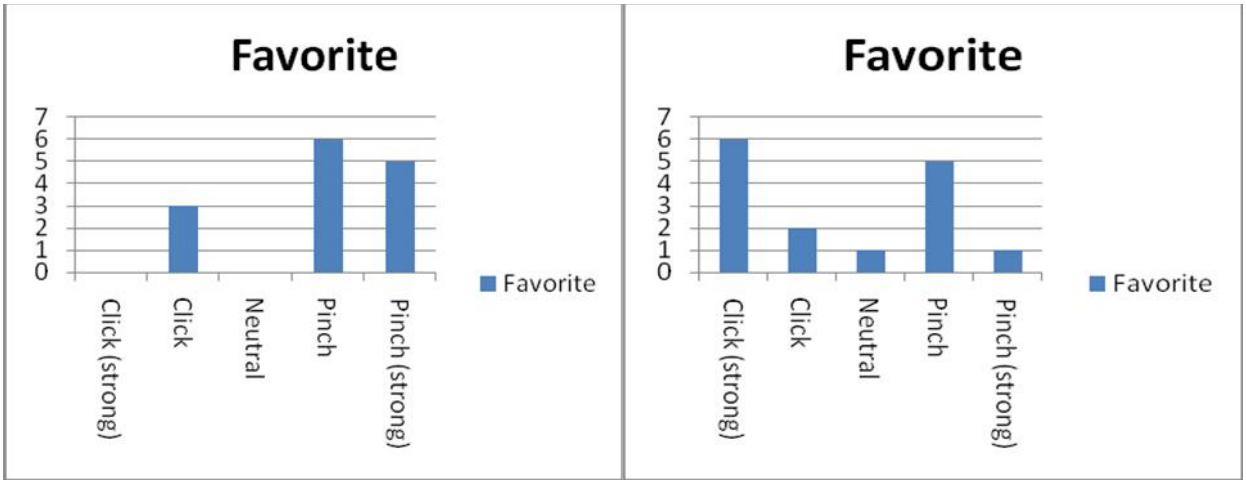


Figure 10: The overall preferences of the users in regards
to their favorite method of interaction in the first part of the study (left)
and the second part of the study (right)

The overall trend continues as the users were asked to choose their preferred interface – if the users chosen the pinch over the first version of the click in an overwhelming manner, in the second half of the study, the users voted in favor of the second version of the click.

3.4 CONCLUSION

Following the results of the experiment, we observed that as the gesture used gets more complicated, fewer people are able to use it efficiently, consideration that can also be found in the work of found in Lao et al [14]. This remark is reinforced by the fact that the participants preferred the click over the pinch, while categorizing the latter as a difficult method of interaction.

The results show that the users prefer gestures which are symmetrical, even if this impedes some of their options – they didn't used the back button once an alternative (the close pinch) was offered to them. We conclude that using the back button is an overall bad design choice as it breaks the natural flow of the user's actions.

Although the pinch offers solutions to the problems of the regular click gesture, its results aren't the ones we expected. As we analyzed results differences between the two tests and the questionnaires, we observed that the precision required for performing the pinch gesture is a great impediment in its efficiency of usage. Even with its higher skill requirement, the results of the pinch are comparable those of the click, suggesting that multi-touch gestures can obtain significantly better results, if their difficulty is lowered.

4. STUDY 2: HIERARCHICAL VIDEO BROWSING MODELS

4.1 DESIGN SPACE

Video browsing interfaces work by extracting frames from movies and presenting them to the user in a grid like arrangement. A novel approach built upon such interfaces present the data hierarchically on different layers - called levels of granularity -defined by the time between extracted frames. This method offers a lot more flexibility to the user and it was proven by Guillemot et al [3] that it offers better practical results than classical video players.

A number of different approaches to the hierarchical model can be distinguished:

1) One method is to present each level of granularity on the whole screen. This approach can be observed in the study of Wolfgang Hürst and Dimitri Darzentas [7]. The large area of exploration of each layer of granularity is a potential advantage for this approach, but the navigation between levels is much more difficult. For the purpose of this experiment, we will define this method as the "grid hierarchy".

Figure 11: Overview of the first level in the grid interface

2) In the second method, each level of granularity is shown on only a small portion of the screen, such that multiple layers can be shown at the same time. The effectiveness of this approach was proven by Manfred Del Fabro, Bernd Munzer and Laszlo Boszormenyi [9], which was selected as the winner of the Video Browser Showdown in 2012. This approach offers faster switching between different levels, but at the cost of seeing only a small number of pictures from the current level of granularity at one time. For the purpose of this experiment, we will define this method as the "tree hierarchy".

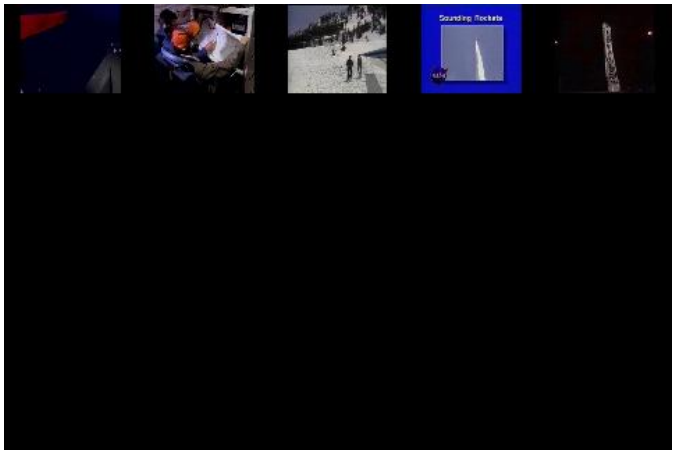


Figure 12: Overview of the first level in the tree interface

Each of the two methods comes with both advantages and disadvantages and from their differences we can derive an interesting research question: which of the two approaches is better suited for video browsing. The goal of this experiment is to try and answer the above question. For this purpose we set up a number of goals:

1. To test which of the two interfaces offers better results in the context of Known Item Search tasks
2. To test which of the two approaches offers better precision, by measuring the differences between the point in time requested by the search and the time selected by the user
3. To test which method will help the user obtain the end requested result faster once the last level of granularity is on screen
4. To test which of the two interfaces gets the user faster to the level of granularity required to complete a task

In order to make the experiment more focused, we decided to restrict the usage of the swiping gesture for browsing through the frames on the screen. This will force the user to change between the different levels of granularity, thus making the results more conclusive. In order to fully define the experiment, we must also decide on the values of the parameters associated with the two approaches:

- **Image size** – represents the width in pixels of one extracted frame on the screen. Following the findings of W. Hürst, Cees G. M. Snoek, W. J. Spoel and M. Tomin in "Size Matters! How Thumbnail Number, Size, and Motion Influence Mobile Video Retrieval", we decided to select a size of 90 pixels for our images in order to be above the 70 pixels threshold set in the paper.
- **Granularity factor** – is the difference in time between two adjacent frames extracted from the video. The initial factor must be selected in such a way that it covers all the content of the movie. This factor is different for each method due to the different number of frames presented at each level. For the grid hierarchy we selected a factor of 75 seconds, while for the tree hierarchy we selected a factor of 180 seconds.
- **Granularity jump** – the difference between the granularity factor of two successive levels. As the grid hierarchy shows a larger number of frames from the same level at one time, its granularity jump can be greater than the one of tree hierarchy. This leads the grid hierarchy to get from the highest to the lowest level of granularity in three steps, while the tree hierarchy goes in four steps.

4.2 SETUP

The goal of the second experiment is to compare two approaches to interface design based on hierarchical browsing. The two competing interfaces are based on previous work done by Wolfgang Hürst and Dimitri Darzentas [1] and Manfred Del Fabro, Bernd Munzer and Laszlo Boszormenyi [2] as presented in the section above.

➢ Participants

A total number of 18 people took part in the experiment. The participant's age is within the range of 16 to 46 years old, with an average of 26 years old. They have varying levels of experience in using mobile tactile devices: all of them used a smart-phone at some point, but only 11 own and use one regularly. Out of the 18 test subjects, half were male and half were female.

➢ Apparatus

The experiment was conducted on a Samsung GT-S7500 running Android version 2.2. Each of the tasks performed by the users is recorded and saved as a log file on the phone's memory card.

➢ Data Set

The data was taken from the example video file used in the Video Browser Showdown competition held during the International Conference on MultiMedia Modeling 2012. Seventeen segments were extracted from the example video file to serve as task goals, with a duration ranging from 8 to 20 seconds. These clips are extracted in such a way that the user has to go to the deepest level of granularity in order to get a perfect answer, and to explore at least a sub-level in order to get a correct answer. The key-frames used in the interfaces were also extracted from the movie before-hand, one at every 3 seconds, for a total of 180 pictures.

