Image metadata as a means to gather image relevant information

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

Roger Bruun Asp Hansen
Department of Computer Science
University of Tromsø, Norway
roger.b.hansen@uit.no

Abstract

This paper presents an approach for collecting image relevant information, where the goal is to provide the user with information about i) the object or event depicted in an image, and ii) the location where the image was taken. The approach aims at improving the image viewing experience by presenting supplementary information such as location names, tags, weather information, placement on map, geographically nearby images, Wikipedia articles and web pages. The information is automatically collected from various sources on the Internet based on the image metadata *gps latitude/longitude* values, *date/time* of image capture and a *category* keyword provided by the user. The paper describes a prototype system that demonstrates and tests the approach.

1 Introduction

As the number of digital photos grow, an increasing problem is managing them so that they can be found and displayed in an efficient manner. The use of supplementary information (such as tags and annotations) is in many cases important, not only for retrieving images, but also for providing users with information about what an image depicts.

A current trend is that digital photos are displayed together with more and more additional information. Digital albums, such as Google Picasa[1] and iPhoto[2], display images on a map and provide tools for identifying faces, adding tags and editing metadata. Web album and photo sharing sites, such as Flickr[3] and Panoramio[4],

This paper was presented at the NIK-2012 conference; see http://www.nik.no/.

[1] http://picasa.google.com/
[2] http://www.apple.com/ilife/iphoto/
[3] http://www.flickr.com/
[4] http://www.panoramio.com/

allow collaborative tagging and commenting. These and other examples show that the image alone is no longer sufficient, and that users in many situations would like their images displayed together with different types of related information.

We believe that image relevant information can be automatically collected from sources on the Internet based on image metadata. This paper describes InfoAlbum, a prototype system that demonstrates our approach for automatically collecting a variety of image relevant information from sources on the Internet based on the image metadata $\{gps\_coordinates, date/time, category\}$.

Location and date/time of image capture are used for gathering location names, weather information, Wikipedia articles, placement on map, and geographically nearby images. Category is in our approach a keyword representing a general description of the image content (for example "tower", "church", "bridge", "concert" or "festival"), and is used for focusing the search for image related information. The information collection done in InfoAlbum is, to our best knowledge, unique in the way *category, location* and *date/time* metadata is used as a basis for gathering image relevant information.

The main objective of InfoAlbum is to provide users with information about i) the object or event depicted in an image, and ii) the location where the image was taken. Performance of InfoAlbum with respect to these objectives is tested and evaluated in this paper.

2 Related work

In recent years approaches for automatically combining images with related information have emerged. Many of these approaches use content analysis of the image, sometimes in combination with location information (such as gps coordinates), to find similar images or automatically determine image tags.

Displaying images on a map based on gps coordinates, is a popular approach used in systems such as Flickr, Panoramio and Google Picasa. Also [10] describes how images uploaded to Flickr are placed on a World map, based on textual tags provided by users.

Automatic or semi-automatic annotation of images is the focus in a number of publications. In systems such as MonuAnno [7], ZoneTag[5] [1] and iPicca [8] images are given location relevant tags by collecting tags from existing images in Flickr and Panoramio. Relevant images are selected based on gps coordinates or other location information. Zonetag suggests tags based on past tags from the user, the user's social network, and names of nearby real world entities. SpiritTagger [6] uses Flickr to assemble visually relevant images weighted by geographic distance from the image that is to be annotated. A set of geographically representative and frequently used tags is collected and suggested to the user. Tag Suggestr [5] expands user provided tags of an image by incorporating tags from other images which are visually similar. The work [9] and [11] describe how to collect and cluster images of landmarks, which are subsequently automatically annotated and enriched with Wikipedia information.

Google Goggles[6] and Nokia Point & Find[7] are mobile applications that let us use pictures taken with a mobile phone to search the web for information. Image

[5]http://zonetag.research.yahoo.com/

[6]http://www.google.com/mobile/goggles/

[7]http://europe.nokia.com/services-and-apps/nokia-point-and-find

recognition is used to find information such as books and DVDs, landmarks, logos, contact info, artwork, businesses, products, and for barcode scanning.

