Keypics: free–hand drawn iconic keywords

Andrea Cerri, Massimo Ferri *Member, IEEE,* Patrizio Frosini, Daniela Giorgi

***Abstract*— We propose an iconic 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, in the retrieval task, already exists: Size Functions.

An experiment on 494 Keypics with Size Functions based on three measuring functions (distances, projections and jumps) and their combination is presented.

***Index Terms*— Image Retrieval, Size Function, Measuring Function, Iconic Metadata.**

I. Introduction

The number of images and multimedia digital shapes populating the Web is rapidly growing, resulting in a huge amount of information potentially available. At the same time, handling this data traffic in a way to ease the access to the desired visual resources is widely acknowledged as one of the primary contemporary challenges. In the current search engines, users can simply browse the resources, or perform searches relying on the use of textual metadata associated to the items. This paper proposes a new technique for indexing images for data retrieval. The main idea is to associate each image with a subjective, non-standardized *visual* description, that we shall call "Keypic" (as alternative to "keyword"), chosen by the dataset manager (DM). This kind of iconic indexing places itself at an intermediate level between semantic and geometrical descriptions, while preserving the capability to support retrieval purposes. Indeed, in this way semantic information can be handled by using geometrical–topological tools. We empirically demonstrate that such iconic descriptions can be dealt with by using a mathematical tool already available for shape matching, and provide experimental evidence on the use of Keypics for image retrieval.

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 DM equips each image with a simplified drawing, the "Keypic", or even more than one. 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 (see Figure 1). Keypics could also be used for indexing Web pages, sites and all sort of files.

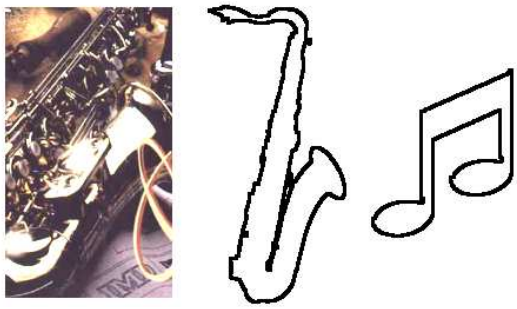

Fig. 1. An image and two possible Keypics for it.

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.

The main difficulty inherent to the "search–by–sketch" paradigm is bypassed 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. What we need is a tool capable to perform non-exact shape matching, in the absence of a specified shape grammar. The shape descriptors we propose, namely Size Functions (see Section II-D), 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".

We support our proposal with an experiment on 494 images equipped with Keypics (see Sections III, IV). We would like to stress that we do not claim Size Functions to be 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.

II. Iconic metadata

The idea of using iconic or graphical metadata is surely not new (see Figure 2). 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

Manuscript received ****; revised ****.

A.C., M.F. and P.F. are with ARCES and Dept. of Mathematics, University of Bologna, Piazza di Porta S. Donato 5 I–40126 Bologna, Italy. E–mail: {cerri, ferri, frosini}@dm.unibo.it

D.G. is with IMATI–CNR, Via de Marini 6, I–16149 Genova, Italy. E–mail: daniela@ge.imati.cnr.it

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.

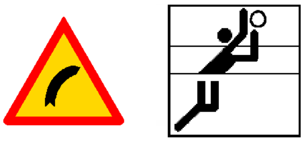

Fig. 2. Two well–known iconic signs.

We suggest that images on the Internet should be equipped with simplified sketches representing the essentials of the images themselves (see also [5], [8]). The sketches should be provided by the image owner or manager. This graphical indexing might be extended to whole Web pages.

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 [1]. 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".

A. 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 [4].

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 content that they consider important for recognition and retrieval. In this way semantic comparison is partially reduced to geometrical comparison of icons. This could be a partial reply to the warning, contained in [14], that "information is not only in the pixels". 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 (see Figure 1, which also hints that a Keypic can give evidence to what is unclear or incomplete in the original image).

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 III-A). 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.

B. 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 DM 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.

C. 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 [2]. Success might also be granted, if some research engines made a search–by–Keypics option available.

A second problem (well–known for keywords) 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.

D. A possible tool for Keypics retrieval: 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. They are modular transforms based on the geometry and topology of the shape. 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, white blood cells, the sign alphabet [15], [10] and retrieval of trade–marks [6] and 3D shapes [3].

