Thesis Proposal: Using Collages as an Interface to Large Photograph Collections

Michael Wallick
University of Wisconsin, Madison
Advisor: Michael Gleicher
Date: December 6, 2005

1 Introduction

Digital cameras are quickly replacing traditional film cameras for the average photographer. With a digital camera it is possible to capture, store and transfer more photographs than ever before. While it is simple to acquire digital photographs, new techniques need to be developed in order to organize, share, and generally interact with large collections of digital photographs.

A collage is a composite of a several different *media elements* combined to form an artistic creation. The media elements used in a collage vary by artistic preference and the message the collage is meant to convey. A very popular media to use is images or photographs. For the purposes of this proposal, I will refer to a collage, and photo collage interchangeably as I am not considering collages designed with other media elements.

A collage is a unique medium, in that it can convey an overall sense of the artist's intent, and/or several messages within a single two-dimensional picture. In this sense, the photo collage is able to convey an overall sense of the photographs in a collection (or a subset of that collection).

The photo elements in a collage present a sampling over all the elements in all of the photographs in the collection. If the sampling is performed in an intelligent manner, that sampling in turn can represent a summary of the photographs in the collection. In other words, we can say that a collage provides an *importance sampling* of the set of photographs. In such a sampling scheme, photos are taken based on their perceived importance relative to other photographs in the collection. In the simplest case, this importance sampling can be a human choosing their favorite photos.

If a collage is well designed with certain properties in mind, then I believe that the collage can be used as a tool for dealing with large photograph collections. Some of those properties are:

- **Use an importance sampling to summarize the collection.** As discussed above, if the photographs that were selected by an importance sampling, then the resulting collage acts as a summary of the photo collection.

- **Comprised of elements of photographs, rather than entire photographs.** If elements, rather than whole photographs, are used then more elements can be included in the final collage without increasing the size of the physical layout.

- **No implied or forced ordering of elements.** The collage should convey a gestalt impression the photographic set, and should be taken in at once.

- **Can be thematic.** We may want to restrict the sampling to only pick from photographs that have specific properties. In this way we further reduce the size of the collection and give a topic to the resulting collage.

It is my thesis that a system which produces collages with the above mentioned properties can be used as an interface to large collections of photographs. The following is a list of some of the ways that such a scheme may be used:

- **Visual navigation of photo collection.** A collage can "sit above" a photo collection, quickly giving the viewer a sense of what photographs are underneath the collage.

- **Display search results.** A collage can be used as a medium for displaying results from a photograph search. This would give the user a sense of what results were returned. The user can navigate through the results in a similar manner as described above.

- **Sharing and Publishing Photo stories.** Collages can be used as a means of showing off photographs. It is much simpler to give (and receive) a few collages rather than several hundred photographs. In this context, sharing and publishing can mean any form of allowing others to view the photographs. This includes showing the photos (screen or printout), e-mailing, publishing on the web, etc. In this way the collages can also be used to tell a story in the same way as photographs.

The remainder of this document is organized as follows. Next, in Section 1.1, I will further motivate the idea of using photo collages through a case study of a personal example, a recent family vacation in which approximately 1000 photographs were taken. Then in Section 2, I will discuss the research efforts that are related to this topic. In Section 3, I will lay out the different challenges involved in creating such collages and using them as an interface tool. In Section 4, I will describe the work that I have already completed towards this goal. After that, I will lay out a tentative time schedule for completing each remaining task in Section 5. Finally, Section 6 gives additional information, and source code.

1.1 Case Study: A Family Vacation

In August 2005, I went on a week long cruise as a family vacation. In total 11 members of my family attended, bringing along five digital cameras. Each camera took about 200

Figure 1: Nine photographs taken at the same time of the same subject, an iguana. Any one of the photographs can be representative of this sighting.

photographs. At the week's end, I copied all of the contents of each camera to have a record of the photographs. I will use this collection as an example to motivate the use of photo collages, and will refer to it for concrete examples throughout this proposal.

The only useful interaction with the entire collection that I have found is to have the collection cycle as background imagery as a slide show. The sheer volume of photographs makes it a daunting task to attempt to organize the photos by any specific manner. Even if I did construct some organizational hierarchy, it is it unlikely that the hierarchy would be useful in every scenario in which I would want to interact with the photos. If the photographs were labeled, then hierarchies can be dynamically generated based on the labeling and my desires. However, again, the sheer volume of photographs makes attempting to label the individual photographs a daunting task.

In addition to being large, the collection of photographs also contains a large amount of redundancy. For example, Figure 1 shows nine different photographs that were taken at roughly the same time, of the same iguana. Unless the viewer specifically wants to see all of the shots of that particular iguana, it is unlikely that all nine photographs would be needed at the same time. This redundancy among digital photographs is common, as the cost of taking a single digital picture is very low, and a bad photograph can be deleted on the spot or at any time. As a result, people will often take multiple pictures of the same subject under differing conditions.

The collage techniques that I describe in this proposal can be applied to help organize my vacation photos (as well as other photo collections). An example of such a collage is shown in Figure 2. This hand generated collage exhibits the properties described above for a good collage. First, the collage consists of nine elements from the original 1000 photographs. The importance sampling, in this case, was to hand select the photographs and their elements. It has a theme that each element contains or makes an allusion to alcohol. Finally, the elements

Figure 2: A simple collage from family vacation photos. Each of the nine elements either contains or alludes to the theme of alcohol.

were placed without regard to any ordering, but rather to look aesthetically pleasing.

The particular collage in Figure 2 is only one possible creation in such a scheme. One could easily image constructing other collages with themes showing events at mealtimes, shore excursions, time by the pool, etc.

2 Related Work

This thesis proposal covers many different aspects of computer science, including areas in computer graphics, vision and user interfaces. In this section, I will briefly describe some of the related work in each of the areas that I plan to investigate.

2.1 Organizing Photo Collections

There are several research projects and commerical ventures that help consumers deal with large sets of images. Three such commercial systems are Flickr [17], Kodak Easy Share Gallery [6], and Tag World [19]. These are webpages which allow users to upload the pictures taken, label the photographs and share them with others around through the web interface. Flickr and Tag World both allow community labeling of photographs. This means that anyone (with permission) can apply labels to a picture, regardless of ownership.

Similarly there are several pieces of commercial software for organizing photo collections. Most notably is Picasa [5], which is a free photo management program provided by Google.

It is a designed to store the photos and allow for quick searches. Adobe's Photoshop Elements [4] is another piece of commercial software that includes tools for storage and organization of photographs. Elements is designed to help the user with the task of organizing photographs; for example, it employs face detection and then has the user manually label each found face in the photo set.

