HAL open science

Graphic-Based Concept Retrieval

Massimo Ferri

To cite this version:

Massimo Ferri. Graphic-Based Concept Retrieval. 1st Cross-Domain Conference and Workshop on Availability, Reliability, and Security in Information Systems (CD-ARES), Sep 2013, Regensburg, Germany. pp.460-468. hal-01506774

HAL Id: hal-01506774

https://inria.hal.science/hal-01506774

Submitted on 12 Apr 2017

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.

Distributed under a Creative Commons Attribution 4.0 International License

Graphic-based concept retrieval

Massimo Ferri

Dip. di Matematica and ARCES, Univ. di Bologna, Italy
massimo.ferri@unibo.it,

Abstract. Two ways of expressing concepts in the context of image retrieval are presented. One, Keypics, is on the side of an image owner, who wants the image itself to be found on the Web; the second, Trittico, is on the side of the image searcher. Both are based on the paradigm of human intermediation for overcoming the semantic gap. Both require tools capable of qualitative analysis, and have been experimented by using persistent homology.

Keywords: concept retrieval; image retrieval; persistent homology; size functions; keypics

1 Introduction

A problem, which is obviously central in information retrieval, is that making a machine understand the content of an image (or any other document, for that matter) is a hard task; making it understand a concept is still tougher. There is ongoing research on extracting concepts from texts, based on occurrences (Bag-Of-Words) and on various, ingenuous stucture analysis (ontologies), and on statistics on very large data sets [6, 11, 17]. Similar ideas have been exported and adapted to concept extraction from images (e.g. within ImageCLEF [15] and in general in the Content-Based Image Retrieval community).

The meaning of words drifts in time. The semantic content of images undergoes even more rapid and drastic changes. The link between *signifier* and *signified* depends very much on the cultural, spatio-temporal environment, on the specific tasks of the user, even on age and gender [19]. Therefore we believe that conceptual annotation will be possible — at least for some more years — only with a human intermediation bridging the "semantic gap" [18]. This is presently performed in various forms, but mostly by textual or numerical indexing [16].

We present here two research lines developed by our Vision Mathematics group at the University of Bologna in the last few years; they are not new, but we think that they might be little known by and of some interest for the Human-Computer Interaction community; both find a place in Semiotic Engineering [14, 13, 12]. In both, the human intervention is essential and is expressed graphically. The first research line, "Keypics", consists of an annotation by hand-drawn sketches. It was suspended because its effectiveness depends on a social factor: for it to be successful, it ought to be adopted by a very large community. The second line, "Trittico", is based on the idea that a visual concept can be "spanned" by

different images, the concept itself being the *quid* in common. It was abandoned as a commercial project but is now being revived in the paradigm of relevance feedback [10].

2 Keypics

A first way of transferring (or condensing, or stressing) concepts contained in a document, in particular an image, is by indexing. However, indexing and retrieving by words suffers from several drawbacks: the language barrier, the existence of synonims and namesakes, and above all rigidity. In fact, concepts have fuzzy and moving boundaries, but words crystallize them into discrete objects among which movement occurs by leaps. On the contrary, drawings are more dynamic. The solution we propose is indexing by *Keypics* (as opposed to key*words*). In practice, we suggest that images on the Internet should be equipped with simplified sketches representing the essentials of the images themselves (see also [5, 4]). The sketches should be provided by the image owner or manager. This graphical indexing might be extended to whole Web pages.

This might be performed by use of simple drawing and processing tools, or by hand. The Keypic should represent what is felt as essential by the image owner. So it could be an outline of the relevant shapes in the image, or a symbol semantically referring to its content. E.g., the picture of a symphonic orchestra (or its home page) might be indexed by a note, meaning that the picture concerns music. Several images might be associated to the same Keypic, and more than one Keypic might be associated to the same image.