➢ Procedure

- **Tasks**

Before each task, a short goal movie is shown on screen to the user. After watching the short clip, the user has to search through the given interface and select the frame closest to the beginning of the clip he just viewed. If needed, the user has the possibility to view the goal video again, using the Show Movie option button.

Figure 13: The goal movie for the user to search
is presented to the user before the start of the test

If the selected image is closer than 15 seconds to the one requested, the test will be deemed as successful and the time difference will be recorded. If the difference is higher than 15 seconds, the log will indicate that the task was failed. In both cases, the search time is also recorded.

- **Hands-on Task**

Before starting the recorded experiments, the users will receive explanations about each interface. After the description is given, the users will receive a try-out task, when they are free to interact with the interface and ask any related questions. The correctness and completion time of this task are not recorded.

- **Goals**

Out of the seventeen goal segments selected from the movie to be browsed, one is used for the hands-on task as described above. The other sixteen movies are split into two groups. The first nine users were given clips from the first group to be found using the first interface and clips from the second groups in the second interface. For the last nine users the groups were switched around. Movies from each group are selected in a random order.

- **Questionnaire**

After finishing the interactive tests, each participant must fill a short questionnaire where they have to select which of the two interfaces is easier to use and which is more useful in KIS tasks.

4.3 RESULTS

4.3.1 RECORDED DATA

In the next section we will show and analyze the data gathered from the practical tests.

- **Time –** recorded in milliseconds, it represents the duration each user took to finish the tasks on one of the two given interfaces

In the following table we present the total time – the sum over all 8 tasks – as well as the average time per task, for each of the 18 users on the two interfaces:

User	Total Time	
	Tree	Grid
1	713130	699570
2	680751	934556
3	892410	835693
4	460441	399777
5	438158	297381
6	519687	538545
7	293307	241536
8	525022	428238
9	629965	415962
10	691515	631763
11	246168	345531
12	314661	579480
13	364357	529363
14	492309	372961
15	428500	510146
16	209432	559918
17	449449	380110
18	982715	973480
Average	**518443,1**	**537445**
Standard deviation	**206962,5**	**204736,9**

Table 4: Task completion time difference between the tree and the grid interface

We can see in the table above that the tree interface scores slightly better than the grid in regards to the time it takes to complete all the tasks. This result is not surprising, as the tree interface allows faster switching between different levels of granularity in the interface.

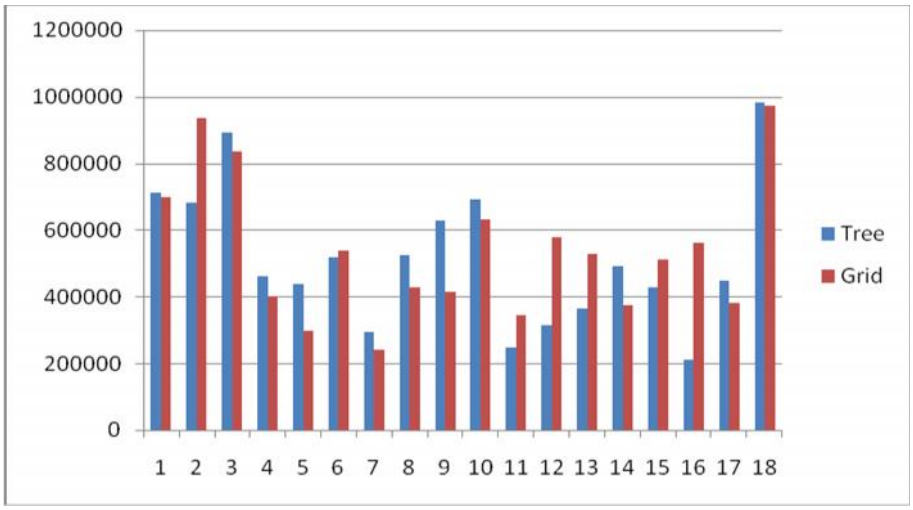


Figure 14: the completion time of all tasks on each of the two interfaces for each user

Eleven out of the eighteen users scored worse on the interface they used first – seven out of the first half and five out of the last half. Although the number of users that scored worse on the grid as the first interface is lower than in the tree case, their results were significantly worse. This makes the grid slower than the tree interface.

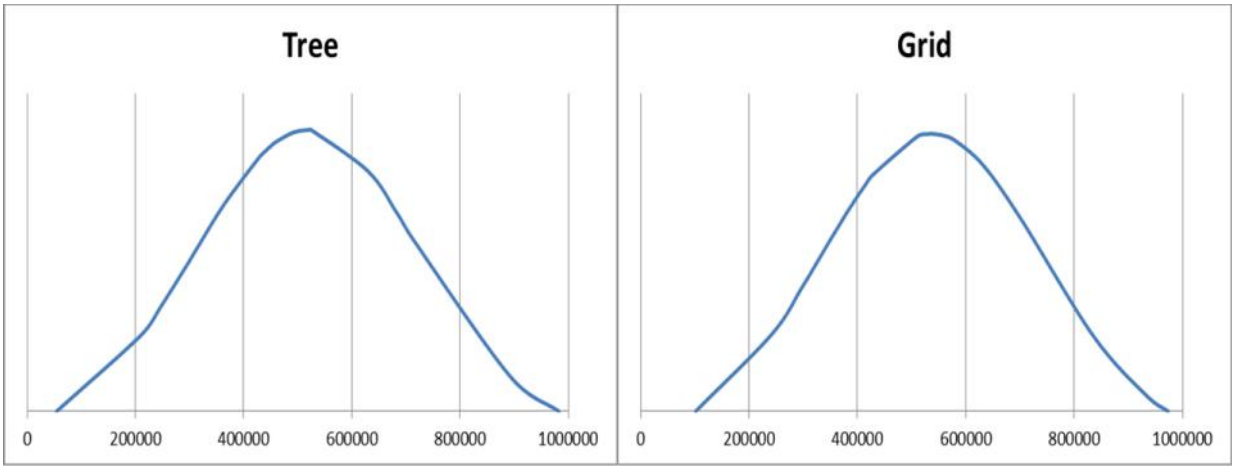


Figure 15: The normal distribution of the task completion time for the tree (left) and the grid (right) interface

The normal distribution of the completion time of the two interfaces is about the same, with relatively small differences – 95% of the tree interface scores reside in the interval between 104 and 932 seconds, while in the grid case these are situated in between 127 and 946 seconds.

In order to verify the statistical significance between the two interfaces we performed a Paired P-test on the alternate hypothesis that the grid scores significantly worse than the tree (H_a: $D_{grid-tree}>0$), while the null hypothesis states that the difference between the two is insignificant (H_a: $D_{grid-tree}=0$). We perform the test with a significance level of 95%. The results of our data with 17 degrees of freedom are the following: a T-Value of 0.53 and a P-Value of 0.7. As the computed probability is smaller than the one expected we reject the alternate hypothesis.

- **Mistakes** – the number of tasks where the difference on the timeline between the chosen image and the target one is larger than fifteen seconds.

User	Total Time	
	Tree	Grid
1	2	0
2	2	1
3	1	0
4	6	1
5	2	0
6	1	1
7	3	3
8	1	1
9	2	3
10	0	1
11	5	5
12	2	4
13	1	4
14	0	0
15	0	0
16	1	1
17	1	1
18	0	0
Total	30	26

Table 5: The total number of errors performed by all the users during the usage of the tree and the grid interfaces

Following the test results, the grid hierarchy got a lower number of mistakes than the tree. This result is in correlation with our expected results as the grid offers a better overview of a single level of granularity, thus making searching for a particular image easier once the user gets to the desired level.

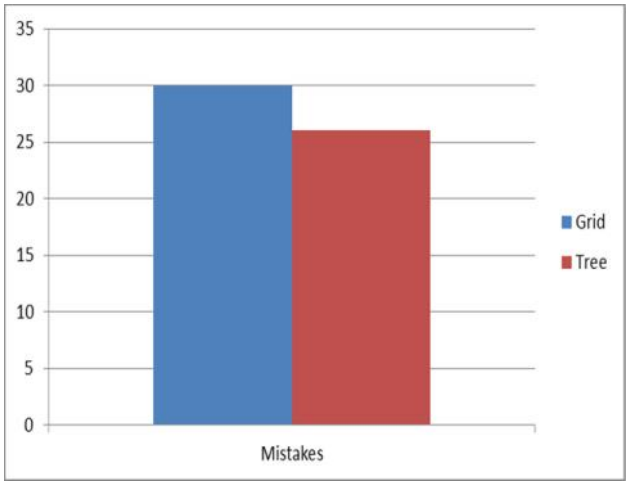


Figure 16: Comparative view of the number of mistakes performed, between the grid and the tree interface

- **Time Difference** – the difference (in seconds) on the timeline between the picture selected by the user and the goal image.

