

Timothy E. Obert. Combining User Tagging and Automatically Generated Storyboards for Digital Video Retrieval: An Exploratory Study. A Master's Paper for the M.S. in L.S degree. April, 2010. 45 pages. Advisor: Gary Marchionini

This is an exploratory study investigating the concept of combining social and automated digital video surrogates with the intention of improving user understanding and retrieval of video. Videos were shown to participants who were then asked to tag the videos according to two systems. The control group tagged videos in the traditional manner of tagging the whole video in one input form. The experimental group tagged a storyboard with multiple input forms corresponding to the storyboard.

The findings of this study suggest that the implementation of a system that combines automatically generated storyboards with social tagging will significantly increase the number of tags that each user will add to a video. In all aspects of the results, there was significant increase in the amount of tags created in the storyboard system when compared to the traditional system of tagging an entire video. The results suggest that it is worth investigating in future work.

Headings:

Digital Video

Indexing – Video recordings

Information Retrieval

Information Systems – Special subjects – Video recordings

Internet Video

Subject Access

Video Surrogates

COMBINING USER TAGGING AND AUTOMATICALLY GENERATED
STORYBOARDS FOR IMPROVED DIGITAL VIDEO RETRIEVAL:
AN EXPLORATORY STUDY

by
Timothy E. Obert

A Master's paper submitted to the faculty
of the School of Information and Library Science
of the University of North Carolina at Chapel Hill
in partial fulfillment of the requirements
for the degree of Master of Science in
Library Science.

Chapel Hill, North Carolina

April 2010

Approved by

Dr. Gary Marchionini

Contents

1. INTRODUCTION.....	4
2. LITERATURE REVIEW	7
A. SEARCHING BEHAVIOR FOR DIGITAL VIDEO	8
B. AUTOMATIC AND NON-USER GENERATED SURROGATES.....	8
C. USER-GENERATED SURROGATES AND TAGGING	13
D. SUMMARY	18
3. METHODOLOGY	19
A. SELECTION AND PREPARATION OF VIDEOS	21
B. PARTICIPANTS	22
C. SUMMARY	23
4. RESULTS	24
A. CONTROL GROUP (USERS TAGGING THE ENTIRE VIDEO).....	24
B. EXPERIMENTAL GROUP (USERS TAGGING THE STORYBOARD)	25
5. DISCUSSION AND ANALYSIS	28
6. LIMITATIONS AND FUTURE WORK.....	30
7. CONCLUSION	32
8. BIBLIOGRAPHY	33
9. APPENDICES.....	36

1. Introduction:

According to *ComScore* (2009), nearly 140 million people viewed more than 200 minutes of online, digital video in the United States in January of 2009. Each person's 200 minutes of video comes at a significant cost of time and bandwidth. Multimedia and, especially, video consist of multiple channels (image, audio, text) streamed to a user; therefore, the amount of data being sent to a user is large when compared to text or most image files (Christel, 2009). Video is also temporal and continuous, which forces users to watch it on a timeline (Christel, 2009). Because of the nature of video, a significant amount of time, besides that spent watching, is spent searching for relevant videos because the current methods used for video retrieval are limited. Therefore, providing users with the gist of a video before they spend resources downloading an irrelevant video is worthy of research. Many of the current methods of video browsing and summarization do not provide the user with enough information to adequately understand the gist of a particular video. Searching for a particular object becomes especially problematic for users when a video is too long to view quickly and the metadata does not adequately summarize the video.

Research can be divided into two camps in regards to video summarization techniques: automated metadata and user generated metadata. Christel (2009) shows that there has been a significant amount of research in areas of automated metadata creation where various forms of computer-generated video surrogates like keyframes, storyboards, fast forwards, speech and text recognition, and keywords have shown to be effective

means of improving video retrieval by end users. Keyframes are images that are meant to represent the entire video. Fast forwards are typically an entire video that has been compressed into a very short amount of time (Wildemuth, 2003). Speech and text recognition software allow a program to gather text from a video that will be indexed to improve search. Storyboards are a group of keyframes collected from a video along a timeline.

On the other end of the spectrum lies manually generated metadata such as user-created keywords or social tagging. Describing content with keywords and descriptive terms is the typical way to provide added meaning to searchable content. Traditionally, these terms are created using a controlled vocabulary by professionals or authorities such as librarians, but the development of collaborative tagging has opened up this process to the general public. With real world examples, such as YouTube, using user-generated tags to add meaningful metadata to digital video is one of the most widely used summarization techniques (Geisler, 2007). Golder and Huberman (2005) define collaborative tagging as "the practice of allowing anyone – especially consumers – to freely attach keywords or tags to content." Tagging can be an effective means of providing meaning when searching for digital video partly because it is a lightweight mechanism that can be applied to existing systems or digital libraries (Melenhorst, Grootveld, Van Setten, and Veenstra, 2008; Treude and Storey, 2009).

Despite a large amount of research, a gap still exists between the implementation of metadata and surrogates for the summarization of digital video and the user's understanding or gisting of the general idea of the video before viewing it. Combining some of these summarization tools may increase the end-user's understanding even

further (Christel, 2009). Combining surrogates to increase user gisting is an area that will require more refinement and experimentation. A possible combination that merits exploration is the use of automatically generated storyboards with social tagging.

Storyboards are multimedia surrogates, meaning they are not text based but image based. Media-based surrogates are useful for providing visual overviews and description of video that can be richer than text alone. By showing thumbnails of a video chronologically, storyboards can contribute to a user's understanding of the narrative of a video. The narrative is an important concept for the user to understand while he or she is determining the value of a video. The narrative provides a way for the user to organize and make sense of the information (Wilkins, Hughes, Wildemuth, and Marchionini, 2003). Storyboards are created by automatically displaying every *i*th frame of a video or by automatically selecting a frame from the middle of every sequence in a video. By displaying a sequence of thumbnails from a video in chronological order, the user can get a brief glimpse into the contents of a video from beginning to end in a very short amount of time, much like a table of contents (Christel, 2009). Combining this technique with collaborative tagging to create a system where the user has the opportunity to tag not just the entire video but to tag each thumbnail in the storyboard could provide an increase in user gisting. Tagging systems are traditionally simple to implement (Treude and Storey, 2009). Input forms could easily be attached to each thumbnail in a storyboard to develop this system.