Fig. 1. A clip-art image of a toucan and its Keypic

The idea of using iconic or graphical metadata is surely not new. The most common example is perhaps that of road signs; although some text often ac-

companies them, road signs are generally conceived as neutral with respect to language. Their shape is not necessarily related in a semantic way to the message they carry: It is mostly conventional, although the choice of the shape may be dictated by psychological considerations. Another noticeable situation in which shapes substitute or at least accompany a textual indication is sports: as far as we know, the universally accepted signs for the different specialities, were designed for the 1964 Olympics in Tokyo for overcoming the obvious linguistic problems.

We propose that the icons for picture indexing should be simple, easy to draw, easy to process; they should either refer to the geometric aspects of the indexed pictures, or to their semantic contents, or both. They should preferably be expressed with a compact, standard code. They should be plastic, in the sense that they should not be limited to any pre-defined set. They should be, in terms of an image, as synthetic, meaningful and free as keywords are in general use. Actually, they would be superior to keywords, in that they would not suffer from the linguistic barrier, they would allow much more freedom of expression, they would be less severely affected by errors.

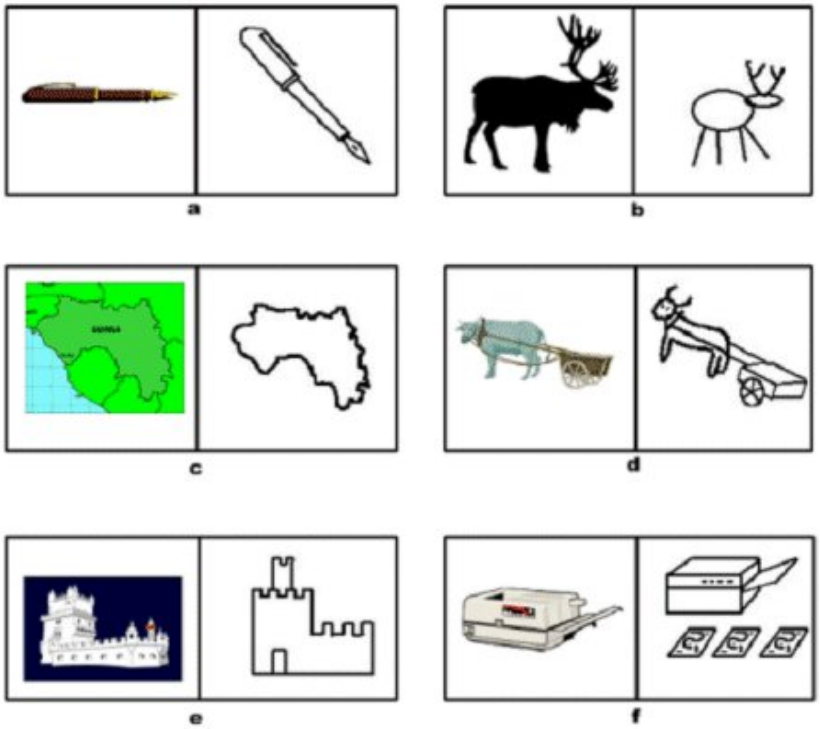

Fig. 2. Different strategies in drawing Keypics

We are aware that a likely and easy solution, which we consider deeply wrong, would be the creation of a fixed set of icons. This would imply that only a limited — even if wide — set of ideas might be conveyed. Moreover, users should depend on the choices of external authorities and maybe even on the claims of copyright owners. Updating would be necessary and frequent, with all problems related to version compatibility.

For these reasons, we stress the importance of leaving the highest freedom of expression to the image owner. This does not mean that stereotypes should be avoided; only that they should not be imposed. Actually, we believe that attractors will arise spontaneously by imitation. As naturally as new words are continually created and subjected to the natural selection of use, new Keypics would arise first in special circles, then possibly spread out to a wider community. They would be left free to appear, evolve (in a far smoother way than words) and eventually disappear.