In addition to commercial ventures, the problem of dealing with large sets of images remains an open one that has been investigated by several user interface researchers. Drucker et. al. [8] developed MediaBrowser. In this system users label individual photographs and videos. The system can then put together thematically related sets, as well perform searches on the set of images. Similar to MediaBrowser is the MiAlbum system [31]. It uses user labeling to help manage a "typical family's" digital photographs.

Shaft and Ramakrishnan [23] developed a system which uses image classifiers and a database to organize images. The images that are placed in the database have information, such as edge map and color histogram, automatically extracted to help provide information about the photographs. In addition, the user can apply labels to objects within the image allowing the user to carry out queries to search for the images.

2.2 Image Clustering and Automatic Labeling

There are several systems that provide some type of automatic labeling/image clustering based on computer vision. It should be noted that labeling and clustering are two different areas of research. However, if a labeling of all photographs exist, then a clustering can be approximated, i.e. group all photos with the same label as a cluster. For this reason, I treat labeling and clustering in the same discussion.

AutoAlbum, developed at Microsoft Research, by Platt [20] is a system for clustering photographs. Like my proposed work, it takes the time stamp of each photograph in order to generate a clustering. In this scheme, the photographs are only organized into a single level. In Section 3, I argue for a multi-level event scheme. The single level works for AutoAlbum since only albums are being created; there is no concept of searches or more in-depth organization

Beyond using metadata for clustering and labeling, others have employed computer vision to do this task. Jeon et al. [14] developed a system for automatic labeling of images. The system takes a training set of manually-labeled images. When it encounters a new image, it attempts to match the image based on the training data. The shortcoming of this system is that it is only as good as its training set. For example, if the system encounters an image of a lion but only has images of cats in its training set, it will label the lion image as a cat. If the system encounters a picture of a lion, but only has a limited type of training data, such as architectural images, then the labeling will be non-deterministically wrong.

In 2000, Shneiderman and Kang [24] developed a labeling system for drag-and-drop labeling of images. In their approach, the user selects a photograph and labels, and simply drops the labels in place. This provides a simple and quick method for labeling photographs.

Schaffalitzky and Zisserman [22] present a system for clustering images based on computer

Figure 3: 15 Images from a church in Valbonne, France. from [22]. Despite having a wide baseline and differing features, the system presented by Schaffalitzky and Zisserman is able to cluster these as one group.

vision. Unlike previous work in computer vision, their system will cluster images of the same scene even if there is a large disparity between the two images, i.e. it does not require a small baseline. This approach works well for clustering if the photographer returns to the same place through multiple points in time. In the vacation example described in Section 1, this method would cluster all of the images taken in the dining room. Figure 3 shows an example of several different photographs taken (at different angles and orientations) of a church. Their method is able to cluster these images together despite the wide baseline and other differences.

Much research has also been done to use classifiers [11, 12] (such as face classifiers) in order to label images. Wei and Sethi [30] present an algorithm for detecting faces in images, which can in turn be used for labeling. In the most recent edition of Adobe Photoshop Elements [4], the photo album tool includes a tool that locates all of the faces in the set of photographs. The user can then label all of the faces individually.

It should be noted that while face detection works quite well, face recognition (determining the person) is still in early development. Despite this, very recently a new web service, Riya [18], entered the market. This service allows users to train the system to perform face recognition, rather than simple face detection. The system also reads text on signs (and other features) in the photograph. This gives users a way to automatically label many of their photographs. At the time this proposal was written, the Riya system is able to interact with some other web-based image databases, with more on the way. Once it is more developed, I will look at incorporating Riya output into my system. At the writing of this proposal, Riya was not far enough along to perform a meaningful test.

In "Show & Tell," Srihari and Zhang [25] describe a system for semi-automatic annotation of images. They use a combination of image classifiers along with natural language processing

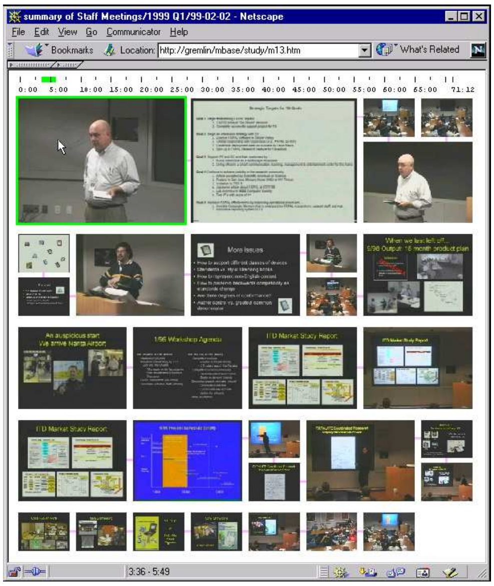

Figure 4: Summarized video from [2]. The more important the key frame, the larger it is in the final display.

to create the labeling. In their system they concentrate on medical images, as doctors are used to dictating information about imagery.

2.3 Layout and Collage Generation

The problem of laying out many images or video frames is one that has been explored by several researchers. Work carried out at FX Palo Alto Laboratory, Inc. [2, 27] looked at summarizing a video in a comic book (or Japanese Mangra) style. In this system key frames are selected from the source video. The algorithmically-determined importance of the key frame dictates how much space the final image would take up. A unique packing algorithm is used to determine the final layout. Figure 4 shows an example of a summarized video.

Many programs and researchers create a very simple collage by laying out thumbnails of each image. This is done in Photoshop Elements, and Picasa [4, 5]. In Photomesa [1], Bederson employs a ZUI or "Zoomable User Interface" to display the photograph thumbnails. In this system all of the photographs are displayed as thumbnails. The more photographs being displayed, the smaller they appear. What makes the system unique is that as the user mouses over and clicks different parts of the display the interface will zoom in photographs in that area. The user can then drive down to show a photograph in full zoom. Using all of the thumbnails to create a collage does not satisfy the requirements that I laid out in 1; most notable is that it is not a sampling of the pictures, but all of them, and entire photographs rather than individual elements, and there is an order implied by how the photographs are organized.

In [3], Christel et al. created a collage as a summary of news reports. In their work, they used a recorded corpus of several thousand hours of news footage. Key frames and pieces of video were used as an importance sampling of the different stories that appear in the corpus.

Figure 5: Geographic news collage from [3]. In this collage a map is shown with different reports coming from that area.

