Shape Based Image Retrieval and Classification

João Ferreira Nunes
Escola Superior de Tecnologia e Gestão
Instituto Politécnico de Viana do Castelo
Viana do Castelo, Portugal
joao.nunes@estg.ipvc.pt

Pedro Miguel Moreira
Escola Superior de Tecnologia e Gestão
Instituto Politécnico de Viana do Castelo
Viana do Castelo, Portugal
pmoreira@estg.ipvc.pt

João Manuel R. S. Tavares
Faculdade de Engenharia
Universidade do Porto
Porto, Portugal
tavares@fe.up.pt

***Abstract* — Content based retrieval and recognition of objects represented in images is a challenging problem making it an active research topic. Shape analysis is one of the main approaches to the problem. In this paper we propose the use of a reduced set of features to describe 2D shapes in images. The design of the proposed technique aims to result in a short and simple to extract shape description. We conducted several experiments for both retrieval and recognition tasks and the results obtained demonstrate usefulness and competiveness against existing descriptors. For the retrieval experiment the achieved bull's eye performance is about 60%. Recognition was tested with three different classifiers: decision trees (DT), k-nearest neighbor (kNN) and support vector machines (SVM). Estimated mean accuracies range from 69% to 86% (using 10-fold cross validation). The SVM classifier presents the best performance, followed by the simple kNN classifier.**

Keywords - content based image retrieval, image classification, shape descriptors, data mining, machine learning.

I. INTRODUCTION

With the fast growth of multimedia data, mainly due to the wide spread of digital devices, multimedia repositories became very common, and some of them extremely large. It became natural that with this amount of archived information it would come out the need of indexing and retrieving this unstructured data. For that, we can find several tools for managing and searching within these collections that are based on textual information about the images, requiring humans to describe every image, resulting not viable for large repositories. Another approach involves extracting the hidden useful knowledge embedded on the images, as for instance, tools to discover relationships between them, to classify images based on their content and tools for extracting data patterns.

Content Based Image Retrieval (CBIR) [1] builds on the image analysis to extract information that is used to retrieve the images that best match the query (which can be itself an image or set of images) using some sort of similarity distance. Image analysis can use several distinct features such as color, texture, shape or any other information that can be derived from the images.

Shape of the objects represented in images is one of the most significant properties used in CBIR and in recognition tasks. This is particularly due to the fact that shape is perceptually very relevant in order to recognize objects. In some circumstances shape contains more intrinsic information about the represented object than color, texture or other features. From a geometric point of view, shape can be informally defined as the result of removing color, texture, and effects due to affine transformations such as scale, translation and rotation from a representation of an object in an image [2].

This paper proposes a reduced set of features which intend to describe 2D shapes in images. To validate our approach we conducted experiments on image retrieval and image recognition. 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. For the retrieval experiment the achieved bull's eye performance is about 60%. For recognition tasks, three distinct classifiers were tested with an estimated mean accuracy ranging from 69% to 86% (using 10-fold cross validation).

The paper is organized into six sections: the first one is the Introduction, where we present our motivation and objectives. The state of existing knowledge in the area of Image Retrieval is written in the second part, followed by section Image Features where we present and discuss the features computed in order to describe shapes. The sections four and five, Retrieval Experiments and Classification Experiments, explain the procedures of the experiments that were followed, and then, finally, conclusions and future work are discussed in the last section.

II. RELATED WORK

A. 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 [3], such as Fourier descriptors [4][5], Wavelet descriptors, grid-based, Delaunay triangulation [6], among others. The study in [3] 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.

