Chapter 1

Geospatial Web Image Mining

Dirk Ahlers
OFFIS Institute for Information Technology
Susanne Boll
University of Oldenburg
Philipp Sandhaus
OFFIS Institute for Information Technology

Abstract: One commonly asked question when confronted with a photograph is "Where is this place?" When talking about a place mentioned on the Web, the question arises "What does this place look like?" Today, these questions can not reliably be answered for Web images as they typically do not explicitly reveal their relationship to an actual geographic position. Analysis of the keywords surrounding the images or the content of the images alone has not yet achieved results that would allow deriving a precise location information to select representative images. Photos that are reliably tagged with labels of place names or areas only cover a small fraction of available images and also remain at a keyword level.

Results can be improved for arbitrary Web images by combining features from the Web page as image context and the images themselves as content. We propose a location-based search for Web images that allows finding images that are only implicitly related to a geographic position without having to rely on explicit tagging or metadata. Our spatial Web image search engine first crawls and identifies location-related information on Web pages to determine a geographic relation of the Web page, and then extends this geospatial reference further to assess an image's location. Combining context and content analysis, we are able to determine if the image actually is a realistic representative photograph of a place and related to a geographic position.

Xian-Sheng Hua, Marcel Worring, and Tat-Seng Chua (Ed)

1.1 Introduction

The role of geographic location for images and the relation of Web information to a physical place has gained much attention in recent years. Web search is increasingly becoming "local", search engines not only index content by keywords but also by geographical locations. Users can have information mapped to a certain place or region of interest and search accordingly. Web image search, on the other hand, is mostly keyword-based or explores basic image features such as color or texture. A geospatial Web image search that works on unstructured Web pages with arbitrary images is not yet available. Searching for images related to a certain location quickly shows that Web images typically do not easily reveal their explicit relation to an actual geographic position. The geographic semantics of Web images are either not available at all or hidden somewhere within the Web pages' content. This lack of explicit location information also applies to the images contained in Web pages. Even though GPS has arrived at the consumer, the photos that actually have a location stamp or are geo-tagged cover only a tiny fraction of all Web images. Furthermore, such images are mostly only available in social communities, but seldom find their way into the interesting asset of local Web pages. This poses an interesting challenge to the field of Web image retrieval, namely, the derivation of reliable location information for Web images from unstructured Web pages.

Currently, a Web search for photos of a certain well-known place name remains a keyword-based search of known geographic place names. This can work for known places, landmarks, or regions such as "New York" or "Singapore" but does not deliver a precise geo-reference for Web image content. We therefore need a spatial Web image search to enable users to search for images from a certain geographical coordinate or its spatial vicinity to complement geospatial Web search engines.

In this work, we develop a system for geospatial Web image search [1]. We aim at a general approach for common Web images that works beyond keyword-search and does not need prepared tags or annotations. We derive possible location information for those images that are contained in Web pages with an established spatial context. The mutual relation of a Web pages' text, location, its contained images, and accociated metadata is a main driving force behind this approach. We employ our spatial search engine for a location-based geo-referenced search down to the address level [2]. It exploits the location information of a Web page that is not explicitly contained as metadata or other structured annotation but is rather an implicit part of the textual content.

Based on this previously identified location of the Web page, we can identify representative photographic representations of the mentioned location. The potential propagation of this location to the images embedded in the page and its reliability is determined by content-based and context-based analysis. In a first step, an analysis of the Web pages identifies photographs, i.e., the image content relevant for a spatial image search for realistic depictions of places. Following the photograph identification, a location-relevance classification process evaluates image context and content to determine if the photograph is actually related to the location already identified for the Web page.

We evaluated and combined several heuristic approaches for analyzing and scoring relevant features. These are categorized into context and content information for both Web pages and embedded images. Feature classifiers are combined to achieve an overall assessment of an image's connection to a location. As a means to reduce the computational cost, classifiers' individual analyzers are ordered by their discriminatory ability and their computational cost to filter out non-promising images early, preferably even without the need to download them. Our test and evaluation show that, even though only a small fraction of the Web pages reveal photos that have a relation to the location, these can be well related to their geospatial context.

With a Web search that reliably finds georeferenced images from unstructured Web pages located at a certain geographic location or region, a large number of potentially geographically-related images on Web pages becomes available for spatial Web image search. This can be applied for a variety of interesting applications such as searching for images along a route when preparing a vacation, extending personal photo collections, gathering visual information about an area, or providing representative images for spatial search results.

1.2 Related Work

This work researches the connection of Web images and location from an unusual angle. While the usual direction is to start from an image and examine its content and metadata to estimate a location, our approach takes a different route in that it starts with a document as the image's context which already contains a known location and tries to extend and propagate it to matching images. Using Web pages as a special case of context and drawing the connection to the image content, our hybrid approach thus attacks the problem from both sides simultaneously. A large body of literature exists on techniques for image retrieval, geographic retrieval, Web search and various combinations thereof. The following examines the different fields of related work and compares and applies their findings to the goal of our approach.

Data sources for geospatial image search

Location is a fast growing topic in the Web today, generating huge interest for users and a growing number of location-aware data sources. GPS has arrived at the end-consumer; services such as Plazes, Dopplr, Yahoo! Fire Eagle, or Google Latitude allow users to share their own physical location. Users can use their location to start a local search or link their current location to their photos. Modern cameras with an integrated GPS-receiver can directly embed coordinates in a photo's EXIF metadata. However, such geo-located images are rare in the first place even in large photo collections and are even more rare in a general Web context. This lack of metadata is only partly due to loss of information in editing processes. Especially in the Web, images are often edited before publication and their EXIF information inadvertently destroyed. Regarding loss of metadata, [41] presents a system that captures metadata at media capture time and preserves it over the production process in a professional environment, which is therefore not yet suitable for common use and cannot address the current state of Web images. However, it is a complementary approach to metadata extraction and derivation on documents without systematically defined metadata.

A richer source for location-based images are Web 2.0 services that allow for organizing and sharing personal photo collections. They provide for an assignment of location-related information by manual tagging and annotation [37]. For example, Flickr (flickr.com) or other tag-oriented services allow adding names of locations to entities such as photos and to find photos tagged with the same placename. More intuitive to the user and more precise in the assignable location are services that provide means to drag digital content onto a map and by this manually geo-code the content: With Placeopedia (placeopedia.com) entries from Wikipedia can be associated with a geographic location and queried on a map; photos on Flickr or locr (locr.com) can also be positioned on a map to geocode them. Panoramio (www.panoramio.com) is a service entirely dedicated to collecting high-quality location-based images. These applications use collaborative efforts to manually geo-code image content and by this contribute to a localized search.

The massive amount of georeferenced images in these collections allows for interesting data mining approaches, for example, to identify places and sights of important landmarks [58], [49]. However, exploiting the long tail of photos of "less-interesting" locations is not as easy. Additionally, tagged media collections or prepared geocoded images cover only a very small fraction of images on the Web; many others still reside unrecognized on arbitrary Web pages.

A complementing approach to the long-tail of locations and their images can come from a different perspective. Several services are based on existing directories such as yellow pages. For their entries, the geographic location is known, enabling local search and visualization of results, e.g., on a 2D map. Such content is mapped by services such as Yahoo! Maps (maps.yahoo.com), Google Maps (maps.google.com), or MSN Live (maps.live.com), enabling map-based spatial search and visualization of results. By enhancing these entries with user-generated or specifically crawled content, images for these lcoations can become available. For example, Google Maps has created the ability for business owners to enhance their yellow pages entry with images of their business. Such images have the added value of being manually selected as representative images for a place, so they match the type of images our approach aims at. Their PlacePages further aggregates information about a specific location. This trend clearly emphasizes the demand for accompanying images to location information.

Geographic Web information retrieval

The field of Geographic Information Retrieval concerns itself with research towards extracting geographic meaning from unstructured documents and spatially indexing it. In many cases, the documents are the Web's textual resources. Central challenges in the field are the identification, extraction, and disambiguation of geographical entities as described in [36], [5], [39], and [38]. Since textual clues for locations are often highly ambiguous, additional disambiguating terms are used to verify found locations. Such other features can be telephone numbers, state names, or similar. These features are verified against domain knowledge taken from gazetteers, geographic taxonomies detailing geographic coverage and location; placenames; physical, political, or demographic features etc. [18]. Further heuristics are used to aid in disambiguation and in improvement of retrieval performance in extracting geospatial entities. For example, when sufficient disambiguation terms are missing, larger locations may be preferred over smaller ones and locations near each other are often more probable than those with higher distance between them. Of specific interest to our work are those systems that aim to identify location references at a high granularity to pinpoint content to individual coordinates and not just a broad region. Exemplary geospatial Web search engines are presented in [43] and [36], addressing issues of extraction, indexing, and query processing.

The basis for the proposed geospatial Web image search is a location-based Web search engine we developed [1, 3] that employs geographically focused Web crawling and database-backed geoparsing at address-level granularity to spatially index Web pages and make them accessible to geospatial search.

