2007

Automation in digital photo management

Benjamin Arthur Brady
Iowa State University

Automation in digital photo management

by

Benjamin Arthur Brady

A thesis submitted to the graduate faculty

in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

Co-majors: Information Assurance; Computer Engineering

Program of Study Committee:
Doug Jacobson, Major Professor
Tom Daniels
Barb Licklider

Iowa State University

Ames, Iowa

2007

UMI Number: 1443142

ProQuest Information and Learning Company
300 North Zeeb Road
P.O. Box 1346
Ann Arbor, MI 48106-1346

DEDICATION

I would like to dedicate this thesis to my parents, Steve and Donna. Thank you for all that you do for me. I would also like to thank the rest of my family and friends for their loving guidance. Special thanks to my friends, Brooke and Faith, for helping me tremendously in the editing and revising process. To my major professor, Dr. Jacobson, thank you for your guidance, thoughts, and direction along the way.

TABLE OF CONTENTS

LIST OF FIGURES vi

CHAPTER 1. OVERVIEW 1

1.1 Introduction 2

1.1.1 Problem 2

1.1.2 Why is this important? 2

1.1.3 Solution 6

CHAPTER 2. BACKGROUND 7

2.1 The photo management process 7

2.1.1 Film prints 7

2.1.2 Digital photos 8

2.2 Terminology 9

2.2.1 Digital photo management 9

2.2.2 Digital albuming 10

2.2.3 Content-based image retrieval 11

2.2.4 Event-based image retrieval 11

2.2.5 Facial recognition 12

2.2.6 Face-based image grouping 14

2.2.7 Manual annotation 15

2.2.8 Automatic annotation 17

CHAPTER 3. LITERATURE REVIEW 19

3.1 Digital photo management 19

3.1.1 Preparation of the Rodden and Wood study 20

3.1.2 Digital vs. non-digital 20

3.1.3 Annotation 21

3.1.4 Browsing and searching 21

3.1.5 Taking and using photographs 23

3.2 Content-based image retrieval 23

3.2.1 Searching 24

3.2.2 Levels of searching 25

3.2.3 Semantic gap 28

3.2.4 Extracting features with CBIR 28

3.3 Event-based image retrieval 28

3.3.1 Automated Event Clustering and Quality Screening of Consumer Pictures for Digital Albuming 29

3.3.2 Home Photo Content Modeling for Personalized Event-Based Retrieval 31

3.4 Face-based image grouping 32

3.4.1 Overview 33

3.5 Automatic image annotation 34

3.5.1 Automatic Linguistic Indexing of Pictures (ALIP) 34

CHAPTER 4. PROBLEM 36

4.1 Details of the problem 36

4.1.1 Room for improvement 37

4.2 Why is this worth doing? 39

4.2.1 People are always looking to save time and money 40

4.2.2 Images have become an essential part of many professions 40

CHAPTER 5. A SOLUTION 43

5.1 How can the automation studies and digital photo management work together? 43

5.1.1 The ideal solution 43

5.2 How to combine the systems 47

5.2.1 Input and output 47

5.2.2 Priorities 48

5.2.3 Interaction 49

5.2.4 Testing 49

5.3 Security 50

CHAPTER 6. CONCLUSION 52

6.1 Conclusions drawn 52

6.2 Contributions 53

6.3 Future research 53

BIBLIOGRAPHY 55

LIST OF FIGURES

Figure 2.1 My mother has to use two such book shelves for all of the family albums. 8

Figure 2.2 An example of facial recognition in Adobe Photoshop Elements 5.0. 14

Figure 2.3 The list of labels one is able to drag to a face in Adobe Photoshop Elements 5.0. 14

Figure 2.4 The properties of a digital photo in Microsoft Windows XP. . . . 16

Figure 2.5 The advanced properties of a digital photo in Microsoft Windows XP. 17

Figure 3.1 The Event Taxonomy from (17). 32

CHAPTER 1. OVERVIEW

"A picture is worth a thousand words." With today's technology and the amount of pictures available, this means an infinite amount of words. Researchers have been working on finding a way to sort through the pictures and the words associated with them for more than 20 years. Until recent years, most people wrote the words on the back of the pictures in order to organize them. With the digital age upon us and digital cameras and photographs becoming the norm, physically writing on them has become less practical. Instead, several ways to label digital photos have been proposed and implemented.

For digital photos, labeling may range from renaming the picture file, to putting a group of pictures in a labeled folder, to adding descriptors in photo management software. However, these ways are all manual. In other words, they involve the person remembering information about the photo or deducing information by viewing it. The current software does not provide intuitive help in the process.

Currently, digital photo management software has saved both home users and varied professional users a great deal of time and money by helping the digital image annotation process (A process of using captions or keywords to automatically label pictures. See Sections 2.2.7 and 2.2.8). Researchers are working on automating many aspects of this process. Many of them focus on one aspect of improving digital photo management. However, with all of the proposed or implemented solutions to the process, they need to be integrated in order to make an ideal system. This thesis aims to show how digital photo management could be improved with a union of automation studies.

1.1 Introduction

The rest of this chapter will give an overview of the challenges in improving digital photo management, why it is important, and what solutions researchers have found. Chapter 2 will give an overview of the digital photo management process. In addition, it will define concepts and terms for the purposes of this paper. Chapter 3 will give a review of recent studies of concepts directly related to digital photo management and automatic annotation. Chapter 4 will describe the problem in more detail. In Chapter 5, a detailed solution is proposed. Finally, Chapter 6 will describe what conclusions were drawn, and future research.

1.1.1 Problem

Currently, digital photo management software aids in organizing digital photos. The software is not limited to home users, as many professions deal with millions of images daily. Section 4.2.2 describes several professions and how they have a need for digital photo management and automatic image annotation. While digital photo management software improves digital photo organization, the software needs to include current research on automatic image annotation. Even though we are in the digital age, the idea of manually annotating digital photos is still comparable to labeling film photos. Several studies and research have proposed solutions, but many of them have focused merely on one detailed solution for one specific concept. Granted, this is nothing new and is how research is supposed to be done, but currently, combining much of the research and solutions would improve digital photo management more than each solution alone.

1.1.2 Why is this important?

The importance of digital photo management and automatic image annotation may not be clear. This section will cover several reasons why digital photo management and automatic image annotation have become essential in recent years. The first section will cover how the average person has become a factor in studies on digital photo management and automatic image annotation. The second section will discuss the emergence of digital cameras and digital photography. The third section will clarify the difficulty of organizing photos. Finally, the fourth section will cover how automation would save time and money.

1.1.2.1 Average person

Analyzing photos and automatically annotating them is not a new topic of interest. The scientific community has used various technologies in order to perform these tasks. However, an average person or organization could also benefit from these technologies for a personal (or at least non-scientific) purpose. Not only could an individual have a significant amount of images, they also may not have the time to analyze all of them for personal purposes. Thus, digital photo management is not directed solely at the scientific community or various professions. One person could have a significant need for a type of digital photo organization, and the current applications could be improved by coordinating the studies in digital photo management.

One significant application for the average person is Internet filtering software. A tool called "Web Guard" has optimized the filtering experience by including visual content analysis with streaming (Internet) data (9). Web Guard looks for skin color in addition to other filtering technology. The same technology could be used to find faces in a photo for facial recognition.

In dealing with personal photo albums, an increasing number of people have digital cameras. As a result, an individual could have hundreds or thousands of pictures from various activities throughout that person's life on one computer. See Section 1.1.2.2 for more details. Automatic image annotation would allow a person to take a picture, upload it to the software, and the software could automatically classify the picture. This would be based on the actual content of the picture instead of only a person's description, which does not have to coincide with actual content.

1.1.2.2 Digital cameras

The camera industry, today, is much different from a few years ago, seen by digital cameras outselling film cameras. As a result, many more people have gathered digital images on their computers. In fact, the International Data Corporation (IDC) has given several statistics for the past few years about digital camera growth both in the United States and worldwide (2).

- 29.8 million digital cameras were sold in the USA in 2006, up 5 percent from 2005. (March 7, 2007)
- 42 percent of Americans projected to have a digital camera by the end of 2004 (February 11, 2004)
- Digital cameras projected to outsell film cameras in the U.S. in 2003, worldwide in 2004. (December 23, 2003)

These statistics show the digital camera industry is not on the decline. As the number of people with digital cameras increases, so will the number of digital photos. With more digital photos, a person is more likely to find the need to organize them.

With digital cameras as the most popular kind of camera today, digital photos are more prominent. With a digital camera, a person can snap a picture and evaluate if they want to keep it instantly, with no expensive film nor time to develop involved. In addition, memory cards on digital cameras hold significantly more pictures than a roll of film. Not only are there more pictures, there is a larger variety of pictures as well. A person does not have to decide whether a scene is ideal for a photo or not. Depending on the capacity and the quality, one memory stick could hold hundreds of pictures. As a result, organizing these photos, after they have been uploaded to a computer, can become a more daunting task.

From a study on how people organize their digital photos, people usually upload all the pictures from a memory stick chronologically (22). In other words, the folders containing the pictures are only labeled by date. By default in Microsoft Windows XP Picture and Fax Viewer, the filenames of the photos themselves are just consecutive numbers as well. Any number of people, events, places, or time periods could exist in each folder. As a result, digital photo management becomes essential.

1.1.2.3 Difficultly with organizing photos

With digital cameras more popular than ever, consumers enjoy taking and storing millions of photos. However, many times these photos remain in the memory stick or original folder into which they were uploaded. Organizing these photos into folders or even ready-to-view albums is a tedious and time-consuming process. Ideally, a person wants to organize photos in a few ways: event, time/date, place, and people [4]. Before digital photos, the potential for automatic organization and albuming did not exist. A person had to go through several steps such as developing and labeling before organization could even begin. No practical way of automatically organizing photos existed, other than finding a way for someone else to do the organization.

