Fine-grained Visual Faceted Search

Julien Champ, Alexis Joly, Pierre Bonnet

To cite this version:

Julien Champ, Alexis Joly, Pierre Bonnet. Fine-grained Visual Faceted Search. MM: Conference on Multimedia, Nov 2014, Orlando, FL, United States. pp.721-722, 10.1145/2647868.2654875 . hal-01088822

HAL Id: hal-01088822
https://inria.hal.science/hal-01088822

Submitted on 23 Jul 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Fine-grained Visual Faceted Search

Julien Champ
INRA, LIRMM
ARCAD FEDER project
Montpellier, France

Alexis Joly
INRIA, LIRMM
Montpellier, France

Pierre Bonnet
CIRAD, AMAP
ARCAD FEDER project
Montpellier, France

ABSTRACT

This paper presents an interactive information retrieval scheme extending classical faceted search with fine-grained visual facets. Consistent visual words are extracted offline by constructing a matching graph at the region image level and splitting the graph in relevant clusters. Visual and textual facets are then indexed in the same faceted search engine, all of them being dynamically reranked according to their filtering efficiency before being presented to the end-user. To enhance their interpretability, each facet value is illustrated by a representative picture automatically selected from the matching graph in an adaptive way.

Categories and Subject Descriptors: H.3.3 [Information Systems]: Information Search and Retrieval

Keywords: search;visual;facet;image;browsing; fine-grained; plant; species; identification.

1. INTRODUCTION

Faceted search has become the de facto search methodology for content rich web applications, in particular online stores and digital libraries. A faceted search system allows each item in the database to be assigned multiple classifications, enabling the classifications to be ordered in multiple ways, rather than in a single, pre-determined, taxonomic order. Facets are typically derived from pre-existing fields in the item's metadata or manually built from the analysis of the items. Using visual facets complementary to textual ones has already been addressed in a few previous works [2, 1, 9]. The principle is to automatically compute the facets by image clustering techniques and to display image links rather than textual links in the graphical user interface. As discussed in [2], visual faceted search might overcome several limitations of text oriented facets including incompleteness and inaccuracy of the meta-data.

In this paper, we stress another interest of visual faceting in the case of expert data to be searched by any user. We focus in particular on the problem of identifying the species of an observed living organism without any expertise in botany or zoology. This challenge, sometimes known as the *taxonomic gap*, is recognized as a major issue towards setting up effective and sustainable ecological surveillance tools. Determining the scientific name of a plant or an animal is actually still the primary key to index or access related data. But unfortunately, the usage of expert identification tools such as dichotomous keys and online floras requires a deep knowledge of the morphological description language. Even high-level categories such as the main types of fruit (achene, drupe, capsule, nut, etc.), leaves (compound ,bipenates, etc.) are not understood by most users.

In this paper, we introduce two complementary visual faceted search mechanisms in order to widen access to such fine-grained attributes. The first one attempts to visually illustrate pre-existing expert facets in the data whereas the other one attempts to discover new facets from the visual contents. Both rely on the offline construction and clustering of a dense visual matching graph computed at the region level (contrary to previous visual faceted search proposals [2, 1, 9] that relied on image-centric clustering and global visual features).

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage, and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s). Copyright is held by the author/owner(s).

MM'14, November 3–7, 2014, Orlando, Florida, USA.
ACM 978-1-4503-3063-3/14/11.
http://dx.doi.org/10.1145/2647868.2654875.

2. AUTOMATIC EXTRACTION OF FINE-GRAINED VISUAL FACETS

Visual facets are automatically extracted from each of the sub-collection of pictures used in our application: leaves, flowers, fruits and trunks (containing themselves categories that might be visually and semantically very similar). This is done through the following steps:

Hash-based K-NN search. The approximate K-Nearest Neighbors (K-NN) of each local feature $\mathbf{x}_j$ belonging to an image I of the dataset are computed efficiently thanks to the hash-based multi-probe search method introduced in [5]. Its principle is to train an adaptive search model at indexing time through kernel density estimates computed on exact K-NN samples. This model is used at search time to select the most probable buckets to be visited so as to retrieve on average a fraction α of the real K-NN's (α was set to 0.80 in all our experiments). The advantage of this method compared to other state-of-the-art methods (such as PQ-code [4]) is that it allows controlling accurately the quality of the retrieved K-NN while being applicable to any quantization function and metrics. In our case, the original scheme is transposed to the use of a more effective hash function (RMMH [6]) and to the application of the Hamming Embedding principle [3].

Matching graphs construction. A consistent matching graph is computed efficiently on top of the raw visual NN's using the RANdom Sampling and Search algorithm introduced in [8]. It allows matching efficiently feature sets with local geometric constraints by iteratively and adaptively selecting features and their neighborhood in the whole collection. As plants are not rigid objects, we however relaxed the affine geometric constraints used in the original method and used instead neighborhood constraints with an adaptive radius computed as the distance of the m-the spatial neighbor of the primary searched local feature. A matching graph is computed in this way for each of the used sub-collection of pictures (i.e. leaves, flowers, fruits and trunks).

