CONTENT ENRICHMENT FOR MOBILE CONTEXT AWARE IMAGING APPLICATIONS WITH A SOCIAL ASPECT

Victor Y.W. Ng

Department of Electrical and Computer Engineering

McGill University, Montreal

August, 2013

A thesis submitted to the Faculty of Graduate and Postdoctoral Studies in partial fulfillment of the requirements for the degree of Master of Engineering

Abstract

With technological advancements in data collection techniques allowing increasing amounts of contextual metadata to be appended to everyday image files, it has become a daunting task to effectively display such supplemental data alongside original image content without overloading users with information. I propose a unique approach to actively display geo-tagged image content that embeds the images in a navigable 3D environment in a way that makes explicit the geographical context and spatiotemporal relationships between the images. This approach enhances the viewer's comprehension of the image's context and content thus supporting my hypothesis that context extracted from metadata can enhance image content absorption rather than hinder it.

The 3D environment is built by mapping Google Street View images onto a spherical tessellation within which user image content is overlaid. The proposed geographical browser and social-networking system is implemented on an iPad, using the iPad's built-in compass, gyroscopes, and accelerometers to provide real-time gesture control and spatial orientation. User studies were performed on the proposed system as well as on a standard social-networking application for comparison purposes. The results were used to evaluate the relative performance of the system in enabling users to absorb and comprehend image information. Test subject were found to consistently answer questions more accurately on images viewed on the proposed system as compared to the images viewed on a typical social-networking application.

Abstrait

Les avancées technologiques dans les techniques de collection de données augmentant la quantité de méta données contextuelles à associer aux images de tous les jours, afficher ces données supplémentaires de manière effective avec l'image originale est devenu une tâche ardue. Je propose une approche unique permettant d'afficher de manière active du contenu géo taggué qui renferme des images dans un environnement 3D navigable d'une manière qui rend explicite le contexte géographique et la relation spatiotemporelle entre les images. Cette approche augmente la compréhension du contexte et du contenu de l'image par l'observateur, supportant ainsi ma thèse selon laquelle le contexte extrait des métadonnées peut améliorer l'absorption du contenu de l'image plutôt que de le cacher.

L'environnement 3D est construit en cartographiant des images de Google Street View dans une tessellation sphérique sur laquelle le contenu de l'image est superposé. Le navigateur géographique et outil de réseau social proposé est implémenté sur un iPad, utilisant la boussole intégrée à l'iPad, ses gyroscopes et accéléromètres pour fournir en temps-réel des contrôles par gestes et orientation spatiale. Des études utilisateurs ont été effectuées sur le système proposé, de même que sur une application de réseau social standard à fin de comparaisons. Les résultats ont été utilisés afin d'évaluer la performance relative du système à permettre aux utilisateurs d'absorber et comprendre l'information de l'image. Les sujets de test ont donné de manière consistante des réponses plus précise aux questions répondues sur des images observées sur le nouveau système comparé aux images observées sur une application typique de réseau social.

Acknowledgements

I am heartily thankful to my supervisor, James Clark, whose encouragement, guidance and support from the first to the final days of my master degree at McGill. He has always treated me with respect and kindness. My time spent under his mentoring has been truly enriching.

The Visual Motors Research lab with all its members has made my graduate studies an enjoyable one. Qing Tian, Amin Abolhassani, and Carmen Au were always generous with their time, support and encouragement.

Jan Binder, Marlene Grey, and Cynthia Davidson and all CIM personnel have been extremely helpful with administrative processes. They have made my time at CIM a truly smooth and enjoyable experience.

Thank you Google Street View for your images. The project wouldn't be as aesthetically pleasing without your help.

I would also like to thank Vei Vern for being there for me. Her unwavering support and belief sustained me through thick and thin. I will be forever grateful to her.

Finally, I would like to thank my family, who means the world to me. My parents' and sisters' never-ending advice and pride in me has been my foundation for self-improvement ever since I could remember. I could not have dreamed for a better family.

Table of Contents

Table of Figures

Chapter 1
Introduction

1.1 Context Aware Imaging & Display Methodology

Ever since the first Autographic Kodak Camera invented in 1914 allowed users to manually record the time and date on film negatives (Kodak, 1999), data collection technologies have boomed, introducing more types of information into pictures. This additional information is usually obtained the second an image is captured, and gives more background information related to the image content. Since this type of information gives context to images, it has been coined the name “contextual information”.

Towards the end of the 20th century, advancements in microprocessor and charge-coupled device (CCD) technology digitized the photography industry. Metadata is digitized contextual information stored directly into an image file, thereby transforming the image file into a contextualized image. With a wealth of sensors now commonly available on commercial digital cameras and smartphones, a typical digital image may contain metadata such as time, date, and location of capture.

The field of research on contextualized images has expanded and evolved to incorporate not only the collection of metadata, but also the displaying and sharing of metadata rich images. Many applications used to display and share contextualized images have been developed in various labs across the globe, but none has been able to fully format such images for effective cognitive consumption. Like Miller stated in his earlier studies, humans have limited information processing capabilities or "Cognitive Load Limits" (Miller, 1956). A recurring issue with displaying contextualized images is that the display of metadata alongside the image content can easily increase a person's Cognitive Load past acceptable limits. This was noted as early as the mid-90's as indicated by the following quote from Lamming and Flynn.

> *"When these extra data start flowing in, the task of organizing them [cognitively] will become even more daunting."* (Lamming et al., 1994)

In this case, the use of contextual information would hinder image content absorption rather than enriching it.

The goal of this thesis is to create a novel system to display contextualized images using metadata. Rather than passively displaying contextual information, metadata will be actively displayed catering to the way human beings observe and perceive their surroundings. After a full evaluation of my method, the hypothesis that context from the use of metadata can indeed enrich content absorption will be examined.

1.2 Motivation for the System Presented in the Thesis

As discussed previously, many forms of contextual information can now be introduced into digital images as metadata. The ability to display the data in a meaningful and effective manner however, poses a problem. Based on current methods, there are two main ways to display contextualized images: passively and actively.

Passively displaying a contextualized image is to simply show the metadata in raw form alongside the image. An example would be printing the time and date next to the image displayed. This is the most common and historical form of contextual image displays, but does little to enhance the user's experience.

Actively displaying contextual images on the other hand, require the images to be modified through the use of metadata. An example would be to display images in sequence according to the time stamp. In this case, users would not be able to see the actual time stamp. However, by flipping through a sequenced image set, it is apparent which images are taken before and after others. This gives users a sense of time without reading the time stamps of each picture. It is important to note that actively displaying metadata does not necessarily require it to be physically shown, as long as it is used to impact the way images are displayed. The challenge of this type of display however, is how to combine forms of metadata in a way that enhances the user's experience rather than overloading them with information.

The goal of my thesis is to search for a display format that allows for contextualized information to be actively displayed enriching content absorption. To achieve this goal, I will create a 3D environment to house the images. Since people observe the world best when images are presented in a seamless matter, according to a study by (McConkie et al., 1995), creating a 3D environment using metadata allows users to view all images at the same time continuously. To control my test environment however, I have only employed limited metadata which are aimed to answer the most elemental questions of: *who, when, and where*. The types of metadata used will therefore be the photographer ID, time/date stamp, and location information. The selected metadata will be used to enrich the final elemental question "*what*", which corresponds to the picture content. Since people make relative judgments rather than absolute judgments

(Miller, 1956), image content will be actively displayed in a relative format. An example would be that images will be displayed from earliest to latest. Although users will no longer know the exact time and date of each image, the sense of time is preserved through knowing which image is taken before or after others.

My system was designed with users in mind, psychophysical and usability testing will be employed to quantitatively gauge the effectiveness of my system on human test subjects. A mock passive image display system will also be created and tested on subjects for comparison purposes. Finally, applying statistical methods on psychophysical test results will evaluate the statistical significance of the primary hypothesis.

1.3 Thesis Overview

This thesis describes the development of creating a novel contextualized image display based on metadata to effectively enrich image content.

Based on the studies from Miller and Sweller, Chapter 1 is intended to give the reader a brief background on the ways humans perceive and process information. Though not an extensive study, this brief overview will help the reader better understand the motivation for this thesis as well as the test methods.

Chapter 2 will give the reader a summary of the related research fields of context aware imaging under the themes of data collection, image display, and sharing. Past applications will be contrasted against each other in order to justify the design choices of my test system.

Chapter 3 summarizes the overall approach and design choices for creating the demonstration systems. The hardware and software used in the testing process will be illustrated in detail.

Chapter 4 will present psychophysical and usability testing methods employed to support my hypothesis. Test results will be displayed and trends will be identified. Test results will also be put through statistical analysis to determine the significance of observed trends. Significant trends will be combined with test subject feedback and analyzed.

In Chapter 5, limitations and areas of possible improvements will be presented and examined. Potential future applications of my findings will also be discussed, followed by a conclusion.

Chapter 2
Previous Work on Context-Aware Imaging

2.1 Background

Context aware imaging has evolved into a vast field of research which has shown the potential to improve on the process of taking, displaying, and sharing images. This research field has been more practical ever since metadata was introduced. In the following subsections, applications which aid in the collection of metadata, the displaying of metadata rich images and the sharing of such images will be illustrated.

2.1.1 Metadata Collection Techniques

Interviews conducted by (Holleis et al., 2005) showed that the general consensus favored supplementing photos with additional contextual data. Many types of metadata can be combined in order to create context. For example, looking up a certain date stamp on a photographer's calendar can show which event an image was taken at. But first, it may be appropriate to categorize conventional types for metadata as shown in Figure 2.1.

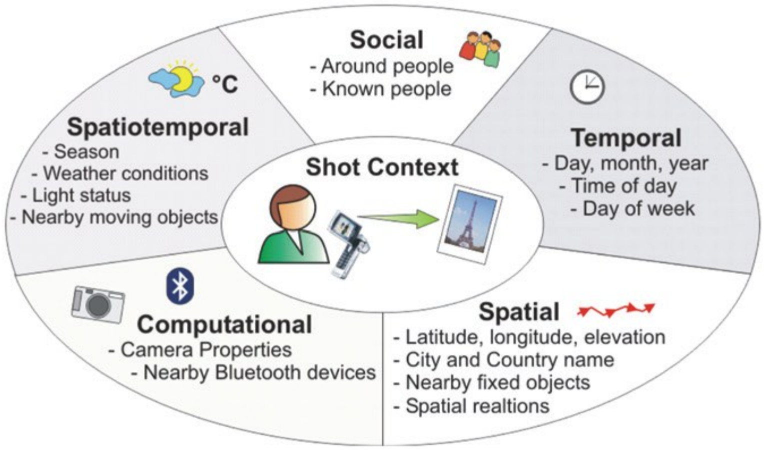


Figure 2.1: Typical Metadata Categories (Viana et al., 2009)

The metadata categories as depicted above do not include the entire range of metadata, but only tries to categorize conventional types. An example of non-conventional metadata is shown in an application named StartleCam (Healey et al., 1998) which uses the skin conductivity signal of the photographer. The skin conductivity reading is useful to judge how "startled" the photographer is when an image is taken. Images can then be sorted in order of the amplitude of the startle response thus allowing the creation of an "importance" filter. This is based on the idea that the level of skin startle reaction directly relates to the importance of the image being taken. Since StartleCam is a life logging application, where images are taken as the photographer lives his life; a method to determine which images are important is needed with mass amounts of pictures taken each day.

Experience Explorer (Belimpasakis et al., 2009) is another context based life logging application like StartleCam, but the metadata used is from the conventional spatiotemporal category (date stamp, geo location, etc.). Using images with spatiotemporal metadata, Experience Explorer displays images in order of a journey through geographical locations and social proximity to other photographers. Although Experience Explorer and StartleCam are both context aware life logging applications, they utilize different types of metadata. Therefore, the purpose and the abilities of the two systems are vastly different.

Since usability tests show that the spatiotemporal attributes are useful clues that can be both used for search purposes and to increase the described information of a photo (Viana et al., 2009), I will address mostly spatiotemporal metadata types to maintain the scope of this thesis. The most popular spatiotemporal metadata utilized is location metadata, as used in typical geo-tagged photos. However, geo-tagged photos only containing GPS coordinates are not a reliable resource to extract spatial information as their orientation is unknown (Chippendale et al., 2009). Therefore, in the research field of context aware imaging, spatial metadata has evolved to not only include GPS coordinates and orientation, but even compass heading and elevation.

Time is an important and natural context for many applications (Chen et al., 2000). This is due to the fact that time is a necessity to track change, may it be on a Point of Interest (POI) or simply determining the sequence in which a series of images were taken. In a social sharing environment, with images updated continuously through time, a context history can be generated via collecting and processing time stamps of past images.

Another significant context clue is the metadata related to the presence of people (Viana et al., 2009). This is especially important for applications like Experience Explorer or any social image sharing application. Having user identification reinforces cognitive connections between images when one can identify the photographer and even the individuals in the picture.

Since any reliance on the user to explicitly provide contextual information proves to be obtrusive and inconvenient to the user (Chen et al., 2000), combining metadata will allow the automated generation of high level context. Simply knowing the location, and current time, together with the user's calendar, applications will have a sufficient idea of the user's current social situation. Armed with such high level context, it is possible to use this context to aid in displaying image content.

2.1.2 Contextualized Image Display Applications

Having numerous ways to collect metadata through everyday consumer electronics, the research field has produced many display applications to employ the use of contextual information. There are infinite ways one can attempt to enrich content via context, but applications generally have taken up three trending approaches: visually modifying images via its context, using context to morph image content into an entirely different product, or creating a contextualized environment to house images. Although each trending approach has their own advantages, drawbacks are also inherited for each method.

Figure 2.2: Visual Effects Using Temperature Metadata (Hakansson et al., 2003)

Many applications in the research field use context awareness to create visual effects on images such as the one created by (Hakansson et al., 2003). An example of this can be seen in Figure 2.2. Color parameters of images were modified according to temperature metadata values collected when images were captured. By modifying the color parameters correctly, one can potentially tint the image cooler or warmer to reflect the temperature. This could achieve a more immersive experience when users view the resultant image and even "achieve artistic views of the world and capture the moment" (Hakansson et al., 2003). Although visual effects allow users to feel more immersed in single images, the effectiveness is reduced as the number of photos in a collection increases. Visual effects lack the ability to deliver enrichment when multiple images tinted differently can potentially confuse viewers.

Depending on how metadata is used, the effects can be diverse. Applications exist where image content is morphed using metadata into an entirely different form, for instance in the terrain tracking application by (Chippendale et al., 2008). Images taken by multiple individuals of a mountainous area are combined using metadata to create a 3D spatiotemporal attractiveness Geographic Information System (GIS) layer as seen in Figure 2.3. Not only is this application useful for seeing which part of a landscape is popular amongst photographers, images taken over time can also be used to track landscape changes. This application with its landscape tracking abilities has an important role to play in environmental observations (Chippendale et al., 2009).

