Hand-draw sketching for image retrieval through fuzzy clustering techniques

Ugo Erra[1] and Sabrina Senatore[2]

[1] Università della Basilicata, Dipartimento di Matematica e Informatica, Viale Dell'Ateneo 10, Macchia Romana - 85100, Potenza, Italy
ugo.erra@unibas.it
[2] Università di Salerno, Dipartimento di Informatica Via Ponte Don Melillo - 84084 Fisciano(SA), Italy
ssenatore@unisa.it

Abstract. Nowadays, the growing of digital media such as images represents an important issue for multimedia mining applications. Since the traditional information retrieval techniques developed for textual documents do not support adequately these media, new approaches for indexing and retrieval of images are needed. In this paper, we propose an approach for retrieving image by hand-drawn object sketch. For this purpose, we address the classification of images based on shape recognition. The classification is based on the combined use of geometrical and moments features extracted by a given collection of images and achieves shape-based classification through fuzzy clustering techniques. Then, the retrieval is obtained using a hand-draw shape that becomes a query to submit to the system and get ranked similar images.

1 Introduction

The idiom 'a picture is worth a thousand words' refers to the concept that a single image can be used to quickly describe an idea but it also suggests us that we can not describe succinctly an image based on few words. Humans tend to describe an image using short sentences with keywords and pointing out different parts of the image, according to their cultural and professional background. On the other hand, finding an image that is close to a hand-draw sketch is a natural approach that fits well in the age of digital information where the growing amount of large-scale image repositories in many application domains emphasize the need for effective and practical means for retrieving digital images.

In general, two different approaches have been applied for image retrieval. The first one consists of attaching textual metadata to each image, then a keyword-based query is submitted to the database in order to retrieve relevant images [9]. This approach requires an initial time-consuming activity of images annotation activity; moreover, it is a laborious, human driven process that can affect the performance of the keyword-based image search, according to the naming and terminology used for annotate the images. The second approach, named Content-Based Image Retrieval (CBIR) exploits the features which characterize *objective* image properties such as color, texture, and shape. These techniques improve the effectiveness of image retrieval through multi-features combination [5] and then, by measuring similarity to a required query image [11]. But, the user does not always have a such image. An alternative to an image as a query is using a line-based hand-drawing, i.e., a sketch as a natural way to make a query [10]. Comparing a rough sketch to an image is a natural yet difficult task. Few approaches deal with sketch-based image retrieval (SBIR).

In [3] query-by-sketch exploits the spatial relationships between shapes in an image. The approach represents shapes, their spatial arrangement, color and texture attributes in the user sketch; then the image content is extracted starting from basic features and combining them in a higher-level description of spatial layout of components. Also interesting is the work in [7], which exploits wavelet-based indexing and query by sketch for color images retrieval. In some approaches, the use of fuzzy techniques can support the image description. In general, fuzzy retrieval models offer more flexibility in the representation of the terms' index, preferences among terms in a query which involve concepts and linguistic expression, through fuzzy values rather than crisp features values [8].

In this paper, we specifically discuss our approach to image retrieval based on the hand-draw sketch that represents a rough approximation of query image contour. The approach exploits fuzzy clustering techniques for image classification. The purpose of using the fuzzy clustering technique is twofold. First, the fuzzy clustering technique can reflect better the imprecise nature of images yielding an adequate classification of images collection that enable the search on a reduced search space. Second, the retrieval based on a query-by-sketch provides accurate results, evidencing the efficacy of such approach.

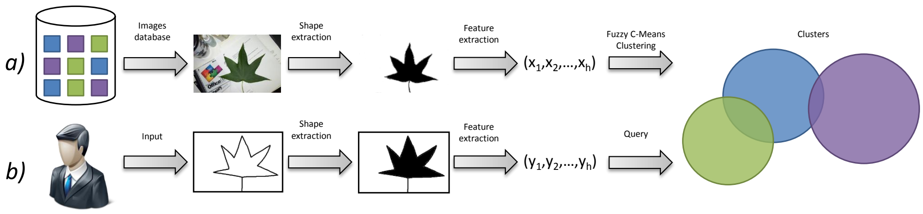

Fig. 1. Overview of the proposed approach.

