Image Retrieval at Memory’s Edge: Known Image Search based on User-Drawn Sketches

Michael Springmann  Ihab Al Kabary  Heiko Schuldt
Databases and Information Systems Group | University of Basel, Switzerland
{michael.springmann|ihab.alkabary|heiko.schuldt}@unibas.ch

ABSTRACT

With the increasingly growing size of digital image collections, known image search is gaining more and more importance. Especially in collections where individual objects are not tagged with metadata describing their content, content-based image retrieval (CBIR) is a promising approach, but usually suffers from the unavailability of query images that are good enough to express the user’s information need. In this paper, we present a system that provides CBIR based on user-drawn sketches. The system combines angular radial partitioning for the extraction of features in the user-provided sketch, taking into account the spatial distribution of edges, and the image distortion model. This combination offers several highly relevant invariants that allow the query sketch to slightly deviate from the searched image in terms of rotation, translation, relative size, and/or unknown objects in the background. To illustrate the benefits of the approach, we present search results from the evaluation of our system on the basis of the MIRFLICKR collection with 25,000 objects and compare the retrieval results of pure metadata-driven approaches, pure content-based retrieval using different sketches, and combinations thereof.

Categories and Subject Descriptors

H.3.3 [Information Storage And Retrieval]: Information Search and Retrieval—query formulation, retrieval models; H.5.2 [Information Interfaces and Presentation]: User Interfaces—input devices and strategies

General Terms
Design, Human Factors

1. INTRODUCTION

Searching for information can be categorized in two fundamentally different classes with respect to the knowledge a user has about the information item(s) and also in the condition under which the search task can end successfully. Searching for known items: a user knows that the items exist and has probably seen them before; the search task will end successfully only if the user has found all these items. Searching for (potentially) novel items: a user is looking for items that satisfy her information need; the search task ends successfully as soon as the user is provided with some items that reflect the information she is looking for.

Depending on the interaction intentions [7] associated with these classes, different information seeking strategies have to be applied. The probably most effective strategy for finding a known image in an image collection is to remember an access path to it. This of course only works for static collections in which the access path is kept stable over time.

A shift in information seeking strategies [7] to the strategy of browsing can be applied for known item search, but it is certainly not as effective and may also not lead to a successful result if it is not known in advance whether or not the collection contains the searched item.

For finding both, known images and novel images, content-based image retrieval (CBIR) provides powerful tools. However, CBIR requires a query image to start with that is sufficiently close to the final result, i.e., that precisely expresses the user’s information need. Without such a query image, it is difficult or nearly impossible to achieve good retrieval quality, even if sophisticated relevance feedback mechanisms are available. Query by Sketching addresses this problem and takes user generated sketches as query images. So far, two main problems have significantly impacted the successful application of query by sketching to known item search and novel item search. First, the mouse as most widely available input device limits the user-friendliness and expressiveness for drawing sketches. Second, users usually do not sketch complete images but concentrate on the parts which are most interesting for them. Thus, when comparing user-drawn sketches with images, there will usually be parts of the images to be queried that do not have any corresponding part in the sketch as the user may not remember or not be able to draw all details of image in the sketch. The user may also place the sketch not exactly at the right position, with proper scale, and/or orientation. Therefore, the corresponding parts may not be at the same coordinates in sketch and image. In order to successfully apply query by sketching to CBIR, both problems need to be solved jointly.

In this paper, we present the QbS approach to query by sketching that exploits novel user interface technologies and supports various invariances in the comparison of a user-generated sketch and the images to make the query process robust against translations, rotations, and different scales.
We have applied the results of our work to use cases from known image search and provide detailed evaluation results of the QbS system based on the MIRFLICKR-25000 image collection [3]. A more detailed description of the system and the evaluation can be found in [5].

The paper is organized as follows: Section 2 discusses related work. In Section 3, we introduce the QbS approach and its implementation. Section 4 summarizes the results of the experimental evaluation. Section 5 concludes.

2. RELATED WORK

As the absence of a good example to start a query has been a problem from the very start of CBIR, query by sketching has always been of interest in this domain. However, it has not been as successful as other approaches to CBIR, mainly due to unavailability of appropriate input devices. A more detailed discussion of existing approaches is presented in [5].

Our approach is based on Angular Radial Partitioning (ARP) proposed in [1] which uses spatial distribution of edges. We extended ARP to support additional invariances.

