

A Complete Keypics Experiment with Size Functions

Andrea Cerri, Massimo Ferri, and Daniela Giorgi

ARCES and Dept. of Mathematics, University of Bologna,
Piazza di Porta S. Donato, 5 I-40126 Bologna Italy
{cerri, ferri, giorgid}@dm.unibo.it

Abstract. Keypics are graphical metadata intended for indexing of images on the Internet. They are conceived as hand-drawn sketches, not restricted to a definite set. An obvious difficulty when dealing with keypics is that they elude rigid geometric treatment.

A proposal of solution comes from Size Functions. This paper is the report of a complete experiment on 494 keypics with Size Functions based on three measuring functions (distances, projections and jumps) and their combination.

1 Introduction

The recent, original idea (by P. Frosini and our Vision Mathematics Group in Bologna) [7] of “Keypics” poses serious problems to an Image Retrieval (IR) System. In fact, the main difficulty inherent to the “search-by-sketch” paradigm is by-passed by Keypics: “search-by-sketch” doesn’t work well because it tries to match sketches with real images; with Keypics, an IR System needs to match sketches with sketches. This seems to be more promising; still, the problem of comparing hand-drawn sketches for similarity is hard, as for any recognition and retrieval system which has to deal with “natural” images (like signatures, biological images, faces, the sign alphabet etc.).

Size Functions (SF’s) have proven to be particularly apt to this kind of settings, because of their geometrical-topological nature and their modularity. The mathematical core of SF’s was exactly conceived for formalizing qualitative aspects of signals (images, but also 3D data, sounds, etc.). Modularity allows the user to fit a SF to the specific nature of the objects to be recognized or retrieved, through the choice of a “measuring function” (see Section 3).

When proposing [7], our team had just performed a preliminary study on a set of Keypics, with only one type of measuring functions: distances from a set of fixed points. Here we enhance that experiment by using two more sets of independent measuring functions: “projections” and “jumps”, and finally by integrating the three subsystems (see Section 5).

Two necessary clarifications: 1) The adjective “complete” in the title refers to the fact that the present experiment considers not only the retrieved sketches,

but also the images that they are indexing. So it is really an experiment on keypics, whereas our preliminary studies were only on retrieval of sketches by sketches. 2) We want to show that image retrieval through keypics has at least Size Functions as a possible solution. We do not claim that Size Functions are a better tool than the competitors'; in fact, we invite other researchers to try their methods on keypics. We shall be very glad, for instance, to make our dataset available for comparison and integration of retrieval methods.

2 Keypics

In [7] it was proposed that the owner or the manager of a set of images on a Web site, 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 [1]. The Keypic should be representative of what is felt as essential by the site manager. 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.

Keypics 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. 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. This could be a partial reply to the warning, contained in [16], that “information is not only in the pixels”. A dataset manager, e.g., might wish to index the image of a saxophone by its geometrical outline, but also (or only) with a musical note (see Figure 1, which also hints that a Keypic can give evidence to what is unclear or incomplete in the original image).

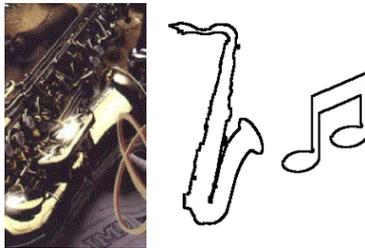


Fig. 1. An image and two possible Keypics for it

3 Size Functions

The key idea underlying SF's is that of setting metric obstructions to the classical notion of connectedness, thus conveying informations both on the geometry and the topology of the viewed shape.