The time difference is recorded only if it is smaller than 15 seconds (5 frames apart), in all other cases it is marked as a mistake.

User	Time difference	
	Tree	Grid
1	1,00	2,25
2	5,00	3,42
3	1,28	5,25
4	1,50	1,28
5	3,00	7,12
6	2,14	0,00
7	3,00	6,00
8	3,75	3,75
9	6,00	7,20
10	0,00	0,00
11	7,00	3,00
12	3,00	3,00
13	2,14	4,50
14	0,00	1,85
15	0,30	0,30
16	1,28	0,85
17	0,30	0,00
18	0,37	0,37
Average	**2,28**	**2,79**
Standard deviation	**2,01**	**2,37**

Table 6: The time difference between recorded during the study on both the grid and tree interfaces

In table 6, we can see that the tree hierarchy got better scores in the time difference category. The result is surprising as we expected that the interface that offers a better view over one level would allow the user to make more precise selections.

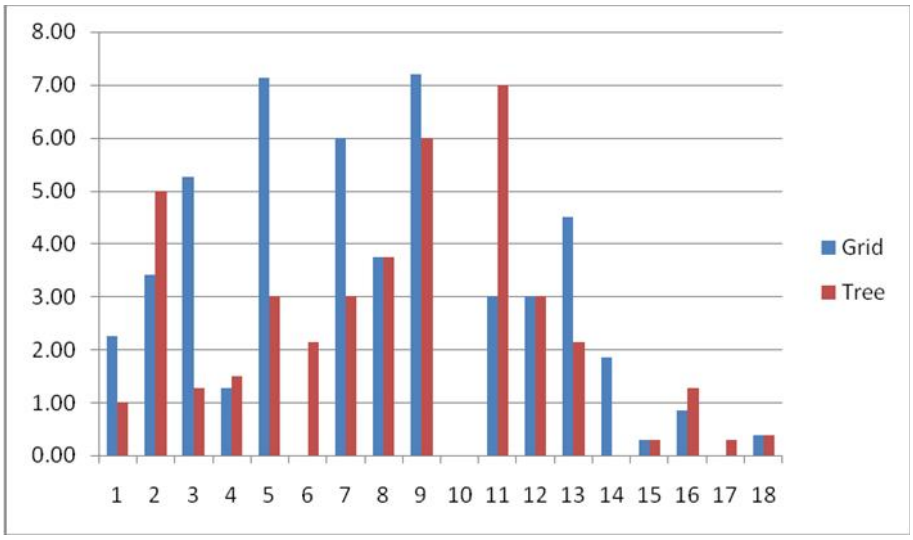


Figure 17: The time difference between the expected result and the user selection recorded during the practical tasks on both interfaces

If we look at the data in Figure 17, we can see that the order of the interfaces is not important when measuring the time difference between.

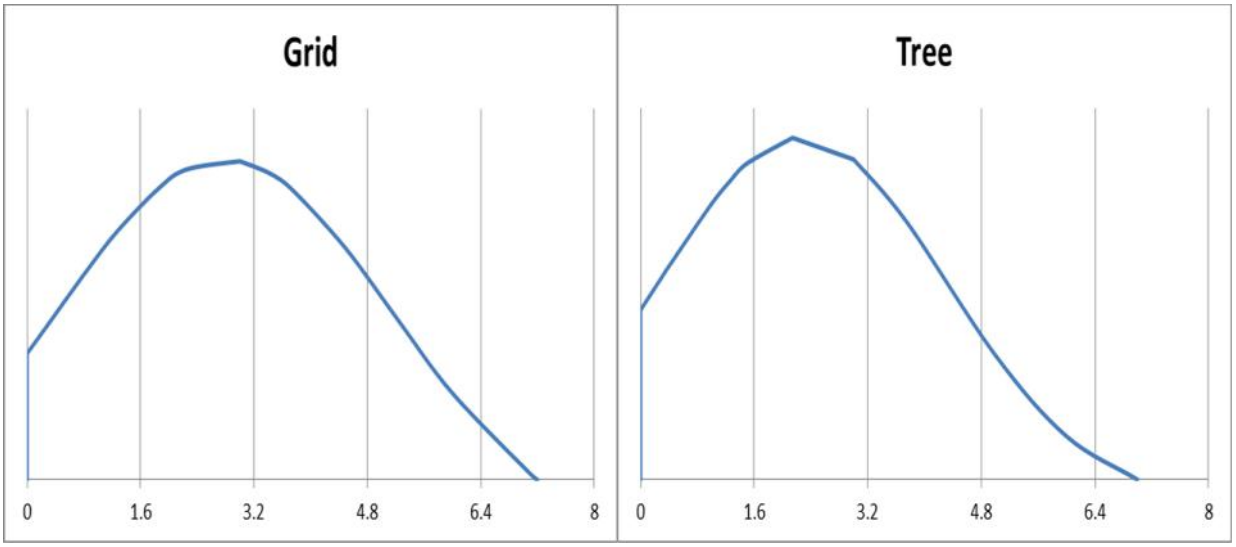


Figure 18: The normal distribution of the time difference for the tree (left) and the grid (right) interface

We perform a significance analysis in order to test whether the grid scored worse (H_a: $D_{grid-tree}>0$) due to chance or if there is no statistical difference between the two interfaces (H_0: $D_{grid-tree}=0$) on the results coming from 18 participants (17 degrees of freedom). The T-Value obtained following the analysis is equal to 1.06, while the P-Value is situated at 84% (0.84), thus making us reject the hypothesis and accept that the difference between the two interfaces occurred by chance.

4.3.2 QUESTIONNAIRE

This section will describe the data gathered from the questionnaires that the test users filled out after completing the practical test.

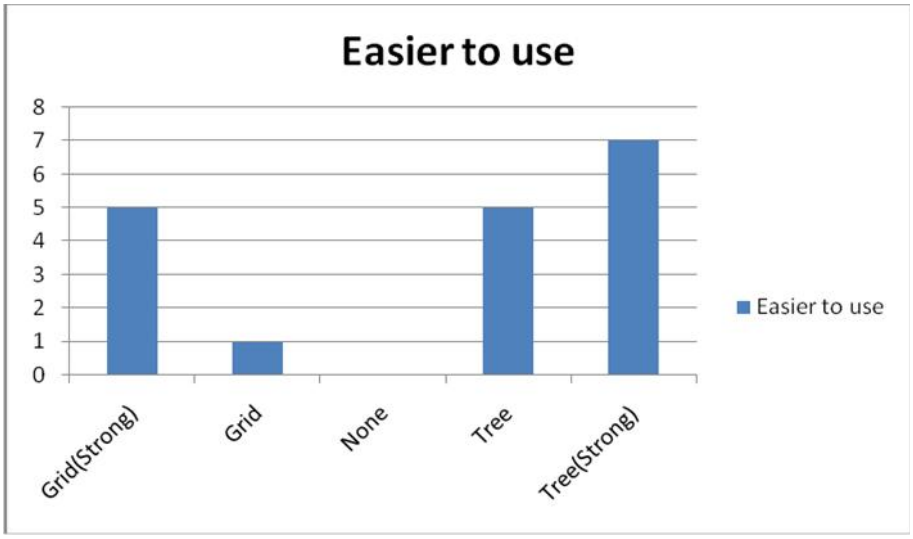


Figure 19: The ease of use for each of the two interfaces according to user feedback

Most of the users complained that the way the grid handles switching between levels is disorienting, thus the tree was the interface of choice in regards to the ease of use.

