

Turning Photo Annotating Tasks into Instant Messaging Fun: Prototyping, User Trials, and Roadmapping

Yuechen Qian and Loe M.G. Feijs

Eindhoven University of Technology, 5600 MB Eindhoven, Netherlands
{y.qian, l.m.g.feijs}@tue.nl

Abstract. In this article we report on our research that integrates photo annotation tasks into online chatting. Users of our system can share and annotate digital photos online while chatting. There are two major innovations: first, users can add annotations and comments to photos in a collaborative manner and secondly, the software itself extracts information from conversations to generate extra annotations. The boring and tedious task of annotating photos is turned into an essential part of an attractive fun activity, viz. online chatting. This article also provides a roadmap towards a systematic analysis of linguistic aspects of automated interpretation of message conversations.

1 Introduction

Many systems have been prototyped to facilitate the easy and attractive use of digital photos. Zoomable user interfaces are used in PhotoMesa Image Browser for users to search and browse photos [1]. Advanced systems support similarity-based image browsing [2], automatic image clustering [3], and photo concept browsing [4]. Unfortunately, the lack of meaningful metadata that describe the place, time, and event in which photos were taken remains one of the problems that hinder the easy use of photos. MSN Messenger and Yahoo Messenger provide various technical solutions for social communication. Picasa's Hello system even allows online users to browse photos simultaneously while chatting. The information passing by instant messaging systems can be used in the quest for meaningful metadata of digital photos. We have developed an innovative chatting system that not only allows online users to share and annotate photos together, but also can automatically extract meaningful data from messages. We developed a roadmap towards automated interpretation of messages.

2 Talkim: Sharing and Annotating Photos Online

As shown in Fig. 1, the user interface of our system consists of the message area on the left, the photo sharing area in the middle, and the photo search area on the right. Participants of a chatting session have the same view on the message and photo

sharing areas. When a user drag-n-drops an image onto his photo sharing area, the others will see the same image in their photo sharing areas, as in Picasa's Hello system.

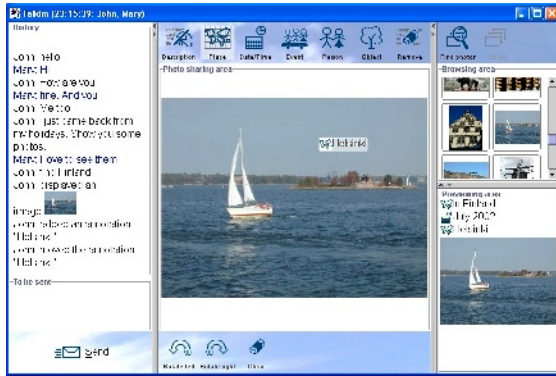


Fig. 1. A snapshot of Talkim's user interface.

Different from the Hello system, Talkim's photo sharing area also functions as a whiteboard. Users can simultaneously put text annotations to anywhere in the photo sharing area and move them freely. This direct annotation function [5] is essential here: it helps to raise attention and create a focus of discussion. A user would otherwise have to type lengthy sentences to explain interesting elements in a photo. In our system users can easily pinpoint interesting elements such as a building, a person, or a pet in photos. To the best of our knowledge, it is for the first time that the photo annotating task is achieved by using the online whiteboard concept.

Talkim also extract annotations from messages. Four annotation types were worked out in our system. They are *location of taking*, *time/date of taking*, *event of taken*, and *persons in photo*. First, Talkim parses messages that are complete fact sentences. For example, "This photo was taken in Finland". Once recognized, such a message is directly added to the photo sharing area as an annotation. Secondly, Talkim recognizes messages that are complete questions, such as "where was it taken?" Such messages provide contexts for processing subsequent messages. Thirdly, Talkim handles ellipsis sentences such as "in a restaurant". A message "in a restaurant", as a reply to a message "where was it taken", can be treated as a location-related annotation. We defined several schemas to recognize and categorize different types of messages.

3 User Trials

We conducted several user trials on our system. Four pairs of people were invited. Persons in a pair know each other well. Each pair used our system for 30 minutes. We recorded all messages that were sent by users. We identified occurrences of question-and-answer dialogues. As shown in Fig. 2, our system managed to identify "at home" as an annotation of the type *location of taking*.