Web image retrieval

Web image retrieval and classification aims to open up images on the Web for multimedia retrieval, often by textual queries that rather work on the surrounding content of the Web page than image content. Since the HTML code of a page provides rich context information for embedded images, most related approaches use text-based image retrieval as a basis by examining textual content such as surrounding text, image metadata, or take hints from the HTML code and especially its structure. Only secondarily, if at all, the actual Web image content is used. Similarly, current keyword-based Web image search engines such as Google Image Search (images.google.com) or Yahoo! Search (images.search.yahoo.com) employ mostly surrounding text features and the image name and path which allows for keyword-based queries. As an addition several approaches exist to combine traditional text-based queries with content analysis of images.

Looking at the field of image retrieval in general, [12] gives a good overview over current trends and approaches for image retrieval in general. An earlier work provides a good survey over the field on content-based image retrieval at *the end of* *the early years* [51] while [34] describes combinations of content- and context-based image retrieval specifically for photographs.

A survey and comprehensive discussion of existing technologies in Web image retrieval and how they address key issues in the field can be found in [28] along with application scenarios and a comparison of different existing systems. An early example for a system also considering the image content is ImageRover [50], combining text-based queries to provide initial example images following a content-based query-by-example approach to refine the result set. Other approaches combining text-based queries and content-based image similarity are [30, 42, 33]. Besides these works for general web image search more specialized approaches exist concentrating on specific domains (e.g. celebrity search by face recognition [6]).

Purely text-based Web image retrieval approaches are solely taking the HTML code into account to derive annotations for the images contained in the page. As one example, [54] uses an approach for image retrieval examining the attributes of the image tag itself, surrounding paragraph and title of the page. However, further examination of the the effectiveness of features [40] leads to the conclusion that the HTML source contains the most valuable initial clues, but that results improve if images themselves also are examined. Aiming to understand the role of an image on a Web page, [35] uses a range of features to assign roles to Web images. The features are taken from structural document data and image metadata but not from actual image content. Based on these features, various roles of Web images are derived and used to preprocess images for display on a mobile device. By means of structural analysis of HTML documents, [16] segments Web pages into semantic blocks. The authors propose that embedded images inherit the semantics from their surrounding semantic block. Other approaches consider both content and HTML context of an image for semantic annotation and clustering. An early work [13] uses the HTML content and a few selected image content features for Web image search. Building upon this work, [7] developed a classifier to distinguish images into two classes of either photographs or graphics using decision trees. The iFind system [17] considers context and content of images and examines their layout on a page for semantic annotation.

Aiming at a more intuitive presentation of Web multiple ambiguous image search results, approaches such as Igroup [25] as well as iFind [14] additionally consider the surroundings and the content of an image to organize Web image search results into semantically coherent clusters. The latter additionally examines the spatial layout of images on Web pages to identify nearness for groups of images. Closely arranged images are then considered to also be semantically related.

Content-based location detection

The use of only content-based image analysis to derive a location has produced some interesting results. Often, such systems are developed to work on known locations and are independent from Web pages. IDeixis [53] is a system to compare pictures taken by a mobile device to previously taken photos to initiate a search by finding similar photos. By combining image analysis and comparison methods, the most likely location of the given image can be determined. Image comparison is done rather naively by emplyong euclidian distance according to color histograms. Photo-to-Search [24] approaches a similar use case, but with an indexing and search system based on SIFT features provides a more sophisticated approach enabling for more accurate and faster search results. The PhotoTourism system [52] demonstrates how, with an extensive set of overlapping photographs of a place, a 3-dimensional reconstruction of that place is possible by estimating viewpoints of photographs and stitching them together for a 3D-scene working on image-based rendering.

These approaches are relevant for our work as they show that location-recognition based on image content is possible in case the possible location is known and the retrieval scope is thus narrowed down. However, these systems require an already previously established collection of georeferenced images as the basis for comparison.

Landmark Recognition

Taking a complementary approach, annotated Web images can also be used to find places of interest and thus create a rich database of high-ranking places. Several groups have employed web images to extract popular world landmarks: The goal of *Tour the World* [58] is to build up a global landmark database by mining GPS-Tagged images from the net and online tour guides. Landmarks are extracted by unsupervised visual clustering of candidate images resulting in a database of currently 5312 landmarks from 1259 cities in 144 countries. [45] follows a similar goal by mining images from community photo collections and clustering these images in an unsupervised fashion regarding visual, textual and spatial proximity. Extracted landmarks are verified and semantically enriched by combining them with matching Wikipedia articles. [23] mines famous city landmarks from blogs with the help of a graph modeling framework, fusing context, content, and community information. [10] employs only tag and location information from community images to identify popular places of a specific area. [27] tries to find representative photos of places by analyzing large collections of community-contributed photos. The work leverages context information like location and tags as well as content-based features. An earlier work [21] solely works on the location information of the photos and their tags to find popular places and corresponding photos of a specific area or to also couple events and locations by more powerful algorithms [46]. This allows the visualization of photo's concepts by grouping them by their location and time.

Semantic Image Annotation

Many attempts have been made in the past to narrow the so-called *semantic gap* in image retrieval by trying to derive high-level semantic annotations from low-level image features. These attempts either try to identify specific kinds of objects in an image or try to categorize images according to semantic concepts. Most works rely on a combination of content-based features and supervised learning.

In the first group, one on the most heavily studied applications is face detection where [57] provides a comprehensive overview. Accordingly, other works focus on other object types such as on text characters [32], vehicles [29] or persons [9]. Considering the second group one can find many works trying to build classifiers for a limited number of categories. For example, [56] describes one of the first examples of simple concept detection by categorizing images as showing parts of a city or landscapes based on low-level image features. In a follow-up, [55] describes a hierarchical classification to detect high level concepts such as indoor or outdoor, city or landscape, sunset or forest etc. with high accuracy by performing supervised classification on the basis of low level features. A more recent contribution is [11] which tries to automatically tag or retag images according to general concepts and thus making them available to text based searches. A real time image annotation engine is presented in [31] and is available as a real-world system on the web (www.alipr.com). Work about inferring semantics by multimodal analysis of personal photo collections and their spatial and temporal properties was done by, e.g., our group in [48].

Regarding the relation of keywords and location, [20] analyzes the dependence of images and text on a page. [49] uses a combination of metadata and content-based retrieval methods in an approach for landmark clustering. With a strong geographic information retrieval background, [44] uses a combination of content-based image retrieval and text-based geographic retrieval to integrate a geographical facet into image search.

Following this line of work, we aim at an approach that can reliably derive a connection between images and their location, working on arbitrary images on the Web, useable on a Web-scale. The remaining open issue therefore is that explicit location information for Web pages and their images is given only for a small fraction of the Web, which leaves a large number of potentially geographically-related images on Web pages as yet unrecognised for spatial search. However, they could be uncovered by better use of implicit location information. This opens up a potential for other methods of location assessment and additional sources of data for geo-referencing Web images.

1.3 Geospatial Web search

Geospatial search allows querying for Web pages related to a certain given location. Most current search engines are already very efficient for keyword-based queries, but lack geospatial capabilities since the search for geospatial information poses some special requirements that they cannot completely fulfil. For example, certain location-based queries cannot be answered with a purely textual index, such as vicinity queries within a certain radius, results in a specific district of a city, or other geospatial constraints. Furthermore, the result presentation should consider and cater for the geospatial dimension.

Nowadays, location-based Web search and also services operating on prepared geo-referenced data receive widespread attention. This is consistent with the high relevance of location to the user. A study in 2004 [47] finds that as much as 20% of user's Web queries have a geographic relation while [26] estimates this at 5-15% for search on mobile devices. On the other hand, the Web contains a massive amount of resources that exhibit geospatial references. Web pages with the required location-related information include home pages of businesses, restaurants, agencies, museums etc. These are excellent sources to answer user's queries for relevant spatial information.

Even though these pages only represent an estimate of 10-20% of all Web pages [19, 38, 22], their relation to a physical location is a yet widely unused asset for an interesting set of location-based applications. The main goal of a spatial search engine is to properly discover and identify those Web pages that bear a relationship to a geographic location and process this information for a spatial search. For our own spatial Web search engine as presented in [1], we use common Web crawler and indexing technology and extended it with our own location specific components. Our spatial crawler and indexer identify Web pages that exhibit a relation to a physical location, extract the desired information, and assign a geo-reference to the pages. They are designed to exploit the location information of a Web page that is not explicitly contained as metadata or structured annotation but is rather an implicit part of the textual content [2]. Based on our research in mobile pedestrian applications [8], we tailor our search to a mobile user who needs precise geo-references at the level of individual buildings.

