# Small Objects Query Suggestion in a Large Web-Image Collection

Pierre Letessier
INRIA ZENITH
34095 Montpellier, France
pierre.letessier@inria.fr

Nicolas Hervé
INA
94360 Bry-sur-Marne, France
nherve@ina.fr

Julien Champ
INRIA ZENITH
34095 Montpellier, France
julien.champ@inria.fr

Alexis Joly
INRIA ZENITH
34095 Montpellier, France
alexis.joly@inria.fr

Olivier Buisson
INA
94360 Bry-sur-Marne, France
obuisson@ina.fr

Amel Hamzaoui
INRIA Rocquencourt
78153 Le Chesnay, France
amel.hamzaoui@inria.fr

## ABSTRACT

State-of-the-art visual search methods allow retrieving efficiently small rigid objects in very large image datasets (e.g. logos, paintings, etc.). User's perception of the classical query-by-window paradigm is however affected by the fact that many submitted queries actually return nothing or only junk results. We demonstrate in this demo that the perception can be radically different if the objects of interest are rather suggested to the user by pre-computing relevant clusters of instances. Impressive results involving very small objects discovered in a web collection of 110K images are demonstrated through a simple interactive GUI.

**Categories and Subject Descriptors:** H.3 [Information Storage And Retrieval]: Miscellaneous

**Keywords:** Mining, Scalable, Small Objects, Hashing

## 1. USER'S PERCEPTION ISSUES

Large-scale object retrieval systems have demonstrated impressive performance in the last few years. The underlying methods, based on local visual features and efficient indexing models, can retrieve accurately small rigid objects such as logos, buildings or manufactured objects, under varying points of view and illumination conditions [2, 4, 5, 6, 9]. Therefore online object retrieval is now achievable up to 1M images with a state-of-the-art computer [4]. From the usage point of view, these methods are often combined with a query-by-window search paradigm. The user can freely select a region of interest in any image, and the system returns a ranked list of images that are the most likely to contain an instance of the targeted object of interest [10]. This paradigm has however several limitations related to user's perception: (i) When no (or very few) other instances of the query object exist in the dataset, the system mostly returns false positives making the user uncomfortable with the results. Indeed, he does not know if there are actually no other instances of the query object or if the system did not work correctly. (ii) When the user selects a deformable or complex object that the system is actually not able to retrieve, the system mostly returns false positives as well. As the user can freely select any object, this appears very frequently leaving the user with a bad impression of the effectiveness of the tool. The second remark is even more critical if the user believes that the system can retrieve any semantically similar objects (e.g. object categories or visual concepts such as cats or cars). We do not argue here that such queries will never be solved effectively in the future. We just emphasize that bridging the gap between a user's understanding of the system and the actual capabilities of the underlying tools is essential to make it successful in a real world search engine. A first possible solution to address these limitations would be to use some adaptive thresholding method, allowing only relevant results to be filtered, and possibly returning no results if none are found. The a contrario method of [5], for instance, allows the actual false alarm rate of rigid object instances retrieval to be controlled very accurately. But still, as the user can select any region of interest, the system might return no results in many cases and leave the user disappointed.


*MM'13,* October 21–25, 2013, Barcelona, Spain.
ACM 978-1-4503-2404-5/13/10
http://dx.doi.org/10.1145/2502081.2502248 ...$15.00.

## 2. PROPOSED APPROACH

We proposed in [3] to solve these user perception issues by a new visual query suggestion paradigm. Rather than letting the user select any region of interest, the system will suggest only visual query regions that actually contain relevant matches in the dataset. By mining off-line object instances in the dataset, it is indeed possible to suggest to the user only query objects having at least a predetermined number of instances in the collection. When a user clicks on a highlighted region, the system returns only the images containing other object instances of the same discovered cluster. From a user perception point of view, the proposed paradigm is very different from the window query paradigm. Indeed, since all suggested objects mostly return correct results, the user might rather perceive them as visual links (or hyper-visual links by analogy to hypertext links). It is important to notice, that unlike existing approaches, the links produced by our method are not links between images, but links between automatically localized image regions containing instances of the same rigid object. An image can thus contain several suggestions belonging to different object's clusters and the user can navigate in the collection by moving from an object to another, step-by-step.