Google Search by Image[8] is a new service from Google that uses computer vision techniques to match an image to other images in the Google Images index. Based on the matches, the service tries to generate a "best guess" text description of the image, as well as to find other images that have the same content as the search image.

Some systems provide augmented reality by using the camera of a mobile device as a "looking glass", where information about the user's surroundings is retrieved and displayed as an overlay on the camera screen view. Two examples are Wikitude World Browser[9] and Layar Reality Browser[10], which both collect overlay information from a supporting server based on gps location information.

Linked Data[11] [2] and the Linking Open Data Community Project[12] take a general approach to combination of information. Linked Data is about using the web to connect related data that is not currently linked, or using new methods for linking already connected data. It includes initiatives to publish various open data sets with useful metadata so that data from different sets can be related and linked.

Our work is also related to image-based question answering [12] in the sense that we assume an implicit query: "Give me information related to this image", where the relationship can be with respect to both location and image content. As opposed to much of the related work, we are not using the image itself as query, but rather rely on image metadata, either extracted directly from the image EXIF header (such as date/time and gps location) or derived from external information sources (such as location names obtained from Flickr). Also, the use of a category keyword, to focus the information search, is novel compared to other approaches described here. An early prototype implementation of our approach is presented in [4].

3 Collecting image relevant information

Architecture

The InfoAlbum architecture, depicted in Figure 1, shows the three components of the system; the Interface, the Information database and the Context Information Collector (CIC). InfoAlbum is implemented as a web service with an interface that accepts an image as input, allows the user to specify a category for the image and displays the image together with all collected information after retrieval is done by CIC. The image together with collected information is stored in the Info database.

An image represents one query to the InfoAlbum system, and the system responds with information about both image content and the location where the image was taken. The CIC component collects relevant information from different Internet sources based on the set of image metadata $IM = \{D, G, c\}$, where D represents date/time of image capture, G is a pair of numbers, $G = \{lat, lon\}$, representing latitude and longitude coordinates, and c is a keyword representing the image category. Date/time and gps coordinates are extracted from the image EXIF header, while category is provided by the user.

[8] http://www.google.com/insidesearch/searchbyimage.html
[9] http://www.wikitude.org/
[10] http://www.layar.com/
[11] http://linkeddata.org/
[12] http://esw.w3.org/SweoIG/TaskForces/CommunityProjects/LinkingOpenData