They use computer vision to pick shots and text understanding (from the news transcript) to perform the importance sampling in order to build collages. In the work, they make different collages to represent different information. The collage in Figure 5 is a map showing different reports superimposed over the area that the report originates from. In this system, the user can drive down in the map to narrow where the stories come from. Another display they discuss is a collage showing the same story subject as it develops over a time-line or all of the stories over a span of time. This system looks explicitly at news footage, rather than personal photographs as I intend too. Also the sampling is performed by using the text of the news transcript, information that does not exist for the problems I am addressing.

Fogarty et al. [9] built a system for making collages that are both aesthetically pleasing and convey information. They describe having a large digital display that is suitable to be hung as a piece of art. The system collects information that can be displayed with information that does not require constant attention, such as e-mail or news group headers. Most of the time, the collage functions as decorative artwork, however when the viewer wants to give it full attention, other information (new e-mail) can be gleaned from the collage. The collages they design are very different from the collages I intend to build. Most notable is that they are not using images to design the collage; further this system is more interested in the artistic side of collage medium, rather than the informational properties.

Recently, Diakopoulos and Essa [7] presented a system for creating a photo collage. In their system, the user selects a set of photographs and a template such as the one in Figure 6. The system will optimize the layout of the photographs based on the selected template. Figure 7 shows an example of the completed collage. Unlike what I am proposing, this system uses entire photographs, requires the user to select the photos to include, and limits the input size for the collage to that of the template.

Work at Microsoft Research Cambridge has lead to the “Digital Tapestry” [21] system. This

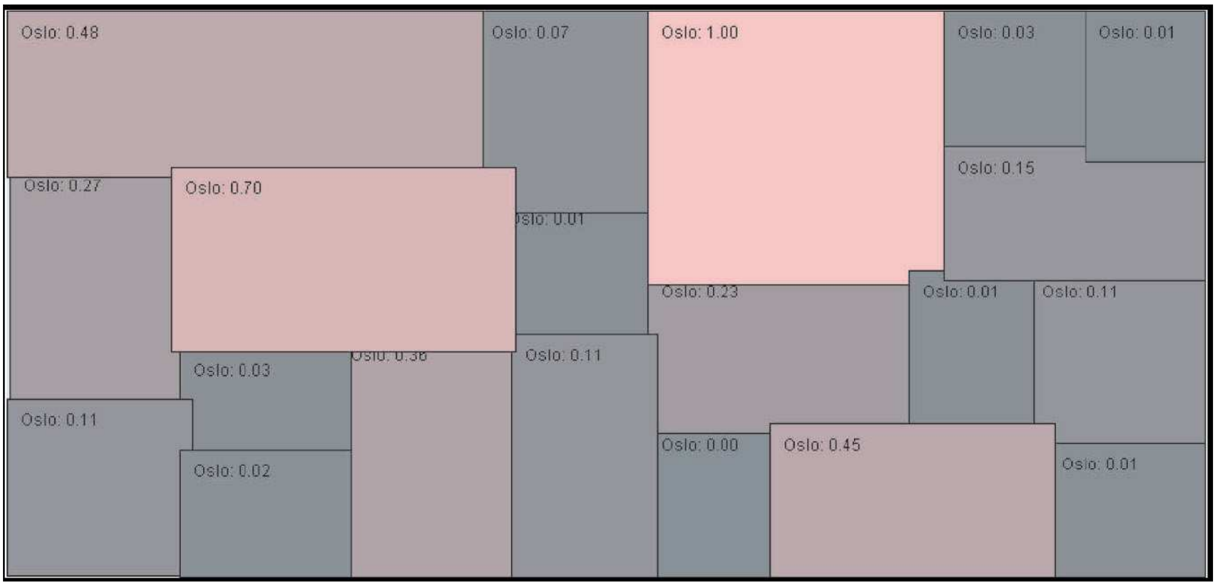

Figure 6: Collage template from [7]. The system runs an optimization to best place the selected images.

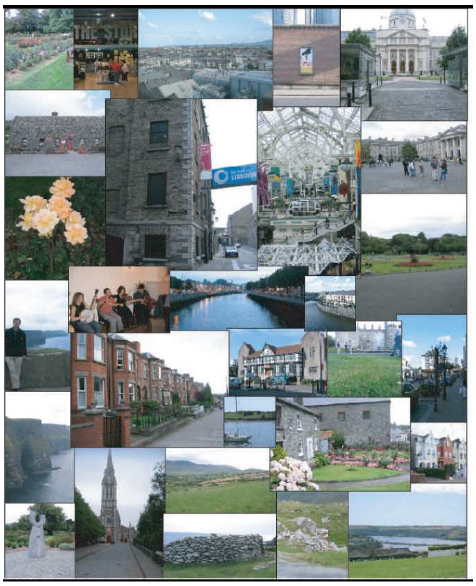

Figure 7: Collage from [7].

Figure 8: Digital Tapestry from [21].

approach uses saliency to identify the important features in an image. The salient regions are rendered together using a graph cut algorithm to minimize the energy (or difference) between different elements and create a reasonable looking composite. Their system is different from what I propose in several ways. While the collages do show elements, the elements can sometime be cut off. This is because their method for collecting the images just uses saliency rather than an input labeling. This can be seen on the horse/waterfall in Figure 8. Second their system does an importance sampling only by image epitome, so two very related subjects may appear different and be included. Finally since there is no notion of labeling of the images, if the user wants to design a thematic tapestry the collection must be compiled by hand.

Companies have also begun to offer collage systems. Kodak Easy Share Gallery [6] allows users to upload photographs and create a collage. The system randomly lays out the images in an $n \times n$ grid. If there are not enough pictures to complete this grid (i.e. a perfect square), then some pictures are repeated. If a picture is not the correct aspect ratio, then it is cropped to fit. Figure 9 shows an example of a user-created collage. This program has several drawbacks. First, when the pictures are cropped, important information may be lost. This is seen in Figure 9 (bottom row/center picture) where the child's face is partially cropped. A similar failing is when the images are rotated and perturbed (for artistic purposes); these translations can cover important information in other photos; this is seen in top row/center picture of Figure 9. Again, in this system the entire photograph is shown rather than individual elements.

2.4 Image Retargeting

As small screen devices increase in popularity, research in the area of retargeting images to these devices has intensified. Simply down-sampling images does not work. If the image

Figure 9: Collage generated automatically in the Kodak Easy Share Gallery.

is down-sampled too much, it will be become unrecognizable. Additionally, if the device displaying the image is a different aspect ratio than the photograph, it can look distorted because the screen is small it does not make sense to give up screen space to make the image fit.

Two notable efforts for retargeting images are by Suh et al. [26] and Liu and Gleicher [16]. Both of these systems only operate on single images rather than large sets. It is my belief that a collage can be used as an interactive object for retargeting large collections of images. The collage can be shown on a small screen as described in Section 3.6.