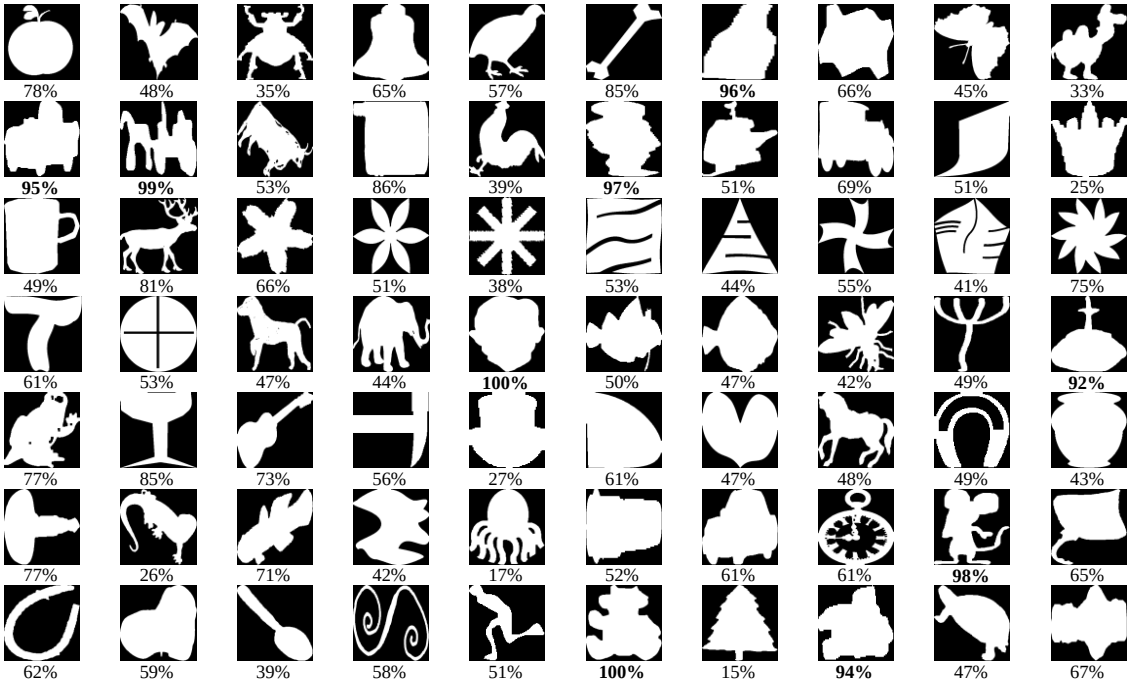

Figure 1. Sample images from the70 classes in the dataset. Values below the images are retrieval BEP performance obtained per class (cf. Section IV).

B. Image Retrieval

Many works have been done in the field of image retrieval, known as Content Based Image Retrieval (CBIR), see e.g. [7] [8]. 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 [9][10] or they simply provide a preexisting image [11]; 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.

C. MPEG-7 Dataset

The dataset used in our experiments is the "MPEG-7 Core Experiment CE-Shape-1 Test Set". It was created by the MPEG-7 committee. The Motion Picture Expert Group (MPEG) [12] is a working group of ISO/IEC and has defined the standard for description and search of audio and visual content. This image collection includes 1400 binary images grouped into 70 categories by their content. Each category contains 20 samples.

Since these images represent 2D objects that are projections of 3D objects, their silhouettes may change due to: (i) the change of a view point with respect to objects; (ii) non-rigid object motion (e.g., people walking or horse running); and (iii) noise (e.g., digitization and segmentation noise). Also, few additional characteristics of the dataset to be mentioned: some images have holes in them, while others do not, and some images have experienced a number of transformations, such as scales, cuts and rotations and, finally, the image resolution is not constant among them. Fig. 1 illustrates a sample image of each one of the 70 classes in the dataset.

In this dataset there are some categories, like classic-car, guitar or spoon, which include images corresponding to the same concept, but showing visible different shapes. As an example, Fig. 2 shows the 20 guitar samples of the guitar class.

