New Strategy of Relevance Feedback based on Specific Feature Selection

X.Muñoz, J.Freixenet, J.Martí, X.Cufí and J.Batlle

University of Girona. Institute of Informatics and Applications.
Campus de Montilivi s/n. 17071. Girona, Spain
{xmunoz,jordif,joanm,xcuf,jbatlle}@eia.udg.es

Abstract

Content-Based Image Retrieval has become one of the most active research areas in the past few years. However, most current systems are designed following a computer centric approach. In contrast in this paper we present an image retrieval system which focuses on the importance of the user. A new relevance feedback strategy based on the selection of the most appropriated feature subset for each specific query is presented. This proposal breaks with the habitual feature weighting to capture the user's goal, while attacks the issue like a feature selection problem.

Keywords: computer vision, image retrieval, user-centric approach, relevance feedback, feature selection.

1 Introduction

Current content-based image retrieval (CBIR) systems often rely on a computer centric approach. The corresponding system design strategy for theses systems is to first find an optimized feature set that permits to obtain good results for any query. Then, during the retrieval process, the user must specify the weights for each feature which characterizes the query. In these systems, the result of the retrieval relies on selected features (system design) and specified weights (user).

However, the performance of the computer centric approach is not always satisfactory due to: 1) The existent gap between high level concepts and low level features, and 2) the subjectivity of human perception of visual content. Because of the fixed weights, this approach can not effectively model high level concepts and user's perception subjectivity [10]. Often there are no mechanisms, or a great effort is needed, to determine what is important in a query image. Furthermore, specification of a precise set of weights imposes a heavy burden on the user as it requires the user to have a comprehensive knowledge of the low level features representations used in the retrieval system which is not normally the case. Motivated by these limitations, recent research in CBIR has moved to an interactive mechanism which involves the user as part of the retrieval process.

Relevance feedback, based on an interactive retrieval approach, was proposed to take into account the above two characteristics in CBIR. During the retrieval process, the user's high level query and perceptual subjectivity are captured from the user's feedback about the relevance of previously retrieved images. The user only needs to mark which images he thinks are relevant to the query. Furthermore, the burden of specifying the weights is removed from the user.

In addition, CBIR have to emphasize the role of the user to define what the exact content to be retrieved is [2, 13]. Users generally want to find images based on the objects they contain. Therefore, the goal is to retrieve images which contain some concrete objects, eg. all images with a `red car`. Obviously, the user is the one who knows best what is relevant for his query, and consequently, he is who has to specify the object query. While in Blobworld [2] the user specifies his interest by marking a region over a previously segmented image, in the system of Tian, Wu and Huang [13] the user uses a bounding rectangle. Here, we propose an approach that combines both main advantages: the simplicity of the bounding rectangle and the accuracy of a segmented region.

This paper focuses on the interaction between the user and CHAMELEON [6], a region based image retrieval. CHAMELEON uses a fast gross-segmentation technique integrating region and boundary information [8] to on-line segment the query image. The paper presents a new strategy which emphasizes interaction between user and system in two different aspects:

- A friendly interface using snakes allows the user to specify which is exactly the object of interest in the query image, and
- a new relevance feedback strategy based on the selection of low level features is proposed.

The last issue breaks with the habitual feature weighting to capture the user's goal, while attacks the issue like a feature selection problem. The main idea is that the system learns how to describe an object by selecting the feature subset which best characterizes this object. These features constitute the query space in which the object is more easily distinguished of the other ones. So, the identification of the object in the posterior process of object retrieval is significantly improved.

The rest of this paper is structured as follows: Section 1 finishes with a brief introduction to relevance feedback and some examples of the most habitual strategies to carry out the retrieval refinement. The user interface to select the desired object is described in Section 2, while Section 3 presents the relevance feedback strategy using feature selection and argues its convenience in front of the traditional proposals. Experimental results are presented in Section 4 and conclusions are given in Section 6.

1.1 Relevance Feedback

Image retrieval systems often rely heavily on the success of one-shot queries using optimised feature sets to obtain the best possible results. However, if there is a significant discrepancy between the similarity as calculated by the system and the notion of similarity in the user's mind, the results are destined to be unsatisfactory. Certain features or feature subsets may have varying degrees of importance with respect to the user, the query image, and the particular retrieval goals of the user [7]. This problem has served as the impetus of what is known as *relevance feedback*.

