A proposal for image indexing: "keypics", plastic graphical metadata

Ferri M.[a] and Frosini P.[b]

Dept. of Mathematics and ARCES, Univ. Bologna,
Piazza Porta S. Donato 5 I-40126 BOLOGNA Italy

[a] ferri@dm.unibo.it
[b] frosini@dm.unibo.it

ABSTRACT

We propose a graphical indexing of images to be exposed on the Web. This should be accomplished by "keypics", i.e. auxiliary, simplified pictures referring to the geometrical and/or the semantic content of the indexed image.

Keypics should not be rigidly standardized; they should be left free to evolve, to express nuances and to stress details. A mathematical tool for dealing with such freedom already exists: Size Functions.

We support the idea of keypics with some experiments on a 498 images dataset.

Keywords: Content–based image retrieval, semantic gap, Size Functions

1. INTRODUCTION

This paper proposes a new technique for indexing a dataset of images. The main idea is to associate each image with a subjective visual description chosen by the dataset manager (DM). This kind of iconic indexing places itself at an intermediate level between semantic and geometrical descriptions. As a consequence, semantic information can be handled by using geometrical-topological tools.

We start from five assumptions. 1) Whoever puts images on the Internet (the DM), wants them to be retrieved by other users; 2) textual clues are incomplete and suffer from the linguistic barrier; 3) a general purpose segmentation system is beyond present technology; 4) the semantic content of an image is often wider than its geometrical content; 5) it is undesirable to confine shapes and concepts to a finite, fixed set.

Then we propose that the dataset manager equips each image with a simplified drawing, called "keypic" (as alternative to "keyword"). This might be performed by use of simple drawing and processing tools, or by hand, but preferably in SVG. The keypic should be representative of what is felt as essential by the DM. So it could be an outline of the relevant shapes in the image, or a symbol semantically referring to its content. Several images might be associated to the same keypic, and more than one keypic might be associated to the same image. Keypics could also be used for indexing Web pages or sites.

We believe it unavoidable that the link between the semantic and geometric levels be realized by a human operator. We also stress that this should not lead to a definite set determined by an external authority. It is probable — and even desirable — that preferred keypics arise spontaneously within the Internet community, but such attractors should be left free to appear, modify and disappear in time.

But is there a technical tool capable of retrieval of such plastic drawings? A possibility (surely not the only one and perhaps not the best) is provided by Size Functions. They are geometrical-topological shape descriptors which have proven useful for qualitative comparison, i.e. comparison performed when the intrinsic metric between shapes is either unknown or not completely clear. We support our proposal with some examples of retrieval from a dataset of 498 clip-art images, each equipped with a keypic.

2. GRAPHICAL METADATA

The idea of using graphical metadata is surely not new. The most common example is perhaps that of road signs; although some text often accompanies 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. The seat (for the first time in Asia) and the fact that it was going to be a massive TV event, suggested the use of a well–defined set of symbols.

We suggest that images on the Internet should be equipped with simplified sketches representing the essentials of the images themselves. The sketches should be provided by the image owner or manager. This graphical indexing might be extended to whole Web pages.

So, our use of the term "graphical metadata" is not to be confused with that, e.g, of graphical modelling in statistics[1, 2] or with the schemes of NBII[3] and of FGDC.[4]

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, as, e.g., SVG. 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 (e.g. for this very paper). 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. Still, we think of them as the graphical analog of keywords; this is why we call them "keypics".

2.1. Automatic graphical indexing?

The ideal situation would be that the semantic content of an image were directly understood and automatically extracted by a computer program. While this is still science fiction (at least for a general purpose software), it could be hoped to have software which automatically extracts at least the relevant low-level features: The meaningful edges or, dually, the meaningful regions. But this is again beyond the present possibilities of any edge detector and of any segmentation tool.[5]

It is not just a matter of state–of–art. Placing data on the Internet and retrieving them is a human–to–human event; it is a form of human communication. The semantic content of an image is a highly subjective matter; reproducibility and objectivity, which are extremely important, e.g., in medical diagnosis and would make a smart machine even preferable to a human, are here a drawback.

We think that the drawing of keypics should definitely be performed by human operators, focusing the aspects of shapes that they consider important for recognition and retrieval. In this way semantic comparison of shapes is partially reduced to geometrical comparison of icons. A DM, e.g., might wish to index the image of a saxophone by its geometrical outline, but also (or only) with a musical note.