Consider a continuous real-valued function $\varphi : M \to \mathbf{R}$, defined on a subset M of a Euclidean space (called a *measuring function*). The *Size Function* (SF) of the pair (M, φ) is a function $\ell_{(M,\varphi)} : \{(x, y) \in \mathbf{R}^2 \,|\, x < y\} \to \mathbf{N}$. For each pair (x, y) in the domain, consider the set $M_x = \{P \in M \,|\, \varphi(P) \le x\}$. 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 equivalence relation in M_y.

A SF actually condenses the behaviour of a measuring function in a function defined on the half–plane $x < y$ with values in the natural numbers. The discontinuities of the SF mark the birth and merging of different connected components of the excursion sets $\{P \in M \,|\, \varphi(P) \le x\}$ of the measuring function $\varphi : M \to \mathbf{R}$ while x varies in $\mathbf{R}$. More information on the theory can be found in [9], [7, Sections 8.4, 9.1].

Figure 3 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.

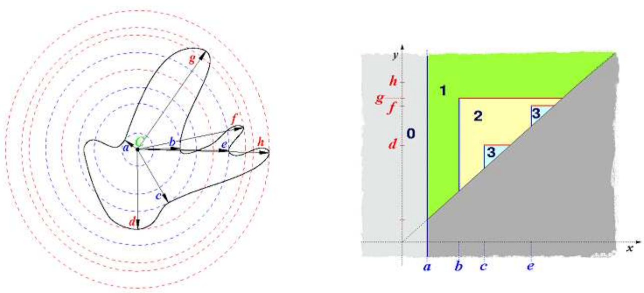

Fig. 3. A curve and its SF with respect to the distance from point C.

As Figure 3 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.

Of the many available distances between formal series, the one we use in this paper is the Hausdorff distance. Although we are fairly satisfied with the results (see Section IV), 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 techniques are possible and welcome (see [16] for a discussion on such methods). E.g., it has been remarked that Keypics might convey more information if colored; Size Functions are not yet equipped to deal with color, while other techniques are.

III. Experimental Setting

In order to validate our proposal of Keypics as iconic metadata, we built an experimental setting consisting of two different steps. The first phase involves human operators producing Keypics to label a given dataset of images. The second step aims at evaluating the performance of Keypics for image retrieval when Size Functions are used.

A. 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 4a. Sometimes the imitation was very rough (Figure 4b); in other cases (e.g. in Figure 4c) the use of an edge detector was evident. Some draftsmen thought it necessary to stress details (Figure 4d), or to ignore them (Figure 4e), but sometimes even to add nonexisting ones (Figure 4f).

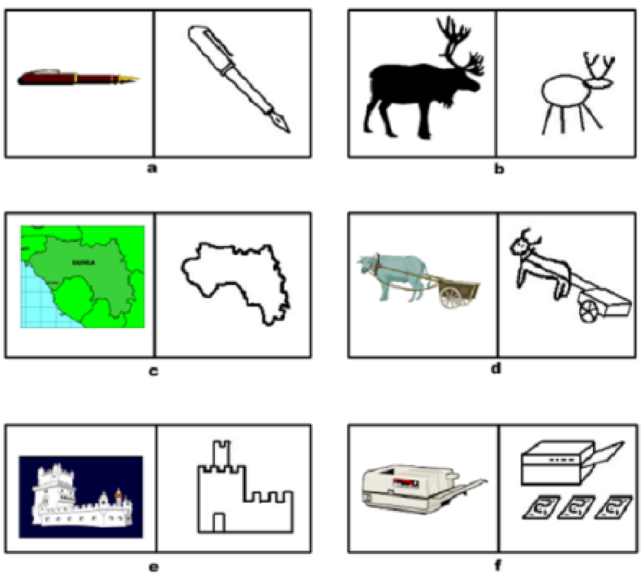

Fig. 4. Different strategies in drawing Keypics.

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.

B. Measuring functions

Three different and independent sets of measuring functions were used for Keypics description and comparison through Size Functions.

The first set consists of sixteen distances from points suitably chosen on the plane of the image. Every input binary image is beforehand 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 is invariant by scale change and translation; as a consequence, the corresponding SF's turn out to be invariant by the same transformation group.