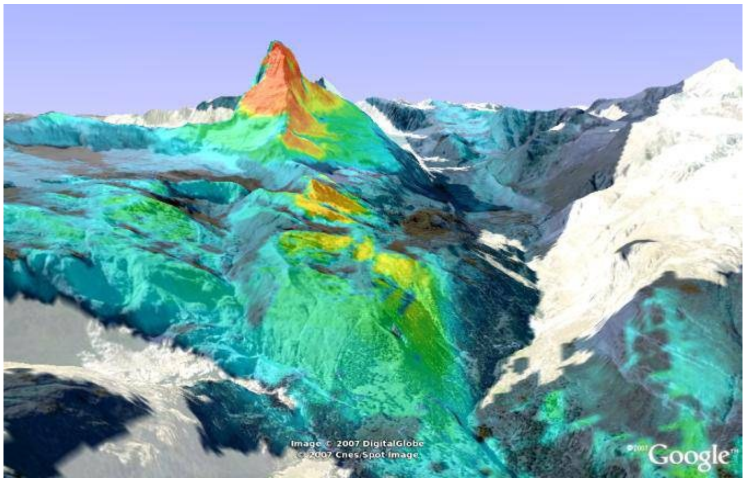
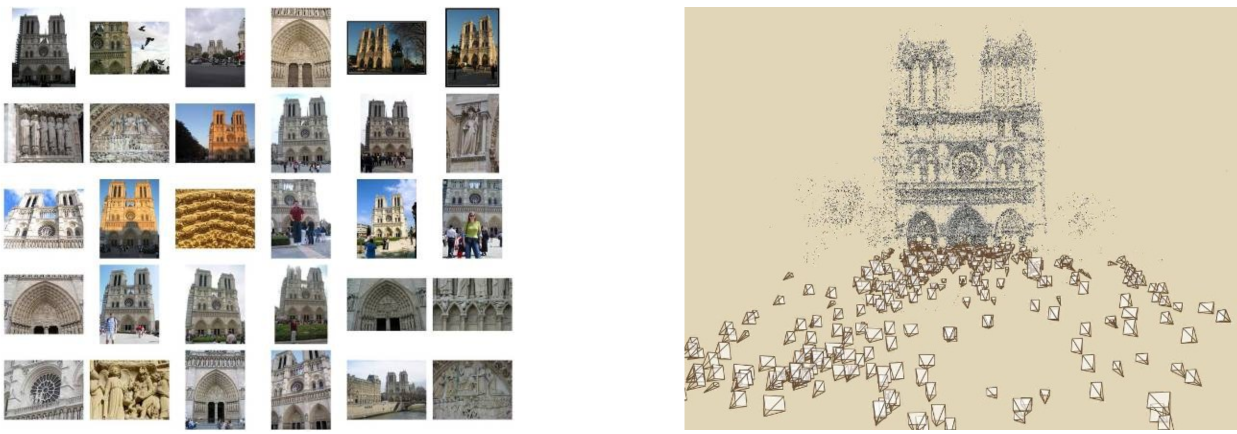


Figure 2.3: 3D spatiotemporal attractiveness GIS layer generated from 2D images (Chippendale et al., 2008)

3D reconstruction using contextualized images is also extremely popular for photo tourism applications. A photo tourism application by (Snavely et al., 2006) reconstructs POIs in a 3 dimensional space by combining numerous photos via position metadata. The model is used to answer questions such as, "Where was I, and what was I looking at when I took this picture?" An example of the Photo Tourism application can be seen in Figure 2.4. Both the Photo Tourism application and the application by (Chippendale et al., 2008) generate a new product by molding contextualized images into a 3D form, but the resultant products serve vastly different purposes. It is apparent that several possible useful products can be created from contextualized images to satisfy different criteria. One common issue with this approach however, is the diversion of attention away from the initial image content. Once relevant image content is processed and extracted, the remaining image data deemed irrelevant is either archived or discarded. This is evident in Figure 2.4 where many images used to reconstruct the cathedral contain people standing in front of the building. The constructed 3D model does not include these people because they have been filtered out in order to reconstruct the cathedral. Although both the Photo Tourism application and the GIS application by (Chippendale et al., 2008) create extremely useful 3D products from spatiotemporal metadata images, they both removed extraneous image content to achieve their purposes.



(a) (b)

Figure 2.4: Photo Tourism takes unstructured collections of photographs such as those from Google Images (a) and reconstructs POI in 3D (b). (Snavely et al., 2006)

A recently proposed method to effectively display multiple images while keeping all images intact in their entirety is to house the images in a contextualized environment. The use of common metadata across different images can be considered as implicit links between them (Belimpasakis et al., 2009). These implicit links can be used to connect image content via a web-like contextual environment. An example of this application is Image Space by (Lucero et al., 2009) where contextualized images are housed in a 3D environment. Images concentrated near a certain location or POI are shown in a proportional 3D space so users can "fly" from one photo to another allowing enrichment of content by using the sense of direction. A screenshot of Image Space can be seen in Figure 2.5. A common concern with this type of display is the issue of expressing the sense of space correctly. The connection between images diminishes as the distance between them increases, since far away images are usually too small to view effectively. Also, if two pictures are so close that they are overlapping, “neither of them is suitable for viewing” (Lucero et al., 2009). Since images spaced too close together or too far apart could render the display unusable, careful adjustment of space representation in image display environments directly impacts the effectiveness of contextualized image displays.

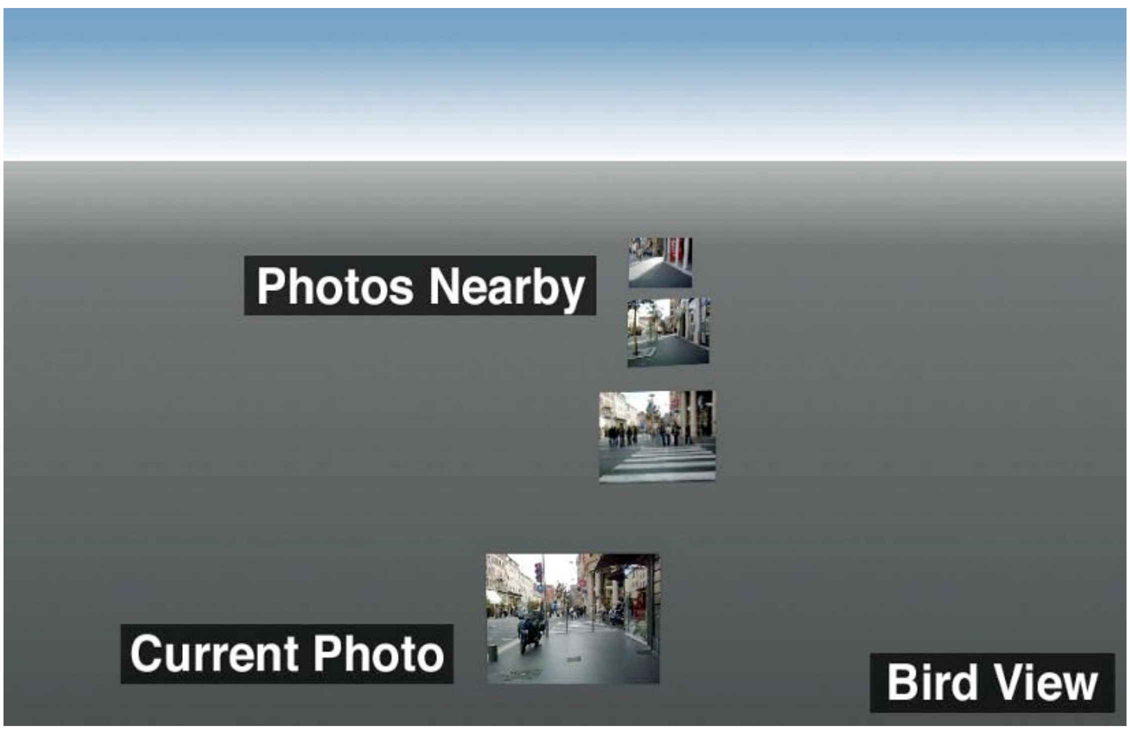


Figure 2.5: Image space's top-down viewpoint of the 3D world. The current photo is shown at the front while the other photos are shown in the surrounding 3D space. (Lucero et al., 2009)

2.1.3 Context Aware Images in Social Media Applications

Contextualized environments used to house many images are the norm for displaying pictures in an image sharing environment. As users upload numerous photos, they are often unmanaged, and thus image collections can become messy. Large collections of images are normally uploaded to a single online folder for later sorting by the user. However, most image collections remain unsorted due to the lack of user motivation in going through images and annotating them with metadata or keywords as Sarvas et al. pointed out (Sarvas et al., 2004). Unorganized image collections are not only confusing for users to view; they are even confusing for the photographer themselves. In order to circumvent messy image collections, a logical solution is to automatically sort images via readily available metadata. As Ahern et al. stated, *"allowing metadata to be shared creates an opportunity to leverage community effort to benefit each of the individuals within the community*." (Ahern et al., 2006)

Besides automatically sorting photo collections, applications like Image Space by (Lucero et al., 2009) use a 3D contextualized environment instead of the typical file and folder environment to house images. Although browsing through images sorted into folders is still the most widely used picture viewing method, flipping between files and folders creates a sporadic and segmented environment which hinders information absorption (McConkie et al., 1995). Since 3D displays allow users to look from one image to another seamlessly, 3D applications have increased in popularity for their cognitive advantages.

Automated 3D displays are extremely effective at displaying images without user input, but the effectiveness of the display is dependent on the spatial representation as mentioned in Section 2.1.2. In an image sharing environment where many images are being displayed simultaneously, displays are often overcrowded, making it difficult for users to understand what is being presented (Chippendale et al., 2010). Therefore, designers of 3D image viewing systems are required to create a scalable spatial representation which is consistent and informative. If the requirement is not met, 3D contextualized image viewing systems used to house shared images could potentially become less effective than the traditional method of using image folders.

2.2 Usability Testing and Psychophysical Methodology

The concept of usability is often used to judge the 'quality' of various software interface designs (McGee, 2003) which is directly linked to how effectively a user can utilize a system. Numerous research applications mentioned in Section 2 contained usability tests components which evaluate the effectiveness of their systems on human subjects. Within all of these applications, usability testing consisted of either questionnaires or interviews in order to record user feedback. Although these usability testing methods are widely used, they are susceptible to social desirability bias. Social desirability biases occur when users generally tell you what they think you want to hear, and are less likely to say disparaging things about people and products (Sauro, 2012). This can potentially skew results thus reducing the usefulness of usability tests. Although interviews and questionnaires can illustrate how pleased users are about a system, it can only show the perceived effectiveness, which can differ from the actual effectiveness. This is possible when a test subject states that a new interface is easier to use increasing his productivity, when it actually took the user longer to accomplish tasks.

Recently, usability testing has been augmented with psychophysical testing methodology to more accurately gauge the effectiveness of interfaces. Psychophysical methods are very efficient in obtaining numerical results free from criterion bias and parametric assumptions (Klymenko et al., 1990). Although psychophysical testing can be used to find actual efficiency instead of perceived efficiency, it is not sufficient by itself. This is due to the fact that psychophysical testing does not include addressing how pleased a user is with an interface. Therefore, researchers have merged psychophysical methods with the classic interview and questionnaire approach to gain the advantages from both testing methods.

Chapter 3
Demonstration System

Imagine a friend of yours has returned from a trip to Paris, and would like to share a collection of photos with you. In the traditional Facebook style of social networking interface, your friend would post the photos on his website page and you would view them one after another in a passive manner. It is very difficult to absorb useful information because there are no apparent connections between the images spatially or temporally unless image annotations were available and memorized. Apart from memorized annotations, the only information learned from purely viewing the images would probably be that your friend went to Paris, if a snapshot of a distinguishing landmark existed in the image set.

In order to enrich the experience of viewing image sets, I propose that social networking interfaces would benefit from actively displaying image sets housed in contextualized environments. One way of accomplishing this is to insert the images into a geographical browser similar to Google Street View. The viewer can then explore the environment around the location of a particular image set for a spatial comprehension of the images. To gain temporal context, one could add time navigation to the browser, for example with a timeline control that shifts the time period in which images are displayed. In this way, one would only see images that were acquired at certain times. Such a system is presented in this thesis.

3.1 Approach

In order to prove that context derived from metadata can enrich image content, it is necessary to create a 3D contextualized active viewing system as a proof of concept prototype and a simple passive image viewing system for comparison purposes. The two demonstration systems will henceforth be referred to as the active system and the passive system respectively. It is important to note that it is not the goal of this thesis to create a commercial image viewing product. Therefore, all systems created for this thesis consists of only the essential components to support my hypothesis. The demonstration system's setup is composed primarily of three blocks; the main active viewing system, a passive viewing system for comparison, and a server where both systems access images.

The active viewing system attempts to display images in a natural and fluid way to enrich information absorption of pictures. My approach is to program a 3D environment where shared images are housed. The environment must be interactive so viewers can explore the images selected via three axes of information; namely when and where the images were taken and who took the images. I believe that users viewing images enhanced with metadata will allow them to learn the content of each image more efficiently compared to passive viewing methods.

To test the effectiveness of my active system, a passive image viewing system is also created for comparison purposes. This system simply displays images from different social media users alongside passively displayed contextual information to mimic the typical image viewing component of social media websites such as Facebook. It is not viable to simply use a readily available social media website as my passive display platform. This is due to the fact that they contain other components, such as advertisements and continuously updated news feeds.

These extra components pull users' attention away from the task of viewing images, resulting in an unfair test when compared to the active system. Although it is possible to add advertisements and newsfeeds to the active system for comparison to social media systems, measuring how distracting extra components are to users is difficult as they may impact individuals differently, potentially skewing results.

Images retrieved by both active and passive systems are located on a central server. Images are uploaded to the server via a computer with relevant metadata filtered from images and saved in a lookup table. Images stripped of metadata are then stored in the image database. When images with a specific metadata property are needed, the lookup table can be scanned and corresponding images can be downloaded by both active and passive systems. The system data paths are shown below in Figure 3.1.

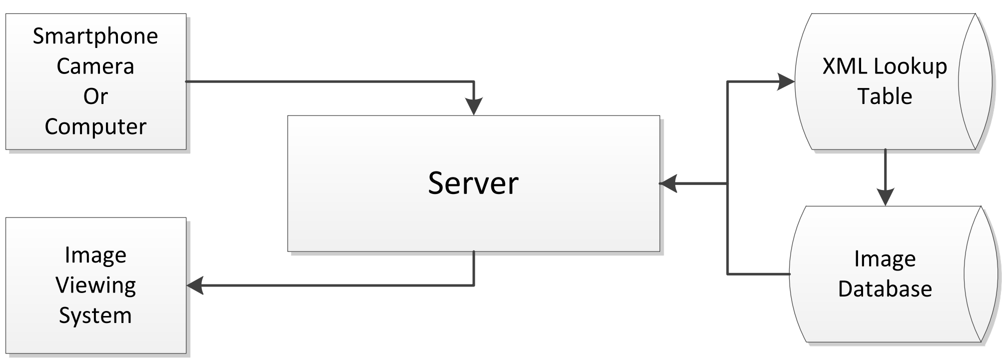


Figure 3.1: System data path

Both systems must encompass secondary design objectives. Due to client and server side limitations, both active and passive image viewing systems should embody secondary design criterion which was used as guidelines illustrated below.

1) **Fluidity:** The system should be able to run smoothly even if unrelated background applications are running. This is especially important on mobile systems.

2) **Scalability:** Since a social network environment is involved, users may upload as many images as they wish. The system will be required to accommodate this increase of information.

3) **Gesture Adaptability:** In the case that a certain gesture or action to navigate between images is not preferred by a user, alternate gestures should be readily available.

4) **Minimize storage load on system:** The system should retrieve images online to allow for easy posting and retrieval of data.

5) **Minimize computation load on the system:** The system should be simple and fast as to not slow down when multiple images are simultaneously required to be displayed.

