Strategies of Shape and Color Fusions for Content Based Image Retrieval

Paweł Forczmański and Dariusz Frejlichowski

Szczecin University of Technology
{pforczmanski,dfrejlichowski}@wi.ps.pl

Summary. The aim of this paper is to discuss a fusion of the two most popular image features - color and shape - in the aspect of content-based image retrieval. It is clear that these representations have their own advantages and drawbacks. Our suggestion is to combine them to achieve better results in various areas, e.g. pattern recognition, object representation, image retrieval, by using optimal variants of particular descriptors (both, color and shape) and utilize them in the same time. To achieve such goal we propose two general strategies (sequential and parallel) for joining elementary queries. They are used to construct a system, where each image is being decomposed into regions, basing on shapes with some characteristic properties - color and its distribution. In the paper we provide an analysis of this proposition as well as the initial results of application in Content Based Image Retrieval problem. The original contribution of the presented work is related to the fusion of several shape and color descriptors and joining them into parallel or sequential structures giving considerable improvements in content-based image retrieval. The novelty is based on the fact that many existing methods (even complex ones) work in the same domain (shape or color), while the proposed approach joins features from different areas.

1 Introduction

Content-Based Image Retrieval (CBIR) has been a very attractive topic for many years in the scientific society. Although there are many academic solutions known (for example IBM's QBIC and MIT's Photobook), there are almost no commercial or industrial applications present. It is mostly caused by not trivial problems, developers of such systems should solve. The most important are the way the images are represented in the database and the way they are being compared. The automatic recognition of objects, which are placed in the image plane, can utilize various features. The most popular and widely used are: shape, texture, color, luminance, context of the information (background, geographical, meteorological etc.) and behavior (mostly movement). It is possible to use more than one feature at the same time, but such an approach is rather rare. Usually, each recognition method is limited to only one feature. On the other hand, the literature survey shows that combining images coming from different sources (instead of different

M. Kurzynski et al. (Eds.): Computer Recognition Systems 2, ASC 45, pp. 3–10, 2007.
springerlink.com

features of the same image) is gaining the popularity among researchers [5]. The image fusion has been widely accepted in diverse fields like medical imaging, aircraft navigation guidance, robotic vision, agricultural and satellite imaging. Image fusion is a necessary stage for these applications to achieve better understanding of the observed phenomena as well as improving decision making. The approach is motivated by weaknesses of individual imageries, which can be eliminated by joining their unique features. However, there are many situations when different sensors are not available and only one type of image is a source of features. Hence, we should use as much object information as it is possible. It is obvious that every type of object representation has its advantages and drawbacks, and the choice is not always easy and evident. It depends on many conditions, e.g. application and user requirements, situation during the image acquisition process (especially the hardware parameters, weather or lighting). In the paper we focus on visual descriptors related to shape and color, since they are the most popular in the literature and guarantee good efficiency when it comes to single-feature type of recognition [1, 2, 9, 13]. To improve the CBIR efficiency we join them into parallel or sequential structures.

2 Visual Descriptors

2.1 Shape Descriptors

The shape (silhouette, contour) is a very widely used object representation. It is very useful, when we are going to identify a particular object in the image, for example in medicine (e.g. cell shapes in microscopic images), optical character recognition and many other systems, where signs are being used, robotics (e.g. machine vision, searching for faults, machine positioning, orientation in distance), ecology (e.g. pollution monitoring), criminology (e.g. fingerprints), military application (e.g. aircrafts and tanks recognition, tracking, maps identifying), visual data retrieval. Shapes of different objects are usually easy to localize and distinguish. For example in a problem of object recognition in remote sensing, the extracted shape is more stable than other features, influenced by changes in illumination, color or contrast. The shape used for recognition can be considered as a binary object, which is represented by a whole, including its interior or as a boundary (contour). It is crucial to uniquely characterize the shape and stay invariant to translation, scale and rotation [8].

Shape descriptors can be classified in various ways [4, 8, 10, 12, 13]. The first classification is based on mentioned earlier distinction between object boundary and the whole shape. The second very popular manner (as described e.g. in [13] is based on whether the shape is represented as a whole (global approaches) or by a set of primitives (structural methods). The third one distinguishes between spatial and transform domain [10].

Every shape descriptor should be resistant to as many shape distortions as possible. These distortions are considered as differences between object under recognition and the reference object belonging to the same class. In real recognition tasks one has to take into consideration the following problems divided into three main groups. The first one includes spatial transformations of an object, mainly translation, rotation in the image plane, the change of its size and the influence of the projection into resultant two-dimensional shape. The second group covers distortions of the object itself: varying amount of points, noise, discontinuity and occlusion, which is equivalent to lack of some parts or added parts to shape. The third group of problems is related to the contour representation only and includes, among others, the selection of starting point and the direction of tracing the outline. Of course any other problem, which is typical to particular representation and is a consequence of using it, will belong to this group. The elements of the second group are the most challenging and difficult to solve. Few examples of real influences of above problem on objects boundaries are presented in Fig. 1.

