Ideas and Considerations for Digital Photograph Sharing

David Merrill
MIT Media Lab
20 Ames Street, E15-320B
Cambridge, MA 02139 USA
+1 617 253 5027
dmerrill@media.mit.edu

ABSTRACT
In this paper, I discuss the digital photo search and annotation problem, with ideas for solutions that incorporate natural language, direct manipulation, and leveraging a digital representation of the user's social network. The social and technological implications of these ideas are also discussed.

Categories and Subject Descriptors
C.5.3 [Computer System Implementation]: Microcomputers – *portable devices.* H.3.1 [Information Storage and Retrieval]: Information Search and Retrieval – *clustering, information filtering, query filtering, relevance feedback, retrieval models, search process, selection process*. H.5.2 [Information Interfaces and Presentation]: User Interfaces – *graphical user interfaces (GUI), input devices and strategies, natural language.*

1. INTRODUCTION
The sharing of photographs is easier now than it has ever been with the advent of digital image capture. However, there are still many ways that technology can help us capture, catalogue, and share our digital images that reach far beyond the manual way in which we accomplish these tasks today. I will examine possibilities for new ways to acquire digital images, both individually and from others, possible ways to annotate and search existing photo collections, as well as improved ways to select and find images that we want to share with people in our social network. The creation and maintenance of a digital representation of a user's social network and personal profile lies at the center of many of these ideas, and the social and privacy implications are also considered.

2. ACQUISITION: GETTING THE SHOT
The traditional way to capture photographs – pointing and shooting a camera - hasn't changed much with the advent of digital. At present, digital photos in the typical person's collection are most likely to be taken by the person, or taken by their friends and family and shared by email, with any number of photo-sharing websites ([4] [11] [22] [23]) or by writeable media (CD-R, DVD-R). Before looking at ways to improve the status quo for photo sharing, I will briefly examine two other paradigms for directly capturing photos that may become popular in time.

Ubicomp'05, September, 2005, Tokyo, Japan.

2.1 Automatically capturing everything
Some research projects have removed the requirement of intentionality in picture-taking, such as the SenseCam project [1] and Brian Clarkson's work [2] which utilize wearable devices to capture images or video continuously throughout the user's day for later processing and/or retrieval. These platforms collect a comprehensive visual trace of the user's day, but they are oriented more towards documentation for research purposes than towards capturing well-composed pictures that the user might want to frame, share, or otherwise utilize in socially meaningful ways.

2.2 Grabbing the shot from a friend
With the right sensors and wireless communication, tomorrow's cameras will determine when their owner has just been captured in another person's photo. The owner's camera will then negotiate with the other camera to retrieve either a copy of the photo or a reference to it for later online retrieval if local storage space is short. This negotiation between devices could also help the other camera properly annotate the photo (see "Annotation" below for more on this), adding the first camera's owner to the list of people who appear in the shot.

Even without the technology to detect when a given camera is in another camera's photo, camera owners will immediately transfer photos between devices wirelessly. Early implementations of this feature exist today in camera-phones featuring Bluetooth, email, or SMS messaging capabilities, and in convergence devices like the Palm Treo [19] that have an integrated camera as well as infrared communication capabilities. Additionally, a few companies have recently made press announcements about standalone digital cameras with WiFi or Bluetooth built-in.

Despite the growing number of ways that cameras are going wireless, the following important social question is still wide open: How will these devices best facilitate immediate sharing of the desired photos with the desired people. Short-range directional communication like infrared (IRDA) is a fairly convenient way to initiate a photo transfer on-the-spot, but it still consumes the user's focused attention at the time of share for a few seconds to select the photo and initiate the transfer. This can be made easier: For instance, perhaps during the 20 seconds after I take a shot, my camera should send an outgoing invitation to share my new photo over IR to any other cameras that are within short-range line-of-sight. If a friend wants a copy, they just have to put their camera in front of mine and confirm the transfer. An ad-hoc wireless radio network among nearby cameras would remove the line-of-sight requirement, but would make it necessary to verify my social connection to any recipients, in order to avoid sharing my new photo with nearby strangers. A further twist this exchange would be to initiate the photo transfer (or confirm the social connection between people) by touching the two cameras to

each other physically. This physical touch would ensure the intention of both parties to transfer the photo.