Of course, current image processing programs can be used in a fruitful way as a tool for indexing. This was actually the choice of some of our volounteers while drawing some keypics (see 3.1). This is in conflict with our suggestion to use SVG or a similar standard, but this divergence is likely to be smoothed in near future.

2.2. The importance of plasticity

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. Of course, a dictionary of icons with a number of items comparable with that of a language dictionary would be of no practical use. 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 DM. This does not mean that stereotypes should be avoided; only that they should not be imposed.

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.

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

2.3. Social issues

A first problem is: How to make the idea of keypics work? In order to be effective, it should be adopted by literally millions of users. This might of course be the case, if the idea were made concrete in a commercial product, but we prefer the scenario of a free trend, possibly driven by an organization such as the Free Software Foundation. Success might also be granted, if some research engines made a search–by–keypics option available.

A second problem could be a malicious use of keypics: Some particular icons might turn out to be frequently retrieved even if the user is looking for something else (we think, e.g., of a single dot). Then, an opportunistic DM might want to use such icons, independently of any semantic or geometric connections with the offered images. Since we think that keypics might be used to index Web pages, and not only images, this might very well be the case. Possible countermeasures might be some loose sort of control, as with the Wikipedia, or simply the elimination or penalization of such icons in the search engines.

2.4. A possible tool: Size Functions

The choice — that we insist to consider unsatisfactory — of a fixed set of icons, would have the advantage of an easy retrieval. Simple superimposition would yield an immediate distance by the mere count of pixels in the symmetric difference. On the other hand, besides the drawbacks we pointed out in the previous section, all images carrying the same standard keypic would be retrieved with the same score.

Nonstereotyped keypics would allow for finer distinctions. But there is the problem of comparing the shapes of sketches, which could also be very rough, and in any case would present great variability even within the same represented category.

There is a tool specifically developed for comparison of “natural” shapes: Size Functions (SF’s). They are modular transforms based on the geometry and topology of the image. They are best suited to catch qualitative features in a quantitative way: Their application is particularly useful when no standard, geometric templates are available and when the intrinsic metric between shapes is either unknown or not completely clear. Examples of applications are recognition of tree-leaves, hand-drawn sketches, monograms, hand-written characters, white blood cells and the sign alphabet.

SF’s depend on the choice of measuring functions, i.e. real functions defined on the image M; a SF actually condenses the behaviour of a measuring function in a function defined on the plane with values in the natural numbers. The discontinuities of the SF mark the merging of different connected components of the excursion sets $\{P \in M \,|\, \phi(P) \leq x\}$ of the measuring function $\phi : M \to \mathbf{R}$ while x varies in $\mathbf{R}$. For the definition of SF’s and for the main theorems of the theory, see Frosini and Landi.[6, 7]

Although we are fairly satisfied with the results (see 3.2), we do not believe SF’s to be the definitive answer to the problem of keypic retrieval: They are nonetheless the expression of a possibility. Alternative or complementary methods are possible and welcome.

3. EXPERIMENTAL RESULTS

We selected 498 images out of a clip–art collection, and distributed them among seven nonprofessional draftsmen. These were invited to draw, by means of a common drawing program and using a mouse, a keypic for each picture. The only constraints were the indication to draw black on white on a 150×150 canvas. They were explicitly left free to choose whether to give a semantic representation or a geometrical simplification of the picture.

