

The InfoAlbum

Image centric information collection

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

Børge Jakobsen
Computer Science Department
University of Tromsø
Tromsø, Norway
borge.jakobsen@gmail.com

ABSTRACT

This paper presents a prototype of an image centric information album, where the goal is to automatically provide the user with information about i) the object or event depicted in an image, and ii) the surrounding where the image was taken. The system, called InfoAlbum, aims at improving the image viewing experience by presenting supplementary information such as location names, tags, temperature at image capture time, 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 coordinates*, *date/time* of image capture and a *category* keyword provided by the user. Collected information is presented to the user, and some is also stored in the EXIF header of the image and can later be used during image retrieval.

Categories and Subject Descriptors

H.3 [Information Storage and Retrieval]: Information Search and Retrieval, Systems and Software, Online Information Services

General Terms

Algorithms, Design, Experiment, Theory

Keywords

Image centric information collection, Image metadata, Information retrieval

1. INTRODUCTION

As the number of digital images grows, the challenging task of managing them so that relevant images can be found and displayed in an effective manner, gets increasingly important. With the huge number of available images, supplementary information (such as tags and annotations) is important, not only for answering high level, semantic queries

for images, but also for providing users with information about what an image depicts.

The current trend is that digital photos are displayed together with more and more additional information. Digital albums, such as Google Picasa¹ and iPhoto², display images on a map (if gps coordinates are available) and provide tools for identifying faces, adding tags and editing metadata. Web album and photo sharing sites, such as Flickr³ and Panoramio⁴, 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.

With the multitude of information available from numerous sources on Internet, it is currently possible to automatically provide the user with supplementary information that can enhance the image viewing experience, and also support image retrieval.

Consider the following scenario. A tourist comes home with a lot of images taken in different locations. She usually wants to transfer the images to a computer for storage and viewing, and when viewing the images she wants access to additional information related to the image she is currently looking at. She finds the additional information useful, both since it may be difficult in retrospect to remember exactly what the image depicted, and also since she wants to know more about what she saw or experienced. She may for instance not know the name of the artifact (the statue, building or church) or have forgotten the name of the village, or she may like some historical information and current facts about the depicted artifact or event.

To support the needs of our tourist, we have developed a prototype for image centric information collection, called InfoAlbum. 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. The system collects automatically a variety of information from sources on the Internet based on the image metadata {*gps_location*, *date/time*, *category*}. The information is presented to the user, but can also be stored in the EXIF header of the image and later used during image retrieval.

Location and date for image capture are used for finding location names, weather information, Wikipedia arti-

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

WIMS '11, May 25-27, Sogndal, Norway

Copyright 2011 ACM 978-1-4503-0148-0/11/05 ...\$10.00.

¹<http://picasa.google.com/>

²<http://www.apple.com/ilife/iphoto/>

³<http://www.flickr.com/>

⁴<http://www.panoramio.com/>

cles, placement on map, and geographically nearby images. By allowing users to specify a category for each image, we provide a basis for collecting a wider range of information. A category is typically a keyword representing a general description of the image content, for example "tower", "church", "bridge", "concert" or "festival". Using category proves useful for detecting information that is relevant to image content, for instance through general information search using Google, Bing and Yahoo.

This paper describes the InfoAlbum system, the automatic extraction of information and the evaluation of system performance with respect to relevance of the collected information. We especially concentrate on the information collected based on category information.

2. AN IMAGE CENTRIC INFORMATION COLLECTOR

2.1 InfoAlbum

This section describes the InfoAlbum prototype and how it automatically collects relevant information to an image from sources on the Internet. The InfoAlbum architecture is depicted in Figure 1, where we see the three components of the system; the Interface, 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 the collected information after the information is retrieved by CIC.

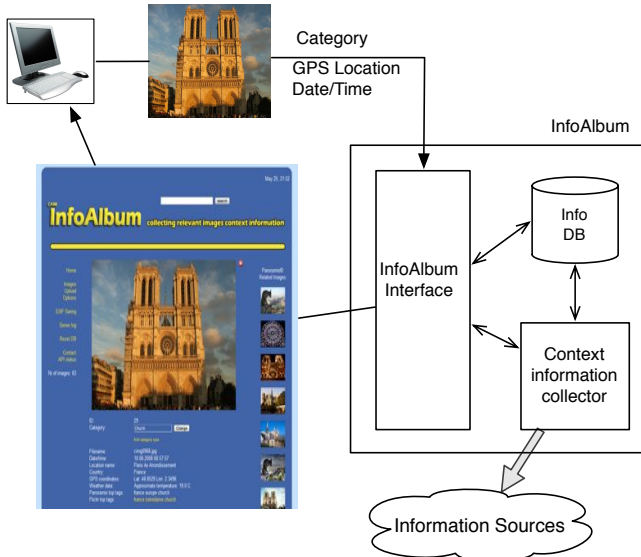


Figure 1: The InfoAlbum system

The image represents a query to the InfoAlbum system, from which the user expects 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 a 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 EXIF header, while category is

provided by the user. After the category is once determined, it will be stored in the EXIF header from where it can later be extracted. The system also allows the category to be changed, in case there are more than one object of interest in the image and the user wants to change focus.