1.1.2.4 Automated process would save time and money

Many areas of our everyday lives have benefited from automation. The telephone, microwaves, and mail service were all invented to automate tasks. All three saved people time and money. Another example is computers. In the last ten years, computers have automated many daily tasks. In addition, communication between people has become easier and more cost effective with e-mail. These few examples show how automated processes in general have helped save time and money. Section 4.2.1 describes this concept in more detail.

1.1.3 Solution

The solution to these challenges in improving digital photo management is to combine previous research. Areas such as content-based image retrieval, event-based image retrieval, and face-based image grouping markedly improve digital photo management. However, if those areas and others could work together in one application, the improvement would be greater. In addition, many of the applications, based on the research in the specific areas, do exactly what they are supposed to accomplish, but it seems something is missing from each of them. Chapters 2 and 3 will describe the main research areas, and Chapter 4 will analyze each of the areas.

CHAPTER 2. BACKGROUND

In this chapter the concepts related to digital photo management and automation are explained. The first section will cover the history with photo management. The second section will describe several terms needed in order to be able to understand the proceeding chapters.

2.1 The photo management process

For film cameras, the photo management process takes a considerable amount of time and money. In the following sections, the process of organizing photos from a film camera are described. Then, the same process with a digital camera is listed.

2.1.1 Film prints

After having pictures developed, consumers had the daunting task of figuring out what to do with them. My mother, Donna Brady, stated that she usually kept the packets of developed photos for viewing for a short period of time (usually a month), during which she would spend a couple days writing information on the back of the photos in each packet. This usually included the date the photo was taken, the event the photo was taken for, who is in the photo, and where the photo was taken. When there are almost fifty pictures in one packet (a roll of 24 with doubles), this process took a few hours per packet. Then, once a year, she put the best photos into the decided albums. Since she had at least one album for each of my two brothers and me and at least one for my dad and herself, this was not a short process. The photos had to be organized in chronological order as the albums were already in that order. Figure 2.1 shows an example of the amount of physical storage space it takes to store family albums.

Figure 2.1 My mother has to use two such book shelves for all of the family albums.

My mother is only one example of what a person does with pictures after they have been developed. Various other ways of organizing developed photos exist. With more photos, it appears more difficult and more time consuming to make sure all of them are organized correctly. In addition, space becomes an issue when the amount of albums exceeds the storage space.

2.1.2 Digital photos

Although digital photos are still often printed (developed), a photographer has much more flexibility after digital photos are taken. Users have the choice of having a business print them directly from the memory stick, leaving them on the camera (memory stick), or uploading them to their computer. Currently, many businesses that offer digital photo printing allow users to choose which photos and the number of each photo to print. By leaving photos on the camera, users are able to show others the most recent photos they have taken.

Digital photo management is not involved in the first two choices. Many times, the first action users take with digital photos is to upload them to their computer. This is where digital photo management can be applied. Several applications aid users in the upload process. Simple applications allow users to create a filename for all of the photos, make a new folder for the photos, and choose which photos to upload. More advanced applications may allow users to perform more organization tasks both during and after the upload process. The focus of this paper is for organization after the upload process.

2.2 Terminology

This section describes several terms directly related to digital photo management and automation. A few of the sections will offer a general overview with more detail included in Chapter 3 (Digital photo management, content-based image retrieval, event-based image retrieval, face-based image grouping, and automatic image annotation).

2.2.1 Digital photo management

The phrase digital photo management covers a broad area in both research and terminology. It involves research on how to organize digital photos including the habits of an average person. It also encompasses the details on the process of recognizing objects or faces in a photo. How people organize their digital photos, Content-based image retrieval, event-based image retrieval, face-based image grouping, digital albuming, facial recognition, and automatic image annotation are all research areas in digital photo management. The term management itself includes organizing and planning in addition to editing. Hence, digital photo management includes any type of manipulation of photos and the decision process in choosing photos for various purposes.

A digital photo management system is a way for a user to organize photos on a computer. Many applications exist in digital photo management, but they have a variety of functions. Most computers include some type of digital photo management software. In Microsoft Windows XP, for instance, Windows Picture and Fax viewer could be considered a simple form of digital photo management software. It has limited functionality, though. Its main function is for viewing photos, but also includes deleting, printing, and copying. One could use it to manually view photos and save them to various folders on a computer.

More complex digital photo management systems offer more organizational utility such as labeling folders, manual annotation, and archiving. In addition, developers have added more functionality to the simple utilities such as viewing or browsing. Many of these systems are built for larger numbers of digital photos (thousands to millions).

Users are not limited to printing photos before they are uploaded to their computer. Instead, they may upload pictures to web sites to order prints from various businesses. Digital photo management software may also include options to order prints. Users can buy photo printers and special paper to print digital photos themselves.

2.2.2 Digital albuming

For the purposes of their study on event clustering and quality screening for digital albuming, authors Alexander C. Loui and Andreas Savakis (18) define digital albuming as "a collection of processes for segmenting pictures into events and subevents and generating an album page layout after screening low-quality and duplicate images." In other words, digital albuming goes a step further than organizing digital photos. It creates a layout to present a group of photos based on the photos themselves. This is similar to a small physical photo album for a specific event such as the birth of a child or a family vacation.

2.2.3 Content-based image retrieval

Content-based image retrieval is one area of research pursued in the past 20 years. In (6), John P. Eakins and Margaret Graham give a definition of content-based image retrieval based on the first mention of it in (11), "The process of retrieving desired images from a large collection on the basis of features (such as color, texture, and shape) that can be automatically extracted from the image themselves." In other words, content-based image retrieval uses images and various parts of the images to automatically select other images. One of the key words in the definition is automatic. The retrieval process has to be automatic in order to be considered content-based image retrieval. In addition, the query needs to be an image; retrieval by manually annotated keywords is not content-based image retrieval.

One other distinction Eakins and Graham make is between automatic face recognition systems and content-based image retrieval. In content-based image retrieval, one method searches a database for close matches to the image on the camera. Another method compares the image on camera (in a security system) with one recorded to verify identity. This is facial recognition and would not be considered content-based image retrieval. The details of how content-based image retrieval works and the different types of queries are discussed in Section 3.2.

2.2.4 Event-based image retrieval

As described in Chapter 1, one of the four ways to search for and classify images is event-based. When a person takes a digital photograph at a particular event (vacation, party, etc.), the set of those photos could possibly be similar in image background, date, and time. For an application, event-based image retrieval is difficult because no prior information exists about the photos. In a study on this subject, researchers combine date and time information with how similar the content in a group of photos appears

(18). The first level in this process is clustering events by date and time (18). Many times, if a person takes a vacation, most of the photos will be from a certain time period. If there is a large time difference between two groups of photos, one group could be from a different event. Since most people usually arrange their photos in some kind of chronological order, this information is quite useful in clustering (18). In addition, if this is a family or another group of people at the same event, more than one person's pictures could be grouped for the same event based on the date and time. This second level of analysis confirms content differences when the two groups are analyzed for similarity. See Section 3.3 for more details on event-based image retrieval.

2.2.5 Facial recognition

Technology has included facial recognition for several years now. For example, in law enforcement, facial recognition in security cameras began to gain popularity in the early 2000s (1). The system can recognize certain faces when they appear on the camera. One main application for this is exemplified in casinos when certain people are flagged to be watched. Facial recognition has also been gaining momentum in the digital photo management field. Recognizing specific faces, such as family members and friends, is still in research and development, but recognizing faces in a general form has already appeared in software applications.

For instance, Adobe Photoshop Elements 5.0 has a facial recognition feature. In this application, a user uploads photos to the program, and then it will recognize all the faces in the photos. From there, the user has to annotate to what people they belong. In addition, faces that have already been annotated are not automatically annotated in other photos. The user still has to manually annotate them. This is in contrast to facial recognition described in Section 2.2.3, it is merely a recognition of the presence of a generic face.

To advance available applications, in 1999, researchers at Hewlett Packard Laboratories proposed an experimental system called FotoFile (12). FotoFile will not only recognize specific faces, it will also try to match that face to all other photos. Some manual annotation is still involved, though. When a face FotoFile does not recognize appears in a photo, it calls for the user to name that person. This kind of manual annotation is not avoidable. Even humans cannot automatically recognize the face of someone they have never seen nor met. The facial recognition feature in FotoFile is closer to the current ideal than many applications, but it still requires manual annotation. When it recognizes a specific face, the user needs to confirm FotoFile was correct. In other words, FotoFile uses an educated guess for facial recognition, leaving it still short of ideal.

2.2.5.1 Current applications

As discussed in 2.2.5, Adobe Photoshop Elements 5.0 has a facial recognition feature, which is used to find faces in photographs. However, after the program finds the faces, it is up to the user to determine and tag the name of that person. The software also does not analyze the rest of the pictures in the folder for the same person. Instead, the user has to drag the face manually to that specific person's user-created tag folder or vise versa. See Figure 2.2 for an example of the facial recognition feature in Adobe Photoshop Elements 5.0. The original picture is a Christmas card photo of my family.

The next step in the process is dragging a label to each picture. This is shown in Figure 2.3. The facial recognition inside the program is useful, but it has more potential. In the future, a software program similar to Adobe Photoshop Elements 5.0 may be able to analyze a given folder and automatically tag what (or who) is in the picture. Currently, tagging is still mostly a manual process. The software has automated a time consuming part of tagging digital pictures, though.