The basis of relevance feedback is the fact that the user knows considerably more about the query being made than can be conveyed in a set of low level features [14]. Also, the mechanism exploits the user's ability to rapidly recognise images which match the particulars of the query. The user can apply feedback to previously retrieved objects indicating their relevance or irrelevance to the user's requirements. This information will allow adjusting the query so that the result is a better approximation on the user's needed information.

Various ways of user feedback have been considered. During the session, the system updates the query space, attempting to learn from the user's feedback. A common approach to the implementation of relevance feedback for a system using image descriptors in numerical form is that of feature weighting and is based on the vector model used for textual documents. In the proposal of Tian et al. [13], the higher weight is given to the feature that has the smaller average distance based on the relevant images. This is the feature that has a most constant value in relevant images, so is considered a good descriptor of the user's goal. A similar approach can be found in the work of Rui et al. [11] who proposed a standard deviation based weight updating approach. The authors argue that intuitively, if all the relevant objects have similar values for one feature , it means that this feature is a good indicator of the user's information need. On the other hand, if the values for one feature are very different among the relevant objects, then this is not a good indicator. Based on this analysis, the inverse of the standard deviation of each feature is a good estimation of the weight.

There is not doubt about the great impact that the relevance feedback and the interaction with the user have had in content-based image retrieval system. As Smeulders et al. affirm in a recent review about content-based image retrieval [12], any information the user can provide in the search process should be employed to provide the rich content required in establishing the meaning of a picture. So, the interaction should form an integral component in any modern image retrieval system.

2 Object query

The first step in the query is the specification of what is the interest of the user. This action permits identification of the object or scene element which should be present in the images to be retrieved.

There have been some proposals that try to automatically determine what is the user's interest from a selection of images which ensemble defines the goal. One option is that the user selects $m > 1$ images, which all them are representative of the target. The common characteristics of the m images query are capable of defining the user's goal [4]. This process can be improved further by adding negative examples into the image query set. This is achieved in [1] by constructing a query q best describing positive and negative examples indicated by the user.

However, this process is always a not easy goal and requires that the user provides a set of images containing the object to be retrieved. For example, if the user is interested in images containing cars he needs to supply to the system a set of few images with cars, and then waiting that the system is really capable of understanding his query. An easier way is simply leave that the user is who indicates what he is interested. So, in CHAMELEON the selection is carried out by the user in a simple way, indicating the object with a polygon which entirely contains the object. This rough selection is used to initialize the initial placement of a snake. Some researchers have emphasized the need to incorporate information supplied by region segmentation in the snake itself, more concretely in its energy functional [3]. In this approach a term derived from region information is added to the external part of the energy functional. As a result, points on the contour are allowed to expand or contract according to the fit between contour and region information. The objective function proposed in this work is a function of conditional probability $P(p|I_r, I_p)$, or the probability of obtaining the p-contour given the segmented image I_r and the image of the potential of the gray level gradient I_p which is extracted from the magnitude gradient image and measures the proximity of a point to high gradient values. The function is constituted by the sum of two terms

$$M(p, I_r, I_p) = M_{region}(I_r, p) + M_{potential}(I_p, p) \tag{1}$$

The first term in the equation, $M_{region}(I_r, p)$, measures the goodness of the match between the contour and the perimeter of the segmented interior of the object. This method rewards the boundary which contains as much of the inside region and as little of the outside as possible. The second term in the equation, $M_{potential}(I_p, p)$, depends on the coincidence of the parameterized boundary with the image edges.

The optimization of the energy functional, which is performed using a Greedy algorithm, molds the snake to the boundary of the object. An example of the object selection process is showed in Fig. 1. This process allows to extract an accurate representation of the placement and shape of the object of interest. Also, it is important to state that little effort is required of the user, as the process is reduced to a simple and coarse selection. The more demanding labor of obtaining the exact shape of the object is carried out entirely by the system.

The selected object is then described by its color, texture and shape descriptors (details of the features extracted can be obtained in [6]). This constitutes a 20-dimensional feature vector which stores the information about the object query.

3 Relevance feedback by feature selection

The feature vector above mentioned describes the object to be retrieved. However, the use of the complete feature set often does not allow to capture the user's query requirements. An example of this is the query of a car. Although the shape, texture and size might be similar across the spectrum of examples, the colour can vary greatly. The similarity measure the user has in mind only takes into account shape and texture and does not give importance to the colour of the object. So, the features which best describe the object is a subset of the complete feature set. The use of other features (colour in the example of the car) can only damage the retrieval because it leads to a discrepancy between the similarity as calculated by the system and the user's notion of it. Hence, we plan the relevance feedback as a features selection problem, in which the goal is to find the subset of features which best captures the particular retrieval goals of the user.