One question that has not been researched yet is: Would the development of a system of combining social tagging with automatically generated storyboards improve video retrieval as opposed to storyboards alone or tagging alone? User generated tagging

may be implemented with storyboards to create a vivid overview of the content of specific parts of a video. Social tagging or ethnographic classification provides an informal way of providing efficient descriptive metadata to objects (Melonhorst et al., 2008). Each thumbnail created for the storyboard becomes an object that could be tagged, and since every thumbnail represents a specific chunk of time in the video, the tags could be used to describe not only the image itself but any aspect of the video for the amount of time that each storyboard frame represents. The tagged storyboard may provide several functions of summarizing meaning in a video: description of the thumbnail itself, description of action or content at a specific time in the video, and technical description of different channels in the video (audio and visual). The interface of tagging a storyboard may yield more tags than the typical system since the users will have more opportunities to describe the video.

The research questions:

- Would tagging thumbnails arranged into a storyboard be more effective than tagging the video as a whole?
- Would tagging a storyboard generate more tags (thus more searchable keywords) than the traditional way of tagging a video?
- Would there be significant or relevant differences between the tagging of an entire video and a video that has been chunked into a storyboard?

Hypothesis: The underlying hypothesis is that the synthesized use of storyboards and social tagging may provide a valuable and efficient way to ascertain the context and content of a video by describing chunks of video visually and textually and that tagging a

storyboard will yield more tags than the traditional method of tagging a video where the user tags the video as a whole.

2. Literature Review:

Since O'Connor (1985) proposed the need for a surrogate developed specifically for digital video due to the inherent nature of video as moving pictures, researchers have been trying to improve video retrieval by enhancing and developing new summarization techniques. Much of the research in this area has been divided into two areas: those working with automatically created surrogates and those working with user-generated surrogates, such as tagging. The focus of this section is to describe the two sides and suggest a point of merger for the improvement of video retrieval. This point of merger is the synthesis of user created or socially created tags with automatically created storyboards.

The aim of this literature review is to identify the current and major articles that relate to the preceding the research questions. This review is split into three sections: searching behavior, non-user created surrogates, and user created surrogates. Each article was selected because of relevance, importance, and currency. The works included below should reveal the gap in the current state of video retrieval and suggest that the marriage of tagging and storyboards could lead to a viable improvement in video searching.

A. Searching behavior for digital video:

Cunningham and Nichols (2008) researched how users search and browse for videos. They collected data from 27 “autoethnographies” and 27 ethnographies that were collected by the first 27 participants. A total of 98 search session summaries were gathered. The findings suggest that most users browse and view videos rather than manually constructing a query or reading the text associated with a video. The text includes metadata. Therefore titles, comments, and tags may be largely ignored by users. Since browsing typically included viewing the poster frame or the single thumbnail that represents a particular video, most users, in this study, based their video retrieval process on a single image of a particular video.

This study has several limitations; the most obvious being that since the ethnographies are qualitative in nature the results are limited to the participants and cannot be easily generalized to a wider population. However, this study is one of very few dealing with the behaviors of users searching specifically for video. One important implication of the study is that despite participants' reliance on a single, static thumbnail, the researchers came to the conclusion that because of average search time and the amount of steps it takes to find a video, a single thumbnail does not often provide an adequate summarization of a video. Therefore, in addition to thumbnails, more gisting techniques need to be implemented in video search systems.

B. Automatic and non-user generated surrogates:

One problem with much of the automatically and non user-generated metadata or surrogates is the usefulness of systems that implement these techniques in broad,

heterogeneous contexts. A significant amount of the research in this area is from Carnegie Mellon and TRECVID. Nearly all of the studies based around TRECVID used news footage (Christel, 2008). Many of the techniques developed like ASR (automatic speech recognition), facial recognition, and VOICR (video optical character recognition) are proven to be useful for news footage, but these techniques are not tested on user generated content, and the nature of many of the automatic techniques like ASR is that they need to be trained according to the context of the videos they are being used on (Christel, 2008; Christel, Yan, 2007). News footage provides a standardized and homogenous framework to train automated techniques; however, a video on YouTube can range from a static shot of an individual talking into the camera to a movie trailer with many cuts and shots, making the automated surrogates and summarization techniques less accurate (Christel, 2008). The bulk of this portion of the review will deal with automatically and non-user created surrogates that are most pertinent to this study.

Christel and Moraveji (2004) developed a system using storyboard views for video retrieval. A storyboard consists of thumbnails of individual shots of a video displayed sequentially. Christel and Moraveji create their storyboards by automatically selecting every *i*th frame of a video and then running algorithms on the selected frames to reduce redundancy. In a study to test the effectiveness of their system on novice users, the researchers tested the search performance of 12 undergraduate and graduate students that were deemed as novices. Half of the participants used a system with storyboards, and the other half used a system without storyboards. Performance was based on the users' length of search by having them find a series of individual shots within predetermined videos and recording the time of each search. The results indicated that

their system for incorporating storyboards into video retrieval improved the effectiveness of the novice searchers. Christel and Moraveji's study implies that storyboards provide system designers with a visual way of informing users of contextual information and allow users to understand specific content of a video in a deep way. Christel and Moraveji's work shows the usefulness of using storyboards in video retrieval; however, because of the nature of the study, there is little focus on the effectiveness of storyboards providing the overall gist of a video, which is an important goal to strive for in video retrieval. This study deals with searching for specific shots within a video not the overall content.

Expanding on the previous study, Christel (2008) takes on the task of video retrieval based more on story and context rather than on searching for shots or specific content. This article provides an overview of the work that he has been involved with on the TRECVID and Informedia projects. In his paper, he examines the use of storyboards in this role, determining that the storyboards are excellent for use in videos that rely largely on visual content and for videos that contain a large amount of shots; however, storyboards begin to fail when a video is static in terms of shots and when the audio track of a video contains useful information that would otherwise be missing from a series of thumbnails. Christel's work ignores the use of social tagging in retrieving digital video and focuses entirely on automatically created metadata. The result of much of this work is inconclusive because of varying results when conducting usability studies. However, Christel suggests that despite the effectiveness of storyboards in video retrieval, there is clear room for improvement by adding other retrieval techniques to the storyboard format.