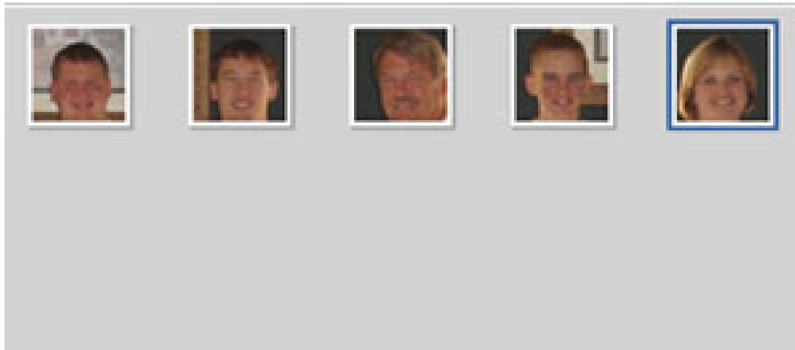

Figure 2.2 An example of facial recognition in Adobe Photoshop Elements 5.0.

Figure 2.3 The list of labels one is able to drag to a face in Adobe Photoshop Elements 5.0.

2.2.6 Face-based image grouping

Another area of digital photo management classification is face-based image grouping. Typically, a person may want to find all of the photos with a particular person or a large group of people in them. One way to accomplish this task is to cluster similar faces for a person to label. A study from the Eastman-Kodak company describes a system for face-based image grouping and automatic albuming (4). Authors Madirakshi Das and Alexander C. Loui, present the difficulty of identifying faces due to large differences in lighting, orientation of the faces, and the passage of time. The differences in lighting may be from photographs from both indoor and outdoor events. In addition, the faces in each photograph may not always be looking directly at the camera when a picture is taken. Plus, in the case of a family, the faces change over a period of time. These problems and the details of face-based image grouping will be discussed in Section 3.4.

2.2.7 Manual annotation

As described in Chapter 1, manual annotation of digital photos is similar to labeling a non-digital photo by writing on the back. Manual annotation aids users with searching and browsing digital photos. With a variety of photos in each folder on a computer, manual annotation seems like a necessary task, but it may not be any better than the annotation of developed or printed photos. In many cases, the manual annotation of digital photos is no better at saving time and money than labeling non-digital photos. For instance, without even simple software, a person cannot do advanced annotations. Granted, the digital annotations from the digital camera are present, but they are only simple text boxes. See Figure 2.4 for the text boxes in the properties of a digital photo.

These properties only help when a person wants to see the details about a photo. The properties can be sorted from the folder menus. However, no real grouping can happen, only alphabetical or numerical sorting. When viewing the advanced properties, this information is not editable. See Figure 2.5 for a view of the advanced properties.

A person could manually change the filenames of all the pictures to give annotations, but this is no different than writing on the back of a developed photo. The tedious process of manually changing the properties of a photo often includes more time per photo than from developed photos. A person may have to switch from viewing the properties window to viewing the picture multiple times.

Manual annotation also leads to ambiguity. For instance, a person can be identified in many different ways. Initials, nicknames, and various permutations of full names are all types of annotations a user could put for one digital picture (ex. Ben, Benjamin, Ben B., etc). Also, specific photos may require greater differentiation, i.e., a picture of two

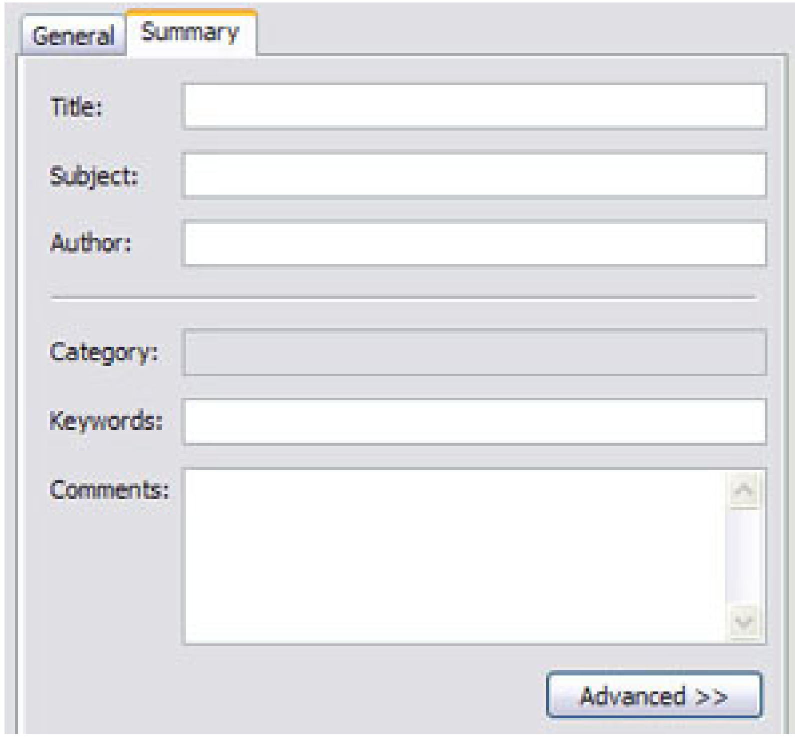

Figure 2.4 The properties of a digital photo in Microsoft Windows XP.

different people with the same first name. This is especially the case if multiple users are annotating pictures. Ambiguity occurs even with objects such as cats or kittens, lights or lamps, and houses or homes. Even if a user is meticulous with manual annotations, discrepancies are inevitable. Thus, inconsistency is one of the major faults with manual annotation. Granted, automatic annotation will likely not be perfect, but it should be easier to be more consistent with such annotations.

Time is a constraint with manual annotation of digital photos as well. Many software applications have improved this constraint through user interfaces. Several applications allow the user to drag-and-drop pictures in order to annotate them. This does cut down on the time it takes for each picture, but the user still has to go through each picture. In many cases, one picture needs several annotations for various objects, scenes, and people. This is not only a concern with time but in consistency as well. The user has to decide what objects and how many objects to annotate.

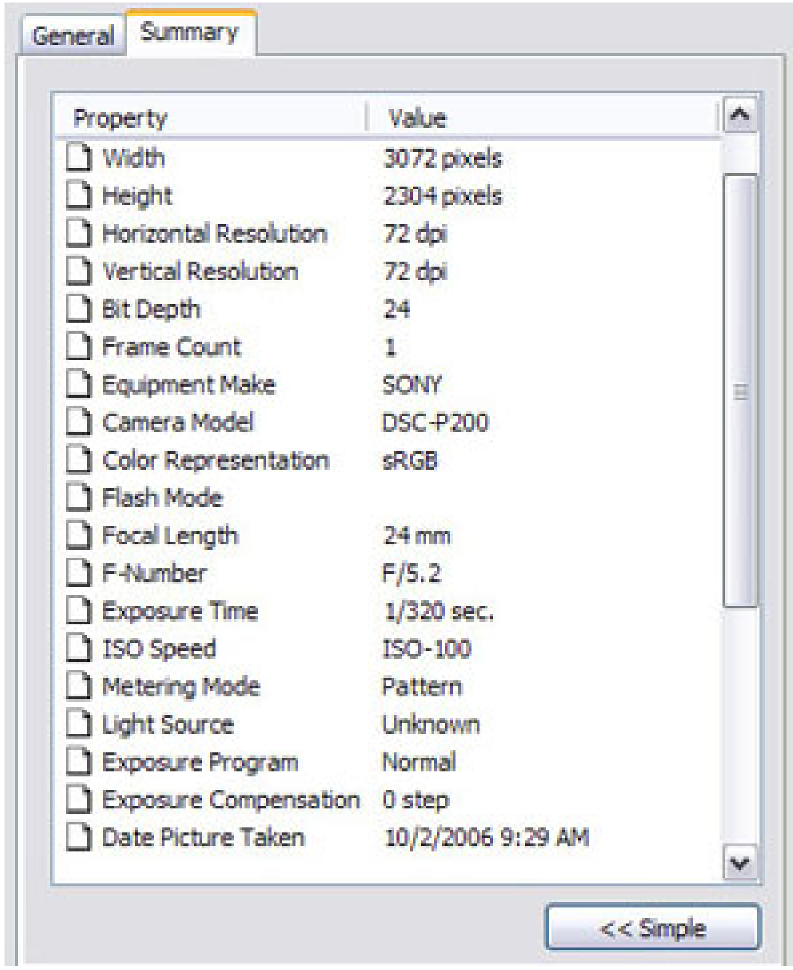

Figure 2.5 The advanced properties of a digital photo in Microsoft Windows XP.

2.2.8 Automatic annotation

Many people have likely wished they could have paid someone to organize all of their non-digital photos in the past. People want to organize them, but they are constrained by many factors. Automatic annotation automates the manual annotation process. Instead of remembering where and when a picture was taken, who was in it, and why the picture was taken, a computer application has the potential do it automatically. Automatic annotation applications can analyze photos for objects, people, scenes, or other parts of a photo and label the photos with little user intervention. Thus, it can aid in the browsing, searching, and general organization process.

With digital photos, time is still a factor. Automatic annotation for digital photos is one solution to the constraint of time discussed in Section 2.2.7. Other than the aforementioned essential annotations above, the program could easily add descriptive annotations such as trees, cars, night, day, happy, sad, etc. These annotations could be used in the future to find specific pictures quickly. Instead of having to remember which album or folder a picture is in, the program will allow more advanced searching capabilities. The annotations can be used to organize the pictures, as well as, aid in finding them. Instead of sorting by the default categories of the properties of a digital photo, the annotations can be used for more advanced groupings.

CHAPTER 3. LITERATURE REVIEW

This chapter will give the current state of research on digital photo management and automation. The first section will cover the management of digital photos. The study includes a comparison on how the participants organize both digital and non-digital photos and their preferences with annotation, browsing, and searching. The second section will discuss details with content-based image retrieval such as the levels of search queries and the semantic gap. The third section will cover two studies on event-based image retrieval and clustering. They include quality screening for digital albuming, event taxonomies, and visual keywords indexing. The fourth section will give an overview of a study on face-based image grouping including its clustering and lableling process, with the final section covering a study on an automatic image annotation application.