3 Research Overview

It is my goal to create a system that generates and uses collages as an interface to a large collection of digital photographs. In this section, I will lay out the challenges involved in reaching that goal. In order to do this, I will first need to obtain a simple hierarchy of the photo collection (3.1). This hierarchy will allow a user to label or annotate groups of photos at once rather than having to mark each photo individually (3.2). After the photos have been given some labeling, it will be possible for the user to search for individual or groups of photographs (3.3). An interface can be developed for creating collages from subsets of the collection (3.4), i.e. the search results or other partitions. Once generated, these collages can then be interacted with in various applications (3.5 and 3.6).

3.1 Automated Clustering

In order to be able to do importance sampling of the image collection, a crucial piece of this proposal, it is necessary to have some type of organization applied to the images for

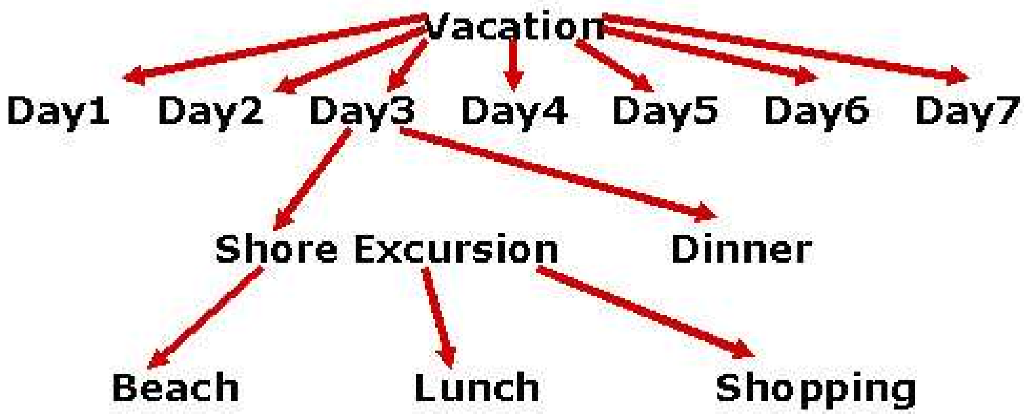

Figure 10: One possible organization of the photographs from the vacation example.

the sampling to operate over. Ideally, this organization can come about by having the user supply it for each image; however, with large image collections, this is not feasible. In my vacation example (from Section 1.1) there are approximately 1000 photographs. It would be too time consuming to label each of those photos individually.

A means of simplifying the organization task is to put the photographs into a simple, but effective hierarchy. One such hierarchy is a temporal hierarchy. Using the metadata stored in the photographs, a temporal hierarchy can be built without requiring any user input.

All photo organization tools offer a time-line organization of pictures; the hierarchy that I propose goes beyond the one-dimensionality that a simple time line offers. Photographs that are taken close together in time are related. Using this fact we can automatically separate the photographs into groups of super=events, events, sub-events, sub-sub-events, etc. If we apply this scheme to the vacation photo collection then the following is an example of such a grouping:

Super-event: All of the photographs in the vacation collection constitute a single super-event. Another collection, such as pictures taken during a weekend trip, are a different super-event.

Event: The events are the happenings within the super-event. Some examples of events include shore excursions, time by the pool, meals, etc.

Sub-event: The sub-events are happenings within the events. An example of a sub-event is shopping or beach time during a shore excursion.

A good clustering algorithm will provide a multilevel hierarchy which will approximate a logical grouping of the photographs. Figure 10 shows an example of a potential grouping from the vacation example. Notice that different branches in the hierarchy have different depths depending on how deep the event structure should go. Figure 12 shows results from the automated clustering.

In Section 1.1 I stated that an organizational hierarchy of any type will most likely not suffice for all possible scenarios of interacting with the photo collection. This temporal

Figure 11: An individual photograph (left) with a single element that may be labeled by a user (right).

hierarchy is no different, as it should not be difficult to construct several scenarios in which this hierarchy would fail. It is important to note that this hierarchy is used simply to boot strap the annotation process (described next). While it may be useful for tasks involving temporal searches, it is by no means meant to be the final organizational method.

3.2 Event and Photo Annotation

If there is a good user-supplied annotation of the photographs in the collection, the task of organizing them becomes much simpler. In this context, organizing means to be able to accurately search the collection and build arbitrary partitions or hierarchies based on the user's desires. It should be pointed out that in this context the terms annotation and label are considered to be the same and will be used interchangeably.

Labeling needs to be accomplished without requiring the user to tag each individual photograph. In order to achieve this I will use the temporal hierarchy that was developed in the last section (3.1). In this way, the user is able to label each node in the hierarchy. The labels will be carried down to all of the photographs in the child nodes. It is much faster for a user to be able to label groups of several photos rather than each individual photograph. For example, if we were using the hierarchy in Figure 10, all of the photographs in the collection would have the label "vacation," photographs underneath the label "lunch" would also have the labels "Shore Excursion" and "Day 3."

In addition to labeling groups of images, it should also be possible for the user to label individual elements of photographs. A simple way to achieve this is through a lazy snapping interface [15], or some other interactive method. An example of this is shown in Figure 11.

I assume that the user will take the time to label the event hierarchy as a means to boot strap the organization. However, I will make no assumptions about labeling individual photographs or elements of photographs. One possibility is to allow the user the ability to label photographs and photo-elements at this point, or when the system returns search results (as described in Section 3.3). Alternatively, the user may never label the individual elements.

In order for this step to be successful an interface needs to be built to allow the user to label the collection of photographs. A successful completion of this step would be the design of an interface which uses the event hierarchy. After some practice with the interface, a user should be able to label a collection of several hundred photos in a few minutes, rather than the hours that it would normally take.

3.3 Photograph Search and Retrieval

In order to deal with large sets of photographs, the user needs to be able to quickly search and retrieve photographs matching their desires. I expect that the user will provide some form of labeling. This can be in the form of labeling the nodes of the event hierarchy (discussed in 3.2). Alternatively, some other labeling may be provided; for example, if the images are stored on-line then the labeling may be a community effort. However, no assumptions are made that photographs (or photograph elements) were individually uniquely labeled.

The search function should perform an importance sampling in order to return up to a desired number of photographs (or elements). Figure 1 shows nine photographs of the same subject. If the user carries out a search which would include those photos among other results, then it may not be wise to show all nine results. It should be noted that by using collages as a form of image navigation, the omitted photographs can be thought of as residing underneath the result; this will be discussed in greater detail in Section 3.5.