In a study by Yang, Wildemuth, and Marchionini (2004), the effectiveness of content-based retrieval systems and concept-based retrieval systems was tested to understand whether combining certain features, in this case, text and visual, would improve searching. The researchers used three systems: text only, features only, and a combined system. They then had 36 users search for specific topics. The users were divided among the three systems. The results were based on precision and recall of the participants searches. The original hypothesis was that the combined system would produce more effective searches; however, the results indicated that there is little to no difference between the results of the text-only system and the combined system, and the features-only system had the poorest performance. These results indicated that the content-based system (the features-only system) performs worse than the concept-based systems (text and combined text systems). This has implications that extend from Christel's research and that are relevant to future research. Since concept-based searching may provide more productive results than content-based searching, systems like storyboards, which are usually content-based need to be supplemented by concept-based systems. Concept-based systems could incorporate social tagging, giving the user more of a general idea of the overall story of a video than just the content.

Wildemuth et al (2003) administered a study that sought to fine-tune a new surrogate specific to digital video. This system created a fast forward of a video. The hypothesis of this study was that users could extract a significant amount of knowledge by viewing a short, fast forward of video, and that different speeds will affect the user's ability to recognize objects and actions in a video as well comprehend the overall gist of a video. Forty-five users were asked to use four different systems that provided variations

on the speed of a fast forwarded video. The fast forwards were created by selecting every n th frame of a video. This method is identical to that method used to create storyboards with the exception that the fast forward allows the user to view many more images or frames. Wildemuth et al concluded that a default speed of 1:64 of the original video should be used to create fast forwards. This speed provides the largest amount of user gisting in the shortest amount of time. She also recommended that the fast forwards should also incorporate a system of user-controlled variable speeds. Different content may call for different speeds.

Marchionini, Song, and Farrell conducted a user study to gauge the effectiveness of using audio surrogates for digital video summarization (2009). The audio surrogates were used both exclusively and in combination with fast forwards for the purpose of answering several questions:

- Do automatically generated descriptions and keyword surrogates approach the effectiveness of manually generated ones?
- How do spoken descriptions and spoken keywords compare for gisting tasks?
- How do fast forward surrogates compare with the spoken surrogates?
- What are the effects of combining fast forwards and spoken surrogates?

Marchionini et al (2007) address these questions by conducting a study of 48 participants. Twenty-five similar videos were selected from an educational series made by NASA. The researchers created five surrogate tests: fast forwards alone, spoken description alone, spoken keywords alone, fast forwards combined with spoken descriptions, and fast forwards combined with spoken keywords. The participants were

then asked to perform several controlled gisting tasks by directly experiencing the surrogates, not the actual video, in order to see which of the five conditions provided the best results. The study concluded that, when used alone, both manually generated and automatically generated spoken descriptions and keywords are nearly as effective as when combined with the visual surrogate, fast forward (Marchionini et al., 2009). Though the study suggests that combining surrogates may not increase user comprehension, it provides a template for conducting similar research and shows the relevance of studying video metadata or surrogates. This study also reveals that there is still a gap in the effectiveness of video surrogates, and the researchers recognize that there is a need to investigate new methods or systems for enhancing video retrieval.

It is evident from these articles that automatically created metadata is not, at this time, functional enough to provide effective and efficient video retrieval. Researchers are beginning to combine different forms of surrogates to increase the functionality of retrieval (Marchionini et al., 2009, Christel 2007). In a study comparing video retrieval systems, Christel found that allowing the user to control what kind of results he or she receives leads to better search performance (2007). This study shows that there is still room for more research, specifically in the realm of user controlled search features. The next section will introduce and describe the possible significance of using socially created metadata and summarization techniques in video retrieval.

C. User-generated surrogates and tagging:

Another major area of research in the field of video surrogates focuses on user - created surrogates, such as keywords and, specific to this study, social tagging. Golder

and Huberman (2005) studied the tags created by the users of del.icio.us. They concluded that stable patterns emerge when a significant amount of users of a community are tagging items and experiencing those tagged items. Though, there will always be a small number outliers and variations in user-generated tags. They also determined that the majority of tags are created for the personal use of the tagger and as a result, the usefulness of tags to others is often a byproduct of tagging. Tags are only understood by other users when they can derive meaning from the tags. Therefore successful tags are semantically and ontologically understandable by the user community.

Golder and Huberman also outlined certain advantages and disadvantages of tagging. Tags create more flexibility, facilitate annotation, and facilitate retrieval and discovery. They also suggested that there are possible disadvantages, such as ambiguity of terms that have multiple meanings and variations in syntax and synonymy, which occurs when many terms with different forms describe the same thing. When tags are used by an active community over time the results are not at odds with traditional, controlled description, but tagging provides an example of emergent consensus (Golder and Huberman, 2007; Halpin, Robu, and Shepherd, 2007).

In a similar study, Chi and Mythkowicz (2008) used information theory and del.icio.us to show the usefulness of tagging and to determine whether tagging provides an efficient means to locate information objects or whether tagging provides a “noisy” or disruptive signal hindering the user's ability to locate an information object. The researchers tested this question by developing a web-crawler that gathers user tags and the relevant websites. Using this information, they developed a model to mimic the encoding and retrieval processes. Their findings suggest that there is a threshold in the

amount of tagging that can be productive and that more research needs to be made on finding this threshold.

One problem with this study is that the crawler that the researchers used randomly walked the contents of del.icio.us. Therefore, there is no way to know if certain users or tagging types were over sampled, and there is also no way for the researchers to know the extent of the del.icio.us websites. Thus, there is no way for the researchers to know if their sample is an accurate portrayal of the site's contents. However, this study does give a framework for how tags work and why they are effective. They also advocate the use of tagging in search systems because it is an efficient way for a group to encode objects for later retrieval and that over time a collective vocabulary forms potentially creating an effective tagging system.

In a study that builds upon the work of the previous article, Heckner, Neubauer, and Wolff (2008) study differences of tagging in different settings. They compare tagging systems in four different areas: scientific articles, images, bookmarking sites, and digital video. They found that users portray different tagging behaviors when using different types of content. Data was collected by gathering content from each of the four categories for a specified length of time and then randomly selecting 1000 entries from each of the respective categories. The researchers then compared the data in categories like words per tag and average number of tags. The results indicate that users do tend to tag differently. The researchers suggested that users tagging video tend to use many more words, tag more extensively, and tag for people. Users also tag for content more than tagging for resources, such as type of camera used. Since Heckner, Neubauer, and Wolff suggest that tagging for digital video is primarily based on content and users tend to tag

more extensively, implementing a tagging system as a supplement to using storyboards as surrogates may be a valuable combination for the retrieval of video.