Let us recall the formal definition of a SF. Consider a continuous real-valued function $\varphi : M \rightarrow \mathbb{R}$, defined on a subset M of a Euclidean space. The *Size Function* of the pair (M, φ) is a function $\ell_{(M, \varphi)} : \{(x, y) \in \mathbb{R}^2 \mid x < y\} \rightarrow \mathbb{N}$. For each pair $(x, y) \in \mathbb{R}^2$, consider the set $M_x = \{P \in M \mid \varphi(P) \leq x\}$. Two points in M_y are then considered to be equivalent if they belong to the same connected component of M_y . The value $\ell_{(M, \varphi)}(x, y)$ is defined to be the number of the equivalence classes obtained by quotienting M_x with respect to the previous equivalence relation in M_y .

SF's have a simple geometric interpretation: $\ell_{(M, \varphi)}(x, y)$ is equal to the number of connected components of M_y containing at least one point of M_x .

A discrete version of the theory exists, which substitutes the subset M of the space with a graph $G = (V, E)$, the function $\varphi : \mathcal{M} \rightarrow \mathbf{R}$ with a function $\varphi' : V \rightarrow \mathbf{R}$, and the concept of connectedness with the usual connectedness notion for graphs.

It is important to remark that SF's are easily and fast computable; see [4] for details.

Figure 2 shows a simple example of SF. In this case the topological space M is a curve, while the measuring function φ is the distance from point C .

As Figure 2 shows, SF's have a typical structure: They are linear combinations (with natural numbers as coefficients) of characteristic functions of triangular regions. That implies that each SF can be described by a formal linear combination of *cornerpoints* and *cornerlines*. Due to this kind of representation, the original complex issue of comparing shapes can be turned into a simpler algebraic problem: Each distance between formal series naturally produces a distance between SF's. A detailed treatment of this subject can be found in [8].

Of the many available distances between formal series(see, e.g., [5]), the one we use in this paper is the Hausdorff distance.

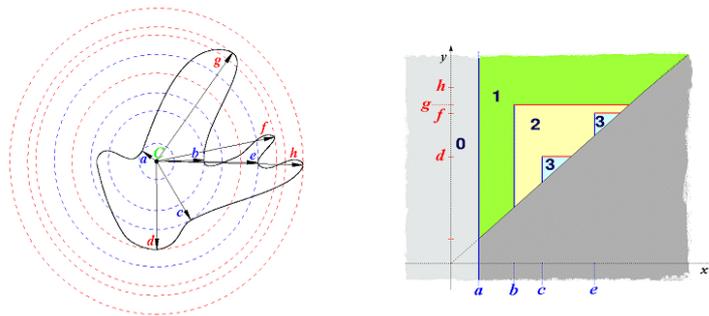


Fig. 2. A curve and its SF with respect to the distance from point C

4 The Dataset

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.

The strategies adopted were very heterogeneous. Some drew a fairly accurate imitation as in Figure 3a. Sometimes the imitation was very rough (Figure 3b); in other cases (e.g. in Figure 3c) the use of an edge detector was evident. Some draftsmen thought it necessary to stress details (Figure 3d), or to ignore them (Figure 3e), but sometimes even to add nonexisting ones (Figure 3f).