3.1 Digital photo management

Kerry Rodden and Kenneth R. Wood present a study on how people manage digital photographs (22). The study is based on a small group of participants and a prototype digital photo management system. They answer two questions in the study.

- How do people organize and browse their digital photo collection (22)?
- Is content-based image retrieval useful in the context of personal photo collections (22)?

The analysis of the first question also addresses the difference between organizing and browsing digital photo collections and non-digital collections. Included in the second question are other forms of retrieval such as using speech recognition, however, this is beyond the scope of this paper.

3.1.1 Preparation of the Rodden and Wood study

In their study, Rodden and Wood give the users Shoebox (20), an application to browse and organize photos. Results from a previous study on the different ways people organize existing photos were used to design Shoebox. The Rodden and Wood study used thirteen employees at AT&T Laboratories Cambridge (22). The members were both male and female, and they ranged in age from 24 to 38 (22). The average size of the participant's non-digital photo collection was roughly 1000 photos (22). Finally, the participants were given digital cameras with which to take photos during a period of six months in 2000 (22).

Two interviews were conducted, one before and one after the study. In the interviews, the participants were asked to discuss digital and non-digital photo management and how Shoebox aided them throughout the study. This resulted in several ideas and points of interest relative to this paper as will be referenced in the succeeding sections.

3.1.2 Digital vs. non-digital

In the interviews before the study, the participants discussed their practices with non-digital photos. They stated they have put photos into albums but not consistently nor often. Most of them separated low-quality or undesirable photos from the photos they put in the albums. In addition, most of the albums had a specific event associated with them and were kept in chronological order (20). Many of the participants kept their photos in packets for long periods of time.

In organizing digital photos, the participants either separated them into folders by event or just put all the photos in the camera's memory in a folder. The interviews and surveys showed that the participants felt more organized with digital photos even though they did not put any more effort into organizing the digital photos.

3.1.3 Annotation

In Section 2.1.1, the example was givien of my mother who writes on the back of all of her photos. However, this study confirms that not everyone writes on every photo. The authors point out that not all of the participants write on the back of photos, and those that do only occasionally do so. Some of the participants only write on the packet to describe a group of pictures in essence naming only the file.

Even though Shoebox aids in the ability to annotate photos, the participants used it sparingly. Most of them changed the name of a roll (a folder in Shoebox), but only a few changed the name of a photo or added annotations to a photo (22). When one of the participants did annotate a photo or change the name, they tended towards adding names of people and places. The date was usually left out based on the fact that the digital photos contain the date and time by default.

The reason the participants did not annotate more photos, the authors propose, is because most of the photos had been taken recently (22). In other words, the locations of most of the photos were so fresh in the participants' minds they did not feel the need to annotate the photos. Finally, most of the participants stated that they would only want to annotate some and not all of the photos because annotating all of them would not be worth their effort (22).

3.1.4 Browsing and searching

During the study, Rodden and Wood also discussed the browsing and searching process with the participants. First, for both digital and non-digital photos, the participants stated that they tended to look at recently taken photos less frequently as time passed (22). For non-digital photos, this meant leaving the photos in the packets in a convenient place for a period of time. This is similar to leaving photos on the camera or memory stick for a period of time to show others. Once uploaded to the computer, photos may require browsing and searching after a lengthy period of time.

The authors outline three main types of queries (22).

- photos from a specific event
- one specific photo
- photos from multiple events with something in common (person, object, etc.)

The queries listed are in order of most frequent to least frequent. Searching for non-digital photos varies in difficulty depending on the owner's organization. For photos from a specific event, a search with non-digital photos is relatively easy when a person has specific albums for significant events. Searching for a specific photo is slightly more complicated.

In the study, the participants gave the steps they took to locate a particular photo. They had to recall the event, the date it occurred, the particular album, and then, they had to pick a point in the album and browse forward or backward (22). This process becomes much more difficult when the photos are unorganized or in multiple locations. The authors mention that the motivation for organizing photos is not primarily for searching but for presentation to others.

In contrast, digital photos do not have these problems. Even without any additional software, photos are easily sorted in chronological order and copied to multiple folders. The authors cite (14) to point out that this makes it easier to search because it is easier to remember the events relative to other events rather than its actual occurrence (22).

Finally, the authors discussed the possible reasons behind a more general search. This could consist of a specific person in multiple pictures or finding a high quality photo from among several similar photos (22). They point out that having pictures in chronological order offers little or no help regardless of whether it is a digital or non-digital collection (22). In addition, a search for all the photos of a person or object is not as common such as in the case of a funeral following a death in the family, a wedding, significant birthdays, retirement, or an anniversary of marriage (22).

3.1.5 Taking and using photographs

Rodden and Wood also observed the difference between taking digital and non-digital photos. In this short study alone, the number of digital photos the participants accumulated was half of the average size of their entire previous non-digital collection. They confirm the reason for this is that it costs relatively nothing to take a digital photo, and that many more images can be accumulated in the memory than on a roll of film (22). Thus, the participants took multiple digital photos where they normally would have taken one photo (22). They also reasoned that because people do not risk anything by taking a poor quality digital photo, they are more likely to take that risk. It is understandable then, why the participants' digital collections were already at half the amount of their non-digital collections.

3.2 Content-based image retrieval

An average search for specific photos on a computer normally consists of either running the file searching program from the operating system or remembering in which folder the photo might be located. The searching program is flexible only in basic file properties (filename, date created, date modified, etc). Once the photo is located in a certain folder, a person has to browse each photo for the desired one. With a larger collection this type of search is not feasible.

In content-based image retrieval (CBIR), a different type of search is proposed. According to Eakins and Graham (6), content-based image retrieval is "the process of retrieving desired images from a large collection on the basis of features (such as color, texture, and shape) that can be automatically extracted from the images themselves." Instead of searching based on manually created annotations or on the file properties, the search is based on the contents and qualities of the photo itself.

As humans, this is what we typically do when we look for photos anyway. Sometimes we may look for a person, an object, or even a color in a photo. The only reason we use the average searching methods is because it was the only method available. CBIR has not been perfected yet. It is still not consistently available in commercial photo management software, but there are several areas of research with regard to CBIR.

3.2.1 Searching

Commonly, if users needs to find a photo, they have folders labeled by the date they uploaded the pictures, with which they browse for the correct photo. Many times, they have not annotated the pictures nor have they made any effort to organize them. Without actually seeing the photos, they do not know what types of pictures are in each folder since the dates could be from several months ago. Section 3.1 discussed how people manage their digital photographs and describes several reasons why a person would want to search for a photo or set of photos. People tend to agree that searching for digital photos is easier than searching for non-digital photos (22). Even though the study in (22) had a small amount of participants, they were all focused on comparing digital and non-digital photos collections. Even without CBIR, digital photos are easily sorted chronologically or by name, whereas with non-digital photos, the same sorting could take hours or days.

3.2.2 Levels of searching

A CBIR search could consist of three levels. The first level is by primitive features, these features may consist of color, texture, shape, or the location of different elements or objects in the photo (6). Eakins and Graham list some examples of search queries from this level.

- Find pictures with long thin dark objects in the top left-hand corner
- Find pictures containing yellow stars arranged in a ring
- Find more pictures that look like this (a more general form)

These types of features are derived from the photos themselves without prior knowledge (6). For example, color is easily extracted from a picture based on each pixel and the RGB scheme. The program or person does not need to know any information about the photo in order show or infer the colors in the photo.

3.2.2.1 Level 1

As CBIR attempts to extract information from photos or images themselves, several techniques have been used to do so. The most common techniques are color, texture, and shape based on mathematical measures (6).

Color retrieval For color retrieval in Level 1 CBIR, a color histogram is computed for each image. According to (6), "this shows the proportion of pixels of each color within the image." This allows a person to search for a percentage of color in an image or even submit an example image for comparison (6). This is especially useful when searching for images with the same background or an object of a certain color.

Texture retrieval Another technique for retrieval of images in Level 1 is texture retrieval. One significant use of texture retrieval is in the case of areas with similar color. For example, recognizing the difference between green carpet and green grass. In order to accomplish this task, second-order statistics are calculated from the images (6). This consists of calculating the brightness of pairs of pixels relative to each other (6). As a result, several areas of measuring image texture are formulated such as degree of contrast, coarseness, directionality, regularity, periodicity, and randomness. The queries for texture retrieval are similar to color retrieval in that a percentage or example is used to match images.

Shape retrieval One of the most intuitive techniques for CBIR is shape retrieval. Researchers have done several studies to show that natural objects have been recognized by their shape including a study by Irving Biederman (3). For a CBIR system to accomplish this task, several characteristics of the shape of each object in an image are calculated. These object characteristics or features include aspect ratio, circularity and movement invariants, and consecutive boundary segments (26) (19). Shape retrieval queries differ from the previous two techniques in that a user-drawn sketch can also be submitted.

Position retrieval One technique that has been around the longest is retrieval by spacial location. In other words, the image is analyzed based on the position of various data within the image (6).

3.2.2.2 Level 2

A second level search consists of derived features (6). In contrast to the first level, which simply requires a content feature, the program or person requires some prior knowledge such as "more glass" or "more concrete" for the features of a building. A program or person would have to use logic to infer something about the picture (6). The program or person would also need to distinguish between two similar objects such as a truck and a car. Eakins and Graham's examples divide the level into two parts.

- Find pictures of a type of object (a skyscraper)
- Find pictures of a specific object (The Empire State Building)

Semantic feature retrieval Researchers today are still working on narrowing the gap between level one and level two. They have been focusing on two main areas, scene recognition and object recognition. One way to help identify objects is to be able to classify the overall scene of an image. This scene recognition can also be a filter used in searching (6). One system developed by Hermes et al uses color, texture, and spatial information to interpret a scene in an image (10). This system, called IRIS, also generates text descriptors (annotations) for searching (6).

