International Journal of Computer Science and Applications
© Technomathematics Research Foundation
Vol. 9 No. 1, pp. 50 - 69, 2012

UNCOVER WHAT YOU SEE IN YOUR IMAGES
The InfoAlbum approach

RANDI KARLSEN
Computer Science Department
University of Tromsø, Tromsø, Norway
randi@cs.uit.no

BØRGE JAKOBSEN
Computer Science Department
University of Tromsø, Tromsø, Norway

ROGER BRUUN ASP HANSEN
Computer Science Department
University of Tromsø, Tromsø, Norway
roger.b.hansen@uit.no

This paper presents InfoAlbum, a novel prototype for image centric information collection, where the goal is to automatically 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 system aims at improving the image viewing experience by presenting supplementary information such as location names, tags, weather condition 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 latitude/longitude values*, *date/time* of image capture and a *category* keyword provided by the user. Collected information is presented to the user, and can also be stored and later used during image retrieval.

Keywords: Image centric information collection, Image metadata, Information retrieval, Image information album

1. Introduction

The huge amount of digital images has lead to new ways of using and sharing visual information. Managing images so that they can be found and displayed in an efficient manner, is a challenging and important task. 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[a] and iPhoto[b], 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[c] and Panoramio[d], 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 on the Internet, it is currently possible to automatically provide the user with supplementary information that can enhance the image viewing experience. Information can also be stored and later used to support image retrieval. We have developed a prototype for image centric information collection, called InfoAlbum, where the main objective 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 InfoAlbum system automatically collects a variety of 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 finding location names, weather information, Wikipedia articles, placement on map, and geographically nearby images. By allowing users to specify a category for each image, we provide a basis for collecting more information, for instance through Google search. A category is typically a keyword representing a general description of the image content, for example ”tower”, ”church”, ”bridge”, ”concert” or ”festival”. Category is used for focusing requests for information and proves useful for collecting information that is relevant to the image content.

To our best knowledge, information collection in InfoAlbum is unique in the way *category, location* and *date/time* metadata is exploited to enhance images with additional information. Moreover, we are not aware of any other systems that automatically provide image relevant information from such a variety of information sources.

This paper describes the InfoAlbum 2.0 system, the automatic extraction of information and the evaluation of system performance with respect to relevance of the collected information. In [Karlsen (2011)] we reported on our first implementation of the InfoAlbum system. In this paper we describe InfoAlbum 2.0, in which both functionality and system performance is improved.

2. InfoAlbum

2.1. *Scenario*

Consider a tourist coming home with a lot of images taken in different locations. She transfers the images to a computer for storage and viewing, and when viewing

[a] http://picasa.google.com/
[b] http://www.apple.com/ilife/iphoto/
[c] http://www.flickr.com/
[d] http://www.panoramio.com/

the images she wants access to additional information related to the image she is currently looking at. The additional information is useful, 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 object in focus (the statue, building or church) or may have forgotten the name of the village. She may also like some historical information and current facts about the depicted object or event.

2.2. *An image centric information collector*

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.

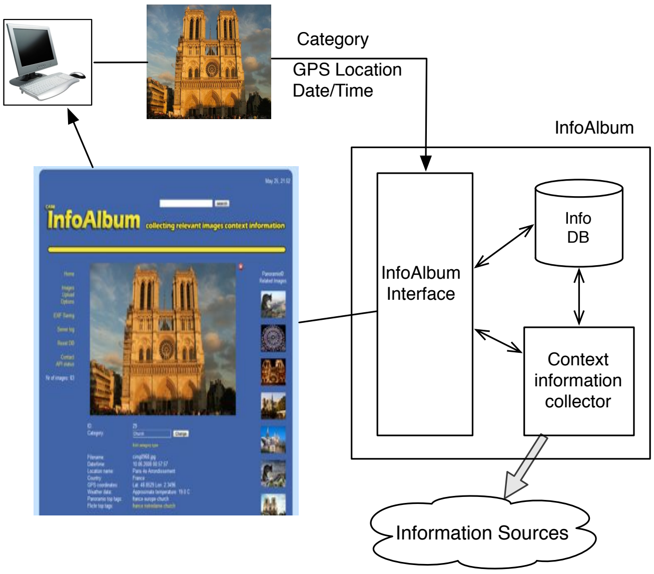

Category
GPS Location
Date/Time
InfoAlbum
InfoAlbum collecting relevant images context information
InfoAlbum Interface
Info DB
Context information collector
Information Sources

Fig. 1. The InfoAlbum system

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.