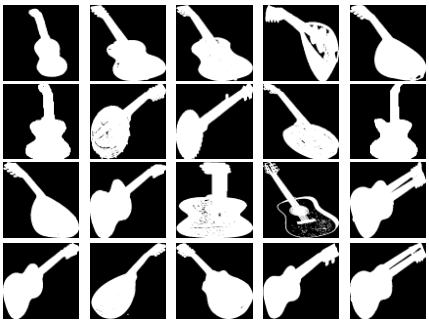

Figure 2. The 20 images for the guitar class.

The dataset is accessible from various sources in the World Wide Web [13] [14], since it has been used on other researches, and as a result some of their authors also make it available.

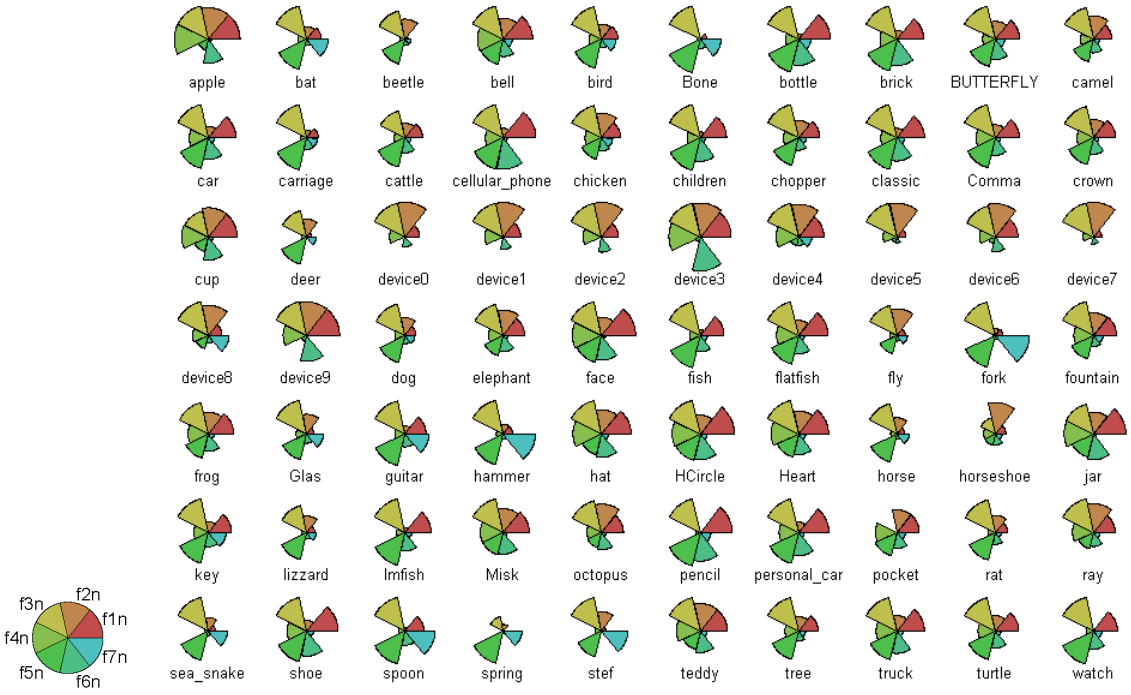

Figure 3. Average values of the proposed features (normalized between 0 and 1) for each of the 70 classes of the used dataset.

III. IMAGE FEATURES

According to [15] preprocessing is always a necessity whenever the data to be mined is noisy, inconsistent or incomplete and it significantly improves the effectiveness of the data mining techniques. Therefore, this section introduces the preprocessing techniques that we have applied to the images before the feature extraction process. We intended to reduce their noise by removing the irrelevant information. This was accomplished by detecting and extracting the images' region of interest, cropping them through their bounding box. Another preprocessing technique that we have also applied is the morphological close filter. This filter closes morphologically the image and it is defined as the dilation of the image followed by the erosion of the dilated image. The closing filter operation smoothes boundaries, reduces small inward bumps, joins narrow breaks and fills small holes caused by noise.