Along with scene recognition, object recognition can also aid the annotation process. A technique developed by Forsyth et al, in 1997, recognized naked human beings within images (8). Since then, the technique has been applied to a broader range of objects (6).

Not all of the techniques in semantic feature retrieval are fully automatic, though. FourEyes, a system from MIT allowed a person to annotate different regions of an image (21). Then, the system looked for similar areas in other images. This semi-automatic process is still a common research area today.

3.2.2.3 Level 3

A level three search is abstract. In other words, a significant amount of knowledge about the photograph has to exist both before and after the search. Eakins and Graham also divide Level 3 into two groups.

- Retreival of named events or types of activity
- Retrieval of pictures with emotional or religious significance

These types of searches deal with the photograph as a whole.

3.2.3 Semantic gap

According to Eakins and Graham (6), "the most significant gap [with CBIR] at present [1999] lies between levels 1 and 2. Many authors refer to levels 2 and 3 together as semantic image retrieval." Thus, the largest semantic gap is between level 1 and level 2. Smeulders et al (24) gives more of a summary description of a semantic gap, "the semantic gap is the lack of coincidence between the information that one can extract from the visual data and the interpretation that the same data [has] for a user in a given situation."

3.2.4 Extracting features with CBIR

Another research area is how a CBIR system could extract more information from the photographs. This information, such as objects in the picture, color, texture, and faces, could be used to further annotate the picture. Then, pictures are easier to search and organize. While intuitive to humans, extracting level 2 features is not as straightforward for a computer. A human has to tell the computer what to look for. In one application, many photos are used to train the program what to look for (15) (See Section 3.5.1 for more details).

3.3 Event-based image retrieval

Two studies on event-based image retrieval describe two different systems. One study from authors Alexander C. Loui and Andreas Savakis propose algorithms for an automated event clustering and quality screening system. The main focus of their study is how to automatically classify events. Section 3.3.1 desribes this system in more detail. The other study from Joo-Hwee Lim, Qi Tian, and Philippe Mulhem describes a system based on home photo content modeling and event-based image retrieval. This system is described in more detail in Section 3.3.2.

3.3.1 Automated Event Clustering and Quality Screening of Consumer Pictures for Digital Albuming

Loui and Savakis stress that automatic event-based classification is difficult because the context of the pictures is limited (18). Even if a tested, capable system is able to group all of the pictures, it still maintains the ability to commit errors. Sometimes groupings are not based on anything other than the interpretation of the person. Hence, this system is not made to replace all organizational methods but is created to help automate the digital albuming process.

3.3.1.1 Process

Loui and Savakis's main approach to this problem is to combine date and time information with picture content information. As discussed in Section 2.2.4, this is because many users naturally group pictures by event. This includes date and time because pictures from the same event are close together chronologically. In addition, pictures from the same event are usually similar in content and style.

The first step in the process of event-clustering is to rank the pictures chronologically. This step is preparation for the first round of clustering. The authors use a date/time algorithm to cluster the pictures based only on dates and times. For example, the algorithm would cluster an event based on date and time when a group of pictures is taken on a Saturday. The algorithm looks for pictures on Friday and Sunday to determine whether the cluster covers multiple days. The authors present another reason for this algorithm in that the time differences are smaller between pictures in one event than between other events (18). In addition, the date/time algorithm is useful when pictures from more than one camera are combined because it will not matter which camera the pictures are from.

The next step in the process is to cluster the pictures based on image content. This step allows the system to differentiate between two different events close in time. The technique the authors use is a block-based histogram correlation, derived in part from the research done by Michael J. Swain and Dana H. Ballard (25). For this technique, the image is divided into blocks and then a color histogram is computed for each block. It represents the dispersal of colors in each block. The color histogram from the images in question are then compared for similarities. Not only does this allow the system to find similar images, but also similar objects in different places or in different images. The final step is to refine the clusters using both algorithms to determine if some clusters should be merged or separated.

Loui and Savakis also describe the screening of low quality images. These are images that are normally discarded or not included in albums. Such images can suffer from underexposure or blurriness. Another type of low quality image is one containing problematic content, such as when a person's eyes are closed or face is turned (18). These types are not included in the Loui and Savakis's system.

In order to discover low quality images, an edge histogram is used to determine the sharpness and contrast of a picture. One problem with this is in images that normally have low sharpness or contrast, but need to be kept, i.e., images of the sky. The authors solve this problem by using a blueness measure. This measure determines how many blue or green pixels exist in the image. Furthermore, the darkness of an image is another factor in the screening process. Hence images are screened based on four factors: sharpness, contrast, blueness, and darkness.

3.3.2 Home Photo Content Modeling for Personalized Event-Based Retrieval

In this study, the concept of event-based image retrieval is similar, however, the main focus is on an event taxonomy. Furthermore, this study gives several other reasons for performing event-based image retrieval.

Citing a study on how users manage digital photos (See Section 3.1) and a study the authors conducted themselves, Loui and Savakis found that users do prefer to organize photos semantically such as organizing by event, people, time, and place. The problem with this kind of organization is that it requires manual annotation. As explained in Section 2.2.7, this is a time-consuming and tedious process. While content-based image retrieval has emerged as one solution, authors Lim et al argue that it is focused on low-level features such as color, textures, and shapes (17). They point out that the semantic gap (See Section 3.2.3) has not been bridged (17). Their suggested bridge is event-based image retrieval.

3.3.2.1 Event taxonomy

The authors propose an event taxonomy to aid in the event-based image retrieval process. While specific to home photos, photos that consumers take for memories, the taxonomy starts at a broad classification of an event and gets more detailed at lower levels. See Figure 3.1 for a picture of their event taxonomy. The authors suggest this can be customized based on an individual. For instance, types of gatherings could be a baby shower, a LAN party, or a study session.

3.3.2.2 Visual keywords indexing

In addition to an event taxonomy, Lim et al also use visual keywords indexing from (16). According to (16) and (17) visual keywords are "intuitive and flexible visual

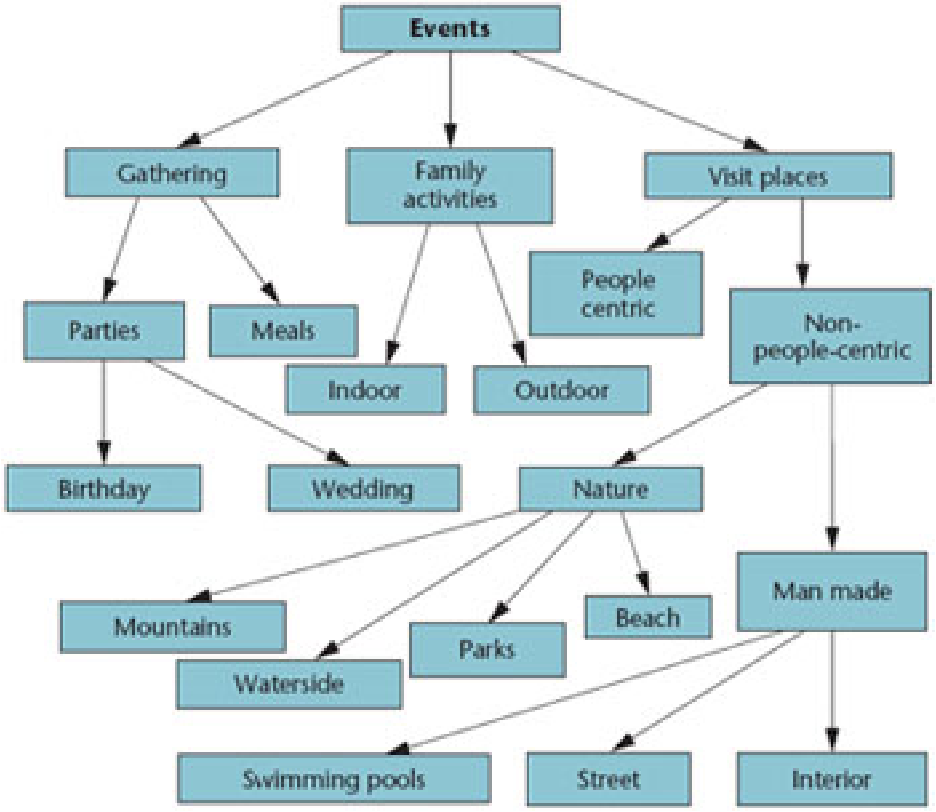

Figure 3.1 The Event Taxonomy from (17).

prototypes extracted or learned from a visual content domain with relevant semantics labels." In other words, visual keywords are several words a person may use to describe an object visually. One example the authors give is "water: pool, pond, river." Hence, if water is recognized in a picture, it could be a picture of a pool, a pond, or a river. The authors use visual keywords in addition to event taxonomies in order to describe regions of a picture. The system learns how to describe a set of pictures. Then, when additional pictures are added, it uses previous pictures to describe the new one.

3.4 Face-based image grouping

As first discussed in Section 2.2.6, the problems involved with face-based image grouping are presented by a study from the Eastman Kodak Company. Identifying faces in family photos presents problems of lighting, pose, image quality, expression, and facial changes due to the passage of time (4). This section will discuss the solution to these problems proposed by authors Madirakshi Das and Alexander C. Loui.

3.4.1 Overview

Face-based image grouping is unique in that not only does it incorporate people in the grouping, it uses events as well. In Das and Loui's approach, the user first groups images by event or people. Then, these groups are broken into further groups by the opposite category. In other words, if the user groups by event, that group is further separated by people and vice versa. This approach is based on the Das and Loui's observation that only a small set of individuals recur in most family photo collections. The problem and solution in face-based image grouping is different from the facial recognition used for security in that, in this case, the application needs to find all faces belonging to a person and not just one close match. The following sections discuss the process of face-based image grouping.