3. SHARING: A PRIMARY REASON TO CARE ABOUT DIGITAL PHOTOGRAPHS

Sharing photos is a fundamentally social activity that strengthens ties between individuals by facilitating the review and communication of past experiences with others [7]. Additionally, mobile and flexible photograph-sharing tools can stimulate the creation of playful stories, the expression of affection, and the creation of art [15]. A recent study about behaviors associated with the sharing of digital webcam photos in an instant messaging context corroborates the social importance of photograph sharing, finding that digital pictures were used for a variety of social functions including the amplification of communicative intent, to create a sense of remote presence, and to illustrate narratives [18]. A host of so-called "photoware" applications for the desktop and the web have emerged in the years since consumer photography went digital, and these applications – informed by findings about how photographs are used in the context of storytelling and reflection – have been increasingly oriented to facilitate the sociable sharing of photos [4].

4. THE SEARCH PROBLEM

Frohlich et al. found that a primary activity that photograph users want to engage in is the creation of "reprint mini-albums" – collections of the user's photographs that are selected for a particular person and include personalized messages about the photos [7]. The less inherent organization and search-ability a user's photo collection has though, the more time is required to create these personalized collections. The problem is that organizing photos into albums is a low-priority task for many users, and is likely to be left undone [20]. The typical user's lack of upfront effort to organize their photographs reveals the need for a good search strategy to solve the following key problem: How can a system best help a user to easily search for and select the most appropriate photos to share with another person, while requiring minimal or at least particularly enjoyable organizational activities of the user.

4.1 Search today

Searching for digital photos to share with others is still an incredibly primitive process for most users today. With default file names like "DSC0327.JPG", and file headers stuffed with aperture and exposure settings there just isn't much socially relevant information connected to the raw photo itself to search on. Many users group their photos in a rough chronological ordering by filing each "batch" of photos downloaded from their digital camera into a folder together, or into a few folders related to the events [20]. Some users give these folders descriptive names and may organize folders hierarchically [21], and the rare user may even rename each photo in a descriptive way, but the amount of time required for naming each photo is prohibitive for most. Frohlich wrote the following generalization about his user group [7]:

> "The filtering and arrangement of 'favorite' prints into albums was seen as the best way of archiving conventional photos for future sharing. However, this activity was seen as complex and time consuming. In addition, it appeared to be an isolated task without any immediate emotional payoff…"

A few commercial and research platforms have begun to develop more sophisticated ways of searching one's collection of digital photos [4] [17]. These platforms attempt to provide improved ways for users to annotate their photos in searchable ways. These groups are on the right track: fundamentally, solving the search problem will require innovative ways to produce useful annotations for digital photos, since meaningful search requires meaningful annotation.

4.2 Annotation

Thorough annotation is the most direct way to achieve robust image search and retrieval. If a collection of images is annotated well, it becomes easy to find desired images with keyword searches or database-style queries. Old wisdom tells us that "a picture is worth 1000 words" and there are lots of good ways to leverage semantically relevant annotations to search for relevant photos [10], so the question becomes: How can we get those words that correspond to the picture in a way that is agreeable to the user?

Manual textual annotation of digital photos is prohibitively time-consuming, and is not considered a practical solution for most consumer applications. The Aria system [13] attempts to glean annotations by observing the text that the user types as they attach images to emails. Even with features to augment this type of scheme – for example, by using time-of-capture information to chunk photos into "sessions" [3] and by probabilistically associating text associated with a single photo with the other photos from the same session – photos are being used and shared in many ways today (websites, writeable media, SMS messages between mobile phones) that bypass email attachments completely.

One solution could be to build more direct manipulation into an interface for annotation and categorization. Historically, sorting photos was done with physical photographs, and physical photos have some convenient affordances that digital photos lack: Physical photographs can be quickly flipped through serially, spread out spatially and sorted into piles, or placed in boxes for on-the-spot classification. All of these activities take advantage of the fact that we have two hands with multiple fingers that can quickly grasp and manipulate physical objects in parallel. It would be interesting to re-introduce a greater degree of direct manipulation into the digital photo sorting/annotation process to speed things up. For instance, the user could "click-and-dip" a photo in and out of any number of category "bins" in order to associate semantic concepts with the photo. Photos could be classified by spatially grouping them in a way that keeps them clustered into groups, but simultaneously visible (at least partially) on-screen, similar to shuffling physical photos into piles on a floor or table. Borrowing the affordance of piles would mitigate the current problem of digital photos disappearing from sight when dragged into a folder. Demonstrating the advantage of mouse-based direct manipulation over keyboard annotation, Schneiderman and Kang's PhotoFinder interface allowed the user to drag-n-drop labels directly onto photos in order to annotate photos with the names and locations of people therein [21]. The mouse only gives a single point of control though, so a multi-finger touchscreen or multiple touchpads could allow a direct-manipulation style pile-sorting or label-dispensing process to mimic the affordances of physical photographs even more closely.