The system collects information from a number of sources, $S = \{S_1, \dots, S_n\}$. For each source S_j , where $1 \leq j \leq n$, the information collection can be described as $S_j(P_j) = I_j$, where P_j is a set of parameters used in the information search and $I_j = \{i_1, \dots, i_l\}$ is the set of collected information. We further have that $P_j = \{p_1, \dots, p_m\}$, where $p_k \in IM$ or $p_k \in I_l$ for $1 \leq k \leq m$. This means that input parameter p_k can take its value from either IM (i.e. a date, time, coordinates or category) or information collected from a different source.

In Figure 2 all information sources used in the current version of the InfoAlbum system are listed. For each source, the figure shows collected information together with the image metadata needed to do the extraction. The image together with collected information is stored in the Info database.

Source (S)	Collected information (I)	Parameters (P)
Flickr	location names, images, tags	gps coordinates, date
Panoramio	images, tags	gps coordinates
Airport Code database	Airport codes	gps coordinates
Weather Underground	temperature at image capture time	airport code, date/time
Geonames	geo-tagged Wikipedia articles	gps coordinates
Google	map, position on the map	gps coordinates
Google, Yahoo, Bing	Internet pages	category, loc. name, date
Wikipedia	Wikipedia articles	category, loc. name, date

Figure 2: Information sources and collected information

2.2 Image metadata

We initially use three types of image metadata in the information collection process. This is 1) date and time of image capture, 2) gps coordinates (giving the position of the camera at image capture), and 3) image category.

Date and time of image capture is normally recorded by every digital camera, and is found in the header of the image file, usually stored in the EXIF header.

GPS coordinates can be obtained either by using a camera that automatically captures and stores coordinates, an external gps tracking device, or by manually adding latitude and longitude values (for instance by dropping the image on Google Map). Use of gps coordinates for images has gained popularity, and the number of devices that automatically store gps coordinates on an image is growing, including mobile phones, consumer cameras and gps navigation devices.

Category can be defined as "any of several fundamental and distinct classes to which entities or concepts belong"⁵. An *image category* is in our system a keyword that represents a general description of the content of the image, and is used for narrowing down the information search and detecting information that is relevant to the image content. Categories are user specified, and the InfoAlbum system al-

⁵<http://www.merriam-webster.com/>

allows users to define and store categories, so that they are easily available for image categorization.

A category always belongs to a *type*, that reflects the type of content seen in the image. The current implementation of InfoAlbum supports two types of images, which are images depicting i) objects and ii) events. Categories of type object can for instance be "tower", "church", "monument", "mountain" and "river", while event categories can be "concert", "football match" or "festival". The reason for introducing the concept *type* is to be able to automatically determine how to collect information from various sources. Images of objects and events are handled differently with respect to information collection in that date/time is an important parameter when searching the Internet for information about an event, while it (normally) is not used for object images.

Generally, a set of image types, $T = \{t_1, \dots, t_n\}$, may be used. In InfoAlbum the set currently includes two types, Object and Event, i.e. $T = \{Object, Event\}$. A set C will at any time consist of the categories defined by users of InfoAlbum, and by defining new categories the users will add elements to the set. For a set of categories $C = \{c_1, \dots, c_n\}$, each element c_i must belong to exactly one image type in T , i.e. $Type(c_i) \in T$ for $1 \leq i \leq n$.

2.3 Using Image Category and Type

Image category is used for different purposes in InfoAlbum; i) as a keyword in information retrieval searches on Google, Yahoo, Bing and Wikipedia, and ii) for ranking information collected by InfoAlbum.

The type of an image determines the search parameters used when collecting information from the different sources. For an image of an event, the objective is to include information about the specific happening at the given date and location. For instance, for an image of a football match, we would like the InfoAlbum to present information such as news articles describing the match, names of the teams, names of players, number of goals, more images of the same match, and more. To obtain this, information sources are searched based on category, location and date/time information.

Given an image O with category c and a set of search parameters P . If $Type(c) = Event$ then $P = \{Location, Category, Date/Time\}$. To find information related to an image of an object, such as a bridge or a church, sources are searched based on category and location information; i.e. if $Type(c) = Object$ then $P = \{Location, Category\}$.