3.4.1.1 Clustering

The system for face-based image grouping first clusters similar faces into groups. Most of the focus is on the larger groups. The authors use the method in (23) to detect the faces. This method uses relative eye locations to find the faces. It is based on distance between features on a face and not necessarily the shape of the face itself. Next, the system groups all pictures twice. The first grouping is by age and gender. For gender, the authors use facial measurements from (7). In order to group by age, the system uses ratios of measurements from (13). The second grouping is face similarity. Since the system does not know how many different people are in the group of photos, the number of clusters is not known beforehand. In addition, as no ultimate database of labeled faces exists, training the system is not an option (4). With the clusters, the system further links clusters together using the algorithm from (5). This happens in order to find multiple clusters of the same person. The authors give the example of finding clusters of a person at different ages.

3.4.1.2 Labeling

After the system finishes creating and linking clusters, it sorts the clusters by the number of members and gives them a unique label (4). Das and Loui propose that the clusters could be given to the user for correction and labeling. This would allow the user to search by name if they were labeled with a specific person. As a result, image annotation for faces would not be as needed.

3.5 Automatic image annotation

Most digital photo management applications allow a person to label photos. As discussed in Section 2.2.8, automatic annotation would expedite this process. A user would essentially be able to upload a batch of photos and an application would automatically rename the photo or add descriptive text based on the contents of the photo. While many studies do not specifically discuss automatic image annotation, many discuss several concepts which would aid in automatic image annotation. Event-based image retrieval and face-based image grouping could be used to aid the process in both annotation and organization. The following study demonstrates the beginnings of automatic image annotation.

3.5.1 Automatic Linguistic Indexing of Pictures (ALIP)

Authors Jia Li and James Z. Wang proposed and implemented an application, ALIP which recognizes objects in a photo and adds annotations automatically. Although currently web-based, the application is one solution to manually annotating a photo.

Initially ALIP is based on manual annotation. In order to train ALIP, both training sets and image categories are manually annotated (15). For instance, if ALIP is going to recognize a 'car' within an image, it needs several images of cars in a training set. Each image is manually annotated, and an image category of 'car' is created. The image category holds descriptions of what an image may look like when a car is present in the image such as street, city, man-made, and house. When a new image is uploaded to ALIP, that image is compared to each image in each image category. Annotations are then selected from the image categories with photos most similar to the uploaded image. For example, if one object in the photo is a car, the image category of vehicles is likely to have similar photos. Then, the application uses the annotations from that image category to describe the image.

3.5.1.1 Current work

Currently, ALIP is web-based because the authors need users to help train the application. Users can either upload photos from their computer or link to photos on another web site. When ALIP is finished analyzing the photo, it gives the user a list of words it found to be relevant to the picture. Many times one or two words can describe the picture. However, the words are more general such as man-made or nature. As a result, ALIP is still learning.

CHAPTER 4. PROBLEM

The studies in Chapter 3 all answer the challenges they propose. However, they all seem to fall short in addressing one or two aspects of digital photo management. Digital photo management is not simply designing one application to perform one task. It involves creating many applications first and, eventually, combining them in one digital photo management application. This chapter addresses this issue. The first section restates the problem with details on what should be automated and how the studies only cover one or two of them. The second section describes why addressing this issue is essential.

4.1 Details of the problem

A digital photo management system should address all aspects of digital photo management. It should include improvements on each step in the process. In addition, it needs to account for different user preferences such as how to organize, search for, or display photos. Many users have different preferences on organization such as by event, object, overall similarity, and person. They also may use various searching methods such as a text-based query, or an image-based query. Users may prefer to display photos differently, as well. This includes photo orientations, cropping, and space for annotating, which are covered in digital albuming. Thus, the problem is not the studies themselves, it is their lack of scope. Some of the studies cover a broader scope than others, but none of them cover everything.

Thus, the following is not a criticism of the solutions within the studies in Chapter 3 or others, instead, a criticism that the solutions have not been combined. The studies each solve their proposed problem, but more collaboration is necessary to attain the ultimate goal. The following sections discuss the remaining room for improvement.

4.1.1 Room for improvement

Each study related to digital photo management thoroughly answers its own question as expected. However, each of them leaves something different out of the picture. Content-based image retrieval addresses searches that are based on an image, but lacks organization by event, or text-based searches. This works well when a person is trying to find the same object, the same scenery, or the same general look in other pictures. The studies on event-based image retrieval address the intuitiveness of organizing photos based on date, time, and similarity using content-based image retrieval, but lacks automatic image annotation. Face-based image grouping addresses how faces can be used to generate photo albums, but lacks searching for other objects or adding annotations. Digital photo management should include concepts from all of these studies. The following sections discuss more details about what could be improved from each study.

4.1.1.1 How people manage their digital photographs

How people manage their digital photographs is one of the few existing studies with a focus on the user. This allowed Rodden and Wood to analyze how people would normally organize their digital and non-digital photo collection. The criticism remains that the application in this study (Shoebox) could be improved with an integration of event-based image retrieval, face-based image grouping, and other studies.

One object of note is the number of participants in this study. With only thirteen participants, the authors had a chance to have more discussions with each participant. However, as with any type of survey, more participants could yield more accurate results.

The authors did realize this and mention it in the study. Also, the study took place in 2000, when digital cameras were only beginning to grow in popularity. A larger number of participants today may not yield different results, but they may be better prepared given the current popularity of digital cameras.

4.1.1.2 Content-based image retrieval

As mentioned in the previous section, content-based image retrieval is based on an image. In order to query photos, a person has to use an example image. However, this is only one way people may want to query a database. A person may not always have an idea of what a general object in various photos looks like, i.e., a car may be green or blue grass may be wide or narrow. In addition, they may not have a similar image ready for searching.

4.1.1.3 Event-based image retrieval

The concept of event-based image retrieval is significantly intuitive. Many people have already organized the pictures manually by event. As discussed in Section 3.3, people usually group pictures based on the date, event, time, or place they were taken. Event-based image retrieval does not cover searching for one object or person. Instead, the user would have to remember all of the events or dates with that object or person in order to find every picture. Again, this is not in the scope of event-based image retrieval studies, but it is, potentially, a common query by a user.

4.1.1.4 Face-based image grouping

Face-based image grouping includes a broader scope than the content-based image retrieval and event-based image retrieval studies. Its main idea is to use faces in pictures to create various albums or borders. In this study, the system differentiates between types such as a baby and adult or a man and a woman, number, and size. It implements parts of event-based image retrieval in solution to its issue and includes digital albuming. However, the study and application leave out other important aspects of existing research such as content-based image retrieval to improve digital photo management.

The fact that faces are also content shows that it does touch on content-based image retrieval. However, the study is focused mainly on recognizing faces and not other content in the photo. In addition, it would not be considered a full digital photo management system because of the focus solely on faces.

4.1.1.5 Automatic image annotation

Studies on automatic image annotation are different from the other studies in that its main focus is describing photos and not retrieving or clustering them. A digital photo management application should not leave out automatic image annotation because it can make the other aspects easier. It allows text-based searching, which may be more intuitive to users. Instead of having to find a photo with a specific object in order to find more photos with that object, the user can merely type a description of the object as a query. One way is not necessarily better than the other, a user may only prefer one way or another. Automatic image annotation alone would not be ideal. Users may prefer other methods of organization and searching such as event-based image retrieval or face-based image grouping.

4.2 Why is this worth doing?

As stated in the introduction, the importance of digital photo management and the union of automation studies may not be clear. The following sections discuss the reasons in more detail.

4.2.1 People are always looking to save time and money

Automating digital photo management can save time and money. Section 2.1 describes the time it took to organize photos before digital cameras. Even though many people have digital photos as described in Section 1.1.2.2, many either do not bother with annotation or spend the same amount of time annotating as in the past. Granted, with automation, a person would need to prepare the software, especially, if most of the annotations were family and friends. Software that recognizes new people automatically is not currently feasible. If the software was prepared, the only time consumed in the digital photo management process would be in uploading pictures and deciding what album to display.

An individual may not have much cost involved in organizing photos. The main cost with organizing personal photo collection is time. However, many professions spend a great deal of money making sure photos are organized. For instance, the advertising and publishing profession searches for various images everyday. In order to make them searchable, someone had to manually annotate all the photos. If the process were automated, no cost would be associated with annotating (with the exception of the cost of the software).

4.2.2 Images have become an essential part of many professions

Several different professions use photographs and images on a daily basis. Content-based image retrieval would help several of these professions in saving both time and money. In contrast to home photography, professional photography has a broader range of photos, a larger number of photos, and the potential for more abstract searches. The next several sections address different professions and the benefits of CBIR.

4.2.2.1 Crime prevention

Law enforcement agencies gather an enormous amount of information at a crime scene. Much of this information is in the form of photos. For instance, many photographs are taken at the scene of a crime. Police need to record what the scene looked like at the time of the crime, any witnesses, and where each piece of evidence was found. In addition, if anyone is arrested as the result of the crime, that person's photograph, fingerprints, and even shoe prints are taken. All the photographs are then both linked together themselves and to any paperwork filed. As a result, content-based image retrieval could help law enforcement in several areas.

For example, CBIR could help police piece together a crime such as finding the similarities between two different crime scenes. It could recognize similar faces, evidence, or even overall layout. Event-based image retrieval may be able to aid in piecing together evidence from multiple crimes. Security of the evidence is also a factor, since access to evidence needs to be restricted. Security features in digital photo management can regulate and log access to photographic evidence. This can also be applied to restrict access in other professions, such as in the medical profession, and for home users. See Section 5.3 for more details on how digital photo management could help with security.