Another solution is to build the annotation process into an activity that is more ubiquitous – the in-person showing and talking about photos with another person. The verbal description that a person gives when showing photographs to a friend or relative who was not present at the depicted event [7] is an important potential channel of annotation information. Important annotation information such as *who* is in the picture, *what* the activity was, *where* it was taken, and other socially relevant data are often discussed verbally during an in-person or telephone-based sharing session. An annotation interface that utilized continuous speech recognition [17] might feel less cumbersome than one featuring only textual annotation, but the expectation of complete speech-to-text annotation may still be optimistic. Basic speech recognition could at the very least create a probabilistic representation in semantic space of a photo or batch of photos using word-spotting techniques, even with recognition errors. A more sophisticated scheme might leverage contextual information from a range of camera sensors, or produce a 'bootstrapped" context from previous word recognitions in order to dynamically tune probabilities for ongoing word recognitions. A system like WordNet [16] or ConceptNet [14] could extend this bootstrapping process beyond actual words spotted to include their surrounding neighborhood of words or concepts. Of course, the speech recognizer for a personal mobile device could be trained on the user's speech to improve upon a default user-independent recognizer. This one-time configuration is still much less cumbersome than ongoing manual textual annotation.

A keyword search over a collection of photos annotated probabilistically with speech recognition could use the actual search terms, as well as collection of other terms from their immediate word-net or concept-net neighborhoods. Categories assigned by any method (such as the direct-manipulation tagging described above) could also be used to screen search results. In addition, good natural language parsing of a user's query could allow much more natural searching based on time-of-capture or other contextual features than we have today for queries like "I'm looking for a photo from last year or the year before, taken on a Sunday morning". Enabling a search interface to produce correct results on natural language queries would be a matter of careful grammar design based on observation of a large corpus of language-specific user data. Eventually, the best interfaces for searching a collection of digital photos may involve several, tens, or even hundreds of separate heuristics, each heuristic contributing to overall search quality much like popular search engines for the web today [9]. Much like a web search, the result of a multi-heuristic photo search could potentially be a very large set of a user's photos, but the search results would be ranked in order of probable relevance to the query. In addition, the most relevant search results could be a starting point for more traditional, less probabilistic forms of exploration like looking at photos that are chronologically near to a selected photo. A photo management application with a modular plug-in framework or a scripting language for search heuristics could become very powerful by leveraging the sensibilities of an entire community of digital photo enthusiasts who could contribute individual search modules under an open-source license.

4.3 Search using the social network

Sharing photos is a fundamentally social activity, and there are many patterns governing how a person decides which photos to share with whom and how to narrate a particular photograph. Some factors involved in photo selection for sharing include the relation of the other person to the sharer (i.e. is the other person a friend, mother, sibling, co-worker, boss, ex-girlfriend, etc.?), the degree of closeness between the two people, the existence of other relations in common between the two, the shared experiences between the two people, and the interests of the other person. Much of this information would be contained in a fully-defined representation of a person's social network, and a set of rules operating on the social network data could provide useful suggestions about what to share with whom.

For instance, photos of my sister's recent visit to Boston ought to be shared with my mother and brother. Photos of me and my good friend Sean on a bicycle ride ought to be shared with my friend Justin, since Sean is a mutual friend. Photos of my ex-girlfriend ought not to be shared with my current girlfriend, and I probably only need to view a subset of her recent "girls weekend in the city" collection. These examples are a sample of the kinds of rules that could operate on a digital representation of a person's social network. A basic set of "default" sharing rules could be distributed with a social-network based photo search system, and a simple language for modifying or adding to these rules would be an important feature. Equipped with a digital representation of a user's social network and annotated photos, a system could suggest a ranked list of possible photos to share with another member of the network. The ordering of this list could be informed both by the social network rules, and any other heuristics as mentioned in the previous section. Presented with the system's suggestions for sharing, the user could then confirm the sharing of particular photos with people directly from the system's suggestions, or by exploring related images and people. The online trading of social-network photo sharing rules could become popular, as users would probably develop rules that are more perceptive or specifically targeted than the ones that the designers of such a system would invent.