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 $45°$.

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.

C. Evaluation parameters

As stressed in several papers (e.g. [11], [12] and [13]) 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 [13]:

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

where R_i is the rank at which the 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.

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,

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

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

	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

TABLE I

EVALUATION OF RESULTS.

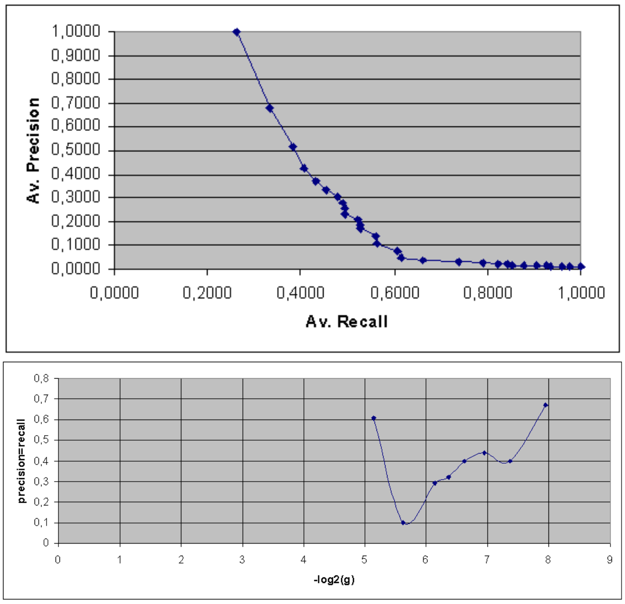

Fig. 5. Precision–recall (top) and GRiP graph (bottom).

IV. EXPERIMENTAL RESULTS

20 queries were submitted, in the form of sketches belonging to the dataset. The following table gathers the results for the combination of the three measuring functions teams. 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").

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 5 (top) refers to the combined distance. Figure 5 (bottom) 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].

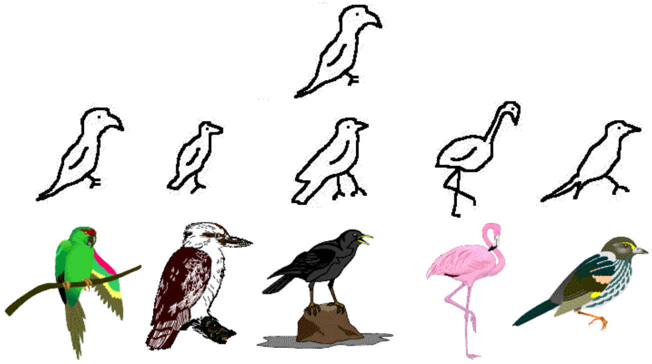

Fig. 6. A successful query.

Browsing the actual outputs of the queries is rather interesting. For instance, the query consisting of a stylized bird yields a sequence of equally uninteresting bird sketches (the Keypics); things turn interesting if we look at the real images to which the Keypics point (Figure 6): 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 7).

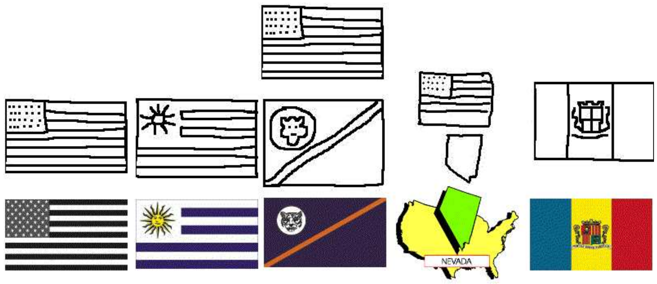

Fig. 7. An unexpected output.