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 format. 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 Maps). Using gps coordinates for images has gained popularity, and the number of devices that automatically store latitude and longitude values on the image file is growing, including mobile phones, consumer cameras and gps navigation devices.

The system collects information from a set of sources, $S = \{S_1, \ldots, S_n\}$. For each source S_j, where $1 \leq j \leq n$, information collection can be described as $S_j(P_j) = I_j$, where P_j represents a set of parameters used in the information search and $I_j = \{i_1, \ldots, i_l\}$ represents the set of collected information.

In Table 1 all information sources used in the InfoAlbum 2.0 system are listed. For each source, the table shows collected information together with the parameters needed to do the information extraction.

Table 1. Information sources and collected information

Sources (S)	**Collected information** (I)	**Parameters (P)**
Flickr	location names	gps(lat,lon)
	nearby images	gps(lat,lon), radius
	tags	category, synonyms, date, gps(lat,lon), radius
Google	map, position on map	gps(lat,lon)
	Web pages	category, location name, tags, date/year
Panoramio	nearby images	gps(lat,lon), radius
Weather Underground	temperature, weather condition at image capture time	gps(lat,lon), date/time
Wikipedia	Wikipedia articles	location names
Wikipedia via GeoNames	geo-tagged Wikipedia articles	gps(lat,lon), radius
WordNet	synonyms	category

2.3. *Image Category*

Category can be defined as "any of several fundamental and distinct classes to which entities or concepts belong"[e]. An *image category* is in our system a keyword that represents a general description of the image content, and is used for narrowing down the information search and for detecting information that is relevant to the image content. Categories are user specified, and the InfoAlbum system allows users to define and store categories, so that they are easily available for image categorization.

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

The system also allows the category of an image to be changed, in case there are more then one object of interest in the image and the user wants to change focus.

A category always belongs to a *type*, that reflects the type of content seen in the image. Generally, a set of image types, $T = \{t_1, \ldots, t_n\}$, is defined by the system. In InfoAlbum the currently set is $T = \{Object, ShortEvent, LongEvent, Place, Personal\}$. A set C will at any time include the categories defined by users of InfoAlbum. When defining a new category, the user must determine a type $t \in T$ to which the category belongs.

Categories of type *object* can for instance be "tower", "church", "statue", or "river". *Short event* describes an event that lasts 0-2 days, and may cover "concert", "football match" or "festival". A *long event* lasts more then 2 days, and may for instance include "tournament" or "carneval". *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 you can not expect to find any public information.

In InfoAlbum, image *type* is used to automatically determine how information is collected. Images of objects and events are for example handled differently in that *date/time* is important when searching for information about an event, while it is not used for object images. For personal images, location names, weather information and nearby images are collected, but since public information is not likely to be found, Wikipedia and Google are not searched.

2.4. *Collected information*

This section describes the information collected in InfoAlbum 2.0.

Location names. Based on gps latitude and longitude values, the system extracts location names from Flickr through the Flickr API. The function flickr.places.findByLatLon returns a list of location names on the form [neighborhood, locality, county, region, country]. Among those, the system choses country, county and locality (which in most cases equals a city name) as location names for the image. For images taken outside of cities, locality may not be returned. In that case, only county and country names are used.

Weather information. Information about weather conditions at image capture time is collected from Weather Underground[f], 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 and ID of the weather station. Weather Underground allows retrieval of an HTML document containing information from the given date. This document is parsed to find the weather conditions closest to image capture time. The current approach for collecting weather information is implemented based on the work in [Sundby (2011)].

