A Context Centric Approach for Semantic Image Annotation and Retrieval

Najeeb Elahi, Randi Karlsen
Department of Computer Science
University of Tromsø
Tromsø, Norway
najeeb.elahi@uit.no, randi@cs.uit.no

Sigmund Akselsen
Group Business Development and Research
Telenor
Tromsø, Norway
sigmund.akselsen@telenor.com

Abstract— **In this preliminary research, we discuss techniques to improve the quality of image retrieval and image management with the help of context information over the web. Our hypothesis is that leveraging the semantic annotated contextual metadata of the image would yield the relevant search results and facilitate building a consistent, unambiguous image knowledge base. Inferencing capability of semantic technology is very useful to deduce the new metadata from the existing metadata about images. We discuss the three different aspects of image context such as spatial, temporal and social context. Spatial and temporal context of the image can easily be gathered by taking advantage of modern digital camera technology. In order to acquire social contextual metadata, we benefit from social network sites.**

Keywords: image retrieval, context-aware, social networks, semantic web

I. INTRODUCTION

The invention of digital cameras and portability offered by mobile phone digital cameras has considerably fueled the popularity of digital images. Moreover, the affordability of these devices has given the common man the opportunity to capture his world in pictures and conveniently share them with others. Therefore, people are now capturing and sharing far more images than ever before. It indeed confirms the Susan Sontag's vision of a world where "everything exists to end up in a photograph" [3]. As a result, billions of searchable image data exist with diverse semantic and visual contents, geographically disparate locations, and is continuously growing in size [4]. However, these collections are inherently difficult to navigate, due to their size and lack of semantic understanding of the content of images by machine. To illustrate the advantages of context centric image retrieval approach, it is important to shed some lights on the limitations of traditional approaches being practiced.

The most common approaches used for image retrieval are Text-Based Image Retrieval (TBIR) and Content-Based Image Retrieval (CBIR). TBIR retrieval is achieved by matching text to annotation associated with the image. The drawbacks of TBIR are that manual annotations varies from person to person and it is challenging to build an efficient indexing for a text based image database. CBIR uses an input image, which is matched to the visual content of the image archives. One shortcoming with CBIR is that prior to a search, the user must have a similar image at hand. Datta et al. [4] has identified two core research problems with CBIR, i) representing an image with mathematical functions, and ii) finding out the similarities between two images based on their abstract descriptions. In a nutshell, both approaches suffer from the semantic gap between user's comprehension of the image and machine capability to understand the semantics of images.

We believe that the current state-of-the-art in context-based image retrieval holds adequate promise and maturity to be useful for real-world application, if the semantic approach is adopted. In this work, we take the merits of social networks in order to acquire social contextual metadata. The image context will be acquired either implicitly by automatic annotation service with the help of inferential mechanism over the existing semantic metadata, or explicitly by requiring the user to specify it with the assistance of system suggestions.

The paper is organized as follows: Section II discusses the use case scenario, section III briefly describes the significance of image annotation and suggests the research directions. The paper concludes with the main features of the work and comments on the future work.

II. USE CASE SCENARIO

Ragnhild is living in Lillehammer and is a software engineer by profession. Because of a heavy workload, she always has very tight schedule. She gets five weeks of vacation every year, which she usually takes in summer and spends traveling around the globe. She has heard a lot about the beautiful Greek islands but has never managed to visit them. This summer, Ragnhild is thinking to explore an exotic destination, an off the beaten track island called Icaria in the east Aegean Sea. Before she finalizes her plan she wants to get relevant non-commercial information and images from her trusted circle of friends to have a realistic impression of the place. She certainly wants to avoid commercial images and unreliable information. Ragnhild is wondering if any of her friends have visited the island before, and if they have shared any informative images that can be useful for her. She is an active member of a social network called Facebook[1], and she has a number of friends who share similar interests. Ragnhild makes a text query, "summer vacation spots Icaria", and finds several very specific

[1] http://www.facebook.com/

images from her social contacts, automatically prioritized by the system.

On the other hand, Kostas is a medical doctor living in Tromsø. Kostas is an Internet friend of Ragnhild. They have never met, however they share the same interests that led them to join the "Travel Globe" community and they became friends. Kostas is a very energetic guy with a zest for traveling. He visited Greece last summer. He enjoyed it a lot and now he would like to share his joy of visiting wonderful Greek islands. In order to make his images more visible, he annotates the event in the album and the activities happening inside each image, and he tags all the people depicted. Kostas adds descriptions and other viewers add notes and comments, all of which is viewable to everyone who views the image. Besides, the system automatically adds the time and geographical position information, which is also confirmed by Kostas.