The architecture detailed in Figure 1.1 comprises the main components of our spatial Web search engine prototype, separated into function groups of crawling, indexing, and search. These groups are modeled after the general architecture for Web search engines, but are extended with spatial capabilities to support the spatial search approach. The indexer can integrate several parsers, for this work we added an image analysis component. The architecture supports an efficient geographically focused crawling, meaning that the crawler aims at retrieving mostly only relevant pages [3]. For this, a seed generator feeds the focused crawler a number of seed URLs. These seeds are selected to be location-oriented and enable the crawler to

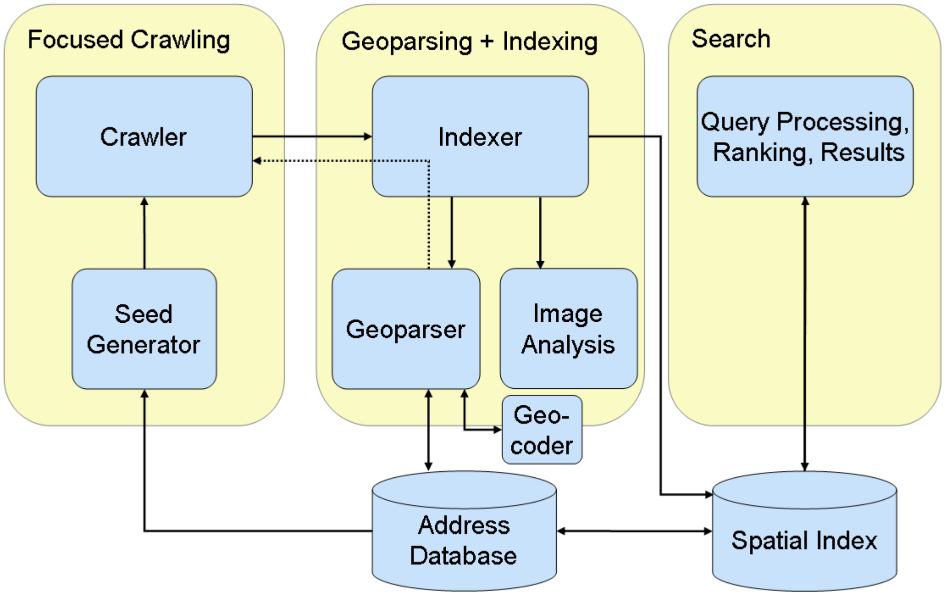


Figure 1.1: Architecture of the spatial search engine

start with pages that have a high relevance. The seeds alone, however, only form the starting point for focused crawling. A key aspect of focused crawling as opposed to general crawling is a feedback loop on downloaded documents to steer the crawler towards pages with the desired topic. Besides the analysis of the textual Web content as known from search engines, a geo-parser analyzes the page and tries to extract geographic information from it, if available. This extraction process is also supported by an address database used as a gazetteer [18]. Whenever a relevant location-related information is found it is notified to the indexer for a later geo-coding and insertion into a spatial index. At the same time the information about a location-information on a crawled Web page is fed back to the crawler to support the focused crawling. If the geo-parser determines one or multiple addresses on a page, each single address is geo-coded, i.e., the hierarchical textual description is mapped to a coordinate of latitude and longitude. The coordinate is then stored along with the textual address, the page and its URL in our spatial index. For this process, we use a commercially available address reference database to enable the geo-coding of found address to a geo-referenced position [4]. The index of the spatial search engine can then be used for the known keyword-based search now extended with spatial search capability.

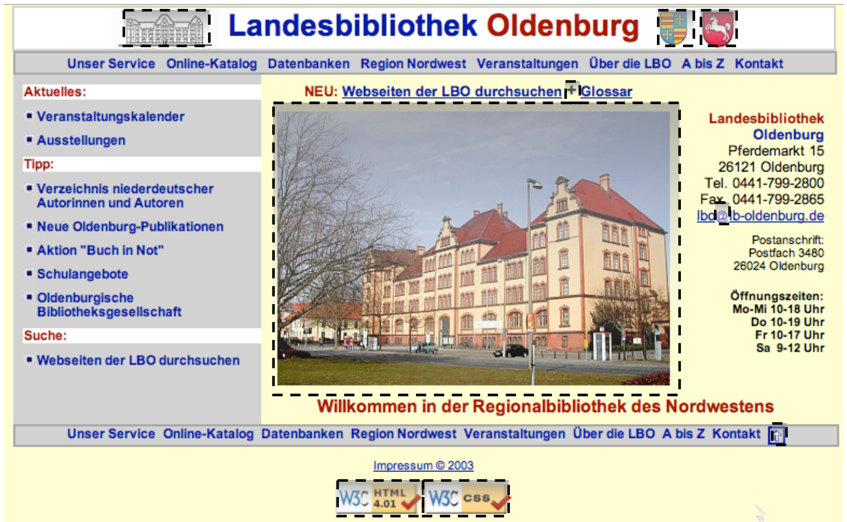
Location references can be very diverse, ranging from mentions of countries, broad regions, city names down to exact addresses. As our approach naturally restricts the amount of pages we can identify to those with precise geographical references, we might miss location-related photographs that are on pages without such references. On the other hand, as our location-based indexing delivers highly relevant and precise spatial information, this places the location-based Web image search on good grounds to deliver images at very high location granularity. The architecture and design of the location-derivation from Web pages has been presented in [1]. As one of several additional and supportive features for the geospatial Web search, this paper proceeds to examine its image-oriented aspects to anhance the search results by the extraction of representative images for the found results.

1.4 Web image location-assessment

While today an interesting set of semantic annotations can be derived from an image's content, a geographic location is usually not part of it. Apart from well-known geographic landmarks, existing annotations, or a prepared environment where images of all objects are already catalogued, a randomly chosen Web image showing some part of a town or some building cannot be processed by current techniques to pinpoint an actual location. This is of course different for images that come directly from a modern camera with GPS-information directly embedded in the EXIF metadata. However, such geo-located images are rare in the first place even in large photo collections and are even more rare in a general Web context. We therefore need other methods of location assessment and additional sources of data for geo-referencing Web images.

The goal of our spatial Web image search is to be able to automatically illustrate locations with relevant, representative images of themselves or their surroundings and for a given place to answer questions like “What does this place look like?” or “What is a representative image of this place?” by retrieving them from the Web. We aim to derive location-information for those images that are contained in Web pages that reveal some spatial context. As discussed in the literature [40, 20, 17], the proposed system will use a combination of content and context features from both Web page and images. The extension of spatial annotations from a page to its embedded images is made by the use of images in the content of a page. A page's spatial semantics are transitively extended to the embedded images which then will, to a certain degree and depending on detected features, inherit the location of the page, thus implementing a location relevance propagation. For this, the implicit location contained in some images has to be made explicit, which is both an extraction and a verification problem. It will have to examine different features from multiple aspects of an image such as its content, its context, available semantic annotations, and possibly other sources of information.

As one example of our approach, consider the Web page of a public library in Germany as shown in Figure 1.2. The page features various images – surrounded by dashed lines – with the one in the center prominently displaying the library


Unser Service Online-Katalog Datenbanken Region Nordwest Veranstaltungen Über die LBO A bis Z Kontakt
W3C HTML 4.01
W3C css

Landesbibliothek Oldenburg

NEU: Webseiten der LBO durchsuchen Glossar

Aktuelles:

- Veranstaltungskalender
- Ausstellungen

Tipp:

- Verzeichnis niederdeutscher Autorinnen und Autoren
- Neue Oldenburg-Publikationen
- Aktion "Buch in Not"
- Schulangebote
- Oldenburgische Bibliotheksgesellschaft

Suche:

- Webseiten der LBO durchsuchen

Landesbibliothek
Oldenburg
Pferdemarkt 15
26121 Oldenburg
Tel. 0441-799-2800
Fax 0441-799-2865
lbo@lb-oldenburg.de

Postanschrift:
Postfach 3480
26024 Oldenburg

Öffnungszeiten:
Mo-Mi 10-18 Uhr
Do 10-19 Uhr
Fr 10-17 Uhr
Sa 9-12 Uhr

Willkommen in der Regionalbibliothek des Nordwestens

Impressum © 2003

Figure 1.2: Images and location on a sample Web page

building. Its address – *Pferdemarkt 15, 26121 Oldenburg* – is present just to the right of the image and its name and city in some of the metadata such as keyword and title. A location-based image search focused on representative images of places should yield the photograph in the center. In the example, only this photograph should be retrieved since it is the only one descriptive of the location. The image analysis should be able to filter out all other, rather decorating images on the page. An analysis of the images' context and content should determine that only this one in fact has a location relevance and a strong connection to the specific location.

Method for location assessment

For the identification of a possible location-relation of images from Web pages, we propose a method for image location assessment based on processing chains. As mentioned before, using the Web page itself, only preliminary conclusions can be drawn about the image content of embedded images as no high-level annotations about these images exist and only keywords and structure of the HTML content is available. Thus, the content itself has to be taken into account as well. This is mirrored in the general process, which comprises two main logical components, a classifier for photographs and one for location relevance, respectively. These are arranged as weighted filter chains and utilize various criteria taken from the HTML content and metadata, the metadata of embedded images, image content, and the address previously found to arrive at a weighted classification result. This process is outlined in Figure 1.3. The starting point is a Web page with an annotated known location from, e.g., our spatial search. This Web page is examined and all its image references are extracted. Each image is then analyzed and processed by a filter chain in the photo classification component. This analyzes each image to validate whether it satisfies the conditions to be considered a photograph. If the condition is met, the surrounding page as well as image features are input to the location classification. It examines image and Web page for hints towards a location reference, which are weighted and summarized. Some filters have can cast a veto on images where a direct negative classification is possible, effectively truncating the chain. Such cases can be 1x1 sized images, single-colored images etc. If a location could be reliably assigned, the process results in a geo-coded image. For both photograph classification and location classification, we utilize different information sources: the Web page itself, the HTML content surrounding the image, metadata of embedded images and their content, and the address previously found. With these criteria, we can have spatial information inherited and automatically annotate it to the images, based on a previously derived or known location of the Web page.