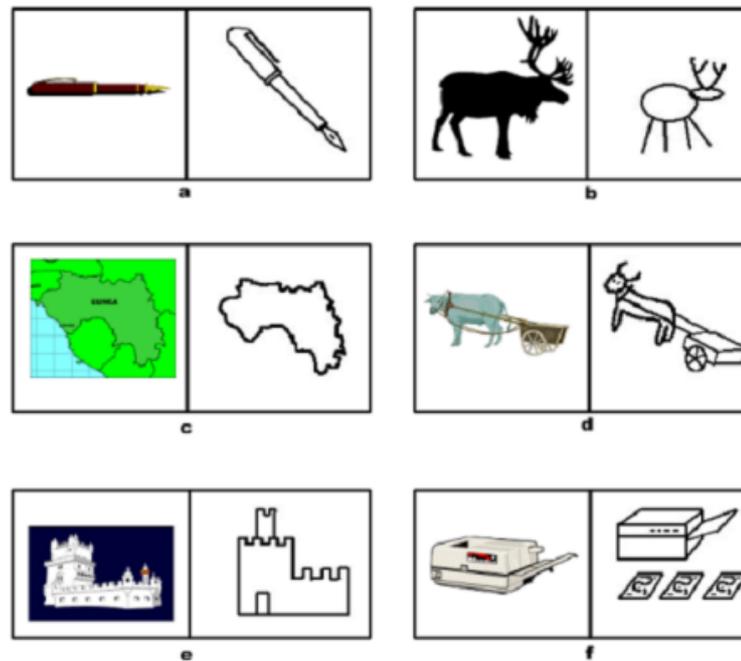


Fig. 3. Different strategies in drawing Keypics

After a moment's perplexity, we accepted this variety of approaches. In fact, we think that a dataset manager 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.

5 Measuring Functions

Three different and independent sets of measuring functions were used. Of these, only the first set had already been used in [7]. The first set consists of sixteen distances from points [6].

Let us fix a Cartesian reference frame (O, e_1, e_2) in the plane. From now on points will be identified with their coordinate pairs. Let $p = (p_x, p_y) \in \mathbb{R}^2$. We define the measuring function $\varphi_p: \mathbb{R}^2 \rightarrow \mathbb{R}$ as $\varphi_p(x, y) = d(p, (x, y))$ with d the Euclidean distance.

Every input binary image is normalized (but without resolution loss) and translated so that its center of mass is taken to the origin of the reference frame. Therefore each measuring function φ_p is invariant by scale change and translation; as a consequence, the corresponding SF's turn out to be invariant by the same transformation group.

Here is the formal definition of the first set of measuring functions used in this research:

$$\begin{aligned} \Phi = & \\ & \{ \varphi_p \mid p = \frac{\bar{r}}{2} (\cos(\bar{\alpha} + i\frac{\pi}{2}), \sin(\bar{\alpha} + i\frac{\pi}{2})) \ i = 1, \dots, 4 \} \cup \\ & \{ \varphi_p \mid p = \bar{r} (\cos(\bar{\alpha} + i\frac{\pi}{4}), \sin(\bar{\alpha} + i\frac{\pi}{4})) \ i = 0, \dots, 7 \} \cup \\ & \{ \varphi_p \mid p = \frac{3}{2}\bar{r} (\cos(\bar{\alpha} + i\frac{\pi}{2}), \sin(\bar{\alpha} + i\frac{\pi}{2})) \ i = 1, \dots, 4 \}, \end{aligned}$$

where the constants \bar{r} and $\bar{\alpha}$ take value respectively 0.8 (all images are scaled with respect to average radius) and 0.349 (approximately corresponding to a 20 degrees phase-displacement).

The next two sets of measuring functions are much less rigid. The second set contains five measuring functions, having each a segment as domain. One of the five is a "projection" of the image on the horizontal base segment: The whole image is fibered into a set of vertical pixel segments; for each of these, the number of black pixels contained in it is counted. The corresponding pixel of the horizontal base segment receives this number. The final measuring function is obtained by convolving these values with a narrow Gaussian. The other four measuring functions are its variations built by projecting along the horizontal direction and along the three at $\pi/8, \pi/4, 3\pi/8$.

The third set consists of four functions. One counts "jumps" along the vertical direction. Again, the whole image is fibered into a set of vertical pixel segments; for each of these, a counter is incremented each time two consecutive pixels of the vertical segment are of opposite color. The corresponding pixel of the horizontal base segment receives this number of black-to-white and white-to-black jumps. Again, convolution with a narrow Gaussian yields the final measuring function. In this case, the other three measuring functions are its variations built by counting jumps along the horizontal direction and along the two at 45 degrees.

Retrieval was performed with each of the three sets of SF's by computing the average of the normalized distances coming out of the different SF's of the set. A final distance combines the contribution of the three.

6 Evaluation

As stressed in several papers (e.g. [14], [15] and [11]) evaluation is a very critical issue for IR Systems. Apart from the common problem of possessing a reliable and objective ground truth, all most common parameters have some drawbacks.