The information provided by Weather Underground includes temperature, hu-

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

midity, 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 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 query image. For both Flickr and Panoramio we use gps coordinates and a radius of 200 meters as input, and we display 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. InfoAlbum provides automatic image tagging by collecting tags from relevant images on Flickr. To retrieve relevant tags, the flickr.photo.search function in the Flickr API is used, with category, synonyms of the category, longitude/latitude and radius as parameter. A time interval is also used for event images. The radius determines the area from which images are selected.

The category keyword and its synonyms are compared agains Flickr image title, description and tags, and a match is required to have a relevant image. Synonyms of the category keyword are, prior to tag retrieval, collected from Princeton WordNet[g].

An algorithm for dynamic tag collection is implemented, based on the work in [Evertsen (2010)]. The algorithm first selects relevant images and, secondly, gather tags from the set of selected images. During selection of object images, the algorithm ensures that only one image from each user is selected. For event images, where availability of images from a specific event may be low, several images from the same user are accepted.

The total number of images retrieved and number of tags gathered are dynamic, depending on the availability of relevant images and the frequency of tags. Initially the system requires a minimum of 50 images as a basis for tag collection. If few images are available, the number of required images is lowered. To ensure that infrequently used and possibly irrelevant tags are avoided, we only collect tags that are used on at least 30% of the selected images.

Geotagged Wikipedia articles. Geonames[h] keeps a database of geotagged Wikipedia articles that are pinned to locations by gps coordinates. We are using the function FindNearbyWikipedia in the Geonames API to find articles about object,

[g] http://wordnet.princeton.edu//
[h] http://www.geonames.org/

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 given as input to the function. 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. Finally we use category, location name, tags and date information to search Google for relevant web pages. The main objective of the search is to collect information that may be relevant to the content of the image. For an event image, where we seek information about the specific happening at a given date and location, temporal information is included in the search parameters.

To retrieve content relevant information, image category and tags are important for focusing the search. The category keyword represents reliable information about image content provided by the user, while for many images the collected tags may include the name of object or event. Queries sent to Google are built on the following form, based on image *type*:

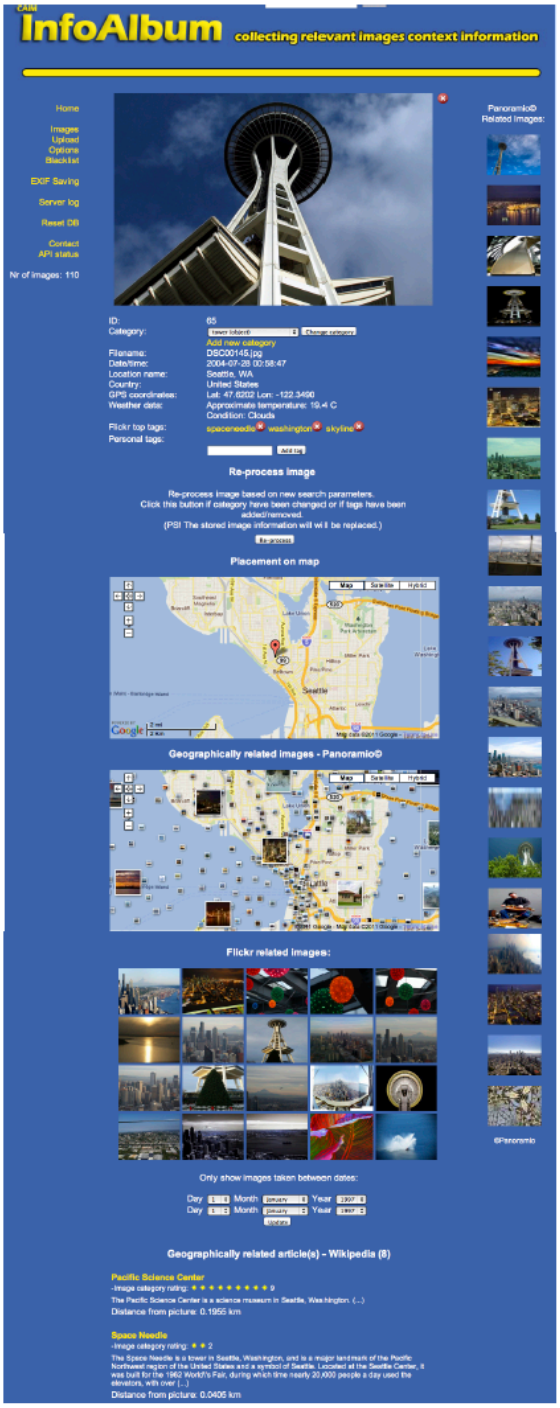
Table 2. Words in Google search queries

Image type	Words in search query
Object	location category (tag1 OR tag2 OR ... OR tagN)
Short event	location category date (tag1 OR tag2 OR ... OR tagN)
Long event	location category year (tag1 OR tag2 OR ... OR tagN)

To ensure that obviously irrelevant information is not presented to the end-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.

The result page presented by Google is crawled, and the articles URL, title and summary are extracted. If words in the filter list are found in the URL or in the header of an article, these are considered as irrelevant and filtered out. The remaining 20 top ranked hits are presented to the user and stored in the InfoAlbum database. 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.

Presenting the collected information. Figure 2 displays a screenshot from InfoAlbum that shows how the collected information is presented to the user. At the top of Figure 2 we see the query image (here Space Needle). Below the image we find textual information about temperature, weather condition, tags, location names and other information. To the right is a row of images collected from Panoramio, while below the maps we see images collected from Flickr. Two maps are shown; one which pinpoint the gps position of the image, the other shows Panoramio images



Nr of images: 110

ID: 65
Category: tower (object)
Filename: DSC00145.jpg
Date/time: 2004-07-28 00:58:47
Location name: Seattle, WA
Country: United States
GPS coordinates: Lat: 47.6202 Lon: -122.3490
Weather data: Approximate temperature: 19.4 C
Condition: Clouds
Flickr top tags: spaceneedle washington skyline
Personal tags:

Re-process image

Re-process image based on new search parameters.
Click this button if category have been changed or if tags have been added/removed.
(PS! The stored image information will wi ll be replaced.)

Placement on map

Geographically related images - Panoramio©

Flickr related images:

Only show images taken between dates:

Day 1 Month January Year 1997
Day 1 Month January Year 1997

Geographically related article(s) - Wikipedia (8)

Pacific Science Center
-Image category rating: ★ ★ ★ ★ ★ ★ ★ ★ ★ 9
The Pacific Science Center is a science museum in Seattle, Washington. (...)
Distance from picture: 0.1955 km

Space Needle
-Image category rating: ★ ★ 2
The Space Needle is a tower in Seattle, Washington, and is a major landmark of the Pacific Northwest region of the United States and a symbol of Seattle. Located at the Seattle Center, it was built for the 1962 World\'s Fair, during which time nearly 20,000 people a day used the elevators, with over (...)
Distance from picture: 0.0406 km

Fig. 2. Screenshot from InfoAlbum

on a map. This information is followed by a list of Wikipedia articles and finally a list of web pages. We only see the start of the Wikipedia articles list in Figure 2. By clicking on the title of an article or web page, the selected information appears 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 based on gps coordinates, is a popular approach used in systems such as Flickr, Panoramio and Google Picasa. Also [Serdyukov (2009)] 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 [Popescu (2009b)], ZoneTag[i] [Ahern (2006)] and iPicca [Pro (2009)] 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 [Moxley (2008)] 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 [Kucuktunc (2008)] expands user provided tags of an image by incorporating tags from other images which are visually similar. The work [Quack (2008)] and [VanGool (2009)] describe how to collect and cluster images of landmarks, which are subsequently automatically annotated and enriched with Wikipedia information.

Google Goggles[j] and Nokia Point & Find[k] 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 such as books and DVDs, landmarks, logos, contact info, artwork, businesses, products, and for barcode scanning.