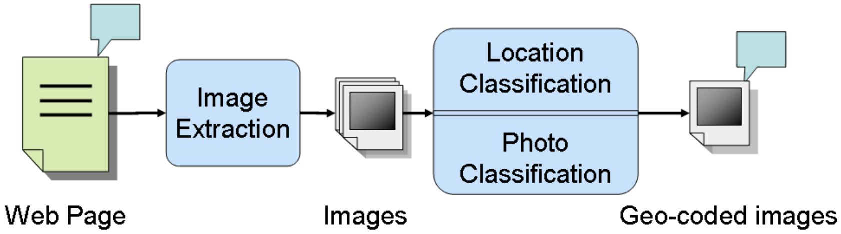


Figure 1.3: Logical process of the location assessment

Photograph classification

Images on Web pages are used for different reasons. Distinguishing their different roles supports filtering certain images by their roles. A first large group of images are those that serve an *illustrative* purpose: photographs, diagrams, graphics, sketches, titlebars, logos. A second group serves as *formatting objects*: bullets, borders, menu items, buttons, headings. Also, Web images can represent *advertisements* by means of banners, animated images or are used as "invisible" images used for *user tracking.* Other embedded media such as videos, flash or Java applets are outside the scope of this work since they have quite different semantics than images. We obviously miss some finer distinctions as needed by other approaches, but these classes suffice for the exclusion of non-location-bearing images according to our goal. Of the roles identified, we are only interested in photographs being the main bearers of visual location information. Such photographs serve illustrative purposes. Page authors use them with this in mind, and they clearly depict objects or scenery that is in most cases related to the text they accompany. This corresponds to the results of our initial surveys and probings.

Therefore, we developed methods to identify images with a high probability to be a realistic photograph of a place. To this end, we use different information sources and combine them into an overall classification by drawing on image metadata as well as image features. The different features and their relevance for photograph classification are discussed in the following. Each of these features are inconclusive, except in the direct negative case of a veto. Yet, in combination they yield satisfying results.

- *Image size* and *image ratio* is used to remove small images such as bullets, menu items and other formatting images as well as illustrative images that lack sufficient size or that have an unusual ratio that deviates too much from usual photos, in this case images with one side being less that 1/3 of the other. We found 100 pixels to be a good minimum size. This removes most advertisement banners, borders, rulers, headings etc. These filters are employed twice, once pre-download, if size information is available from the img-tag, and then post-download when size information is always available from the image itself.

- *size comparison* is an analysis over all images within a page to sort images according to their size. The assumption is that larger images often are more related to the content of a page. This can be observed in the case of news articles [15] which are illustrated with a teaser image, but can also – while with less weight – be used to search for photographs depicting locations. The analysis runs once in the pre-download phase, but since many images do not carry size information in the HTML source, is run again when all images of a page are downloaded. Since many images can already be removed by vetoes, its results may suffer from missing values and are thus weighted less.

- *filename comparison* is another page-scale analysis running only in the pre-download stage on the page content. Since image size information often is unavailable, it employs a different method to mainly detect formatting objects. First, it counts *identical image* filenames within the page to identify images that have multiple occurrences in a page. Second, all filenames are compared to each other. By using editing distances, *similar filenames* can be grouped and images contained in a group weighted less since they are more likely to constitute formatting objects.

- *Color count and histogram analysis* is used to distinguish photographs with their high number of colors from illustrations or drawings. We calculate the ratio of used colors in an image to the maximum possible colors depending on file format. Black & white images are detected to be treated with different thresholds. For GIFs with the restricted palette of only 256 colors we cannot give a reliable threshold but rate them generally lower; JPGs and PNGs with 16777216 possible colors were found to need a minimum of 0.18% of possible colors to qualify as a photograph; black and white images were invariant for gifs, the other formats needed 0.0014% of possible colors. In the case of ver few colors, images are most probably simple logos or similar and can receive a veto. Similarly, a histogram analysis aims at the same goal, but by using a greyscale histogram with its limited number of bins, it basically allows to work on a reduced palette, which can pool similar colors which often are present in jpeg-artifacts and might misguide a pure color count. The histogram therefore smoothes over such outliers and can provide a more aggregated view. For discrimination, bins are ordered by their value and the necessary number of bins to contain half the image's pixels are evaluated. Furthermore, if less than three bins are used, a veto discards the image.

- *Animation or transparency* in images usually indicates non-photographs. We also found many part-transparent click maps.
- *EXIF-data* in JPG files is a strong evidence for a photograph. Digital cameras today generate a multitude of camera-specific metadata such as camera type, focal length, flash usage, exposure etc. We use the fields of Flash (whether the flash fired), Model (the camera model) and ExposureTime as sufficient evidence that the image was taken by a digital camera and is thus a photograph. The reverse is not true; converting or manipulating images may destroy or remove the EXIF information.
- *Vertical edges* in above-average quantity can indicate photographs of a place showing buildings, pieces of architecture or trees. For our purposes, a modified Sobel filter operating on greyscaled images for edge detection worked best for urban buildings but is inconclusive for rural scenery.
- *Face detection* is used to distinguish mere snapshots of people at a location from images truly depicting the location itself. We found the number of such snapshots quite high and use this feature to focus on location-depicting photos.

Location-relevance classification

With the bearers of visual location information found to be in and around photographs, we have to make the connection to the actual location of the Web page. An image related to a location will exhibit certain characteristics which we exploit to create a set of measurements of how well the location of a Web page can be inherited by the images embedded in it.

Already knowing the location referenced on a Web page, we need to verify whether an image—or more precisely its content—is related to this location as well.

We assume a certain intention of a Web page's author about images embedded in the pages. Images identified as illustrative do usually indeed often serve to visually describe the content present in the page—in our case, the location of the address present on the page or the referenced building or place. This relation is not always very clear and it is also not always present. Though, an image related to a location will exhibit certain characteristics which we exploit to create a set of measurements of how well a given location of a Web page can be inherited by the images embedded in it. To establish a relation of an image and an address on the Web page, we rely on content and structure of the page in question. From the photograph classification step, we know all photographs on the page; from our search engine, we also know the geographic location of a Web page as a full address including street, number, zip code, city name and also geo-coordinate. We assume this location reference does not necessarily stand for the entire page but rather may determine a location for at least part of the page decreasing in relevance by increasing distance from the location reference. Then, an image in the vicinity of an address has a higher probability for location relevance than one further away. Also, if we can identify a page as generally dealing with the location, it's probability to contain relevant images increases.

For the implementation of the location-relevance classification, we consider the distance of the image to each address on the page and the matching of image descriptors and key elements on the page to parts of the address. We use an in-page location keyword search. We interpret the given address as a small subsection of a geographic thesaurus. This allows for generalization of an address by traversing its hierarchy for keywords. We can broaden the search from full street address to only city or region. An address such as "*Escherweg 2, 26121 Oldenburg*" could be matched in decreasing levels of relevance as "*Escherweg, Oldenburg*" or "*Oldenburg*" alone. This truncation matching is already useful for searching the page, but absolutely necessary for searching features such as metatags, page title, or image attributes. They are very unlikely to contain full addresses, but will often contain at least the city name. This method also allows for more general approaches to describe location than just addresses. In theory, arbitrary topics could be searched this way. Matches of this search are rated by the amount of matched keywords and—where applicable—the distance from the image based on the DOM-representation of the document. The process would repeat for each location associated to a Web page.

The examined image tag attributes are *alt*, *title*, and *name*. These descriptive fields for an image may contain descriptions of the image contents and are therefore highly relevant as well as the *src* attribute for meaningful image or path names. The page's metatag fields *description*, *keywords* and *page title* as well as Dublin Core *DC.Subject*, *DC.Description* can hint that the page as a whole is concerned with the location and therefore contained elements receive a higher rating. Other well-known tags such as dc.coverage, geo.location or icbm coordinates were not encountered during our crawls and were therefore not included at this time. The set of evaluation features is presented in the following.

- *image-attributes* Checks whether a keyword appears within a textual, descriptive attribute of the image tag such as *alt*, *title* or *name*. These fields of an image may contain descriptions of the image contents and are therefore highly relevant.
- *DOM-Distance* Checks whether identified address keywords appear near the image on the page by using a DOM-representation of the HTML source. This decreases by distance to give text near the image a higher rating. Within many of the other systems discussed in Section 1.2, this is their only or major source for a similarity measure of keywords and image.
- *Source-Path* Checks the *src* attribute of an image for meaningful image file names or, to a lesser degree, path names.
- *Page title and metadata* Checks the page's metatag fields *description*, *keywords* and Dublin Core *DC.Subject*, *DC.Description*. This would mean that the page as a whole is concerned with the location and therefore the parts on it get a higher rating. These fields were selected for their possible general location relevance. Other well-known tags such as dc.coverage, geo.location or icbm coordinates were not encountered during our crawls and were therefore not included at this time.

