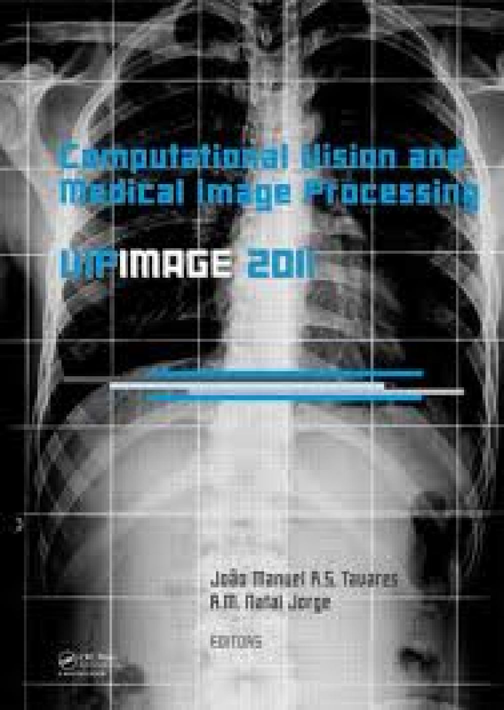
Computational Vision and
Medical Image Processing
VIPIMAGE 2011
João Manuel R.S. Tavares
R.M. Natal Jorge
EDITORS

Simple and Fast Shape Based Image Retrieval

João Ferreira Nunes
Escola Superior de Tecnologia e Gestão, Instituto Politécnico de Viana do Castelo, Portugal
Departamento de Engenharia Informática, Faculdade de Engenharia Universidade do Porto, Portugal

Pedro Miguel Moreira
Escola Superior de Tecnologia e Gestão, Instituto Politécnico de Viana do Castelo, Portugal
Laboratório de Inteligência Artificial e Ciências de Computadores, Universidade do Porto, Portugal

João Manuel R. S. Tavares
Faculdade de Engenharia Universidade do Porto, Portugal

ABSTRACT: Content Based Image Retrieval (CBIR) is a challenging and active topic of research. This paper focuses on the shape of the represented objects as the main criterion in respect to evaluate the relevance of the retrieval results. There are several shape descriptors described in the literature, which are reported to achieve good results. However some of them are not obvious to implement or are computationally demanding. In this paper we propose a simple and fast to compute set of features to achieve shape based image retrieval. We conducted retrieval experiments on the MPEG-7 Core Experiment CE-Shape-1 test set and the results obtained demonstrate usefulness and competiveness against reported results from other more elaborated descriptors. Results demonstrate that our approach is also valuable when objects represented in the images share similar shapes, although being conceptually different. Another interesting result is that users tend to be very stringent when a good result set is presented (very similar shapes) whilst they are more permissive when the result set does not present a very high level of similarity between the shapes.

1 INTRODUCTION

The “democratization” of technology increased the access to digital devices and consequently amplified the amount of multimedia contents produced, raising significantly the size of the multimedia repositories. Simultaneously several applications in entertainment, art and commerce arise, needing to accurately access the unstructured data stored on those repositories. For example, for a distinctive trademark registry application (Eakins et al. 2007), one might need to ensure that a new registered trademark is sufficiently unique from the existing marks by searching the database.

To accomplish this need for managing and searching within multimedia collections we can find several tools. On one hand the tools based on textual information (keywords) describing the multimedia content. These tools require humans to label all the content stored. This task turns out to be very subjective since it depends on the interpretation of the person that catalogs and also impractical in circumstances where the amount of information is enormous, as in situations where the contents are generated automatically, like a surveillance cameras system. On the other hand the tools that comprise extracting the hidden useful knowledge embedded on the multimedia content, and then attempt to discover relationships between them, to classify them based on their content and to extract data patterns. Having regard only to images, this last approach is the basis of a Content Based Image Retrieval system (CBIR), which builds on the image analysis to extract information that is used to retrieve the images.

Most researches on CBIR have contributed to color/texture based indexing and retrieval. Comparatively, little work has been done on image retrieval using shape. In fact, among all the visual features, shape is the most valuable feature to identify or describe objects represented in images since it is easier for users to describe in the query, either by example or by sketch rather than sketch a colored or a textured image as query. In some circumstances shape contains more intrinsic information about the represented object than color, texture or any other feature.