In order to keep track of the photo and labels, as well as perform the search (before importance sampling), I will employ the use of a preexisting database, e.g. "Flickr" [17]. Flickr is an on-line database that stores images and allows the user to set permissions on who may annotate, view and download images. Additionally, Flickr provides an API for accessing the database so that third party applications can be built without exposing the user to the database underneath. Sections 2.2 and 4.3 describe Flickr in greater detail.

The challenge for this step is to come up with rules for performing importance sampling. As mentioned, most search results will return a much larger set of images than necessary due to the nature of digital photo collections and the fact that the labeling puts the same labels on many images. For this step, I expect to have an interface that allows the user to search a photo collection and specify a maximum number of returned results. The results should give a good representation of the collection based on a calculable measure of each photograph's importance.

3.4 Photo Collage Generation

In addition to the importance sampling, another contribution of this proposal is that the collage can be used as a medium for presenting the results of importance sampling as a single image. The collage can also be used as an interface tool for interacting with a collection of digital images. This will be discussed in greater detail in Section 3.5.

It is currently an open ended problem on how to lay out the elements of the collage. Diakopoulos and Essa [7] use a template to guide the layout process. Other methods, such as the one employed by Kodak [6] in their collage program, simply use a random layout. Currently I am working with saliency as a means of determining photograph layout, such as ordering the images based on total saliency (see Section 4.4).

Current research in automatically producing collages (including those listed above and in the related works sections) fails to meet the criteria for a collage to act as a good interface tool that I laid out in Section 1. Mostly, these collages are comprised of entire images, rather than elements. Further, the importance sampling for these collage systems must be done by hand, in other words the user must select which elements to include.

In my research, I will explore methods for generating layouts of collages. An important piece of this will be to select appropriate elements from the collection of images. Recall that I make no assumption that the user will mark individual elements of images. If there are no sufficient labels, other approaches, e.g. saliency [13, 21], may be used to help approximate these elements.

This step will be considered complete when a collage can automatically be generated from a selected set of photographs. The set of photographs will be automatically picked by the search and importance sampling function described in the previous section. The individual elements to be used will come from either labels provided by the user or from some method that automatically extracts important regions, e.g. saliency.

3.5 Applications of Photo Collages

It is my thesis that the photo collage can be used as a interface to large collections of digital photographs. Therefore, it is necessary to build applications that are aided by the use of collages. In Section 1 I listed three different applications that can use collages as an interface: visual navigation of a photograph collection, mechanisms for displaying image search results, and as a means of sharing or publishing photo stories.

The first application is visual navigation of a photo collection. I have alluded to such an application throughout this document. I will now explain it more concretely. Assume the photographs are organized into some type of hierarchy (temporal, label-based, etc.). At the bottom of the hierarchy is the photographs broken into groups. Each node in the hierarchy can be represented by a collage showing the summary of the photographs below it. In the vacation example the top of the hierarchy would be a collage showing a summary of all the photographs. The user could then drive down through the hierarchy as a means of navigating through their photo collection.

The second application is a mechanism for displaying search results. In this case a collage can be used to show the elements of the photographs that were found as the search results. The user may be able to further refine the search by interacting with the collage, specifying which elements better match their desires, and generate a new collage.

Using a collage as a means of sharing or publishing photo stories is another application. The other two applications require collages that are quickly rendered, so that the user may interact with them. If the user wants to share a collage (by printout, e-mail, publishing on the web, etc.) a richer interface for adding elements, and high-quality rendering must be developed.

This step will be considered complete after I have a system that is able to do each of the above outlined tasks. Ideally, there will be a single program to handle all of the tasks described thus far. This includes creating the hierarchy, an interface for annotating the images, constructing collages, and interacting with those collages.

3.6 Retargeting Collages

Small screen devices with photo and video capability have become very popular in the last few years, and even more so with the recent introduction of the iPod Video. It is challenging to show off a single image on a small screen device (as discussed in Section 2.4). It is even more challenging to effectively display many images at once on a small screen device. I believe that the methods I describe in this proposal may be used to address the problem of displaying multiple images on a single screen.

Time permitting, I plan to investigate up to three models of displaying photo collections on a small screen. The first model is to use the selected elements that would appear in a collage. Rather than putting the elements into a collage, they would be displayed individually on the small screen. An important question to answer in this case is would the context of the photograph be lost without the other supporting elements?

The second model is to create very small collages, so only a few elements would be shown. It is my expectation that most collages will have many elements (tens to hundreds), whereas a small collage would only contain a few (around 5) elements in it. The important question to answer with this model is, does such a small number of elements still represent a reasonable importance sampling of the images, or is it simply under sampled?

The final model to investigate is to use a full-sized collage showing only as many elements as fit on the screen at one time. The user can then move around the collage. There are several questions that can be posed in this model of interaction:

- What is a good method of navigation through the collage?
- How much navigation control should the user have?
- Does that control change depending on the device?
- How many elements (or pieces of elements) must be shown at once in order to maintain the context of the collage?

In this step I would like to construct examples of each model, and compare them. Ultimately I would like to be able to offer answers, or at least insight, for each of the questions posed by each model of retargetting.

3.7 Evaluation

In this thesis there are some pieces that will need to be evaluated. The first piece to be evaluated is the labeling mechanism. It is my belief that by labeling the groupings rather than individual photographs, the process can be drastically reduced. I would like to test this by conducting user studies. Users will be asked to supply two similarly sized sets of their own photographs. For one set they will be asked to label all (or some subset) of the photographs individually using a commercial system. For the second set, they will be asked to produce a similar labeling scheme using the system that I design. The time it takes to label each set will be recorded.

The second piece to be evaluated is the actual collages. Two of the collage qualities that I stated in Section 1 are that the collage should be able to tell a story and have a theme. In order to test that the collage does this, I will again carry out a user study. This time users will be shown either a stack of photographs or a collage containing elements from the photograph stack. The user will be presented with a list of words from the set of labels, and be asked to pick the words that they would apply to the photographs. A correct answer is including a word that is a label in that set. An incorrect answer is including a word that is not a label of the set, or omitting a word that is a label. I expect that the photo collage should have similar results to that of the stack of photographs.

Third, I would like to test the usability of the applications (3.5). For this evaluation I will ask users to use the collage interface to search for a specific image either by navigating the collage hierarchy or by using the collage as a display of search results. For both cases I will record the number of moves that the user makes to get to their destination and time that it takes them. This will be compared to the task of finding the photograph(s) with commercial photo management software.

A final evaluation, time permitting, will be to evaluate the different models of the retargetted images. In this case I will carry out a test similar to the second one described above. In these tests, I will show the user the different methods of retargeting the collection on a small screen device.