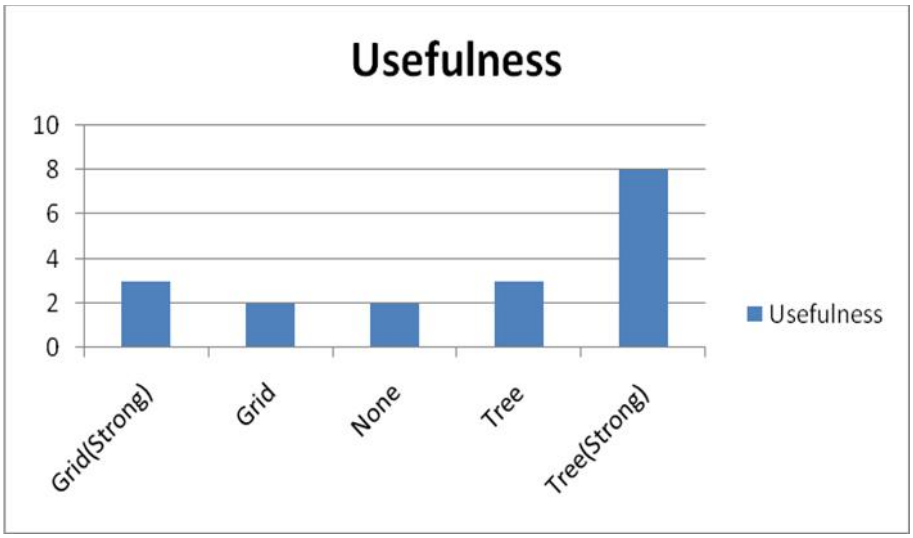


Figure 20: Interface usefulness according to user feedback

Similarly to the ease of use, most users also favored the tree hierarchy in terms of usefulness. When asked why they preferred the tree over the grid, they answered that the fast switching between levels allows for faster searching.

4.4 CONCLUSION

The tree scored better in terms of tasks completion time than the grid. Even if the differences between the two were statistically insignificant, we observed that by allowing the users to view multiple levels of granularity and, as a result, switch faster between them they tend to perform the tasks faster. The appreciations made by the subjects in the questionnaires seem to support our considerations, as the majority voted for the three as the more useful interface.

Because we don't allow swiping in the two interfaces, it is easier for a user to navigate to a part of the movie close to the one he is currently viewing by using the tree than by using the grid. We assume that this is the cause that leads to better results for the tree in terms of precision, but this has to be further tested by adding the option to swipe inside the two interfaces and performing another study.

We also observed that, as the testers got a better overview over one particular part of the movie, they have a higher chance of successfully finding the target picture, hence the smaller number of mistakes for the grid interface.

Our observations seem to suggest that the tree interface is better suited for high level search, where the goal is to determine the position of a scene in a larger movie, while the grid is more efficient at more in-depth searches, such as determining the exact starting and ending frames of the scene. Each of the two gestures seems to perform better at different steps of known item search tasks. If this assumption is correct, better results can be obtained by combining the two interfaces, theory that we shall put to the test in the third study.

5. STUDY 3: COMBINING THE BENEFITS OF THE TREE AND GRID INTERFACES

5.1 DESIGN SPACE

One of the goals of this paper was to design an interface based on the findings of the previously conducted tests. One such interface would combine the advantages of both interfaces tested during experiment 2, by showing off as many levels of granularity at one time on the screen, but with the emphasis on a single level. In order to do this we must ensure that all the levels show the same point in the timeline at any given time – whenever a user performs an action (swipe left/right on the timeline) on a given level, it will also influence all the other layers of granularity found on screen.

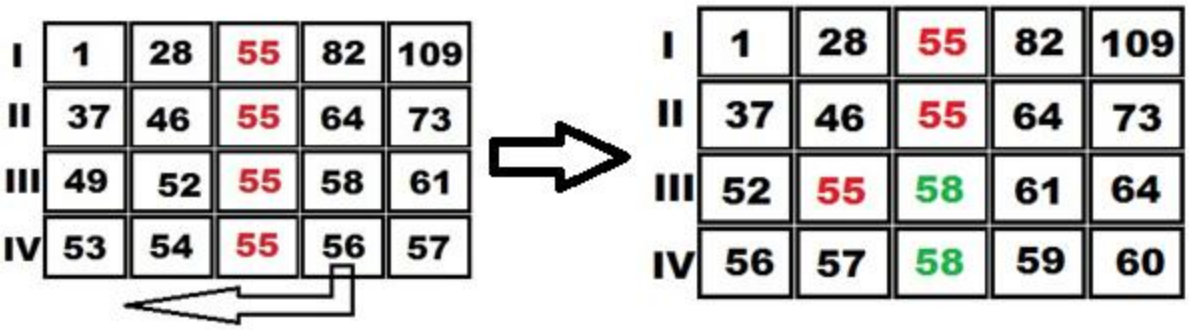


Figure 21: The connection between levels is made on the middle column – if we move the pictures from row III to the right, picture number 58 will appear on the center of the screen, the row below (IV) will also be moved to the right by 3 positions

As the user explores the movie, he does so by interacting with only one level of granularity at one time, so all the other levels lose their utility until the user decides to switch focus to them. Immediately after the user decides to start interacting with another level, an element of confusion occurs due to the fact that the two levels can show a different time span from the main movie. This can be avoided by interconnecting the levels as shown in Figure 21 and practically presenting the same spot in time on all levels, but at different granularities.

We would also like to give the user the opportunity to switch fast between the multi-level view and a more detailed overview of a single level of granularity. To achieve this, we allow the users to get to the secondary detailed view by tilting the device.

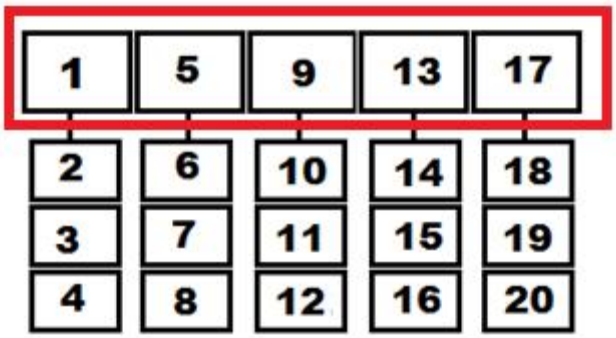


Figure 22: The detailed view – the examined level is placed on top of the screen, while beneath it its sublevels are shown on columns

Another topic of discussion related to this occurs, where we should place the detailed view inside the interface. For reasons of continuity (to offer a more intuitive understanding of this process to the user) we selected to place the detailed view directly below the main part of the interface, on the same distance from the screen. As we want to perform a study focused on certain aspects of interfaces, we will neglect other options for positioning the detailed view (ex: at a 90^{o} angle relative to the main interface, …) as these will be the topic of research of another study.

Another important design aspect is the way a user would interact with the interface. The movement through the timeline can be easily defined using left and right swiping gestures, but the interaction used for switching between different levels of granularity is harder to design. As seen in the first experiment, a simpler gesture is preferred over a gesture more complex interaction, even if it offers more advantages. In the same experiment, we also observed that users prefer a gesture for getting back to other granularity level instead of the back button. For our implementation, we can easily solve this problem by using the placement of the levels on the bottom of the screen:

- After the user touches the screen, if he moves the finger left or right, the level on which the finger is placed will move left and right, but will also influence the higher and lower levels of granularity. If the user does not touch any level while placing his finger on the screen, no interaction will occur.

- If the user will drag his finger up, a new level of granularity will be shown on the bottom of the screen, given there are still lower levels of granularity to explore.

- If the user will drag his finger down, the level of granularity that was touched with the finger will immediately become the lowest level on the bottom of the screen, and all the levels beneath it will disappear. If no level is selected with the finger, no action will be performed.

- If the user will not move his finger, if he then lifts it off the screen, it will result in a selection action centering all the levels over that position.

5.2 EXPERIMENT SETUP

The goal of the third experiment is to evaluate the performance and user appeal of the designed interface by comparing it with the classical approach to present a movie in a tree-like hierarchical manner. The time a user takes to complete one task and the correctness of the completed task will be recorded during the practical experiment, while the users opinions of the interface's usefulness and appearance will be gathered with the use of a questionnaire.
A secondary goal is to compare if the detailed view proves a useful tool during the search tasks.

- **Participants**

A total number of 24 people took part in the experiment with the ages between 19 and 47 years old. The test subjects had varying levels of experience in using mobile tactile devices, but all of them used one such device at least once.

- **Apparatus**

The experiment was conducted on a Samsung GT-S7500 running Android version 2.2 Froyo. The data recorded during the completion of the given tasks was recorded on the memory card of the device.

- **Data Set**

As in the previous experiment, the data was selected from a one hour long example video file used in the Video Browser Showdown competition held during the International Conference on MultiMedia Modeling 2012. The pictures used for browsing were selected from the movie at every 13 seconds, with a total of 256 frames. The goal videos were also extracted from the movie and selected in such a way that the user needs to explore the video to at least on the third level of granularity.

- **Procedure**

- **Tasks**

Prior to starting each task, the user will be presented with the goal movie on the screen. When the user is done with visualizing the movie, he will have to press either the back button or anywhere outside the frame of the presented movie.

Figure 23: The goal video is presented to the users before starting each task

If the user needs to review the goal movie, he can do so by pressing the "View Movie" button from the options.