4.2.2.2 Medicine

The medical profession also uses images in their daily routine (6). They record x-rays, ultrasound, etc., for both diagnosis and monitoring. CBIR can be used to help these two areas by initiating some of the detection of problematic areas (such as tumors or fractures). The system could also be used to analyze images to find out better diagnosis techniques. Similar to the law enforcement profession, images in the medical profession need to be secure. Laws and regulations prevent medical documentation, including images, from being accessed without permission from the patient. Section 5.3 discusses more details on how security features, such as access restriction, can be applied to digital photo management.

4.2.2.3 Publishing and advertising

In the publishing and advertising industry, photographs are used as illustrations. CBIR could be used to aid a person in searching for specific content in photographs. In addition, automatic annotation of photographs will save time by allowing text-based searches for finding appropriate illustrations for text segments. Manual annotation often hinders the searches when multiple words for the same object or scene are used interchangeably. Many newspaper, magazine, and book publishers have large databases of images for their use.

4.2.2.4 Historical research

Visual information is vital in historical research, as well. To have an image of a historical event is precious to historians in art, sociology, medicine, and other disciplines. Manual annotation also presents a unique conflict in historical research. In areas where certain concepts or parts of history are disputed, images could be annotated with a different bias. As a result, automatic image annotation could help with consistency. In addition, the number of images throughout history is large. The process of manually annotating such a variety of images takes much time and money.

CHAPTER 5. A SOLUTION

5.1 How can the automation studies and digital photo management work together?

One automation study involved automatic image annotation. It is imaginable that were image annotation intuitive, manual annotation would be obsolete. Image annotation can help for a few reasons. One, it will help the person. If the computer organizes without describing the classification, the user will be confused. Two, if a picture is automatically annotated correctly, it saves the user time. However, if the user still has to annotate pictures manually, it does not save time compared to the time it takes to write on film photos. Automatic image annotation is only part of the ideal solution to the challenge of digital photo management. This chapter describes an ideal solution, the union of the studies, and the application of security features in digital photo management.

5.1.1 The ideal solution

Many researchers have proposed various ways to create automatic image annotation, content-based image retrieval, and other automation methods in digital photo management. Ideally, many of those procedures could be incorporated into one piece of software. In the following proposition, the assumption is perfection in the application. While perfection may never be the case, the following solution attempts to approach it.

5.1.1.1 Automatic image annotation

First, the users of digital photo management software should not ultimately be limited to the home user, but here the focus is on the home user. For security concerns, this software will need to be stand-alone since web-based would incur more of a security risk. Stand-alone software is more versatile for the user. However, it could be broadened, later, to have a web-based component. (See Section 5.1.1.4.)

Next, when a user uploads pictures through the program, each picture is analyzed for both annotations and image quality. The technology for automatically annotating images is currently available, but has not been sufficiently perfected. Authors Jia Li and James Z. Wang have recently created a web-based version of their application called ALIPR. In this version, the software is still learning from training sets and users since anyone can use it online. However, the developers mention "it is still a child" (15). In addition, ALIPR is not currently designed to learn and recognize specific named objects such as specific cars. It is, instead, concerned with concepts such as landscapes or general objects. If a user searches for a specific object and it returns no pictures, the software asks for a picture to upload. This could result in pictures for named objects but in a round about way.

In the ideal case, a user could upload several pictures to the computer and the software would recognize specific people in the picture (friends and family), various objects, the location, and possibly a feeling about or from the picture. Some of the recognition will always have to be learned, especially in the case of annotating specific people. Then, even if the user does not do anything more in preparing the photos, they should still be ready for searching by the software.

In the end, the user could allow the integrated software to do much more than annotate, assuming the annotations were mostly correct. With the annotations completed, sorting and organizing would be much easier. For example, the software could automatically sort uploaded photos into various folders similar to sorting them into physical photo albums. The technologies for the sorting component of the software are already available, but they rely on manual annotations from the user.

5.1.1.2 Image retrieval

This paper has discussed several ways of image retrieval including content-based image retrieval, event-based image retrieval, and face-based image grouping. All of these studies deal with automating different processes of organization in digital photo management. Event-based image retrieval significantly aids in the organization process. To most users, organizing by event is intuitive. Most non-digital albums are in some type of chronological order. To a person, having photos in an album lacking chronological order (possibly an album about a single person) seems to be unorganized and difficult to consume. Granted, some albums that seem unorganized may have order to them for the creator, but it would not be obvious to a stranger viewing the album. Thus, event-based image retrieval is essential to digital photo management.

Part of event-based image retrieval relies on content-based image retrieval. Photos from the same event may also be similar in content due to location. As a result, content-based image retrieval aids event-based image retrieval to make query results more accurate.

Content-based image retrieval is a part of the foundation for automatic image annotation. In this case, objects in one image are used to locate similar objects in another. When an application can learn to identify objects, it is able to automatically annotate new images. Content-based image retrieval can also recognize other similarities such as a general style of pictures, color, and layout.

Face-based image grouping and retrieval is also essential for automated digital photo management. Many people already annotate non-digital photos by labeling people in the photo. Face-based image grouping would allow the application to learn faces of specific people to aid in annotation of those people in digital photos.

Many times, users are only concerned about the people in a photo. If an application could automatically recognize the faces in a photo, it would save the users a significant amount of time. Face-based image grouping could also aid in finding photos with one specific person for projects surrounding various special events or purposes.

5.1.1.3 Customization

As discussed in Section 3.1, people may have different preferences as to the methods of digital photo management. Hence, customization would be welcomed in digital photo management. In many applications this is already emerging. However, in this solution, the user would be able to use all of the ways of automation. This flexibility would greatly improve the accuracy and perceived worth of a digital photo management application. Furthermore, manual annotation tools would also be included in order to cover a broader range of preferences and for corrections of less adequate annotations.

5.1.1.4 Web-based versus stand-alone applications

Web-based applications and stand-alone applications have many benefits, but a stand-alone application would be better suited for digital photo management. Many times, security is an issue with web-based applications. Currently, Automatic Linguistic Indexing of Pictures Real-time (ALIPR) is best suited as web-based in order for many people to provide input for it learn how to annotate photos. It would be much harder for only a few people to upload photos to ALIPR and make sure it annotated them correctly. However, this is mainly for ALIPR to learn and not act as a fully functional web-based image annotation application. The researchers may have implemented ALIPR in order to aid in training for a stand-alone application such as ALIP.

A potential downfall to the web-based application is that people are cautious about (or, perhaps, should be) uploading personal photographs to the web solely for annotation purposes. That is not to say ALIPR should not be trusted, only that many people may not want to upload as many pictures to it since it is web-based. In other words, privacy is a concern.

5.2 How to combine the systems

The previous section describes how the combination of systems is essential in digital photo management. This section will describe how to combine the image retrieval and annotation systems into one application. Each system will be evaluated on its input and output, priority within the combined application, interactions, and how it should be tested. Concepts and results from the study in Section 3.1 will be beneficial in designing the user interface for the main application but not for image retrieval or annotation, since its focus is on user preferences.

5.2.1 Input and output

Content-based image retrieval is flexible with its input. A user has a choice of three different types of input.

1. One whole image

2. A user-defined region of an image

3. An image drawn by the user

The output is a group of images similar to the input from the database of images. The CBIR system needs to be trained on a decent number of images, and the images need to exist in a database or other type of group, i.e., a folder system.

In event-based image retrieval and clustering the user inputs a group of images to the system. For the event taxonomy, the user needs to annotate a few images initially.

The system uses the annotations to build models for the events. Groups and subgroups of events are the output for this system.

The user input for face-based image grouping is initially a group of images for training purposes. Once trained, the system allows the user to query using an image or a name (if clusters of similar faces have been annotated). The output of this system is a group of images. However, the user also has the option of specifying a layout for the images, in which case the output will be in the form of that layout.

For automatic image annotation, the input is either a single image or a group of images, assuming the system has been trained initially. The output is a list of words describing both the image in general and objects within the image.

5.2.2 Priorities

In combining these systems into one application, the user needs to be able to choose how to sort photos. Each system should be able to perform on its own. However, if the user merely wants the application to perform a general organization, all of the systems need to run. As a result, the following discusses the priority for each system.

Content-based image retrieval needs to run first because the other systems can benefit from the results. It will be able to group similar images for use in event-based image retrieval, recognize the presence of faces for face-based image grouping, and recognize objects for automatic image annotation.

The automatic image annotation system is next since it benefits the other systems as well. If many of the photos are annotated before they are sent on, they will be easier to organize. The annotations may give clues to the type of event or the names of people in the photo.

Event-based image retrieval and face-based image grouping are interchangeable in priority. It is higher in priority in this list because face-based image grouping uses event-based image retrieval in its system. In addition, this system also benefits from content-based image retrieval through content similarity.

Thus, face-based image grouping is fourth because many images include people, especially images from home users. It is able to use the previous three systems to aid in the grouping process. In addition, the images without faces are already weeded out from the analysis from content-based image retrieval.

Automatic image annotation should be performed again in order to annotate groups from event-based image retrieval and face-based image grouping. Ideally, this will be the default configuration. However, users will want the ability to choose preferences on which system to use and its priority.

5.2.3 Interaction

The interaction between systems in this combined application will be focused on speed, performance, and memory consumption. One way to speed the interaction is to allow referencing in dealing with images. In other words, instead of making copies of images to send to each system during the organization process, references to images should be sent. If the systems send images between each other, it would use up CPU time and memory, especially since new digital cameras allow for increasing quality (and therefore size) of images. With each system, multiple copies of images are required since one image may belong to multiple groups. Referencing will allow the application to use CPU and memory resources for analyzing images instead of manipulation between systems.