Kostas subscribes to the Facebook Travel Calendar application [2] on his profile, which manages his travel schedule and shares this information among his friends. His friends and the system can easily track Kostas's location, see the sequence of events happening, and by appending taken images with those calendar events broaden the image understanding.

Ragnhild finds Kostas's informative images quite fascinating along with other people's comments, notes and suggestions. Ragnhild also receives the system's inferred information about the social event details, hotel expenses, beaches and other islands nearby. Finally, Ragnhild successfully makes a vigorous plan and looks forward to having a great summer vacation.

III. RESEARCH DIRECTIONS

This research will be directed towards improving the quality of the image retrieval and image management with the help of context information. We have the following hypothesis

- Leveraging the semantically annotated contextual metadata of the image, it would yield the relevant search results and facilitate building a consistent, unambiguous image knowledge base.
- Including the social network sites as an application domain would provide enormous contextual metadata of the image.

Following is the detailed description about the background analysis of our hypothesis along with suggestions.

A. Automatic Image Annotation and Semantic Retrieval

From a general viewpoint, annotations can be well thought-out as metadata. They associate remarks with existing contents [18]. A first step towards building semantically enabled image retrieval system is to have an infrastructure that handles and associates metadata with image contents. In order to reach this goal, we will develop an automatic image annotation system. The development of an automatic image annotation system will help us to build a Semantic Web accessible image repository that contains images with rich semantic metadata, so that they can be integrated with other resources and consumed by machines.

Smeulders et al. discuss in the survey of content-based image retrieval [12] that "semantic gap" is the major problem with all approaches that rely on the visual similarity of image for judging semantics of image. The semantic gap occurs between low-level image features that machine can parse and higher-level real world concepts. In order to bridge this gap, there is a distinctive need to build a sequential mapping between image contents and real world concepts. Such problems can be handled by introducing the efficient algorithm for defining the image annotation [4,6,7,8,17]. In Marc Davis' research dealing with camera phone image annotation [15], he coined the idea of Context-to-Contents inference for image retrieval that is closely in line with our need, using context of the images to infer the content of the images.

An essential part of this research work is an inference engine that leverages the context of image creation to infer metadata about image content. In order to achieve machine-based reasoning about the evidence image's context, we need additional interpretive semantics that must be attached to the data. A shift from data-intensive approach to a semantic-rich model of evidence is suggested [17]. Using semantic reasoning over the already gathered evidence, to infer image contextual metadata has shown quite a few advantages [14]. One of the main advantages is, reuse of existing metadata in order to generate more semantic metadata for non-annotated images. For example, in contrast to the majority of the images in a same album, only one of the images has rich contextual metadata about the location, time, main activity, tagged people, weather, and ownership[3]. This metadata can be used for other images as a reference metadata. Consequently, by using the reasoning capability of semantic technology, the reference metadata can be unambiguously classified and associated with other images.

Our approach would be grounded upon image ontology specifying the domain knowledge and description logic taxonomy. The inferential engine would be responsible for semantic reasoning and image retrieval. Figure 1 shows the abstract view of this approach.

The objective in particular is to build an infrastructure based on OWL-DL. Though OWL DL lacks in expressivity power compared with OWL Full, it maintains decidability and regains computational efficiency. The computational efficiency is an important feature since the mechanism has to handle scores of complex social, spatial and temporal contextual metadata. It comprises of all the OWL language

[2] http://www.facebook.com/apps/application.php?id=7040314722&ref=s#

[3] An individual who uploaded and shared image

constructs with restrictions and it is based on the Description Logics (hence the suffix DL).

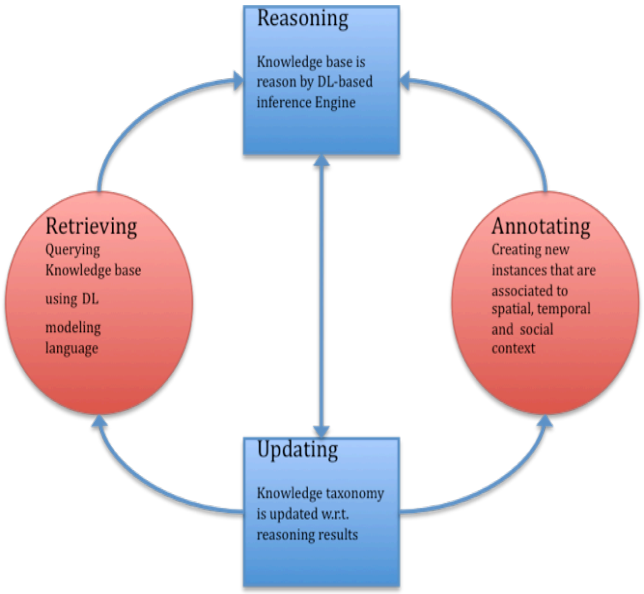

Figure 1: Abstract view of image annotation and retrieval

These are the decidable parts of the First Order Logic[4] and are therefore amenable to automated reasoning. It makes sure that all its entailments are computable and the computations will be finished within a finite time. In order to achieve more expressivity and decidability, we will use Semantic Web Rule Language, which is designed as an extension of OWL DL, but this may come at the cost of additional complexity.