Feature selection can also be seen as a special case of feature weighting where each weight is either 0 or 1, and thus weighting methods are potentially more powerful. However, because they have more degrees of freedom, they can also be harder to apply succesfully, especially when are few training examples [5]. Obviously, in an image retrieval system the user can not be required to provide a lot of images to effect (or refine) the query. Hence, the set of training examples will be always reduced and the feature weighting approach can not be correctly

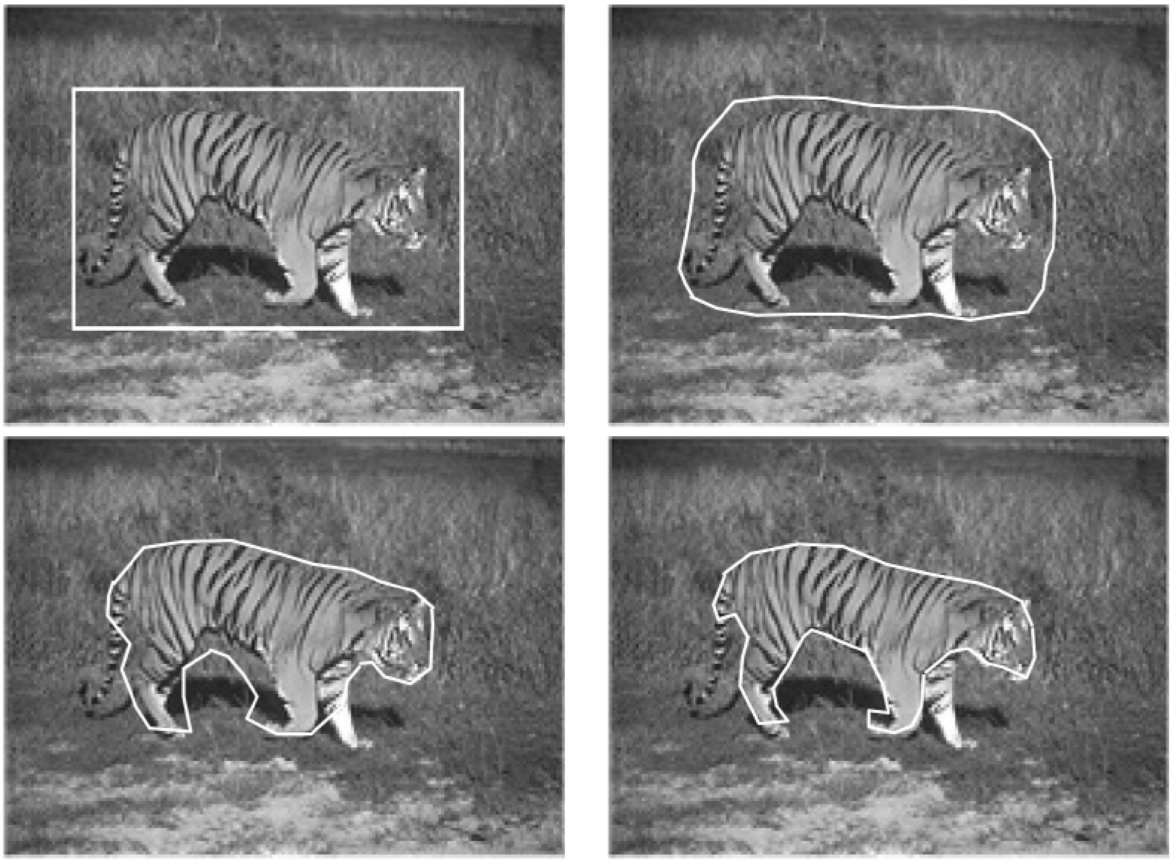

Figure 1: Object query: the user roughly indicates the object of interest, surrounding it whit a polygon that is used to initialize the snake. Then, the optimization energy process provides an accurate boundary of the object.

applied.

Furthermore, facing the relevance feedback as a feature selection problem has two main advantages. The first one, as has been above described, is the similar-human way that the problem is solved. The relevance feedback looks for a similarity measure close that the own user would use to find the object in the image database. The second one is the a priori analysis of the query space. The feature selection has the goal to obtain the space in which the object retrieval can be carried out with the best results. So, the different spaces are analyzed to decide what is the most adequate to effect the posterior retrieval. This is a principle that the classical feature weighting approach does not follow because the convenience of the query space for the retrieval and the effect of the weight updating in the posterior results are not analyzed. In other words, traditional weighting approach change the query space without knowledge about the consequences of the updating. On the other hand, the proposed feature selection based strategy tries always a priori to determine the convenience of the query space.