3.2 Hardware

For the demonstration systems, an Apple iPhone 3GS, iPad 2, and MacBook Pro were used to take pictures and display them. Images to be displayed were taken by an iPhone 3GS camera. A smartphone was used as a camera was because of the automatic metadata collection capabilities of the phone using the numerous sensors onboard the device. Images are taken as a JPEG format with metadata stored as Exchangeable Image File (EXIF) data. EXIF data can be easily extracted after the iPhone has been connected to a computer running an EXIF extraction program (Renaud, 2012). The iPhone image properties are displayed in Table 3.1.

Camera Properties

Sensor size	3.15 Megapixels
Image Resolution	3048 X 1536
Fixed Aperture	f/2.8
Focal Length	3.85mm

Image data Collected

- **Longitude Latitude**
- **Date**
- **Compass Heading**
- **Image Angle**

Table 3.1: Related camera and image properties.

The active system is an active geographical social networking interface located onboard an Apple iPad 2 as an IOS application. The iPad was chosen due to its large array of sensors but also for its 3D OpenGL capabilities. The active system requires viewers to interact with a 3D environment which requires the system to have real-time data input and 3D rendering. The iPad has both WIFI and 3G capabilities so it can download needed images quickly and render them in the environment accordingly. The application could potentially have been used on the iPhone, but the small screen size was considered to be a hindrance for effective image viewing. The relevant specifications of the iPad 2 are located in Table 3.2.

Display	**1024 X 768 Resolution**
Storage	16GB
WIFI	802.11a/b/g/n
Sensors	• Digital Compass • 3-Axis Gyro • Accelerometer • Multi-touch Surface

Table 3.2: Relevant iPad 2 specifications.

Both active and passive viewing systems were programmed on an Apple MacBook Pro using Objective C and HTML programming languages respectively. The User Interface (UI) for carrying out psychophysical testing was also programmed in Visual C# with the MacBook and used during test sessions. An Apple computer was used because applications for Apple mobile devices can only be programmed with Apple computers.

3.3 Software

The demonstration systems consists of 3 main blocks as described in Section 3.1. The construction of all three blocks is presented in the next sub sections.

3.3.1 Image Server

The use of an online server to store context aware images is employed for quick access of image data across the world. However, before the images are uploaded and stored, metadata is first extracted from each image offline. The extracted information is archived in an XML file serving as a look up table for selecting images of interest without downloading the entire database. The metadata extractor written in C# by Ferret Renaud (Renaud, 2012) was used to extract image metadata as an XML file. A reduced sample of the output XML file is shown below in Figure 3.2.


```
        <?xml version="1.0" encoding="ISO-8859-1" ?>

<metadataExtractor nbFile="1">
<file name="C:\Program2\ImagesStrip\ImagesStrip\bin\Debug\image\photo.JPG">
        <directory name="Exif Makernote">
                <tag type="0x010F">
                <tagLabel>Make</tagLabel><tagDescription>Apple</tagDescription></tag>
                <tag type="0x0110">
                <tagLabel>Model</tagLabel><tagDescription>iPhone 3GS</tagDescription></tag>
                                                <tag ty..
                                              </directory>
        <directory name="GPS Makernote">
                <tag type="0x0001">
                <tagLabel>GPS Latitude Ref</tagLabel><tagDescription>N</tagDescription></tag>
                <tag type="0x0002">
                <tagLabel>GPS Longitude Ref</tagLabel><tagDescription>W</tagDescription></tag>
                <tag ty..
        </directory>
        <directory name="Jpeg Makernote">
                <tag type="0x0000">
                <tagLabel>Data Precision</tagLabel><tagDescription>8 bits</tagDescription></tag>
                <tag type="0x0001">
                <tagLabel>Image Height</tagLabel><tagDescription>2448 pixels</tagDescription></tag>
                <tag type="0x0003">
                <tagLabel>Image Width</tagLabel><tagDescription>3264 pixels</tagDescription></tag>
                <tag ty..
        </directory>
</file>
</metadataExtractor>
```


Figure 3.2: Sample of extracted EXIF data in text form

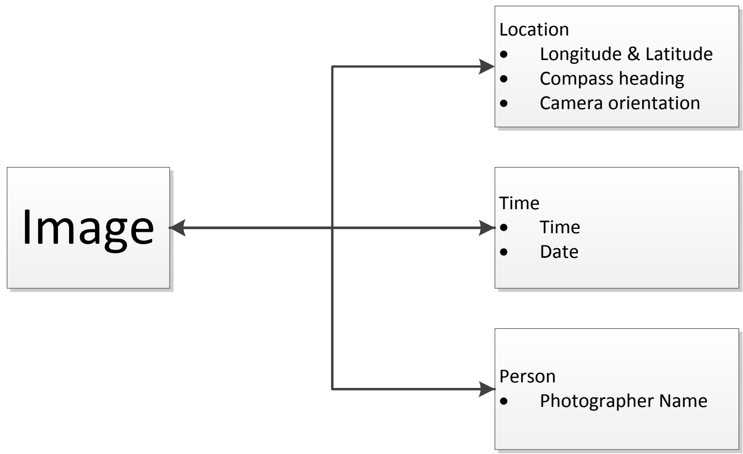


Figure 3.3: Image metadata of interest

Knowing the needed types of image metadata shown in Figure 3.3, the relevant labels of XML data was extracted using an XML text reader. When users upload images, their user ID is also included as an XML entry. XML values are then entered into the xml archive used as a lookup table. A sample of a new image entry into the XML file is shown below in Figure 3.4.

```
<file name="\image\newPhoto.JPG">

<directory name="Location">
        <tag><tagLabel>LatitudeRef</tagLabel><tagDescription>N</tagDescription></tag>
        <tag><tagLabel>Latitude</tagLabel><tagDescription>45,30.49,0</tagDescription></tag>
        <tag><tagLabel>LongitudeRef</tagLabel><tagDescription>W</tagDescription></tag>
        <tag><tagLabel>Longitude</tagLabel><tagDescription>73,34.64,0</tagDescription></tag>
        <tag><tagLabel>ImgDirection</tagLabel><tagDescription>30.1709</tagDescription></tag>
        <tag><tagLabel>Orientation</tagLabel><tagDescription>1</tagDescription></tag>
</directory>

<directory name="Time">
        <tag><tagLabel>DateTime</tagLabel><tagDescription>2012:03:18 4:22:28</tagDescription></tag>
</directory>
<directory name="Photographer">
        <tag><tagLabel>User</tagLabel><tagDescription>John Doe</tagDescription></tag>
</directory>
</file>
```

Figure 3.4: Sample of new image database entry

There is only one XML file located on the server, which contains metadata for all images in the database. Any image viewing device may access and read the XML file located at the Uniform Resource Locator (URL):

http://cim.mcgill.ca/~VNG/imageDatabase/LookUp.xml

Images of interest can be found by scanning metadata information for each image entry in the database and then retrieving corresponding file names. Files can then be downloaded by inputting the file name into the following URL:

http://cim.mcgill.ca/~VNG/imageDatabase/FileName

Initial plans were to automate the file uploading process but due to time restrictions, image uploading with metadata was uploaded manually.

3.3.2 Active Viewing

The main active viewing system uses metadata to form a contextualized 3D environment allowing users to interact with it. The interface allows user to navigate between images seamlessly because all images are shown simultaneously in a 3D virtual world. Besides viewing images by virtually navigating through a 3D world, users can also select which images to view by selecting photographers and time range. This continuous navigation between images can organize clutter and also allow users to develop cognitive connections as one navigates the interface. For example, by selecting only one photographer and a point in a timeline, users will be able to know which images are taken by the selected photographer including the sequence in time which the images were taken. Finally, navigating the virtual 3D panoramic world would spatially connect the images. It is worthwhile to note that whatever function a user wants to perform, he should be able to find it instantly (Dubuis et al., 1995) without the use of menus or complex actions to maintain the usability of this system. To achieve this, axes of contextual information used to navigate pictures, are designed as separate layers of the interface enabling users to learn features in a meaningful sequence (Shneiderman, 2009). A screen shot of the active viewing system is shown in Figure 3.5.

Figure 3.5: Sample of active image viewing system.

Users can view images located in a 3D panoramic environment by picking up the device (e.g. iPad or smartphone) and physically point at different directions by using the iPad as a portable window, or by simply swiping a finger across the screen to change the viewing direction. The primary way to navigate the 3D environment is the virtual window method. It is difficult to correctly perceive 3D object position and orientations on an unmoving flat screen (Houde et al., 1992). Therefore, moving the surface physically in the real world can allow users to apply 2D virtual information to the 3D reality. In case users become fatigued using the virtual window method, touch gestures are also implemented to navigate the active system. Making use of the ability for such gestures tap into a very basic part of the human syntax (Yee, 2009) and require little to no training to use. Both input methods are gesture based methods involving only the iPad, thus saving time and effort in using conventional pointing devices (Cabral et al., 2005).

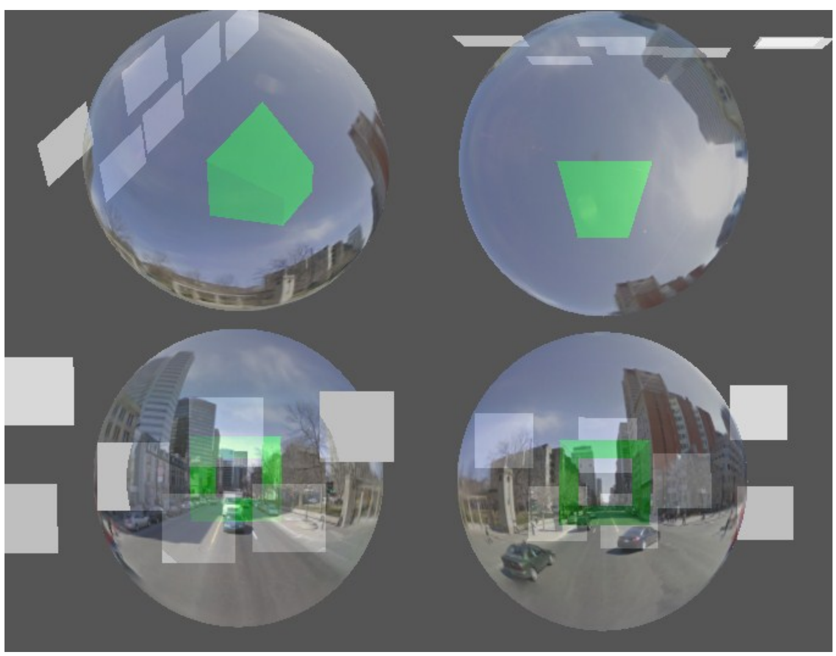


Figure 3.6: 3D environment structure of the active viewing system. The green object represents the camera surrounded by a spherical panoramic shell. Images are represented by the grey rectangles on a perpendicular plane with respect to the camera.

The 3D environment of the active system is programmed in XCode 4 for the Apple iPad 2. The environment consists of a panoramic sphere surrounding the camera with images located on a plane perpendicular to the camera's viewing direction as shown in Figure 3.6. Although the optimal method to simulate a 3D environment is to model physical locations in virtual space, using a panoramic sphere reduces the rendering load on the iPad while providing an appropriate level of detail (Chen, 1995). The 3D sphere is textured with a spherical panoramic image and rotates around its center where a camera, represented by a green trapezoidal prism, is fixed. When a different viewing direction is determined, the panoramic globe rotates to the new direction whereas the camera does not move. The viewing direction, which includes compass heading and pitch angle, can be changed by either tilting the iPad or touching the screen and physically dragging the viewing direction. The panoramic globe is programmed to move in the exact opposite of the selected viewing direction. If the user would like to turn 15 degrees to the right and 30 degrees up from the horizon, the sphere would rotate 15 degrees to the left and rotate 30 degrees down from the horizon. This achieves the illusion that the user is looking around a fixed virtual world without actually moving the camera in 3D space.

The spherical panorama in Figure 3.6 is retrieved from combining two 512X512 images retrieved from the Google Street View database and wrapped around a 3D globe as a texture. To retrieve the spherical panorama of selected GPS coordinates, the panorama identification code must first be retrieved from the Google servers. To do so, an XML file is retrieved by adding longitude and latitude coordinates of a POI and sending the link:

http://cbk0.google.com/cbk?output=xml&ll[latitude, longitude]

An XML file will then be returned from Google. One of the parameters within the XML file is the panorama id located under a property called "pano_id". A sample XML file with a panorama identification code highlighted is shown below in Figure 3.7.

```
<data_properties image_width="13312" image_height="6656" tile_width="512" tile_height="512"
image_date="2008-03" pano_id="sLaiF6Jex7mJmNol7tdOoA" num_zoom_levels="3" lat="51.495078"
lng="-0.147004" original_lat="51.495073" original_lng="-0.146987">
  <copyright>© 2012 Google</copyright>
  <text>Eccleston Place</text>
  <region>London, England</region>
  <country>United Kingdom</country>
</data_properties>
```

Figure 3.7: XML file containing panorama identification code.

With the correct panorama id, the two images can be retrieved as JPEG images with the following two image requests:

```
http://cbk0.google.com/cbk?output=tile&panoid=[pano_id]&zoom=[1]&x=[0]&y=[0]
http://cbk0.google.com/cbk?output=tile&panoid=[pano_id]&zoom=[1]&x=[1]&y=[0]
```

The two images retrieved are the two halves of a spherical panorama. After combining the two images together horizontally as one single image, the image is ready to become a texture for the panoramic sphere in the 3D environment. A sample of a spherical panorama retrieved from Google of the McGill University campus is show below in Figure 3.8.

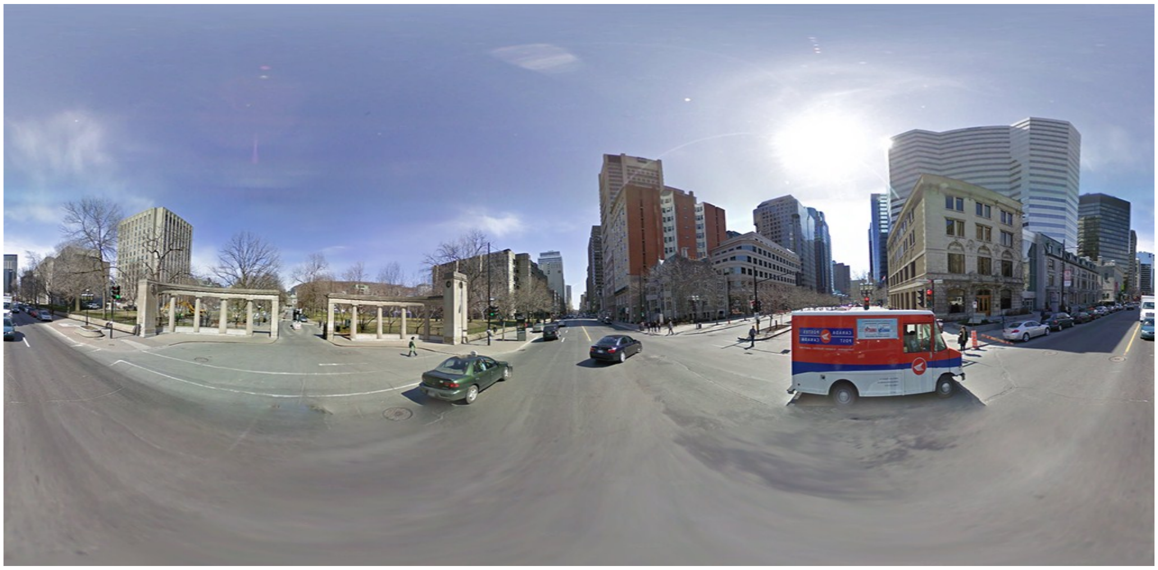


Figure 3.8: Sample of a spherical panorama retrieved from Google Street View databases.