The next step was to compute the image features we intended to extract. The features chosen had to be discriminative and sufficient in describing the object presented in each image. Therefore we developed another Matlab procedure using the *regionprops* function from the Image Processing toolbox. This function measures a set of properties of the image, like its area, Euler number, bounding box, perimeter, centroid, etc. At this stage we had to guarantee that all the features computed were normalized, so that they would be weight balanced. The resulting vector of features was stored on an n x f matrix, where n corresponds to the number of images in the dataset and f is the number of features. The features that were computed are:

F1: Solidity. This feature results from the ratio between the image Area and its ConvexHullArea, where Area is the number of pixels in the foreground region and the ConvexHullArea is the number of pixels of the area of the smallest convex polygon that can contain the same region.

F2: Axis Ratio. It's the ratio between the MinorAxisLength and the MajorAxisLength. The MinorAxisLength gives the length (in pixels) of the minor axis of the ellipse that has the same normalized second central moments as the region, while the MajorAxisLength gives the length (also in pixels) of the major axis of the ellipse that has the same normalized second central moments as the region.

F3: Areas Ratio; It's the ratio between the image Area and the image FilledArea. The attribute Area gives the number of pixels in its foreground region while the FilledArea gives the number of on pixels in FilledImage. This ratio feature gives a notion if the image has holes on it or not, where values close to one indicate that the image has very few holes.

F4: Perimeter-Area Ratio; It's the ratio between the image Perimeter and the image Area.

F5: Eccentricity; Specifies the eccentricity of the ellipse that has the same second-moments as the region. The eccentricity is the ratio of the distance between the foci of the ellipse and its major axis length. An ellipse whose eccentricity is zero is actually a circle, while an ellipse whose eccentricity is one is a line segment.

F6: Extent; Specifies the ratio of pixels in the foreground region with the pixels in the total bounding box. It is computed as the Area divided by the Bounding Box area. To reduce the Bounding Box area, the image is previously rotated by its Orientation attribute.

F7: *Invariant moment*; This is a useful measure to describe objects because image properties that are found via image

moments are invariant under translation, changes in scale, and also rotation. From the Hu [16] set of invariant moments, we've chosen the *skew invariant* and to compute this feature we used the *momentsupto3* function from the LISQ toolbox.

On Fig. 3 we can see the seven extracted features for the 70 sample images formerly presented in Fig. 1, presented in a radar plots form.

Analyzing the radar plots in Fig. 4, each one representing the seven features of the 20 sample images of the guitar class, it is noticeable that this category contains images that are very similar between them, and consequently their radar's shape are identical, but it also contains images whose shapes are quite different, although being labeled into the same category.

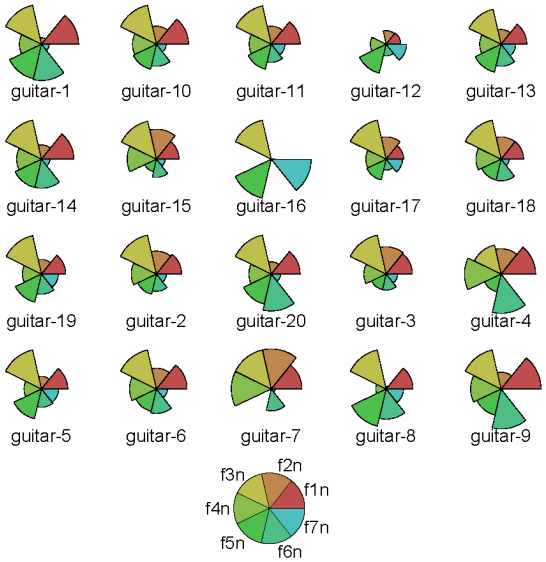

Figure 4. Feature values (normalized between 0 and 1) for the each of the 20 images of the guitar class.