- **Demo**

Before starting the practical tests, each user will get time to get used with the interface and is free to ask questions about the interaction method. When the subject feels comfortable with the use of the given interface, he can press the "Start Test" button to begin the practical tasks.

- **Goals**

The goal movies were split in two groups (G1 and G2), prior to starting any test. The first half of the users will randomly receive tasks for the designed interface from G1 and for the comparison interface from G2. The second half of the users will receive tasks from G2 for the designed interface and tasks from G1 for the comparison interface. This will ensure an equal distribution of the tasks over the two interfaces, such that none of the two would gain an unfair advantage.

- **Questionnaire**

After a subject finished all the tasks handed to him in the practical test, he will be given a questionnaire to complete, where he is asked to choose which of the two interfaces was easier to use, more intuitive, useful and better looking. All the results from the questionnaire will be stored in digital format in the form of .doc documents.

5.3 RESULTS

5.3.1 RECORDED

- Time – the time the user takes to complete all the tasks on each interface

User	Time difference	
	Combined	Tree
1	1231744	1434417
2	554354	630013
3	1010974	1084331
4	551782	1019036
5	606000	1205460
6	791768	722783
7	649328	792294
8	486326	389697
9	773889	1052249
10	804042	922640
Average	**746020,7**	**925292**
Standard deviation	**219823,7**	**286471,4**

Table 7: The task completion time of the combined interface (tree with detailed view) and the tree interface

The results show that the users got better scores when using the combined interface over the simple tree interface.

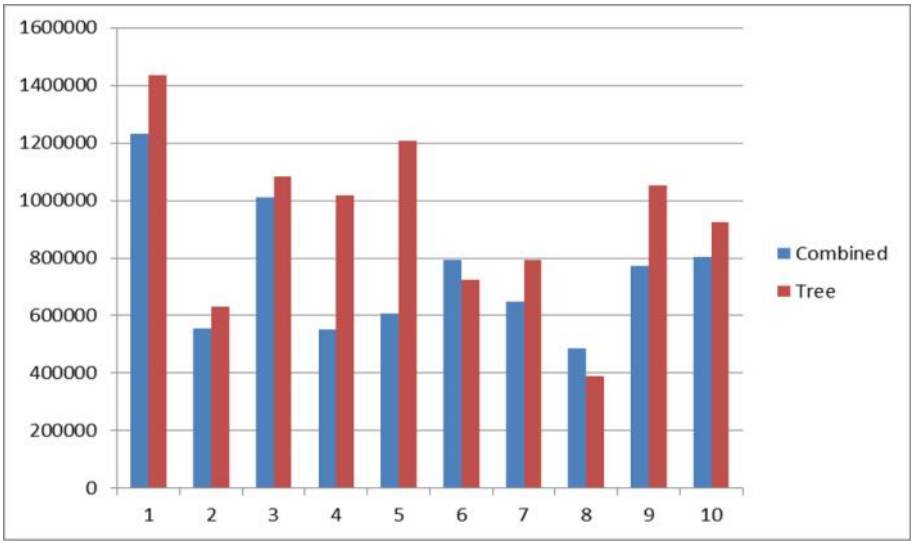


Figure 24: The total task completion time for the tree and combined interfaces of each user

Although all ten users were shown the combined interface first, such that when using the tree interface they were more experienced with the overall movie, only two of them got better scores on the tree interface. This shows that the strengths of the combined interface can overcome the advantage the tree interface gains from being shown second and allow the users to get better scores.

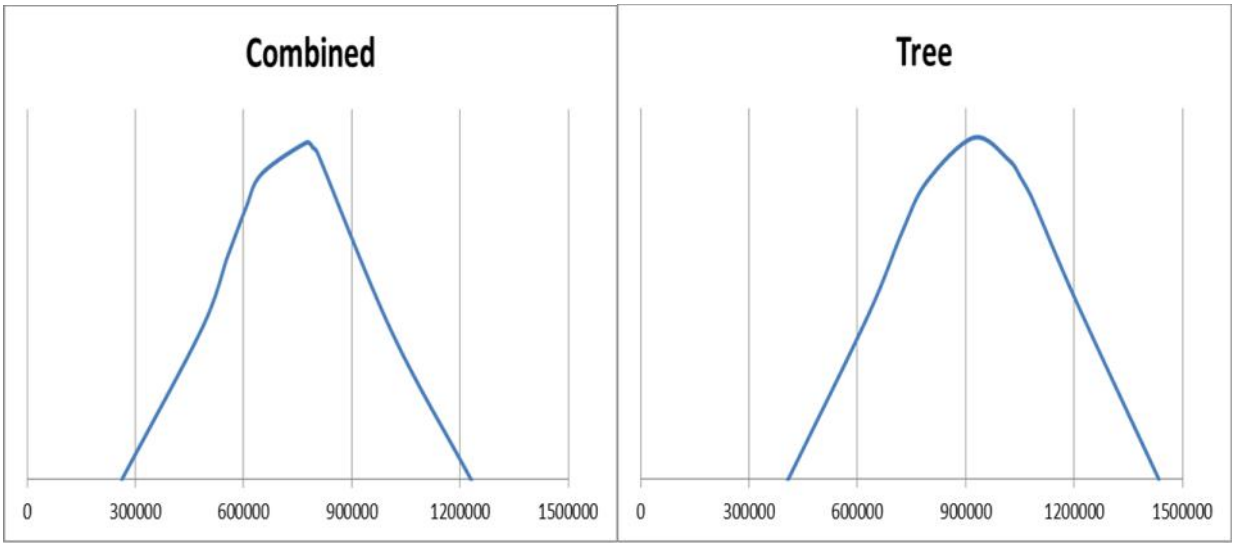


Figure 25: The standard distribution of the total time for the combined (left) and the tree (right) interface

The normal distribution of the combined interface is centered around the value 746 seconds and 77% of all the values for the combined interface are smaller than the center point of the tree interface (925 seconds). To confirm the results we conduct a Paired T-Test on the hypothesis that the tree scored significantly worse than the combined interface (H_a: $D_{combined-tree}>0$). The expected significance level is 95%. For the 9 degrees of freedom provided by the ten testers, we obtained a T-Value of 2.719 and a P-Value of 0.98. As the probability of the hypothesis is higher than the expected 95%, we accept it as valid and we conclude that the better score obtained by the combined interface did not occur by chance.

As we detailed in the design space section, the general part of the combined interface – where a user can view multiple levels of granularity at the same time – is designed to work together with the detailed view. We would like to test if the good results obtained by the combined interface are due to the addition of the detailed view, or if the general part of the interface can stand on its own. In order to do this we choose to disable the option to access the detailed view and compare the general part of the combined interface with the tree interface, in a new set of tests with different participants.

User	Time difference	
	Combined	Tree
1	1144239	1579472
2	1328993	1041518
3	806011	973628
4	396096	430858
5	626756	928861
6	475544	563748
7	885491	692748
8	461207	583114
9	613495	678455
10	675424	763599
11	855832	864828
12	436031	569297
Average	**725426,5**	**805843,8**
Standard deviation	**279589,4**	**293881,7**

Table 8: The task completion time of the combined interface
-after disabling the detailed view - and the tree interface

The “combined” interface obtained better task times than the tree interface even after we disabled the detailed view, but the difference between the two is smaller than in the previous case (80 seconds compared to 179 seconds). This shows that the detailed view potentially has a positive effect on the average search time of a task.

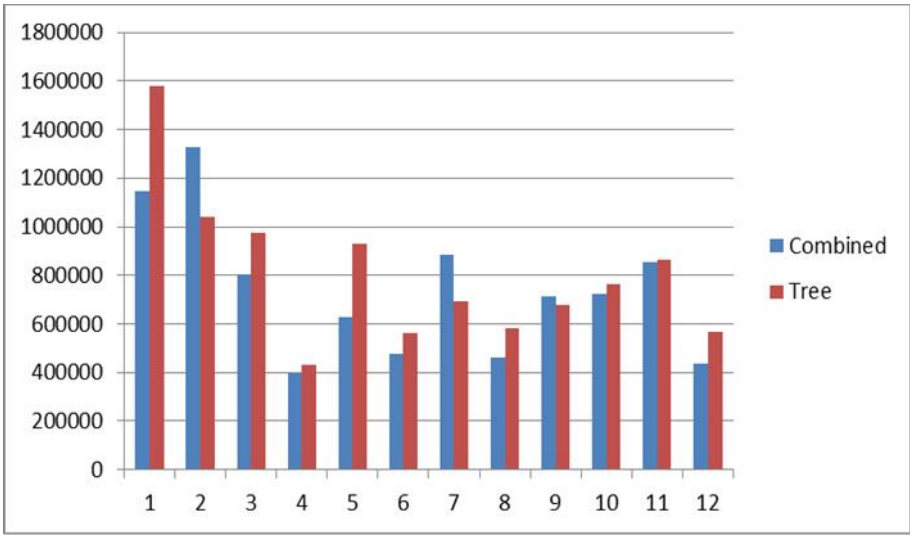


Figure 26: The task completion time of each user for both the tree interface and the combined with tilting disabled

Only two users got better scores when using the tree interface, out of which, one used the combined interface first (7) and one used it second (2). This seems to suggest that the order of the interfaces does not have such a high impact on the performance of the users.