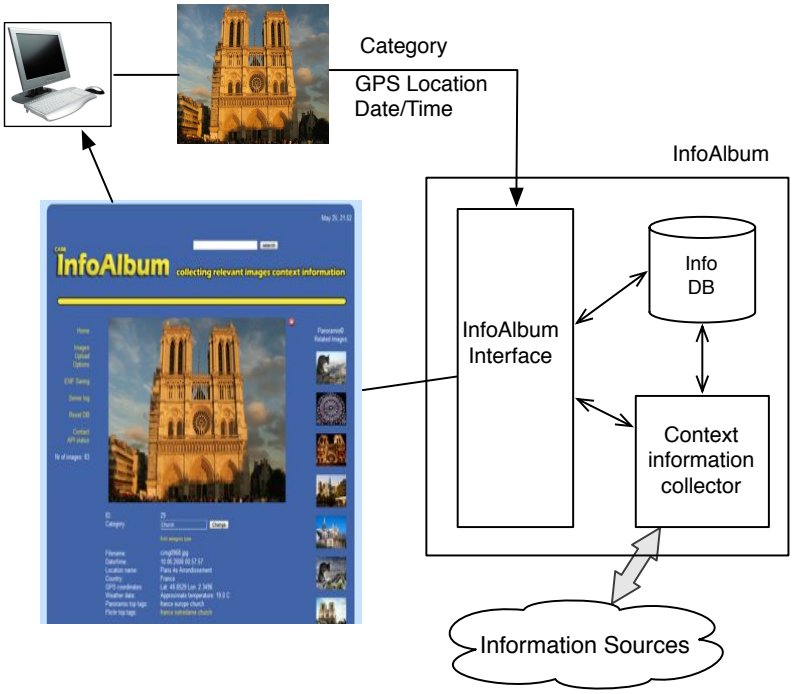

Figure 1: The InfoAlbum system

Image type and category

The InfoAlbum system identifies a set of *image types*, $T = \{Object, ShortEvent, LongEvent, Place, Personal\}$, that reflects the content of images. An *object image* typically depicts a thing, for example a building, statue or a natural thing (such as a river or mountain), while an *event image* depicts some sort of happening, such as a football match, concert or festival. A *Short Event* lasts 0-2 days, while a *Long Event* lasts more the 2 days. *Place* is typically used for landscape or panorama images with no specific object or event in focus, while *Personal* is used for images (of for instance family or friends) for which one cannot expect to find any public information. InfoAlbum is primarily designed to automatically collect information relevant to object and event images.

Category can generally be defined as "any of several fundamental and distinct classes to which entities or concepts belong"[13]. An *image category* is here a user defined keyword that represents a general description of image content (for instance "bridge" or "statue"), and determines the user focus in the image. Category is used to narrow down the information search to detect information that is relevant to image content and the specific user focus.

When an image is submitted to InfoAlbum, the user can choose a category from an image category list, L. To add a new category to L, the user enters the name of the category and selects the type of content the category refers to. For example will "bridge" be of type *Object*, "concert" of type *ShortEvent* and "festival" of type *LongEvent*. For each category c in L, a type t is determined such that $t \in T$.

If there are more then one object of interest in the image and the user wants to change focus, the category of an image can easily be changed and the image resubmitted to InfoAlbum.

The reason for using a general category keyword, is to relieve users from the

[13]http://www.merriam-webster.com/

burden of having to provide specific terms, such as name of the object or event, as people do not necessarily know or remember the names of all depicted objects (for instance attractions visited during a holiday), and may have forgotten or confused the names of places they have visited. However, if image content is known, the user is free to provide a specific keyword identifying the images content by name (for instance "Big Ben" or "Eiffel Tower").

Collected information

Image relevant information is collected from a set of sources, $\mathcal{S} = \{S_1, \ldots, S_n\}$, available on Internet. For each source $S \in \mathcal{S}$, information collection can be described as $S(P) = I$, where P represents a set of parameters used in the information search and $I = \{i_1, \ldots, i_n\}$ the set of collected information.

The information collected by InfoAlbum is listed in table 1, where we also see information sources and the parameters needed for information extraction. The search parameters include date/time (D), GPS coordinates (G), and category (c) together with location names (Ln), synonyms (S) of the category keyword, tags (T), and a radius (r), that together with G give an area of interest.

Given a query image, Q, the category, c, of Q, is used to automatically determine sources and parameters for information search. Images of objects and events are for example handled differently in that *date/time* is important when searching information about an event, while it is not used for object images. A {start_date, end_date} value-pair determines the period of interest when collecting tags and web pages for an event image. In InfoAlbum, the period of interest for *Short Events* is set to 2 days, while it is one month for *Long Events*. For *Place* and *Personal* images, location names, weather information and nearby images are collected, but since a public object or event is not focused, Wikipedia and Google are not searched and Flickr tags are not collected.

Collected information (I)	**Source** (S)	**Parameters (P)**
Location names (Ln)	Flickr	G
Nearby images	Flickr, Panoramio	G, r
Position on map	Google	G
Temperature, weather condition	Weather Underground	G, D
Synonyms to category (S)	WordNet	c
Tags (T)	Flickr	G, r, D, c, S
Wikipedia articles	Wikipedia	Ln
Geo-tagged Wikipedia articles	Wikipedia via GeoNames	G, r
Web pages	Google	c, Ln, T, D

Table 1: Collected information

Location names of country, county and locality (which in most cases equals a city name) are collected, using the flickr.places.findByLatLon function with gps latitude and longitude values as parameters. For images taken outside of cities, locality may not be returned by the function. In that case, only county and country names are used.

Weather information is gathered from Weather Underground[14], by first finding the closest weather station based on latitude and longitude values, and secondly by finding historic information based on the date of image capture

[14]http://www.wunderground.com/

and ID of the weather station. Weather Underground provides information of temperature, humidity, barometric pressure, wind direction, wind speed and condition. Currently, InfoAlbum collects, stores and displays temperature and condition (such as "overcast", "mostly cloudy", "clear" and "light snow").