The image distortion model has been used for optical character recognition and automated classification of medical images [4] for which it uses images in reduced resolution. Deformations are allowed within the warp range for individual pixels and the area around them (local context). For our approach, we added IDM as a distance measure which is able to measure much finer deviations than ARP. However, we replaced the Sobel filter for edge detection with the same variant of the Canny edge detector used for ARP.

3. THE QBS APPROACH

QbS provides an application that can be used in an interactive setting to retrieve known images. The prototype [6] runs on a Tablet PC as shown in Figure 1, thus allowing the user to draw edges with a stylus directly on the screen.

The QbS system provides a fast query mechanism to retrieve some of the top-ranked results to give the user a fast feedback about the retrieval result that can be expected from the final or full result. Then, the user is able to add further details to refine the search and remove misleading parts of the sketch without having to redo the complete sketch.

To generate edge maps [1] from the images, a variant of the Canny edge detector is applied. It takes a single threshold, $\beta$, to control how many details are preserved. As a single value of $\beta$ might not be sufficient to cope with different lighting conditions or depth of field of different images in a collection as well as the level of detail in the user sketch, we extract edge maps from the images at several values between 2 and 50 as shown in Figure 2.

We incorporate two different sets of features and corresponding distance measures: Angular Radial Partitioning (ARP) [1] is used as a compact, fast way to retrieve images, when rough sketches and spatial layout is sufficient to separate the desired known item from all other images in the collection. An adapted version of the Image Distortion Model (IDM) [4] on the edge maps is used as a more complex, computationally more expensive solution whenever the user needs more thorough comparison between a good-enough sketch and the other images. In addition, the user can in both cases restrict the comparison to images selected on their metadata, e.g., only consider images with certain tags which have been indexed using Apache Lucene.

3.1 Similarity using ARP

We support ARP at various resolutions. By default ARP is used with 8 angular and 4 radial partitions (as depicted in Figure 2(e)) in which the number of edge pixels inside sketch and edge map are counted, which will generate a 32 dimensional feature vector. The user is given the possibility to distinguish areas that shall remain free of edges or areas where the content is unknown (e.g., the user does not remember the content). Internally this is handled by comparing the vectors using a weighted Manhattan distance: For areas which should get treated as “unknown”, a weight of zero is assigned; all non-empty partitions have a weight of one – as well as areas which are intentionally left empty.

ARP can handle small misplacements and rotation, as long as not many edge pixels are moved from one partition to another. If the user expects even more rotation, invariance to rotation is provided by applying the 1D FFT on the features as proposed in [2].

For translations that are further off the position of the sketch, a heuristic can be applied: The original image is
cut into several regions that still cover enough pixels to give meaningful search results. In addition to the image as a whole and a bounding box covering all the area with edge pixels, we also extract slightly smaller regions shifted towards each of the image corners. During search, the sketch will get compared with all regions, thus compensating to a certain extent for translation. As most of these regions are smaller than the full image, comparing the sketch or parts of the sketch can also provide some invariance to scale. Two additional regions in the center, one horizontally with a wide aspect ratio and one vertically with a narrow aspect ratio, can assist when the scaling does not maintain proportions.

Features of these multiple sub-images with and without 1D FFT are stored separately for each chosen $\beta$ value. For simple, regular searches the query feature vector from the user’s sketch will be compared to exactly one feature vector for each image in the collection. Invariant searches will compare the distances with all corresponding representations, but only select the best-matching version of the image features for an individual image to compute its distance.

### 3.2 Similarity using IDM

For IDM, the edge map image is scaled down to smaller sizes. As default size, we use 32 pixels for the longer side. As the user is expected to draw the edges in the sketch, edge detection is never applied to the sketch. This means that the scaling has to be performed after the edge detection - and it is preferable to use an interpolation for scaling down and to apply thresholding afterwards to return to a binary image (edge / non-edge) rather than not using interpolation in scaling, which may drop a lot of the edges.

The warp range can be defined at query time. Therefore the user can specify how many pixels (in the reduced resolution) the edges are allowed to be misplaced by translation, scaling or rotation. Big warp ranges thus result in higher invariance. The size of the local context defines whether individual pixels (local context of 0) or patches (local context > 0) are matched. Big local contexts result in small invariance as bigger patches must fit to achieve low distances.