Geisler and Burns (2005) developed a study to determine conventions and strategies of tagging behaviors in the YouTube community. They used data from over 500,000 contributors from over 1,000,000 videos. Over 500,000 tags were found with a median number of 6 tags per video. This study suggests that YouTube users are tagging nearly half of the videos that they encounter. They also concluded that 66% of tags were of words that did not appear in the title of the video, suggesting that the tags are adding value. However, they did find that a significant portion of the tags were deemed non-descriptive. Geisler and Burns concluded that new strategies need to be developed in order to increase the effectiveness of tags in providing descriptive meaning to video.

Melenhorst, Grootveld, Van Setten, and Veenstra (2008) studied the effectiveness of social tagging compared to professionally generated metadata and automated metadata of digital video. To conduct this study, the researchers developed a tagging system for video and allowed 194 participants to produce three types of tags for 15 videos. To test the effectiveness of the tags when compared to other types of metadata, they created five categories to test: searching only professional metadata, searching only user tags, searching in both professional metadata and user tags, searching in derived keywords, and searching in smart tags. They divided a separate group of 153 participants into the five groups and tested their searching. Since the tagging application was designed for the study, the researchers were able to exercise some degree of control over the study. They concluded that tagging is just as effective if not more so than professionally created metadata (Melenhorst et al., 2008). This article informs future studies by describing how

the testing took place and also giving advantages and disadvantages of the methods that they chose to use. This study is significant because it shows the potential power of tagging and the need for implementation of tagging into other systems like automatically created metadata. This study also suggests that a union of tagging and video storyboards as surrogates could be a beneficial area of study, since both areas have been proven to improve searching.

Srinasan and Kukreja (2008), in *Tagboards for video Tagging*, propose a system that combines tagging and keyframes that would typically make up a storyboard. Their system would enable users to tag images not temporally like a typical storyboard but by image similarity. They ignore temporal tagging because a video may contain several similar shots; therefore, the images in a storyboard may repeat, creating redundancy. By grouping images by similarity and then allowing users to tag these clusters, they are able to side step the problem of image repetition in scenes. Though Srinasan and Kukreja have not yet tested their system, they believe that tagging clusters of images will provide better summarization of video.

This study is very similar to the one I propose, and shows the potential for tagging keyframes of a video as a way to gain a deep understanding of the content in a video. Since video is necessarily dynamic and difficult to tag as a whole, tagging individual keyframes allows the user to understand the gist of specific parts and specific content of a video. However, the Srinasan and Kukreja system does have one large flaw. By ignoring the temporal nature of videos and tagging only clusters of specific images, they lose the ability of tagging videos to summarize the narrative or story. Content may be tagged more efficiently and with less redundancy, but providing a user with the overall

gist of a video may be lost. In the Srinasan and Kukreja system, tagging is clustered and limited to description of the specific content of an image, not the sequence of events.

D. Summary:

Video inherently needs surrogates that are designed specifically for the unique qualities of video, such as moving images and synchronized sound. Today, search behaviors suggest that users typically use thumbnail images to determine relevancy of a digital video (Cunningham and Nichols, 2008). There is room for improvement in this area by supplementing video search results with more surrogates. Research in creating more useful surrogates can be placed into two camps: automated metadata and user-created metadata. Storyboards, sequential keyframes or thumbnails meant to represent the many visual parts of a video, have been shown to be an effective surrogate for the improvement of video retrieval (Christel, 2008). However, research has shown that to provide more data to the user other surrogates need to be used with storyboards, especially when the overall gist or story of a video needs to be communicated - rather than specific content (Christel, 2008). Studies suggest that social tagging may be another effective form of providing relevant metadata to video (Melonhorst, 2008). Work is beginning to appear that suggests that tagging significant thumbnails generated from a video could be an effective means of improving video retrieval (Srinasan & Kukreja, 2008). It might be the case that tagging a storyboard could also be an effective way to provide the user with a deep understanding of the content and context of a video.

3. Methodology:

This is an exploratory study, meaning that the results are preliminary. Twenty participants were randomly divided into two groups and asked to tag videos. The tags were then collected and used to test the original hypothesis. Quasi-experimentation, a method of data collection based around causal relationships and a derivation of true experimentation, was the method used to conduct this study (Cook & Campbell, 1979, p. 1-2). The difference between quasi-experimentation and true experimentation centers primarily around the concept of control (Caporaso, 1973, p. 5). True experiments should allow the researcher to completely control all aspects and variables of a study. A quasi-experiment is used when some of those variables may not be controlled for various reasons – usually because of a functional issue (Caporaso, 1973, p. 5). The typical area where researchers are unable to control their study when using quasi-experimental methods occurs when there is an inability to randomize samples (Caporaso, 1973, p. 12). For example, a researcher may not be able to randomly select and sample participants or subjects and then, must resort to using a convenience sample. This study uses a convenience sample of participants and, therefore, will be quasi-experimental in nature. Participants were randomly selected undergraduates at the University of North Carolina—Chapel Hill.

The purpose of an experiment is to ascertain causal relationships or set up a situation where observations can be made to determine whether or not a variable will change due to an intervention of another variable. The variable that intervenes in some

way causes a change in the original variable (Cook & Campbell, 1979, p. 12). The variables that cause change are typically referred to as independent variables. The original variable and the variable that results from the change caused by the independent variable are the dependent variables. Therefore, an experiment should involve an independent variable causing change to a dependent variable.

The specific form of quasi-experimentation used in this study will be a between-subjects nonequivalent control group, posttest only design. Because of this design, it is difficult to infer causality without statistically significant results. The study consisted of two groups: one used as a control and the other used as the experimental group.

Experimental group: R X O

Control group: R O

Since tags must be generated for this study, twenty participants were asked to tag five videos. Half of the participants, used as a control group, were asked to tag the entire video in the traditional manner of tagging videos. The other half of the participants were asked to tag the same videos; however, they were asked to tag a storyboard. The storyboard segmented each video into fifteen-second chunks. Therefore, two separate groups of tags were generated for the same five videos. The two groups of tags were then analyzed, using quantitative analysis, to determine whether a significant difference exists between the two forms of tagging. Any observed differences were used to determine whether the implementation of a tag-able storyboard is an effective method of surrogation for digital video retrieval and user gisting.

A. Selection and preparation of videos:

Five similar videos that already had storyboards were chosen based on content and length of time from *Open Video* (open-video.org), a site that provides access to videos in the public domain (Appendix 9.1). Each video was stripped of all metadata and surrogates to eliminate internal variables or influence when the users tagged the videos. The storyboard was kept for use in the experimental group only.