Matching graphs clustering The matching graphs produced by the previous step are then processed by the MCL algorithm [10] in order to group image regions in relevant clusters. To derive facets potentially useful at different granularities, MCL is run several times on each matching graph but with different inflation values (typically at 3 granularities).

3. FACETED SEARCH ENGINE AND GUI

Automatically extracted visual facets are indexed through Apache Solr faceting module as any structured field (the cluster identifier being used as field value). Plant species are considered as the main index entries (i.e. as *documents* using Solr's terminology). Each species is then associated with a number of multi-valued visual facets, corresponding to the cluster id's of the images of that species in each of the MCL partition (and each organ). Complementary, available text-based facets are also integrated in the faceted search module (in our case <Inflorescence type>, <Infruitescence type>, <Flower colour>, <Leaf type>). Both the automatically extracted visual facets and the text-based facets are displayed to the user using representative images for each facet value (rather than texts like in classical applications). The representative image of each facet value is dynamically computed from the fine-grained matching graph so as to adapt it to the previous user search actions. This is simply done by selecting the most connected node (i.e. image region) among all the nodes belonging to the species that were not filtered before. The images containing the selected nodes are roughly cropped before being displayed on the GUI according to the position and size of the selected image region.

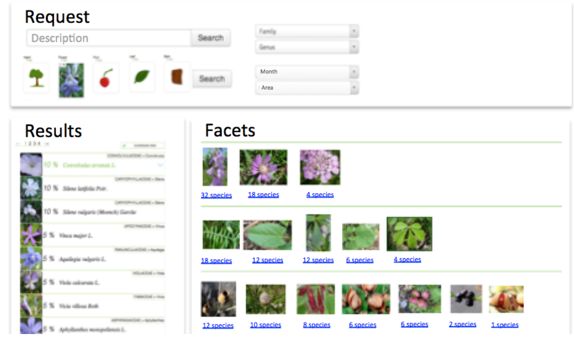
Before being presented to the user, all facets (textual and visual) are dynamically reranked according to their filtering efficiency on the remaining species. This is done by computing the following normalized entropy as relevance score of each facet:

$$S(X) = -\sum_{i=1}^{n} \frac{p(x_i).log_2(p(x_i))}{n}$$

where n is the number of values observed in the search results for facet X and x_i is the i-th observed value (e.g. value *red* for facet *flower color*). Ranking all facets according to that score allows to highlight the most informative while penalizing the ones with too much possible values. The number n of observed values is displayed to the user on the GUI as well as the number of species matching each facet value.

4. DATA AND DEMO

Our fine-grained visual faceting scheme is demonstrated on a set of 500 plant species living in France area correspond-



ing to the species included in CLEF plant identification task 2014 (within LifeCLEF *lab* [7]). Textual facets were extracted from TelaBotanica[1] eFlore sheets (leaf, flower and fruit type, flower color, vernacular name). Full text search is performed on the 500 wikipedia articles of the selected species. The demonstration of the system presented at ACM MM 2014 conference will be organized as a user trial, involving a volunteer tester among the audience for each demo session. The volunteer tester will be provided a picture of a plant or a plant organ and will be in charge of determining what it is. The amount of time to succeed and the logs will be analyzed afterwards and contribute to the evaluation of the system.

5. REFERENCES

[1] I. Bartolini and P. Ciaccia. Integrating semantic and visual facets for browsing digital photo collections. In *SEBD*, pages 65–72, 2009.

[2] M. J. Henry, S. Hampton, A. Endert, I. Roberts, and D. Payne. Multifacet: A faceted interface for browsing large multimedia collections. In *Multimedia (ISM), 2013 IEEE International Symposium on*, pages 347–350. IEEE, 2013.

[3] H. Jegou, M. Douze, and C. Schmid. Hamming embedding and weak geometric consistency for large scale image search. In *Computer Vision–ECCV 2008*, pages 304–317. Springer, 2008.

[4] H. Jegou, M. Douze, and C. Schmid. Product quantization for nearest neighbor search. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 33(1):117–128, 2011.

[5] A. Joly and O. Buisson. A posteriori multi-probe locality sensitive hashing. In *Proceedings of the 16th ACM international conference on Multimedia*, pages 209–218. ACM, 2008.

[6] A. Joly and O. Buisson. Random maximum margin hashing. In *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on*, pages 873–880. IEEE, 2011.

[7] A. Joly, H. Müller, H. Goëau, H. Glotin, C. Spampinato, A. Rauber, P. Bonnet, W.-P. Vellinga, and B. Fisher. Lifeclef: Multimedia life species identification. In *ACM Workshop on Environmental Multimedia Retrieval*, 2014.

[8] P. Letessier, O. Buisson, and A. Joly. Consistent visual words mining with adaptive sampling. In *Proceedings of the 1st ACM International Conference on Multimedia Retrieval*, page 49. ACM, 2011.

[9] W. Muller, M. Zech, A. Henrich, and D. Blank. Visualflamenco: Dependable, interactive image browsing based on visual properties. In *Content-Based Multimedia Indexing, 2008. CBMI 2008. International Workshop on*, pages 568–575. IEEE, 2008.

[10] S. Van Dongen. A cluster algorithm for graphs. *Report-Information systems*, (10):1–40, 2000.

[1] http://www.tela-botanica.org/

View publication stats