Combining features for location assessment

The process shown before in Figure 1.3 depicts two logical blocks for location assessment, photograph classification and location classification, each realised by an analysis chain. Inside each chain, several filters serially evaluate the aforementioned features, calculate a relevance score, and add an annotation for the assessed image.

With this partition into two processing chains it quickly becomes clear that this might not be the most efficient arrangement for use within a search engine, especially when the wanted images are rather rare. For an applicability within a resource-constrained search engine, it is desireable to reduce the processing time, which is induced mainly by the need for analysis of features from the image file itself. Since we currently have no further need for images beyond those with a derived location, the solution is to reduce expensive image operations. The most expensive of these is the I/O overhead induced by the necessary download of images and to a lesser degree some image analysis steps such as face detection. However, for most analysis steps, the download and decoding of images is the major part of processing time. Therefore, the ability of the system to filter out unwanted and unpromising images has to be improved. This is achieved by adding a further result state to individual filters. While for the overall system, only the combination of multiple

Table 1.1: Combination of evaluation features

	stage, source		*scope*		*classification*		
feature	*pre-dl (html)*	*post-dl (image)*	*image*	*page*	*location*	*photo*	*veto*
image size/ratio	●	●	●			●	●
size comparison	●	●		●	(●)	●	
filename occurrence	●			●		●	●
filename similarity	●			●		●	
img-attributes	●		●		●		
DOM-distance	●		●		●		
source-path	●		●		●		
metadata	●			●	●		
animation/transparency		●	●			●	●
color count/histogram		●	●			●	●
EXIF-data		●	●			●	
edge detection		●	●		(●)	●	
face detection		●	●		(●)	●	

feature classifiers can provide reliable estimates on the positive classification for photograph or location, some features have the ability to directly judge an image as negatively classified for the overall chain. That is, a filter can have its own threshold below which the image is considered a non-match for the whole chain, making further processing unnecessary. Such filters are thus equipped with a veto ability for the remainder of the filter chains. This allows for a fast veto on unpromising images, freeing up performance for further processing.

Using the veto mechanism, the filter chains are rearranged from their logical arrangement of photograph and location classification into pre- and post-download operations on the images. Within each group, filters are ordered by their veto ability, discriminative ability, and computational expense. The resulting properties of the feature filters and their order is shown in Table 1.1. The stage column denotes whether a filter works before or after the download of images, i.e., on the html page or the image file; the scope denotes whether the filter works on image or Web page properties; and the classification indicates whether this goes into the location or photograph classification. The veto field indicates whether a specific filter can cast a veto on the processing. Note that the image size filter can be run twice, as often embedded images do not carry size attributes in the img-tag and size can only be evaluated when images are available after their download.

With the reordered chains, the location assessment is still mostly decoupled from the photo classification and run in the pre-download stage. Thus, if the evidence for location lies below a threshold, it is possible to abort the processing chain before the

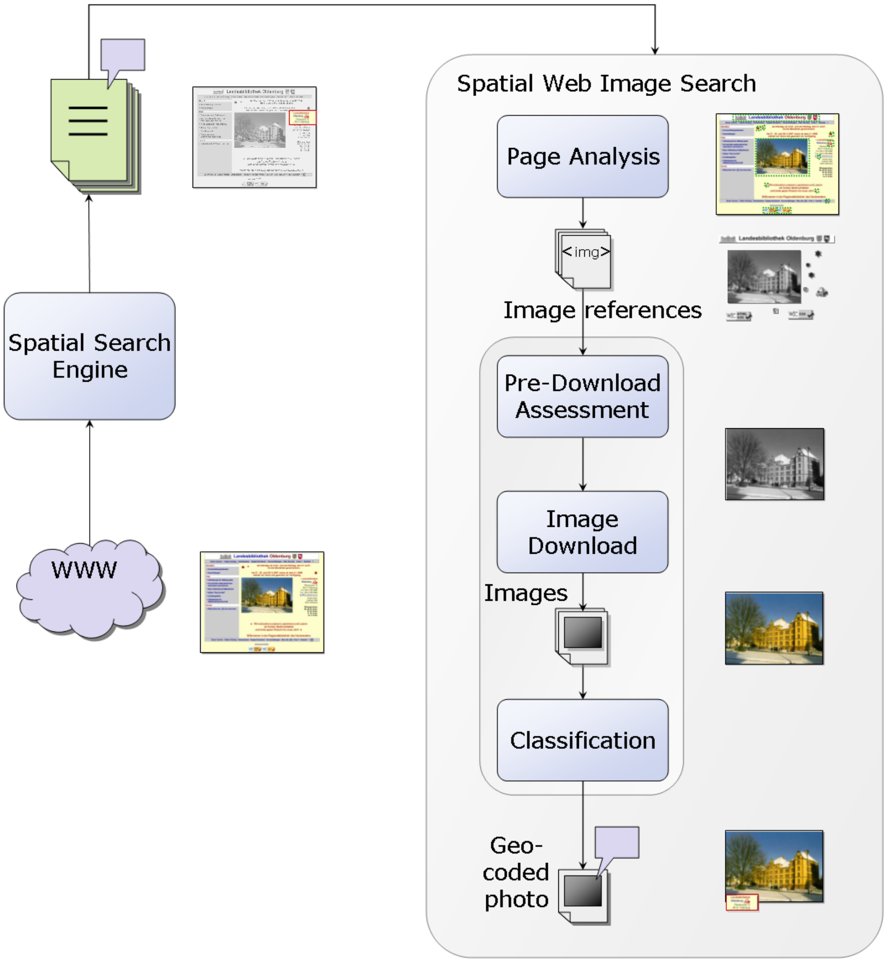


Figure 1.4: Optimized process of the location assessment

download as well, further improving efficiency. Furthermore, a cache of downloaded image's analysis results is kept so that once downloaded images can be reused if they appear in other pages again. Also, vetoed images are identified in subsequently analyzed pages and trigger a shortcut towards a faster veto. The optimized process architecture with an arrangement of pre- and post-download filter sets is shown in Figure 1.4. The page analysis extractes all image references from the page and feeds these to the pre-download operations, where unpromising images can be filtered out by their context features. The remaining image references are then downloaded, resulting in actual images which undergo the content.based classification steps and a relevance score evaluation. When the classification suceeds, the process results in geocoded photographs.
]

Formally, the assessment is defined as follows. Two actual chains *location* and *photo* are installed for location assessment and photograph detection respectively. These span the whole process, independent of the stage, as defined in Table 1.1. Let n be the number of feature filter modules in the chain. Each filter module $evaluation_i$ is assigned a weight w_i to express its relevance relative to other modules and allow for easy parameterization and tuning. The relevance score for an image *img* per chain is calculated by summing up the evaluation scores for each feature and noting potential vetoes:

$$relevance(img)_{chain} = \sum_{i=1}^{n} w_i \; evaluation_i(img) \tag{1.1}$$

$$veto(img) = \exists i \in 0..n \colon evaluation_i(img) = veto \tag{1.2}$$

For each chain, the final relevance score of an image is compared to an empirically determined threshold $threshold_{chain}$ to decide whether it is in or out of scope after checking for a veto:

$$pass(img)_{chain} = \begin{cases} 0 & veto(img) = true \\ 1 & threshold_{chain} \leq relevance(img)_{chain} \\ 0 & otherwise \end{cases} \tag{1.3}$$

An image is classified as location-relevant if it passes both chains; its overall location score is given as $relevance(img)_{location}$ and the relation between image and location is established with the given relevance score value.

$$locRelevant(img) = pass(img)_{location} \wedge pass(img)_{photo} \tag{1.4}$$

Separating the classification into location classification and photograph classification allows setting different thresholds to pass the chain for the location classification and the subsequent photograph classification. This can be seamlessly integrated with the aforementioned separation into a pre- and post-download stage by simply collecting evaluation scores along the process and doing the calculation of the overall relevance at the end.

In the evaluation section, we will present the experimentation with different thresholds and discuss the quality of the results for different threshold values. The initial thresholds were determined experimentally with a known testset of images using the abovementioned features. During the experiments, we allowed each image to pass through each evaluation module to gather complete data. For later use, the veto mechanism and thresholds introduce efficient truncation for images which are rated well below passing thresholds in the assessment chains. With the location relation thus established, the image is associated with the location on the Web page and is accessible to spatial Web image search.

1.5 Evaluation

To evaluate our approach and assess its validity, we ran a series of test crawls on our system. These served to answer the questions of first, how reliable the photograph-classification is, and second, whether the overall location reference is correctly assigned. As a testset we chose a crawl of Rügen, Germany's largest island and a famous tourist area, and of the city of Oldenburg, Germany. The resulting testset comprises Web pages along with the addresses that were identified on them. Both regions deliver similar results, but the rural region of of Rügen has a slightly different composition of image types. We show the performance of photograph classification on a smaller testset and then present the location classification in a second step.