2 Overview

Our approach accomplishes two stages. The first one performs an off-line images classification, through a fuzzy clustering technique (Fig. 1a), the second one instead, performs the query by hand-draw image sketch (Fig. 1b). Preliminary activity is shape and features extraction. The shape extraction separates the object (or region) of interest from other non-important image structures. There are several approaches for the extraction of the shape from a given image based on clustering methods, histogram methods, edge detection, level set methods, graph partitioning methods and so on. There is no a general methodological approach to reach this target: many factors and parameters (i.e., the background color, the contour, etc.) can affect the result. In our implementation, we adopt the k-means clustering algorithm for image segmentation which is suitable when the foreground and background colors contrast sufficiently with each other. Afterwards, a features extraction step gets the set of features from the region of interest that characterizes the image. Efficient shape features must have some essential properties such as identifiability, invariance, noise resistance, statistically independence and so on. We use several geometric description and image moments which are invariant to translation, rotation and scaling (Sect. 3).

The clustering algorithm achieves a partitioning of given images into clusters. In general a partition holds two properties: homogeneity within the clusters (data in a cluster must be similar) and homogeneity between clusters (isolation of a cluster from one another: data of different clusters have to be as different as possible). The images are opportunely translated into a matrix, where each row is a characteristic vector which represents an image described by the extracted features. In this study, we exploit the well-known *fuzzy C-Means* (FCM) algorithm [2], which takes as an input a collection of patterns (in our case, the collection of images) and produces fuzzy partitions of the given patterns (i.e. images) into (prefixed) c clusters (Sect 4). The computed clusters are analyzed in order to discover the nature of the data groups (similar images) and associate to each cluster the expected class name. A this point, the second stage can be applied. Query by image is a technique that generally provides the CBIR system with an example image that will base its search upon. The underlying search approach may vary depending on the application, but resulting images should be similar to the given sample query, sharing its common descriptors.

3 Shape descriptors

Three types of shape descriptions are adopted: geometric description, invariant moments and affine moments. The geometric features discriminate shapes with large difference. They are useful to eliminate false hits and usually are not suitable as single description, in fact they are combined with other shape descriptors to better discriminate shapes. The moment instead, represents a mathematical concept coming from the concept of moment in physics. It is used in computer vision for both contour and region of a shape. The invariant moments [6] are one of the most popular and widely used contour-based shape descriptors. Affine moments invariants are instead, features computed from moments that do not change their value in affine transformation. In the case of geometric features, let P and A denote the shape perimeter and area, respectively. Note that perimeter and area are invariants respect to translation and rotation but when combined, they are not invariant with respect to scale. The features we adopt are:

- Eccentricity E is the measure of aspect ratio. It is defined as the ratio $E = W_{bb}/H_{bb}$ where W_{bb} and H_{bb} are, respectively, the width and height of minimal bounding rectangle of the shape.

- Rectangularity R represents how rectangular a shape is, i.e. how much it fills its minimum bounding box. It is defines as $R = A/A_{bb}$ where A_{bb} is the area of the minimum bounding rectangle.
- Compactness C is a measure that combines area with perimeter. It is defined as $C = L^2/4\pi A$.
- The value π_{gen} is a measure of the compactness of a shape respect to a circle. It is defined as $\pi_{gen} = P/W_{bb}$.

Invariant moments m_{pq} are the simplest and is given as:

$$m_{pq} = \sum_x \sum_y x^p y^q f(x,y) \qquad p,q = 0,1,2,\ldots$$

where $f(x,y)$ is the intensity function at position (x,y) in a 2D gray level image. To obtain translation invariance, the central moments μ_{pq} should be applied:

$$\mu_{pq} = \sum_x \sum_y (x-\overline{x})^p (y-\overline{y})^q f(x,y) \qquad p,q = 0,1,2,\ldots$$

where $\overline{x} = m_{10}/m_{00}$ and $\overline{y} = m_{01}/m_{00}$. Given central moments we are able to compute a set of 7 invariant moments [6], given by:

$$\begin{aligned}
I_1 &= \eta_{20} + \eta_{02} \\
I_2 &= (\eta_{20} - \eta_{02})^2 + 4\eta_{11}^2 \\
I_3 &= (\eta_{30} - 3\eta_{12})^2 + (3\eta_{21} - \eta_{03})^2 \\
I_4 &= (\eta_{30} + \eta_{12})^2 + (\eta_{21} + \eta_{03})^2 \\
I_5 &= (\eta_{30} - 3\eta_{12})(\eta_{30} + \eta_{12})[(\eta_{30} - \eta_{12})^2 - 3(\eta_{21} - \eta_{03})^2] + \\
&\quad (3\eta_{21} - \eta_{03})(\eta_{21} + \eta_{03})[(3\eta_{30} + \eta_{12})^2 - (\eta_{21} - \eta_{03})^2] \\
I_6 &= (\eta_{20} - \eta_{02})[(\eta_{30} + \eta_{12})^2 - (\eta_{21} - \eta_{03})^2] + \\
&\quad 4\eta_{11}^2(\eta_{30} + \eta_{12})(\eta_{21} + \eta_{03}) \\
I_7 &= (3\eta_{21} - \eta_{03})(\eta_{30} + \eta_{12})[(\eta_{30} - \eta_{12})^2 - 3(\eta_{21} - \eta_{03})^2] + \\
&\quad (3\eta_{12} - \eta_{03})(\eta_{21} + \eta_{03})[(3\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2]
\end{aligned}$$

where $\eta_{pq} = \mu_{pq}^{\gamma}$ and $\gamma = 1 + (p+q)/2$ for $p+q = 2,3,\ldots$. These moments are simple to calculate and they are invariant to translation, rotation and scaling. From central moments with a little computational effort we are able to obtain also an affine transform invariance which includes the similarity transform and in addition to that stretching and second rotation. We adopt affine moments as defined in [4] and given as:

$$\begin{aligned}
AMI_1 &= (\mu_{20}\mu_{02} - \mu_{11}^2)/\mu_{00}^4 \\
AMI_2 &= (\mu_{30}^2\mu_{03}^2 - 6\mu_{30}\mu_{21}\mu_{12}\mu_{03} + 4\mu_{30} + \mu_{12}^3 + \\
&\quad 4\mu_{03}\mu_{21}^3 - 3\mu_{21}^3\mu_{12}^3)/\mu_{00}^{10} \\
AMI_3 &= (\mu_{20}(\mu_{21}\mu_{03} - \mu_{12}^2) - \mu_{11}(\mu_{30}\mu_{03} - \mu_{21}\mu_{12}) + \\
&\quad \mu_{02}(\mu_{30}\mu_{12} - \mu_{21}^2))/\mu_{00}^7
\end{aligned}$$

All these features are sufficient to characterize the shape of an image. The rationale behind the choice of these moments is that we are interesting in translation, rotation, scale, and projective transform invariance in order that the location, orientation, and scaling of the shape do not affect the extracted features.

4 Fuzzy clustering

FCM represents the most common fuzzy clustering, particularly useful for flexible data organization. It takes as input a collection of patterns of a universe U in form of matrix and produces fuzzy partitions of the given patterns into (prefixed) c clusters. In fact, after the fuzzy clustering execution, each pattern has associated a c-dimensional vector, where each cell represents the membership (in the range $[0,1]$) of that pattern to each cluster. Compared to the crisp version, the fuzzy clustering generates a flexible partitioning, more intuitive to interpret: a pattern can have some characteristics that are natively representative of more than one cluster. In the fuzzy approach, the membership values better reveal the nature of data set and allow a clearer data analysis. Anyway, it is conceivable to assign a pattern to the cluster, whose membership is the highest. More formally, each row of the matrix is a vector that represents an image $I \longleftrightarrow \underline{x} = (x_1, x_2, \ldots, x_h)$, where each component of vector is a value computed for a feature. The FCM algorithm aims at minimizing the objective function constituted by the weighted sum of the distances $dist_{i,k}$ between data points $\underline{x}_k = (x_{k,1}, x_{k,2}, \ldots, x_{k,h})$ and the centers (or prototypes) $\underline{v}_i = (v_{i,1}, v_{i,2}, \ldots, v_{i,h})$, according to this formula:

$$Q(U,c) \ = \sum_{i=1}^{c}\sum_{k=1}^{n} u_{i,k}^{m}(dist(\underline{x}_k,\ \underline{v}_i))^2$$

where $c \geq 2$ is the number of clusters, $u_{i,k} \in$ [0,1] is the membership degree of $\underline{x}_k$ (k=1, …, n) in the i-th cluster A_i (i=1, …, c), and $m > 1$ is the fuzzifier, which controls the quantity of fuzziness in the classification process (common choice of fuzzifier is $m = 2$) and finally $dist(\underline{x}_k,\ \underline{v}_i)$ represents the euclidean distance between the data $\underline{x}_k$ and the center $\underline{v}_i$ of the i-th cluster.