Some of the collected information, such as web pages and Wikipedia articles, are ranked in InfoAlbum before it is presented to users. Ranking is generally done in order to sort matching documents according to their relevance to a given search query. Since category is used as a parameter when searching information sources, it also has a role when ranking collected information. For example, InfoAlbum ranks collected web pages by using a term frequency method, where category, together with location name, represent terms.

When defining an image category, the user should make sure that the category is reusable, i.e. applicable to a class of images, and that the category is useful as a query for information collection on the Internet. In our testing we have chosen categories such as "bridge", "tower", "church", "football" and "concert".

A user may choose more specific keywords, by for instance

naming the specific object or event (such as "Notre Dame" or "U2 concert"). This approach may give more relevant hits on Internet search. However, the user may not know the specific name of an object or event. Also, this may violate the requirement for having reusable categories suited for a class of images. Our main objective was to test performance of the system for generic categories, so this option was not chosen for any of the images. Tests of InfoAlbum demonstrate that it is possible to find relevant information about both objects and events using generic categories.

2.4 Collected information

This section describes all the information collected in the current version of InfoAlbum.

Location names. To find location names based on gps coordinates, we extract information from Flickr through the Flickr API. With gps coordinates as input, the function `flickr.places.findByLatLon` returns a list of location names sorted by "Country, Region, County, Locality, and Neighbourhood", spread over 16 accuracy levels. The InfoAlbum system displays level 1, which gives country name for each image, and level 10, which in most cases provides the name of the city where the image was taken. The level 10 name is also used as input when searching Google, Yahoo, Bing and Wikipedia for web pages and Wikipedia articles.

Weather information. Information about weather conditions at image capture time, can be grabbed from the Weather Underground⁶ service. Weather Underground provides historical information about for instance temperature, moisture, barometric pressure and wind speed. Currently, InfoAlbum collects and displays temperature at the time and approximate location where the image was taken.

To grab temperature information from Weather Underground we need date/time and a location name as input. However, since location names in general are not unique, we first link the gps location to the nearest airport and its unique ICAO-code. All the worlds airport codes together with gps latitude and longitude values are available in a list created by Greg Thompson⁷ and is used in our system to determine the nearest airport.

Once the code of the nearest airport is determined, the Weather Underground web page is parsed for the required airport, at the date and hour for image capture, grabs the nearest hourly temperature measure, stores the value in the InfoAlbum database, and displays it to the user.

Geographic position. Geographic position for an image is shown by displaying Google Maps with pinpointed location. The Google Maps API let anyone embed Google Maps in their web page, and provides a number of utilities for manipulating maps and adding content to it.

Nearby images. Nearby images can be fetched from a number of collections. We have chosen two sources, Flickr and Panoramio, for finding images from the same area as the current image. For both Flickr and Panoramio we give gps coordinates and a radius of 200 meters as input, and we display a small number of 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

⁶<http://www.wunderground.com/>

⁷<http://www.rap.ucar.edu/weather/surface/stations.txt>

Flickr search. This makes it possible for 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.