4 Work To Date

In this section I will describe the progress that has already been made towards this research goal. This includes photograph clustering, and preliminary photo collage generation.

4.1 Previous Projects

While working on previous projects [10, 28, 29], I have worked on developing tools that will be directly useful for this project. Most notably, the development of our ITImage library. This is an image handling library based on Intel IPL and OpenCV [11, 12], it is written for both C/C++ and Python. Additionally, I created and continue to maintain the ImageMath

Figure 12: Results from the automatic clustering. All nine photographs are from the same event (formal evening), each row is a different sub-event. (Top Row) Getting Ready for Formal Night. (Middle Row) Posing for Photographs on Deck. (Bottom Row) End of Formal Evening.

library (based on ITImage). This library provides several methods for manipulating images in different ways. This tool set will be very useful to me as I investigate photo collages.

4.2 Photographic Clustering

As of the writing of this proposal a photo clustering tool has been developed. The program reads the EXIF [32] data from the photograph and extracts the time stamp. The photographs are clustered together, based on time, using a temporal clustering scheme. The groupings are recursively broken down until there is not enough space left between images. This creates the event structure tree as described in Section 3.1.

Currently the tool has no GUI interface. An interface needs to be built which can show the groupings at different levels of the hierarchy and allow the user to apply labels. Figure 12 shows partial results of the clustering program. All the photographs in the figure were clustered in the same event, and each row is a different sub-event (see Section 3.1). The algorithm for clustering is given in Section 6.2.

4.3 Image Database

I have been investigating different programs and services to use for an image database. In doing this I have come up with the following list of requirements for a good database, shown in the list below:

1. **Globally available.** The photographs stored and labeled should be globally available, ideally through some web interface. It is a desirable feature to show off photographs anywhere, not just your home location. Also, a globally available database can help reduce the task of labeling by splitting the job up, allowing others to label the photographs. Finally, if the database is globally available, the user can retrieve all the pictures stored in the event of a computer crash without having to worry about creating back-ups.

2. **Published API.** Since I want to create new interfaces for working with images, it is necessary to have a published API to access the database.

3. **Free or nexpensive.** The database should be affordable by anyone who wants to use it. If storage of the photographs adds too much to the cost, users would not want to use it.

The table below summarizes the different programs and services previously discussed in this proposal and if it meets each of the requirements from the list above. Since Riya [18] is still in early development, it is not considered in this table.

Program Name	Globally Available	API	Free or Inexpensive
Picasa	×	×	✓ - Free
Adobe Photoshop Elements	×	×	× - Approx. $100
Kodak EasyShare Gallery	✓	×	✓- free
Tag World	✓	×	✓- free
Flickr	✓	✓	✓- free to $25 a year

As can be seen from the table above, Flickr meets all of the requirements laid out. As such, I have begun investigating using the Flickr database as a back end to my research. Photographs will be stored on the webpage and I will write programs using the provided Flickr API to keep track of the photographs and labeling.

4.4 Simple Photo Collage

At this point, I have written a program that works similarly to the Kodak collage. The user inputs a set of photographs and a collage is laid out. Figure 13 shows output from the program.

I am currently working with saliency maps of images in order to drive a more intelligent layout scheme. Saliency can be used in many different ways. For example, we may want to order the images by saliency. Another possible use is if two images are overlapping, put the more salient one on top. Finally, if photographs need to be repeated, we can choose the most salient photos for repeating.

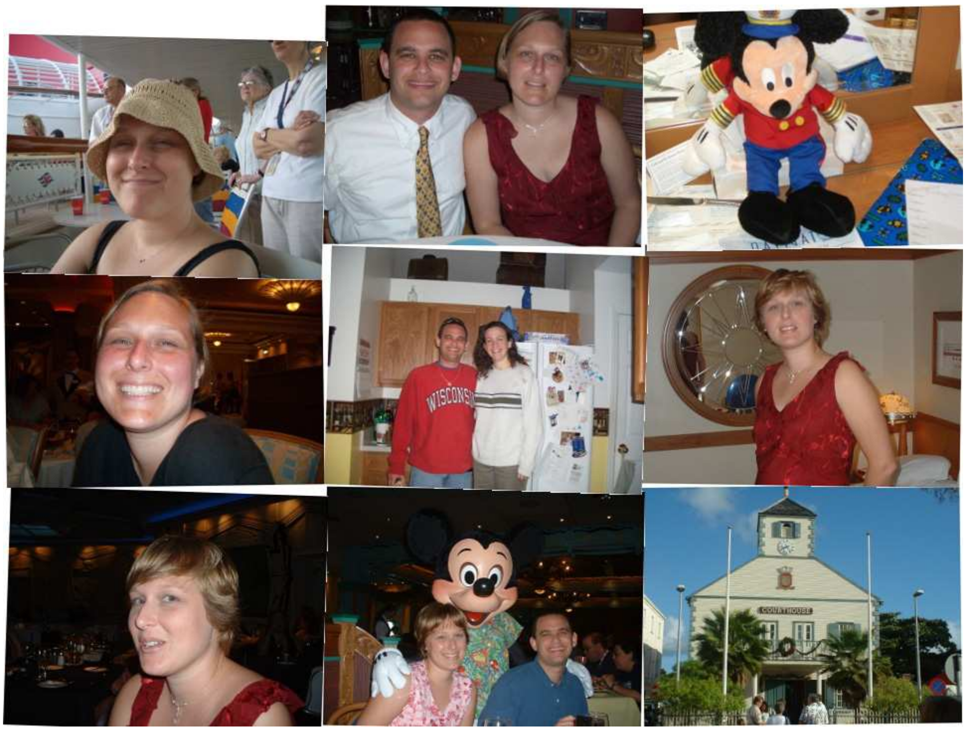

Figure 13: A simple collage generated by my program.

5 Research Plan

The following is a tentative schedule for completing the proposed research and a brief description of how each step will be approached.[1]

5.1 Photograph and Database Interface

Tentative Completion Date: February 15, 2006

Problem Statement. For a large collection of photographs and their associated labels, it is useful to keep track of the images through a database, as stated in Sections 3.3. In addition to the database, I will also need to develop an interface to that database, in order to add photographs and their labels.

Research Objective. For the database service, I will use Flickr [17] for the reasons outlined in Section 4.3. I will first study the Flickr API and develop a simple system for storing images and labels. I will then build an extended interface to easily complete the following tasks:

- Display images in a temporal hierarchy (Section 3.1).

[1] The dates listed are meant to include time for an internship during the summer of 2006 as well as time to complete papers for various conferences including SIGGRAPH, Multimedia, UIST and CHI. As the research progresses, this time line will be adjusted accordingly.