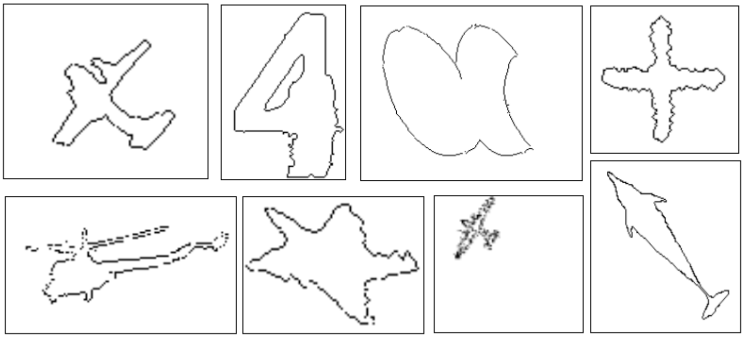

Fig. 1. Few examples of boundaries affected by shape distortions after extracting them from real images: first row Ũ an aircraft on airfield, a digit on car license plate, leaves, a symbol on a banknote; second row Ũ a helicopter, a star on flag, an aircraft in the sky, a dolphin on sea

The most important problem with shape descriptors is the fact that usually when particular method is really 'brilliant' in presence of one or even few distortions, it completely fails in the presence of another one (ones).

2.2 Color Descriptors

Color is the second most important feature taken into consideration, when it comes to content-based image retrieval. The use of color is motivated by the way the Human Visual System (HVS) works. It is known that in good lighting conditions human-being pays attention: first to intensity and color of objects, second to shape and movement, then to texture and other properties. There have been many color descriptors proposed in the past, most of them based on different color-subspace histograms and dominant values. Nowadays, when the MPEG-7 standard [1] is being introduced, the most promising are compact descriptors, which join color information and its distribution: Scalable Color

(SCD), Dominant Color (DCD), Color Layout (CLD). In our recent works we have been also utilized specific simplified representations, like RGB color histogram (RGBHIST), intensity thumbnail (8x8 pixels) of an image (IBOX8), three (R,G,B) thumbnails (RGBBOX8), mean RGB value together with mean intensity (RGBI), dominant values H and V in HSV color model (DHV).

Descriptors presented above were successfully implemented in prototype software realizing CBIR tasks (querying similar images by example). Sample results of image query based on RGBHIST can bee seen in Fig. 2. The most upper-left image is a query image, while the rest was provided by a system. Although all resulting images have similar color characteristics, it is clear, that using of only one feature is not optimal and is not in accordance with a way HVS works. Hence it is obvious that joining descriptors and decisions based on them will lead to the improvement in CBIR efficiency.

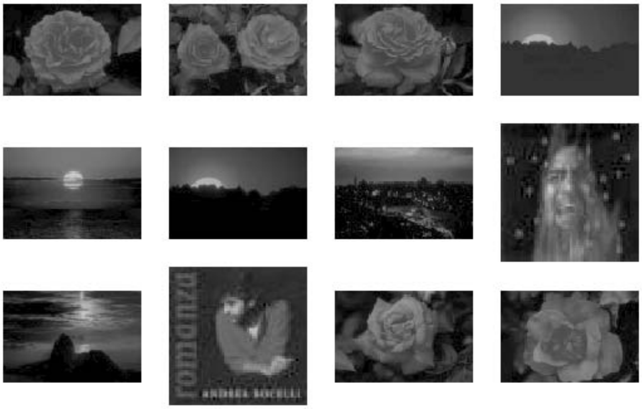

Fig. 2. Result of query involving color-only descriptors (RGBHIST)

3 Fusion of Queries

The specific of large visual data sets (consisting of several hundreds thousands of similar images) is associated with the high probability of the situation, when single descriptor used for query would give false hits. Hence, the problem of joining descriptors or decisions based on them is so important. The idea is not new and there are many examples of its successful implementation [5, 6, 7]. However, most solutions in that field use joint descriptors from the same domain (e.g. color or texture). Since the most real images contain objects that feature not only one kind of attributes i.e. color or shape, the proposed methods are not suitable for them. They are focused on either binary objects or textured rasters. The strategy of joining queries in different domains is, in fact, rather rare.