V. Conclusions

Keypics — i.e. iconic 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 users. Keypics should absolutely be plastic, in the sense that they should be allowed to vary from author to author. Keypics 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 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] http://www.fsf.org/
[3] Biasotti, S., Giorgi, D., Spagnuolo, M. and Falcidieno, B., "Size functions for 3D shape retrieval", *Proc. EG Symp. on Geometry Processing*, pp.239–242, 2006.
[4] Carson, C., Belongie, S., Greenspan, H. and Malik, J., "Blobworld: Image segmentation using expectation–maximization and its application to image querying", *IEEE Trans. on PAMI* vol. 24, n. 8, pp. 1026–1038, 2002.
[5] Cerri, A., Ferri, M. and Giorgi, D., "A complete keypics experiment with Size Functions". *Proc. CIVR 2005, Singapore, Springer LNCS 3568*, W.-K. Leow, M.S. Lew, T.-S. Chua, et al. Eds., pp. 357–366, 2005.
[6] Cerri, A., Ferri, M. and Giorgi, D., "Retrieval of trademark images by means of size functions", *Graph. Models*, vol. 68, pp. 451–471, 2006.
[7] Kaczynski, T., Mischaikow, K. and Mrozek, M., *Computational homology*, Applied Mathematical Sciences 157, Springer, 2004.
[8] Ferri, M. and Frosini, P., "A proposal for image indexing: 'keypics', plastic graphical metadata", *Proc. SPIE Vol. 5670, Internet Imaging VI*, S. Santini, R. Schettini, Th. Gevers Eds., pp. 225-231, 2005.
[9] Frosini, P. and Landi, C., "Size theory as a topological tool for computer vision", *Pattern Rec. and Image Analysis*, vol. 9, n. 4, pp. 596-603, 1999.
[10] Handouyahia, M., Ziou, D. and Wang, S., "Sign language recognition using moment–based size functions", *Proc. Vision Interface '99, Trois-Rivières, Canada, 19-21 May*, pp. 210–216, 1999.
[11] Huijsmans, D.P. and Sebe, N., "How to Complete Performance Graphs in Content-Based Image Retrieval: Add Generality and Normalize Scope", *IEEE Trans. on PAMI*, vol. 27, pp. 245–251, 2005.
[12] Liu, W., Su, Z., Li, S. and Zhang, H. J., "A Performance Evaluation Protocol for Content-Based Image Retrieval Algorithms/Systems" *Proc. IEEE CVPR Workshop on Empirical Evaluation in Computer Vision, Kauai, USA, December, 2001*.
[13] Müller, H., Müller, W., Squire, D.M., Marchand–Maillet, S. and Pun, Th., "Performance evaluation in content–based image retrieval: Overview and proposals", *Pattern Rec. Letters*, vol. 22, pp. 593–601, 2001.
[14] Petkovic, D. and Jain, R. C., "Visual Information systems: lessons for its future", *Proc. SPIE Vol. 5670, Internet Imaging VI*, S. Santini, R. Schettini, Th. Gevers Eds., pp. 16–20, 2005.
[15] Uras, C. and Verri, A., "On the recognition of the alphabet of the sign language through size functions", *Proc. 12th Int. Conf. Patt. Recog., Jerusalem*, vol. 2, pp. 334–338, 1994.
[16] Veltkamp, R.C. and Latecki, L.J., "Properties and performances of shape similarity measures", *Proc. IFCS06, Data Science and Classification, Springer*, Batagelj et al. Eds., pp. 47–56, 2006.

Andrea Cerri received his degree in Mathematics from the University of Bologna in 2003. He is currently a PhD student at the University of Bologna. His research interests include geometrical-topological methods for shape comparison and applications in Pattern Recognition.

Massimo Ferri has been full professor of Geometry since 1981; he is presently at the University of Bologna, Engineering Faculty. He is the coordinator of WP2 of the Advanced Research Center on Electronic Systems for Information and Communication Technologies at the University of Bologna. His research interests smoothly turned from low–dimensional combinatorial topology to the application of geometry and topology to Robotics and Pattern Recognition.

Patrizio Frosini received the PhD degree in Mathematics from the University of Florence in 1991. Since 1993, he has been a researcher in the Faculty of Engineering at the University of Bologna. He is a member of the Advanced Research Center on Electronic Systems for Information and Communication Technologies at the University of Bologna. His research interests include the study of geometrical-topological methods for shape comparison and related applications in Pattern Recognition.

Daniela Giorgi got the Laurea Degree cum laude from the Dept. of Mathematics, University of Bologna, in 2002, and a Doctoral Degree in Computational Mathematics at the University of Padua, in 2006. Currently she's a member of the Shape Modelling Group of the Institute of Applied Mathematics and Information Technology of the National Research Council, Department of Genoa, Italy. Her research interests are related to computational topology and shape abstraction techniques for digital shape analysis, matching and retrieval.