- Allow simple labeling of groupings from the hierarchy and elements of the photographs (Section 3.2).
- Search and retrieve full images from the database (Section 3.3).
- Upload and download photographs from the on-line database.

In this research I will explore ways of quickly labeling photograph sets and elements of individual photographs. The techniques will include a set of manual and automated labeling procedures.

5.2 Importance Sampling

Tentative Completion Date: October 15, 2006

Problem Statement. A major part of this thesis proposal is to be able to perform an intelligent importance sampling over a set of images. Given some search criteria, it is necessary to cull the returned photographs and collection to a manageable size, while still showing off as many of the important elements as possible.

Research Objective. It is in this step that I intend to make my major research contribution. I will develop algorithms for finding the most important elements in the collection of photographs. This will be based on as much information that can be gathered. This may include (but is not limited to):

- Temporal information about the photograph and other similar photograph
- Amount of labeling of the photograph relative to other returned images
- How often and long the photograph was viewed
- How much overall interaction takes place with the individual photograph

For this step, the photographs and elements will only be selected. The actual photo collage layout will not be considered until the next step.

5.3 Collage Layout and Interface

Tentative Completion Date: November 30, 2006

Problem Statement. Once the important elements are selected, they must be laid out in an aesthetically pleasing collage.

Research Objective. In this step I will look at different ways of laying out photographic elements in a collage. At this point, the method that I will use for laying out the collages is unclear. Some existing methods rely on a template-based approach, e.g. Diakopoulos and Essa [7]. A drawback of their approach, however, is that the total size of the collage is strictly limited to the size of the collage template, roughly 100 images. In my system I would like to have a more free-form collage generation technique that allows for any (reasonable) number of elements to be included.

Once the collages have been generated, I will work on using the collages as an interface to interacting with the collage. This includes the ideas discussed in Section 3.5 for directly helping organization, search and sharing of photographic sets.

5.4 Small Screen Devices

Tentative Completion Date: February 15, 2007

Problem Statement. Many people carry devices that have a small screen in it, such as a cellular telephone, PDA, etc. It is difficult to show even a single image on the small screen. It is my belief that photo collages can be used as a means of showing several images on a small screen device.

Research Objective. In this step I will investigate different ways of showing a generated collage on a small screen device. I will treat the collage as a type of object that the user may navigate through, either automatically or semi-automatically. It will be necessary to come up with control mechanisms to allow the user to move around the collage and figure out the best way to display the element of interest appropriately on the small screen.

5.5 System Evaluation and User Testing

Tentative Completion Date: May 15, 2007

Problem Statement. The ideas proposed in this research are primarily user interface in nature. Therefore, it is necessary to be able to determine if a user is able to properly use the different aspects of this system.

Research Objective. I will conduct various studies to determine the validity of the system. These studies were laid out in Section 3.7, however, I will briefly recap them here. First, the system will be tested in terms of its ability to find "important" photographs or elements. In this test, a search term will be provided, with the expectation that some predetermined elements should show up in the search results. The second set of tests will be in the form of a user study. Users will be asked to supply their own set of photographs and given full access to the system. They will be asked to try various tasks with the system and then be given a questionnaire to gauge their opinion of the system.

5.6 Writing the Thesis

Tentative Completion Date: August 1, 2007

This thesis is to be finalized in time for a defense at the end of the 2006-2007 academic year.

6 Appendix

6.1 Exif Metadata Chart

The following is a chart is an example of the information contained in the EXIF file format of a photograph taken by a digital camera [32].

Manufacturer	CASIO
Model	QV-4000
Orientation	top-left
Software	Ver1.01
Date and Time	2003:08:11 16:45:32
YcbCr Positioning	centered
Compression	JPEG compression
x-Resolution	72.00
y-Resolution	72.00
Resolution Unit	Inch
Exposure Time	1/659 sec.
FNumber	f/4.0
ExposureProgram	Normal program
Exif Version	Exif Version 2.1
Date and Time (original)	2003:08:11 16:45:32
Date and Time (digitized)	2003:08:11 16:45:32
ComponentsConfiguration	Y Cb Cr -
Exposure Bias	0.0
MaxApertureValue	2.00
Metering Mode	Pattern
Flash	Flash did not fire
Focal Length	20.1 mm
Maker Note	432 bytes unknown data
FlashPixVersion	FlashPix Version 1.0
Color Space	sRGB
PixelXDimension	2240
PixelYDimension	1680
File Source	DSC
InteroperabilityIndex	R98
InteroperabilityVersion	(null)

It should be noted that EXIF data can also be extended to include additional values. This includes the possibility of storing GPS (or other) data if available at the time the picture is taken.

6.2 Event Clustering Algorithm

i

The following shows the algorithm for clustering the images based on temporal events (Sections 3.1, 4.2. In this algorithm, the user calls the recursive function "Cluster." The parameters for that function are images, time, minTime and interval. Images is a list of tuples, where the first entry is the time the photograph was taken. Time is the maximum length of time allowable between two shots to be considered in the same group, e.g. time=3600 is one hour between shots. MinTime is the minimum amount of time between two shots to be considered in the same sub (or sub-sub, etc.) group, e.g. minTime=300 means that photographs must be taken at least every 5 minutes to appear in the bottom level group. Finally, interval is a control to decrement the time check for clustering the groups, e.g. interval=600 means that "time" is reduced by 10 minutes with each recursive call of Cluster().

```
#Takes the Image List and Times List and does Hierarchical Clustering
def __Cluster( images, maxSpace ) :
    clusters = []
    a = []
    a.append(images[0])
    for i in range(1, len(images)) :
        if images[i][0] - images[i-1][0] < maxSpace :
            a.append(images[i])
        else :
            clusters.append(a)
            a = []
            a.append(images[i])
    clusters.append(a)
    return clusters

# Returns true if __Cluster will actually make any difference,
# i.e. is there more than one group that can be split
def __ClusterTest(images, time ):
    canCluster = False;
    for i in range(1, len(images)):
        if images[i][0] - images[i-1][0] >= time :
            canCluster = True
    return canCluster

# This is a recursive function that will return the clustered
# image set images is a list of tuples, the first entry of
# each tuple (list[i][0]) should be the time.  Any other
# entries are irrelevant, but will get copied along when the
# clustering is performed.
# Default minTime is 5 minutes and Interval is 10 minutes
def Cluster( images, time, minTime=300, interval = 600 ) :
    aList = []
    if __ClusterTest( images, time) and time > minTime and len(images) > 1 :
        newList = __Cluster( images, time )
```

```
    for i in range(len(newList)):
        aList.append( Cluster(newList[i], time - interval, minTime, interval) )
    return aList
else :
   return images
```

References

[1] Benjamin B. Bederson. Photomesa: a zoomable image browser using quantum treemaps and bubblemaps. In *UIST '01: Proceedings of the 14th annual ACM symposium on User interface software and technology*, pages 71–80. ACM Press, 2001.

