Hub, Adrian; Blank, Daniel; Henrich, Andreas; Müller, Wolfgang

Picadomo : Faceted Image Browsing for Mobile Devices

Date of secondary publication: 03.03.2025

Accepted Manuscript (Postprint), Conferenceobject

Persistent identifier: urn:nbn:de:bvb:473-irb-1068773

Primary publication

Hub, Adrian; Blank, Daniel; Henrich, Andreas; Müller, Wolfgang (2009): Picadomo : Faceted Image Browsing for Mobile Devices, in: 2009 Seventh International Workshop on Content-Based Multimedia Indexing, Piscataway, NJ: IEEE, pp. 249–254, doi: 10.1109/CBMI.2009.34.

Picadomo: Faceted Image Browsing for Mobile Devices

Adrian Hub, Daniel Blank, Andreas Henrich
Chair of Media Informatics
University of Bamberg
D-96052 Bamberg, Germany
e-mail: firstname.lastname@uni-bamberg.de

Wolfgang Müller
EML Research gGmbH
Villa Bosch
D-69118 Heidelberg, Germany
e-mail: wolfgang.mueller@eml-r.villa-bosch.de

Abstract

Picadomo combines content-based image retrieval and faceted search on mobile devices. It is designed for finding images with desired visual properties, tags or other known metadata. Due to the limitations of mobile devices such as small screen sizes and low processing power, we had to carefully select the features that come in use (dominant color, GPS data, tags, etc.). With Picadomo the user can pick visualized facets directly via touch screen, while using very little screen size for the facet browsing navigation. We present our architecture, the facets used for image browsing, our new control concept and user experiments.

1. Introduction

More and more mobile devices are equipped with powerful, high-resolution digital cameras and memory cards are getting cheaper. So, people tend to use their cell phone as everyday camera and store a huge amount of photos on their device. Unfortunately, a cell phone is usually not designed to manage thousands of photos and in particular conveniant image retrieval facilities are missing. This problem is intensifed by the fact that, due to the storage capacity and the usage patterns formed by digital natives for mobile devices, huge image collections on mobile devices are common nowadays. Showing images to friends or even transferring images to devices of friends has become a frequent task. But, applications allowing for the search of a target image in mind that is concealed in the huge image collection seem to be missing. Therefore, designing powerful and user-friendly, image retrieval applications for mobile devices is a challenging topic.

We argue that faceted, contend-based image browsing (explained in more detail below) is well suited for exploring large image collections on mobile devices. The combination of tag-based search, search techniques based on EXIF data (GPS and time information) as well as Content-Based Image Retrieval (CBIR) generates a browsing experience and gives a powerful tool to the user for fnding desired images on a mobile device. The approach integrates easy to understand visual features into a search paradigm traditional web users frequently employ (e.g. faceted text-based search on websites such as *http://www.ebay.com*). As our experiments indicate, visual faceted search can be adapted to mobile devices concentrating on the most important facets and using an intuitive browsing interface.

A facet in the context of visual faceted search can be for example the most frequent color, the contrast level or the date the image was taken. As a search example, just imagine that you are looking for images of your pets to show them to your friends. You do not remember when you have taken these images, but you have a rough imagination of how the image looks like. Your dog was playing in the garden, so there must be some shade of green in the lower part of the image. Or, your cat was sitting on the red carpet, so there will probably be some red in a corner of the image. You will be able to fnd images that have given visual properties. In addition, being able to choose images e.g. taken at night and at a weekend surely could help fnding the desired images.

We presented VisualFlamenco, i.e. visual faceted search designed for desktop PCs, in [1]. In VisualFlamenco, a huge amount of facets were displayed on the screen together with the result set. As a consequence, even with a resolution of 1280×1024 pixels, the result page was perceived as being overloaded. To overcome this issue, we reduce the number of facets in Picadomo by only choosing the most promising facets according to the evaluation in [1]. A second drawback of VisualFlamenco was the issue that during navigation users had to click various buttons and links in different areas of the screen. Within the design of Picadomo, we therefore present only relevant information to the user (which facets can be selected, which facets are selected at the moment, etc. together with the result set). The user can quickly select or deselect facets using the touch screen always focusing on the small screen with a resolution of 480×320 pixels. As every selection as well as deselection of facets generates immediate feedback, a browsing experience on mobile devices can be achieved.