B. Contextual Image Metadata

The term *context* has been used in several ways in different areas of computer science, such as contextual search, context-sensitive help, multitasking context switch and so on [21]. In fact, context is a general concept and has a loose definition. Therefore, there are numbers of definitions of context that can be found in a computer's application domain [19,20,21]. Many of them define context in terms of characteristics of the surrounding environment that determine the behavior of user and information relevance to the user. Dey [9] defines context as

> "*any information that can be used to characterize the situation of an entity. An entity is a person, place, or object that is considered relevant to the interaction between a user and an application, including the user and applications themselves*"

Dey further elaborates the context values that generate the powerful understanding of the current situation, by using *primary* context such as location, entity, activity and time as an input into other sources of contextual information. Dey's work shows significance of knowing the current situation of the user in application domain. However, Dey leaves an important question about how information becomes relevant, unanswered [1].

Some advancement has been made by including the social aspect of images as a key element of the context. In our research sphere the main entity is the image. It is important to note that in the arena of social context, there is a strong correlation between the image and the individual who uploaded that image. As described by [15] the spatial, temporal and social context are potential factors for generating contextual metadata. Combining these contextual metadata from a given user and across users makes it possible to infer the target image content.

By taking the advantage of modern GPS-enabled digital camera technology, we can easily gather the three most important spatial-temporal aspects of image context.

- The date and time of image capture
- Location of the site where image was captured
- Who took that image

However, the automatically gathered spatial and temporal contextual metadata is unable to answer the following questions.

- Who is depicted in the image
- What are the relationships between people in the image
- What is the main activity of the image

Recent works [15,10] have addressed the aforementioned question by adding the social aspect as a contextual metadata. In this research work, we will take the merits of social networks in order to acquire social contextual metadata.

Nowadays social networks are one of the most widely used platforms of online communities for sharing text, images and video resources among the users. According to the report from ComScore in July 2007[5], famous social network sites like MySpace, Facebook, Friendster, Orkut, Bebo etc. are enjoying about 65 million daily visitors and the growth rate is 50% to 300%. Also the immense popularity of image sharing communities (e.g., Flickr, Photobucket, Photo.net) has made it imperative to introduce a new approach.

Social network sites are gravely based on user profiles, which offer a description of each member. In addition to the images and videos uploaded by the member, the user profile also contains comments and positive/negative ratings over the uploaded resources. The styles of these sites have emerged out of several diverse services, such as dating, tailored advertisement etc. In order to meet the demands of new services, other detailed descriptions related to the user

[4] First Order Logic (FOL), http://en.wikipedia.org/wiki/First- order logic

[5] ComScore, Social Networking Goes Global, http://www.comscore.com/press/release.asp?press=1555 [accessed on Jan. 14, 2007]

are required, such as: demographic details (age, sex, location, etc.), tastes (Interests, favorite music, favorite star, movies list, etc.), personal photographs, and open-ended descriptions e.g. who the person would like to meet [2]. Furthermore, these sites empower user to add new software applications.

The rapidly growing number of users who are sharing and uploading images, well defined information about the users, and appealing application domains draw our attention to include social networks for acquiring social contextual metadata in our research project.

IV. Conclusion and Future work

This is a preliminary discussion of issues that are related to improve the image retrieval with the help of context. The immense amount of metadata associated with images has made it imperative to exploit novel semantic web approach. We will develop an infrastructure for efficiently storing and retrieving the image with the help of semantically rich contextual metadata. It will also provide an automatic image annotation service that will reuse the existing contextual metadata. Since privacy is a big issue in social networks, image-capturing devices such as digital cameras or camera-phones will be used for the identification of the user.

References

[1] Elgesem,D. and Nordbotten.J. (2007). The Role of Context in Image Interpretation. Context-based Information Retrieval, CIR'07. at CONTEXT'07. Roskilde University, Denmark. Aug.20-21, 2007.

[2] Boyd, Danah. "Why Youth Social Network Sites: The Role of Networked Publics in Teenage Social Life." Youth, Identity, and Digital Media. Edited by David Buckingham. The John D. and Catherine T. MacArthur Foundation Series on Digital Media and Learning. Cambridge, MA: The MIT Press, 2008. 119–142. doi: 10.1162/dmal.9780262524834.119.

[3] Susan Sontag. *On Photography*. Picador, New York, NY, 1977.