[2] John Boreczky, Andreas Girgensohn, Gene Golovchinsky, and Shingo Uchihashi. An interactive comic book presentation for exploring video. In *CHI '00: Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 185–192. ACM Press, 2000.

[3] Michael G. Christel, Alexander G. Hauptmann, Howard D. Wactlar, and Tobun D. Ng. Collages as dynamic summaries for news video. In *MULTIMEDIA '02: Proceedings of the tenth ACM international conference on Multimedia*, pages 561–569. ACM Press, 2002.

[4] Adobe Corporation. Photoshop elements version 4.0. Computer Software, October 2005.

[5] Google Corporation. Picasa. Computer Software, available at http://picasa.google.com/index.html, November 2005.

[6] Kodak Corporation. Kodak easy share gallery. http://www.kodakgallery.com, November 2005.

[7] Nicholas Diakopoulos and Irfan Essa. Mediating photo collage authoring. In *UIST '05: Proceedings of the 18th annual ACM symposium on User interface software and technology*, pages 183–186. ACM Press, 2005.

[8] Steven M. Drucker, Curtis Wong, Asta Roseway, Steven Glenner, and Steven De Mar. Mediabrowser: reclaiming the shoebox. In *AVI '04: Proceedings of the working conference on Advanced visual interfaces*, pages 433–436. ACM Press, 2004.

[9] James Fogarty, Jodi Forlizzi, and Scott E. Hudson. Aesthetic information collages: generating decorative displays that contain information. In *UIST '01: Proceedings of the 14th annual ACM symposium on User interface software and technology*, pages 141–150. ACM Press, 2001.

[10] Michael L. Gleicher, Rachel M. Heck, and Michael N. Wallick. A framework for virtual videography. In *Proceedings SmartGraphics 2002*. University of Wisconsin-Madison, Department of Computer Sciences, June 2002.

[11] Intel. Intel image processing library. URL http://developer.intel.com/vtune/perflibst/ipl/index.htm.

[12] Intel. Intel open source computer vision library (opencv). URL http://www.intel.com/research/mrl/research/opencv/.

[13] "Laurent Itti, Christof Koch, and Ernst Niebur". "a model of saliency-based visual attention for rapid scene analysis". *"IEEE Transactions on Pattern Analysis and Machine Intelligence"*, "20"("11"):"1254–1259", "1998".

[14] J. Jeon, V. Lavrenko, and R. Manmatha. Automatic image annotation and retrieval using cross-media relevance models. In *SIGIR '03: Proceedings of the 26th annual international ACM SIGIR conference on Research and development in informaion retrieval*, pages 119–126. ACM Press, 2003.

[15] Yin Li, Jian Sun, Chi-Keung Tang, and Heung-Yeung Shum. Lazy snapping. *ACM Trans.Graph.*, 23(3):303–308, 2004.

[16] Feng Liu and Michael Gleicher. Automatic image retargeting with fisheye-view warping. In *UIST '05: Proceedings of the 18th annual ACM symposium on User interface software and technology*, pages 153–162. ACM Press, 2005.

[17] Ludicorp Research & Development Ltd. Flickr. http://www.flickr.com, November 2005.

[18] P.C. Magazine. Riya alpha. http://www.pcmag.com/article2/0,1895,1885030,00.asp, November 2005.

[19] P.C. Magazine. Tag world beta review. http://www.pcmag.com/article2/0,1895,1884543,00.asp, November 2005.

[20] J Platt. "autoalbum: Clustering digital photographs using probabalistic model merging", proceedings of ieee workshop on content-based access of image and video libraries (cbaivl-2000) ieee press,". In *Proceedings of IEEE Workshop on Content-based Access of Image and Video Libraries (CBAIVL-2000)*. IEEE Press, 2000.

[21] Carsten Rother, Sanjiv Kumar, Vladimir Kolmogorov, and Andrew Blake. Digital tapestry. In *CVPR (1)*, pages 589–596, 2005.

[22] Frederik Schaffalitzky and Andrew Zisserman. Multi-view matching for unordered image sets, or "how do i organize my holiday snaps?". In *ECCV '02: Proceedings of the 7th European Conference on Computer Vision-Part I*, pages 414–431. Springer-Verlag, 2002.

[23] Uri Shaft and Raghu Ramakrishnan. Data modeling and querying in the piq image dbms. *IEEE Data Eng.Bull.*, 19(4):28–36, 1996.

[24] Ben Shneiderman and H. Kang. Direct annotation: A drag-and-drop strategy for labeling photos. In *IV '00: Proceedings of the International Conference on Information Visualisation*, pages 88–95. IEEE Computer Society, 2000.

[25] Rohini K. Srihari and Zhongfei Zhang. Show & tell: A semi-automated image annotation system. *IEEE MultiMedia*, 7(3):61–71, 2000.

[26] Bongwon Suh, Haibin Ling, Benjamin B. Bederson, and David W. Jacobs. Automatic thumbnail cropping and its effectiveness. In *UIST '03: Proceedings of the 16th annual ACM symposium on User interface software and technology*, pages 95–104. ACM Press, 2003.

[27] Shingo Uchihashi, Jonathan Foote, Andreas Girgensohn, and John Boreczky. Video manga: generating semantically meaningful video summaries. In *MULTIMEDIA '99: Proceedings of the seventh ACM international conference on Multimedia (Part 1)*, pages 383–392. ACM Press, 1999.

[28] Michael Wallick and Michael Gleicher. Magic boards. In *SIGGRAPH Posters*, 2005.

[29] Michael N. Wallick, Rachel M. Heck, and Michael L. Gleicher. Chalkboard and marker regions. In *Mirage 2005 - Computer Vision/Computer Graphics Collaboration Techniques and Applications*, March 2005.

[30] Gang Wei and Ishwar K. Sethi. Face detection for image annotation. *Pattern Recogn.Lett.*, 20(11-13):1313–1321, 1999.

[31] Liu Wenyin, Yanfeng Sun, and Hongjiang Zhang. Mialbum - a system for home photo managemet using the semi-automatic image annotation approach. In *MULTIMEDIA '00: Proceedings of the eighth ACM international conference on Multimedia*, pages 479–480. ACM Press, 2000.

[32] Wikipedia. Exchangeable image file format. http://en.wikipedia.org/wiki/EXIF, November 2005.