Google Search by Image[l] 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. The search results page can show both textual information and related images. InfoAlbum and Google Search by Image have similar objectives, in that more

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

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

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

[l] http://www.google.com/insidesearch/searchbyimage.html

information about images are sought, but apply different techniques when handling images. Because of the similarities in objective, we compare the performance of the two systems in section 3.5.

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 Browser[m] 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 Browser[n] 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 [Popescu (2009a)] 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 Data[o] [Bizer (2009)] and the Linking Open Data Community Project[p] 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 pages), other images, tags, weather information and maps. This information is related to the image through mapping based on location, date/time and category information. As opposed 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 image 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 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 [Yeh (2008)] in the sense

[m] http://www.wikitude.org/
[n] http://www.layar.com/
[o] http://linkeddata.org/
[p] http://esw.w3.org/SweoIG/TaskForces/CommunityProjects/LinkingOpenData

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 has been tested with 97 images, where 69 depict an object and 28 some event. These images represent 97 different queries to the InfoAlbum system. For each image, InfoAlbum collects 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 Note Dame, to less famous objects located in much smaller cities. For events we chose some famous events such as the Carnival in Rio and Octoberfest in Munich, via a U2 concert in Barcelona and a football match with Manchester United, to a small football match and local festivals with only regional interest. The categories used were for instance "architecture", "bridge", "church", "monument" and "tower" for object images, and "concert", "festival" and "football" for event images.

Some of the collected information represents facts. Examples are the location names collected based on gps coordinates and weather condition at the closest weather station at image capture time. The nearby images are collected among the most recent images in the specified area, and will only have the location of image capture in common with the query image. However, in many cases, sharing location is sufficient to get many images of the object seen in the query image.

In this section we evaluate the performance of InfoAlbum with respect to relevance of the collected Wikipedia articles, web pages and tags.

3.1. *Content and Location relevancy*

As we have a dual objective in InfoAlbum to collect information that is relevant to both i) content of the image and ii) location where the image was taken, we must evaluate performance of the system with respect to both objectives. We therefore introduce two 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. 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 in the area. A document that is Content Relevant is also considered 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 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.

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. $Precision_{CR}$ and $Precision_{LR}$ can now be formulated as

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

3.2. ***Wikipedia articles***

Wikipedia articles are collected in two different ways. Firstly, based on the location names, by accessing Wikipedia directly to obtain references to articles describing the place where the image was taken. For each image, articles describing locality, county and country are collected. All these articles are 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 3 we see that 82% of the object images and 89% of event images received a Content Relevant article, while the numbers for Location Relevant articles were 91% and 82% respectively. Only 3% of the object images did not receive any geotagged article. We also found that for 5% of the object images, there do not exist a Wikipedia article describing the object in focus.

Table 3. Geotagged Wikipedia articles in InfoAlbum

	CR article	LR article	no article	CR article 1. ranked
Object images	82%	91%	3%	74%
Event images	89%	82%	0%	82%

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 was 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 geo-tagged 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.

3.3. *Tags*

For each image, tags are automatically collected from relevant images on Flickr as described in Section 2.4. For the 97 images in our test, the system collected between 0 and 5 tags for each image. The majority of tags were location names and object names.

In Table 4 we see that InfoAlbum collects 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 do not get any tags. We believe the reason for this is the lack of a sufficient number of tagged images from events.

Table 4. Automatically collected tags in InfoAlbum

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

As reliable location names can be collected based on gps coordinates, we are here primarily concerned with the Content Relevant tags. All collected tags are presented to the user, and the 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 also used as input to Google search. This in order to automatically construct focused queries that give result sets with relatively high $Precision_{CR}$ scores. The evaluation of this approach is found in section 3.4.

3.4. *Web pages*

As described in Section 2.4, InfoAlbum performs, for each image, textual search in Google, filters the results against the filter list and presents the 20 top ranked hits to the user.

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 the average precision is presented in Table 5.

Table 5. Average precision of collected web pages

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

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 3, where we show the number of images for which $Precision_{CR}$ is within a specific range. From Figure 3 we see that 10 object images and 10 event images have 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.