This paper is organized as follows: In the next section we briefy describe related work, in Section 3 we discuss the facets, the architecture and the control concept used in our prototype Picadomo[1]. Section 4 describes our user experiments. Finally, Section 5 summarizes our work and points towards future research.

2. Related Work

Most of the existing applications on mobile devices browse images only by folder, some applications allow the user to assign tags or notes to images[2] and few let the user classify their photos into categories or albums[3].

The *Zurfer* prototype [2] offers a channel metaphor (channels can be e.g. *Nearby, Contacts, My Stuff, Interestingness*), where the user can browse up and down between channels, and right left to view photos in any one channel with the 4-way navigation key on the mobile phone. This work also presents a study of mobile image application use.

Other approaches additionally show selected EXIF data beside the corresponding photo, but there is no possibility to browse the photos by one of the EXIF features. Being able to do so and to browse also by visual properties is the goal and unique feature of Picadomo.

The administration of personal photo collections on mobile devices has been addressed in various other work. Some of the work that focuses on visualization and interaction uses time information as a main ordering criterion [3]. In our work, time information is only a single aspect within the comprehensive supply of facets.

Zhang et al. [4] present an approach that generates an image browsing strategy that is derived from an image attention model. Certain important details within a single image are identifed by the model and automatic browsing is achieved by presenting these details to the user. Whereas we focus on the browsing of image collections, a strategy like the one presented by Zhang et al. could be integrated into our system for intra-image browsing.

Within this section, we only describe some of the existing applications that deal with the administration of image collections on mobile devices. Mor Naaman et al. give a more detailled overview in [2].

3. The Picadomo Prototype

Our prototype explores the combination of visual and semantic features based on metadata with a new interface, which was designed to be used on mobile devices. We realized Picadomo for the Android platform[4] as a stand-alone application, i.e. automatic feature extraction from

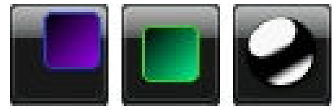


Figure 1. Three example visual facets for Picadomo. Left: Violet should be the dominant color in the upper right of the image. Center: Most pixels in the center are to be green. Right: the contrast of the image should be rather high.

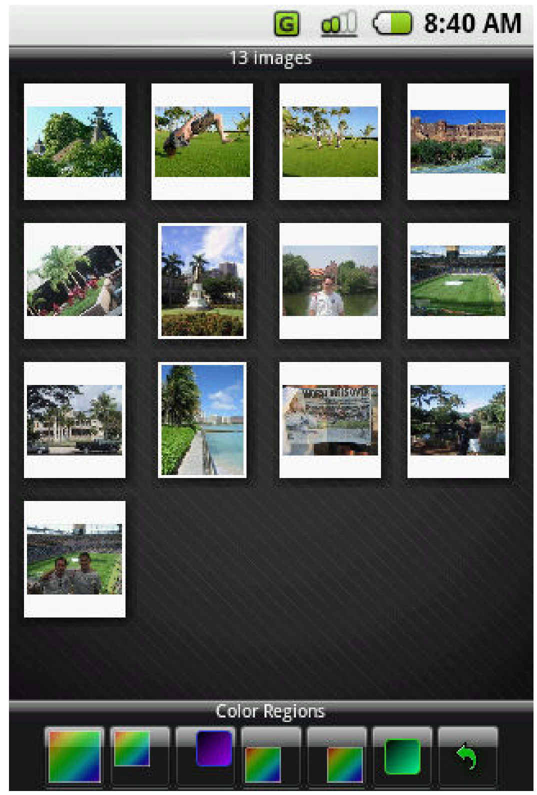


Figure 2. Searching our collection with two of the facets given in Fig. 1 (dominant color is specifed for the upper right and the center part of the image).

images can be done by the mobile device. Picadomo extracts low-level features, as well as EXIF data from JPEG images and stores these features as facets in a database on the mobile device. See some facets in Fig. 1 and a screen-shot of our prototype in Fig. 2.