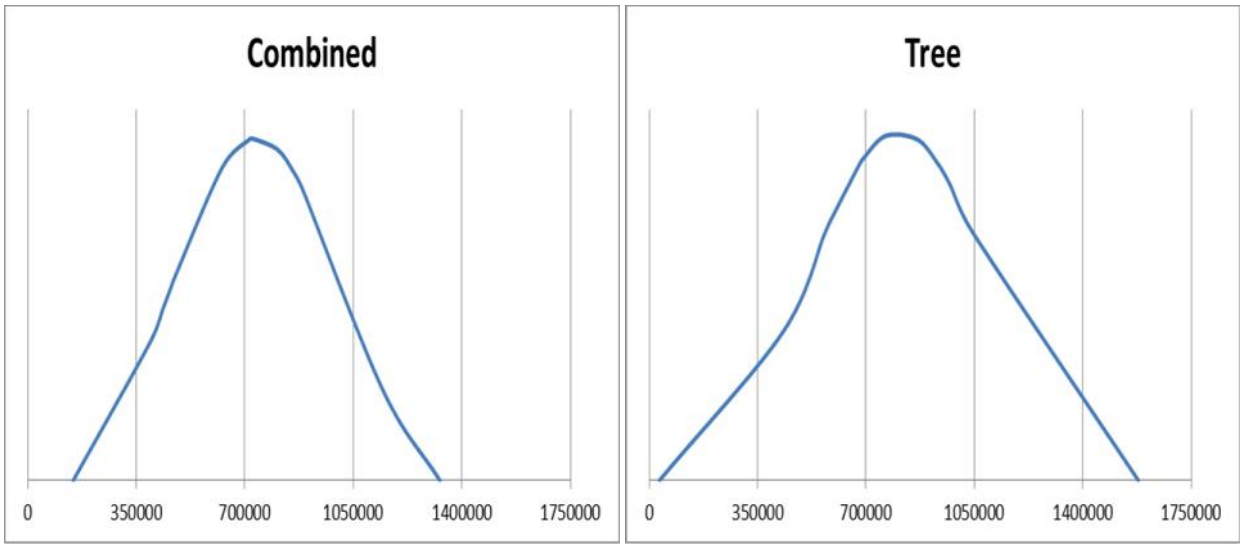


Figure 27: The standard distribution of the total time for the combined without tilting (left) and the tree (right) interface

The small standard deviation of the combined interface places a lot of the values of this interface inside the normal distribution of the tree interface, which has a larger deviation.

Like in the previous case, we conduct a Paired T-Test to test whether our result that the combined interface scored better than the tree interface (H_a: $D_{combined-tree}>0$) due to chance or is it a statistically significant observation, with an expected significance level of 95%. The test yields a T-Value of 1.2611, that translates in a P-Value of 0.88 with the 11 degrees of freedom given by the numbers of testers. The 88% probability of the alternate hypothesis falls below the 95% margin, so we accept the null hypothesis that the two results are not statistically significant.

- Errors – the number of selections made by the users that fall outside the target movie.

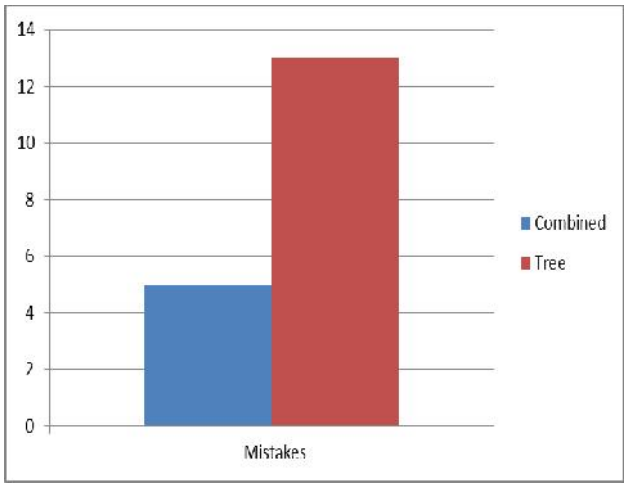


Figure 28: The number of mistakes registered by the tree interface combined with the detailed view and the simple tree interface

As we can see in the figure above, the tree interface registered more than the double of the number of mistakes registered by the combined interface. The interface which had access to an in-depth view over one particular level of granularity allowed less error prone selections, re-confirming the result registered in the second study.

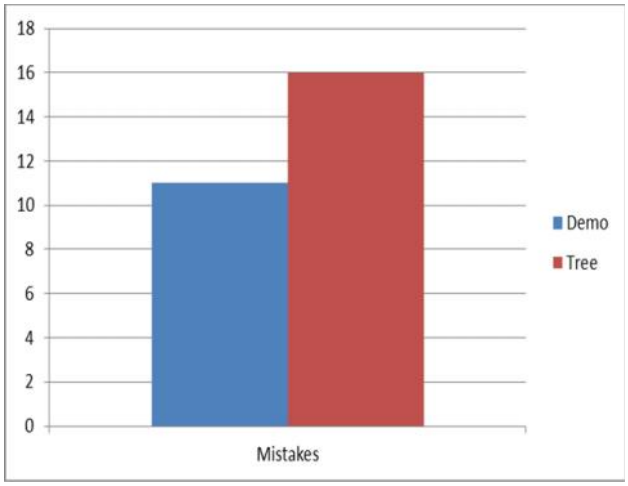


Figure 29: The number of mistakes registered by the tree interface with the detailed view disabled and the simple tree interface

While the unchanged tree interface got a number of mistakes comparable to the first part of the study, the demo with the detailed view disabled got a larger number of mistakes. This goes to show that the detailed grid view has a vital part in the precision of selection.

- Tilting – how many times each user tilted the device in order to gain access to the detailed view

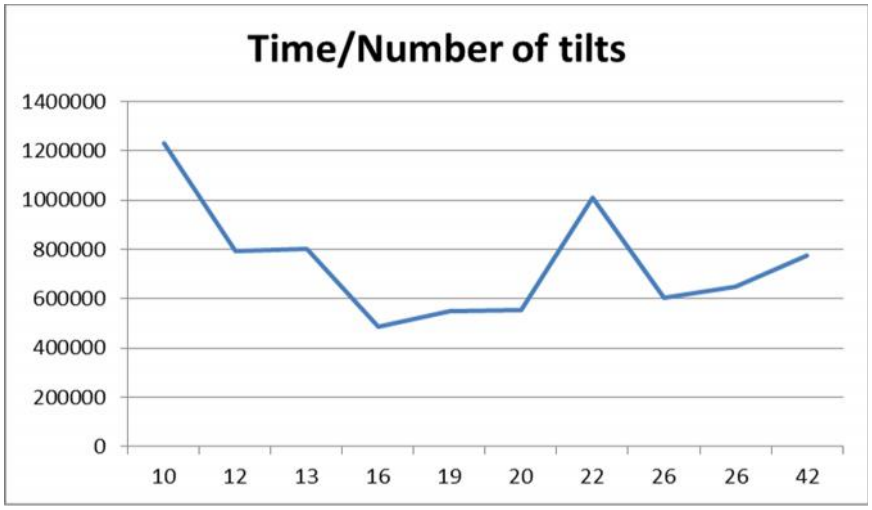


Figure 30: The task completion time/number of detailed view uses

As the graph shows, the overall task completion time dropped as the number of tilts increased inside the range of 16-20, and begun to rise again as the number of tilts go beyond that range. This seems to indicate that an ideal usage of the detailed view is within the interval of 2-3 uses per task. These results need to be verified by performing a statistical significance analysis, unfortunately the data gathered is insufficient to confirm our assumption.