After inspection of the storyboards that were provided by *Open Video*, it was determined that the storyboards were not evenly selected from each video. The thumbnails from the videos were not representative of evenly divided chunks of time, and there were differing amounts of thumbnails for each video. Therefore, it was necessary to generate new thumbnails when needed for each video. Each video was loaded into Final Cut Pro, a professional video editing software. Each video was divided into fifteen-second chunks, and a thumbnail was then generated for every fifteen seconds. The thumbnails were then exported as image files to a spreadsheet, which ordered the thumbnails in chronological order. Next to each thumbnail was a brief caption that contained the times in the video that each particular thumbnail represents. [image here] Therefore the storyboard became a tag-able timeline for each the video.

The five videos were all between 2:20 and 2:50 minutes. The content of each video was instructional, pertaining to sociological or scientific topics. The videos were downloaded and played to the participants using a Quicktime player in a Firefox web browser. Web input-forms were attached to the videos so that the users were able to input tags during or after watching each video. Participants separated each tag they entered by commas in order to delineate where a tag began and ended. The tags were automatically stored in a MySQL database, which was later exported to a spreadsheet.

The videos with storyboards were arranged next to the video in a browser. Each thumbnail had an input form next to it, so that the users could then tag each image in the video (Appendix 9.3). Participants were then asked to watch the videos and tag them as they see fit. Only instructions as to how to tag the video were provided (Appendix 9.4).

B. Participants:

Twenty participants were randomly chosen responders to a mass email targeting undergraduates on a UNC-CH listserv. It can be assumed that the two groups are similar since the participants were all undergraduates. Emails were sent to undergraduate students at the University of North Carolina – Chapel Hill asking for participation in the tagging of the videos. (Appendix 9.2) The participants were gathered and randomly split into two groups, with ten in each group. The participants selected for the control group were asked to tag the entire video, and the participants selected for the experimental group were asked to tag the storyboard of each video (Appendix 9.4). The groups each had an hour to complete the tagging. The location for the study was the Media Lab of the Media Resources Center of the House Undergraduate Library at the University of North Carolina—Chapel Hill. Each station that the participants used had an identical setup. Headphones were meant to isolate any noises distracting the participants. To limit the number of external variables and maximize the amount of control exercised over the experiment, the only difference between the two groups was the method of tagging. The groups were tested separately and on different days. Upon arriving to the study, participants were asked to read and sign a consent form. Each participant was randomly

assigned a station. The interface was already setup on each computer to eliminate loss of time or any confusion.

C. Summary

Quasi-experimentation was used to attempt to find a causal relationship between a system that combines user-generated tags and storyboards and the tagging behavior of a group of random responders to a mass email. Quantitative analysis was used to analyze the data.

4. Results

A. Control Group (Users tagging the Entire Video):

The control group, the participants asked to tag the entire video, produced a total of 290 tags for all five videos. Each participant averaged a total of 29 tags across the five videos, and each participant averaged a total of 5.8 tags per video (Fig 1). The median was 4, and the mode was 1 tag, with eleven instances of 1 tag per video. The range for total tags per participant across all five videos was 60. The range for total tags per video was 21. The standard deviation for all tags for all five videos is 21.71. The standard deviation for average amount of tags per video was 0.82.

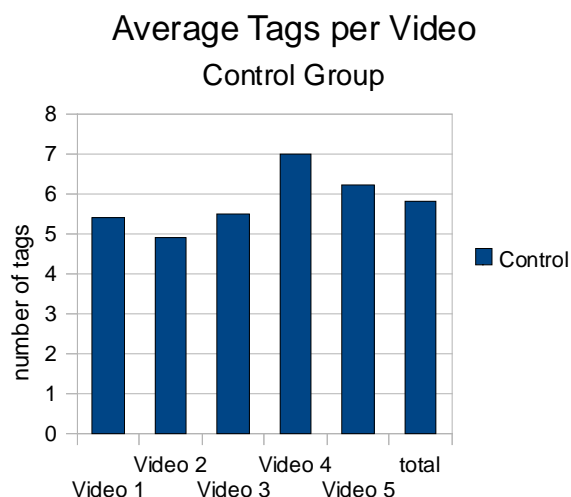


Figure 1: Control Group average tags

Video 1 had a total of 54 tags with an average of 5.4 tags per person (Fig. 1 & 2). The range was 13, and the standard deviation was 4.38. The Video 2 had a total of 49 tags with an average of 4.9 tags per participant (Fig. 1 & 2). The range was 9, and the

standard deviation was 3.25. Video 3 had a total of 55 tags with an average of 5.5 tags per person (Fig. 1 & 2). The range was 11, and the standard deviation was 4.14. Video 4 had a total of 70 tags with an average of 7 tags per person (Fig. 1 & 2). The range was 16, and the standard deviation was 5.85. Video 5 had a total of 62 tags with an average of 6.2 tags per person (Fig. 1 & 2). The range was 13, and the standard deviation was 4.71.

	Video 1	Video 2	Video 3	Video 4	Video 5
P1	7	7	12	10	11
P2	1	1	1	1	1
P3	1	1	2	1	2
P4	14	10	9	16	12
P5	12	10	12	17	14
P6	4	3	5	6	4
P7	4	4	4	4	4
P8	4	5	5	7	6
P9	4	5	4	7	7
P10	3	3	1	1	1
Total	54	49	55	70	62
average	5.4	4.9	5.5	7	6.2
St dev	4.38	3.25	4.14	5.85	4.71

Figure 2: Control Group total tags per participant (P)

B. Experimental Group (Users Tagging the Storyboard):

The storyboard group, the users asked to tag the storyboard, which divided the video into 15-second chunks, produced a total of 699 tags. There were a total of 69.9 tags per video, and an average of 13.98 tags made per participant (Fig. 3). The range for the total amount of tags per participant was 60. The range for total tags per video was 19. The standard deviation for total tags per participant was 18.73. The standard deviation

for average amount of tags per video was 0.74. The median was 14 tags, and the mode was 11 tags.