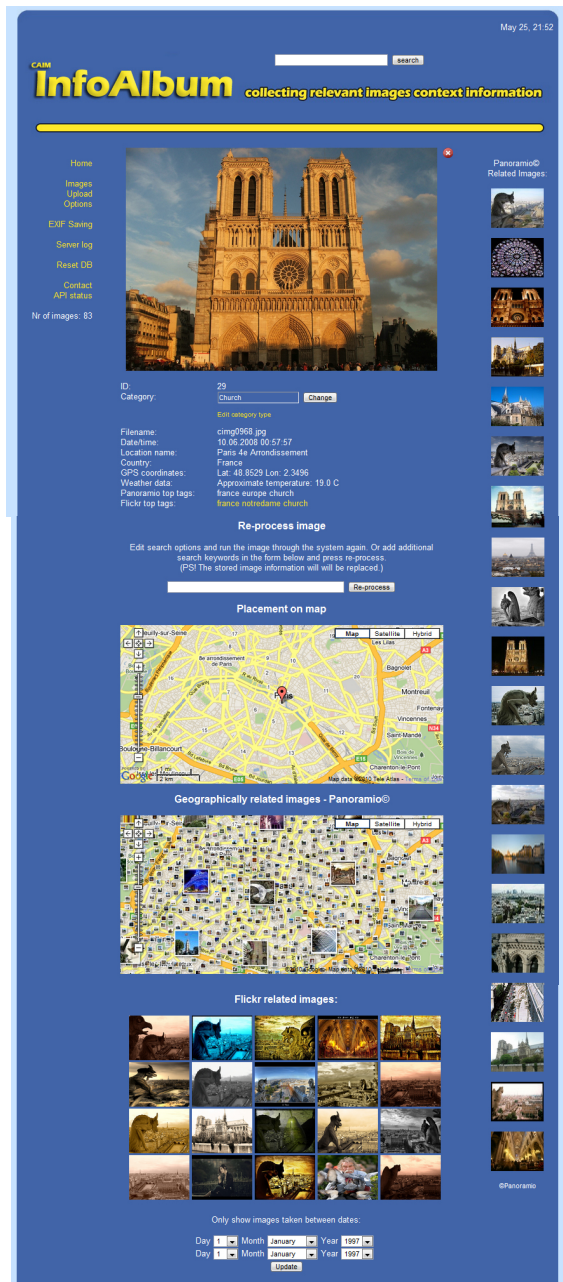


Figure 3: Screenshot Part 1

Geotagged articles. Geonames⁸ keeps a database of geotagged Wikipedia articles that are pinned to locations by gps coordinates. We are using the functions FindNearbyWikipedia and WikipediaSearch in the Geonames API to find articles that are relevant to the location where the image was taken. Gps coordinates and a radius, defining the area of interest, are given as input to the function. As out-

⁸<http://www.geonames.org/>

put we obtain references to Wikipedia articles, and for each article the distance to the given gps position.

Only a small number of articles were obtained through Geonames. In our tests, we used a radius of 1 kilometer and received not more than 5 articles per image. All articles were relevant with respect to location. As we also are interested in identifying articles that are relevant to image content, we rank the articles with respect to category. Evaluation of this ranking is found in Section 3.



Figure 4: Screenshot Part 2

Web pages and Wikipedia articles. Finally we use category, location name and date information to search Google, Yahoo, Bing and Wikipedia for web pages and articles. For each image O , the associated set of search parameters, P , is used to formulate an Internet search through each of the search engines, Google, Yahoo and Bing. Additionally, Wikipedia is also searched based on parameters in P . Wikipedia and the three search engines are in this context considered as four different information sources.

After executing the same search query in each source, we take the highest ranked hits from each source and merge them in InfoAlbum while removing duplicate hits. In the default setting of InfoAlbum, we include the 10 top ranked results from each source. However, this number can be changed by the user.

While Wikipedia articles received through Geonames are

obtained based on gps coordinates only (see Geotagged articles), we are now searching Wikipedia directly using keywords, where both location, date and category may be represented. This makes it possible to search among articles that are not geotagged, and to find articles based on category and date, not only location.

The main objective of using these Internet searches in InfoAlbum, is to collect information that may be relevant to the content of an image. To enable this, image category is used as input to the search, and is of vital importance as it represents information about image content provided by the user. The relevance of collected web pages and articles, with respect to both location and image content, has been evaluated and is reported on in Section 3.

Figure 3 and 4 show screenshots from InfoAlbum, showing how collected information is presented to the user. At the top of Figure 3 we see the image in focus (here Notre Dame in Paris). Below the image we find textual information about temperature, tags, location names and other information. To the right is a row of images collected from Panoramio, while at the bottom we see images collected from Flickr. Two maps are shown; one which pinpoint the gps position of the image, the other shows Panoramio images on a map. Figure 4 is a continuation of the first screenshot, and shows how geotagged Wikipedia articles and web pages are displayed to the user. By clicking on the title of each entry, the selected web page or Wikipedia article is opened in a new window.

2.5 Related work

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

Displaying images on a map is a popular approach used in systems such as Flickr, Panoramio and Google Picasa, where gps coordinates are used for finding the right position on the map. In [9] it is described 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, which describe different approaches to collecting tags from previously tagged images. In systems such as MonuAnno [6], ZoneTag⁹ [1] and iPicca [7] images are given location relevant tags by collecting tags from existing images from Flickr and Panoramio. Relevant images are selected based on gps coordinates or other location information. Zonetag also suggests tags based on past tags from the user, the user's social network, and names of nearby real world entities. SpiritTagger [4] 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 [3] expands user provided tags of an image by incorporating tags from other images which are visually similar.

The work [8] and [10] describe how to collect and cluster images of landmarks, which are subsequently automatically annotated and enriched with Wikipedia information. First geotagged photos are retrieved, from community photo

collections (such as Flickr), and clustered. The resulting clusters are analyzed to find frequently used text labels (for example tags). The selected text labels are used to automatically assign each cluster to a corresponding Wikipedia article.

Google Goggles¹⁰ and Nokia Point & Find¹¹ are mobile applications that let us use pictures taken with a mobile phone to search the web for information. Image recognition is used to find information about for instance books and DVDs, landmarks, logos, contact info, artwork, businesses, products, and for barcode scanning. The information search is based on image recognition techniques.