This paper focuses on presenting a reduced set of image features simple and fast to extract and that can be used to describe 2D shapes in images. To validate our approach we conducted experiments on image retrieval. The reported experiments were conducted with the well-known MPEG-7 Core Experiment CE-Shape-1 test set and the results obtained demonstrate usefulness and competiveness against existing descriptors.

The paper is organized into six sections: after this introduction, on Section 2 we present related work in respect to shape descriptors and CBIR. On the third section we describe the proposed set of features as well as the experiments conducted to assess and validate our approach. Results and conclusions are presented in section four and five, respectively.

2 RELATED WORK

2.1 *Shape descriptors*

Image description consists in one of the key elements of multimedia information description. In the Multimedia Content Description Interface (MPEG-7) images are described by their contents featured by color, texture and shape. The shape descriptor aims to measure geometric attributes of an object to be used for classifying, matching, and recognizing objects. There are available several techniques for shape representation that are summarized in (Mehtre et al. 1997), such as Fourier descriptors (Zhang & Lu 2002; El-ghazal et al. 2009), Wavelet descriptors, grid-based, Delaunay triangulation (Shahabi & Safar 2006), among others. The study in (Mehtre et al. 1997) classifies the shape description techniques into boundary based and region based methods. Boundary based methods use only the contour of the objects' shape, while the region based methods use the internal details in addition to the contour.

2.2 *Image Retrieval*

Many works have been done in the field of image retrieval, known as Content Based Image Retrieval (CBIR), see e.g. (Lew et al. 2006; Wong et al. 2007). The key to a successful retrieval system is to choose the right features that represent the images as accurately and uniquely as possible. We can find different implementations of CBIR with various types of user queries: some fed with queries by example, where users draw a rough approximation of the image they are looking for (Chatzichristofis et al. 2010; Langreiter 2011), or they provide a preexisting image; in other implementations the query is made by direct specification of image features; and in others the query is done by image region (rather than the entire image); or by multiple example images; or even using multimodal queries.

3 METHOD

With this work we intended to define a set of image features that are simple and fast to extract involving very light math operations, and yet robust enough to distinguish 2D shapes in images. To validate our proposal set we conducted some experiments on image retrieval using the Part B of the MPEG-7 Core Experiment CE-Shape-1 dataset. To measure the performance of those experiments we used the metric known as Bulls Eye Percentage (BEP). As we also intended to evaluate how our proposed methods perform in respect to the users' subjective relevance of retrieved shapes, we have developed a web-based tool to support these experiments.

3.1 *The MPEG-7 dataset*

The MPEG-7 Core Experiment CE-Shape-1 was created by the Moving Picture Experts Group to evaluate the performance of 2D shape descriptors under the change of a viewpoint with respect to objects, non-rigid object motion (e.g., people walking or horse running) and noise resulted from digitization and/or segmentation (Latecki & Lakamper 2000).

The dataset consists of 1400 shapes grouped into 70 classes, each class containing 20 similar objects. Some of the shapes have experienced a number of transformations, such as scales, cuts and rotations and also some of them have holes. Finally, the image resolution is not constant among them. The next figure (Fig. 1) illustrates a representative shape image of each one of the 70 classes.

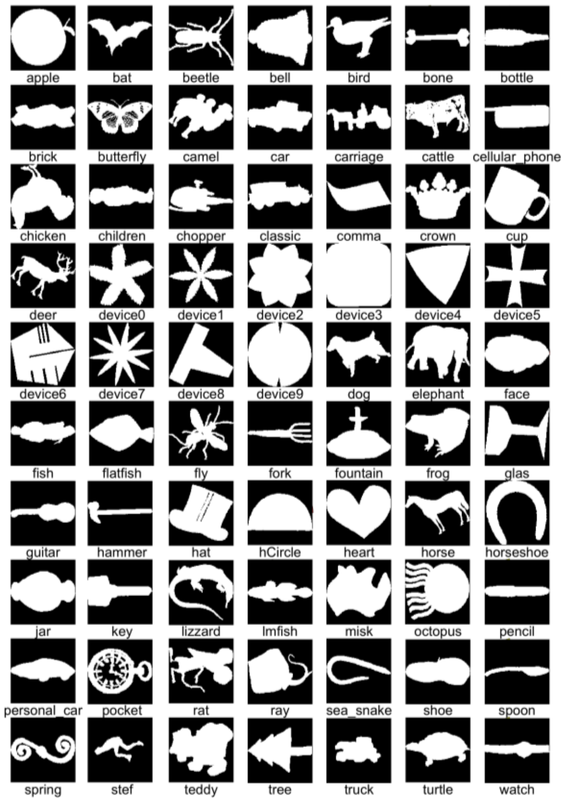

Figure 1. The MPEG-7 Core Experiment CE-Shape-1 dataset.