Evaluation of photograph classification

To create a small testset for the photograph classification we used a random sampling from our spatial index of 100 Web pages with an identified location relevance. Out of these, 78 pages could be used as 22 pages remained unreachable or had parsing problems. Within the remaining pages, we found 1580 images. Our photograph classification resulted in 86 images classified as photographs. Visual inspection of these images revealed that 69 were classified correctly. The remaining 17 images were misclassified graphics which still showed some of the relevant photographic features. 1538 images were classified as non-photographs and mainly comprised decorative items of Web pages. We also went through a visual inspection of these images and identified 1511 as classified correctly. The 27 falsely classified photographs contained only few colors or had untypical sizes or ratios such that their score fell beyond the threshold. The fact that we only found very few photographs is not surprising, as on the Web nowadays quite a lot of images are used for decorative purposes. This still means that $\approx 70\%$ of pages contained a photograph. Within this small testset, the achieved values for $precision \approx 0,80$ and $recall \approx 0,72$ are promising.

Evaluation of location classification

To draw meaningful conclusions for the location assessment, two larger crawls were performed. For the Crawl of Rügen, the testset was enlarged to 1000 pages. Of these, 857 pages could be examined. The result set of photographs confirmed a similar performance as before, but was not manually assessed. A resulting 618 combinations of photos with locations were analysed. Of the 74 images classified as having a location relevance we found 55 classified correctly and 19 with dubious or false results. Of the 544 images classified without location relevance, we have no exact values for the ratio of wrong classification. However, scanning these images seems to confirm that most of them are correctly classified; we estimate recall at above 0.7 and can calculate $precision \approx 0,7$ at a similar value. Examples of the images retrieved by our system are shown in Fig. 1.5. We find some well-related images *a-d*, some even from the interior of buildings, especially for holiday homes. However, some false positives and mismatches are also returned. Images *e* and *g* depict a location, but the connection to the found location is uncertain. For *g* we can assume that it is just a general landscape of the island. Some uncertainties could only be solved by in-depth linguistic analysis of the pages, some are unsolvable even by a human user. Clearly mis-classified is the illustrative graphic *f*. The sketch of a chapel *h* shows good location relevance but is not a photo. This shows an interesting capability of our system: The major part of images falsely classified as photographs were rated lower for location relevance than true photographs. Apparently the location classification compensates for some errors in the photograph detection.

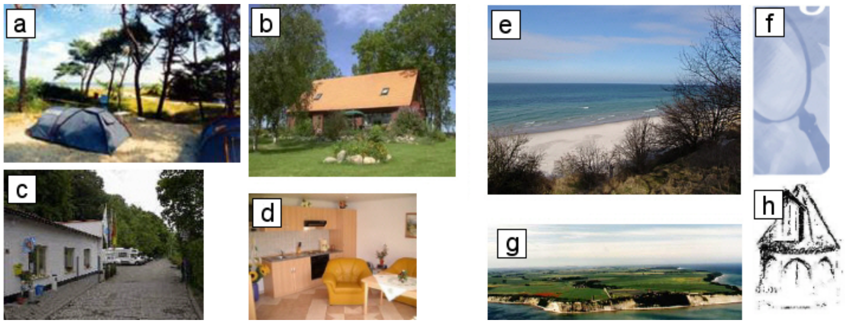


Figure 1.5: Retrieved images Rügen, sample true positives (a–d) and false positives (e–h)

The results from the crawl of Oldenburg show similar results. A noteable difference is first the higher variance of images that were retrieved, manifesting in much more urban buildings instead of landscapes. Second, the images, especially the false positives, have a much higher variance in their types, ranging from actual buildings and interiors to products, people, or logos. Figure 1.6 shows some characteristic results. Images *a–d* and *f* show broad exterior shots of buildings, two of of them museums *a* or banks *f*, other smaller houses of pharmacies or workshops. the latter is also shown in a more close-up shot in *g*. Finally, *g* is the interior of a church. The false positives are rather varied, with *h* a collage from a museum, *m* a lasershow from a respective craftsman. *j*, *n* and *o* are general illustrative photographs. Interestingly, *i* shows a sports club in front of a landmark building and *l* shows another landmark, albeit on a CD cover. Finally, *k* shows a harbour location in a different city.

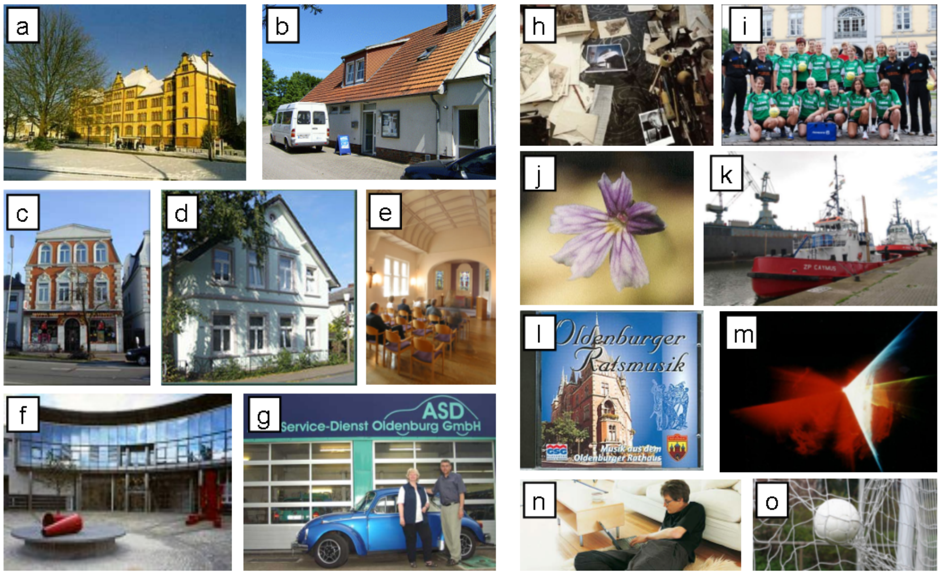


Figure 1.6: Retrieved images Oldenburg, sample true positives (a–g) and false positives (h–o)

Performance evaluation

The performance evaluation was conducted for the crawl of Oldenburg, where more pages were made available. From a total number of 16,072 URLs, the system extracted 629,927 image references. Of these, 458,415 (73%) received a veto in the pre-download stage, a further 164,505 (26%) were vetoed afterwards. That left only 7007 images, a mere 1% of all image references, to receive an evaluation score. Manual random sampling showed no false negatives in the vetoed images but we cannot exclude that case completely. A run-time analysis was made to examine the impact of individual filters. It was seen that the arrangement of filters within a stage has a very low impact, as most performance is consumed by decoding an image. Only the edge and face detection run slightly slower that other filters. However, this shows that the reordering into a two-stage-approach is very effective and can reduce the amount of downloaded images by over 70%.

Discussion of test sets and results

The evaluation shows that the approach is rather promising. However, some issues were encountered. The test set for the island Rügen covers a touristic but rural region. This leads to the problem that a lot of interesting sites do not have any kind of address and thus are not retrieved by the geospatial Web search in the first place. The famous chalk cliffs are an example for this difficulty. Furthermore a lot of images depicting Rügen are pure landscape-photos, which makes location validation difficult. The testset for the urban region of Oldenburg exhibits different images where a lot of places actually have an address and are present in the result set. In this more urban environment most addresses lie within the city, while in a rural area, addresses can be in the countryside, leading to two different sets of photographs to be expected; one showing more buildings, the other more landscape-oriented photos. Our results show a slight decrease in reliability of our photo-detection for urban regions. By examining the filtered images, we recognized that the images from a more urban area contain much more graphics which are produced in a way that makes them more similar to photos. Another issue is the composition of the test set. It contained a lot of classifieds directory sites which mostly have no photos embedded, but often sport entensive graphics.

The system also discovers a substantial amount of maps on Web pages detailling location and directions for the referenced place. Other images that we aim to filter out, but which sometimes still appear in the results, comprise logos, products, people, etc. This can be further improved by respective filters. More difficult are some cases where the image does show a location, but it remains unclear whether it really shows the location present on the page. In a few cases, the location of the page and the one shown in the image are a definite mismatch, as evidenced, e.g., by mountains on the seaside. However, we were also able to recognize an interesting capability of our system: The majority of image falsely classified as photographs were rated with a lower score for location relevance than the real photographs. Apparently our location classification can compensate for some of the photograph classification errors and vice versa. This furthermore confirms our initial assumption that location and corresponding photos already relate to each other on the Web. This seems at least in part be due to the fact that page creators tend to use them in a certain way. In conclusion, the results and our chosen features show that context analysis for Web images in combination with content analysis can deliver useful results for detecting location-bearing photographs.

1.6 Future developments

As the results of the evaluation revealed, it remains challenging to determine if a photo actually carries a location-relation. Still, the results and our chosen features show that while context analysis for images alone is not sufficient, in combination with basic image content analysis it shows promising results in detecting location-bearing photographs. From this achieved state onward, several options are available to advance the system into different directions.

Improvements

A first advancement is of course the improvement of the existing system alongs its previously defined goals, to reliably propagate a Web page's location to embedded images for place illustration. This will include refining the parameters of the heuristics, analyzing in more detail to what extent the filters each influence the results and whether this could be exploited by using machine learning approaches. Since the current approach is rather exploratory, the results could be made more soundly founded by comparing different classification results by, e.g., decision trees, neural networks, SVMs etc. This could be supported by the use of more page and image features for the classification. Since the location of a page is already known, other sources could be used to indicate the type of area this location is in. Such information could be used to validate found images against an expectation for, e.g., rural areas, landscapes, or urban buildings. Such attempts to improve the location assessment would however require the use of content-based retrieval techniques. The aim then would be to better understand the content of images and their use and role on Web pages, which leads to further improvements as discussed in the next section.