Although distorted as a rectangular picture, Figure 3.8 contains enough data to reconstruct a spherical view of a location when wrapped around a 3D sphere using UV-mapping methods. UV-mapping conversion equations map 2D points to 3D spherical location as shown below:

$x = R * \sin(u)\cos(v)$ — R = radius of sphere

$y = R * -sin(u) * sin(v)$ — x,y,z = 3D coordinates of sphere

$z = R * cos(u)$ — u,v = 2D coordinates of panoramic image

When every (u,v) point on the texture is located to every (x,y,z) point on the sphere, the mapping is complete allowing a 2D texture to wrap around a 3D sphere as shown below in Figure 3.9.

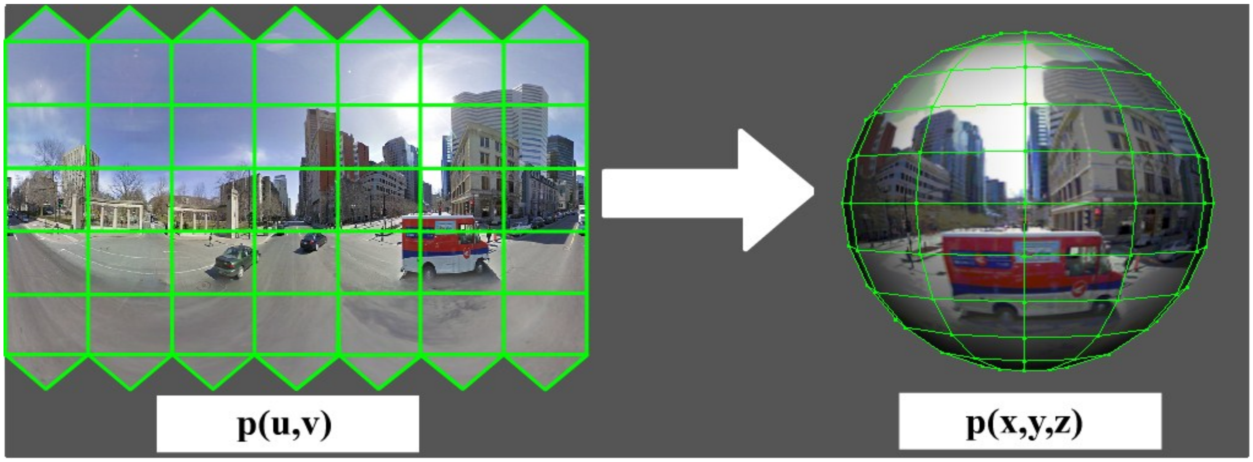


Figure 3.9: Example of UV mapping where every (u,v) point on the grid shown on the left is mapped to an (x,y,z) point on the sphere shown on the right.

Although the spherical panoramas are all retrieved from the Google Street View database, potentially any spherical panoramic picture with a 1024 X 768 resolution can be used as a texture. IOS applications such as Microsoft's Photosynth can be used to generate custom panoramas of any location and then directly imported into the active system.

User images to be viewed are located on a plane perpendicular to the camera's viewing angle in 3D space. Although the panoramic globe's motion is purely rotational, the motions of image sets are translational in the X-Y plane. The plane's location is situated perpendicular to the camera at all times as shown in Figure 3.10.

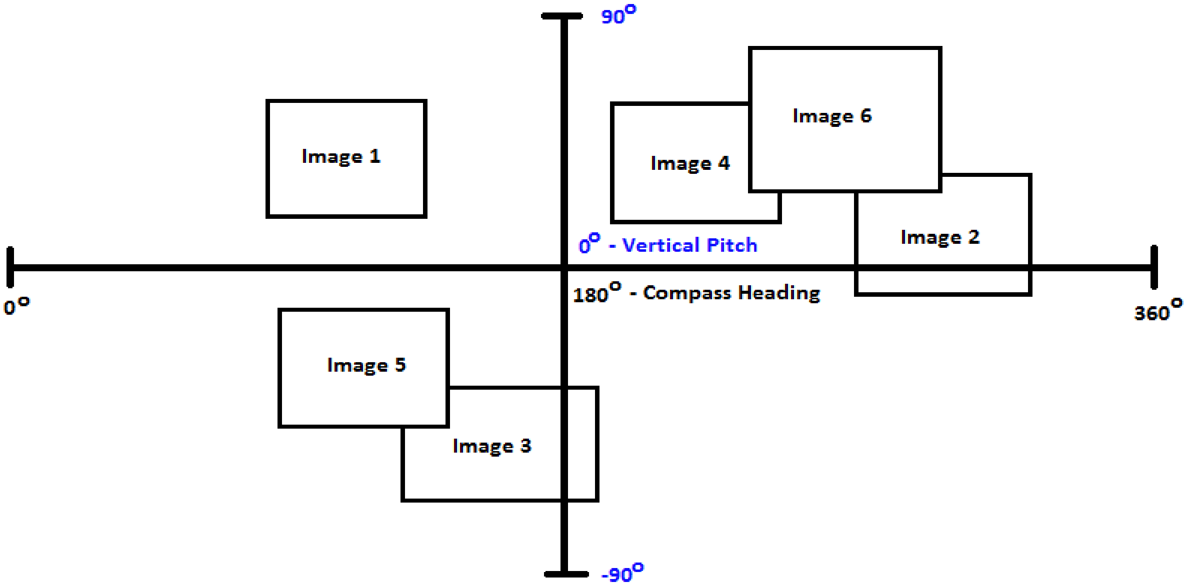


Figure 3.10: Demonstration of image plane.

Each image is a rectangular surface with user images as textures for each surface. Images are displayed only if the image is visible from the user's currently selected viewing scope. Images outside the camera's viewing scope will not be drawn in 3D space to reduce the load on the system. Although visible images are moved horizontally and vertically with respect to the camera, the effect of images rotating with the viewer is achieved when the user is looking in different directions. This is possible due to its proportional mapping with the coordinate system which also controls the panoramic sphere. Images contain two types of positional information, vertical pitch and compass heading. If the user changes the viewing angle to 15 degrees to the right and 30 degrees up from the horizon, all images within viewing range will move 15 units left on the x-axis and 30 units down on the y-axis. This coordinate system is shown in Figure 3.10. Each image changes its X-Y coordinates by contrasting its compass heading and pitch metadata to the compass heading and pitch the user has selected. The reason why images on the X-Y plane are restricted to only translational movement is to avoid image collision. If images were mapped and rotated with the panoramic globe, the same effect could be achieved. However, images close to each other could potentially collide reducing the effectiveness of the interface as discussed in Section 2.1.2. By formatting all images in a 2D plane perpendicular to the camera, all images are parallel to each other and thus image collision is impossible.

Since images are located at different distances from the camera, images must be appropriately sized to maintain the correct spatial representation of the 3D environment. Correct spatial representation ensures that all data presented is distinguishable even if displayed all at once (Travis, 1995). The size of each image is scaled inverse proportionally to the difference in distance between the GPS coordinates of the image and the coordinates of the viewing location. This results in the shrinking of an image if it is located far from the viewer and enlargement if images are close to the viewer. Due to the error in image pitch values from metadata, pitch values were re-entered by hand. The error in pitch comes from the error prone gyroscope onboard the iPhone (Schindhelm et al., 2011). This results in offset images if the pitch values stored in the metadata was used to directly position images without adjustments. Image display variables for position and size are shown below.

$$\text{X} = (\text{Image compass heading}) - (\text{User Compass Heading})$$

$$\text{Y} = (\text{Image Pitch}) - (\text{User Pitch})$$

$$\text{Size} = \frac{\alpha}{\sqrt{((\text{Image Longitude} - \text{Camera Longitude})^2 + (\text{Image Latitude} - \text{Camera Latitude})^2)}}$$

Besides being able to control the viewing direction, users can also select and deselect images based on the relative times they were taken. Automatic temporal synchronization and alignment of images can be achieved when all images are controlled via a single timeline (Dionisio et al., 1996). A scroll bar located at the bottom of the screen and the circle slider represents a relative point in time. If the circle slider has been moved, all images taken after the selected time will be hidden whereas images taken before the selected time will remain visible. Users manipulate the slider by simply dragging the circle slider left and right; left for older images, right for newer. Newer images are offset slightly closer to the viewer and older images are located further in 3D space. Sliding the time divider from right to left achieves an effect of "peeling" back time. Images are sorted and spaced equally on the timeline. This is to avoid small changes in time selection hiding many images taken within a small time frame.

Images can also be selected or deselected via photographer ID. Buttons located just above the time scroll bar can be tapped with a finger to select or deselect social media friends as seen in Figure 3.5. If a social media friend is deselected, images taken by that individual will be hidden and vice versa.

In the active display system, spatiotemporal metadata has been actively displayed via 3D navigation, timeline, and photographer identification. The three axes of information have been formatted to allow for users to browse images in a seamless manner with the purpose of enhancing image content absorption.

3.3.2 Passive Viewing

For comparison purposes, a traditional social networking interface, which henceforth will be referred to as the "passive viewing system", has been implemented. Passive viewing is defined as displaying images with their metadata in its raw form. Time metadata displayed via a time stamp simply printed below a selected image is an example of passively displaying contextual information. Information displayed such a manner can easily be unnoticed by a viewer. Although the passive viewing system is very different from the active system, the interface must contain the same contextual information which includes data on location, time, and photographer ID for each image.

The passive viewing system was made to mimic a typical social media website like Facebook which leads to its ironic name, "Fakebook". As mentioned before, simply using a readily available website such as Facebook is not a viable option. Readily available social media websites contain many components such as sponsored ads, comments from friends, event invitations, etc., which detracts attention away from the action of purely viewing images. Therefore the construction of this mock social media site, Fakebook, is necessary to create a social media website only containing the essential image viewing components that can be fairly contrasted with the active demonstration system. Fakebook consists of three web pages, one for each user in a simulated social media group. Each page consists of five thumbnails which link to the original image stored on the image server. Each website also contains the profile pictures of two other social media friends which link to their corresponding web pages. Users can view five images from each profile and navigate between such friends. The web pages were programmed in HTML and the entire site is located on an online server at the URL:

http://www.cim.mcgill.ca/~VNG/FakeBook/

All images on the Fakebook sites are displayed with three categories of metadata, namely location, time, and photographer. As all 15 images in the passive system are of the same location, viewers are simply told images presented are acquired from near to particular location in Vancouver, British Columbia. GPS coordinates were initially printed underneath the image, but such information would be not usable for users without a digitized map. Each image viewed contains a date stamp shown at the bottom of the picture. Finally, the photographer of an image is implicitly known because the images are shown and accessed through the corresponding photographer's Fakebook profile.

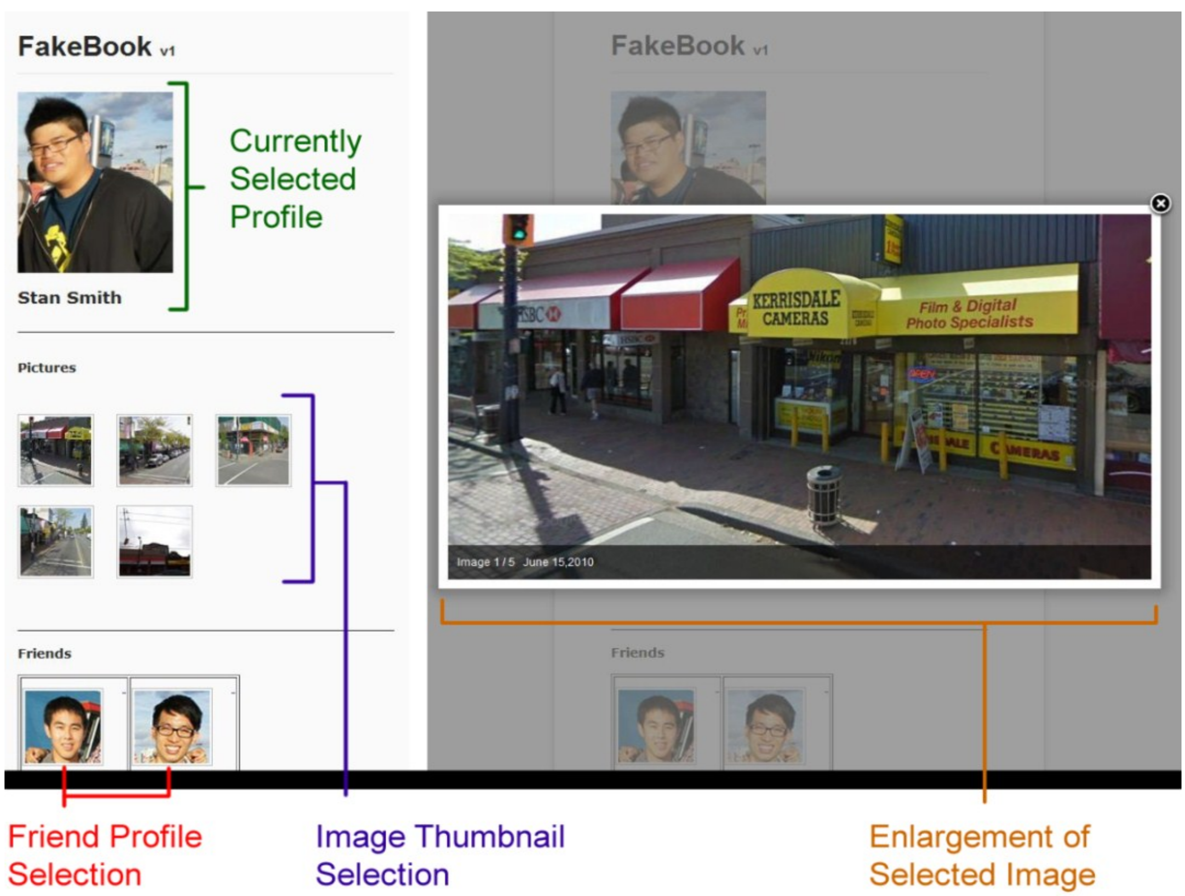


Figure 3.11: Sample page of the passive image viewing system.

An example of the passive image viewing system is shown above in Figure 3.11. Each webpage, as shown on the left half of Figure 3.11 contains the selected photographer's profile picture, his picture thumbnails and his social media friends' profiles. Once a thumbnail is selected, the website becomes shaded grey and the selected image is blown up for viewing, as shown on the right half of Figure 3.11. Users are able to cycle through all of the images in a profile by pressing the left and right arrow keys of the keyboard.

Chapter 4
Validation of Both Systems

4.1 Psychophysical & Usability Testing

With both active and passive systems developed, the next step in this thesis is to validate the hypothesis that active exploration of imagery in a spatiotemporal context improves upon the understanding of content over the traditional passive social networking media presentation. As mentioned in Section 2.2, one of the most effective ways to test an interface is to use a hybrid test environment with psychophysical and usability testing components. The process of using psychophysical testing methodology calculates statistics to quantify the effectiveness of a system whereas the recording of user testimonials quantifies user experience.