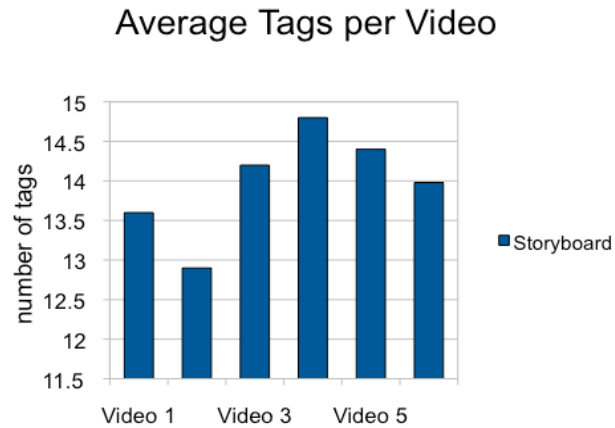


Figure 3: Average tags for the Experimental Group (Storyboard)

Participants viewing Video 1 produced an average of 13.6 tags with a standard deviation of 4.03 and range of 12 (Fig. 3&4). Video 2 had an average of 12.9 tags with a standard deviation of 2.51 and a range of 7 (Fig. 3&4). Video 3 had an average of 14.2 tags with a standard deviation of 3.82 and a range of 12 (Fig. 3&4). Video 4 had an average of 14.8 tags with a standard deviation of 6.6 and a range of 23 (Fig. 3&4). Video 5 had an average of 14.4 tags with a standard deviation of 7.03 and a range of 20 (Fig. 3&4).

	Video 1	Video 2	Video 3	Video 4	Video 5
P1	9	15	14	17	22
P2	11	13	14	26	24
P3	18	17	17	17	15
P4	11	12	12	11	12
P5	17	11	18	15	20
P6	12	14	19	17	19
P7	9	10	12	10	10
P8	15	10	7	3	4
P9	13	11	11	10	4
P10	21	16	18	22	14
Total	136	129	142	148	144
Average tags	13.6	12.9	14.2	14.8	14.4
St. Dev.	4.03	2.51	3.82	6.6	7.03

Figure 5: Total tags per user for Experimental Group

		Statistics									
		TV_W1	TV_W2	TV_W3	TV_W4	TV_W5	TV_S1	TV_S2	TV_S3	TV_S4	TV_S5
N	Valid	10	10	10	10	10	10	10	10	10	10
	Missing	0	0	0	0	0	0	0	0	0	0
	Mean	5.40	4.90	5.50	7.00	6.20	13.60	12.90	14.20	14.80	16.30
	Std. Error of Mean	1.384	1.027	1.310	1.850	1.489	1.275	.795	1.209	2.086	2.039
	Median	4.00	4.50	4.50	6.50	5.00	12.50	12.50	14.00	16.00	17.00
	Mode	4	1 ^a	1 ^a	1	1 ^a	9 ^a	10 ^a	12 ^a	17	4 ^a
	Std. Deviation	4.377	3.247	4.143	5.850	4.709	4.033	2.514	3.824	6.596	6.447
	Variance	19.156	10.544	17.167	34.222	22.178	16.267	6.322	14.622	43.511	41.567
	Skewness	1.213	.574	.703	.749	.537	.632	.403	-.460	-.053	-.618
	Std. Error of Skewness	.687	.687	.687	.687	.687	.687	.687	.687	.687	.687
	Kurtosis	.522	-.684	-.858	-.498	-1.151	-.624	-1.260	-.411	.174	-.340
	Std. Error of Kurtosis	1.334	1.334	1.334	1.334	1.334	1.334	1.334	1.334	1.334	1.334
	Range	13	9	11	16	13	12	7	12	23	20
	Minimum	1	1	1	1	1	9	10	7	3	4
	Maximum	14	10	12	17	14	21	17	19	26	24
	Sum	54	49	55	70	62	136	129	142	148	163

Figure 4: Total Statistics for each Video. TV_W is Control. TV_S is Experimental.

5. Discussion and Analysis:

The results of the study indicate that having users tag a storyboard, as opposed to tagging a whole video, significantly increases the amount of tags generated. The total average amount of tags per user, per video for the control group, was 5.8; the storyboard group averaged 13.98 tags per user, per video (Fig. 6, 7, & 8). Therefore, the experimental group averaged 2.4 times the amount of tags than the control group. The difference between means was 8.18.

After running an Independent Sample T-Test on each video and for the total averages of each group, it could be determined that the results were statistically significant and that the null hypothesis of no change from the control group could be rejected (Fig. 9). Therefore, the data suggests that hypothesis that the implementation of a storyboard tagging system could increase the amount of tags users input by more than 2 times the amount of tags users normally input.

It is also of note that the tendencies of the participants to tag certain videos more or less was nearly identical. Video 4 had the most tags in both groups, and Video 2 had the least tags in both groups. This is evidence that the amount of variables disrupting the relevance of the data was minimal and that the implementation of the storyboard most likely caused the increase in the amount of tags.

The cause of this increase can only be assumed since the research is still in the early stages; however, it is likely that the participants using the storyboard system created more tags because they felt obligated to create more tags due to the increased number of

possibilities of tagging the video. The typical system of tagging has one input field. The storyboard system generates a new field and thumbnail every fifteen seconds.

	Control	Storyboard
Video 1	5.4	13.6
Video 2	4.9	12.9
Video 3	5.5	14.2
Video 4	7	14.8
Video 5	6.2	14.4
Total	5.8	13.98

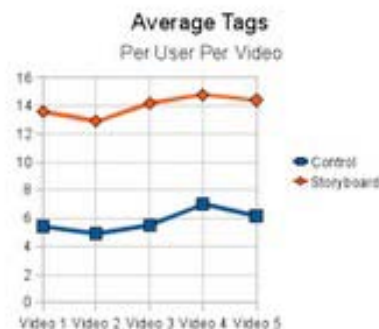


Figure 8

Figure 9

Group Statistics

group		N	Mean	Std. Deviation	Std. Error Mean
test	A	5	5.8000	.81548	.36469
	B	5	13.9800	.74297	.33226

Figure 7: Overall Stats. Control is A and Exp. Is B

		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper
test	Equal variances assumed	.056	.818	-16.580	8	.000	-8.18000	.49336	-9.31768	-7.04232
	Equal variances not assumed			-16.580	7.932	.000	-8.18000	.49336	-9.31939	-7.04061

Figure 6: T-Test Results

6. Limitations and Future Work:

Part of the original research question involves determining whether the increase in tags will result in increased user gisting and video retrieval, but that is impossible to judge at this stage in the research. Future studies should investigate if there is a relationship between the implementation of a tag-able storyboard as a surrogate for digital video and an increase in the user's ability to effectively search and browse for digital video. Data should also be collected in regards to the effectiveness of the extra tags themselves.