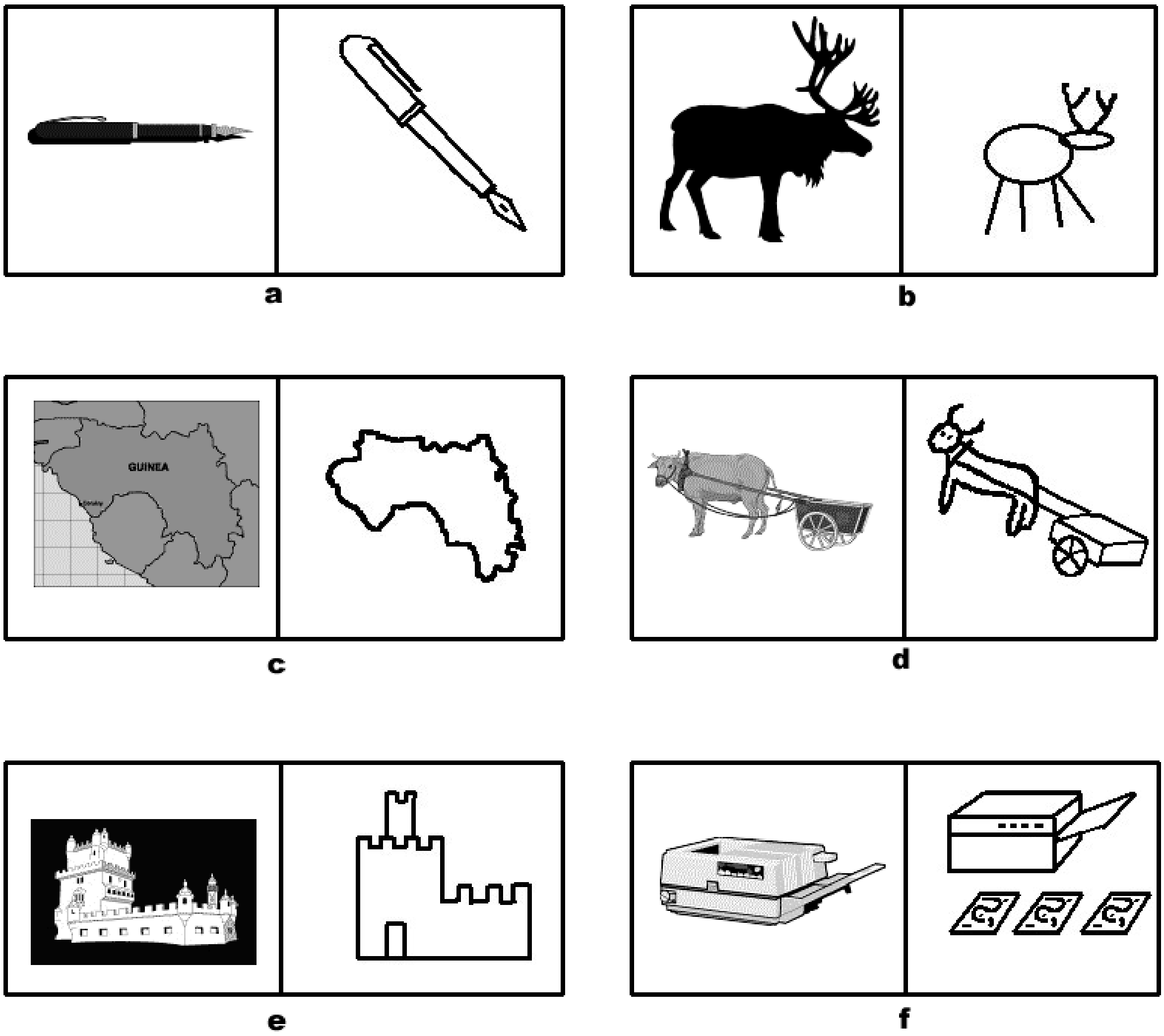

Figure 1. Different strategies in drawing keypics

3.1. The data

The strategies adopted were very heterogeneous. Some drew a fairly accurate imitation as in Figure 1a. Sometimes the imitation was very rough (Figure 1b); in other cases (e.g. in Figure 1c) the use of an edge detector was evident. Some draftsmen thought it necessary to stress details (Figure 1d), or to ignore them (Figure 1e, but sometimes even to add nonexisting ones (Figure 1f).

After a moment's perplexity, we accepted this variety of approaches. In fact, we think that a DM will stress the aspects and cure the details of what he/she considers essential in the images. So his/her keypics will be particularly high in score for the users "tuned on the same wavelength", i.e. interested in the same aspects and the same details.

3.2. Retrieval

The measuring functions chosen for this retrieval experiment were 16 "range" functions[8] yielding, for each black pixel of the image, its distance from a fixed point of a 16 points grid. The SF's of a query image and the ones of a dataset image were then coded by their cornerpoints and cornerlines, on whose sets an adapted Hausdorff

	avg	min	# at min	max	# at max
Rank	0.0829	0	53	0.9960	1
$P(N_{rel})$	0.6434	0	22	1	54
$P(2N_{rel})$	0.3422	0	15	0.5	58
$P(3N_{rel})$	0.2311	0	19	0.33	59
$R(2N_{rel})$	0.6735	0	19	1	79
$R(3N_{rel})$	0.6824	0	19	1	79

Table 1. Evaluation of results.

distance was computed. The maximum of the 16 distances was finally used as a distance between the query image and the one in the dataset.