Invariance w.r.t. the value $\beta$ used in extraction is achieved with the same strategy as for ARP. Again, unknown areas can be handled by ignoring non-edge pixels in the down-scaled sketch. They only affect the distance score as part of the local context of an edge pixel.

### 4. EVALUATION

We use the MIRFLICKR-25000 dataset [3] for our evaluations, a collection with 25,000 images that were downloaded from flickr.com including the tags assigned to the images.

#### 4.1 Preparation

We have performed known item search on 4 different images that have different characteristics with respect to difficulties in drawing a sketch of them.

Figure 3(a) shows a single object located in the center of the image with rather homogeneous background. Therefore at a value of $\beta = 7$, none of the clouds contribute to edges anymore while almost all aircraft details are still available and users can follow the common intuition to draw just the desired object directly in the image center. Moreover, the bounding box can compensate for translation and scaling.

Figure 3(e) is more challenging as it contains several distinct objects (e.g., houses, mountains). There are plain or homogenous areas like the sea or the sky as well as areas with strong contrasts like the rocks and vegetation on the mountains. The spatial distribution may lead some users to place the coastline too high/low and not to match the aspect ratio and relative sizes of objects.

Figure 3(i) contains again one main object, but this time, it is difficult to separate it from the image background. As can be seen in Figure 2, there is no value of $\beta$ at which the details of the bike would still be preserved while the floor and walls would already disappear. The image perspective is also non-trivial as there is no planar view on the bike.

Finally, Figure 3(m) contains several prominent objects (hand, figure of the Eiffel tower) which are neither placed directly at any image border nor directly in the center of the image. No major visible lines are either horizontal or vertical, but slightly rotated. Significant parts of the image are blurred, but colors still differ too much to generate areas which would appear to edge detection as homogenous areas.

For the evaluation, we have collected at least 10 different sketches for each of the 4 images from a group of 7 people (54 sketches in total; some examples shown in Figure 3). The individuals were given time to familiarize with Qo5 system running on a Tablet PC as shown in Figure 1 and were then requested to search for the known items which were shown to them in printed form. For each image, the participants were also shown the tags associated on flickr and
asked, which they would use to search; the two tags selected most frequently were used in the evaluation.

4.2 Retrieval Quality

We measured the individual performance of ARP and IDM. Figure 4 shows the ranks split up by known item. The ends of the bar show the rank achieved by the best and the worst sketch for this image; the thick grey bar starts at the first quartile and ends at the third quartile and indicates therefore the range of ranks which half of all sketches achieved; the blue diamond indicates the mean or –in other words– the average rank of all sketches for this image. Notice that the rank axis is in logarithmic scale.

Only a single sketch had a rank slightly worse than 1000 while the vast majority of sketches achieved a rank below 100, which could already be a number of results a user might be willing to browse. For searching with IDM even the majority is below 10, a number of results that can easily fit on a single result screen and out of which the user would recognize the image almost instantaneously.

For all searches, the choice of the appropriate $\beta$ value depends on both the known item and the user sketch. Therefore it is not possible to come up with a single value of $\beta$ for a diverse collection like the MIRFLICKR benchmark that would suit all users. However, $\beta S$ defines a set of values with reasonable limits and lets the user not only pick a single value, but also a range from this set that he can refine and shift during the search process.

In Table 1 we compare the average ranks achieved by pure content-based retrieval using ARP and IDM with the ranks when combined with keyword search and also with the rank where Lucene would return the item when only the keyword search is used. The column Search Space contains the number of objects that pass the filter (in case of plain CBIR: all images of the collection; in case of a text filter, the number of hits). Notice that 8.51% (2,128 out of the 25,000 images) in the dataset do not have any tags associated to them, so using text-search alone or relying entirely on keywords to be available would not give satisfactory results to the users. More details including the execution times for searches can be found in [5].

5. CONCLUSIONS AND OUTLOOK

We have presented the QbS approach that assists users in searching for known images on the basis of user-drawn sketches, which includes several invariances that allow a user to find the desired result despite of missing background information, misplacements, and/or differences in scale or rotation. Furthermore, the QbS system has been evaluated on the MIRFLICKR-25000 benchmark collection and the results show that query by sketching can in fact help the user to find known images.

We are currently extending the QbS system to also incorporate color information and exploit novel interface types (digital pens and interactive paper).

6. ACKNOWLEDGMENTS

This work was supported by the Swiss National Science Foundation (PAD-IR, contract No. 200020_126829 / 1).

7. REFERENCES