This study was conducted under a controlled setting. Since the participants were asked to tag the videos, it can be assumed that the amount of tags generated for both groups in this study may be higher or lower than videos tagged in the real world. Geisler and Burns found that the median amount of tags was 6 from a sample of more than one million tagged YouTube videos (2007). In this study, control group had a median of 55 tags for 10 participants across five videos. The group using the storyboard system had a median of 142 tags for 10 participants across the same five videos. The controlled test produced many more tags than Geisler and Burns' findings. The data for this study suggests that implementing the storyboard system would generate 2.4 times the amount of tags typically generated. Therefore, a future research question is whether or not implementing a storyboard tagging system in real world circumstances would generate 2.4 times as many average tags per person, per video. It would be beneficial for future

researchers to mimic the circumstances of tagging in a real world-like scenario with other surrogates to determine if the increase in tags will still occur if such a system were used.

Another topic for future research is whether this combined system would be more effective in video retrieval than having the storyboard alone. One form of data that could not be gathered during this study is the difference in the amount of time it took to tag each video. A higher cost of time could reduce the users' willingness to tag the storyboard.

The goal of this research is to eventually develop a simple system that will allow users to tag videos in chunks of time, creating not just more tags but more specificity in the tags. Users would then be able to click on the thumbnail image in the storyboard that corresponds to that tag to jump to that time in the video.

7. Conclusion:

The findings of this study suggest that the implementation of a system that combines automatically generated storyboards with social tagging will increase the number of tags that each user will add to a video. This study does not suggest that the increase in tags will aid in video retrieval and user gisting; however, the results do suggest that it is worth investigating in future work. In all aspects of the results, there was significant increase in the amount of tags created in the storyboard system when compared to the traditional system of tagging an entire video.

8. Bibliography

- Caporaso, J. A., Roos, L. L., & Bohner, H. J. (1973). *Quasi-experimental approaches; testing theory and evaluating policy*. Evanston, [Ill.]: Northwestern University Press.
- Chi, E. H. and Mytkowicz, T. (2008). Understanding the efficiency of social tagging systems using information theory. In *Proceedings of the Nineteenth ACM Conference on Hypertext and Hypermedia*. DOI=
<http://doi.acm.org/10.1145/1379092.1379110>
- Christel, M. (2009). *Automated metadata in multimedia information systems: Creation, refinement, use in surrogates, and evaluation*. [San Rafael, CA]: Morgan & Claypool.
- Christel, M. (2008). Supporting video library exploratory search: when storyboards are not enough. In *Proceedings of the 2008 International Conference on Content-Based Image and Video Retrieval*. DOI=
<http://doi.acm.org/10.1145/1386352.1386410>
- Christel, M. and Moraveji, N. (2004). Finding the right shots: assessing usability and performance of a digital video library interface. In *Proceedings of the 12th Annual ACM international Conference on Multimedia*. DOI=
<http://doi.acm.org/10.1145/1027527.1027691>
- Christel, M. and Yan, R. (2007). Merging storyboard strategies and automatic retrieval for improving interactive video search. In *Proceedings of the 6th ACM international Conference on Image and Video Retrieval*. DOI=
<http://doi.acm.org.libproxy.lib.unc.edu/10.1145/1282280.1282351>
- ComScore. (2009). "YouTube.com Accounted for 1 Out of Every 3 U.S. Online Videos Viewed in January." Retrieved from:
http://www.comscore.com/Press_Events/Press_Releases/2008/03/YouTube_Usage/%28language%29/eng-US
- Cook, T. D., & Campbell, D. T. (1979). *Quasi-experimentation: Design & analysis issues for field settings*. Boston: Houghton Mifflin.
- Cunningham, S. J. and Nichols, D. M. (2008). How people find videos. In *Proceedings of the 8th ACM/IEEE-CS Joint Conference on Digital Libraries*. DOI=
<http://doi.acm.org/10.1145/1378889.1378924>

- Geisler, G., & Burns, S. (2007). Tagging Video: Conventions and Strategies of the YouTube Community. In *Proceedings of the 7th ACM/IEEE-CS joint conference on Digital libraries*. (p. 480).
- Golder, S.A., & Huberman, B.A. (2006). Usage Patterns of Collaborative Tagging Systems. *Journal of Information Science*, 32(2), 198-208.
- Haas, D. & Kraft, D. (1984). Experimental and quasi-experimental designs for research in information science. *Information Processing & Management* (20)1-2. DOI: 10.1016/0306-4573(84)90053-0
- Halpin, H., Robu, V., & Shepherd, H. (2007). The complex dynamics of collaborative tagging. In *Proceedings of the 16th international conference on World Wide Web* (pp. 211-220).
- Heckner, M., Neubauer, T., and Wolff, C. (2008). Tree, funny, to_read, google: what are tags supposed to achieve? a comparative analysis of user keywords for different digital resource types. In *Proceeding of the 2008 ACM Workshop on Search in Social Media*. DOI= <http://doi.acm.org/10.1145/1458583.1458589>
- Macgregor, G., & McCulloch, E. (2006). Collaborative Tagging as a Knowledge Organisation and Resource Discovery Tool. *Library Review*, 55(5), 291-300.
- Marchionini, G., Song, Y., and Farrell, R. (2009). Multimedia surrogates for video gisting: Toward combining spoken words and imagery. *Inf. Process. Manage* (45) 6. DOI= <http://dx.doi.org/10.1016/j.ipm.2009.05.007>
- Melenhorst, M., Grootveld, M., van Setten, M., and Veenstra, M. (2008). Tag-based information retrieval of video content. In *Proceeding of the 1st international Conference on Designing interactive User Experiences For TV and Video*. DOI= <http://doi.acm.org/10.1145/1453805.1453813>
- O'Connor, B. (1985). Access to moving image documents: Background concepts and proposals for surrogates for film and video works. *Journal of Documentation*, 41(4), 209-220.
- Srinivasan, S. H. and Kukreja, M. (2008). Tagboards for video tagging. In *Proceeding of the 16th ACM international Conference on Multimedia*. DOI= <http://doi.acm.org/10.1145/1459359.1459518>
- Wildemuth, B., Marchionini, G., Yang, M., Geisler, G., Wilkens, T., Hughes, A., et al. (2003). How Fast Is Too Fast? Evaluating Fast Forward Surrogates for Digital Video.

Yang, M., Wildemuth, B. M., and Marchionini, G. (2004). The relative effectiveness of concept-based versus content-based video retrieval. In *Proceedings of the 12th Annual ACM international Conference on Multimedia*. DOI=
<http://doi.acm.org/10.1145/1027527.1027613>

9. Appendices

9.1: Videos Used from Open Video.