The psychophysical testing environment consists of 3 sets of 43-question quizzes for each test session on a specific system. All quizzes are identical in terms of format and contain the same number of questions for each question complexity class. Since each test subject will be asked to validate both active and passive systems, subjects will need to complete 6 quizzes in total for the two test systems. Users were asked to complete each quiz after viewing 15 pictures on a test system for a total of 5 minutes and then complete a quiz within a 2 minute limit. Questions in the quizzes are only directed towards the image content of the 15 pictures. This continues until all three quizzes for a single test system is completed. The test subject will then repeat the same process for the remaining system. Half of the test subjects began with the passive viewing system and the other half began their test session with the active viewing system to remove any bias effects of testing. The two systems were referred to as "System A" for the first system tested and “System B” as the second system tested”. Each quiz consisted of 43 multiple choice questions with a total of 6 different question complexity classes. The classes are illustrated below in Table 4.1 and a table of all questions categorized by class is located in the Appendix.

Question Category	Question Frequency	Question Sample
Class 1: Existence	7/43	Was there a picture of a car?
Class 2: Complex Existence	13/43	Was there a picture of a blue car?
Class 3: Location	8/43	Is the dollar store located across the street from the bank HSBC?
Class 4: Photographer	6/43	Did Stan take a picture of a fire truck?
Class 5: Time	3/43	Were most images taken in the summer?
Class6: Complex Scenario	6/43	If you wanted to buy food at a fast food restaurant but needed to withdrawal money, can you get to a bank by crossing two streets from the fast food restaurant?

Table 4.1: Quiz question categories and examples.

Each quiz consists of 43 questions with increasing complexity as questions are answered. To increase the complexity of questions gradually, users would answer class 1-6 questions in order. Existence type questions were used to test if users could recognize objects within images which verify that users were paying attention to image content. Results for existence questions are essential to judge if test subjects were engaged in the testing process. This can be used to filter out subjects who were not interested in the test, potentially skewing test results. Besides simply acknowledging that objects were present in images, complex existence questions were asked in order to gauge the level of detail users have learned in terms of image content. To test the effectiveness of the three axes of metadata separately; location, photographer, and time questions were utilized to test how effectively users can recall image content using contextual information. Finally, all test categories attributes are merged to form the complex category questions. This type is used to judge how deeply users have internalized information and how effectively users can use the newly learned information

Each quiz is evaluated with the following equations:

$$\text{Accuracy} = \frac{\text{Correct Answers}}{\text{Questions Answered}}$$

$$\text{Comprehensiveness} = \frac{\text{Seconds to Complete Quiz}}{\text{120 Seconds}} \quad \text{or} \quad \frac{\text{Questions Answered}}{\text{43 Questions}}$$

Users are ask to complete quizzes for both active and passive system and rated in terms for Accuracy and Comprehensiveness to see on which system they perform better. For each quiz, the 'accuracy' is calculated to judge how accurate the user is in answering questions. Users who achieve a higher score for a particular system simply have learned image content better for that system and vice versa. The 'comprehensiveness' result judges how comfortable the user is with the information learned. If a user completes quizzes on system A faster than system B, then they were more comfortable with answering questions for system A. There are possible cases, however, where users are not able to complete the quiz in the allotted 2 minute time limit.

Therefore, if users have completed the quiz, their comprehensiveness result is calculated via the time taken to finish the quiz. Whereas if users have not completed the quiz, comprehensiveness is calculated by seeing what percentile of the quiz has been completed.

The scores for each quiz across all test subjects are calculated individually and then pooled. Effect size measurements (described in Section 4.4) are calculated to observe if the sample size is high enough to yield significant results. Finally, paired, two-tailed significant tests are calculated to provide statistical confidence values in hopes to support the hypothesis that context derived from metadata enriches image content. It is important to note that any claim must have at least a 95% confidence value to hold statistical merit.

Besides calculating numerical results, users' testimonials were also recorded after all quizzes were completed. Test subjects were asked to verbally compare and contrast active and passive systems against each other. Since both demonstration systems are focused prototypes with limited purpose (Houde et al., 1997), users were asked not to compare test systems with readily available commercial image viewing systems.

4.2 Passive Context Image Viewer Results

The test subject sample size consisted of 7 females and 8 males ranging in ages from 18 – 30 years of age. All participants are partaking in an academic program at McGill University. Although not a prerequisite, every test subject has voiced that they own and regularly use their personal Facebook profile site. Aggregate test results are calculated by pooling all test results and recording the mean and standard deviation for each of the 3 quizzes on both systems. The results are shown below in Figures 4.1 and 4.2.

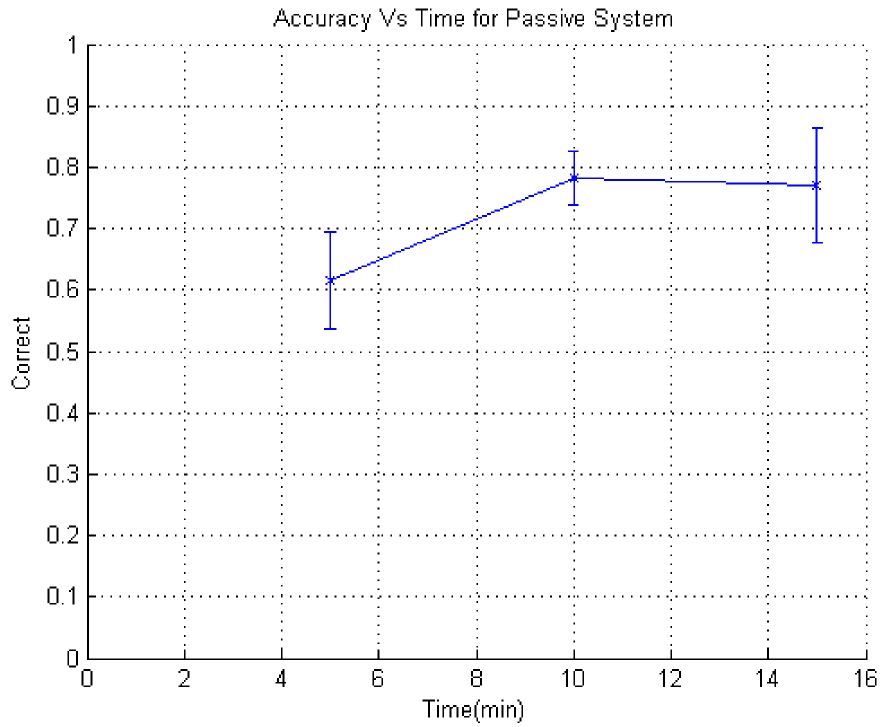


Figure 4.1: Passive viewing accuracy results with error bars indicating standard error.

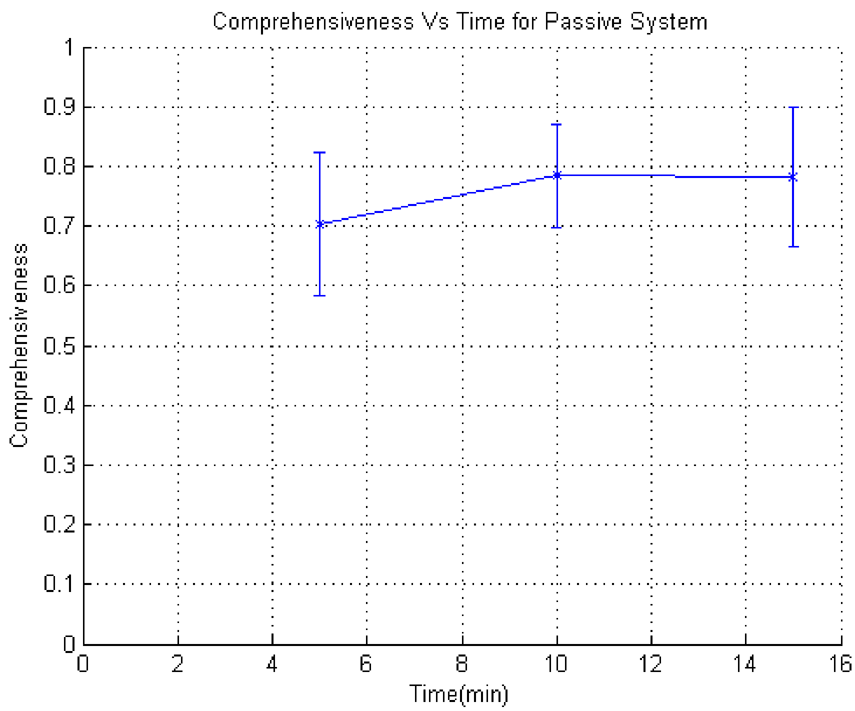


Figure 4.2: Passive viewing comprehensiveness results with error bars indicating standard error.

Figures 4.1 and 4.2 show the pooled accuracy and comprehensiveness as a function of time. Error bars indicate the standard error. The results show that both accuracy and comprehensiveness increase from the first test to the second. The claim for an increase of accuracy has a statistical confidence of 99.98% and an increase of comprehensiveness with a statistical confidence of 99.91%. This reflects the intuition that users become more familiar with image content as time spent on observing images increase.

No claims in the change for accuracy and comprehensiveness results for the 3rd quiz can be made since there is a statistical confidence of 88.32% and 37.07% that there was any change for accuracy and comprehensiveness values respectively from the 2nd to the 3rd quiz results. However, after reviewing comments from test subjects, as shown in the appendix located at the end of this thesis, many consistencies were observed. Test subjects claim that they have learned all the content in the images after the second viewing of the images. However, when the third quiz was given, most subjects voiced that they felt that the third quiz was the same difficulty as the previous two, but were not as efficient in recalling information.

4.3 Active Context Image Viewer Results

Besides the quizzes mentioned in Section 4.2, test subjects were also required to test the active image viewing system with yet another set of 3 quizzes. Test subjects were asked to view 15 images for a total of 5 minutes using the active viewing system then take a 2 minute quiz. Images were displayed at locations relative to the 360 degree panoramic view. Test subjects were asked to view images then proceed to a quiz for a total of 3 times and asked to give comments on the interface at the end of the testing. Graphs representing the pooled results for the active viewing system are shown below in Figures 4.3 and 4.4.

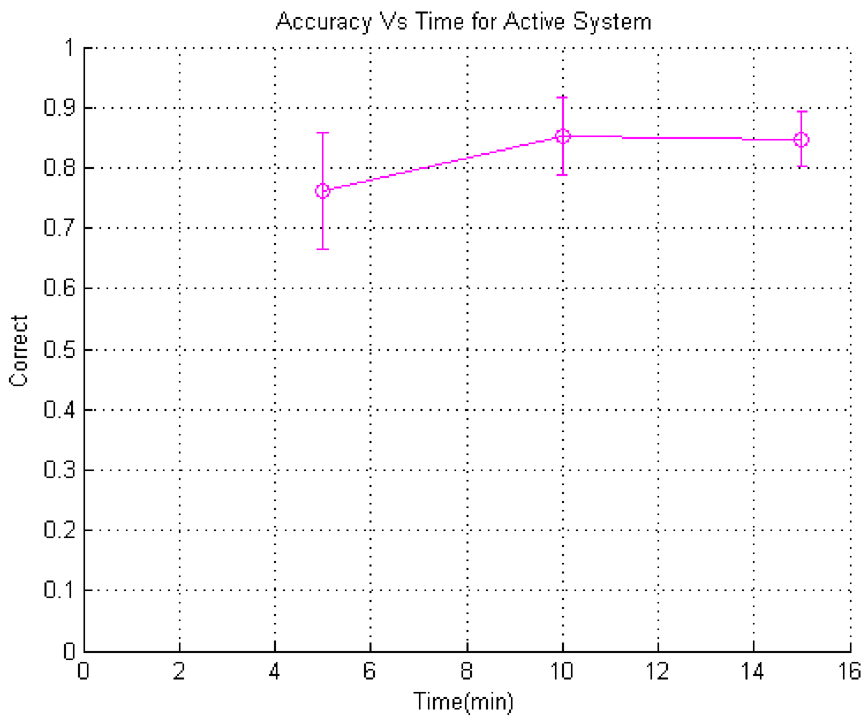


Figure 4.3: Active viewing system accuracy results with error bars indicating standard error.

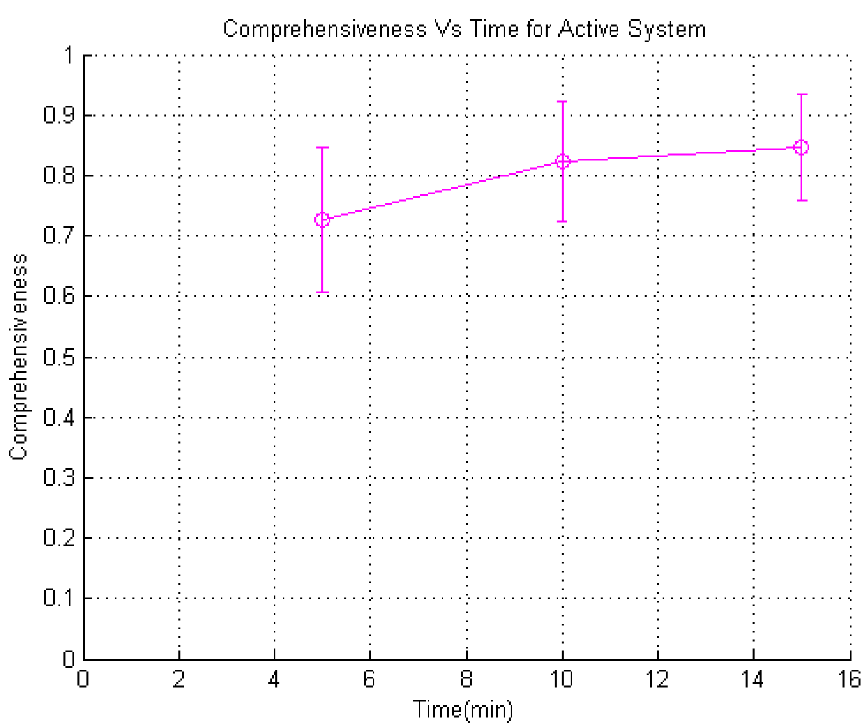


Figure 4.4: Active viewing system comprehensiveness results with error bars indicating standard error.

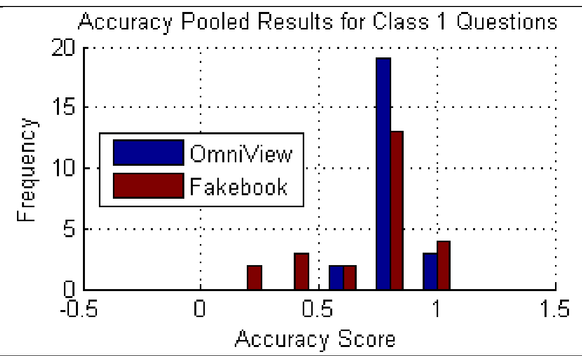
Test subject results in Figures 4.3 and 4.4 show improvement for both accuracy and comprehensiveness especially for the improvements between the first 2 quizzes with statistical confidence values of 99.94% and 99.57% respectively. Once again, no claims with statistical merit can be made for accuracy and comprehensiveness results between the 2^{nd} and 3^{rd} quizzes since the statistical confidence for improvements for accuracy and comprehensiveness is 11.50% and 47.92% respectively. The lack of improvement for the 3^{rd} quiz could either have been due to the error margins of the data, or user learning saturation. Saturation occurs when users can no longer absorb any more information from the image set.