Geographic position for an image is shown by displaying Google Maps with pinpointed location. The Google Maps API is used for embedding Google Maps on the InfoAlbum web page, and for adding location information.

Nearby images are collected from Flickr and Panoramio based on gps coordinates and a radius, r. InfoAlbum displays a small number of the top-ranked images from each source. Additionally, for Panoramio images, we generate a map through Google Maps with embedded photos on geo-locations.

InfoAlbum also allows users to define a date interval that, together with location information, is used as input to a Flickr search. This enables the user to search for images that are nearby, both in time and location, and possibly find more images of the same event, or choose a different time period to search for images from some other event or situation.

Image tags are automatically collected from relevant images on Flickr using the flickr.photo.search function. For object images, parameters $P = \{G, r, c, S\}$ are used, while for event images, P also include a time interval. Synonyms of the category keyword are, prior to tag retrieval, collected from WordNet[15].

Frequently used tags are collected from Flickr images where i) the category keyword, c, or terms in S is found within the image title, description or tags, ii) the image is positioned within the area of interest, and, for event images, iii) the image is taken within the period of interest. An algorithm for dynamic tag collection is implemented, based on the work in [3].

Geotagged Wikipedia articles. Geonames[16] keeps a database of geotagged Wikipedia articles that are pinned to locations by gps coordinates. We use the function FindNearbyWikipedia in the Geonames API to find articles about objects, and sometimes events, that are relevant to the location where the image was taken. Gps coordinates and a radius, defining the area of interest, are input to the function, and as output we obtain references to Wikipedia articles. To identify articles that are also relevant to image content, the articles are ranked with respect to category and tags.

Location Wikipedia articles. InfoAlbum accesses Wikipedia directly to obtain references to articles describing the place where the image was taken. Based on the *locality*, *county* and *country* names received from Flickr, references to the corresponding Wikipedia articles (if available) are included for each query image.

Web pages are collected through a Google search, using category, location name and tags as search query. The main objective of the search is to collect information that is relevant to the content of the image. For an event image, where we seek information about the specific happening, a period of interest is additionally included in the query.

Image category is important for focusing the search, in that it represents reliable information about image content provided by the user. The automatically collected tags are also very useful, since for many images the collected tags may include the name of object or event.

[15] http://wordnet.princeton.edu//

[16] http://www.geonames.org/

To ensure that obviously irrelevant information is not presented to the user, a technique of filtering is implemented. The user can add words to a filter list which later is used for filtering out undesirable information retrieved in the web search. If words in the filter list are found in the URL, title or summary, these pages are considered irrelevant and filtered out. The remaining 20 top ranked hits are presented to the user and stored in the InfoAlbum database.

Content and Location relevance

In information retrieval, *relevance* denotes how well a retrieved document or set of documents meets the information need of the user. In this work we have the dual objective to collect information that is relevant to i) content of the image and ii) location where the image was taken. We evaluate performance of the system with respect to both objectives and introduce therefore two relevance concepts; *Content Relevance* and *Location Relevance*.

A *Content Relevant* document is relevant with respect to what the user sees in the image as indicated through the category keyword. For example, for an image with category "bridge", a Content Relevant document will describe or mention the depicted bridge. A *Location Relevant* document will typically describe the city or neighborhood where the image was taken, or an object or event nearby. A document that is Content Relevant is also considered Location Relevant.

We calculate precision of retrieved web pages and Wikipedia articles, and use two precision measures, $Precision_{CR}$ and $Precision_{LR}$, where $Precision_{CR}$ is defined as the fraction of the retrieved documents which is Content Relevant, while $Precision_{LR}$ is the fraction of the retrieved documents which is Location Relevant. This is shown in formula 1, where $CRset$ represents the set of Content Relevant documents, $LRset$ the set of Location Relevant documents, and A the set of all collected documents.

$$Precision_{CR} = \frac{|CRset|}{|A|} \qquad Precision_{LR} = \frac{|LRset|}{|A|} \tag{1}$$

4 Testing

The InfoAlbum prototype was tested with 97 images, that represented 97 different queries to the system. Of these, 69 images depicted an object and 28 some event. Image content varied from very famous objects and events, such as the Eiffel Tower and Carnival in Rio, to less famous objects/events.