The contents of photos would also be informative to use in concert with the social network. As a simple example, a user could directly tag certain photos as "risqué" by the drag-n-drop or "click-and-dip" direct-manipulation methods mentioned earlier. Or perhaps the phrase "wild party" was recognized in the verbal description with a word-spotting interface. In either case, a social-network sharing rule could suggest showing these photos to close friends, but not to parents or co-workers. A machine-learning algorithm could learn the nuances about what implications certain words or classes of words have to an individual user. Even more interesting would be if the system could use the contents of a picture in concert with the interest profile of people in my social network in order to suggest sharing. For instance, if I have a picture with enough annotation to identify it as a Reykjavik shipyard, the system could suggest sharing the photo with my grandfather, since he was in the United States Navy many years ago and is interested in nautical topics. All of these suggestions by the system could be presented to the user in a ranked list, and whenever social network rules of sharing were utilized in suggesting a photo, the particular rules utilized could be identified alongside the search results, giving the search a degree of reflective capability. In this way, the connectivity of my social

network, as well as the ability to access the personal profiles of its members, could allow a rule-based or probabilistic system (or a combination of the two) to profitably suggest photos to share with other people in my social network.

Recently, the merging of social-network functionality and photograph sharing has appeared in several commercial and research applications ([4], [24], [25]). Flickr has received the most press for the combination of social network and photo sharing, but Flickr's social network – built originally so that users could restrict access to their photos – is used primarily today by users in order to share their photos with a wide audience of interested strangers [8]. In contrast, the Yahoo 360 service incorporates a person's social network and their friend's blogs in order to help users connect with and communicate with the people that they already know, rather than try to meet new people [12]. Microsoft's "Wallop" project [24] seems to offer similar features to Yahoo 360 [25]. The creation of a system that uses a community-contributed set of rules operating over the connectivity and individual profiles in a person's social network in order to actively suggest sharing activity will be an exciting step forward for photo sharing.

5. SOCIAL IMPLICATIONS

The aforementioned ideas on capture, annotation, search and sharing of digital images could have implications on personal privacy and people's willingness to make their social network and personal profile explicit and accessible.

5.1 More explicit definition of social network

In order to take advantage of photo sharing services based on a user's social network, it will be to the user's advantage to maintain an accurate and up-to-date digital representation of their social network. An accurate social network would feature a representation of the category and degree of relationship between the user and others in their network. For instance, one person might be classified as a "good friend" while another person is just a "business acquaintance" or anywhere else in the space of relationships. People might be uncomfortable making the relationships in their social network explicit, or they might fear social repercussion should the information be made public through a system security breach.

Popular social networking websites like Friendster [6] are increasing public comfort with the idea of a digital representation of one's social network, and it is likely that successive generations of technology users will have fewer worries about maintaining an accurate digital rendering of their social network.

In addition to keeping their social network current, users will benefit from an accurate and publicly-query-able personal profile. This profile would have much in common with the profile in current social-networking websites (type of humor, eating habits, favorite books, etc.), but might also include special entries related to photos. For instance, if I am a still-life aficionado, I want to be sure to see still-life photos that my friends and acquaintances capture. I may also want to meet new people if they share my interest in still-life photography, either to talk shop about still life issues or for dating/friendship purposes. As the practice of keeping an accurate and query-able personal profile becomes more common, people may be presented with an increasing number of personalized services and dating opportunities.

6. TECHNICAL IMPLICATIONS

There are hardware and system architecture implications of the ideas stated thus far, and the following sections will investigate a few of these implications.

6.1 Hardware Implications

Cameras (both mobile-phone cameras and single-purpose ones) will be fitted with more and different sensors. GPS is making its way into mainstream mobile phones, and infrared/Bluetooth are already a standard feature on so-called "smartphones" today. A digital compass combined with GPS and wireless radio would enable the directional sensitivity needed for instant sharing with people who are in a recently-taken picture. For enabling the touching of phones together to initiate photo sharing, phones will need to have some way to sense the ID of the other phone. This could be close-range RFID, or even just a one-wire serial data connection. Moreover, the addition of new sensors to tomorrow's digital cameras (temperature, humidity, altitude, tilt or other inertial, etc..) will augment the contextual information available for annotation and searching purposes.

6.2 Architectural Implications

In order to leverage a user's social network for searching and sharing purposes, a portable, easily malleable, and secure representation of the network and personal profile will be needed.

Portable: This has two meanings here. First, since the photo sharing activity will happen on mobile as well as desktop devices, in scenarios with and without internet connectivity, a representation of the user's social network must be synchronized across the user's desktop and mobile computing platforms frequently. Second, the information would ideally be represented by an open and universal standard, to allow interoperability between diverse applications and interfaces. The Foaf [5] project is an example of an effort to create such a standard.

Easily malleable: The architecture should easily (and perhaps automatically) support the addition of new friends and acquaintances, as well as changes and updates to the personal profile, in a way that isn't cumbersome to the user.