IV. RETRIEVAL EXPERIMENTS

In order to evaluate the suitability and usefulness of the proposed set of features to describe shape of the represented images in a retrieval context we conducted the following experiment. A query was made for each individual image returning a ranked list based on the Euclidean distance evaluated between the query and the archived images.

To measure the performance we used the metric known as Bulls Eye Percentage (BEP). This takes into account a number of results equal to twice the number of relevant results (images labeled as the same class as the query).

The number of relevant images is summed and divided by the maximum number of relevant images. For the given dataset, and as the classes are equally distributed in number of instances (20 per class), for each query the first 40 results were taken into account. The resulting BEP is then the total number of relevant retrieved images divided by 20 x 1400.

The achieved result was a BEP of 59%. Although not the best result, when compared to those reported in the literature, we reinforce the fact that our shape description is short (7 numeric values) and easily to extract and implement.

A comprehensive comparison of shape descriptor methods is reported by Veltkamp and Latecki [17] where distinct shape descriptors were compared, re-implemented and tested against the same data set we used. This comparison is partially reproduced in Table 1. As the authors notice, 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.

TABLE I. BEP PERFORMANCE (REPORTED AND RE-IMPLEMENTED) FOR SEVERAL SHAPE SIMILIRATY MEASURES USING THE MPEG 7 PART B DATASET (SHOWN RESULTS ADAPTED FROM [17])

method	BEP reported	BEP reimpl
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**	**59**

To better understand our result we present in Fig. 5 the mean BEP performance achieved for each class of shapes. As it can be observed there are relevant differences. The BEP performances range from 15% to 100%. This result indicates that our method is not able to distinguish between some shapes. Results per class are also presented below each sample image in Fig. 1.

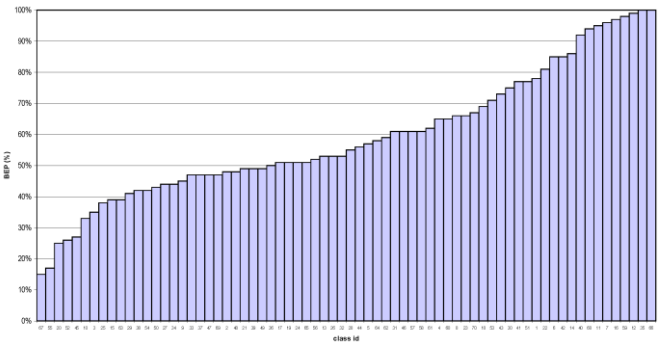

Figure 5. BEP performances per class.

A visual inspection of the retrieved results suggests that although not able to distinguish between the ground truth class labels the retrieval process results in a subset of visually similar shapes, even for those queries (images) with lower BEP performance.

An illustration of the behavior of the retrieval is illustrated in Fig. 6, where three sample results are presented. We choose to illustrate this behavior for queries with images from three classes with very dissimilar performances. For each group of images in Fig. 6 the query is an image similar to the top left image (as the first result is always the corresponding archived image). The first ten retrieved images are shown. The first group a) illustrates a query from an image from one of the best performing images. For this all the first ten results are images from the same class. The middle subset illustrates a query from a class with a BEP similar to the overall BEP (59%). As it can be observed four not relevant images are retrieved in the first ten. Although, the results visual inspection revealed a clear level of shape similarity between them. For the third and last row, from the *TREE* class, the results are poor, albeit visual similarity remains amongst the retrieved subset.