A main goal when designing InfoAlbum was to develop a system that could take an unknown or forgotten image as query and uncover where it was taken and what it depicted. When testing InfoAlbum, we consequently used generic category keywords for all images. We used categories such as "bridge", "church" and "tower" for object images, and "concert", "festival" and "football" for event images.

In this section we evaluate the performance of InfoAlbum with respect to relevance of the collected Wikipedia articles, web pages and tags. Other types of information, such as location names collected based on gps coordinates, position on map and weather condition, represent facts and were not evaluated here.

Wikipedia articles

Wikipedia articles are collected in two different ways. First by accessing Wikipedia directly, and based on location names, obtain references to articles describing the place where the image was taken. For each image, articles describing locality, county and country are collected. These articles are all location relevant to the image.

Secondly, InfoAlbum collects, through Geonames, geotagged Wikipedia articles from the area where the image was taken. In our test, we used a radius of 1 kilometer from the image capture position and received between 0 and 5 articles per image. These articles typically describe a point of interest in the area (such as a building or construction), and were all relevant with respect to location.

In Table 2 we see that 82% of the object images and 89% of event images received a Content Relevant article. Only 3% of the object images did not receive any geotagged article. We also found that for 5% of the object images, there does not exist a Wikipedia article describing the object in focus.

	CR article	no article	CR article 1. ranked
Object images	82%	3%	74%
Event images	89%	0%	82%

Table 2: Geotagged Wikipedia articles in InfoAlbum

A Content Relevant Wikipedia article for an object image, is an article that describes the object depicted in the image. Articles describing specific events, such as football matches or concerts, are not normally found in Wikipedia. Therefore, an article describing the stadium or arena where the event took place is in this context considered Content Relevant for the event image. We noticed, however, that for event images taken at some annual festival (such as the Munich October festival and Roskilde festival), Wikipedia articles describing the specific event were found.

InfoAlbum uses category keyword and tags to rank articles, and for 74% of object images and 82% of event images the first ranked article describes the image content. This result indicates that by automatically choosing the top ranked geotagged article, we will, with a high probability, receive an article that describes the content of the image.

Tags

Tags were automatically collected from relevant images on Flickr. For each image, the system collected between 0 and 5 tags, where the majority of tags were location names and object names.

	CR tags	LR tags	irrelevant tags	no tags
Object images	59%	78%	6%	4%
Event images	21%	39%	0%	39%

Table 3: Automatically collected tags in InfoAlbum

In Table 3 we see that InfoAlbum collected Content Relevant tags for 59% of object images and 21% of event images. A Content Relevant tag may in many cases give the name of the depicted object (for example Space Needle or Big Ben) or, in some cases, name of festival or artist. Location Relevant tags (mostly location names) are collected for 78% of object images and 39% of event images. Irrelevant tags were collected for only 6% of the object images. We further see that a relatively

high number of event images did not get any tags. We believe one reason for this is the current lack of a sufficient number of tagged event images on Flickr.

Web pages

To evaluate the relevance of collected web pages, we manually inspected each web page to determine if it was a) Content Relevant or b) Location Relevant. Precision with respect to Content and Location Relevance were calculated for each image and is presented in Table 4.

	$Precision_{CR}$	$Precision_{LR}$
Object image	0.44	0.66
Event image	0.31	0.47

Table 4: Average precision of collected web pages

The average precision measures do not show the differences between images with respect to amount of relevant web pages. These differences are illustrated in Figure 2, where we show the number of images for which $Precision_{CR}$ is within a specific range. From Figure 2 we see that 10 object images and 10 event images had a low $Precision_{CR}$ score (between 0 and 9), while for some of the images we received much better results. Even for 3 images, all collected web pages were Content Relevant.