Some systems provide augmented reality by using the camera of a mobile device as a "looking glass". Information about the object or place seen through the camera is retrieved and displayed to the user as an overlay on the camera screen view. An example is Wikitude World Browser¹² which displays information about the user's surroundings. The application calculates the user's current position and accesses the Wikitude data set to provide geographic information, history, and contact details of points of interest. Another example is Layar Reality Browser¹³ that provides overlay information about for instance location of ATMs, houses for sale and restaurants. The overlay information is in both systems collected from a supporting server based on gps location information.

The work [5] describes how to automatically create a multilingual geographical gazetteer by mining heterogeneous sources of geo-referenced metadata. The work is based on detecting explicit geographic concepts in place names. Place names are extracted, localized, categorized and ranked by merging information from Flickr, Panoramio, Wikipedia and AllTheWeb.

Linked Data¹⁴ [2] and the Linking Open Data Community Project¹⁵ 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 goal is to present users with a variety of information relevant to an image. This includes textual information (such as Wikipedia articles and web sites), other images, tags, weather information and maps. This information is related to the image through mapping based on location, date and category information. As apposed to much of the related work, we are not using the image itself in the query, but rather rely on image metadata, either extracted directly from the exif header (such as date/time and gps location) or derived from external information sources (such as location names obtained from Flickr).

Our use of category information is novel compared to other approaches described here. The category, giving a general description of the image content, is an important parameter when retrieving information through general search

¹⁰<http://www.google.com/mobile/goggles/>

¹¹<http://europe.nokia.com/services-and-apps/nokia-point-and-find>

¹²<http://www.wikitude.org/>

¹³<http://www.layar.com/>

¹⁴<http://linkeddata.org/>

¹⁵<http://esw.w3.org/SweoIG/TaskForces/CommunityProjects/LinkingOpenData>

⁹<http://zonetag.research.yahoo.com/>

engines, while the type to which the category belongs, is important for determining how information retrieval should be done to gain the best possible outcome.

Our work is related to image-based question answering [11] in the sense that we assume an implicit query: "Give me information related to this image". The relationship between image and information can be with respect to i) location, and ii) content of image. To answer the query with respect to location, we can search the internet based on gps coordinates with a specific bounding box or radius, or based on location names. Detecting information that is relevant to the content of the image is much harder, in that there is not a reliable way of determining the semantics of an image. However, the category is highly useful conveying what the user sees in the image and for narrowing down the search.

3. EVALUATION

The InfoAlbum prototype implementation has been tested with 80 images, where 50 depicted an object, while 30 images depicted some event. These 80 images represent 80 different queries to the InfoAlbum system. For each image, InfoAlbum collected information, from sources on the Internet, as described in the previous section.

We chose test images where the content varied from very famous objects, such as the Eiffel Tower and Notre Dame, to less famous objects located at much smaller cities. For events we chose some famous events such as Carnival in Rio and Oktoberfest in Munich, via U2 concert in Barcelona and a football match with Manchester United, to a small football match with only regional interest. The categories that were used was "architecture", "bridge", "church", "monument" and "tower" for object images, and "concert", "festival" and "football" for event images.

In this section we evaluate the relevance of web pages and Wikipedia articles, and evaluate the effect of using category in information retrieval and ranking.

3.1 Content and Location relevancy

As we have a dual objective in InfoAlbum of collecting information that is relevant to both i) content of the image and ii) location where the image was taken, we also need to evaluate performance of the system with respect to both objectives. We therefore introduce two different relevancy concepts; *Content Relevance* and *Location Relevance*.

In information retrieval, relevance denotes how well a retrieved document or set of documents meets the information need of the user. When distinguishing between Content and Location Relevance, we explicitly focus on specific needs when considering relevance. Thus, a *Content Relevant* document is relevant with respect to what the user sees in the image as indicated through the category keyword, while a *Location Relevant* document is relevant to the location where the image was taken. A document describing the place or neighborhood where the image is taken is Location Relevant. The same is a document describing a nearby object not seen in the image or an object depicted in the image but not referred to by the category keyword.

For example, for an image with category "bridge", a document describing or mentioning the specific bridge is Content Relevant, while a document describing a nearby building (which may also be seen in the image), the neighborhood or city where the image is taken is Location Relevant.

In the following sections we calculate precision of retrieved

web pages and Wikipedia articles with respect to both Content Relevance (CR) and Location Relevance (LR). We will use two precision measures, which are defined as follows.

$Precision_{CR}$ 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.