5 Query by hand-draw image

After the FCM execution, the generated fuzzy partitioning emphasizes that each pattern, i.e., each image is assigned to a clustering according to its highest value of membership. Then, output of FCM is the prototypes that are the centers of clusters and represent the more representative elements for each cluster. A query is considered a new image that must be placed in the clusters space, in order to find which cluster it belongs to and which images it is more similar to. Thus, it is processed to extract its features and then translated into a vector-based representation to be homogeneously compared with images in the clusters space. A simple way to get this evaluation is to measure the euclidean distance between the query and the prototypes. The prototype with minimal distance from the query represents the cluster which contains the most similar images. This approach has a drawback due to nature of the clustering, that is fuzzy and thus, the memberships of the images may be affected by some characteristics (features) that are natively representative of more than one cluster. Then, in order to enable user to yield a more accurate results from the hand-draw query, we focus only on patterns that lie inside a given space around the prototype. Formally, given a query s, a number c of centers v_i, and a fuzziness threshold ϵ, we use the following criteria to obtain the cluster membership. The potential cluster j is found by taking into account the minimal euclidean distance between the center and the query, that is

$$i = min_{1 \leq j \leq c} \{j | dist(s, v_j)\}$$

Because we do not have a way to establish the membership of s inside the cluster chosen, we select a reference image x_i belonging to the cluster i whose membership is immediately above the fuzziness threshold. It is obtained as follows.

$$x_i = min_{1 \leq k \leq n} \{x_k | u_{k,i} \geq \epsilon\}$$

where, n is the size of cluster i. Then, if $dist(v_i, s) \leq dist(v_i, x_i)$, the query is in the space computed as a sphere, with radius equals to $dist(v_i, x_i)$. Then, we return all the images contained in the cluster i sorted based on euclidean distance and whose membership is above the fuzziness threshold. This ranked list represents the answer to the given query.

6 Experimental results

The experimentation consists of two testing phases. The first one is related to the classification of the images. The second one evaluates the retrieval through query by sketched image. For the classification experiment, we consider a collection of images downloaded exploiting Google [1]. This testbed consists of a sample of 965 images, composed of six classes of about 160 images.All the features presented above has been exploited in the experiment. A 965 $\times$ 14 data matrix as been given as an input of FCM. Let note that the main peculiarity of the fuzzy clustering is that the membership of an image may result distributed on more than one clusters. Typically, an image (in general, a pattern) is associated to the cluster in which its membership is the highest. In order to emphasize the natural flexibility of the fuzzy clustering methods in the distribution of data, we have processed the dataset on crisp k-means too.

Table 1 synthesizes the results, associated to the experiment for the two clustering algorithms. Each cell indeed, contains resulting values computed for k-means and FCM clustering algorithms, respectively. Each row of the matrix represents a cluster obtained by the algorithm execution. Once identified which images are mainly representative of each cluster, the name of the category/class (bottle, guitar, etc.) is associated to the cluster. The first cell of each row represents the cluster/category. Thus, in Table 1, column headings contains the name

Table 1. Cluster-based evaluation, for k-means and FCM clustering, respectively

Classes	*# Misclassified.*	*# Undecided.*	*Recall. %*	*Precision. %*
bottle	21-6	0-0	86-96	93-100
guitar	7-3	0-0	95-98	95-96
leaf	92-58	0-7	44-64	41-78
apple	94-16	0-2	41-90	79-92
motorcycle	17-22	0-5	89-86	57-70
gun	13-13	0-3	91-91	92-93

of the class, the *misclassified* images, i.e. those images that have the highest membership in a class, which is not the expected one, the *undecided* images, viz. all the images which membership is almost equally distributed among two or more clusters. Note that, due to the crisp nature of the k-means algorithm, there are no resulting *undecided* images. Then, *recall* and *precision* is evaluated inside each cluster. Particularly, we define *local recall* and the *local precision*, as follows:

$$Recall = \frac{relevant\ retrieved\ images}{relevant\ images} \quad Precision = \frac{relevant\ retrieved\ images}{retrieved\ images}$$