Six users were invited to submit queries by drawing with a tool similar to the one used for producing the images. Of them, three had also been authors of the original images. No pre–processing of the images was needed. 100 queries were submitted. The number of relevant items N_{rel} for each queried class is greatly variable: it goes from a minimum of 1 (for 70 queries) to a maximum of 14 (for 8 queries).

Just for the sake of completeness, in Table 1 we report the results, although the scope of this article is not centered on the effectiveness of this particular retrieval system, but on the idea of graphical indexing.

We adopted the *normalized average rank* introduced by Müller *et al.*[9]:

$$Rank = \frac{1}{NN_{rel}}\left(\sum_{i=1}^{N_{rel}} R_i - \frac{N_{rel}(N_{rel}+1)}{2}\right)$$

where R_i is the rank at which the ith relevant image is retrieved, N is the dataset size, and N_{rel} is the number of relevant images for a given query. It is 0 for perfect performance and approaches 1 as performance worsens. We have also computed $P(k)$ and $R(k)$, respectively *precision* and *recall* on the first k retrieved images, with $k = N_{rel}, 2N_{rel}, 3N_{rel}$. (Of course, $R(N_{rel}) = P(N_{rel})$). We have considered these parameters meaningful, in view of the variability of N_{rel} (from 1 to 14). Explicitly,

$$P(k) = \frac{NRI(k)}{k} \qquad R(k) = \frac{NRI(k)}{N_{rel}},$$

where $NRI(k)$ is the number of relevant items among the first k retrieved.

For each evaluation parameter, the average, minimum and maximum value are given; the columns denoted by "# at max" and "# at min" show the numbers of queries reaching the minimum and maximum score respectively of each computed parameter. Note that $P(2N_{rel})$ and $P(3N_{rel})$ can reach at most values 1/2 and 1/3.

Figure 2 shows the first five retrieved keypics, together with the corresponding images, of two successful queries. We remark that, given those two query sketches, it would have been very difficult to reach the retrieved images without the intermediation of keypics.

4. CONCLUSIONS

Keypics — i.e. graphical metadata synthesizing the images to be indexed — might build the bridge over the semantic gap in image retrieval. In our opinion, they should be simple and possibly coded in a compact, standard way. They should be drawn by humans, who would catch and stress the relevant semantic or geometric features of the indexed images; this would also perform a broad selection of the target user. Keypics should absolutely be plastic, in the sense that they should be allowed to vary from author to author.

The technical problem of keypic retrieval has at least one candidate solution: the theory of Size Functions. This tool has been used rather successfully in the explanatory experiments of this paper.