3.1. Facet Description

Some of the facets are the same as in our prior work VisualFlamenco [1], but we also introduce some new facets. We frst describe the facets currently integrated in Picadomo starting with the most important ones.

Dominant color – a simple visual facet: The dominant color in a region is a simple visual facet, which is easy to derive and easy to use even on a mobile device. In order

1. *Picado* is a guitar playing technique used within *Flamenco* and gives a clue of the origin of our prototype. *Picadomo* refects that the tool administers Pictures without further ado on mobile devices

2. http://www.spbsoftwarehouse.com/products/imageer/?en (last visited on: January 14th, 2009)

3. http://iwindowsmobile.com/mobile-photo-viewer.html (last visited on: January 14th, 2009)

4. http://developer.android.com/ (last visited on: March 11th, 2009)

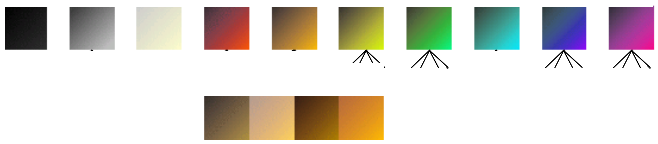


Figure 3. Color feature hierarchy used in our prototype [5].

to extract this facet, the most frequent color is extracted for the whole image and each image region (center, upper left quadrant, upper right quadrant, lower left quadrant, lower right quadrant). For this purpose we obtain 36-dimensional color histograms based on the HSV color space [5] and pick the bin that contains the highest number of pixels. Each color bin is a leaf in the two-level dominant color hierarchy (see Fig. 3). The user can specify the most frequent color of an image region from seven coarse colors or a gray level on the frst level of the hierarchy and then refne his choice further on the second level if this seems necessary.

One important point for usability is the visualization of the color facet. Each dominant color facet corresponds to an image region but also to a region in color space. While it is straightforward to visualize the region, we found that we gain in usability if we visualize the color region by showing a color gradient in the facet visualization as depicted in Fig. 1 on the left and in the middle.

Tags – associated keywords: Whenever an image is presented to the user he can use keywords for tagging an image to fnd it later searching for these keywords. The tags are stored in a separate database table, as they are an exception from our facet data structure. Any image can have as many tags as the user associates to it, whereas for the color an image matches exactly one facet for each region.

GPS data: If the EXIF data contains the latitude and longitude of the place where a photo was taken, we store the coordinates in our database and show the photos on an optional displayable map.

A short description of the other facets integrated in Picadomo:

EXIF facets: The following facets are all extracted from EXIF data.

Date: The date the photo was taken. The granularity of the facet is chosen to be by year. Further extensions in hierarchy are possible (year, month, week and so on).

Time of day: The time the photo was taken. Here we differentiate only between day and night, a more fne grained differentiation would be possible, but might be fussy.

Season: The season (spring, summer, autumn, winter) naturally depends on the date the photo was taken.

Weekend or weekday: As it would make sense to be able to look for photos taken on weekends only, we added this facet.

Full image visual facets: The following visual facet was extracted from the full image.

Coherent color: Coherent colors [6] of irregular image regions[5]. We classify coherent regions of a color into three classes: big, medium or small.

Full image and image region visual facets: The following visual facets were extracted from the full image, the center and the four quadrants of the image.

Dominant color: The dominant color as described above.

Contrast: This facet enables the user to choose between images with many short-range changes of brightness and images with rather soft changes of brightness.

3.2. Feature Tree

With VisualFlamenco, the predecessor of Picadomo, we used the Flamenco[6] system as backend. For our Android based Picadomo prototype, the Flamenco system was no longer applicable, so we designed a backend of our own. One part is a database that stores all associated facets for each image. Another part is our dynamically created feature tree depending on the feature extractors that come in operation. A feature extractor accepts the input image, extracts its feature(s) and fnds the corresponding facet.