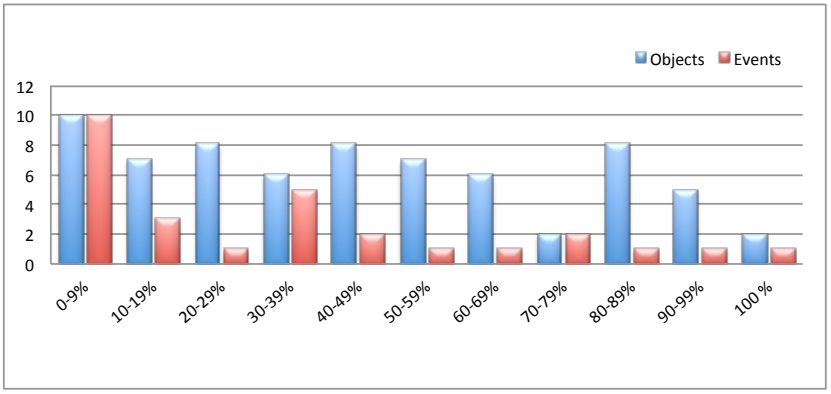

Figure 2: Distribution of content relevance precision scores

To evaluate the impact of using Content Relevant tags in Google search, we compared the average web page precision score of all images with the average precision score for the subset of images where Content Relevant tags were available. We found that the average $Precision_{CR}$ score increased from 0.40 to 0.61 when tags were used. This is a statistically significant improvement, which shows that the use of automatically collected image tags is useful for collecting relevant web pages.

Discussion

InfoAlbum demonstrates automatic collection of information about publicly known objects and events based on image metadata. In our tests, the system collected for example name and description of objects, such as churches, towers, monuments and bridges, and were for events able to identify name of concert and festival, name of band and band members, tour schedule, names of football teams and players and the match result.

Our tests revealed some challenges when using generic category keywords. We see for example that the category *architecture* in many cases is too broad, making it difficult to collect content relevant web pages for these images. Categories such as *tower*, *church* and *bridge* are more narrow, and provide in general more content relevant information. A second challenge is that there may be a number of objects of the same type within a small area. For example when using *church* as category, InfoAlbum may collect information about other churches in the area.

InfoAlbum allows users to re-process the information collection task using a different category. This is useful if there are more than one object of interest in the image and the user wants to change focus, or if the user wants to try a more focused (i.e. narrow) category that may give more relevant information as result.

InfoAlbum does not impose any restrictions on the defined categories. If the user knows exactly what the image depicts, she may well use a very specific keyword instead of a generic category keyword. By naming the specific object or event, information collection through InfoAlbum may give more relevant hits and better $Precision_{CR}$ results than reported here.

During testing we found that content relevant tags are highly useful in that they in most cases provide the user with a name of the depicted object or event. Tags are used as input to Google search, to automatically construct focused queries. Our tests show that content relevant tags can significantly improve the average precision scores for web page retrieval. Thus, by improving the automatic tag collection algorithm, and thereby retrieving content relevant tags for more images, information retrieval through Google search can also be improved.

Information collection is done in an iterative manner, in that the result of one information search can be used as input to a new search. We are, for instance, using gps coordinates, category, synonyms and date/time to collect tags from Flickr. The tags are later used as input to Google searches. To avoid receiving irrelevant information through iterative information collection, it is required that the automatically collected search parameters (e.g. tags) with a high probability are relevant to the image. In the current implementation of InfoAlbum, we seek to avoid irrelevant tags by using a threshold of 0.3 for tag selection, meaning that a tag must appear in at least 30% of the relevant images in order to be used in InfoAlbum.

When searching Google and Wikipedia for information, the outcome, as seen in InfoAlbum, relies on the information retrieval performance of the search engine and the available information in Wikipedia. However, the design of InfoAlbum heavily effects the outcome of the searches, in that InfoAlbum should provide the best possible search parameters to cover the information need. Currently we have identified category, location name, tags and (for event images) time period as the most useful image metadata for information collection from Google.

In the current version of InfoAlbum we have mostly used general information sources that can provide a wide variety of information. Weather Underground is an exception, in that this source is designed for providing world wide weather information. We believe that more specialized information sources targeted to specific categories of information, for example football, concerts, bridges or castles, may give better results compared to our tests. When having specialized information sources available, image category may be used for automatically selecting the sources that are of specific interest to an image.