A particular fault of several evaluation methods, is that they don't take sufficiently well into account the position of the retrieved relevant objects within the scope (i.e. within the whole retrieved set). In what follows, we try to overcome this problem in two ways. First, we adopt the *normalized average rank* \widetilde{Rank} introduced by [15]:

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

where R_i is the rank at which the i th 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.

Second, we have also computed $P(k)$ and $R(k)$, respectively *precision* and *recall* on the first k retrieved images, with $k = N_{rel}, 2N_{rel}, 3N_{rel}$, so adapting the scope to the (varying) number of relevant objects, rather in the line of the normalizations supported by [11]. (Of course, $R(N_{rel}) = P(N_{rel})$) Explicitly,

$$P(k) = \frac{NR(k)}{k} \quad R(k) = \frac{NR(k)}{N_{rel}},$$

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

7 Experimental Results

20 queries were submitted, in the form of sketches belonging to the dataset. The following tables (1 to 4) gather the results for the three measuring functions teams and for their combination. For each evaluation parameter, the average, minimum and maximum value are given. These values are followed by the number of queries reaching the minimum and maximum score respectively (indicated as “# at min” and “# at max”).

Table 1. Evaluation of results for the distances

	avg	min	max	# at min	# at max
\widetilde{Rank}	0.1970	0.0	0.6839	2	1
$P(N_{rel})$	0.4831	0.07	1.0	3	6
$P(2N_{rel})$	0.3558	0.04	1.0	3	4
$P(3N_{rel})$	0.2682	0.02	1.0	1	1
$R(2N_{rel})$	0.4314	0.10	1.0	1	2
$R(3N_{rel})$	0.5046	0.10	1.0	1	3

Table 2. Evaluation of results for the projections

	avg	min	max	# at min	# at max
\widetilde{Rank}	0.2360	0.0171	0.5655	1	1
$P(N_{rel})$	0.4795	0.07	1.0	5	7
$P(2N_{rel})$	0.2701	0.04	1.0	2	4
$P(3N_{rel})$	0.2096	0.02	1.0	2	1
$R(2N_{rel})$	0.3543	0.10	0.67	1	1
$R(3N_{rel})$	0.3942	0.10	0.75	1	1

Table 3. Evaluation of results for the jumps

	avg	min	max	# at min	# at max
\widetilde{Rank}	0.1539	0.0010	0.4622	7	1
$P(N_{rel})$	0.4962	0.07	1.0	4	7
$P(2N_{rel})$	0.2886	0.04	1.0	2	1
$P(3N_{rel})$	0.2376	0.02	1.0	2	1
$R(2N_{rel})$	0.3887	0.07	0.75	1	1
$R(3N_{rel})$	0.4274	0.14	1.0	1	1

Table 4. Evaluation of results for the combination

	avg	min	max	# at min	# at max
\widetilde{Rank}	0.1794	0.0	0.4852	1	1
$P(N_{rel})$	0.5117	0.07	1.0	2	7
$P(2N_{rel})$	0.3719	0.04	1.0	2	3
$P(3N_{rel})$	0.2703	0.02	1.0	2	1
$R(2N_{rel})$	0.4749	0.10	1.0	1	2
$R(3N_{rel})$	0.4856	0.10	1.0	1	2

The number of relevant items N_{rel} for each queried class is greatly variable: it goes from a minimum of 2 to a maximum of 14. The reader should keep in mind that good ranks have low values, while good precision and recall have high scores.

The precision–recall graph of Figure 4 (left) refers to the combined distance. The system runs on an Athlon 900 MHz based PC, under Linux OS. Response time for a query is, for the moment, between 7 and 10 seconds.

Figure 4 (right) depicts the GRiP graph, plotting the value of precision=recall versus $-\log_2(g)$, where the *generality* g is the ratio of the number of relevant items for each query (2 to 14) by the total size of the data set (494) [11].