In the following part the structure of the feature tree is depicted showing some possible values:

```
(01) -(root node)
(02)  --dominant color      color extractor
(04)   |-full image
(05)   | --gray level
(..)   |  ---...
(14)   | --red
(15)   |  ---dark pastel red
(16)   |  ---bright pastel red
(17)   |  ---dark saturated red
(18)   |  ---bright saturated red
(19)   | --orange
(..)   | --...
(50)   |-top left image
```

5. These irregular regions were identifed in the whole image, whereas within the rest of the paper we assume an image region to be predefned like e.g. top left or bottom right.

6. http://famenco.berk eley.edu/ (last visited on: January 14th, 2009)

```
(..)    |-...
(279) --contrast            contrast extractor
(281)   |-full image
(282)   | --low
(283)   | --medium
(284)   | --high
(286)   |-top left image
(...) --...                 other extractors
```

The color extractor for instance fnds the dominant color for the full image and the regions. Assuming that an image is mostly (dark saturated) red, the color extractor will return facet ID 17 as dominant color for the full image. Every extractor is also responsible for creating its part of the feature tree of an image. We only administer a list of extractors we want to use and the tree will be created from the extractors in the list beginning with the root node. So, we can easily add, remove or change an extractor. The GUI relies on the feature tree as well and is generated in a dynamic fashion. Every node corresponds to one button in the GUI. As shown in Fig. 2 we have a button for the full image (facet ID 04) and for all regions (e.g. facet ID 50 for the top left quarter of the image) in our navigation bar.

3.3. Control Concept

The control sequence of our image browser when searching for images with certain properties can be viewed as a navigation through the feature tree while selecting and deselecting single nodes or leafs whereby the corresponding facets and those in the subtrees will be activated or not. Fig. 4(a) shows the initial navigation bar when starting Picadomo. Every click on a button directs deeper in the feature tree, the dominant color node is assigned to the frst button and leads to a regional choice, where the user can select the desired image region (see Fig. 4(b)).

After having chosen a region, the user can specify the desired shade of color for the corresponding region of the image. He can either select the main shade of any color or refne this shade further with one of the four sub color buttons (see Fig. 4(c) with shades of blue as example) according to the color hierarchy in Fig. 3. The numbers in this view show the size of the result set, if the user would activate this facet.

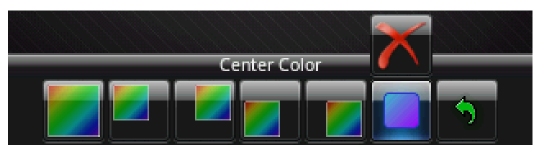
Fig. 4(c) depicts the buttons for the fner color shades overlaying the image grid. This approach is used for all leaf nodes from our feature tree. All non-leaf nodes correspond to a button within the navigation bar, whereas all leaf nodes will be painted over the image grid, in order to save screen size. In our opinion Picadomo improves the clarity and usability in comparison to VisualFlamenco, because the selection of the facets can be done without scrolling. However, it takes the user some more "touches" to achieve the same results as in VisualFlamenco, but it is still an

(a) The initial navigation bar of our prototype, every button matches a child of the root node from our feature tree.

(b) The navigation bar appearing when the user has clicked the very left button (*Color Regions*) in Fig. 4(a).

(c) View after pressing the *Center Color* button (second from right in Fig. 4(b)) and clicking once on the button for a shade of blue.

(d) When having chosen a color for a region the region button will change to the selected color to indicate an active facet.



(e) If the user clicks an already active button again, he can delete his earlier choice with a click on the red X.

Figure 4. The control sequence of our navigation.

improvement, if you consider the limited screen size and the ease of use of performing a "touch" on a touch screen.