3.1 Query space selection

The system selects a first set of images projecting the region's features onto the indexes space and looking for the nearest neighbors. More formally, the system returns the images with potentially similar regions by minimising the Euclidean distance measure from the region's feature vector. Using this method, an ordered list of the best matches is created. This first retrieval is not intended to be perfectly accurate but establishes a starting point to begin the construction of a class representing the user's query object. The user must mark some of the returned images as positive or negative examples based on the nature of the query. This allows the system to form a small training set from which to improve the retrieval.

The goal of the feature selection is to correctly choose those features which allow examples belonging to different classes to occupy disjoint regions in a m-dimensional feature space. This process, given a set of N features, selects a subset of size m (where $m < N$) which obtains the highest value of a criterion function $J(X)$, assuming that a higher value indicates a better feature subset. The $J(X)$ is calculated as the Euclidean distance between the set of examples marked as positive and negative. More concretely, the system tries to find the feature space in which the closest positive and negative examples have a larger Euclidean distance. The criterion function is computed as

$$J(X) = \min d(x, y) \qquad x \epsilon A, y \epsilon B \tag{2}$$

where A is the set of positive examples and B is the set of negative examples, while $d(x, y)$ is the normalized Euclidean distance between the points x and y in the X feature space of dimension $m < N$.

The search for the best feature space could be accomplished by an exhaustive search. However, this technique may be too costly and practically prohibitive even for a medium sized feature set size. Furthermore, one of the main constraints of retrieval systems (especially in user centric systems) is time. In order to solve this problem, other methods have been proposed in the literature to attempt to reduce computational complexity by compromising performance. We have adopted the SFSS (Sequential Forward Floating Search) heuristic search method [9], which is based on a typical bottom-up approach. It starts from the empty set and, in each iteration, generates new subsets by adding the best feature. Besides avoiding the nesting of features, the method incorporates a backtracking process.

After the feature selection process is performed, the next set of images is obtained by comparing regions to the examples from the user in the selected feature space. The Euclidean distance between a given region's feature vector and each of the examples is calculated and example E_{min} with the minimum distance is found. If the example E_{min} is marked as negative, the region is rejected. Those regions closest to a positive example are chosen as the best matches for the new set of retrieved images. The user will mark some of these images as relevant or irrelevant, and they will be added to the image training set to repeat the process. Thus, the iterative refinement can continue until the user is satisfied with the resulting images.

3.2 Off-line training

In order to facilitate posterior queries, at all stages of the refinement process the user has the option of saving the positive and negative examples to set a class database. A key image and textual tag are assigned to the class to facilitate future identification. In this way, when a user wants to retrieve images belonging to a class upon which a previous query has already been based, the relevance feedback process need not be repeated.

Moreover, after a query sessions is completed, and off-line training phase can be performed based on the information stored during the user's interaction with the system. With less emphasis on high-speeds queries and without time constraint, a longer search is used to improve the results of the feature selection process. The exhaustive search of all the feature subset spaces can guarantee that the optimal solution will be find, so the best feature subset for the query can be used in next times.

4 Experimental results

To test the system, we have used a database which contains more than 10.000 images which come from 100 categories in the COREL photo galleries. Two example queries are shown in Fig. 2. The first column shows the results of the initial query, while the second column shows the results of relevance feedback process after a few iterations.

Images corroborate the difficulty in capturing the user's goals with a simple image query. These problems are carried by systems without relevance feedback (Blobworld system of Carson et al. [2] is a well know example). The first column of the figure shows a `golf player` who has been retrieved in the `rose` query. The golfer's red sweater easily provokes this confusion. As stated in the second column, the interactivity with the user refines the results adjusting the query to the user's requirements. Images marked as negatives (and their similar ones) are rejected in the next iterations of the refinement process, while positive images allow the retrieval of some new images which initially had not been returned. In the example of the `rose` query, the system selects the features which best distinguish positives from negatives; that being the roses from the red pullover and red car. A weak point of the proposal is the possibility that new false positive appear after the feedback from the user. An example of this problem is shown in Fig. 2, in which a `coffee cup` image appears in the refined results of the `tiger` query. However, this situation might be solved with some new iterations of the relevance feedback process. Besides, the obtained results have to be considered as positives.