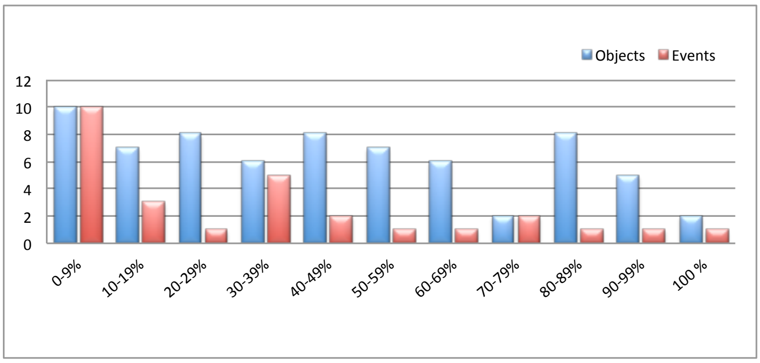


Fig. 3. Distribution of content relevance precision scores

In Section 3.3 we saw that 59% of objects and 21% of event images were given a Content Relevant tag through automatic tag collection. To evaluate the impact of using Content Relevant tags in Google search, we compare the average web page precision score of all images with the average precision score for the subset of images where Content Relevant tags are available. This comparison is presented in Table 6, where we see that the average $Precision_{CR}$ increased from 0.40 to 0.61, which is a statistically significant improvement.

When collecting web pages, we rely on the ability of Google to retrieve and rank information. A main objective in InfoAlbum is to identify and collect the best possible search parameters, so that relevant information is retrieved. The better these parameters are, the better the result from Google. The current testing shows

Table 6. Precision for images with content relevant tags

	All images	CR tagged images
Average $Precision_{CR}$	0.40	0.61

that content relevant tags can significantly improve the average precision scores for web page retrieval. This means that by improving the automatic tag collection algorithm, and thereby retrieving content relevant tags for more images, information retrieval through Google search is also improved.

3.5. *InfoAlbum vs. Google Search by Image*

Google Search by Image (GSbI) is a new service from Google that allows a user to search for information related to a specific image. The service is different from InfoAlbum in that it is based on image analysis to answer the query. We choose, however, to compare with this service since the objectives are to some extent similar.

We have tested GSbI on the same set of images used in the InfoAlbum test. Based on our testing, we find that the strength of GSbI is the ability to detect identical images on the Internet and guess image content based on the found information. GSbI does also provide good results if the query image depicts some distinct and well known structure with a lot of similar images available on Internet.

Table 7 summarizes the test of GSbI, and we see the amount of images for which GSbI provided i) a Content Relevant (CR) guess, ii) a Location Relevant (LR) guess, iii) no guess, and iv) an incorrect guess. A Content Relevant guess is typically the name of the object seen in the image, while a Location Relevant guess is the name of the city or area where the image was taken.

Table 7. Test results of Google Search by Image

	CR guess	LR guess	no guess	wrong guess
Object images	51%	10%	38%	1%
Event images	14%	0%	82%	4%
Public images	69%	10%	18%	3%
Private images	21%	5%	72%	2%

When testing, we first distinguished between *object* and *event* images, and secondly between *public* and *private* images. A *public* image is in this context an image for which an identical copy can be found on the Internet, while a *private* image is an image we have taken ourselves and where an identical image is not publicly available. Among the 97 test images, we had 39 public images (that was copied from freely available images on Internet) and 58 private images.

In GSbI there is a close correspondence between the ability to make a correct guess and present relevant textual information. Among the images with a Content Relevant guess, 95% had a response page that included one (often two) Content Relevant articles. For images without a guess, no articles were collected.