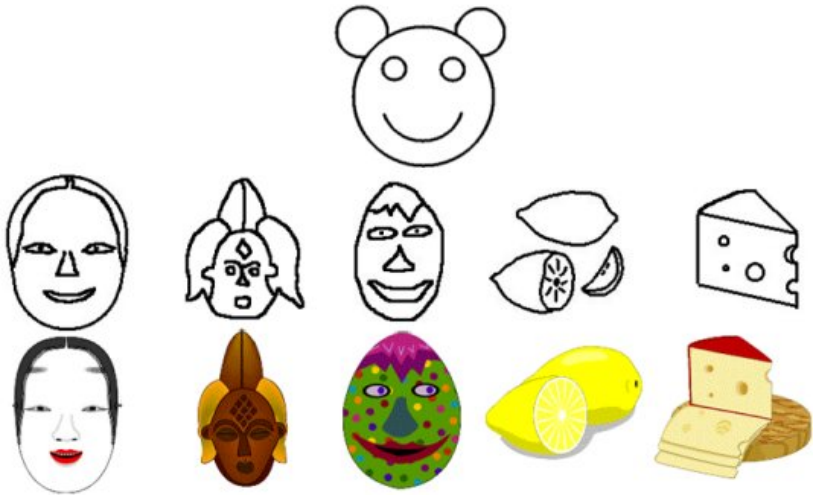

Fig. 3. A (partly) successful query

Another advantage of the plasticity we propose, lies in the rendering of morphological (and possibly semantic) nuances. As an example, the image owner who uploads a toucan image should be so provident as to detail the large beak as in Figure 1. Then, the image would be retrieved both by a user looking for birds, and (with greater priority) by one strictly interested in toucans.

How to process such drawings? The solution we adopted in an experiment is the geometrical-topological tool of Persistent Betti Numbers in degree zero

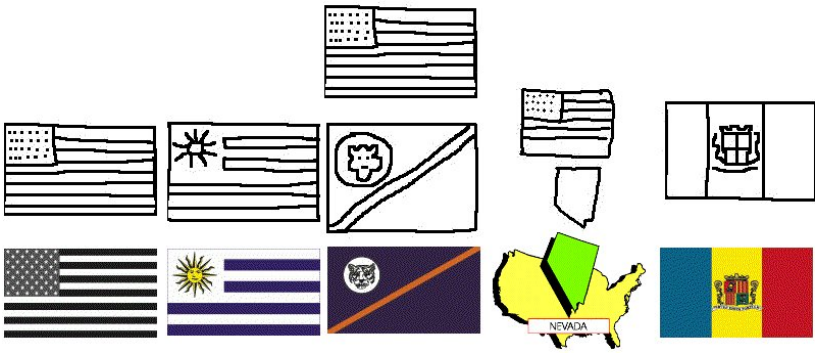

Fig. 4. An unexpected output

(also called Size Functions) [7, 8, 1, 3, 9]. They are modular shape descriptors particularly apt to capturing qualitative aspects of images.

Seven nonprofessional draftsmen were given templates chosen within very heterogeneous pictures of a commercially available clip–art collection; the stated aim was to depict the essentials of the given template, not to reproduce it accurately. A standard drawing program was used by all of them, endowed with standard tools as free–hand, straight–line or ellipse drawers, thresholding and edge detection. A set of 494 drawings resulted of it, all of a standard size, all black on white.

We were surprised by the very heteogeneous strategies adopted. Some drew a fairly accurate imitation as in Figure 2a. Sometimes the imitation was very rough (Figure 2b); in other cases (e.g. in Figure 2c) the use of an edge detector was evident. Some draftsmen thought it necessary to stress details (Figure 2d), or to ignore them (Figure 2e), but sometimes even to add nonexisting ones (Figure 2f).

We obtained rather satisfactory results in term of precision-recall curves in a set of retrieval experiments, but what we would like to stress here is that the semantic gap was actually filled effectively by the Keypic producers (see e.g. Figure 3. Sometimes this happened in unexpected ways. It was, e.g., the case of a query with the USA flag, where the map of Nevada popped up, because the operator had decided to add the Stars and Stripes — absent in the original image — in order to convey a meaning to the Keypic (Figure 4).

