Annales UMCS Informatica AI X, 1 (2010) 41-49
DOI: 10.2478/v10065-010-0034-8

Annales UMCS
Informatica
Lublin-Polonia
Sectio AI
http://www.annales.umcs.lublin.pl/

Current challenges in content based image retrieval by means of low-level feature combining

Paweł Forczmański*, Dariusz Frejlichowski†

Division of Multimedia Systems, West Pomeranian University of Technology, Żołnierska 49, 71-210 Szczecin, Poland.

Abstract – The aim of this paper is to discuss a fusion of the two most popular low-level image features - colour and shape - in the aspect of content-based image retrieval. By combining them we can achieve much higher accuracy in various areas, e.g. pattern recognition, object representation, image retrieval. To achieve such a goal two general strategies (sequential and parallel) for joining elementary queries were proposed. Usually they are employed to construct a processing structure, where each image is being decomposed into regions, based on shapes with some characteristic properties - colour and its distribution. In the paper we provide an analysis of this proposition as well as the exemplary results of application in the Content Based Image Retrieval problem. The original contribution of the presented work is related to different fusions of several shape and colour descriptors (standard and non-standard ones) 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 single domain (shape or colour), while the proposed approach joins features from different areas.

1 Introduction

Computer vision has been explored for many years in the scientific society. This specific domain of computer science is related mainly to pattern recognition, image

*pforczmanski@wi.zut.edu.pl
†dfrejlichowski@wi.zut.edu.pl

analysis, and image processing. One of its most promising applications is content-based image retrieval, which focuses on managing large sets of visual data, e.g. images and video streams. The problem of retrieving an image based on its content by a computer is not trivial since it requires to model the human visual system (HVS). The most important are the way the images are represented in the database and the way they are being compared. The automatic recognition of the objects, which are placed in the image plane, can utilize various low-level features. The most popular and widely used are: shape, texture, colour, luminance, context of the information (background, geographical, meteorological etc.) and behaviour (mostly movement). The comparison can utilize one of similarity assessment methods, ranging from classical to those incorporating artificial intelligence. Moreover, it is possible to use more than one feature and similarity metric 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 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. For these applications, image fusion is a necessary stage in order 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).

Fig. 1 shows a typical scheme of CBIR system, where a query image is compared to the images stored in the database. The comparison is performed in several feature spaces independently. In this case, an image database (Image DB) collects all the images, yet is not utilized directly in the comparison stage. It is processed in the offline manner in order to extract visual descriptors (based on low-level features). The same descriptors are extracted from a query image. Comparing the descriptors can be performed according to a strategy, which simulates the way humans compare images.

In the paper we focus on visual descriptors related to shape and colour, since they are the most popular in the literature and guarantee good efficiency when it comes to a single-feature type of recognition [1, 2, 9, 13]. To improve the CBIR efficiency we join them into parallel, sequential or mixed structures.

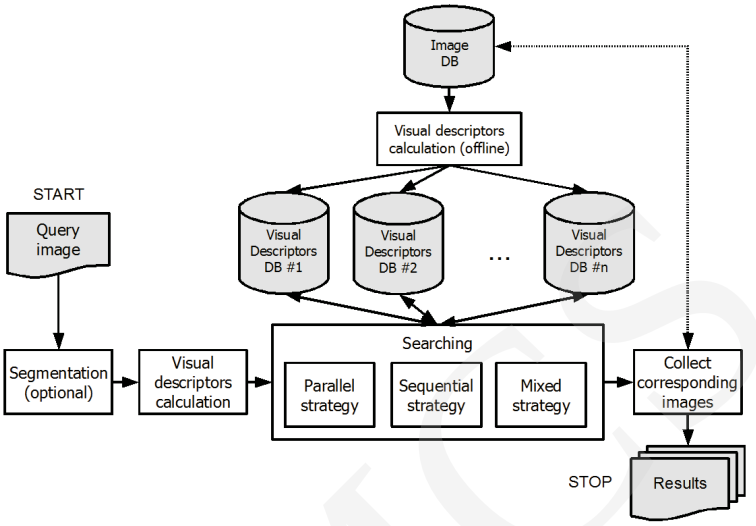

Fig. 1. General scheme of a CBIR system based on features fusion.