This dataset offers the possibility for experimental comparison of the existing approaches evaluated based on the retrieval rate. In (Veltkamp & Latecki 2006) we can find a wide-ranging comparison of shape descriptor methods that were tested against the MPEG-7 dataset, also used by us to compare our results.

3.2 *Image features*

Since the images in the MPEG-7 dataset were inconsistent in terms of resolution, scalability and rotation we had to apply some preprocessing operations in order to produce a feature set as generic as possible. Thus, using some Matlab functions we cropped the images by their bounding box. We also applied a morphological close filter smoothing boundaries, reducing small inward bumps and filling small holes caused by noise.

We also developed a procedure that computed several geometric features and to guarantee that all of them were weight balanced, we had to assure that they were normalized (between 0 and 1). After a correlation study between all the features, our initial extensive set of image features was reduced into the following:

f1: Solidity – it is the ratio between the image area (number of pixels in the foreground region) and its convex-hull area (number of pixels of the area of the smallest convex polygon that can contain the same region).

f2: Perimeter-Area Ratio – results from the ratio between the image perimeter and the image area.

f3: Eccentricity – specifies the eccentricity of the ellipse that encloses the shape having the same second-moments as the shape. It finds how much the conic section deviates from being circular. An ellipse whose eccentricity is zero is actually a circle, while an ellipse whose eccentricity is one it is a line segment.

f4: Extent – it is the result from the ratio between the number of pixels in the foreground region of the shape with the number of pixels in the bounding-box area.

f5: Area vs. Contour – this feature intends to distinguish shapes in respect to their ratio of the occupied area in respect to the contour, normalized by the area and perimeter of the minimal axis aligned bounding box (AABB) enclosing the shape image.

f6: Compactness – the compactness of a shape represents the degree to which a shape is compact. It is computed by the ratio of the shape's area to the area of a circle (the most compact shape) having the same perimeter.

f7: this feature intends to capture the relation between the elongatedness and the complexity of the contour in respect to the minimal bounds given by the enclosing circumference and bounding box.

The features f1, f2, f4, f5, f6 and f7 according to the taxonomy reported in (Zhang & Lu 2003) are classified as a region-based techniques while f3 is classified as a conventional contour-based technique.

In Figure 2 we can see the seven extracted features in the form of radar plots of each shape image that was illustrated in Figure 1.