A typical system used for visual material querying and indexing consists of four main elements: the feature extractor, the comparison and classification block, the storage sub-system, and the front-end. The efficiency of such system depends mainly on the performance of the comparisons and the accuracy of description. Our proposal, presented here, is to combine few queries (using several descriptors in the shape and color domains) at the same time in order to utilize the advantages of particular, single-feature methods while reducing the influence of their drawbacks.

The idea of combining few methods in the field of pattern recognition is not new. Fusion at decision level is employed to increase classification accuracy of an image beyond the level accomplished by individual classifiers. Rank-based decisions provide more opportunities compared to other numerical score measurements. Fusion at the level of features, on the other hand, is much simpler to implement, but suffers from the incompatibility of individual scales and requires applying universal classificators (instead of feature-specific, which is undoubtedly better). For example, in the domain of shape recognition, the UNL-F descriptor is a combination of UNL and Fourier transforms [11]. In [10] two combined approaches were also explored, namely: moments and Fourier descriptors, and moments and UNL-F features. In [3] a combination of global (moments) and local (Structural Decomposition) was successfully proposed. These examples were given to show the recent tendency of combining various approaches to achieve better performance of resultant descriptor.

Our main idea, presented in this paper, is a presentation of two strategies of query fusion: sequential and parallel one using both color and shape descriptors. In this work, we propose a framework system that provides for both feature-types comparison and spatial query for unconstrained color images. The idea is based on the two-tier approach. First, each image is decomposed into regions (based on their shapes) which have properties, such as color and its distribution. Second, a color-based comparison is applied. In this way, these images are compared by comparing their individual objects. The system accommodates partial matching of objects that are the most important parts of an image not only using their shape, but also a color. The elementary query processes can be joint in sequential or parallel way.

The first approach (presented in Fig. 3) consists of n iterative queries. The first one uses $D1$ descriptor which limits the initial dataset and creates a subset $IS1$ by means of seeking and sorting images according to some similarity measure (e.g. distance metrics or correlation). Next, the $IS1$ subset is used as a initial dataset for query with $D2$. The whole process is repeated n times giving the resulting images. In this approach it is important to create the sequence of descriptors in a way that each consecutive query produces successive approximation of the required result.

In the second approach (presented in Fig. 4) we assume parallel use of descriptors to get n results. After that we apply certain voting rules (i.e. two-out-of-three, three-out-of-five and similar) and select the resulting images. It should be noticed that in both approaches we can use different classificators adequately to different features, which give distinctly better and more trustworthy results. Each strategy has its own advantages and drawbacks. The first one (sequential) utilizes an intuitive flow, which is similar to iterative way of seeking images by

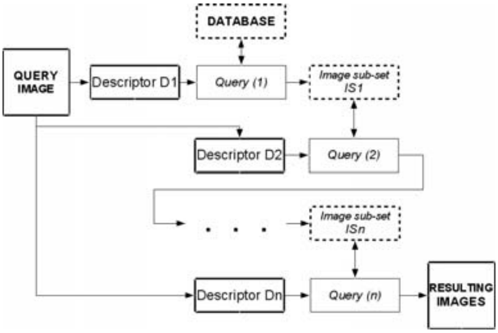

Fig. 3. Sequential structure of joining queries

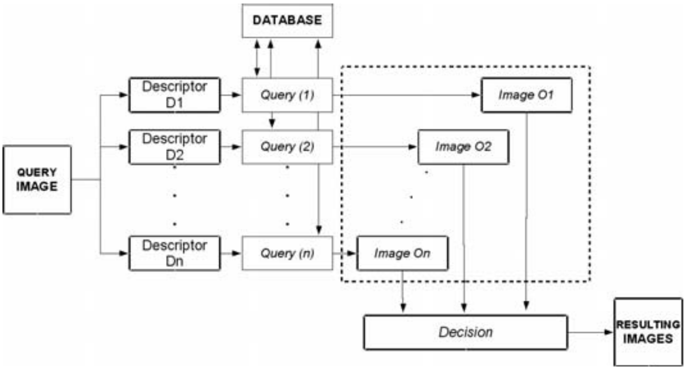

Fig. 4. Parallel structure of joining queries

humans, while the second one (parallel) makes it possible to avoid a situation, when images correctly found at the first stage are eliminated during the process of reducing the dataset at the further stages. In practice each strategy has its own specific application. According to several experiments we have performed, the sequential order is better for image retrieval based on examples, while the parallel one tends to be more appropriate for seeking images used for composing photo-mosaics, when only one resulting image is needed. Sample results of image query employing two stage, sequential seeking (based firstly on shape, then on RGB histogram) are presented in Fig. 5. As it can be seen, in comparison to the results presented in Fig. 2, the results has been distinctly improved.