3 Spanning concepts

A completely different idea for implicit concept description assigns the responsability of filling the semantic gap to the querying person. This was done in the *Trittico* experiment [2] (and is presently evolving in the relevance feedback paradigm [10]). The query is expressed by means of three images that the user ei-

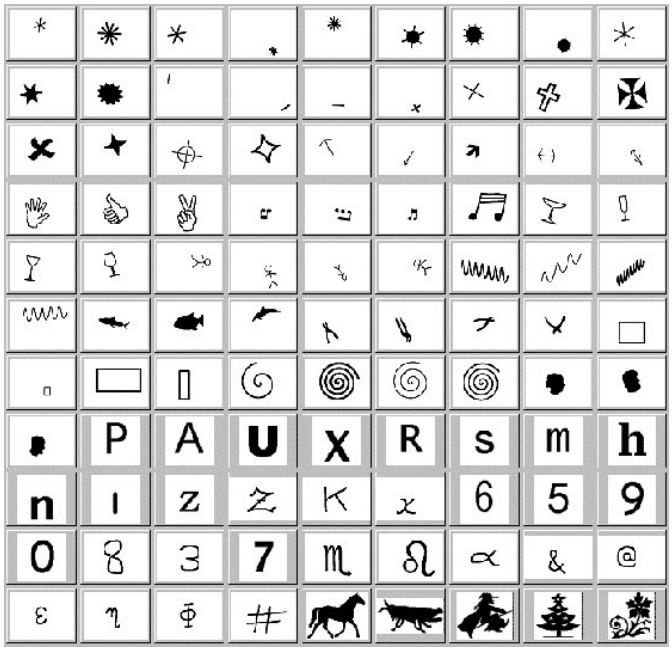

Fig. 5. A sample of the dataset

ther uploads or draws directly. The user is requested to propose three "extreme" instances of the shape he/she has in mind.

The mathematical core of the system is again a set of size functions (i.e. Persistent Betti Number Functions in degree zero). They depend on the choice of so called *filtering* (or *measuring*) functions (continuous real maps defined on the set). The definition of particular filtering functions allows to isolate those aspects of the image shape, which are of interest for the particular application goal.

Here is how the system should extrapolate the abstract shape concept from the three images. For each filtering function embedded in the program, the system computes the size functions of the three images, and evaluates their reciprocal distances. Then a comparison is done with the probability that these distances occur in a random triple. The lower the probability, the higher the weight that is given to the measuring function. The classifiers corresponding to the measuring functions, co-operate — with contributions depending on these weights — in the determination of a pertinence factor. On the base of the pertinence factor a set of images is extracted; finally, this set is sorted by a much finer comparison of the size functions.

In other words, the user puts a common feature into the three pictures, which for all the rest should be as different as possible. Then, those filtering functions

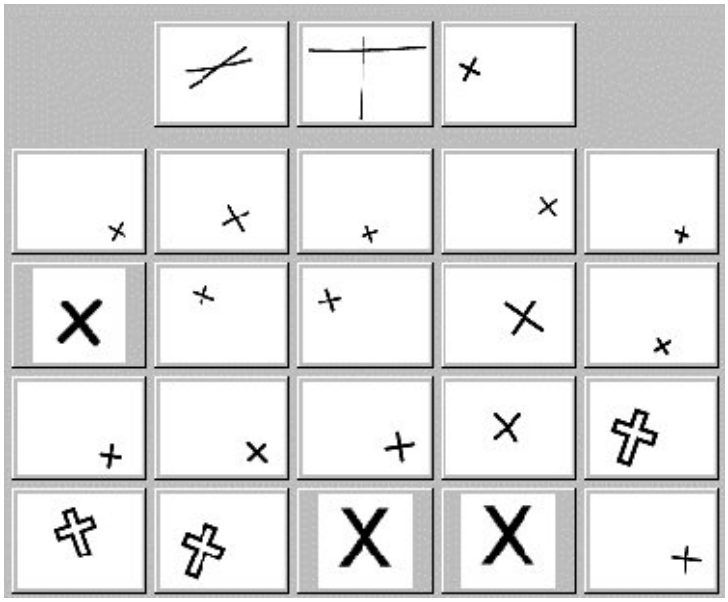

Fig. 6. A query and the first 20 output images

which are best fit to recognize that feature, are stressed in the comparison of the three input images (or, better said: of their size functions) with the database.