Test subjects have commented that the active viewing system was extremely fluid, easy to use, and thus allowing them to learn image content quickly. Most test subjects have described the active viewing system as "a giant picture in which you could modify what you wanted to see". It was observed through testimonials that users felt very confident when answering the quiz questions for the active system. Some users even made the comment that the task of viewing images then answering quiz questions was "easy". Additional reoccurring testimonials from test subjects on the active system are available in the appendix.

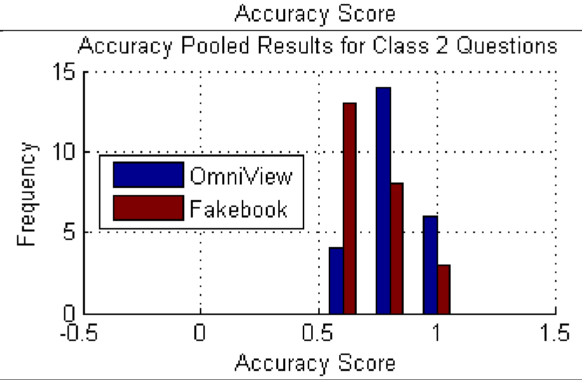
4.4 Compare and Contrast Between Systems

After all test results have been collected, it is possible to compare and contrast the results of both systems. Before aggregate results for both systems are contrasted, the relative scores for each question classification on both active and passive systems are analyzed. Relative scores for each question classification is used to see if a certain question classification has greatly affected the aggregate accuracy score presented in the previous sections. If inconsistencies are observed in performance between question class scores, then aggregate results cannot be compared to make high level observations as they do not represent the data correctly. The relative score is calculated with the following formula and shown in Figure 4.5:

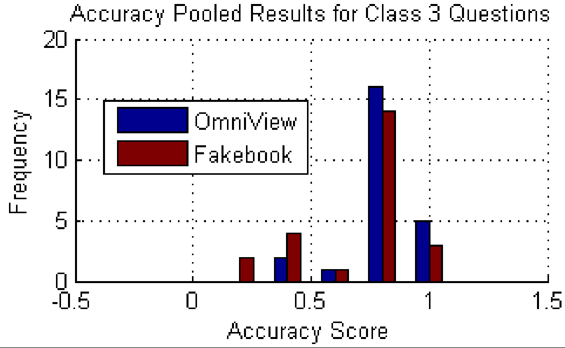
$$\text{Relative Score for Specific Class} = \frac{\text{Questions Answered Correctly for Specific Class}}{\text{Question Class Frequency}}$$



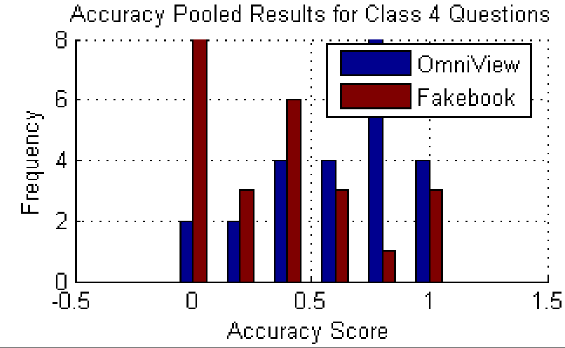
	OmniView	Fakebook
Mean	0.798	0.696
Std. Dev.	0.118	0.212
Statistical Confidence (OmniView > Fakebook)	95.60%	



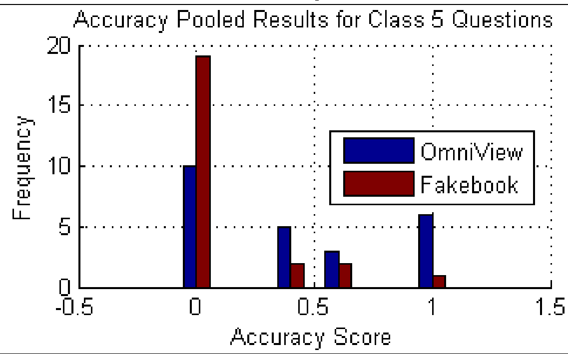
	OmniView	Fakebook
Mean	0.811	0.731
Std. Dev.	0.088	0.118
Statistical Confidence (OmniView > Fakebook)	99.83%	



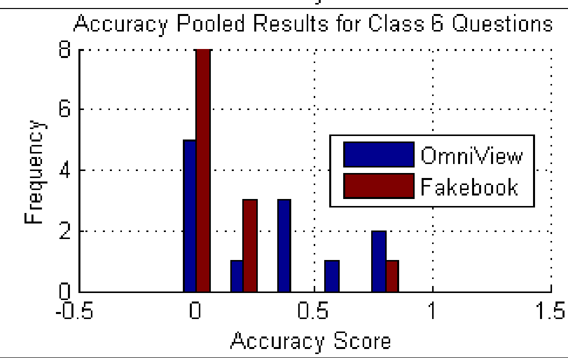
	OmniView	Fakebook
Mean	0.828	0.708
Std. Dev.	0.142	0.220
Statistical Confidence (OmniView > Fakebook)	99.56%	



	OmniView	Fakebook
Mean	0.639	0.375
Std. Dev.	0.317	0.359
Statistical Confidence (OmniView > Fakebook)	99.971%	



	OmniView	Fakebook
Mean	0.403	0.125
Std. Dev.	0.417	0.275
Statistical Confidence (OmniView > Fakebook)	98.80	



	OmniView	Fakebook
Mean	0.292	0.111
Std. Dev.	0.327	0.239
Statistical Confidence (OmniView > Fakebook)	90.97%	

Figure 4.5: Relative Score for each classification of questions on both active (OmniView) and passive (Fakebook) systems. The histogram contains 6 bins between normalized accuracy values ranging from 0 to 1.

As observed in Figure 4.5, question Classes 1-5 showed that accuracy for the active system is higher than the passive system with over 95% statistical confidence whereas question class 6 observed the same result but with a 90.97% statistical confidence. With an overall increase of performance for the active system over the passive system across all classifications of questions, it is thus possible to compare and contrast aggregate results of both systems to form high level observations.

The aggregate results for both systems are computed by averaging quiz accuracy and comprehensiveness scores as illustrated in Figures 4.6 and 4.7 below. Numerical results from Figures 4.6 and 4.7 have also been presented in Tables 4.2 and 4.3.

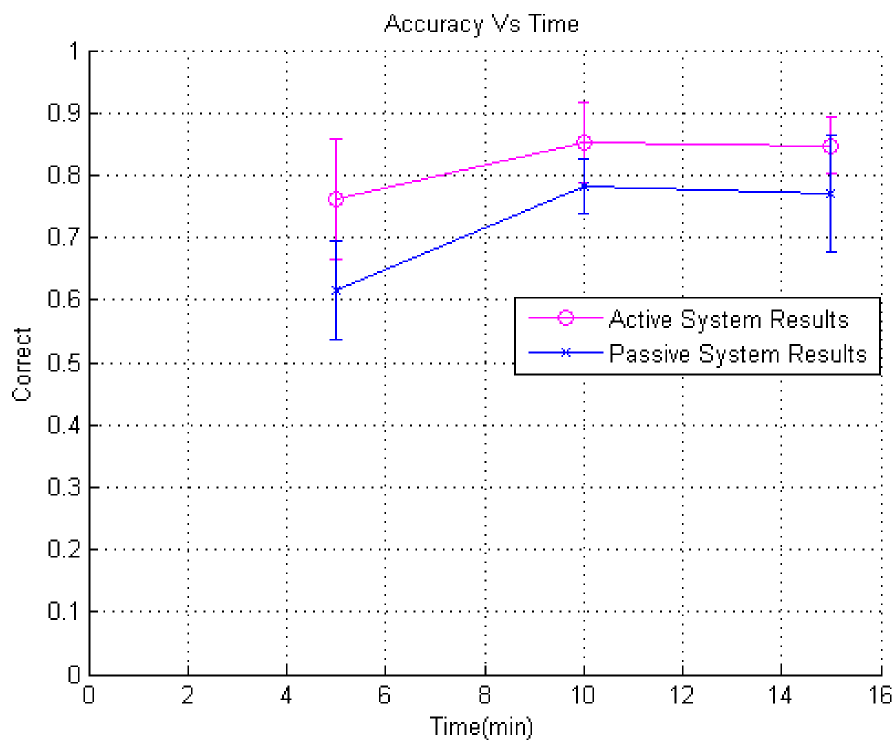


Figure 4.6: Accuracy results for both systems with error bars indicating standard error.

	Active Viewing System			Passive Viewing System		
	Quiz 1	Quiz 2	Quiz 3	Quiz 1	Quiz 2	Quiz 3
Mean	0.7617	0.8523	0.8479	0.6162	0.7812	0.7706
Standard Deviation	0.0965	0.0650	0.0441	0.0792	0.0438	0.0930

Table 4.2: Numerical values for accuracy.

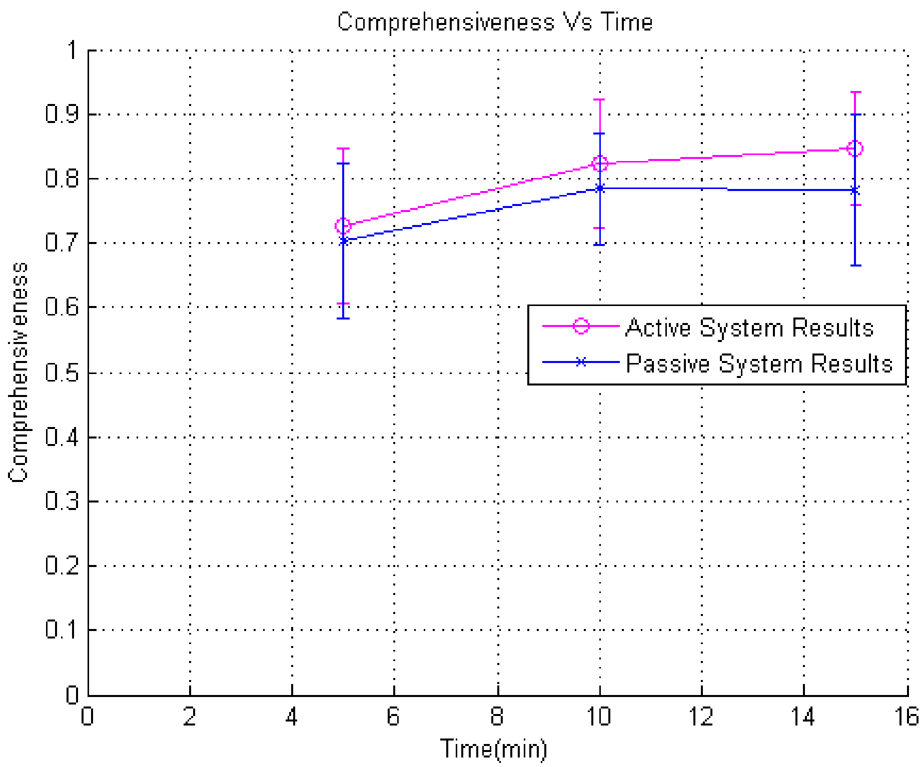


Figure 4.7: Comprehensiveness results for both systems with error bars indicating standard error.

	Active Viewing System			**Passive Viewing System**		
	Quiz 1	Quiz 2	Quiz 3	Quiz 1	Quiz 2	Quiz 3
Mean	0.7256	0.8233	0.8465	0.7023	0.7837	0.7814
Standard Deviation	0.1195	0.0988	0.0872	0.1200	0.0850	0.1172

Table 4.3: Numerical values for comprehensiveness.

Figures 4.6, 4.7 and Tables 4.2, 4.3 show that the active viewing system enables users to view and remember contents of images more accurately and comprehensively. Although the passive system results on accuracy are fairly high in terms of total percentile, all three points for the active viewing system score observably higher in comparison.

Since the error bars on the accuracy graph overlap, a significance test is required to judge how statistically significant the results are before any conclusions can be made. The Null hypothesis is defined to be that there is no change in the mean performance for both active and passive systems. Our test is a two-tailed paired significance test with an alpha value of 5%. If the p-value of the significance tests for all three quizzes falls below 5%, then the null hypothesis is false thus proving that my results are statistically significant. The significance test p-values for accuracy are shown below in Table 4.4.

Quiz 1	**Quiz 2**	**Quiz 3**

p-values	0.00243	0.005913	0.00957

Table 4.4: Significance test results.

As observed in Table 4.4, p-values for all three quizzes are well below 5% thus the test results presented are assumed to be statistically significant.

During the course of testing, a total of 15 subjects participated, but the results from 5 individuals were removed since confessions were made by the 5 individuals that they randomly answered questions. This resulted in only 10 test subjects contributing to the presented results. Although, statistical significance tests tells us the likelihood of the observed results due to chance alone, it is also important to calculate the effect size to determine if my statistically significant results are non-trivial and observable. Effect size values were calculated using the following Cohen's d formula (Cohen, 1988):

$$d = \frac{\overline{x_1} - \overline{x_2}}{s}$$

The d value is the difference of two averages divided by a standard deviation of data. If the d values are above 0.8, then the results are said to have a large effect and sufficient to establish my hypothesis. The d-values for all three tests are shown in Table 4.5 below.

	Quiz 1	**Quiz 2**	**Quiz 3**

d-values	1.6482	1.2823	1.0623

Table.4.5: Cohen's d effect size test results.

As observed in Table 4.5, all 3 d-values for the experiment are significantly higher than 0.8 which states that my results are non-trivial and viewable with the naked eye (Walker, 2008). Even though the sample size was lowered to 10 test subjects instead of 15, it was enough to yield statistically significant non-trivial results. It is important to note that although 10 test subjects are enough to sustain the statistical significance of this experiment in theory, attempts to increase the sample size were made but could not be achieved due to time and resource constraints.

4.5 Psychophysics and Usability Analysis

Psychophysical results have yielded statistically significant trends, but user testimonials must also be analyzed in order to explain why users perform better when using the active system as compared to passive system. After test subjects have completed quizzes for both systems, they were asked to give general comments on their personal comparison for both viewing systems. Reoccurring comments have been recorded and displayed in the appendix. As shown in Section 4.4, both Accuracy and Comprehensiveness graphs have yielded observably better results for the active system over the passive system. After reviewing documented comments by test subjects, it is evident that there are potential reasons for increases in both accuracy and comprehensiveness.

A trend in the user testimonials illustrated that users were able to bring forth forgotten image content more efficiently using the active system. If a question required the user to recall and relate image content from multiple pictures, users would normally answer incorrectly if needed content from one image has been temporarily forgotten. However, the active system has allowed users to recall forgotten information by cognitively reconnecting image content using the 3 axes of contextual information. Although both active and passive systems contain the 3 axes of contextual information, the active system allows users to relate image content with the use of contextual cues, whereas the passive system requires users to memorize contextual data and content for each image separately. Accuracy is measured by the total amount of correct answers per quiz over the questions answered. Therefore, if users can recall information better, they can answer more questions correctly thus increasing the accuracy. This can explain why the improvement in accuracy for the active system over the passive system is greater for the higher complexity questions illustrated in Figure 4.5. Class 1 questions only require subjects to recall content from single pictures whereas higher class questions require users to recall information from multiple pictures simultaneously. Although all questions types illustrate an increase of performance for the active system over the passive system, it is observable that the increase in accuracy for Class 5-6 questions is greater as compared to the increase in accuracy for Class 1-2 questions.