2 Visual Features

Amongst many types of visual features describing single objects two seem to be especially attractive and popular in use. Those are shape and color. Both are strongly influenced by HVS and the way it works. In image analysis or recognition the shape as a description is regarded especially applicable when a particular object (or objects) is identified. The areas are numerous but regarding the importance and number of applications special attention is paid to: medical problems, optical character recognition, machine vision, ecology, forensic, military and visual data retrieval.

The importance of shape in the discussed problem comes from its easy localisation and segmentation from the background (when compared to other features). 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 most popular classification is based on the mentioned distinction. The second (as described e.g. in [13]) is based on whether the shape is represented as a whole (global) or by a set of local primitives (structural methods). The third one distinguishes between spatial and transform domains [10]. Every shape descriptor should be robust to as many shape distortions as possible. They are considered as differences between the object under recognition and the reference object belonging to the same class, but stored in the database. Firstly, spatial transformations have to be described. Those are the most obvious problems and therefore for years the most explored and solved ones. One can take into consideration translation, rotation in the image plane, the

change of its size and the influence of the projection of a three-dimensional object into a two-dimensional plane. On the other hand, the description algorithm has to take into account more challenging problems, e.g. varying amount of points, noise, discontinuity and occlusion, which is equivalent to a lack of some parts or added parts to a shape. If the contour representation is used, one has to solve also the problems of selecting the starting point and the direction of tracing the outline.

Colour is the second mentioned feature. In good lighting conditions human-being pays attention: firstly to intensity and colour of objects, secondly to shape and movement, then to texture and other properties. Therefore, the colour seems to be the second very popular and promising feature and it is easy to point out many colour descriptors in the literature. Usually they are based on colour-subspace histograms and dominant values. Nowadays, when the MPEG-7 standard [1] is being introduced, the most promising are compact descriptors, which join colour information and its distribution: Scalable Color (SCD), Dominant Color (DCD), Color Layout (CLD). In our recent works we have been using also specific, simplified representations, like RGB colour histogram, 8x8 pixels intensity thumbnail of an image, three (R,G,B) thumbnails, mean RGB value together with mean intensity, dominant values H and V in the HSV colour model.

The traditional scheme for a system applied in querying and indexing images is composed of four main elements. The feature extractor is the first one. The second is the block devoted to comparison and classification. Separately, the storage element has to be considered. Finally, the front-end has to be provided. The results achieved by the above scheme strongly depend on the two parameters: the efficiency of comparisons and accuracy of description. In order to improve the retrieval results we propose to combine queries using descriptors from the shape and colour domains, used jointly. That gives the possibility of increasing the influence of advantages of particular methods and reducing the influence of the drawbacks. Hence, in our work we concentrate on the combination of the two features briefly described in this section. It comes from the obvious observation that in the case of large visual databases the usage of single descriptors can not be enough. The idea is promising but not new. One can find examples of its successful implementation [5–7]. However, the existing approaches are usually based on joint descriptors from the same domain (e.g. colour or texture). For many real image data sets this is not suitable, because they contain significantly richer and heterogeneous information. Therefore our approach emphasizes the possibility of joining queries in different domains.

As it was mentioned, the idea of combining a few methods in the field of pattern recognition is definitely not new. Fusion at the 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 on 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 classifiers (instead of feature-specific, which is

undoubtedly better). The elementary query processes can be joint in a sequential or parallel way. However, mixed strategy is also possible. The first approach consists of n iterative queries. Each single query limits the initial dataset and creates an image subset by means of seeking and sorting images according to a similarity measure (e.g. distance metrics or correlation). Next, the obtained subset is used as an initial dataset for the next query. 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 we assume parallel use of descriptors to get n independent 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 classifiers adequately to different features, which gives 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 the iterative way of seeking images by humans, while the second one (parallel) makes it possible to avoid a situation, when images correctly found in the first stage are eliminated during the process of reducing the dataset in further stages. In practice, each strategy has its own specific application. According to several experiments we have performed, the sequential order is better for the image retrieval based on examples, while the parallel one tends to be more appropriate, when only one resulting image is needed. However, in the experiments presented in this paper, some other properties are also explored (see section 3. for examples). Joining queries may significantly improve the retrieval accuracy. It can be observed using a simplified example of the CBIR system which consists of three descriptors and three comparators, respectively. We denote the mean retrieval rate of the pairs descriptor comparator as: P1, P2 and P3. In this case we take each single retrieval as an independent event and assume that each one works under its optimal conditions. It is a very basic approach, yet a more sophisticated decision taking solutions can be found in [6, 7]. The total rate P of such a combined system can be calculated (on the interval <0,1>) according to the following formula:

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

where: $P_i = 1 - \bar{P}_i \ \forall i \in \{1, 2, 3\}$. For example, if the mean retrieval accuracy of a single descriptor-comparator pair is equal to 0.8 (80%), which is a typical rate for the state-of-the-art methods, then the combined accuracy will increase to 0.9.

3 Experimental flags retrieval

In order to explore the properties and possible applications of algorithms based on fusion of low-level image features an experiment with flags retrieval was performed. The idea here was to verify the influence of our strategies on retrieval results.

The images used in experiments were collected from the "Flags of the World" database, which consists of over 74,000 of pictures (flag images of various kinds - national, regional, provincial, state, municipal, historical, civil, war, international, organisation, naval, political, sport, personal, positional, fictional, etc.). The experimental database was limited to flags only (24,000 images). Other symbols, coats of arms, etc. were eliminated.

Every single experiment was conducted based on the same idea. The query image (taken from the database) was presented to the retrieval system that used combination of descriptors and a few closest base elements were presented as an output. The stress was put here on the results of joining methods of different kinds, mainly colour and shape descriptors. It was interesting for us what properties will be emphasized in that case. It is obvious that if only the colour descriptors are used, the result from the above 20 thousand flags will be rather homogeneous. For example providing the Polish flag to the system will result in several dozen of exactly the same white and red flags, because one can find many of them across the world (Austrian states: Tirol and Upper Austria, Czech cities: Bubanec, Holesovice-Bubny, Karlin, German city Bad Durrenberg, etc.). Therefore the results of that kind are not even presented here. The special emphasis is put on co-operation of completely different features (descriptors).

Two exemplary results from many performed ones are described in this section. In our opinion, they are the most interesting and confirm the relevance of combination of various descriptors'.

The first case (see Fig. 2) was based on fusion of Colour Layout, Scalable Colour and Edge Histogram descriptors. As it can be seen in the figure, thanks to this, various characteristics were emphasized by the system. On one hand, the colours presented in the query image were important. On the other hand, the star was also influencing the results. This is an obvious result of the parallel strategy, where every element of the scheme has its contribution to the result (in this case equal for every component).

The second presented test (see Fig. 3) used the United States flag, a very characteristic one, as a query. This time combination of RGB Histogram and Moments Invariants was applied. The appearance of flags other than the mentioned one is very interesting because, as it turned out, in the tested fusion scheme the most important property of the query image was the domination of red-white stripes in the image. Probably it was the result of the sequential strategy, where firstly the histogram is used and secondly, in the selected subset, the shape descriptor becomes more important.

Fig. 2. Sample retrieval results. The most up-right image is a query object.

Fig. 3. Sample retrieval results. The most up-right image is a query object.

4 Conclusions

In the article we showed some aspects of multi-tier content-based image retrieval, employing visual description provided by more or less standard low-level features. The

ideas presented here can be applied in many different fields of digital image processing and pattern recognition. This approach is universal and, as it was proved, can be successfully implemented. Moreover, its 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.

In the article we showed some aspects of multi-tier content-based image retrieval, employing visual description provided by more or less standard low-level features. The ideas presented here can be applied in many different fields of digital image processing and pattern recognition. This approach is universal and, as it was proved, can be successfully implemented. Moreover, its 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.

References

[1] Bober M., MPEG-7 visual shape descriptors, IEEE Transactions on Circuits and Systems for Video Technology 11(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 10(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 17(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 15(4C) (2006): 111–114.

[6] Kuncheva L. I., Combining Classifiers: Soft Computing Solutions. 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(2) (2002): 281–286.

[8] Loncaric S., A survey on shape analysis techniques, Pattern Recognition 31(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 11(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 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 29(1) (1996): 1–17.

[13] Zhang D., Lu G., Review of shape representation and description techniques, Pattern Recognition 37(1) (2004): 1–19.