10 Mar 2004 09:30:04 GMT	Youpca	displayed an image. 00fabb9e6818_036589_Car 1.jpg
		
10 Mar 2004 09:30:10 GMT	Youpca	ShowMessage nice car
10 Mar 2004 09:30:11 GMT	Youpca	ShowMessage hu
10 Mar 2004 09:30:14 GMT	Youpca	ShowMessage h
10 Mar 2004 09:30:21 GMT	wouter	ShowMessage:where did you take the picture?
10 Mar 2004 09:30:34 GMT	Youpca	ShowMessage:at home
10 Mar 2004 09:30:34 GMT	Youpca	abstracted an annotation "at home"
10 Mar 2004 09:30:37 GMT	Youpca	moved the annotation "at home".
10 Mar 2004 09:30:37 GMT	Youpca	added an annotation "mitsubishi"
10 Mar 2004 09:30:43 GMT	Youpca	ShowMessage:cool
10 Mar 2004 09:31:01 GMT	wouter	ShowMessage:yes
10 Mar 2004 09:31:03 GMT	wouter	ShowMessage:it is
10 Mar 2004 09:31:25 GMT	Youpca	ShowMessage:ask me more

Fig. 2. Part of the log file that recorded the conversation done by two users.

Participants all liked the integration of photo sharing and chatting. One pair continued to use our system for a while after their session was finished. All participants liked the functionality that they can put annotations onto the photos and move annotations around. Our analysis shows that large amount of metadata were added to photos. Two participants found even played a “grabbing” game: one moves an annotation while the other tries to grab it and move it to another place. They thought it was fun and they played this game many times throughout the user trial. They often used the on-screen annotating functionality to chat, not only for annotating photos.

4 Roadmap

Linguistic analysis is crucial to improve the automated analysis and understanding of message conversation. Meaning can be derived by following the syntactic structure of the text [6]. We distinguish three levels: words, sentences and dialogues.

At the level of words, words have to be assigned a grammatical category and if possible aspects of their meaning have to be retrieved. Errors deserve special attention. If the conversation language is known, a minimization of the Levenshtein edit distance to the closest dictionary word should suffice for the errors like “picuter”. For long un-recognizable words, such as “girlfriendshouse”, we propose a separate dictionary-based splitter. In categorization, two important categories are pronouns and prepositions. Personal pronouns are easy to recognize. They indicate persons and will be a rich source of annotations. The “wh” pronouns indicate questions and they can be used to determine the annotation type of the subsequent answer. Prepositions can help to determine the annotation type (“in” is either place or time). Dis-ambiguation is doable by name if the following word indicates a place, time, or event.

At the level of sentences, sentences have to be assigned a sentence type, the phrase structure has to be parsed, and if possible aspects of their meaning have to be retrieved. Though messages are short and simple, semantic analysis still is challenging because of frequent ellipsis and linguistic pointers. For syntax we propose

to use a straightforward phrase structure grammar. Any bottom-up parsing technique or backtracking top-down technique will suffice. Prepositional phrases are particularly important because they contain many clues. For semantic analysis it is necessary to develop underlying models, e.g. for personal preferences and for relations between people.

At this level of dialogues, dialogues are used to determine whether recognized sentences should be treated as annotations or not. Such dialogic analysis is necessary to categorize ellipsis and ambiguous sentences. Asynchrony is yet another challenge in dialogic analysis. When chatting online, participants can talk about things in an interleaving manner, having parallel threads of topics. This makes it difficult to extract annotations. We expected that formal dialogue theory will help to solve these issues.

5 Concluding Remarks

Whereas taking photos, editing them and sharing them with others is considered a creative and highly rewarding process by many people, the difficulty of classifying and grouping them has always been difficult and tedious. Most of the propositions that promise to automate these tasks rely on the assumption that photos are annotated, which in fact they are not, thus shifting the problem from one tedious task to another even more tedious task. Precisely the latter task is turned into fun by the type of application investigated here. The result of the user trials of our system shows the enthusiasm that the participants had on our system. The integration of photo browsing and chatting facilitates online photo sharing and discussing activities. Putting this annotating task in a social context makes the photo annotating task fun and appealing. Allowing users to freely annotate photos on the photo itself produce a large amount of annotations without many computational difficulties. Such information can be directly used in photo search even without further processing. Finally, the roadmap represented here shows the way ahead to improve automated annotation extraction.

References

1. Bederson, B.B.: PhotoMesa: a zoomable image browser using quantum treemaps and bubblemaps. In Proceedings of ACM UIST 2001. 71-80
2. Rodden, K., Basalaj, W., Sinclair, D., Wood, K.: Does organisation by similarity assist image browsing? In Proc. SIGCHI CHI 2001. 190-197
3. Platt, J.C., Czerwinski, M., Field, B.A.: PhotoTOC: Automatic clustering for browsing personal photographs. MSR-TR-2002-17, Microsoft Research, 2002.
4. Qian, Y., Feijs, L.M.G: Stepwise Concept Navigation. In de Moor, A., Ganter, B. (Eds.): Using Conceptual Structures of ICCS 2003. Germany. 2003. 255-268
5. Shneiderman, B., Kang, H.: Direct Annotation: A Drag-and-Drop Strategy for Labeling Photos. Proc. International Conf. on Information Visualisation, 2000.
6. Rosetta, M.T.: Compositional Translation. Kluwer International Series in Engineering and computer science: Volume 273. (1994).