Browsing the actual outputs of the queries is rather interesting. For instance, the query consisting of a stylized bird yields a sequence of as uninteresting bird

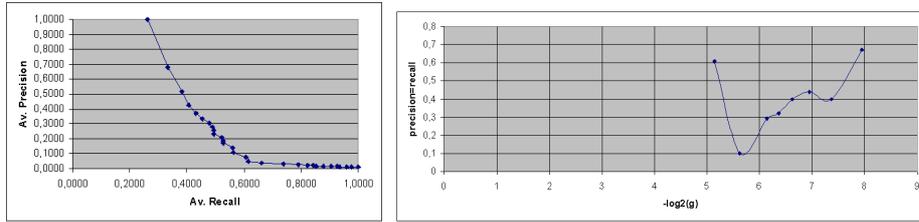


Fig. 4. Precision–recall (left) and GRiP graph for the combined distance

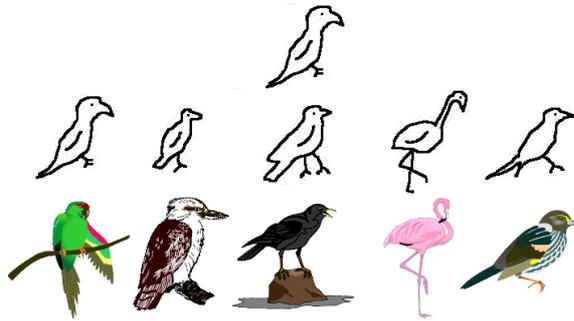


Fig. 5. A successful query

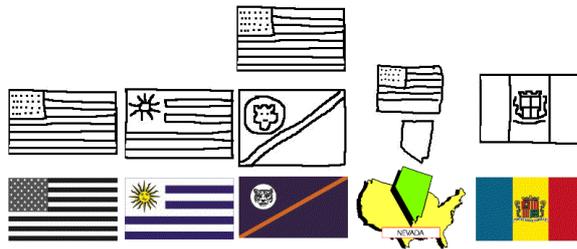


Fig. 6. An unexpected output

sketches, as Keypics; things turn interesting if we look at the real images to which the Keypics point (Figure 5): Without the intermediation of the Keypics — rough and childish as they may appear — it is unlikely that a “normal” query would have retrieved such heterogeneous images. More remarkable is a query with the USA flag, where the map of Nevada pops 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 6; by the way, that image has *not* been considered as relevant in the statistics of this Section).

8 Other Descriptors

Of course, there are several shape descriptors and shape matching methods (see, e.g., [17]). A comparison of methods is really possible only when performed on the same data. We have actually done it recently on a data set of trade marks [3]. For the moment, let us examine similarities and differences with competitors.

Take the classical Fourier descriptors, for instance [9]. A coincidence in the first few coefficients would grant that two shapes are roughly superimposable, differences being limited to the higher frequencies. SF's (with distance from center of mass as a measuring function) recognize the presence and size of comparable bumps even if they are differently disposed in the two images to compare; this is a simple case of similarity with no (even rough) superimposition. We remind that, anyway, also SF's enjoy completeness theorems like Fourier descriptors [6].

Other descriptors, like order structure [2], turning function [13], chain code histogram [12], need a sort of local superimposition and are mainly limited to silhouettes. A much closer relative to size functions is the Reeb Graph [10], which has been used so long — as far as we know — only for 3D objects.

9 Conclusions

Keypics — plastic graphical metadata — cannot have a chance of diffusing and succeeding as universal bridges of the semantic gap, unless powerful, qualitative tools are developed for comparing and retrieving hand-drawn sketches.

The feasibility of Keypics is shown by the experiment reported here. Size Functions propose themselves as a possible candidate for retrieving images through Keypics. Our research shows that different measuring functions can integrate together effectively. Of course, integration with still different methods should give Keypics an even better chance.