Assume that $CRset$ represents the set of Content Relevant documents collected by InfoAlbum, $LRset$ represents the set of collected Location Relevant documents, and A represents the set of all collected documents. A document that is considered Content Relevant is not also Location Relevant, i.e. $CRset \cap LRset = \emptyset$. $Precision_{CR}$ and $Precision_{LR}$ can now be formulated as

$$Precision_{CR} = \frac{|CRset|}{|A|}$$

$$Precision_{LR} = \frac{|LRset|}{|A|}$$

When testing InfoAlbum, one of our main concerns was the ability to retrieve content relevant web pages and Wikipedia articles to images. Search parameters were therefore chosen so that the retrieved documents could include information about the object or event depicted in the image. For object images this meant that information were retrieved using $\{category, location_name\}$ as parameters. Fortunately, this kind of search provides both Content and Location Relevant documents, and both $Precision_{CR}$ and $Precision_{LR}$ were therefore calculated for object images.

For event images we also included date as parameter, with the objective of finding web pages relevant for the specific event. However, we found that this type of search is not well suited for finding location relevant information, and $Precision_{LR}$ was therefore not calculated for event images.

3.2 Ranking

Ranking of web pages and Wikipedia articles were done based on term frequency of the two search parameters; category keyword and location name. For each web page the system counts occurrences of a) the category keyword and b) the location name within the first page returned by the search engines. As both location and category information is needed in order to identify the specific object in an image, we combine the term frequency score for the two parameters before ranking the pages.

The geotagged Wikipedia articles were collected based on gps coordinates, and are all Location Relevant to the image. Ranking of these articles were done with regard to Content Relevance, and we used a term frequency measure based on the category keyword.

3.3 Web pages and Wikipedia articles

As described in Section 2.4, InfoAlbum performs, for each image, textual search in Google, Yahoo, Bing and Wikipedia, merge and rank the 10 top ranked hits from each source, while removing duplicate hits. For the 50 object images, the textual search is based on location name and category keyword, while date is used as a third parameter for the 30 event images. InfoAlbum ranked between 26 and 39 articles for each image.

To evaluate the performance of InfoAlbum, we manually inspected every web page and article collected by InfoAl-

bum, 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 the average precision is presented in Figure 5.

	object images	event images
content relevance, average precision	16%	14%
location relevance, average precision	31%	--

Figure 5: Average precision of ranked documents

The average precision with respect to Content Relevance was 16 % for object images and 14% for event images. This means that on average every 7th web page contained information about the object or event depicted in the image. The average precision with respect to Location Relevance was 31%, meaning that approximately every third article contained information about the location where the image was taken. As described in Section 3.1, $Precision_{LR}$ has not been measured for event images.



Figure 6: Content relevance precision measurements

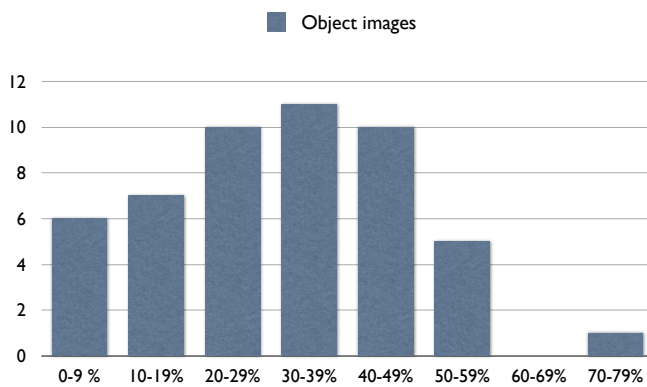


Figure 7: Location relevance precision measurements

The average precision measure does not show the difference between images with respect to amount of relevant web

pages. This difference is illustrated in figure 6 and 7, where we show the number of images for which precision is within a specific range. From Figure 6 we see that 24 object images and 16 event images have a low Content Relevance precision score (between 0 and 9), while for some of the images we received much better results. For Location Relevance we have higher average precision and a different distribution of precision values, as shown in Figure 7.

The testing shows that InfoAlbum is capable of collecting both Content and Location Relevant web pages for images. In our test, InfoAlbum retrieved at least one Content Relevant web page for 90 % of the object images, and for 82% of the event images. The system presented Location Relevant web pages for all object images.

The web pages selected from Wikipedia, Google, Yahoo and Bing, were automatically ranked as described in Section 3.2. To test this ranking we compared the average precision presented in Figure 5, with a calculation of average precision among the 10 top documents for each image. As shown in Figure 8, we found that precision scores for the 10 top ranked documents were better then the precision score including all documents.