Classification of illustrative images

A second development would be to turn some of the current misclassifications into new challenges. Instead of trying to remove them, they might actually be good results if only the question were framed a little different. This requires a redefinition of the notion of a ”representative” image. A substantial amount of found images could be said to have a location relevance, but often rather because they have a very high relevance to the entity described on a Web page instead of depicting the location itself. So if the goal might not be to actually find a picture of a location but one that is characteristic for the described entity, a lot more opportunities open up. This is especially true since the actual amount of location-depicting images our approach finds is quite low.

A complimentary approach can be to extend the concept of image roles to be able to identify the type of photograph for more content-based image understanding. Some of the found images that are currently being discarded include driving directions (which still have a stong location aspect), logos, representative products, graphics design, characteristic tools, or pictures of individuals or groups These could be a valuable asset in some scenarios. For example, logos and similar images can serve as as mnemonic identifiers for entities. Similar to a favicon in a bookmark list, such images could be used to identify geospatial results on a map. These different semantics of images could be identified by image type classification using a content- and context-based hierarchical classification system.

The Web page context can be a further opportunity to feed specialized methodswith additional data. These could work on basic characteristics of the underlying Web pages, such as the type of page or the type of entity described on them such as cultural entities, landmarks, restaurants, activities, shops, etc. Since different types of pages often carry different types of representative images, such knowledge could be gainfully used in a classification. A Web page about a person could then be illustrated with that very person, a shop with its logo or defining product. Depending on the type of entity, further preferences and expectations for image types could be defined. Complementing this development, the next section examines how the original goal of finding location-depicting images can be improved.

Geospatial images

Susan Sontag once famously wrote that "everything exists to end up in a photograph". Today, one might add, everything exists to ends up on the Web. But still, the exact extraction of what the "everything" in a given photo might be and, following, where it is located, still poses strong research challenges. Some of these challenges are currently being solved, as discussed in Section 1.2. Simultaneously, a gworing number of sources for geospatial images exist. While in some cases these do not hold images georeferenced at a building level, they can still be used for a gainful advantage. First, the large amount of geospatially-enabled photo sharing sites such as flickr, locr, panoramio etc. can be used to find images from the position of interest and to try to match them to the current location and to a larger dataset, possibly identifying established logos or names of entities within some pictures. Another interesting source are large-scale high-resolution mappings of cities to retrieve images for pages without photographs. For example, Google Street View offers high-resolution street-level imagery in areas where coverage is available. Other navigation vendors have similar data, but usually do not release it. It can also be used to indentify individual buildings and could be a suitable source to depict a building, museum, or landmark. For that, the individual boundaries of buildings and their spatial extent must be known and these then annotated in the image. At lower resolution, building or facade reconstruction from the oblique view of bird's eye imagery has been undertaken in several projects and can also provide low-resolution building and environment images. A use in car navigation systems could even benefit more from an abstract view of buildings for better visual integration. However, indoor images could not be retrieved this way and we remain dependent on the business owners to provide them on their pages. We were able to find an interesting set of such images, from recording studios over hotels or workshops to libraries or churches.

These are just some examples to demonstrate how rich datasources could be employed to retrieve spatially-enriched images. The main argument for the approach followed in this chapter is that the geospatial images on Web pages often better selected by the page's authors and therefore are better suited to be representative images for the page in question. However, to offset their rare occurrence, methods as described here can be employed for a more complete picture over aggregated sources.

1.7 Conclusion

In this chapter, we presented an approach to automatically geo-reference images embedded in Web pages on the address level. Based on our work on location-based Web search, we presented our method and heuristics to propagate the location of a Web page and assign it to embedded photographs. We could show that a combined multi-source analysis of content and context is feasible for this task. Automatic location detection of images on unstructured Web pages now allows images to become part of the results of a geospatial Web search without previous manual tagging. The results of our evaluation are promising regarding classification and performance. However, for borad applicability, the approach lacks sufficient data in the form of geospatial Web images. To counter this effect, we also proposed several improvements. Depending on the application scenario, other types of images or other sources of data might be considered to gather representative images for a place. For such cases, the system described here could be adapted to arrive at a deeper understanding of representative Web images and their associated location.

Acknowledgements

We thank our students Sebastian Stockfleth for the initial implementation of the image search system and Timo Scheffler for substantial improvements and suggestions. Part of this research has been carried out as subproject of the Niccimon project and was supported by the State of Lower Saxony, Germany.

References

[1] D. Ahlers and S. Boll. Location-based Web search. In A. Scharl and K. Tochterman, editors, *The Geospatial Web. How Geo-Browsers, Social Software and the Web 2.0 are Shaping the Network Society.* Springer, London, 2007.

[2] D. Ahlers and S. Boll. Retrieving Address-based Locations from the Web. In *GIR '08: 5th Workshop on Geographic Information Retrieval*, pages 27–34, New York, NY, USA, 2008. ACM.

[3] D. Ahlers and S. Boll. Adaptive Geospatially Focused Crawling. In *CIKM '09: Proceeding of the 18th ACM Conference on Information and Knowledge Management*, pages 445–454, New York, NY, USA, 2009. ACM.

[4] D. Ahlers and S. Boll. On the Accuracy of Online Geocoders. In W. Reinhardt, A. Krüger, and M. Ehlers, editors, *Geoinformatik 2009, Osnabrück*, volume 35 of *ifgiprints*, pages 85–91, Münster, 2009.

[5] E. Amitay, N. Har'El, R. Sivan, and A. Soffer. Web-a-Where: Geotagging Web Content. In *SIGIR '04*, pages 273–280, New York, NY, USA, 2004. ACM Press.

[6] Y. Aslandogan and C. Yu. Diogenes: A Web search agent for content based indexing of personal images. In *Proceedings of ACM SIGIR*, pages 481–482, 2000.

[7] V. Athitsos, M. J. Swain, and C. Frankel. Distinguishing Photographs And Graphics on the World Wide Web. In *CAIVL '97: Proceedings of the 1997 Workshop on Content-Based Access of Image and Video Libraries*, page 10, Washington, DC, USA, 1997. IEEE Computer Society.

[8] J. Baldzer, S. Boll, P. Klante, J. Krösche, J. Meyer, N. Rump, A. Scherp, and H.-J. Appelrath. Location-Aware Mobile Multimedia Applications on the Niccimon Platform. In *IMA'04*, 2004.

[9] H. Bischof. Robust person detection for surveillance using online learning. In *WIAMIS '08: Proceedings of the 2008 Ninth International Workshop on Image Analysis for Multimedia Interactive Services*, page 1, Washington, DC, USA, 2008. IEEE Computer Society.

[10] W.-C. Chen, A. Battestini, N. Gelfand, and V. Setlur. Visual summaries of popular landmarks from community photo collections. In *MM '09: Proceedings of the seventeenth ACM international conference on Multimedia*, pages 789–792, New York, NY, USA, 2009. ACM.

[11] R. Datta, W. Ge, J. Li, and J. Z. Wang. Toward bridging the annotation-retrieval gap in image search by a generative modeling approach. In *IEEE Multimedia: Special Issue on Advances in Multimedia Computing*, volume 14, pages 24–35, New York, NY, USA, 2007. ACM Press.

[12] R. Datta, D. Joshi, J. Li, and J. Z. Wang. Image Retrieval: Ideas, Influences, and Trends of the new age. *ACM Computing Surveys*, 2007.

[13] C. Frankel, M. J. Swain, and V. Athitsos. WebSeer: An Image Search Engine for the World Wide Web. Technical report, Chicago, IL, USA, 1996.

[14] B. Gao, T.-Y. Liu, T. Qin, X. Zheng, Q.-S. Cheng, and W.-Y. Ma. Web image clustering by consistent utilization of visual features and surrounding texts. In *MULTIMEDIA '05: Proceedings of the 13th annual ACM international conference on Multimedia*, pages 112–121, New York, NY, USA, 2005. ACM Press.

[15] W. Gong, H. Luo, and J. Fan. Extracting informative images from web news pages via imbalanced classification. In *MM '09: Proceedings of the seventeen ACM international conference on Multimedia*, pages 1123–1124, New York, NY, USA, 2009. ACM.

[16] Z. Gong, L. H. U, and C. W. Cheang. Web image semantic clustering. In *Proceedings of the 4th International Conference on Ontologies, Database and Applications of Semantics (ODBASE 2005)*, Lecture Notes in Computer Science, Agia, Napa, Cyprus, 2005. Springer.

[17] X. He, D. Caia, J.-R. Wen, W.-Y. Ma, and H.-J. Zhang. Clustering and Searching WWW Images Using Link and Page Layout Analysis. In *ACM Trans. on Multimedia Comput. Commun. Appl.*, volume 3, 2007.

[18] L. L. Hill. Core elements of digital gazetteers: Placenames, categories, and footprints. In *ECDL '00: Proceedings of the 4th European Conference on Research and Advanced Technology for Digital Libraries*, pages 280–290, London, UK, 2000. Springer.