Test subjects demonstrated a general speed increase in taking active system quizzes over passive system quizzes. Users appear to treat the active system's environment as a single picture with small changes whereas the passive system was treated as a set of 15 separate images. Treating an image set as a collection of separate images hinders the efficiency of information recall since users would need to cognitively flip through all 15 images when answering each question for the passive system. As mentioned, comprehensiveness is measured by calculating the time taken to complete a quiz or the amount of questions finished at the end of the 2 minute time limit. As users take more time to recall information, they require more time to complete the quiz or they will have answered fewer questions when the 2 minute limit has been reached. This theory can explain the increase in comprehensiveness illustrated in Figure 4.7. By needing less time to recall information while answering more questions, users make fewer mistakes thus also increasing accuracy.

Chapter 5
Conclusion

In this thesis I have presented a method to actively display image content to encourage information absorption performance over the passive methods which most consumer image viewing interfaces employ. In the next section, limitations of the novel active interface will be presented along with possible improvements. Finally, a conclusion section summarizes the previous chapters and presents potential future applications for the active viewing methodology.

5.1 Overall Limitations and Possible Improvements

Although the active system presented enriches the content absorption of image content to a statistically significant degree, there are a few shortcomings which can be improved.

Upon construction of the active viewing system, camera orientation information extracted from images was found to be very inaccurate. To counteract this problem, the orientation angle was input manually into the database. A possible improvement is to automate this process using computer vision techniques. David Lowe's SIFT (Lowe, 2004) features can be employed to match the image with the panoramic background. The resulting data can be used to resize and position the image correctly to maintain accuracy in visual consistency of the environment. The location metadata could thus be used as an estimation and be fine-tuned via automated SIFT positioning algorithms.

The primary method to navigate the active system's 3D environment was to hold the iPad at arm's length and point it in different directions as if the device was a window to another location. Test subjects loved this functionality and voiced how innovative it was. Soon after however, many of the test subjects began to feel slightly nauseous. The use of the virtual window effect was to amplify the viewer's sense of direction to learn location context cues and image content more effectively. To avoid producing a group of nauseous test subjects, an alternative way to navigate the globe via touch gestures was employed so users can explore the environment by dragging a finger across the screen. Although dragging the globe was not as effective to utilize a viewer's sense of direction as compared to the virtual window method, the method would be unproductive if it made test subjects too ill to use the system. A possible explanation for nausea is the latency of the active system's 3D environment. A reduction in latency can be achieved by using better hardware for the onboard sensors and increasing frame refresh rates of the interface.

A comment most test subjects had on the active system was that they could not enlarge the images. Although the ability to resize image via touch gestures was present, it was disabled intentionally because the passive system used to mimic a typical social media site did not have image enlargement functionality. This was to ensure a fair comparison between active and passive systems for testing. However, since images were proportionally sized in terms of the distance from the viewer, far away images are relatively smaller than images taken close to the viewer. It would be greatly beneficial for viewers to have the ability to resize images via gestures to further increase the effectiveness of the active system.

Although the active system has many limitations as an image viewing application, it was designed as a proof of concept. With that regard, I believe my system was very successful in presenting the hypothesis that "context enriches content absorption" with statistical merit. My concept could potentially be implemented into future works or improve on present applications as discussed in the next section.

5.2 Future Implementations and Conclusion

This thesis presents the hypothesis that "context extracted from metadata enriches image content absorption." In my approach, metadata extracted from digital images was used to form a contextualized environment to house image content in hopes to improve users' ability to absorb such content. This contextual environment consists of a 3D panoramic social networking interface in which images were placed according to their relative location and orientation to the user's viewing position. This user interface was named the "active system". For comparison purposed, a "bare-bones" passive image viewing system constructed similarly to social media sites like Facebook or Flickr was created and nicknamed the "passive system". Both systems displayed a set of 15 images with three axes of contextual information namely location, time, and photographer. Although both systems displayed similar data, the method in which images were displayed was vastly different.

Test subjects were requested to participate in a hybrid psychophysical-usability test in order to gauge how effectively users were able to recall and use information learned from both systems. Subjects were asked to explore both image viewing systems and complete a series of 6 quizzes over strict time constraints. Their performance was evaluated in terms of accuracy and comprehensiveness. Accuracy is determined by how many correct questions the test subjects have answered whereas comprehensiveness is measured by how quickly users completed the quiz or how many questions were completed if the test subjects ran out of time. After quizzes were completed, subjects were asked to comment on their comparison of both systems and the results were compiled. Based on a statistical significance analysis on the psychophysical test results, it can be concluded that the active system has allowed users to absorb and utilize image content faster and more accurately than the more traditional passive system. Users' comments yielded the consensus that the active system encouraged faster image content recall and also provided multiple means to remember information as contrasted to the passive system. Results from both the psychophysical and usability tests show that my hypothesis is statistically valid.

Although the active viewing methodology has been applied to the proof-of-concept application of image viewing, active viewing can be potentially implemented for any situation in which a user is being presented information passively. The accuracy and speed in which people absorb information can be potentially enhanced whether in an educational setting or a professional environment.

The process of learning information passively from textbooks can be potentially improved by displaying the same information via active methods. Educational information in textbooks is typically segmented via chapters making it difficult to recall learned information as the learned information is also cognitively segmented. In a history book, if events of a certain time period have been forgotten, there would be no other cues to recall this information. However, if the history book has been formatted into a 3D environment with a timeline located on the horizon connecting all historical events, users can utilize their sense of direction to recall information even if the sequence of historical events has been forgotten. By using active display methodology in an educational setting, the benefits of improved accuracy and speed of information recall could also potentially be compounded school year after school year.

Actively viewing methodology can be applied to professional environments to improve productivity. In a professional work setting, high accuracy in produced work is usually a standard. However, high accuracy standards usually sacrifices time efficiency. For example, in a crime scene investigation setting, detectives must produce accurate judgments which can be costly in terms of time. Detectives usually passively learn information about a crime scene from a folder containing image clips and reports (Byrd, 2012). Such information can be formatted into a 3D environment so detectives could cognitively relate pieces of information without flipping through the entire folder every time. Although the detective would probably make the same conclusions regardless of display methodology, active display methods can potentially increase investigation speed which is essential to catch criminals.

Although the statistically significant hypothesis for this thesis is applied to the task of memorizing image content using contextual data extracted from metadata, my methodology can potentially be applied to situations which require users to learn information via passive methods. Regardless if informational content is being enriched via context within an educational or professional setting, an increase in information recall accuracy and speed is beneficial to every individual.

Glossary

EXIF	Exchangeable Image File Format
GPS	Global Positioning System
GSR	Galvanic Skin Response
HCI	Human Computer Interaction
JPEG	Joint Photographic Experts Group
PNG	Portable Network Graphics
POI	Point Of Interest
UI	User Interface
URL	Universal Resource Locator
XML	Extensible Markup Language

References

Ahern, S., Davis, M., Eckles, D., King, S., Naaman, M., Nair, R., et al. (2006). Zonetag: Designing Context-Aware Mobile Media Capture to Increase Participation . *Proceedings of the Pervasive Image Capture and Sharing: New Social Practices and Implications for Technology Workshop (PICS 2006) at the Eighth International Conference on Ubiquitous Computing (UbiComp 2006).*

Apple. (n.d.). Retrieved March 1, 2012, from Apple: http://www.apple.com/

Au, C., Ng, V., & Clark, J. J. (2011). Mirrormap: augmenting 2d mobile maps with virtual mirrors. *Proceedings of the 13th International Conference on Human Computer Interaction with Mobile Devices and Services.* New York, NY, USA.

Belimpasakis, P., Roimela, K., & Yu, Y. (2009). Experience Explorer: A Life-Logging Platform Based on Mobile Context Collection. *Next Generation Mobile Applications, Services and Technologies, 2009. NGMAST '09. Third International Conference on*, (pp. 77-82).

Byrd, M. (2012). *Written Documentation at a Crime Scene*. Retrieved 4 10, 2014, from Crime Scene Investigator Network: http://www.crime-scene-investigator.net/document.html

Cabral, M., Morimoto, C., & Zuffo, M. (2005). On the usability of gesture interfaces in virtual reality environments. *CLIHC* , 100-108.

Chen, G., & Kotz, D. (2000). *A Survey of Context-Aware Mobile Computing Research.* Darmouth Computer Science Technical Report, Dartmouth College, Department of Computer Science.

Chen, S. (1995). QuickTime VR: an image-based approach to virtual environment navigation. *SIGGRAPH* , 29-38.

Chippendale, P., Zanin, M., & Andreatta, C. (2009). Collective Photography. *Conference for Visual Media Production. CVMP '09*, (pp. 188-194).

Chippendale, P., Zanin, M., & Andreatta, C. (Nov 2010). Mobile Augmented Photography. *The 7th European Conference on Visual Media Production (CVMP 2010).* London, UK.

Chippendale, P., Zanin, M., & Andreatta, C. (2009). Re-photography and Environment Monitoring Using a Social Sensor Network. In *Image Analysis and Processing* (pp. 34-42). Springer Berlin/Heidelberg.

Chippendale, P., Zanin, M., & Andreatta, C. (July 2008). Spatial and Temporal Attractiveness Analysis through Geo-Referenced Photo Alignment. *International Geoscience and Remote Sensing Symposium (IEEE).* Boston, Massachusetts, U.S.A.

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences.* New York: Lawrence Erlbaum Associates, Publishers.

Dionisio, J. D., Cardenas, A. F., Taira, R. K., Aberle, D. R., Chu, W. W., McNitt-Gray, M. F., et al. (1996). A unified timeline model and user interface for multimedia medical databases. *Computerized Medical Imaging and Graphics* , 333-346.

Dubuis, E., Gilmour, D., & Waters, R. (1995). Graphical User Interfaces: Mess them up!

Facebook. (n.d.). Retrieved March 1, 2012, from http://www.facebook.com

Flickr. (n.d.). Retrieved March 1, 2012, from http://www.flickr.com/

Google Images. (n.d.). Retrieved March 1, 2012, from http://www.google.ca/imghp

Google Maps and Street View. (n.d.). Retrieved March 2012, from http://www.google.ca/maps

Hakansson, M., Ljungblad, S., & Holmquist, L. E. (2003). Capturing the invisible: designing context-aware photography. *Proceedings of the 2003 conference on Designing for user experiences.* New York, NY, USA.

Healey, J., & Picard, R. (Oct 1998). StartleCam: a cybernetic wearable camera . *Wearable Computers, 1998. Digest of Papers. Second International Symposium on* , (pp. 42 - 49). Pittsburgh, PA , USA .

Holleis, P., Kranz, M., Gall, M., & Schmidt, A. (2005). Adding Context Information to Digital Photos. *Distributed Computing Systems Workshops, 2005. 25th IEEE International Conference on*, (pp. 536-542).

Houde, S. (1992). Iterative Design of an Interface for Easy 3-D Direct Manipulation. *CHI* , 135-142.

Houde, S., & Hill, C. (1997). What do prototypes prototype. *Handbook of HumanComputer Interaction* .

Klymenko, V., Pizer, S., & Johnston, R. (1990). Visual Psychophysics and Medical Imaging: Nonparametrix Adaptive Method for Rapid Threshold Estimation in Sensitivity Experiments. *IEEE Transactions on Medical Imaging* , 353-365.

Kodak. (1999). *History of KODAK Cameras*. Retrieved 1 5, 2013, from http://www.kodak.com/global/en/consumer/products/techInfo/aa13/aa13.shtml

Lamming, M., & Flynn, M. (1994). "Forget-menot" Intimate Computing in Support of Human Memory. *Symposium on Next Generation Human Interface.* Meguro Gajoen, Japan.

Lowe, D. G. (2004). Distinctive Image Features from Scale-Invariant Keypoints. *International Journal of Computer Vision* , 91-110.

Lucero, A., Boberg, M., & Uusitalo, S. (2009). Image space: capturing, sharing and contextualizing personal pictures in a simple and playful way. *Proceedings of the International Conference on Advances in Computer Enterntainment Technology* . New York, NY, USA.

McConkie, G. W., & Currie, B. C. (1995). Visual Stability Across Saccades While Viewing Complex Pictures. *Journal of Experimental Psychology: Human Perception and Performance* , 563-581.

McGee, M. (2003). Usability Magnitude Estimation. New York, NY, USA: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems.

Microsoft Photosynth. (n.d.). Retrieved March 1, 2012, from http://photosynth.net/

Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review* , 81-97.

Renaud, F. (2012, March 1). *Meta Data Extractor in C#*. Retrieved from Renaud91: http://renaud91.free.fr/MetaDataExtractor/index.html

Sarvas, R., Viikari, M., Pesonen, J., & Nevanlinna, H. (2004). MobShare: Controlled and Immediate Sharing of Mobile Images. New York, NY, USA: ACM international conference on Multimedia.

Sauro, J. (2012). *9 Biases In Usability Testing*. Retrieved 1 3, 2013, from http://www.measuringusability.com/blog/ut-bias.php

Schindhelm, C. K., Gschwandtner, F., & Banholzer, M. (2011). Usability of Apple iPhones for Inertial Navigation Systems. *Personal Indoor and Mobile Radio Communications (PIMRC)*, (pp. 1254 - 1258). Munich, Germany .

Shneiderman, B. (2009). Promoting universal usability with multi-layer interface design. *CUU* , 1-8.

Snavely, N., Seitz, S., & Szeliski, R. (2006). Photo tourism: exploring photo collections in 3D. *ACM SIGGRAPH.* New York, NY, USA.

Sweller, J., van Merrienboer, J. J., & Paas, F. G. (1998). Cognitive Architecture and Instructional Design. *Educational Psychology Review , 10* (3), 251-296.

Travis, D. (1995). Applying Visual Psychophysics to User Interface Design. *Bahaviour & Information Technology* , 425-438.

Viana, W., Filho, J. B., Gensel, J., Villanova-Oliver, M., & Martin, H. (2009). PhotoMap: From Location and Time to Context-Aware Photo Annotations. *Journal of Location Based Services , 2* (3).

Walker, I. (2008). *Null hypothesis test and effect sizes*. Retrieved April 1, 2013, from Statistics for Psychology: http://staff.bath.ac.uk/pssiw/stats2/page2/page14/page14.html

Yee, W. (2009). Potential Limitations of Multi-touch Gesture Vocabulary: Differentiation, Adoption, Fatigue. *Human-Computer Interaction* , 291-300.

Appendix

Summary of consistent comments voiced by test subjects

* Comments have been modified into System A representing the Active system and System B representing the passive system

- "System A was much easier and fun to use than System B."
- "It is hard to flip through images in my head when using System B."
- "Quiz 3 for System B felt more difficult but had the same type of questions."
- "I feel like I know everything in the pictures after quiz 2 for System B"
- "It took me longer to recall details in System B compared to System A"
- "It is much more fun playing with the pictures in System A than trying to memorize images in System B."
- "System A is like a giant picture in which you could modify what you wanted to see."
- "I could not relate two objects in my head for System B when I forget sequential objects that link the two objects in question together."
- "I could close my eyes and see the environment in System A right now!"
- "It got easier to answer the questions the longer I spent looking at the environment in System A'
- "I cannot tell you the year an image is taken from System B but I am pretty confident I can tell you which images were taken before or after each other from System A"
- "It would be nice for System A to have a zooming function"

Psychophysical Testing Questions

Active System Questions

Simple Existence

1. Was there a mattress store?
2. Was there a fast food restaurant?
3. Was there a coffee house?
4. Was there a bank?
5. Was there a Dollar store?
6. Was there a picture of a bus?
7. Was there a picture of an old lady crossing the street?
8. Was there a picture of a woman with an umbrella
9. Was there a picture of a car?
10. Were there fencing in any of the pictures?
11. Were there blue banners on street lamps?
12. Was there a picture of rows of trees?
13. Were there large apartment buildings?
14. Were there power lines?
15. Were there green bushes?
16. Was there a drug store?
17. Was there a fine dining restaurant?
18. Was there a hardware store?
19. Was there a sports store?
20. Was there a picture of a bird
21. Was there a picture of a hot air balloon?