The next step is shown in Fig. 4(d) where the prior chosen region button changed its color to the opted color. Here, the center of the image should be in some shade of light blue. If the user is not satisfed with the result, he can delete and change the choice of an active facet by clicking again on the corresponding button, deleting his earlier choice and choosing another value (see Fig. 4(e)). The same scheme is also used for the remaining facets. As can be seen in Fig. 4(a), the dominant color uses only the left outer most button. The next button adjusts the contrast for a given region in three different levels, the third button lets the user choose coherent color regions for each shade of color, and button four is for date and time facets. The ffth button allows for searching the database via tags and keywords, the sixth button opens a map view, on which all the images will be displayed at their GPS location. The right outermost button deselects all active facets.

4. User Experiments

We did two user studies – the frst with 5, the second with 15 people – with a total of 12 feasible outcomes. Except for one participant, no-one had any experience in content-based image retrieval. The knowledge of digital image processing ranges from *very well* to *none* and all users said that they had some more or less fxed scheme for organizing their private image collections.

In the beginning the idea of faceted search was described and demonstrated. Then, the features used for Picadomo were explained and all participants agreed in having understood faceted search. In all cases the introduction took less than 10 minutes and afterwards the users were allowed to explore the system and its control concept at will, which took from one to 16 minutes depending on the individual user. All experiments were conducted on the Android emulator[7], since real devices were not yet available.

4.1. Category Search

Within all assignments we used a collection of more than 600 images from Flickr (http://www.fickr.com) and private sources. We confronted the users with the task of fnding some images out of the following categories: *sunset/sunrise, cityscape, outdoor sports, winter landscape, Egypt, forests/trees*. Each user had to fnd at least one image from every category with less than 30 images in the total result set. A search via tags was not possible due to our untagged image database.

Every participant was able to solve the category search within a reasonable time for all six categories. The median lies at exactly one minute for each category search assignment, including reading and solving the particular task. The fastest participant solved all six searches within 3:40 minutes. The facets used to fnd the desired images signifcantly vary among the users, what supports individual user intuitions. As an example, a *cityscape* image could be found by choosing one of the following facet sets:

- Dominant Color: full image: *gray*
- Coherent Color: blue shade: *medium*
- Contrast: top-right image part: *medium*
- EXIF: time: *day*; Coherent Color: gray shade: *high*

In the current system setting with only 600 images, one or two well-chosen facets were often enough to reduce the result set to less than 30 images. But the visual faceted search also works with a collection of more than 11 000 images, as we showed in our prior work on VisualFlamenco. In our current work we use fewer images, because the emulator for the Android platform offers only limited performance in comparison to a real mobile device running the Android platform.

4.2. Target Search

The next block of exercises for the participants consisted of a target search for ten given images. Table 1 gives an example of four images, together with the number of clicks and the duration it took our participants to fnd it, which includes looking up the image on our questionnaire, refecting on suitable facets, choosing them, checking the result set for the desired image and possibly deselecting or changing facets. Here we count every single click the user made, so high numbers of clicks indicate a user searching for the right facets. Our experiments furthermore showed that after 9.5 clicks a facet was selected on average, whereby every single facet selection cannot be done with less than $3-5$ clicks. The median for the duration of a target search task lies at 1:03 minutes and the fastest participant found all ten images within 3:04 minutes, which is almost the same as for category search.

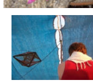
Target Image	Number of clicks			Duration	
	Min	Max	Median	Min	Median
	5	32	16.5	0:15	0:51
	6	38	14	0:23	0:49
	5	40	5	0:12	0:14
	9	49	20	0:32	0:58

Table 1. Some sample images used in our target search task with corresponding number of clicks and duration statistics.

The feature mostly used to fulfll a given task was Dominant Color with approximately 66%, where the color for a region was used twice as often as the color for the full image. Surprisingly, the Coherent Color facet came in second with 18%, the Contrast features were used for 13% of all queries and EXIF features could be neclected with only 3%. EXIF features like the date taken would be probably more useful for own image collections, where the user has more chronological knowledge about the images. The contrast could be helpful for professional users, as normal users are not profcient enough with image contrast and prefer to use the easy to understand color features, which

7. http://developer.android.com/guide/developing/tools/emulator.html (last visited on: March 10th, 2009)

also confrms our strategy to use only easy understandable features.