5 Conclusions

We have presented an approach for automatically collecting image relevant information, based on the image metadata *gps coordinates*, *date/time* and *image category*, from a variety of information sources on the Internet. To test the approach, we have implemented InfoAlbum, a prototype system where the objective is to provide the user with information about the content of an image and the location of image capture. The system collects and presents information such as location names, tags, temperature and weather condition at image capture time, placement on map, geographically nearby images, Wikipedia articles, and web pages.

In this paper we evaluated the ability to collect image relevant information and calculated precision scores for Wikipedia articles, web pages and automatically collected tags. For both object and event images we found that Wikipedia articles was a very good source of information, where InfoAlbum collected many content relevant articles. We found that automatically collected tags were very useful as input parameter when collecting relevant web pages, as they in many cases contributed to focus the information search.

InfoAlbum demonstrates the ability to collect relevant web pages based on the selected image metadata. However, the precision scores showed mixed results, with very good precision scores for some images and not so good scores for others. To further improve the system we believe that better precision can be achieved by choosing specialized information sources targeted to specific categories of information. We also seek to improve the algorithm for tag collection so that more images receive a content relevant tag, which may subsequently improve precision of for instance information searches through Google. Based on the good precision scores for geotagged Wikipedia articles, we will also investigate using information from these articles as basis for new information searches.

6 Acknowledgments

The authors appreciate support from the Norwegian Research Council, NFR, through project nr. 176858, Context-aware Image Management, CAIM.

References

[1] Ahern, S., *et al*, ZoneTag: Designing Context-Aware Mobile Media Capture to Increase Participation. In *Proceeding of the Pervasive Image Capture and Sharing Workshop (PICS) 2006*, Orange County, California, Sept. 2006.

[2] Bizer, C., Heath, T. and Berners-Lee, T., Linked data–the story so far. *International Journal On Semantic Web and Information Systems*, IGI Publishing, 2009.

[3] Evertsen, M.H., *Automatic Image Tagging based on Context Information.* Thesis in Computer Science, University of Tromsø, Norway, June 2010.

[4] Jakobsen, B., *Collecting relevant images context information.* Thesis in Computer Science, University of Tromsø, Norway, January 2010.

[5] Kucuktunc, O., *et al*, Tag Suggestr: Automatic Photo Tag Expansion Using Visual Information for Photo Sharing Websites. In *SAMT '08: Proceedings of*

the 3rd International Conference on Semantic and Digital Media Technologies, Berlin, Heidelberg, 2008. Springer-Verlag.

[6] Moxley, E., Kleban, J.and Manjunath, B., SpiritTagger: A Geo-Aware Tag Suggestion Tool Mined from Flickr. In *MIR '08: Proceeding of the 1st ACM international conference on Multimedia information retrieval*, Vancouver, Canada, Oct 2008.

[7] Popescu, A. and Moëllic, P.A., MonuAnno: Automatic Annotation of Georeferenced Landmarks Images. In *CIVR '09: Proceeding of the ACM International Conference on Image and Video Retrieval*, Island of Santorini, Greece, July 2009.

[8] Proß, B., Schöning, J. and Krüger, A., iPiccer: Automatically retrieving and inferring tagged location information from web repositories. In *Proceedings of the 11th International Conference on Human-Computer Interaction with Mobile Devices and Services*, Bonn, Germany, 2009.

[9] Quack, T., Leibe, B. and Gool, L., World-scale mining of objects and events from community photo collections. In *CIVR '08: Proceedings of the 2008 international conference on Content-based image and video retrieval*, Ontario, Canada, July 2008.

[10] Serdyukov, P., Murdock, V. and Zwol, R., Placing Flickr Photos on a Map. In *SIGIR '09: Proceedings of the 32nd international ACM SIGIR conference on Research and development in information retrieval*, Boston, Massachusetts, July 2009.

[11] Van Gool, *et al*, Mining from large image sets. In *Proceeding of the ACM International Conference on Image and Video Retrieval*, CIVR '09, pages 10:1–10:8, New York, NY, USA, 2009. ACM.

[12] Yeh, T., Lee, J.J. and Darrell, T., Photo-based question answering. In *Proceeding of the 16th ACM international conference on Multimedia*, MM '08, New York, NY, USA, 2008. ACM.