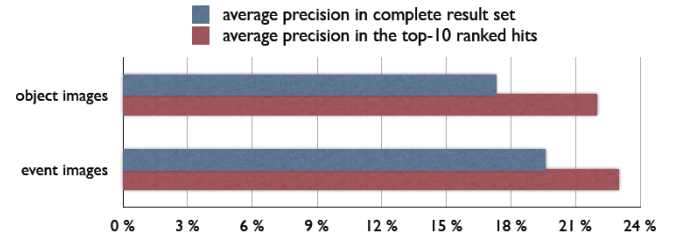


Figure 8: Effect of ranking

3.4 Geotagged articles

Through Geonames we obtained geotagged Wikipedia articles relevant to the location where the image was taken (see section 2.4). By using a radius of 1 kilometer from the image capture position, we received between 0 and 5 articles per image. These articles described either a place, a building or some other point of interest in the area.

For the 80 test images, we received a total number of 257 geotagged Wikipedia articles, that were all relevant with respect to location. For 35 images we received 5 articles each, while only 5 images did not receive any geotagged articles. The number of collected articles obviously depends on the availability of geotagged articles in the area of interest.

For 90 % of the object images we received articles that were Content Relevant to the image. This means that the article described the object depicted in the image and focused through the category.

InfoAlbum uses the category keyword to rank articles, and for 84 % of the object images (42 images) the first ranked article described the image content. This means that among the 47 object images where geotagged articles were found, 89 % had a Content Relevant article ranked first. 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. Also, by parsing the article and selecting the title, we obtain, with an equal high probability, a name for the object in focus. Category ranking of geotagged articles thus proves very useful

for selecting the content relevant article, and in fact determining the content of an object image.

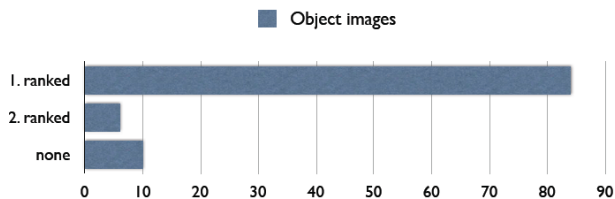


Figure 9: Ranking of directly content relevant Wikipedia articles

As Wikipedia does not normally contain articles about specific events, such as football matches or concerts, we can not expect geotagged articles obtained through Geonames to be Content Relevant to the event images. However, 4 out of 30 event images were taken at some annual festival (the Munich October festival, Carnival in Rio and Roskilde festival), and for those images InfoAlbum collected Content Relevant articles through Geonames. Also for these 4 images, category based ranking were useful, in that the Content Relevant articles were ranked first among the geotagged articles.

3.5 Discussion

When searching Google, Yahoo, Bing and Wikipedia for information, the outcome, as seen in InfoAlbum, relies on the information retrieval performance of the search engines and the information available in Wikipedia. However, the information collected by InfoAlbum is effected by using search criteria that as good as possible cover the information need. Currently we have identified category, location and (for event images) date/time as the most useful image metadata for doing information collection.

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 to collect location names, which are later used when searching Google for information related to a specific place. To avoid receiving irrelevant information through iterative information collection, it requires that the automatically collected search parameter (name in this example) with a high probability is relevant to the image. For example, the mapping between gps coordinates and location names, results in a number of names that are all relevant (facts) for the given position. A challenge is then to choose the name with the right accuracy level in order to collect more relevant information.

In InfoAlbum we distinguish between *types* that are used by the system for determining search approaches, and *categories*, where the keyword is used as one parameter in information searches. While *types* are predefined in the system, the user is allowed to freely specify the *category* for an image. The user can then not only choose the category keyword that best reflects what the user see in the image, but also use a keyword that he/she believes to be useful in an information retrieval request.

The choice of keyword language is important for the information collection. English is suited in Wikipedia search and for ranking Wikipedia articles. The local language might be best for textual search on the Internet. This is specifically true when searching for information about events. To

search for reports and news articles about a specific event, local or national news papers might be the best source, and for finding these we need a keyword in the local language. InfoAlbum do currently not support synonyms or automatic language translation in the information collection process. This may be included in a next version of InfoAlbum.

In InfoAlbum we are both concerned with i) presenting information describing the content of the image, and ii) presenting information related to the location of the image, i.e. such as a description of the town where the building stands, nearby buildings, history about the street, and so on. For example, for one of our test images depicting a church, InfoAlbum collected two Wikipedia articles. One article described the church while the second was about a nearby bridge. Both articles are relevant and presented in InfoAlbum. The article about the church is Content Relevant to the image, as it describes an object depicted in the image, while the article about the bridge is relevant as information about the surrounding area.

A third objective of InfoAlbum is to collect information to an image that can later be used for image retrieval purposes. This means that information collected by InfoAlbum is selected and stored as image metadata. Since the purpose of this metadata is to describe image content, a challenge would be to automatically select the most appropriate Content Relevant information. Taking the example above, the article about the church would be useful as a basis for image metadata (for instance by selecting terms from the text), while the article about the bridge should not be used. Currently InfoAlbum offers users the possibility to select the information types (for instance location names, tags and terms selected from the top ranked Wikipedia article) that will automatically be stored as metadata to images.