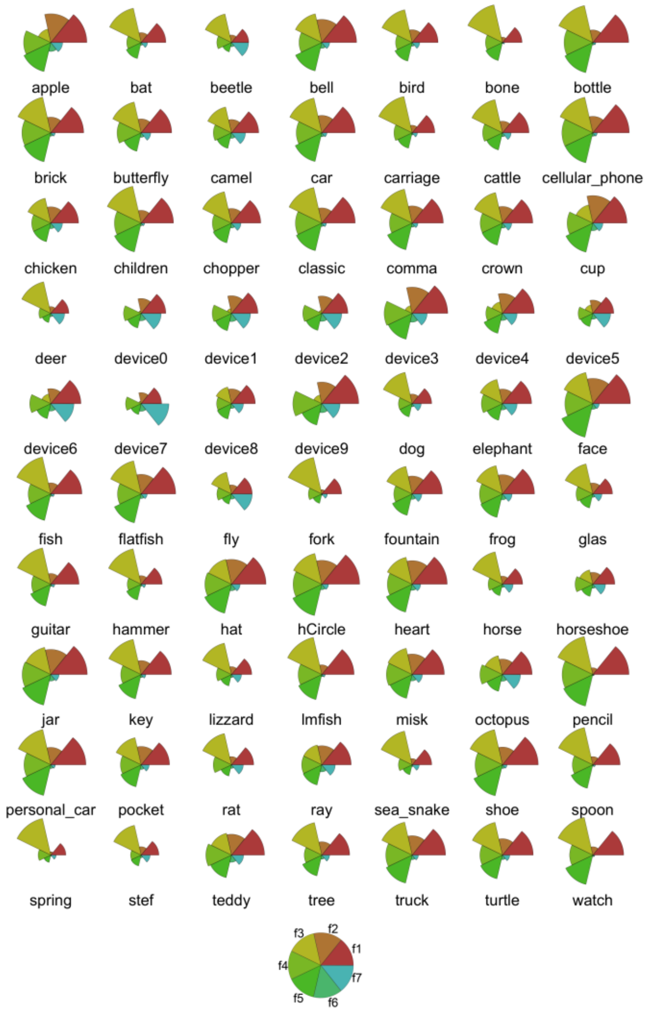

Figure 2. Radar plots of the average feature values.

3.3 *Retrieval experiments*

To measure the performance of our proposed features applied on a shape based image retrieval system we used the Bulls Eye test. This is an automatic and frequently used test in shape retrieval, which enables the comparison of our approach against other performing shape retrieval techniques.

This test is to present each of the images of the dataset as the query image and through a Euclidean-distance, based on the nearest neighbor approach, the top 40 matches are retrieved. The number of relevant images (images that are labeled within the same class as the query image) is summed and then divided by the highest number of relevant images. The resulting Bulls Eye Percentage (BEP) is then the total number of relevant retrieved images divided by 20 (number of instances per class) x 1400.

3.4 *User study*

In some related research projects that also work with the MPEG-7 dataset, like the one reported in (Chen & Huang 2008) the authors claim that a possible reason of their proposed scheme performs poorly in some shapes is because the categorization of those shapes is not consistent with the human perception. In fact, after a visual examination of the dataset we also noticed that some shapes categorized in different classes are graphically similar. Some of the disputed categories are the *guitar*, the *spoon* and the *key*, where their images have a great resemblance in shape as well as in their features' values, here represented in form of radar plots in Figure 3.

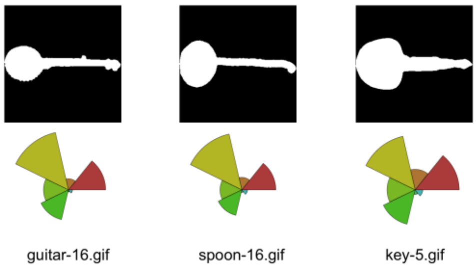

Figure 3. Sample images from the *guitar*, *spoon* and *key* classes (top) and the corresponding extracted features (botton).

This fact motivated us to conduct a study to confirm if there are images of the MPEG-7 dataset that are categorized in different classes, but nevertheless that they could be considered valid in a resulting set of an image retrieval experiment.

Thus we developed a web-based tool that uses a methodology quite similar to the one we used on the automatic retrieval experiments: for each image picked from the dataset and presented as the query image, the tool presents the top forty matches using the same metric – the Euclidean distance, based on the nearest neighbor approach. The main difference between both methodologies lies in selecting the relevant images. While in this method the selection is made by the users' judgment, in the automatic method the relevant images are automatically selected according to the a priori known classification of each represented object.

Once developed, our tool was made available at (Nunes et al. 2011), we invited via email a significant number of volunteers to perform the retrieval experiments. For each experiment the users were asked to distinguish the images that they considered to be shape similar to the reference image (the relevant images) from those that they considered to be shape dissimilar (Figure 4). Both resulting sets of every experiment containing the similar and dissimilar shapes were stored on a database for further analysis.

4 EVALUATION AND RESULTS

The achieved result of our automatic retrieval experiments was a BEP of 59%. We reinforce the fact that our shape description is short, and easy to extract and to implement.