Detailed Existence

22. Was there a blue bus?
23. Was there a white bus?
24. Was there a picture of a red traffic light?
25. Was there a woman with a black umbrella?
26. Was there a man with a black umbrella?
27. Was there a woman with an orange bag?
28. Was there a picture of a blue car?
29. Was there a picture of a white car?*
30. Was there a picture of a green car?
31. Was there a bank called BMO?
32. Was there a bank called HSBC?
33. Was there a bank called TD?
34. Was there a mattress store called The Brick?

35. Was there a mattress store called Sleep Country?
36. Was there a mattress store called Ikea?
37. Was there a fast food store called McDonald's?
38. Was there a fast food store called KFC?
39. Was there a fast food store called Burger King?
40. Was there a coffee shop called My Cup Coffee?
41. Was there a coffee shop called Starbucks?
42. Was there a coffee shop called Second Cup?
43. Was there a brick crosswalk?
44. Was there a cement crosswalk?
45. Was there a dirt walkway?*
46. Were there trees without leaves?
47. Were there trees with green leaves?
48. Were there trees with colorful leaves?
49. Were some pictures taken during a sunny day?
50. Were some pictures taken during a rainy day?
51. Were some pictures taken during a snowy day?
52. Was there a dollar store called Dollar Max?
53. Was there a dollar store called Dollarama?
54. Was there a dollar store called Everything for a Dollar Store?
55. Was there a hardware store called Canadian Tire?
56. Was there a hardware store called Home Depot
57. Was there a hardware store called Home Hardware?
58. Was there a drug store called Shoppers Drug Mart?
59. Was there a drug store called London Drugs?
60. Was there a drug store called Pharmaprix?

Location Based

61. Is the fast food restaurant McDonalds located across the street from the mattress store Sleep Country?
62. Is the fast food restaurant McDonalds located across the street from the mattress store Ikea?
63. Is the fast food restaurant McDonalds located across the street from the coffee shop My Cup Coffee?
64. Is the fast food restaurant McDonalds located across the street from the coffee shop Starbucks?
65. Is the fast food restaurant McDonalds located across the street from the bank BMO?
66. Is the fast food restaurant McDonalds located across the street from the bank HSBC?
67. Is the fast food restaurant McDonalds located across the street from the Dollar Store?
68. Is the fast food restaurant McDonalds located across the street from the tool shop Home Depot?
69. Is the fast food restaurant McDonalds located across the street from the drug store London Drugs?
70. Is the mattress store Sleep Country located across the street from the mattress store Ikea?

71. Is the mattress store Sleep Country located across the street from the coffee shop Starbucks?
72. Is the mattress store Sleep Country located across the street from the bank BMO?
73. Is the mattress store Sleep Country located two streets from the coffee shop My Cup Coffee?
74. Is the mattress store Sleep Country located across the street from the Dollar Store?
75. Is the mattress store Sleep Country located across the street from the bank HSBC?
76. Is the bank BMO located across the street from the Dollar Store?
77. Is the bank BMO located across the street from the coffee shop My Cup Coffee?
78. Is the bank BMO located across the street from the bank HSBC?
79. Is the bank BMO located across the street from the tool shop Home Depot?
80. Is the bank BMO located across the street from the drug store London Drugs?
81. Is the Dollar Store located adjacent from the coffee shop My Cup Coffee?
82. Is the Dollar Store located across the street from the mattress store Ikea?
83. Is the Dollar Store located across the street from the bank HSBC?
84. Is the Dollar Store located across the street from the mattress store the Brick?

Photographer Based

85. Did Stan take a picture of a blue car?
86. Did Stan take a picture of a fast food restaurant called McDonalds?
87. Did Stan take a picture of a coffee shop called My Cup Coffee?
88. Did Stan take a picture of the drug store London Drugs?
89. Did Stan take a picture of the bank BMO?
90. Did Stan take a picture of a Fire Truck?

91. Did Steve take a picture of an old lady crossing the road?
92. Did Steve take a picture of a woman carrying an umbrella?
93. Did Steve take a picture of a fast food restaurant called McDonalds?
94. Did Steve take a picture of the bank BMO?
95. Did Steve take a picture of the hardware store Canadian Tire?
96. Did Steve take a picture of the Dollar Store?

97. Did Roger take a picture of the mattress store Sleep Country?
98. Did Roger take a picture of a Bus?
99. Did Roger take a picture of a coffee shop called My Cup Coffee?
100. Did Roger take a picture of a coffee shop called Second Cup Coffee?
101. Did Roger take a picture of a hardware store called Home Depot?
102. Did Roger take a picture of a green bus?

Time based

103. Were the pictures mostly taken during the summer?
104. Were the pictures mostly taken during the winter?
105. Were the pictures mostly taken during the fall?
106. Were rainy day pictures taken before the sunny day pictures?
107. Were sunny day pictures taken before snowy day pictures?
108. Were snowy day pictures taken before rainy day pictures?
109. Were some pictures taken during the day?
110. Were some pictures taken during the evening?
111. Were some pictures taken during the night?

Complex Scenario

112. If you were eating at the fast food restaurant and wanted to withdrawal money from the bank, could you go to the bank BMO?
113. If you were shopping at the mattress shop Sleep Country but became hungry, could you cross one street to get to the fast food restaurant McDonalds?
114. If you were shopping at the mattress shop Sleep Country but became thirsty, could you cross two streets to get to a coffee shop?
115. If you were shopping at the Dollar store, would you need to cross the street to buy a cup of coffee?
116. If you were at the bank and wanted to shop at the dollar store, could you get there by crossing one street?
117. If you were drinking coffee and wanted to buy a mattress, could you see a mattress store from your location?
118. If you were walking from the bank to the mattress store, would a fast food restaurant be on your right?
119. If you were walking from the coffee shop to the fast food restaurant, would a bank be on your left?
120. If you were walking from the dollar store to the bank, would the fast food restaurant be on your left?
121. If you were walking from the mattress store to the McDonalds, is a bank to your left?
122. If you were eating at McDonalds, could you see a coffee shop from your location?
123. If you were at a mattress store, would a bank or a coffee shop be closer to your location?
124. If you were at a fast food restaurant, would a bank or a coffee shop be closer to your location?
125. If you were at the bank, would the dollar store or a fast food restaurant be closer to your location?
126. If you were at a fast food restaurant, could you go to a hardware shop by crossing one street?
127. If you were shopping at the dollar store, could you get to a bank called HSBC?
128. If you were eating at the fast food restaurant McDonalds, could you see another fast food restaurant from your location?
129. If you were drinking coffee could you see a drug store from your location?
130. If you were at a bank, could you cross the street to get to a hardware shop?

Passive System Questions

Simple Existence

1. Was there a picture of a bank?
2. Was there a bus?
3. Was there a shoe store?
4. Was there a bakery shop?
5. Was there a camera store?
6. Was there a Japanese restaurant?
7. Was there a bookstore?
8. Was there a coffee shop?
9. Was there a chocolate shop?
10. Was there a picture of a lady crossing a street?
11. Was there a picture of a construction worker?
12. Was there a picture of a dog?
13. Was there a picture of a cat?
14. Was there a picture of a house?
15. Was there a picture of a motorcycle?
16. Was there a picture of a fast food restaurant?
17. Was there a picture of a man on a bicycle?
18. Was there a picture of a woman on a bicycle?
19. Was there a picture of a toddler?
20. Was there a brick wall in any of the pictures?
21. Was there a picture of a hardware store?

Detailed Existence

22. Was there a picture of a man wearing a yellow construction vest?
23. Was there a picture of a man wearing a yellow hard hat?
24. Was there a picture of blue hanging banners?
25. Was there a drug store called Shoppers Drug Mart?
26. Was there a bakery shop called Cobs Bread?
27. Was there a camera store called Kerrisdale Cameras?
28. Was there a bank called HSBC?
29. Was there a coffee shop called Starbucks?
30. Was there a picture of a red Scotiabank banner?
31. Was there a bank called Scotiabank?
32. Was there a Japanese restaurant called ASA Sushi?
33. Was there a picture of a white and blue bus?
34. Was there a bakery called Faubourg?
35. Was there a bookstore called Hager Books?
36. Was there a green bus?
37. Was there a red bus?

38. Was there a coffee shop called Tim Hortons?
39. Was there a coffee shop called Waves?
40. Was there a bank called CIBC?
41. Was there a picture of a pink car?
42. Was there a drug store called London Drugs?
43. Was there a bookstore called Chapters?
44. Was there a picture of a blue bench?
45. Was there a picture of a woman wearing a blue dress?
46. Was there a picture of a red and blue bus?
47. Was there a picture of a woman with blonde hair?
48. Was there a restaurant called Minerva?
49. Was there a chocolate store called Purdy's?
50. Was there a picture of a purple banner?
51. Was there a picture of a green banner?
52. Was there a picture of green construction netting?
53. Was there an orange traffic cone in any of the pictures?
54. Were there trees with red leaves?
55. Were there trees with no leaves?
56. Were there trees with green leaves?
57. Was there a hardware store called Home Hardware?
58. Was there a fast food restaurant called McDonalds?
59. Was there a fast food restaurant called A&W?
60. Was there a theatre called Stanley Theatre?

Location Based

61. Is the bank HSBC located adjacent to the camera store Kerrisdale Cameras?
62. Is the bank HSBC located across the street from the coffee shop Tim Hortons?
63. Is the bank HSBC located across the street from the drug store London Drugs?
64. Is the bank HSBC located across the street from the bookstore Chapters?
65. Is the bank HSBC located adjacent to the clothing store Jacob?
66. Is the bank HSBC located across the street from the coffee shop Starbucks?
67. Is the bank HSBC located adjacent to the bank CIBC?
68. Is the bank HSBC located across the street from the chocolate store Purdy's?
69. Is the bank HSBC located adjacent to the hardware store Kerrisdale Lumber?
70. Is the chocolate store Purdy's located adjacent to the coffee shop Starbucks?
71. Is the chocolate store Purdy's located across the street from the bank BMO?
72. Is the chocolate store Purdy's located across the street from the bank CIBC?
73. Is the Japanese restaurant ASA Sushi located adjacent to bookstore Chapters?
74. Is the Japanese restaurant ASA Sushi located across the street from the drug store Shoppers Drug Mart?
75. Is the Japanese restaurant ASA Sushi located adjacent to the bookstore Hager Books?
76. Is the Japanese restaurant ASA Sushi located across the street from the bank CIBC?
77. Is the Japanese restaurant ASA Sushi located across the street from the clothing shop Aritzia?
78. Is the Japanese restaurant ASA Sushi located across the street from the coffee shop Tim Hortons?

79. Is the Japanese restaurant ASA Sushi located across the street from the drug store London Drugs?
80. Is the Japanese restaurant ASA Sushi located across the cinema Oakridge Cinema?

81. Is Hager Books located adjacent to the bakery Cobs Bread?
82. Is Cobs Bread located adjacent to Kerrisdale Cameras?
83. Is the shoe store The Bootery located adjacent to ASA Sushi?
84. Is the shoe store The Bootery located adjacent to clothing shop Gap?

Photographer Based

85. Did Stan take a picture of the bank HSBC?
86. Did Stan take a picture of the coffee shop Starbucks?
87. Did Stan take a picture of the chocolate shop Purdy's?
88. Did Stan take a picture of a blonde woman?
89. Did Stan take a picture of a drug store London Drugs?
90. Did Stan take a picture of a white dog?
91. Did Steve take a picture of a blue banner?
92. Did Steve take a picture of a drug store Shoppers Drug Mart?
93. Did Steve take a picture of a bakery?
94. Did Steve take a picture of a blue bus?
95. Did Steve take a picture of a fire truck?
96. Did Steve take a picture of a hardware store Kerrisdale Lumber?
97. Did Roger take a picture of a bookstore Hager Books?
98. Did Roger take a picture of a Japanese restaurant ASA Sushi?
99. Did Roger take a picture of a bank Scotiabank?
100. Did Roger take a picture of a bank BMO?
101. Did Roger take a picture of a clothing store Gap?
102. Did Roger take a picture of a coffee shop Waves Coffee?

Time based

103. Were the pictures mostly taken during the day?
104. Were the pictures mostly taken at night?
105. Were the pictures mostly taken in the evening?
106. Were there any pictures taken in the summer?
107. Were there any pictures taken in the fall?
108. Were there any pictures taken in the winter?
109. Were rainy day pictures taken before sunny day pictures?
110. Were sunny day pictures taken before rainy day pictures?
111. Were snowy day pictures taken before sunny day pictures?

Complex Scenario

112. If you needed to withdraw money to dine at a Japanese restaurant, would you need to cross the street to get to a bank?
113. If you needed to withdraw money for coffee at Starbucks, would you have to cross the street to get to a bank?
114. If while having coffee at Starbucks, you decided that you need to buy some chocolates, would you have to cross the street to get some?
115. If you needed to withdraw some money to buy bread, would you have to cross two streets to get to the nearest bank?
116. If you were at the bakery store Cobs Bread and decided that you need to buy some cough syrup, would you have to cross two streets to get to the nearest drug store?
117. If you were walking to the bank HSBC from the Japanese restaurant ASA Sushi, would you be passing the bakery store on your right?
118. If you were walking to the bank HSBC from the Japanese restaurant ASA Sushi, would you be passing the drug store Shoppers Drug Mart on your right?
119. If you needed to walk from the shoe store The Bootery to the coffee shop Starbucks, would you need to cross a street?
120. If you were walking from the bakery Cobs Bread to the shoe store The Bootery, would you be passing the bookstore Hager Books on your left?
121. If you were walking from the bakery Cobs Bread to the shoe store The Bootery, would you be passing at least one traffic light on your right?
122. If you were walking from the bakery Cobs Bread to the Japanese restaurant, would you be passing a bank on your left?
123. If you were at a camera shop, would a bank or a Japanese restaurant be closer to your l ocation?!Japanese restaurant
124. If you were at the bank Scotiabank, would the bank HSBC or the coffee shop Starbucks be closer to your location?
125. If you were at the chocolate store Purdy's, would a coffee shop called Starbucks or a drug store called Shoppers Drug Mart be closer to your location?
126. If you were having lunch at the Japanese restaurant ASA Sushi, would you be able to see the drug store Shoppers Drug Mart from your location?
127. If you were exiting the bank HSBC, would you be able to see the chocolate store Purdy's from your location?
128. If you were leaving the bookstore called Hager Books, would you need to cross the street to get to a clothing store called The Gap?
129. If you were leaving the drug store Shoppers Drug Mart, could you see another drug store called London Drugs?
130. If you were walking from a coffee shop called Starbucks to a bank called Scotiabank, would you be passing a cinema called Oakridge Cinema on your right?