[4] Datta, R., Joshi, D., Li, J., and Wang, J. Z. 2008. Image retrieval: Ideas, influences, and trends of the new age. ACM Comput. Surv. 40, 2 (Apr. 2008), 1-60. DOI= http://doi.acm.org/10.1145/1348246.1348248.

[5] A. Smeulders et al. Content-based image retrieval: the end of the early years. *IEEE PAMI*, 22(12):1349–1380, 2000.

[6] Sinha, P. and Jain, R. 2008. Classification and annotation of digital photos using optical context data. In Proceedings of the 2008 International Conference on Content-Based Image and Video Retrieval (Niagara Falls, Canada, July 07 - 09, 2008). CIVR '08. ACM, New York, NY, 309-318. DOI= http://doi.acm.org/10.1145/1386352.1386394.

[7] M.M. Tuffield et al., "Image Annotation with Photocopain," *Proceedings of the Fifteenth World Wide Web Conference (WWW2006)*, 2006.

[8] N. O'Hare et al., "MediAssist: Using Content-Based Analysis and Context to Manage Personal Photo Collections," *Image and Video Retrieval, 5th International Conference*, Tempe, AZ, USA: Springer, 2006, pp. 529-532.

[9] Daniel Salber, Anind K. Dey, and Gregory D. Abowd. The context toolkit: Aiding the development of context-enabled applications. In *Proceeding of the CHI 99 Conference on Human Factors in Computing Systems: The CHI is the Limit*, pages 434-441, Pittsburgh, PA, May 1999. ACM Press.

[10] Naaman, M., Yeh, R. B., Garcia-Molina, H., and Paepcke, A. 2005. Leveraging context to resolve identity in photo albums. In Proceedings of the 5th ACM/IEEE-CS Joint Conference on Digital Libraries (Denver, CO, USA, June 07 - 11, 2005). JCDL '05. ACM, New York, NY, 178-187. DOI= http://doi.acm.org/10.1145/1065385.1065430

[11] Web image repository for biological research images. In web image repository for biological research images. In Proc. of the 5th European Semantic Web Conference, Tenerife, Spain, 2008.

[12] Gevers, T. and Smeulders, A. 2000. Pictoseek: Combining color and shape invariant features for image retrieval. IEEE Trans. Image Process, 9, 1, 102–119.

[13] DATTA, R., LI, J., AND WANG, J. Z. 2007. Learning the consensus on visual quality for next-generation image management. In Proceedings of the ACM International Conference on Multimedia.

[14] Rodrigo Fontenele Carvalho, Metadata goes where Metadata is: contextual networks in the photographic domain, Ph.D. Symposium of the 5th European Semantic Web Conference, 2008.

[15] Davis, M., King, S., Good, N., and Sarvas, R. 2004. From context to content: leveraging context to infer media metadata. In Proceedings of the 12th Annual ACM international Conference on Multimedia (New York, NY, USA, October 10 - 16, 2004). MULTIMEDIA '04. ACM, New York, NY, 188-195. DOI= http://doi.acm.org/10.1145/1027527.1027572.

[16] J. Kahan, M. -R. Koivunen, E. Prud'Hommeaux, R. R. Swick Annotea: an open RDF infrastructure for shared Web annotations *Computer Networks*, *Volume 39, Issue 5*, 5 August 2002, Pages 589-608

[17] Hu, B., Dasmahapatra, S., Lewis, P. and Shadbolt, N. (2003) Ontology-based Medical Image Annotation with Description Logics. In: *The 15th IEEE International Conference on Tools with Artificial Intelligence*, 3-5, November 2003, Sacramento, CA, USA.

[18] Kahan, J. and Koivunen, M. 2001. Annotea: an open RDF infrastructure for shared Web annotations. In *Proceedings of the 10th international Conference on World Wide Web* (Hong Kong, Hong Kong, May 01 - 05, 2001). WWW '01. ACM, New York, NY, 623-632. DOI= http://doi.acm.org/10.1145/371920.372166.

[19] Bill Schilit, Norman Adams, and Roy Want. "Context-aware computing applications". In *Proceedings of IEEE Workshop on Mobile Computing Systems and Applications*, pages 85-90, Santa Cruz, California, December 1994. IEEE Computer Society Press.

[20] Albrecht Schmidt, Kofi Asante Aidoo, Antti Takaluoma, Urpo Tuomela, Kristof Van Laerhoven, and Walter Van de Velde. Advanced interaction in context. In *Proceedings of First International Symposium on Handheld and Ubiquitous Computing, HUC'99*, pages 89-101, Karlsruhe, Germany, September 1999. Springer Verlag.

[21] Chen, G. and Kotz, D. 2000 A Survey of Context-Aware Mobile Computing Research. Technical Report. UMI Order Number: TR2000-381., Dartmouth College.