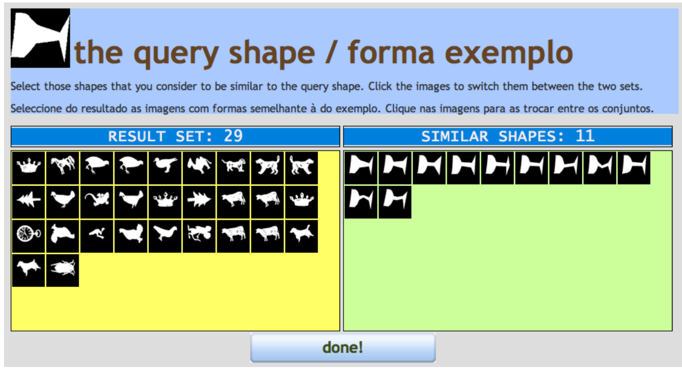

Figure 4. Sample screen interface of the developed web-based tool. Initially, all the images are placed at the left container. Users have to select relevant shapes (right container)by clicking over the images to switch their container.

A comprehensive comparison of shape descriptor methods is reported by Veltkamp & Latecki (2006) where distinct shape descriptors were compared, reimplemented and tested against the same dataset we used. This comparison is partially reproduced in Table 1.

Table 1. BEP performance (reported and re-implemented) for several shape similarity measures using the MPEG-7 Part B dataset (adapted from (Veltkamp & Latecki 2006))

Method	BEP reported	BEP reimp.
Shape context	76.51	
Image edge orientation histogram		41
Hausdorff region		56
Hausdorff contour		53
Grid Descriptor		61
Distance set correspondence	78.38	
Fourier descriptor		46
Delaunay triangulation angles		47
Deformation effort	78.18	
Curvature scale space	81.12	52
Convex parts correspondence	76.45	76
Contour-to-centroid triangulation	84.33	79
Contour edge orientation histogram		41
Chaincode nonlinear elastic matching		56
Angular radial transform		53
Our method		**59**

As the authors observe, there are some important differences between the reimplementation and the reported performances. This can be due to several issues such as: lack of information to devise a proper implementation, some methods are inherently complex and some fine tuning in respect to the datasets for which the performance values were reported. We notice that our proposed approach ranks in fourth place with respect to the re-implemented performances. Another possible reason for this result derives from the fact that we consider the dataset has some images that despite being associated with a particular class, they could also be classified within other classes and therefore being valid in a retrieval experience.

We used a total of 1535 individual experiments via our web-based tool. The overall result was similar to that obtained automatically of 58%. Although, it is clear that in some classes the retrieval rate increased significantly (e.g. class device2) leading to the conclusion that for these classes the resulting set of the top forty relevant images contains images from different classes but considered by users to be shape similar. Both retrieval results are illustrated in Figure 5. As it can be observed there are noticeable differences. The retrieval performances for the BEP range from 14% to 100% while for the user tests range from 18% to 97%. Since we used the same criteria in both methods to get the forty more relevant images, it would not be expected retrieval rates from the user tests above the retrieval rates achieved with the BEP, unless the resulted set of relevant images has images from disputed classes sharing similar shapes. This has turned out to be more clearly when a good result set is presented, as for instance, the *face* class (leftmost), for which users tend to be more rigorous (100% of retrieval rate for the BEP against 72% for the user tests). However, users are more permissive when the result set does not present a very high level of similarity between the shapes, like in the *tree* class (the rightmost) where the retrieval rate for the Bulls Eye test is 14% and the twice for the user tests (28%).

The confusion matrix in Figure 6 presents the detailed, per class, relevant results assigned by users. For each line of the grid, the dark circles, are correlated to the number of shapes from each class that were considered relevant (similar) in respect to the example image. A first conclusion that can be drawn is that the results, although not strictly symmetrical, present a high degree of symmetry. For instance, there are a comparable number of *guitar* shapes considered relevant when the retrieval is based on a *key*, in respect to the number of *key* shapes when the retrieval is based on a *guitar* shape. As it was expected, there are classes that are considered by users to exhibit shape similarity. As a consequence shapes belonging to objects from these classes are considered relevant by users when asked to separate the images from the retrieved result set. For instance, it can be observed that shapes from *deer* class are considered to be similar to those from the *horse* class. A high level of subjective similarity is also present amid the *device* classes of objects.

5 CONCLUSIONS AND FUTURE WORK

We have presented a simple method supported by few set of simple geometric image features to describe shapes that are simple and fast to compute.

We conducted our experiments on the MPEG-7 Core Experiment CE-Shape-1 and the achieved results demonstrate usefulness and competiveness against other reported approaches. We also developed a web-based tool to conduct user experiments in order to investigate subjective relevance of the retrieved shapes. Results indicate that users tend to be very stringent when a good result set is presented (very similar shapes) whilst they are more permissive when the result set does not present a very high level of similarity between the shapes. We intend to further investigate this behavior by conducting more experiments using, for example, a set of identical images with different rotations or scales.