5.3.2 QUESTIONNAIRE

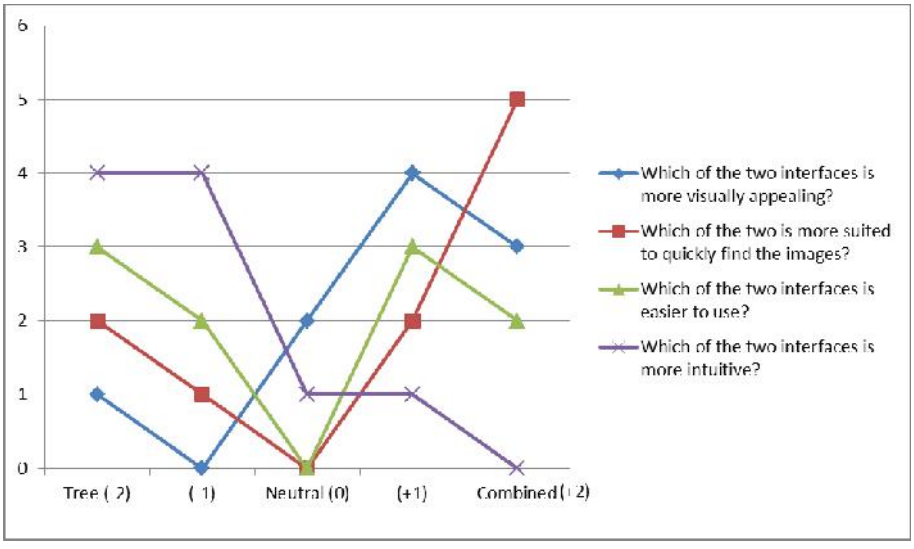


Figure 31: The results of the questionnaire when comparing the
combined tree interface with the detailed grid view and the simple tree interface

The results of the questionnaire show that the users selected the combined interface as the more visually appealing and useful for the search tasks. Although the tree was clearly chosen as the interface that is more intuitive and easier to understand, the interfaces were equally voted as the one easier to use. The users that voted the combined interface as easier to use, noted that although the interface starts slowly due to its complexity, but once they got used to the it, its utility made it easy to use for the given tasks.

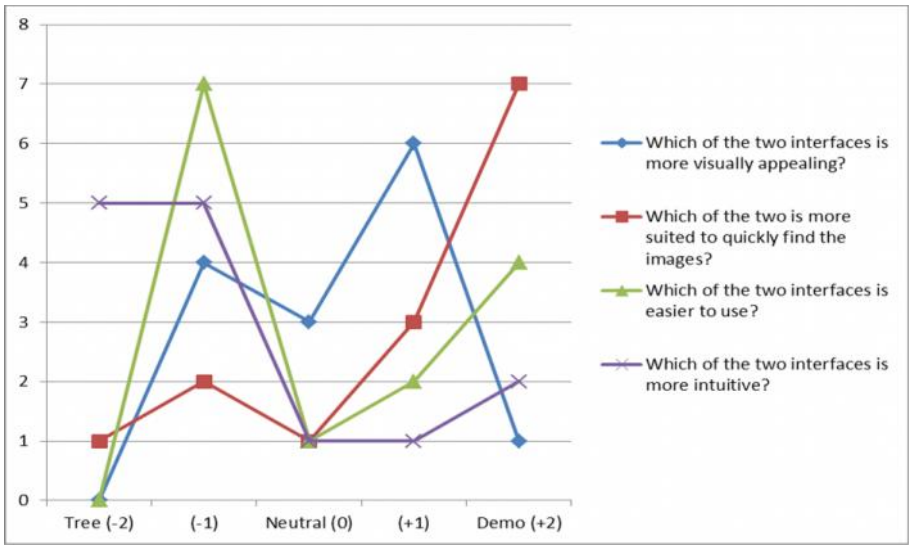


Figure 32: The results of the questionnaire when comparing the combined tree interface
with the detailed grid view disabled and the simple tree interface

Even with the detailed view disabled, the combined interface was viewed as a non-intuitive interface when compared to the tree, but was once more selected as the most efficient and visually appealing interface. The feedback on the ease of use shown once more that despite the initial disadvantage of the combined interface, the two are just as easy to use.

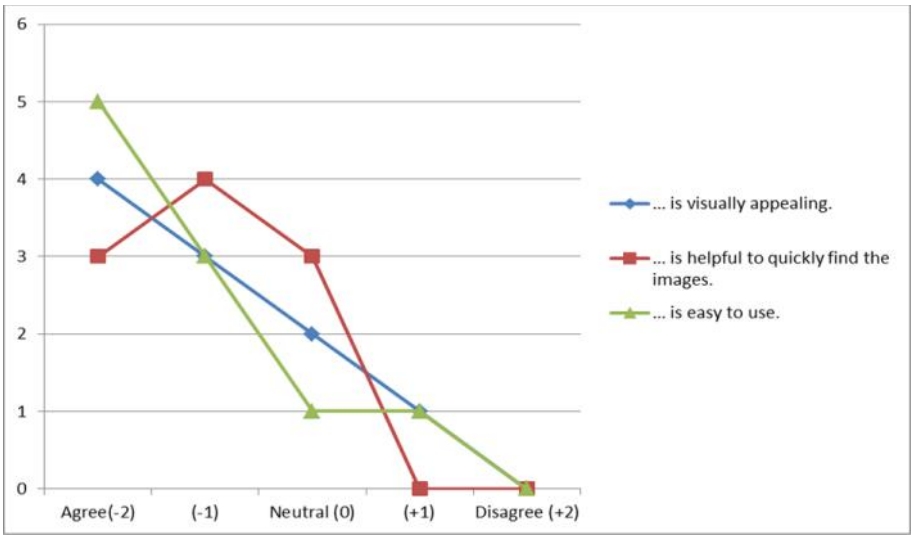


Figure 33: The user's feedback on the detailed view

The detailed view got positive feedback on all three aspects – visually appealing, helpful to find images and ease of use – from the users, showing the added utility that this view adds over the normal tree-like interface.

5.4 CONCLUSION

The results prove our assumptions that, by combining the benefits of the grid and tree in a single interface, it would yield better results than each interface on its own. The combined interface obtained better scores in both task completion time and number of mistakes.

As we disabled the detailed grid view over one level of granularity, in the combined interface, we observed that, even if it still gets better results than the tree, these results were not statistically significant, as it was the case with the combined interface. This proves that the union of the two interfaces is the reason why the combined interface scored better.

One important observation that we made during the study was that the combined interface, although conceived as unintuitive by the users, still managed to score better than the simple tree interface. A longer study can show if the results will improve even further as the participants get more used to the interface.

One of the participants used only the detailed view to complete the tasks. This leads us to believe that the view in discussion can be used a single interface, rather than just an extension to others. We are also interested to find out if this extension offers the same benefits to the regular tree interface.

6. CONCLUSION AND FUTURE WORK

6.1 CONCLUSION AND SUMMARY

Over the last few years, we witnessed a fast growth in the domain of mobile computing – leading to the modern smartphones and tablets. These devices come equipped with touchscreens and a number of sensors that allow a more natural interaction with such apparatus. However, as these contraptions are mobile in nature, they come with less computing power and smaller screen sizes than traditional computers.

Knowing these differences, interface designers must adapt their applications to be better suited for these devices. Such is also the case of media display programs such as image galleries, document viewers or movie browsers. The simplest methods of showing media data is by presenting it in grid-like interfaces called storyboard presentations.

The goal of this thesis was to study the characteristics of such interfaces. To do so, we have chosen to conduct a series of user studies on hierarchical video storyboard interfaces – we have selected video data for our test due to their active nature.

As mentioned before, smartphones allow different means of interaction when compared to traditional computers, but most hierarchical media browsers use a tap gesture for navigation within the interface. As the tap gesture is very similar in nature with the mouse click (except it is performed with the finger) we conducted a test to determine if other gestures that can be performed on touchscreens can be used as a replacement for the tap. We selected the pinch as the gesture to be used for navigation between different levels of granularity on our storyboard video browser interface and compared it with an interface that used the tap as its means of interaction.

Although the pinch gesture did not obtained the results we were expecting, due to its theoretical advantages over the tap, we made an important observation derived from the user questionnaire: using the tap is much more enjoyable for the participants, because the pinch is a difficult gesture to perform. This leads us to believe that a simpler multi-touch gesture with the same advantages the pinch offered, it would score significantly better than the click.

A conclusion we can draw from the results is that users prefer to use a gesture to get back to previous levels of granularity, instead of pressing a button, as the latter breaks the natural flow of their actions. This was clearly demonstrated as the back button was used only 12 times out of 1560 actions of zooming back to a previous level.

In a second study, we tested two different approaches to the implementation of a hierarchical storyboard video browser interface – one that shows one level of the hierarchy on a grid on the whole screen, while the other has a smaller number of pictures in each level, but multiple hierarchy planes are shown on the screen at once.