# 3. OBJECTS MINING ALGORITHM

The batch process allowing the efficient discovery of the suggested objects is based on our previous work described in [8]. It introduces a new scalable method for automatically discovering frequent visual objects in large multimedia collections even if their size is very small. It first revisits formally the problems of mining and discovering such objects, by introducing the concept of *(c,f)-frequent objects* and generalizes two kinds of existing methods for probing candidate object seeds: weighted adaptive sampling and hashing-based methods. It then introduced a new hashing strategy, working first at the visual level, and then at the geometric level. It allows integrating weak geometric constraints into the hashing phase itself and not only neighborhood constraints as in previous works. Experiments did show that using this new hashing-based prior allows a drastic reduction of the number of tentative probes required to discover small objects instantiated several times in a large dataset.
Figure 1 reproduces the result of a comparative study (conducted in [7]) between our method and the popular Geometric min-Hashing approach (GmH, [1]). This experiment is based exactly on the same evaluation protocol and data as the one introduced in [8] (the comparison to GmH was not ready at the time of submitting that paper). The aim is to measure the ability of each method to produce relevant object's seeds in terms of recall/precision on FlickrBelgaLogo dataset[1] (to our knowledge, no other dataset allows such an evaluation). To achieve the best possible run with GmH, its main parameter has been boosted to much more higher values than the one recommended in the paper: 100K seeds per image instead of 60. As in GmH, we used a large vocabulary of 1M visual words. To convert the collisions produced by GmH into a probability mass function comparable with our own hashing scheme, we count, for each descriptor, the number of sketches colliding with the sketches including the considered descriptor.

# 4. DEMONSTRATOR

We built a demo of the proposed scheme on a web corpus of 110K images constructed so as to contain many potentially interesting instances of small objects such as sports or international organizations logos, famous buildings and places, etc. For that purpose, we queried a popular web image search engine with a list of 170 ad hoc keywords. Around 600 millions SIFT features were extracted from this corpus and the mining with our algorithm was completed within 48 hours on a single computer (two hexa-cores CPU Intel X5660). Instance clusters were extracted from the produced matching graph thanks to MCL[11] clustering algorithm (nodes representing the discovered frequent objects instances and edges weights the number of shared matched images). Then the interactive GUI allows four main actions: (i) visualize all resulting clusters sorted by frequency of the discovered object and represented by a representative thumbnail. (ii) visualize all instances of a discovered object by clicking on the representative thumbnail of its cluster. (iii) visualize all suggested objects in a given image by clicking on its thumbnail. Note that since an image can contain several object's suggestions, the instances belonging to the current cluster are displayed with a green bounding box whereas the instances belonging to other clusters are displayed with red bounding boxes. (iv) iteratively move from an object to another one by simply clicking on any red bounding box and visualize the images of the new cluster in a pop-up window. Due to the very small size of the discovered objects, their multiplicity and the precision of the clusters, this allows a very nice and unusual user experience in exploring an image collection. The user really has the impression to follow hyper-visual links and to the best of our knowledge such a navigation tool was never presented elsewhere.

[1] http://www-sop.inria.fr/members/Alexis.Joly/BelgaLogos/FlickrBelgaLogos.html

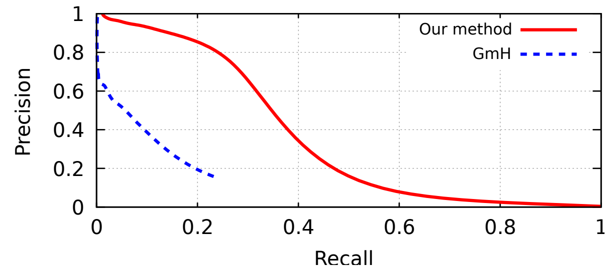


**Figure 1: Comparative study with Geometric min-Hashing**

# 5. REFERENCES


[1] O. Chum, M. Perdoch, and J. Matas. Geometric min-hashing: Finding a (thick) needle in a haystack. In *Proceedings of CVPR*, pages 17–24, 6 2009.

[2] O. Chum, J. Philbin, J. Sivic, M. Isard, and A. Zisserman. Total recall: Automatic query expansion with a generative feature model for object retrieval. In *Proceedings of ICCV*, 2007.

[3] A. Hamzaoui, P. Letessier, A. Joly, O. Buisson, and N.Boujemaa. Object-based visual query suggestion. *Multimedia Tools and Applications*, (11042), 2013.

[4] H. Jégou, M. Douze, and C. Schmid. Improving bag-of-features for large scale image search. *IJCV*, 87:316–336, 2010.

[5] A. Joly and O. Buisson. Logo retrieval with a contrario visual query expansion. In *Proceedings of ACM Multimedia*, pages 581–584, 2009.

[6] Y. Kuo, K. Chen, C. Chiang, and W. Hsu. Query expansion for hash-based image object retrieval. In *Proceedings of ACM Multimedia*, pages 65–74, 2009.

[7] P. Letessier. *Découverte et exploitation d'objets fréquents dans des collections multimédias*. PhD thesis, Telecom ParisTech, 2013.

[8] P. Letessier, O. Buisson, and A. Joly. Scalable mining of small visual objects. In *Proceedings of ACM Multimedia*, pages 599–608, 2012.

[9] J. Philbin. *Scalable Object Retrieval in Very Large Image Collections*. PhD thesis, Uni. of Oxford, 2010.

[10] J. Philbin, O. Chum, M. Isard, J. Sivic, and A. Zisserman. Object retrieval with large vocabularies and fast spatial matching. In *Proceedings of CVPR*, 2007.

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