[19] M. Himmelstein. Local Search: The Internet Is the Yellow Pages. *IEEE Computer*, 38(2):26–34, 2005.

[20] M. Hughes, A. Salway, G. Jones, and N. O'Connor. Analyzing image-text relations for semantic media adaptation and personalization. In *SMAP '07: Proceedings of the Second International Workshop on Semantic Media Adaptation and Personalization*, pages 181–186, Washington, DC, USA, 2007. IEEE Computer Society.

[21] A. Jaffe, M. Naaman, T. Tassa, and M. Davis. Generating Summaries and Visualization for Large Collections of Geo-Referenced Photographs. In *MIR '06: Proceedings of the 8th ACM international workshop on Multimedia information retrieval*, pages 89–98, New York, NY, USA, 2006. ACM.

[22] M. Jakob, M. Großmann, D. Nicklas, and B. Mitschang. DCbot: Finding Spatial Information on the Web. In L. Zhou, B. C. Ooi, and X. Meng, editors, *Proceedings of the 10th International Conference on Database Systems for Advanced Applications DASFAA 2005*, volume 3453 of *LNCS*, pages 779–790. Springer, 2005.

[23] R. Ji, X. Xie, H. Yao, and W.-Y. Ma. Mining city landmarks from blogs by graph modeling. In *MM '09: Proceedings of the seventeenth ACM international conference on Multimedia*, pages 105–114, New York, NY, USA, 2009. ACM.

[24] M. Jia, X. Fan, X. Xie, M. Li, and W.-Y. Ma. Photo-to-Search: Using Camera Phones to Inquire of the Surrounding World. In *7th International Conference on Mobile Data Management, MDM 2006*, pages 46–48, 2006.

[25] F. Jing, C. Wang, Y. Yao, K. Deng, L. Zhang, and W.-Y. Ma. IGroup: Web Image Search Results Clustering. In *ACM Multimedia 2006*, 2006.

[26] M. Kamvar and S. Baluja. A large scale study of wireless search behavior: Google mobile search. In *CHI '06: Proc. of the SIGCHI conf. on Human Factors in computing systems*, pages 701–709, New York, NY, USA, 2006. ACM.

[27] L. S. Kennedy and M. Naaman. Generating diverse and representative image search results for landmarks. In *WWW '08: Proceedings of the 17th international conference on World Wide Web*, pages 297–306, New York, NY, USA, 2008. ACM.

[28] M. L. Kherfi, D. Ziou, and A. Bernardi. Image retrieval from the world wide web: Issues, techniques, and systems. *ACM Comput. Surv.*, 36(1):35–67, 2004.

[29] Z. Kim and J. Malik. Fast vehicle detection with probabilistic feature grouping and its application to vehicle tracking. In *Proc. 9th IEEE International Conference on Computer Vision, Nice, France*, volume 1, pages 524–531, 2003.

[30] M. S. Lew. Next-Generation Web Searches for Visual Content. *IEEE Computer*, 33(11):46–53, 2000.

[31] J. Li and J. Z. Wang. Real-time computerized annotation of pictures. In *MULTIMEDIA '06: Proceedings of the 14th annual ACM international conference on Multimedia*, pages 911–920, New York, NY, USA, 2006. ACM.

[32] R. Lienhart. Video OCR: a survey and practitioners guide. *Video Mining*, pages 155–184, 2003.

[33] G. Lu and B. Williams. An Integrated WWW image retrieval system. In *Proceedings of the AusWeb99*, Lismore, Australia, 1999.

[34] J. Luo, M. Boutell, and C. Brown. Pictures are not taken in a vacuum. *IEEE Signal Processing Magazine*, 23(2):101– 114, 2006.

[35] T. Maekawa, T. Hara, and S. Nishio. Image classification for mobile web browsing. In *WWW '06: Proc.of the 15th Intl. Conf. on World Wide Web*, pages 43–52, New York, NY, USA, 2006. ACM Press.

[36] A. Markowetz, Y.-Y. Chen, T. Suel, X. Long, and B. Seeger. Design and implementation of a geographic search engine. In A. Doan, F. Neven, R. McCann, and G. J. Bex, editors, *WebDB 2005*, pages 19–24, Baltimore, Maryland, USA, 2005.

[37] C. Marlow, M. Naaman, D. Boyd, and M. Davis. HT06, Tagging Paper, Taxonomy, Flickr, Academic Article, To Read. In *HYPERTEXT '06*. ACM, 2006.

[38] K. S. McCurley. Geospatial Mapping and Navigation of the Web. In *WWW '01: Proc. of the 10th Intl. Conf. on World Wide Web*, pages 221–229, New York, NY, USA, 2001. ACM Press.

[39] Y. Morimoto, M. Aono, M. E. Houle, and K. S. McCurley. Extracting Spatial Knowledge from the Web. In *SAINT '03*. IEEE, 2003.

[40] E. V. Munson and Y. Tsymbalenko. To search for images on the web, look at the text, then look at the images. In *Proceedings of the First International Workshop on Web Document Analysis (WDA2001)*, 2001.

[41] F. Nack and W. Putz. Designing annotation before it's needed. In *MULTIMEDIA '01: Proceedings of the ninth ACM international conference on Multimedia*, pages 251–260, New York, NY, USA, 2001. ACM Press.

[42] M. Ortega-Binderberger, S. Mehrotra, K. Chakrabarti, and K. Porkaew. WebMARS: A Multimedia Search Engine, 2000.

[43] R. S. Purves, P. Clough, C. B. Jones, A. Arampatzis, B. Bucher, D. Finch, G. Fu, H. Joho, A. K. Syed, S. Vaid, and B. Yang. The Design and Implementation of SPIRIT: a Spatially Aware Search Engine for Information Retrieval on the Internet. *International Journal of Geographical Information Science*, 21(7):717–745, 2007.

[44] R. S. Purves, A. Edwardes, and M. Sanderson. Describing the where – improving image annotation and search through geography. In G. Csurka, editor, *Proceedings of the workshop on Metadata Mining for Image Understanding (MMIU 2008)*, pages 105–113, Setbal, 2008.

[45] T. Quack, B. Leibe, and L. Van Gool. 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*, pages 47–56, New York, NY, USA, 2008. ACM.

[46] T. Rattenbury, N. Good, and M. Naaman. Towards automatic extraction of event and place semantics from flickr tags. In *SIGIR '07: Proceedings of the 30th annual international ACM SIGIR conference on Research and development in information retrieval*, pages 103–110, New York, NY, USA, 2007. ACM.

[47] M. Sanderson and J. Kohler. Analyzing Geographic Queries. In *ACM SIGIR Workshop on Geographic Information Retrieval*, Sheffield, UK, 2004.

[48] P. Sandhaus, A. Scherp, S. Thieme, and S. Boll. Metaxa – Context- and content-driven metadata enhancement for personal photo albums. In *Multimedia Modeling 2007*, Singapore, 2007.

[49] C. Schlieder and C. Matyas. Photographing a city: An analysis of place concepts based on spatial choices. *Computational Models of Place, Special Issue of Spatial Cognition and Computation: An Interdisciplinary Journal*, 9(3):212–228, July 2009.

[50] S. Sclaroff, M. La Cascia, and S. Sethi. Unifying textual and visual cues for content-based image retrieval on the World Wide Web. *Comput. Vis. Image Underst.*, 75(1-2):86–98, 1999.

[51] A. W. M. Smeulders, M. Worring, S. Santini, A. Gupta, and R. Jain. Content-based image retrieval at the end of the early years. *IEEE Trans. Pattern Anal. Mach. Intell.*, 22(12):1349–1380, 2000.

[52] N. Snavely, S. M. Seitz, and R. Szeliski. Photo Tourism: Exploring Photo Collections in 3D. *ACM Transactions on Graphics (TOG)*, 25(3):835–846, 2006.

[53] K. Tollmar, T. Yeh, and T. Darrell. IDeixis - Image-based Deixis for Finding Location-Based Information. In *MobileHCI*, 2004.

[54] Y. Tsymbalenko and E. V. Munson. Using HTML Metadata to Find Relevant Images on the World Wide Web. In *Proceedings of Internet Computing 2001*, volume 2, pages 842–848, Las Vegas, 2001. CSREA Press.

[55] A. Vailaya, M. Figueiredo, A. Jain, and H. Zhang. Image classification for content-based indexing. *IEEE Transactions on Image Processing*, 10(1):117–130, 2001.

[56] A. Vailaya, A. Jain, and H. J. Zhang. On Image Classification: City Images vs. Landscapes. In *Proc. of the IEEE Workshop on Content-Based Access of Image and Video Libraries*, 1998.

[57] M.-H. Yang, D. J. Kriegman, and N. Ahuja. Detecting Faces in Images: A Survey. *IEEE Trans. Pattern Anal. Mach. Intell.*, 24(1):34–58, 2002.

[58] Y.-T. Zheng, M. Zhao, Y. Song, H. Adam, U. Buddemeier, A. Bissacco, F. Brucher, T.-S. Chua, and H. Neven. Tour the world: Building a web-scale landmark recognition engine. *IEEE International Conference on Computer Vision and Pattern Recognition (CVPR)*, 0:1085–1092, 2009.