Testing of InfoAlbum shows that it is possible to collect information about publicly known objects and events. We found for instance name and description of objects, such as churches, towers, monuments and bridges. For events we were able to identify for example which concert, football match or festival it was, name of band and band members, the tour schedule, names of football teams and players and the result of the match. However, the current problem when collecting web pages is that precision is too low. The user need to browse through too many non relevant web pages in order to find the relevant information. We believe that a higher precision is possible by carefully selecting the most relevant information sources for each image.

In the current version of InfoAlbum we have mostly used general information sources that can provide a wide variety of information. The search engines, Google, Bing and Yahoo, are in particular using sources of all sorts 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 selecting the sources that are of specific interest to an image.

4. CONCLUSIONS

We have described InfoAlbum, a novel prototype for image centric information collection that, based on the image metadata *gps coordinates*, *date/time* and *image category*, au-

tomatically collects a variety of information from sources on the Internet. The information is presented to the user as supplementary information together with the image. The objective of InfoAlbum is to provide the user with information about the content of an image and the location of image capture. The system collects and presents to the user information such as location names, tags, temperature at image capture time, placement on map, geographically nearby images, Wikipedia articles, and web pages collected through Google, Yahoo and Bing.

Testing of InfoAlbum shows that it is possible to collect information about publicly known objects and events. In this paper we specifically evaluated the retrieval performance with respect to Wikipedia articles and web pages. For object images, we found that Wikipedia articles provide a very good source of information. This is particularly true for geotagged articles for which we had a very good precision score. The web pages presented in InfoAlbum also included some useful documents with relevant information about both content of the image and the location where the image was taken.

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 plan to use synonyms and language translation in order to detect relevant articles and web pages where the specific category keyword is not explicitly used, and we are working on an improved approach for automatically collecting tags from photo sharing sites. Based on the good precision scores for geotagged Wikipedia articles, we will also investigate using information from these articles as basis for new information retrieval searches.

5. ACKNOWLEDGMENTS

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

6. REFERENCES

- [1] S. Ahern, M. Davis, D. Eckles, S. King, M. Naaman, R. Nair, M. Spasojevic, and J. Yang. Zonetag: Designing context-aware mobile media capture to increase participation. In *Proceeding of PICS 2006 in conjunction with Ubicomp 2006*, 2006.
- [2] C. Bizer, T. Heath, and T. Berners-Lee. Linked data—the story so far. *International Journal On Semantic Web and Information Systems*, 2009.
- [3] O. Kucuktunc, S. G. Sevil, A. B. Tosun, H. Zitouni, P. Duygulu, and F. Can. 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*, pages 61–73, Berlin, Heidelberg, 2008. Springer-Verlag.
- [4] E. Moxley, J. Kleban, and B. Manjunath. Spirittagger: a geo-aware tag suggestion tool mined from flickr. *MIR '08: Proceeding of the 1st ACM international conference on Multimedia information retrieval*, Oct 2008.
- [5] A. Popescu, G. Grefenstette, and H. Bouamor. Mining a multilingual geographical gazetteer from the web. *Web Intelligence and Intelligent Agent Technologies*, 2009. *WI-IAT '09. IEEE/WIC/ACM International Joint Conferences on*, 1:58–65, 2009.
- [6] A. Popescu and P.-A. Moëllic. Monuanno: automatic annotation of georeferenced landmarks images. *CIVR '09: Proceeding of the ACM International Conference on Image and Video Retrieval*, Jul 2009.
- [7] B. Proß, J. Schöning, and A. Krüger. ipiccer: automatically retrieving and inferring tagged location information from web repositories. *Proceedings of the 11th International Conference on Human-Computer Interaction with Mobile Devices and Services*, page 69, 2009.
- [8] T. Quack, B. Leibe, and L. Gool. World-scale mining of objects and events from community photo collections. *CIVR '08: Proceedings of the 2008 international conference on Content-based image and video retrieval*, Jul 2008.
- [9] P. Serdyukov, V. Murdock, and R. Zwol. Placing flickr photos on a map. *SIGIR '09: Proceedings of the 32nd international ACM SIGIR conference on Research and development in information retrieval*, Jul 2009.
- [10] L. Van Gool, M. D. Breitenstein, S. Gammeter, H. Grabner, and T. Quack. 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.
- [11] T. Yeh, J. J. Lee, and T. Darrell. Photo-based question answering. In *Proceeding of the 16th ACM international conference on Multimedia*, MM '08, pages 389–398, New York, NY, USA, 2008. ACM.