Acknowledgements

Work performed within the activity of ARCES “E. De Castro” and of CIRAM, University of Bologna, under the auspices of INdAM-GNSAGA and of the University of Bologna, funds for selected research topics.

References

1. <http://w3.org/Graphics/SVG/About.html>
2. Carlsson, S.: Order structure, correspondence, and shape based categories. In: D.A. Forsyth et al. (Eds.): Shape, Contour and Grouping in Computer Vision, LNCS 1681 (1999) 58–71.
3. Cerri, A., Ferri, M., Giorgi D.: A New Framework for Trademark Retrieval Based on Size Functions. To appear on: Proc. 2nd International Conference on Vision, Video and Graphics, 7-8 July 2005 Heriot Watt University, Edinburgh

4. d'Amico, M.: A New Optimal Algorithm for Computing Size Functions of Shapes. In: CVPRIP Algorithms III, Proc. Intl. Conf. on Computer Vision, Pattern recognition and Image Processing, Atlantic City, (2000) 107–110.
5. Donatini, P., Frosini, P., Landi, C.: Deformation energy for size functions. In: E.R. Hancock, M. Pelillo (eds.) Energy Minimization Methods in Computer Vision and Pattern Recognition,, LNCS 1654 (1999) 44–53.
6. Ferri, M., Frosini, P.: Range size functions. Proc. SPIE Conf. on Vision Geometry III, Boston, 1994 Nov. 2–3 (1995) 243–251.
7. Ferri, M., Frosini, P.: A proposal for image indexing: “keypics”, plastic graphical metadata Proc. IS&T/SPIE Symp. on Electronic Imaging, Internet Imaging VI, San Jose, 2005 Jan. 16–20 (2005).
8. Frosini, P., Landi, C.: Size functions and formal series. *Applicable Algebra in Engineering Communication and Computing*, 12 (2001) 327–349
9. Granlund, G.H.: Fourier preprocessing for hand print character recognition. *IEEE Trans. Computers*, C-21 (1972) 195–201.
10. Hilaga, M., Shinagawa, Y., Kohmura, T., Kunii, T.L.: Topology matching for fully automatic similarity estimation of 3D shapes. SIGGRAPH 2001, Computer Graphics Proc., Annual Conference Series (2001) 203–212.
11. Huijsmans, D.P., Sebe, N.: How to Complete Performance Graphs in Content-Based Image Retrieval: Add Generality and Normalize Scope *IEEE Trans. on PAMI* **27** (2005), 245–251.
12. Iivarinen, J., Visa, A.: Shape recognition of irregular objects. In: D. P. Casasent (ed.) *Intelligent Robots and Computer Vision XV: Algorithms, Techniques, Active Vision, and Materials Handling*, Proc. SPIE 2904 (1996) 25–32.
13. Leung, M-W., Chan, K.-L.: Object-based image retrieval using hierarchical shape descriptor. In: M.S. Lew, N. Sebe, J.P. Eakins (Eds.): *Image and Video Retrieval*, LNCS 2383 (2002) 165–174.
14. Liu, W., Su, Z., Li, S., Zhang, H. J.: A Performance Evaluation Protocol for Content-Based Image Retrieval Algorithms/Systems. In: Proc. IEEE CVPR Workshop on Empirical Evaluation in Computer Vision, Kauai, USA, December, 2001.
15. Müller, H., Müller, W., Squire, D.M., Marchand-Maillet, S., Pun, Th.: Performance evaluation in content-based image retrieval: Overview and proposals. *Pattern Rec. Letters* 22 (2001) 593–601.
16. Petkovic, D., Jain, R. C.: Visual Information systems: lessons for its future. Proc. IS&T/SPIE Symp. on Electronic Imaging, Internet Imaging VI, San Jose, 2005 Jan. 16–20 (2005).
17. Veltcamp, R.C., Hagedoorn, M.: State-of-the-art in shape matching. In: M. Lew (Ed.), *Principles of Visual Information Retrieval*, Springer (2001) 87–119.