5.2.4 Testing

In order to test this system, one needs a variety of photos. In half of the systems, photos from commercial sources such as the COREL and Kodak corporations were used for testing. Photos from personal collections were used in the other half. In the testing of this application, however, photos need to be used from both places. Both users and the application should organize the same photos. The accuracy of the test is based on the similarity of the groups and annotations from both the application and the users. Ideally, a significant number of users should organize their personal photos along with using the application to organize them.

5.3 Security

An improvement to digital photo management applications is to include security. All of the studies in this paper have left out discussions on security measures. For example, security measures need to be taken with automatic annotation, especially in professions such as law enforcement. A home user may not need as many security features as a profession, but that does not mean that the security features should not be addressed. Many of the studies do not include security as a topic of their research.

The security desired by a home user may be described more accurately simply as privacy. It is unlikely that any photos would be of value so-much-as to have only the photos themselves taken from the computer. The computer would most likely be stolen as a whole and other information on the computer could be more valuable (bank records, personal notes, or other information). The security measures in the digital photo management software may not initially be built to keep out law enforcement, but it is a possibility. Then, the question becomes whether the security would only prevent viewing the annotations or viewing the photos altogether. The digital photo management software could be password protected, but the question remains, what exactly would be protected by a password? Even if the software was secure by high standards, the actual photos may not be secure.

Many digital photo management systems include ways to share, save, and print photos. Sharing is an ideal area for security measures, especially when photos are sent over the Internet. Personal information could be sent inadvertently in a photo. For instance, if a person takes a picture of a brand new Mustang, protecting that photo may not be needed. However, if the picture includes a good view of the license plate number, it could be used to find more information about the person. Granted, there are laws and regulations in place such as the Drivers Privacy Protection Act to make the malicious use of license plate numbers illegal for a person, but it is a concern. Another example is taking a picture of a house. If the right situation occurs, the person's address could be derived from various objects in the picture (House number, street sign, city vehicle in the driveway). All of these components to a photo may not be common, but a person could have reason to be concerned.

Professional photographers, artists, or historians may use digital photo management software for freelance work. Professionals in these areas need security for photos because many photos in these areas are copyrighted. Photographers for newspapers and magazines are depend on compensation for the pictures that are printed. If another person stole a photo and published it in a newspaper, the photographer would not receive the full benefit of his or her work. Similarly, many professional photographers do not want their photos to be mas-produced. One such case may be for family photos or high school senior photos. Not only do the professional photographers want to protect the photos when they are in transit, but they also want to protect the actual photo when the intended recipients have it.

Law enforcement is another profession that has a need for digital photo management with security. Photos are taken at crime scenes for evidence and clues. With each crime, though, the photos included in evidence have to have regulated access and authentication. When evidence passes from one investigator or officer to another, the access has to be logged in order to stand up in court. Even without annotation, photos need to be secure in the law enforcement profession. Annotated photos may even present a higher liability since more information would be available.

CHAPTER 6. CONCLUSION

This paper described how digital photo management could benefit from automation studies. The first chapter gave an overview of how this thesis was organized and some general concepts. Chapter 2 provided more details on the background of this paper. Chapter 3 described five studies for digital photo management in detail. Chapter 4 gave more details on the proposed problem. Finally, in Chapter 5, an ideal solution was proposed. The sections, in this chapter, give specific conclusions from this thesis and ideas for future research.

6.1 Conclusions drawn

The main conclusion of this thesis is that each automation study in combination would benefit a digital photo management system. Researchers may reference (22) as to human behavior in the case of digital photo management techniques in order to create applications for a variety of users. This basis of research can guide developers to a more focused application from the beginning.

Components of the application of content-based image retrieval serve as the basis for image retrieval in general. It aids in recognizing objects in a photo and determining photo similarity. Automatic image annotation directly benefits from content-based image retrieval based the comparisons of objects and images. Event-based image retrieval and face-based image grouping both use concepts and technology from the twenty-something year history of research in content-based image retrieval. Event-based image retrieval provides an intuitive way to manage digital photos. Face-based image grouping aids the user in another intuitive process: annotating people in images. It allows a user to recognize specific faces in photos. Finally, automatic image annotation applications can aid organization and searching by providing the more immediate ability to do text-based searches.

6.2 Contributions

The intended contribution to the field of digital photo management, as set forth here, is in proposing the union of automation studies in one application. Many studies have already implemented multiple studies of automation such as face-based image grouping. However, few have proposed this union of all of the automation studies. Per the reasoning in previous pages, it would seem a software advancement is necessary.

6.3 Future research

It seems implausible that research in automating digital photo management would not slow down. It includes many possible areas for future research. Future studies on how one application could incorporate all of these areas of automation would significantly benefit digital photo management. This, of course, would involve application logistics such as memory, compatibility, resources, and studies, such as face-based image grouping with multiple automation areas, would be very useful in accomplishing this goal.

Each specific area of automation and its algorithms would also require more research to improve efficiency and effectiveness. For example, automatic image annotation (ALIP and ALIPR in particular) currently involves a long training period. Research on ways to shorten the learning process would be beneficial. This efficiency would simplify the union under one application.

As this paper has discussed, with digital cameras increasing in popularity, a person's collection of digital photos will increase. As a result, consumers and professionals alike will require more automation in digital photo management to save time and money.

BIBLIOGRAPHY

[1] Face recognition. http://www.epic.org/privacy/facerecognition. Accessed March 28, 2007.

[2] Itfacts: Digital imaging. http://www.itfacts.biz/index.php?id=C0_35_1. Accessed March 14, 2007.

[3] Irving Biederman. Recognition-by-components: A theory of human image understanding. *Psychological Review*, 94(2):115–147, 1987.

[4] Madirakshi Das and Alexander C. Loui. Automatic face-based image grouping for albuming. *IEEE Conference on Systems, Man and Cybernetics*, 4(5-8):3726–3731, 2003.

[5] Richard O. Duda, Peter E. Hart, and David G. Stork. *Pattern Classification (2nd Edition)*. Wiley-Interscience, 2000.

[6] John P. Eakins and Margaret E. Graham. Content-based image retrieval. Technical Report JTAP–039, JISC Technology Application Program, Newcastle upon Tyne, 2000.

[7] Leslie G. Farkas. *Anthropometry of the Head and Face*. Number 2. Raven Press, 1994.

[8] David A. Forsyth, Jitendra Malik, Margaret M. Fleck, Hayit Greenspan, Thomas K. Leung, Serge Belongie, Chad Carson, and Chris Bregler. Finding pictures of objects

in large collections of images. In *ECCV '96: Proceedings of the International Workshop on Object Representation in Computer Vision II*, pages 335–360, London, UK, 1996. Springer-Verlag.

[9] Mohamed Hammami, Youssef Chahir, and Liming Chen. Webguard: A web filtering engine combining textual, structural, and visual content-based analysis. *IEEE Transactions on Knowledge and Data Engineering*, 18(2):272–284, 2006.

[10] Thorsten Hermes, Christoph Klauck, Jutta Kreys, and J. Zhang. Image retrieval for information systems. In *Storage and Retrieval for Image and Video Databases (SPIE)*, pages 394–405, 1995.

[11] Toshikazu Kato. Database architecture for content–based image retrieval. In *Image Storage and Retrieval Systems*, San Jose, California, 1992.

[12] Allan Kuchinsky, Celine Pering, Michael L. Creech, Dennis Freeze, Bill Serra, and Jacek Gwizdka. Fotofile: a consumer multimedia organization and retrieval system. In *CHI '99: Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 496–503, New York, NY, USA, 1999. ACM Press.

[13] Young H. Kwon and Niels da Vitoria Lobo. Age classification from facial images. *Comput. Vis. Image Underst.*, 74(1):1–21, 1999.

[14] Mark Lansdale and Ernest Edmonds. Using memory for events in the design of personal filing systems. *Int. J. Man-Mach. Stud.*, 36(1):97–126, 1992.

[15] Jia Li and James Z. Wang. Automatic linguistic indexing of pictures by a statistical modeling approach. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(9):1075–1088, 2003.

[16] Joo-Hwee Lim. Building visual vocabulary for image indexation and query formulation. *Pattern Analysis and Applications*, 4(2-3):125–139, June 2001.

[17] Joo-Hwee Lim, Qi Tian, and Philippe Mulhem. Home photo content modeling for personalized event-based retrieval. *IEEE MultiMedia*, 10(4):28–37, 2003.

[18] Alexander C. Loui and Andreas E. Savakis. Automated event clustering and quality screening of consumer pictures for digital albuming. *IEEE Transactions on Multimedia*, 5(3):390–402, 2003.

[19] Rajiv Mehrotra and James E. Gary. Similar-shape retrieval in shape data management. *Computer*, 28(9):57–62, 1995.

[20] T. Mills, D. Pye, D. Sinclair, and K. Wood". Shoebox: A digital photo management system, 2000.

[21] Thomas Minka. An image database browser that learns from user interaction. Master's thesis, Cambridge, MA, 1996.

[22] Kerry Rodden and Kenneth R. Wood. How do people manage their digital photographs? In *CHI '03: Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 409–416, New York, NY, USA, 2003. ACM Press.

[23] Henry Schneiderman and Takeo Kanade. A statistical method for 3d object detection applied to faces and cars. *cvpr*, 01:1746, 2000.

[24] Arnold W.M. Smeulders, Marcel Worring, Simone Santini, Amarnath Gupta, and Ramesh Jain. Content-based image retrieval at the end of the early years. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(12):1349–1380, 2000.

[25] Michael J. Swain and Dana H. Ballard. Color indexing. *Int. J. Comput. Vision*, 7(1):11–32, November 1991.

[26] et al Wayne Niblack. The qbic project: Querying images by content using color, texture and shape. In *Storage and Retrieval for Image and Video Databases (SPIE)*, volume 1908, 1993.