For checking this idea, we have chosen to build a database of simple black-on-white silhouettes. It consists of 2976 images belonging to 18 classes, plus 62 classes corresponding to alphabet characters, plus an extra "jamming" set of 197 unrelated images: see Figure 5 for a sample. Again, we are not interested here in reporting the quantitative assessment of the experiment, which you can find in [2]. What we want to stress, is that the system actually rather often output pertinent images which were very different from the query, "understanding" —so to say — the concept spanned by the three query images.

This is the case of Figure 6, for instance. The query consists of three heterogeneous crosses; this very independence has given the system the possibility to tune well the weights by which to stress some filtering functions against the other ones. So the output yields more crosses, some of which very different from the ones of the query.

A nice example is also given by Figure 7. In the first query, the three images are similar, but with three different orientations. The first ten output images have, coherently, no preferred orientation. In the second query, however, the three sketched little men are all upside down. The first ten output little men are also upside down. Without any need of directly imposing the system a restricted transformation group to be respected, the common attitude of the query images was sufficient to select the filtering functions which privileged this restriction.

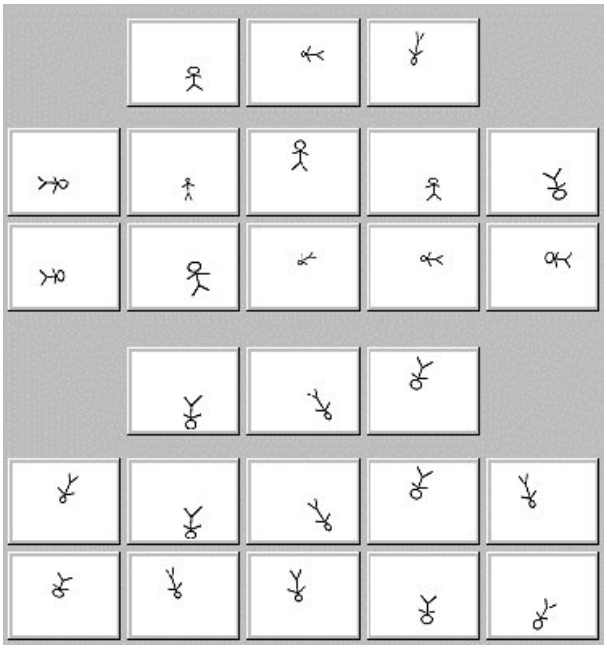

Fig. 7. Two queries and the first 10 output images

This idea is being used in a series of experiments on relevance feedback, with more sophisticated, multidimensional filtering functions and by tuning distances. A first attempt is reported in [10].

4 Conclusions

Two lines of research, in which human intervention bridges the semantic gap between image and concept, have been presented. In Keypics, image (and, more generally, document) annotation is by sketches rather than by words. Images are uploaded to the Web for a reason, and the image owner has all interest in focussing that reason, so he/she will extract from the images the part of content of his/her interest; this can be done by a simple, schematic picture drawn, copied or adapted by the image owner. In Trittico, a pictorial query is given by a triple of examples which span the searched for concept; the examples should be as far apart as possible, within the relevant class. Both lines have been developed by using persistent homology in degree zero (size functions), with successful initial experiments. The first awaits the sharing by a web users community for a true social experimentation. The second is evolving as a relevance feedback tool.

Acknowledgements

Work performed within the activity of INdAM-GNSAGA and of CIRAM and ARCES of the University of Bologna.