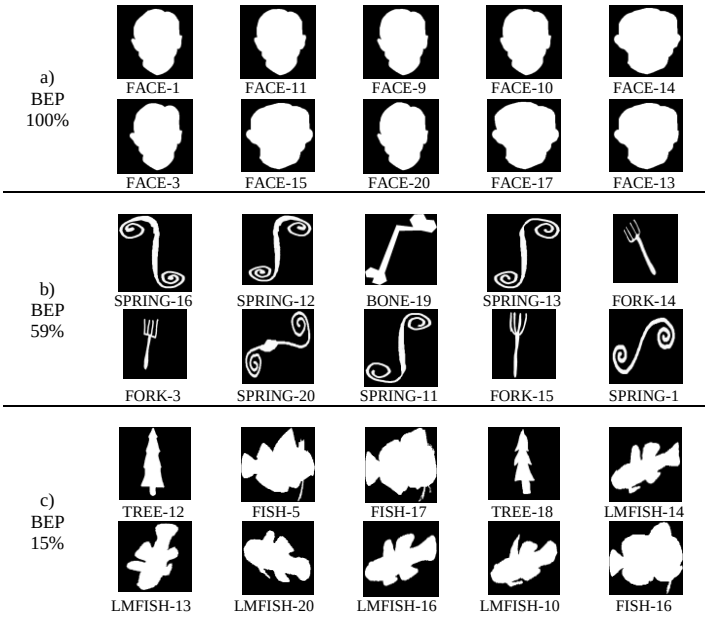

Figure 6. Retrieval examples for classes with different BEP performance (maximal, average and minimal from top two bottom).

V. CLASSIFICATION EXPERIMENTS

As noticed, shape is also a major feature in order to classify (recognize) objects represented in images. In order to test the ability of our proposed shape description in recognition tasks we tested with three different supervised learning algorithms in order to devise classification models (overall results shown in Table III)..

The first experiment was made using a decision tree learner. Decision trees result in a hierarchical classifier from which classification rules can be derived by traversing the tree from the root to each node. The classifier is typically very human legible. Although the induction process, at each node, inspects the corresponding training subset in order find a feature and a partition on that feature that best discriminates between the classes. Thus, partitions are done based on hyper-planes with a fixed value in respect to one of the features (partitions are not based on oblique hyper planes). This can limit the performance of the decision trees as the division between classes is likely to do not depend in one of the features.

A second experiment was made using the simple, but sometimes very effective, k-nearest neighbor algorithm. In this instance based learning scheme, a test case is classified based on the observation of the k nearest training neighbors, given a distance metric. The resulting classification can assume different blends, but a usual rule is to classify based on the prevalent class amongst the observed neighbors.

An important decision is to decide on the number k (number of neighbors). We used a leave-one-out (using each image as the test set and all the remaining as the training set) validation schema to learn the best k based on the smaller error rate. An experience was conducted with k ranging from 1 to 11. Results are shown in Table II and reveal that the best k is equal to one, i.e. a shape is classified with the same class as its nearest neighbor in the feature space.

TABLE II. ESTIMATED ERROR OF KNN USING LEAVE-ONE-OUT VALIDATION TO LEARN THE BEST VALUE OF K

k	1	2	3	4	5	6	7	8	9	10	11
Err	0.20	0.27	0.26	0.29	0.29	0.32	0.33	0.34	0.35	0.36	0.36

The third experiment was conducted using support vector machines, which are known to be able to outperform other classification algorithms in many situations.

TABLE III. ESTIMATED ACCURACCY USING 10 FOLD CROSS VALIDATION (REPEATED 10 TIMES)

classifier	min	max	mean	sd
DT	0.64	0.73	0.69	0.03
kNN (k=1)	0.71	0.84	0.80	0.04
kNN (k=3)	0.66	0.79	0.73	0.04
KNN(k=5)	0.64	0.75	0.70	0.04
SVM	0.81	0.90	0.86	0.03

The results achieved with classifiers learned from the three different algorithms are presented in Table III. As it can be observed, the SVM based classifier has an estimated mean accuracy of 86%. The simple kNN classifier has an estimated accuracy of 80%. Decision trees had the poor results. Notice that all classifiers have some stability as demonstrated by the standard deviation (sd) obtained.

VI. CONCLUSIONS AND FUTURE WORK

Shape is one of the most valuable features to identify or describe objects represented in images. We have presented a simple method based on a few set of image features to describe

shapes. Our method aims to be simple and to result in a short description.