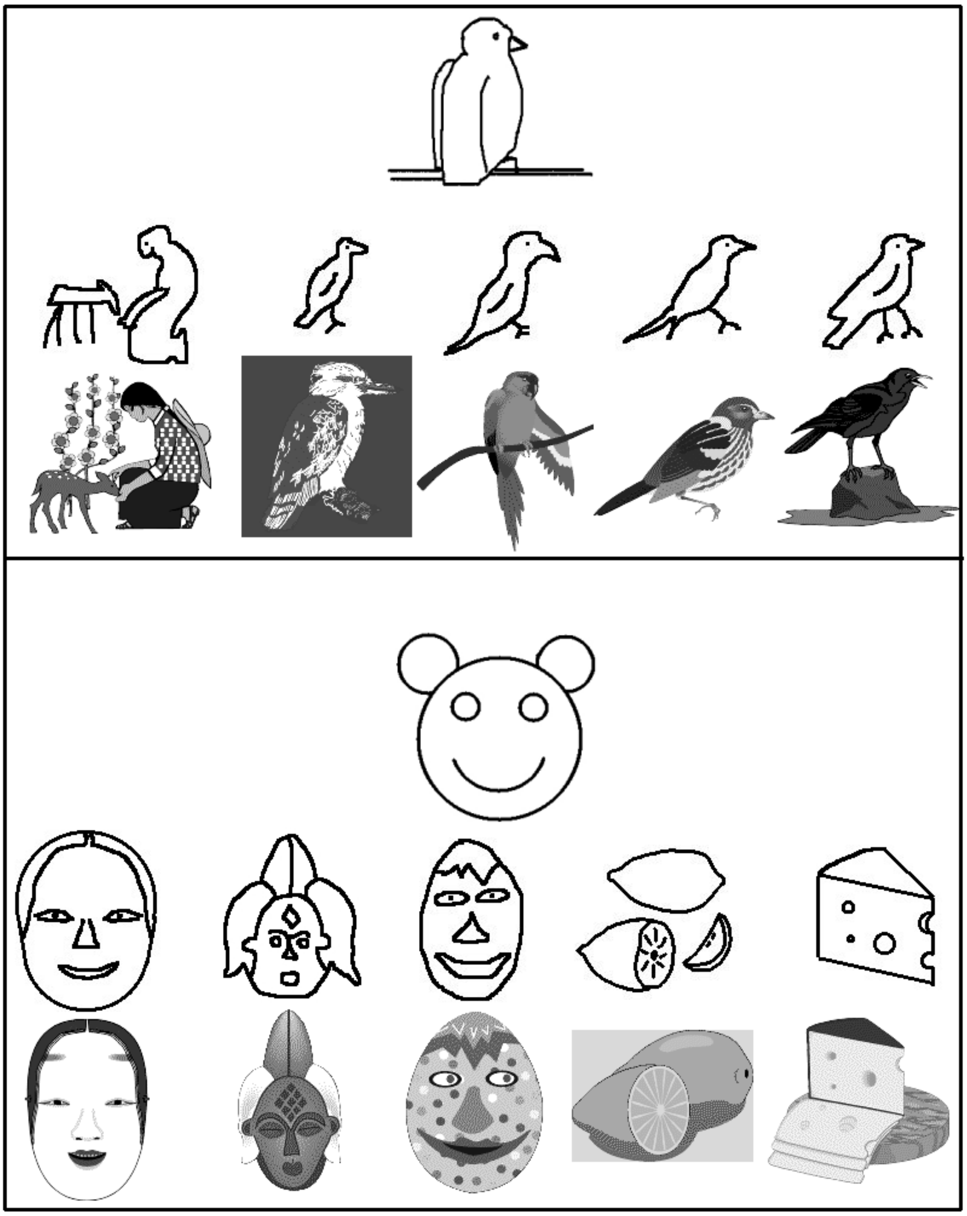

Figure 2. The first five results of two queries

ACKNOWLEDGMENTS

Work performed within the activity of ARCES "E. De Castro" and CIRAM, with funds of the University of Bologna for selected research topics, and of INdAM-GNSAGA, MACROGEO and ASTA2. We are deeply indebted to all friends of the VM group and of ImaVis for their intensive collaboration.

REFERENCES

1. D. Edwards, *Introduction to Graphical Modelling*, Springer Texts in Statistics, Springer, 2000.
2. S.L. Lauritzen, *Graphical Models*, Oxford University Press, 1996.
3. National Biological Information Infrastructure, "Nbii metadata standards." http://www.nbii.gov/datainfo/metadata/standards/, 1999.
4. Federal Geographic Data Committee, "Content standard for digital geospatial metadata." http://www.fgdc.gov/metadata/contstan.html, 1998.
5. C. Carson and S. Belongie and H. Greenspan and J. Malik, "Blobworld: Image segmentation using expectation-maximization and its application to image querying," *IEEE Trans. on PAMI* **24**(8), pp. 1026–1038, 2002.
6. P. Frosini and C. Landi, "Size theory as a topological tool for computer vision," *Pattern Rec. and Image Analysis* **9**, pp. 596–603, 1999.
7. P. Frosini and C. Landi, "Size functions and formal series," *Appl. Algebra in Engin. Commun. and Comput.* **12**(327-349), 2001.
8. M. Ferri and P. Frosini, "Range size functions," in *Vision Geometry III*, *Proc. SPIE* **2356**, pp. 243–251, 1994.
9. H. Müller and W. Müller and D.M. Squire and S. Marchand–Maillet and Th. Pun, "Performance evaluation in content–based image retrieval: Overview and proposals," *Pattern Rec. Letters* **22**, pp. 593–601, 2001.