References

1. Biasotti, S., Cerri, A., Frosini, P., Giorgi, D., Landi, C.: Multidimensional size functions for shape comparison. Journal of Mathematical Imaging and Vision 32(2), 161–179 (2008)
2. Brucale, A., Cesari, F., d'Amico, M., Ferri, M., Frosini, P., Gualandri, L., Guerra, M., Lovato, A., Pace, I.: Image retrieval through abstract shape indication. In: Proceedings of the IAPR Workshop MVA 2000, Tokyo Nov. 28-30. pp. 367–370 (2000)
3. Cagliari, F., Di Fabio, B., Ferri, M.: One-dimensional reduction of multidimensional persistent homology. Proceedings of the American Mathematical Society 138(8), 3003–3017 (2010)
4. Cerri, A., Ferri, M., Frosini, P., Giorgi, D.: Keypics: free–hand drawn iconic keywords. International Journal of Shape Modeling 13(02), 125–137 (2007)
5. Cerri, A., Ferri, M., Giorgi, D.: A complete keypics experiment with size functions. In: Image and Video Retrieval, pp. 357–366. Springer (2005)
6. Egozi, O., Markovitch, S., Gabrilovich, E.: Concept-based information retrieval using explicit semantic analysis. ACM Trans. Inf. Syst. 29(2), 8:1–8:34 (Apr 2011), http://doi.acm.org/10.1145/1961209.1961211
7. Frosini, P.: Measuring shapes by size functions. In: Intelligent Robots and Computer Vision X: Algorithms and Techniques, Boston, November 01, 1991. pp. 122–133. International Society for Optics and Photonics (1992)
8. Frosini, P., Landi, C.: Size theory as a topological tool for computer vision. Pattern Recognition and Image Analysis 9(4), 596–603 (1999)
9. Frosini, P., Landi, C.: Persistent betti numbers for a noise tolerant shape-based approach to image retrieval. Pattern Recognition Letters pp. 863–872 (2012)
10. Giorgi, D., Frosini, P., Spagnuolo, M., Falcidieno, B.: 3D relevance feedback via multilevel relevance judgements. The Visual Computer 26(10), 1321–1338 (2010)
11. Haav, H.M., Lubi, T.L.: A survey of concept-based information retrieval tools on the web. In: Proceedings of the 5th East-European Conference ADBIS. vol. 2, pp. 29–41 (2001)
12. Holzinger, A.: On knowledge discovery and interactive intelligent visualization of biomedical data. In: Proceedings of the Int. Conf. on Data Technologies and Applications DATA 2012, Rome (Italy). pp. 5–16 (2012)
13. Nake, F., Grabowski, S.: Human–computer interaction viewed as pseudo-communication. Knowledge-Based Systems 14(8), 441–447 (2001)
14. Norman, D.A.: Turn signals are the facial expressions of automobiles. Basic Books (1992)
15. Nowak, S., Hanbury, A., Deselaers, T.: Object and concept recognition for image retrieval. In: ImageCLEF, pp. 199–219. Springer (2010)
16. Obeid, M., Jedynak, B., Daoudi, M.: Image indexing & retrieval using intermediate features. In: Proceedings of the ninth ACM international conference on Multimedia. pp. 531–533. ACM (2001)

17. Reiterer, E., Dreher, H., Gütl, C.: Automatic concept retrieval with Rubrico. In: Multikonferenz Wirtschaftsinformatik. pp. 3–14 (2010)
18. Smeulders, A.W.M., Worring, M., Santini, S., Gupta, A., Jain, R.: Content-based image retrieval at the end of the early years. Pattern Analysis and Machine Intelligence, IEEE Transactions on 22(12), 1349–1380 (2000)
19. Ziefle, M., Himmel, S., Holzinger, A.: How usage context shapes evaluation and adoption criteria in different technologies. In: AHFE 2012, Proceeding of Int. Conf. on Applied Human Factors and Ergonomics, San Francisco. pp. 2812–2821 (2012)