We conducted experiments both for retrieval and classification of objects represented in images. The retrieval results demonstrated to be competitive with other more elaborated approaches.

For the recognition task obtained estimated accuracy rate of 86% using a SVM classifier. A simpler to implement kNN classifier is able to get 80% accuracy rate using our set of features.

Several improvements are intended to be carried as 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. This process is expected to increase the accuracy of the classifier for a given dataset. These results can be also valuable for retrieval purposes if these weights demonstrate stability among several datasets.

Another improvement to the retrieval process is to make use of relevance feedback, where the user progressively refines the search results by marking images in the results as "relevant", "not relevant", or "neutral" to the search query, then repeating the search with the new information.

As a major conclusion we stand that our method demonstrated usefulness and effectiveness for both retrieval and recognition purpose, particularly if taken into account its simplicity.

REFERENCES

[1] K. Hirata, K., Kato, T.: Query by visual example-content-based image retrieval. in: Third International Conference on Extending Database Technology. (1992), pp. 56–71.

[2] David G. Kendall, D.G.: A Survey of the Statistical Theory of Shape. Statist. Sci. Volume 4, Number 2 (1989), 87-99.

[3] Mehtre, B., K. M. L. W.: Shape measures for content based image retrieval: a comparison. *Information Processing and Management: an International Journal*, (1997), *33*

[4] Zhang, D., L. S.: Content-Based Shape Retrieval Using Different Shape Descriptors: A Comparative Study. *IEEE International Conference on Multimedia and Expo (ICME2001), (*2001)

[5] El-ghazal, A.; Basir, O. & Belkasim, S.: Farthest point distance: A new shape signature for Fourier descriptors. *Image Commun., Elsevier Science Inc., (*2009), *24*, 572-586

[6] Shahabi, C. & Safar, M.: An experimental study of alternative shape-based image retrieval techniques. *Multimedia Tools Appl., Kluwer Academic Publishers, (*2007), *32*, 29-48

[7] Wong, W.-T.; Shih, F. Y. & Liu, J.: Shape-based image retrieval using support vector machines, Fourier descriptors and self-organizing maps. *Inf. Sci., Elsevier Science Inc., (*2007), *177*, 1878-1891

[8] Lew, M. S.; Sebe, N.; Djeraba, C. & Jain, R.: Content-based multimedia information retrieval: State of the art and challenges. *ACM Trans. Multimedia Comput. Commun. Appl., ACM, (*2006), *2*, 1-19

[9] Retrievr, http://labs.systemone.at/retrievr, 2010.02.11

[10] Anaktisi, http://orpheus.ee.duth.gr/anaktisi, 2010.02.10

[11] FIRE - Flexible Image Retrieval Engine, http://www-i6.informatik.rwth-aachen.de/~deselaers/cgi_bin/fire.cgi, 2010.02.10

[12] The Moving Picture Experts Group (MPEG), http://www.chiariglione.org/mpeg, 2009.12.01

[13] MPEG-7 Core Experiment CE-Shape-1 Test Set, http://www.imageprocessingplace.com,2009.12.01

[14] MPEG-7 Core Experiment CE-Shape-1 Test Set, http://www.ehu.es/ccwintco/uploads/d/de/MPEG7_CE-Shape-1_Part_B.zip,2009.12.01

[15] Jiawei, H.J., Kamber, M., Pei, J., Data Mining: Concepts and Techniques: Morgan Kaufmann Publishers (2006)

[16] Hu, M.K.: Visual Pattern Recognition by Moment Invariants. IEEE Transactions on Information Theory, IT-8, 179-187 (1962)

[17] Veltkamp, R. C., R. C., Latecki, L. J.: Properties and Performance of Shape Similarity Measures. 10th IFCS Conf. Data Science and Classification. Slovenia, July, (2006).