There are several avenues for future work. A first one is to learn feature weights using, as for instance, evolutionary algorithms (e.g. genetic algorithms) to properly tune the used similarity distance metric. Another improvement is to make use of relevance feedback from the users.

REFERENCES

Chatzichristofis S. A. & Boutalis Y. & Papamarkos N. & Zagoris K. 2010. Accurate Image Retrieval Based On Compact Composite Descriptors And Relevance Feedback Information. *Int. J. Patt. Recog. Art. Intel. 24:207.*

Chen C. & Huang P. 2008. A New Method for Image Retrieval Based on Shape Decomposition *In 2008 Congress on Image and Signal Processing IEEE, p. 439-444.*

Eakins J. P. & Boardman J. M. & Shields K. 2007. Retrieval of Trade Mark Images By Shape Feature.

El-ghazal A, Basir O, Belkasim S. 2009. Farthest point distance: A new shape signature for Fourier descriptors. *Signal Processing: Image Communication 24:572-586.*

Langreiter. 2011. Available at: http://labs.systemone.at/retrievr [Accessed June 2011].

Latecki L. J. & Lakamper R. 2000. Shape Similarity Measure Based on Correspondence of Visual Parts. *IEEE Transactions On Pattern Analysis And Machine Intelligence 22:1-6.*

Lew M. S. & Sebe N. & Djeraba C. & Jain R. 2006. Content-based Multimedia Information Retrieval - State of the Art and Challenges. *ACM Transactions on Multimedia Computing, Communications, and Applications:1-26.*

Mehtre B. M. & Kankanhalli M. S. & Lee W. F. 1997. Shape Measures for Content Based Image Retrieval. *Information Processing & Management:1-19.*

Nunes J. F. & Moreira P. M. & Tavares J. M. R. S. 2011. Fast and Simple Shape Based Image Retrieval - Web Based Tool. Available at: http://shapes.estg.ipvc.pt.

Shahabi C. & Safar M. 2006. An experimental study of alternative shape-based image retrieval techniques. *Multimedia Tools Appl 32:29-48.*

Veltkamp R. C. & Latecki L. J. 2006. Properties and Performance of Shape Similarity Measures *In 10th IFCS Conf. Data Science and Classification Slovenia, p. 1-9.*

Wong W. & Shih F. Y. & Liu J. 2007. Shape-based image retrieval using support vector machines, Fourier descriptors and self-organizing maps. *Information Sciences 177:1878-1891.*

Zhang D. & Lu G. 2002. An Integrated Approach to Shape Based Image Retrieval *In ACCV2002: The 5th Asian Conference on Computer Vision, p. 1-6.*

Zhang D. & Lu G. 2003. Evaluation of MPEG-7 shape descriptors against other shape descriptors. *Multimedia Systems 9:15-30.*

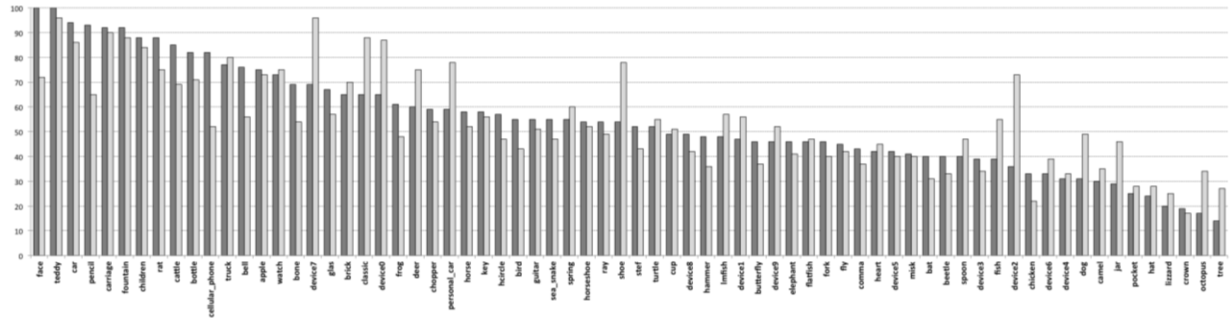

Figure 5. BEP per class for the automatic retrieval (dark grey) and for the retrieval rate for the user experiment (light gray). The classes are ranked left to right in decreasing order in respect to the achieved BEP for the automatic retrieval.