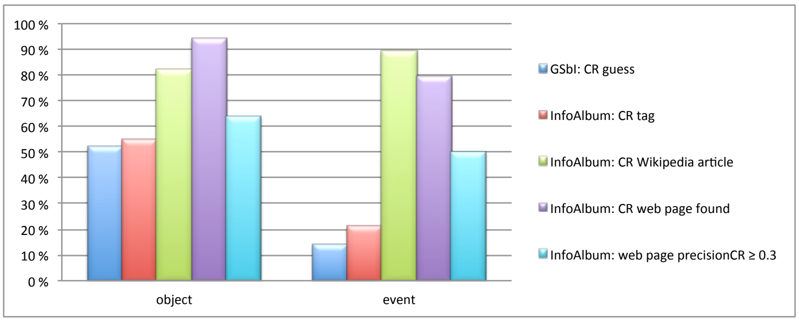


Fig. 4. InfoAlbum vs. Google Search by Image - the object-event image view

We compared InfoAlbum and GSbI with respect to the ability to provide Content Relevant information. In Figure 4 we distinguish between object and event images, and measure the fraction of images that in GSbI were given a Content Relevant guess, and in InfoAlbum were provided with Content Relevant i) tags, ii) Wikipedia articles and iii) web pages. In Figure 5 the same type of comparison is done with respect to public and private images.

From Figure 4 we see that 51% of object images were given a Content Relevant guess in GSbI, while in InfoAlbum, 82% were given a Content Relevant Wikipedia article and 55% a Content Relevant tag. For web pages we first show the fraction of images where at least one content relevant web pages were found, and secondly, the fraction of images with a $Precision_{CR}$ score of 0.3 or higher.

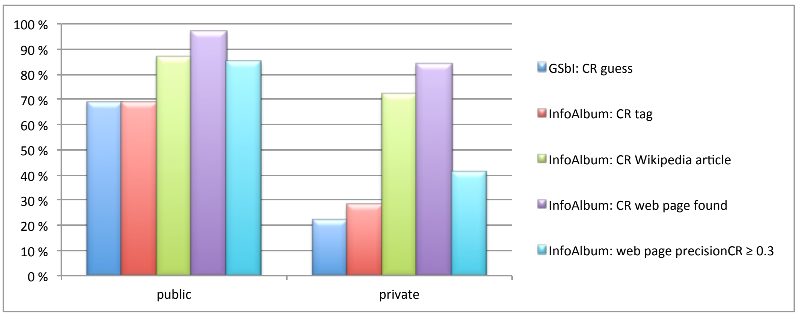


Fig. 5. InfoAlbum vs. Google Search by Image - the public-private image view

InfoAlbum has the advantage of using gps coordinates and thereby the possibility to narrow down a search for information to the specific area where the image was taken. The usefulness of gps coordinates are specifically evident when collecting geotagged articles from Wikipedia, where the InfoAlbum performance is very good. We also notice that the automatic image tagging used by InfoAlbum performs equally well or slightly better than the GSbI guess for all types of images.

Note that InfoAlbum is capable of collecting Content Relevant web pages for a high number of images. However, for some of the images, the $Precision_{CR}$ score is low and the information may consequently be difficult to find. Figures 4 and 5 therefore also display the fraction of images with a $Precision_{CR}$ score of 0.3 or higher.

Both InfoAlbum and GSbI show the trend that it is easier to find Content Relevant information for object images compared to event images. One reason may be that there are much more information available about objects compared to events. At least for those objects that have some public interest.

In Figure 5 we see that information collection was in general much better for public images compared to private images. For GSbI this is explained by the way GSbI uses identical copies on the Internet to guess image content. For InfoAlbum, one reason may be that the majority of event images in the test set are private, and also that a majority of images of less famous objects are private.

As InfoAlbum and GSbI use different techniques for handling the query image, we consider them complementary approaches and techniques from both systems may well be combined to improve information collection.

3.6. *Discussion*

Testing of InfoAlbum shows that it is possible to collect information about publicly known objects and events. As an example, the system collected name and description of objects, such as churches, towers, monuments and bridges. For events, the system was able to identify for instance which concert, football match or festival it was, name of band and band members, tour schedule, names of football teams and players and the match result.

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 information available in Wikipedia. However, the design of InfoAlbum heavily effect the outcome of the searches, in that InfoAlbum should provide search parameters that as good as possible cover the information need. Currently we have identified category, location name, tags 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, 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 is 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.