The strategy of joining queries can be presented on an simplified example of a CBIR system which consists of three descriptors and three comparators, respectively. We denote the mean retrieval rate of the pairs descriptor-comparator as: P_1, P_2 and P_3. We omit a problem of decision itself by taking each single retrieval as independent event and assuming that each pair works in its optimal conditions. More sophisticated approaches can be found in [6, 7]. The total rate

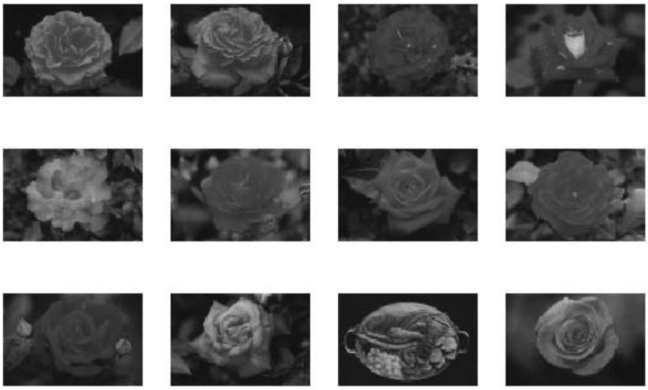

Fig. 5. Result of query involving combined features

P of such a combined system can be calculated (on the interval $\langle 0, 1\rangle$) according to the following formula:

$$P = P_1P_2P_3 + P_1P_2\bar{P}_3 + P_1\bar{P}_2P_3 + \bar{P}_1P_2P_3, \qquad (1)$$

where: $\bar{P}_1 = 1 - P_1, \bar{P}_2 = 1 - P_2, \bar{P}_3 = 1 - P_3$.

For example, if the retrieval accuracy of a single descriptor-comparator pair is equal to 0.9 (90%), which is a typical rate for current methods, then the combined accuracy will increase to 0.972.

4 Summary

In the article we showed some aspects related to multi-tier content-based image retrieval, employing shape and color information. The ideas can be applied to many different fields of digital image processing and pattern recognition. They are universal and, as it was proved, can be successfully implemented. The main advantage over the existing methods is the possibility of joining descriptors from various domains and compare them using specific metrics to get better efficiency. Since the proposed structures are the elements of a framework, they can collect descriptors focused on different image classes, e.g. faces, road signs, logos, etc. However, we should remember that there are still some unsolved problems and questions regarding successful image retrieval, especially reducing the semantic gap between low-level features and high-level meaning.

References

1. Bober M.: MPEG-7 Visual Shape Descriptors, IEEE Transactions on Circuits and Systems for Video Technology, vol. 11, no. 6 (2001) 716-719
2. Deng Y., Manjunath B. S., Kenney C., Moore M. S., Shin H.: An Efficient Color Representation for Image Retrieval, IEEE Transactions on Image Processing, vol. 10, no.1 (2001) 140–147

3. Foggia P., Sansone C., Tortorella F., Vento M.: Combining statistical and structural approaches for handwritten character description, Image and Vision Computing, vol. 17, no. 9 (1999) 701–711
4. Jain A. K.: Fundamentals of Digital Image Processing, Prentice Hall, 1989
5. Kukharev G., Mikłasz M.: Face Retrieval from Large Database, Polish Journal of Environmental Studies, vol. 15, no. 4C (2006) 111–114
6. Kuncheva L.I.: Combining classifiers: Soft computing solutions, in: Pattern Recognition: From Classical to Modern Approaches, World Scientific Publishing Co., Singapore (2001) 427–452
7. Kuncheva L.I.: A theoretical study on six classifier fusion strategies, IEEE Transactions on PAMI, 24, no. 2 (2002) 281-286
8. Loncaric S.: A survey on shape analysis techniques, Pattern Recognition, vol. 31, iss. 8 (1998) 983–1001
9. Manjunath B. S., Ohm J.-R., Vasudevan V. V., Yamada A.: Color and Texture Descriptors, IEEE Transactions on Circuits and Systems for Video Technology, vol. 11, no. 6 (2001) 703–715
10. Mehtre B. M., Kankanhalli M. S., Lee W. F.: Shape measures for content based image retrieval: a comparison, Information Proc. & Management, vol. 33 (1997) 319–337
11. Rauber T.W., Steiger-Garcao A.S.: 2-D form descriptors based on a normalized parametric polar transform (UNL transform), Proc. MVA'92 IAPR Workshop on Machine Vision Applications (1992)
12. Wood J.: Invariant pattern recognition: a review, Pattern Recognition, vol. 29, iss. 1 (1996) 1–17
13. Zhang D., Lu G.: Review of shape representation and description techniques, Pattern Recognition, vol. 37, iss. 1 (2004) 1–19