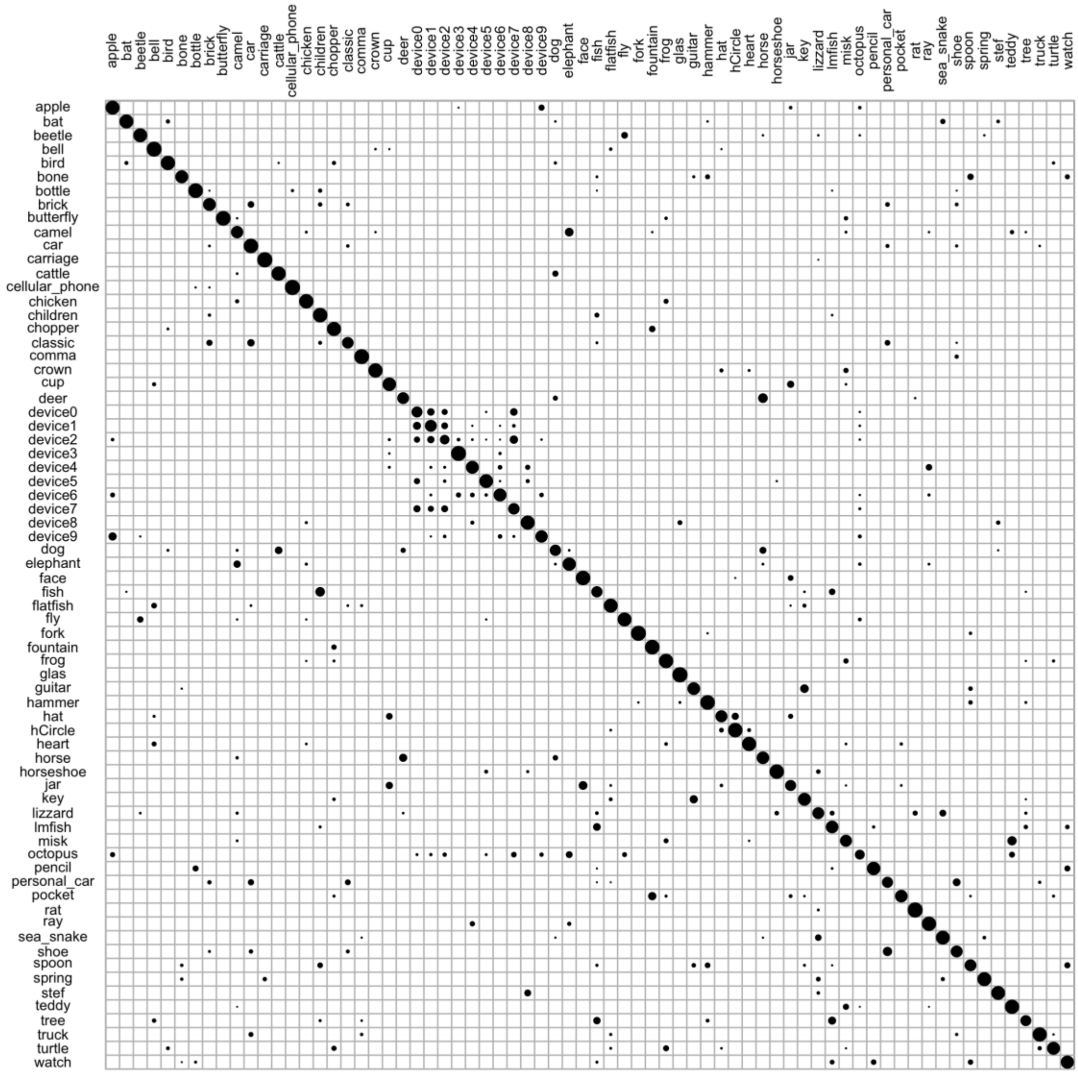

Figure 6. Detailed relevance results assigned by users. For each line of the grid, the dark circles, are correlated to the number of shapes from each class that were considered relevant (shape similar) in respect to the example image (class).