5 Future work

Further work in CHAMELEON targets to test the performance that different generic feature selection methods can offer to the system. In addition to the current SFSS (Sequential Forward Floating Search) heuristic search method, the immediate future consists on the analysis and test of other methods as: SFS (Sequential Forward Selection), SBS (Sequential Backward Selection), SBFS (Sequential Backward Floating Search), and GA (Genetics Algorithms).

Besides, work is already under way to examine how the user can combine classes using boolean operators to retrieve images contain several object types. It can easily be envisaged that answering a query such as "show me images of a car on a beach" could be made possible using this approach, making use of the "car", "beach", "blue sky" and "sea" classifiers in conjunction.

6 Conclusions

In this paper we have presented the interaction user-system in the CHAMELEON project, an image retrieval system which emphasizes the role of the user as basic element of the retrieval process. The interface for object selection, which allows to the user defining the object to be retrieved, has been described. The method is based on the use of a snake to obtain the exact shape of the object from a roughly selection by the user.

Moreover, a new feature selection based relevance feedback strategy has been proposed. The main advantages of this technique have been noticed and also the convenience of using this proposal in front of the most habitual feature weighting based approaches. Experimental results show that this proposal captures the user's requirements and achieves promising results.

Acknowledgements

This work has been partially developed thanks to the support of the Departament d'Universitats, Recerca i Societat de la Informaci de la Generalitat de Catalunya.

References

[1] J. Assfalg, A. Del Bimbo, and P. Pala. Using multiple examples for content based retrieval. *International Conference Multimedia and Expo*, 2000.

[2] C. Carson, M. Thomas, and S. Belongie. Blobworld: A system for region-based image indexing and retrieval. In *Third International Conference on Visual Information Systems*, 1999.

[3] A. Chakraborty, L.H. Staib, and J.S. Duncan. Deformable boundary finding influenced by region homogeneity. *Computer Vision and Pattern Recognition*, pages 624–627, 1994.

[4] I.J. Cox, M.L. Miller, T.P. Minka, and T.V. Papathomas. The bayesian image retrieval system, pichunter: Theory, implementation, and pychophysical experiments. *IEEE Trans Image Processing*, 9(1):20–37, 2000.

[5] P. Domingos. Exploiting context in feature selection. In *Workshop on Learning in Context-Sensitive Domains ICML-96*, pages 15–20, 1996.

[6] J. Freixenet, X. Muoz, X. Cuf, and J. Mart. Chameleon: A region-based image retrieval system. In *Spanish Symposium on Pattern Recognition*, 2001.

[7] S.D. MacArthur, C.E. Brodley, and C. Shyu. Relevance feedback decision trees in content-based image retrieval. In *IEEE Workshop on Content-Based Access of Image and Video Libraries*, 2000.

[8] Muñoz, X. Cufí, J. Freixenet, and J. Martí. A new approach to segmentation based on fusing circumscribed contours, region gowing and clustering. *International Conference on Image Processing*, pages 800–803, 2000.

(a) (b)

Figure 2: Retrieval results (rank from left to right and from top to bottom). (a) Initial query retrieval results, (b) retrieval results after few iterations of relevance feedback.

[9] P. Pudil, J. Novovicova, and J. Kittler. Floating search methods in feature selection. *Pattern Recognition Letters*, 15(11):1119–1125, November 1994.

[10] Y. Rui, M. Huang, T.S. Ortega, and S. Mehrotra. Relevance feddback: A power tool for interactive content-based image retrieval. *IEEE Trans on Circuits and Systems for Video Technology*, 8(5):644–655, 1998.

[11] Y. Rui, T.S. Huang, and S. Mehrotra. Content-based image retrieval with relevance feedback in mars. *International Conference on Image Processing*, 1997.

[12] A.W.M. Smeulders, M. Worring, S. Santini, A. Gupta, and R. Jain. Content-based image retrieval at the end of the early years. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(12):1349–1380, December 2000.

[13] Q. Tian, Y. Wu, and T.S. Huang. Combine user defined region-of-interest and spatial layout for image retieval. In *International Conference on Image Processing*, volume III, pages 746–749, 2000.

[14] M.E.J. Wood, N.W. Campbell, and B.T. Thomas. Iterative refinement by relevance feedback in content-based digital image retrieval. *ACM Multimedia 98*, pages 13–20, 1998.