A main goal when designing InfoAlbum was to develop a system that could take an unknown or forgotten image and uncover where it was taken and what it depicted. When testing InfoAlbum, we have consequently assumed a user that does not know or remember exactly what the image depicts, and we have used generic category keywords for all images. A generic category (such as *tower* and *bridge*) can be selected without any background knowledge about the image. These categories are also reusable (i.e. applicable to a class of images), and once defined by a user, InfoAlbum stores the category for later use.

Our tests have revealed some challenges when using generic category keywords. We have for example seen 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 provides 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 does not impose any restrictions on the defined categories. If the user knows exactly what the image depicts, she may well define a more specific category. By naming the specific object or event (such as "Notre Dame" or "U2 concert"), information collection through InfoAlbum may give more relevant hits and better $Precision_{CR}$ results than reported here.

InfoAlbum allows users to re-process the information collection task using a different category. This is useful if there are more then 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.

The user may also manually add tags to the image, and subsequently re-process the information collection. A scenario may be that the user identifies the depicted object from the information initially collected by InfoAlbum, and requests a re-process after adding the name of the object as a tag. Our test results from the use of Content Relevant tags, show that the effort of manually adding such a tag will significantly improve the relevancy of the information returned by InfoAlbum.

As previously pointed out, we have a dual objective in InfoAlbum to collect information that is relevant to i) content of the image and ii) location where the image was taken. A third objective of InfoAlbum is to collect information that can later be used for image retrieval purposes. This means that a subset of the information collected by InfoAlbum is selected and stored as image metadata. Since the purpose of this metadata is to describe the image itself, and not nearby objects, the selected information must be Content Relevant to the image. 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.

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.

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*, automatically 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 and weather condition at image capture time, placement on map, geographically nearby images, Wikipedia articles, and web pages.

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, and the relevancy and usefulness of automatically collected tags.

To evaluate retrieval performance, we introduced two relevancy concepts; Content Relevance (CR) and Location Relevance (LR), and use two corresponding precision scores; $Precision_{CR}$ and $Precision_{LR}$.

For both object and event images we found that Wikipedia articles provide a very good source of information. This is true for both geotagged articles where over 80% of the images received a Content Relevant article, and for Wikipedia articles collected based on location names, where all articles were relevant with respect to the location of image capture.

We found that automatically collected tags were very useful as input parameter when collecting relevant web pages, as they in many cases contribute to focus the information search. Our test shows a moderate average $Precision_{CR}$ score of 0.40 for the whole set of test images. However, when focusing on CR tagged images only, the average $Precision_{CR}$ score increases to 0.61, which is a significant improvement of performance.

To further improve the system we believe that better precision can be achieved by choosing specialized information sources targeted to specific categories of information. 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 in InfoAlbum. We will further 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.

5. Acknowledgments

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

References

Ahern, S., *et al* (2006) 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.

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

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

Karlsen, R., Jakobsen, B., (2011) The InfoAlbum, Image Centric Information Collection. In *Proceeding of WIMS11, International Conference on Web Intelligence, Mining and Semantics*, Sogndal, Norway, May, 2011.

Kucuktunc, O., *et al* (2008) 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.

Moxley, E., Kleban, J.and Manjunath, B. (2008) 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.

Popescu, A., Grefenstette, G. and Bouamor, H. (2009a) Mining a multilingual geographical gazetteer from the web. In *The 2009 IEEE/WIC/ACM International Conferences on Web Intelligence and Intelligent Agent Technologies*, Milano, Italy, 2009.

Popescu, A. and Moëllic, P.A. (2009b) 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.

Proß, B., Schöning, J. and Krüger, A. (2009) 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.

Quack, T., Leibe, B. and Gool, L. (2008) 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.

Serdyukov, P., Murdock, V. and Zwol, R. (2009) 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.

Sundby, D. (2011) *Summarizing image collections.* Thesis in Computer Science, University of Tromsø, Norway, June 2011.

Van Gool, *et al* (2009) 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.

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