<http://www.open-video.org/details.php?videoid=348&surrogate=storyboard>

<http://www.open-video.org/details.php?videoid=368&surrogate=storyboard>

<http://www.open-video.org/details.php?videoid=434&surrogate=storyboard>

<http://www.open-video.org/details.php?videoid=548&surrogate=storyboard>

<http://www.open-video.org/details.php?videoid=722&surrogate=storyboard>

9.2: Mass Email

Hello,

My name is Tim Obert, and I am a Master's student working under the supervision of Gary Marchionini, PhD. in the School of Library and Information Science. I am looking for volunteers to participate in a research study on Digital Video retrieval.

Participation in this study involves coming into a laboratory and viewing five, short videos and tagging them or describing them in brief, single word statements. You will also be asked to fill out a brief questionnaire. Participation in this study will take approximately 30 minutes.

Four participants will randomly receive a drawing of twenty-five dollars.

If you are interested in participating, please respond to this email. I will then send a confirmation and provide you with further information concerning the location and available times of the study. If you have questions, please email me.


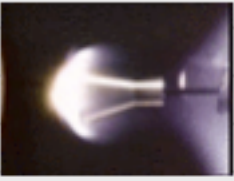



IRB Study # 10-0074
The IRB approved this study on 2/04/2010.

Thank you for your time.

Sincerely,
Tim Obert

This email is sponsored by: School of Library and Information Science

9.3: Storyboard Screenshot

	0:00 – 0:15
	0:16 – 0:30
	0:31 – 0:45
	0:46 – 1:00
	1:01 – 1:15
	1:16 – 1:30

9.4: Instructions:**Control:**

Watch the following videos. Please tag the video by placing the tags in the available input forms. Separate each tag with a comma. It does not matter how many tags you place. Be sure to hit submit after you are done tagging.

Storyboard Videos:

Please tag the video by placing the tags in the available input forms next to each thumbnail. Separate each tag with a comma. It does not matter how many tags you place. Be sure to hit submit after you are done tagging.

9.5: Consent form

**University of North Carolina-Chapel Hill
Consent to Participate in a Research Study
Adult Participants
Social Behavioral Form**

IRB Study # 10-0074
Consent Form Version Date: January 24, 2010

Title of Study: Combining User Tagging and Automatically Generated Storyboards for Improved User Gisting

Principal Investigator: Timothy Obert

UNC-Chapel Hill Department: School of Library and Information Science

UNC-Chapel Hill Phone number:

Email Address: info@ils.unc.edu

Faculty Advisor: Gary Marchionini, Phd

Study Contact telephone number: (706) 255-1206

Study Contact email: obert@email.unc.edu

Faculty Advisor telephone number: (919) 966-3611

Faculty Advisor email: march@ils.unc.edu

What are some general things you should know about research studies?

You are being asked to take part in a research study. To join the study is voluntary. You may refuse to join, or you may withdraw your consent to be in the study, for any reason, without penalty.

Research studies are designed to obtain new knowledge. This new information may help people in the future. You may not receive any direct benefit from being in the research study. There also may be risks to being in research studies.

Details about this study are discussed below. It is important that you understand this information so that you can make an informed choice about being in this research study. You will be given a copy of this consent form. You should ask the researchers named above, or staff members who may assist them, any questions you have about this study at any time.

What is the purpose of this study?

The purpose of this research study is to learn about new methods of digital video retrieval and how these methods may differ from one another.

Are there any reasons you should not be in this study?

You should not be in this study if you are a minor.

How many people will take part in this study?

If you decide to be in this study, you will be one of approximately 26 people in this research study.

How long will your part in this study last?

The study will last approximately one hour and thirty minutes on one evening with no follow up time.

What will happen if you take part in the study?

Before the study begins, you will be randomly assigned to one of two groups. There will be a 50% chance of being assigned to either group. The study will begin with you filling out a questionnaire/survey. You will then be assigned a computer. The computers are all the same and will be randomly assigned. You will have 1.5 hours to watch 5 short videos and tag them according to the instructions provided. You will tag the videos in a form that will be attached to the videos on screen. Once you are done, let the researcher know, and you are then free to leave.

What are the possible benefits from being in this study?

Research is designed to benefit society by gaining new knowledge. You may not benefit personally from being in this research study.

What are the possible risks or discomforts involved from being in this study?

Risks are being minimized by conducting the study in a controlled lab setting. There are no known risks, but there may be uncommon or previously unknown risks. You should report any problems to the researcher.

How will your privacy be protected?

Participants will be identified only by numbers. Names will not be used for the study. Email addresses and names will be held to issue the reward. Afterwards, the names will be destroyed. Any computer with identifying information will be password protected, and any file will be password protected and encrypted. No one will have access to the information but the primary investigator.

What if you want to stop before your part in the study is complete?

You can withdraw from this study at any time, without penalty. The investigators also have the right to stop your participation at any time. This could be because you have failed to follow instructions, or because the entire study has been stopped.

Will you receive anything for being in this study?

You may not receive anything for taking part in this study. However, there will be a randomized drawing to award two participants \$25 each.

Will it cost you anything to be in this study?

There will be no costs for being in the study [other than time.]

What if you are a UNC student?

You may choose not to be in the study or to stop being in the study before it is over at any time. This will not affect your class standing or grades at UNC-Chapel Hill. You will not be offered or receive any special consideration if you take part in this research.

What if you are a UNC employee?

Taking part in this research is not a part of your University duties, and refusing will not affect your job. You will not be offered or receive any special job-related consideration if you take part in this research.

What if you have questions about this study?

You have the right to ask, and have answered, any questions you may have about this research. If you have questions, complaints, concerns, or if a research-related injury occurs, you should contact the researchers listed on the first page of this form.

What if you have questions about your rights as a research participant?

All research on human volunteers is reviewed by a committee that works to protect your rights and welfare. If you have questions or concerns about your rights as a research subject, or if you would like to obtain information or offer input, you may contact the Institutional Review Board at 919-966-3113 or by email to IRB_subjects@unc.edu.

Title of Study: Combining User Tagging and Automatically Generated Storyboards for Improved User Gisting

Principal Investigator: Timothy Obert

Participant's Agreement:

I have read the information provided above. I have asked all the questions I have at this time. I voluntarily agree to participate in this research study.

Signature of Research Participant

Date

Printed Name of Research Participant

Signature of Principal Investigator

Date

Printed Name of Principal Investigator