Secure: Making one's social network explicit comes with the risk of a security breach. In order for users to hand over this personal data to a computerized system, they will have to feel confident that the data will be kept safe and used only in ways that they specify. However, there is evidence to suggest that people are willing to expose amounts personal information, for a reward. For instance, the free Gmail service from Google is supported by targeted advertising informed by the contents of users' personal emails.

7. CONCLUSIONS

Possibilities are great for upcoming innovations in how we capture, search, and share our digital photos. While the inclusion of more sensors and communication capabilities will allow digital cameras to capture and annotate photos for us in more useful and automated ways, new direct-manipulation and natural speech and language processing interfaces will permit the annotation process to be less prohibitive to the user. Modular, scriptable search heuristics that can be traded between users will expand the ways in which photoware applications can find appropriate photographs to share, and the utilization of each user's social network and

personal profile will give the search process a new degree of social intelligence.

8. ACKNOWLEDGMENTS

Thanks to my colleagues at the MIT Media Lab for conversations that have helped in the creation of this paper.

9. REFERENCES

[1] Bell, G. et al. "Passive capture and ensuing issues for a personal lifetime store". In *Proceedings of the 1st ACM workshop on Continuous archival and retrieval of personal experiences*. New York, New York, USA. 2004, pp. 48-55.

[2] Clarkson, B. “Life patterns: structure from wearable sensors” Doctoral Thesis, MIT. 2002.

[3] Cooper, M. et al. “Temporal event clustering for digital photo collections”. In *Proceedings of the eleventh ACM international conference on Multimedia*, pages 364–373. ACM Press, 2003.

[4] Flickr website: http://www.flickr.com/

[5] Foaf Project website: http://www.foaf-project.org/

[6] Friendster website: http://www.friendster.com/

[7] Frohlich, D., Kuchinsky, A. et al. “Requirements for Photoware”. In the *Proceedings of the ACM CSCW*, pp. 166-75, 2002.

[8] Garrett, J. An Interview with Flickr's Eric Costello. webpage: http://www.adaptivepath.com/publications/essays/archives/000519.php

[9] Google website: http://www.google.com/

[10] Hollink, L., Schreiber, A. T., Wielemaker, J. and Wielinga, B.J. ”Semantic Annotation of Image Collections.,” in *Workshop on Knowledge Markup and Semantic Annotation*, KCAP’03, Florida, 2003.

[11] KodakGallery Website: http://www.kodakgallery.com/

[12] Li, Charlene. First Look at Yahoo! 360. http://blogs.forrester.com/charleneli/2005/03/first_look_at_y.html

[13] Lieberman, H., Rosenzweig, E., and Singh, P. “Aria: An Agent for Annotating and Retrieving Images.” *IEEE Computer*, pp. 57-61, July 2001.

[14] Liu, H. and Singh, P. “ConceptNet: A Practical Commonsense Reasoning Toolkit”. *BT Technology Journal*. Kluwer, 2004.

[15] Mäkelä, A., Giller, V., Tscheligi, M., and Sefelin, R. “Joking, storytelling, artsharing, expressing affection: A field trial of how children and their social network communicate with digital images in leisure time”. In *Proceedings of CHI 2000*, 548–555. ACM, 2000.

[16] Miller, G. A. “WORDNET: A Lexical Database for English”. In *Communications of ACM*(11), pp. 39-41, 1995.

[17] Mills, T. J., Pye, D., Sinclair, D., and Wood, K. R. “Shoebox: A digital photo management system”. *Technical Report Technical Report 2000.10*, AT&T Research, 2000, http://ftp.uk.research.att.com:/pub/docs/att/tr.2000.10.pdf

[18] Mynatt, E. and Voida, A. “Six themes of the communicative appropriation of photographic images”. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. pp. 171 - 180. Portland, Oregon, USA. 2005.

[19] Palm Treo, website: http://www.palm.com/us/products/smartphones/treo650/

[20] Rodden, K. and Wood, K.R. “How Do People Manage Their Digital Photographs?” In *Proceedings of CHI 2003*, ACM Press, pp. 409-416, 2003.

[21] Shneiderman B. & Kang H. Direct annotation: A drag and drop strategy for labeling photos. *Proceedings of the International Conference on Information Visualization 2000*, IEEE, Los Alamitos, CA. July 2000, 88-95.

[22] Shutterfly website: http://www.shutterfly.com/

[23] Smugmug website: http://www.smugmug.com/

[24] Wallop website: http://research.microsoft.com/scg/

[25] Yahoo 360. website: http://360.yahoo.com/reg/whatis.html