where the *relevant images* are the images which are expected in a certain class, the *retrieval images* are all the (correct and incorrect) images which are returned in that cluster, while the *relevant retrieved images* are just the images that really belong to the right cluster, associated to the correct class. By analyzing the results, the set of leaf is the worst classified. Let us note that, comparing to each other the classes of images, the class of leaf is composed of very heterogeneous image samples: leaves with one or more tips, maple leaves are put together, without considering the obvious differences in the forms. In fact, most of misclassified data appear in cluster of apples; this is due to the different shapes of leaves: after the image processing, some leaves present rounded shapes (such as walnut and cherry leaves) that can be easily confused with apples. In fact, geometrical features such as π_{gen} and compactness assume values similar to those ones of apples. Similar considerations arise by comparing some leaves with motorcycles shapes. That means, the class of leaves is not adequately representative, because its individual are confused with other classes. A possible consequence is the amount of false positive (i.e., retrieved elements that are wrongly considered to belong to a class) increases, affecting negatively the precision. However, the global performance of FCM clearly overcomes the crisp clustering, that reveals its intrinsic weakness in clustering images. The second stage of this experiment considers the process retrieval, given a query-by-sketch. We have hand-drawn a sketch and according to Section 5, we retrieve the more similar images. In this experiment, shape extraction algorithm plays a key role in retrieving similar images results. We exploit the shape to generate the feature vector associated to an image and then, to return the similar images. Thus, the quality of result depends on the validity of extracted shapes. Figure 2 shows the results of three queries. Note that for the first and second sketches, the matches provide a very reasonable answer to the user query in the data base, while for the third sketch, the matches provide also wrong answers, due to misclassification of leaves dataset.

7 Conclusion

This paper proposes an approach for retrieving image by hand-drawn object sketch. The use of fuzzy clustering allows a better representation of the image domain: the recall and precision measures reveal a discrete performance by combining geometrical and moments features. Then, the shape query submitted as the free hand drawing evidences the effectiveness in the retrieving of returns

Fig. 2. Queries by sketch: ranked list of similar images based on the submitted queries. For each query, similar image and extracted shape are returned.

relevant similar images. Future extensions of this work foresee a development of a GUI-based application which supports the features extraction, the clustering technique; moreover a tool for hand-draw images to submit to the system.

References

1. The dataset is available at:. `http://www.dmi.unisa.it/people/senatore/www/dati/dataset.rar`.
2. J. C. Bezdek. *Pattern Recognition with Fuzzy Objective Function Algorithms*. Kluwer Academic Publishers, Norwell, MA, USA, 1981.
3. E. Di Sciascio, F. M. Donini, and M. Mongiello. Spatial layout representation for query-by-sketch content-based image retrieval. *Pattern Recogn. Lett.*, 23:1599–1612, November 2002.
4. J. Flusser and T. Suk. Pattern recognition by affine moment invariants. *Pattern Recognition*, 26(1):167–174, 1993.
5. J.-Y. Ha, G.-Y. Kim, and H.-I. Choi. The content-based image retrieval method using multiple features. In *Proc. of the 2008 Fourth International Conference on Networked Computing and Advanced Information Management*, pages 652–657, Washington, DC, USA, 2008. IEEE Computer Society.
6. M.-K. Hu. Visual pattern recognition by moment invariants. *Information Theory, IRE Transactions on*, 8(2):179 –187, february 1962.
7. C. E. Jacobs, A. Finkelstein, and D. H. Salesin. Fast multiresolution image querying. In *Proc. of the 22nd annual conference on Computer graphics and interactive techniques*, SIGGRAPH '95, pages 277–286, New York, NY, USA, 1995. ACM.
8. R. Krishnapuram, S. Medasani, S.-H. Jung, Y.-S. Choi, and R. Balasubramaniam. Content-based image retrieval based on a fuzzy approach. *IEEE Transactions on Knowledge and Data Engineering*, 16(10):1185–1199, 2004.
9. H. Lieberman, E. Rozenweig, and P. Singh. Aria: An agent for annotating and retrieving images. *Computer*, 34:57–62, 2001.
10. J. M. Saavedra and B. Bustos. An improved histogram of edge local orientations for sketch-based image retrieval. In *Proc. of the 32nd DAGM conference on Pattern recognition*, pages 432–441, Berlin, Heidelberg, 2010. Springer-Verlag.
11. D. Zhang and G. Lu. Evaluation of mpeg-7 shape descriptors against other shape descriptors. *Multimedia Syst.*, 9(1):15–30, 2003.