4.3. User Comments

All participants were asked to write down their personal pros and cons on a questionnaire. On the pro side they certifed our prototype a very easy and fast to learn user interface, in spite of the numerous amount of facets a good overview is given at any time. Our users liked the idea of faceted search on a mobile device and the possibility to fnd a given image very fast (*target search*) or the simple way to delete unwanted facets. All test persons mentioned as well that the available facets can be easily used to fnd the desired images very quickly.

On the other hand, some users would like to have a faster reacting frontend or the shades of color improved, more textual tooltips and labels for the facets, an even fner choice for the image regions or some kind of undo history for all chosen facets. Furthermore, one participant mentioned the lack of an overview of all currently active facets and another one would have preferred some kind of direct date input form in opposition to the predetermined facets for date and time.

4.4. Evaluation

Our user studies showed properly that even people without image retrieval knowledge can easily use our prototype for image browsing and searching. The overall control concept seems very promising, although there is still place for improvement, as the users wish to have undo functionality or an overview of all active facets, for example. Also, the utilized facets should be revised. Instead of rarely used facets we can introduce other available facets, such as coarseness or directionality, although we omitted them for the current prototype. The breakdown of the colors could be changed to more main colors and less fne colors, to better ft the human color perception. Since our user experiments were done with frst-time users only, we will address more experienced users as well. This second type of users could need some real expert features for even better and more precise image retrieval and an adaptation of the interface to the users seems desirable in future work.

5. Conclusion and Future Work

In this paper we described our faceted image browser Picadomo for mobile devices. The accomplished user experiments, as well as our own experience, show that Picadomo is a useful application for searching image collections. We believe that Picadomo is a useful approach for fnding images, combining search based on visual properties, tagging, EXIF and GPS data. We are planning to combine the local image search with peer-to-peer (P2P) and web search, so people can share their local image collections with friends and search within their images or use a global web-based image database like Flickr as image resource. Of course we will address user suggestions, such as an overview window for all currently active facets or a listing of all facets a certain image is assigned to. Last but not least, we will try to improve the response times with respect to our Android-based prototype.

References

[1] W. Müller, M. Zech, A. Henrich, and D. Blank, “VisualFlamenco: Dependable, Interactive Image Browsing Based on Visual Properties,” *CBMI International Workshop on Content-Based Multimedia Indexing, IEEE, London*, pp. 568–575, 2008.

[2] M. Naaman, R. Nair, and V. Kaplun, “Photos on the go: a mobile application case study,” in *CHI ’08: Proceeding of the twenty-sixth annual SIGCHI conference on Human factors in computing systems*. New York, NY, USA: ACM, 2008, pp. 1739–1748.

[3] S. Harada, M. Naaman, Y. J. Song, Q. Wang, and A. Paepcke, “Lost in memories: interacting with photo collections on pdas,” in *JCDL ’04: Proceedings of the 4th ACM/IEEE-CS joint conference on Digital libraries*. New York, NY, USA: ACM, 2004, pp. 325–333.

[4] H. Liu, X. Xie, W.-Y. Ma, and H.-J. Zhang, “Automatic browsing of large pictures on mobile devices,” in *MULTIMEDIA ’03: Proceedings of the eleventh ACM international conference on Multimedia*. New York, NY, USA: ACM, 2003, pp. 148–155.

[5] L. Zhang, F. Lin, and B. Zhang, “A CBIR Method Based on Color-Spatial Feature,” *IEEE Region 10 Annual International Conference*, pp. 166–169, 1999.

[6] G. Pass, R. Zabih, and J. Miller, “Comparing Images Using Color Coherence Vectors,” in *ACM Multimedia*, 1996, pp. 65–73. [Online]. Available: citeseer.ist.psu.edu/pass96comparing.html