The results show that the multi-level interface scored better in terms of task completion time, while the single-level interface got a smaller number of mistakes. Although these results were not significant from a statistic standpoint, we combined the information gathered with user feedback,

and estimated that the multi-level interface works faster in finding a required part of the movie, but the single level is more useful in quickly finding an exact frame, due to the larger exploration size.

These observations formed the basis for our third study, where we designed an interface that would combine the single-level and multi-levels models of the hierarchy browser. As using both models on the screen at the same time is an impossible task, we choose to show the multi-level interface on the main screen, while the detailed view over a single level is shown at a different position in the virtual space, position that can be reached by tilting the device, in a similar fashion to the shoebox virtual reality.

We also designed a method of interaction based on the findings of our first study. This interaction scheme must be simple and intuitive and provide an easy way to go back to previous levels – as you drag your finger towards the top of the screen up to go to a more detailed hierarchy, while dragging the finger towards the bottom of the screen would show a previous granularity.

We then compared this interface with the regular multi-level interface. The results have shown that the combined browser scores better both in search times and number of mistakes than the regular multi-level one. The users also selected the combined interface as their favorite out of the two, although they described it as less intuitive than the multi-level one.

As we wanted to test whether the combination of the multi-level and single-level models was indeed the factor favoring the positive results, we disabled the detailed view and compared the version of the multi-level interface with independent levels with the version used in the combined interface – with interconnected levels. This time, the difference between the two was statistically insignificant, proving that our design that combines the benefits of two models of hierarchical interfaces provides a fast and efficient way of browsing through video content.

6.2 FUTURE WORK

Although the designed interface was proven to be an efficient implementation of a hierarchical video browsing interface, we can still further examine and improve its characteristics. First, we would like to test how the addition of a single-level detailed view to a state of the art multi-level interface with independent levels, would influence it. This would show us if the connection between varying levels of granularity has any influence on the combined interface.

Another study topic that would originate from our findings is whether the detailed view can stand on its own as an interface or if it works only as a complement to a tree-like interface. This can be tested against a similar, stand-alone interface - like the grid.

In the implementation of our interface, we placed the detailed view in the simplest position possible – directly beneath the multi-level view on the virtual space. We would also like to test which position works best for the detailed view inside the 3D space – example: at an angle beneath the multi-level view, on the left or right instead of the bottom, etc. A better positioning of the detailed view may be a potential solution to one of the major complaints about the combined interface – it is unintuitive and requires a period of accommodation in order to be used efficiently.

7. ACKNOWLEDGMENTS

I would like to thank dr. W. O. Hürst for the support, advice and feedback he provided during the production of this project. I would also like to thank all the people for their participation in the user studies.

8. BIBLIOGRAPHY

[1] Marco Furini ,Filippo Geraci , Manuela Montangero , Marco Pellegrini, "**VISTO: visual storyboard for web video browsing**", Proceedings of the 6th ACM international conference on Image and video retrieval, p.635-642, July 09-11, 2007, Amsterdam, The Netherlands

[2] Zhang, H. J., Low, C. Y., Smoliar, S. W., Wu, J. H., 1995. "**Video parsing, retrieval and browsing: an integrated and content-based solution**". Proceedings of the international ACM conference on Multimedia (New York, NY, USA, 1995), 15-24.

[3] Maël Guillemot, Pierre Wellner, Daniel Gatica-Perez, and Jean-Marc Odobez, "**A Hierarchical Keyframe User Interface for Browsing Video over the Internet**". INTERACT, IOS PRESS, (2003)

[4] W. Hürst and P. Jarvers, "**Interactive, Dynamic Video Browsing with the ZoomSlider Interface**", in Proc. IEEE Intl. Conf. Multimedia and Expo, pp. 558–561, IEEE, Amsterdam, The Netherlands (2005).

[5] Schöffmann, K., Hopfgartner, F., Marques, O. Boeszoermenyi, L. and Jose, J.M., 2010. "**Video browsing interfaces and applications: a review".** SPIE Reviews, Vol. 1, No. 1, 1-35 (018004), SPIE

[6]M. Jansen, W. Heeren, and B. van Dijk, "**Video trees: Improving video surrogate presentation using hierarchy**", in Intl. Workshop Content-Based Multimedia Indexing, pp. 560–567 (2008).

[7] Wolfgang Hürst, Dimitri Darzentas, "**HiStory – A Hierarchical Storyboard Interface Design for Video Browsing on Mobile Devices**", MUM '12 Proceedings of the 11th International Conference on Mobile and Ubiquitous Multimedia, December 4, 2012

[8] M. del Fabro, K. Schoefmann, and L. Boszormenyi. "**Instant video browsing: A tool for fast non-sequential hierarchical video browsing**". In G. Leitner, M. Hitz, and A. Holzinger, editors, HCI in Work and Learning, Life and Leisure, volume 6389 of Lecture Notes in Computer Science, pages 443-446. Springer Berlin/Heidelberg, 2010.

[9]Del Fabro, M., Böszörmenyi, L.: "**AAU Video Browser: Non-Sequential Hierarchical Video Browsing without Content Analysis**". In: Schoeffmann, K., Merialdo, B., Hauptmann, A.G., Ngo, C.-W., Andreopoulos, Y., Breiteneder, C. (eds.) MMM 2012. LNCS, vol. 7131, pp. 639–641. Springer, Heidelberg (2012)

[10] Manske, K. and Mühlhäuser, M. "**OBVI: Hierarchical 3D Video-Browsing**". Proceedings of ACM Multimedia, (1998).

[11] Klaus Schoeffmann, Manfred del Fabro, "**Hierarchical video browsing with a 3D carousel**", Proceedings of the 19th ACM international conference on Multimedia, November 28-December 01, 2011, Scottsdale, Arizona, USA

[12] Google Developers. **Using Touch Gestures**. Website. Retrieved: 2013-11-10. Available at: http://developer.android.com/training/gestures/index.html

[13] Microsoft® Developer Center. **Windows Touch**. Website. Retrieved: 2013-11-10. Available at: http://msdn.microsoft.com/en-us/library/dd562197.aspx

[14] Lao S., Heng, X., Zhang G., Ling Y. and Wang P. "**A gestural interaction design model for multi-touch displays**". In Proceedings of the 2009 British Computer Society Conference on Human-Computer interaction (Cambridge, United Kingdom, September 01 - 05,2009). British Computer Society, Swinton, UK, 440-446.

[15] Dietrich Kammer , Jan Wojdziak , Mandy Keck , Rainer Groh , Severin Taranko, "**Towards a formalization of multi-touch gestures"**, ACM International Conference on Interactive Tabletops and Surfaces, November 07-10, 2010, Saarbrücken, Germany

[16] Hürst, W., Snoek, C.G.M., Spoel, W.-J. and Tomin, M.,2011. "**Size matters! How thumbnail number, size and motion influence mobile video retrieval**". Proceedings of the 17th international conference in Advances in multimedia modeling - Volume Part II (Berlin, Heidelberg, 2011), 230–240.

[17] Hürst, W. and Darzentas, D., 2012. "**Quantity versus quality: the role of layout and interaction complexity in thumbnailbased video retrieval interfaces**". Proceedings of the 2nd ACM International Conference on Multimedia Retrieval (New York, NY, USA, 2012), 45:1–45:8.

[18] W. Plant and G. Schaefer, **"Visualising image databases**", in IEEE Int. Workshop on Multimedia Signal Processing, 2009, pp. 1-6.

[19] Klaus Schoeffmann, David Ahlstrom, and Laszlo Boszormenyi, **"A user study of visual search performance of interactive 2d and 3d storyboards,"** in Proceedings of the 9th International Workshop on Adaptive Multimedia Retrieval , 2011.

[20] Klaus Schoemann, David Ahlstrom, **"Using a 3D Cylindrical Interface for Image Browsing to Improve Visual Search Performance**", in the Proceedings of WIAMIS, IEEE 2012

[21] Martin de Jong, BSc, Supervisor: Dr. Wolfgang Hürst, **"Influence of the Shoebox Model and its parameters on the 3d Perception and Experience of Users**", Master Thesis, 2012

[22] Mathijs T. Lagerberg, BSc, Supervisor: Dr. Wolfgang Hürst "**A proposal for enhancing mobile interfaces using a 3D effect and an investigation on its effectiveness**", Master Thesis, August 15, 2012

[23] Steven Weijden, Supervisor: Dr. Wolfgang Hürst, "**3D video browsing using the shoebox visualization. The value of tilting with respect to the performance and entertainment value**", Small Project, September 24, 2012