A NSGA Based Approach for Content Based Image Retrieval

Salvador Moreno-Picot, Francesc J. Ferri, and Miguel Arevalillo-Herráez

Department of Computer Science, University of Valencia
Avda.de la Universidad s/n. 46100-Burjasot, Spain
salvador.moreno@uv.es

Abstract. The purpose of CBIR (Content Based Image Retrieval) systems is to allow users to retrieve pictures related to a semantic concept of their interest, when no other information but the images themselves is available. Commonly, a series of images are presented to the user, who judges on their relevance. Several different models have been proposed to help the construction of interactive systems based on relevance feedback. Some of these models consider that an optimal query point exists, and focus on adapting the similarity measure and moving the query point so that it appears close to the relevant results and far from those which are non-relevant. This implies a strong causality between the low level features and the semantic content of the images, an assumption which does not hold true in most cases. In this paper, we propose a novel method that considers the search as a multi-objective optimization problem. Each objective consists of minimizing the distance to one of the images the user has considered relevant. Representatives of the Pareto set are considered as points of interest in the search space, and parallel searches are performed for each point of interest. Results are then combined and presented to the user. A comparatively good performance has been obtained when evaluated against other baseline methods.

1 Introduction

Usually, a CBIR system represents the images in the repository as a multi-dimensional feature vector extracted from a series of low level descriptors, such as color, texture or shape. The perceptual similarity between two pictures is then quantified in terms of a distance/similarity function defined on the corresponding multi-dimensional feature space. A major problem with CBIR systems is the so called "semantic gap", which refers to difficulty of translation of user's intentions into similarities amongst low level features. Relevance feedback, a technique inherited from traditional information retrieval, has been used to increase the efficiency of CBIR systems helping to induce high level semantic contents from low level descriptors. When relevance feedback is used, a search is considered an iterative process. At each iteration, the system retrieves a series of images ordered according to a pre-defined similarity measure, and requires user interaction to mark the relevant and non relevant retrievals. This data is used to adapt the similarity measure and produce a new set of results, repeating the process until the desired picture is found.

Relevance feedback has been a major topic of research during the last two decades (see [1, 2]). First methods were based on adapting the similarity measure and moving

J. Ruiz-Shulcloper and G. Sanniti di Baja (Eds.): CIARP 2013, Part I, LNCS 8258, pp. 359–366, 2013.

the query point so that more emphasis is placed on relevant elements and less on irrelevant ones [3–5]. This type of techniques use the user's judgments to dynamically adjust the weights of each feature, and to produce a new query point that represents his/her interest in a more reliable way. In general, these are the fastest techniques, but they assume the existence of a unique query point. A large number of probabilistic methods have also been proposed *e.g.* [6–8]. Most of these are based on estimating posteriori probabilities from the prior probabilities and the relevance judgments provided by the user. One particular way to estimate these probabilities is by using nearest-neighbour estimators [9, 10]. The use of supervised learning techniques has also been a major trend in the development of relevance feedback mechanisms. In this context, SVMs (Support Vector Machines) have been widely used [11–13], despite the difficulties associated with fine-tuning the retrieval systems choosing the optimal set of parameters for the SVM [9]. Other successful approaches to CBIR include the use of fuzzy sets [14], self organized maps [15] or evolutionary computation [16] to determine the degree of relevance of each image in the database.

In this paper, we propose a novel technique which considers the search as a multi-objective optimization problem. Each positive selection constitutes an objective, and the search space is explored to find a representative set of trade-off solutions between the objectives. Then, each member of this representative set is chosen as a seed, and the search proceeds concurrently at each seed. By using this method, the search not only takes place in regions surrounding the relevant selections but also in others areas in between. This is in contrast to many other existing techniques, which concentrate the search only on regions around known positive samples.

The remainder of the paper is organized as follows. First the technique proposed is explained in section 2. Then , the approach is evaluated in section 3. Finally, some conclusions are drawn in section 4.

2 The Algorithm

2.1 Problem Formulation

Let us consider the discrete solution space of all M images stored in the repository and denote it by $\{Im_1, Im_2, \ldots, Im_M\}$. Let us also denote the set of P relevant user selections by $\{Im_1^+, \ldots, Im_P^+\}$, and the non-relevant selections by $\{Im_1^-, \ldots, Im_N^-\}$. Let us assume that a similarity function s exists which produces an estimate of the resemblance between any pair of images.

Let us also consider a set of P objectives $\{o_1, o_2, \ldots o_p\}$ for each image Im_x, and define each objective o_i as the similarity between the image and the corresponding relevant user selection Im_i^+. The similarity function s can then be used to measure the degree of satisfaction of the objective o_i as $s(Im_x, Im_i^+)$. This formulation allows us to consider each of the M images in the repository as a potential solution to the problem, and the similarity to each of the P relevant selections as a different objective which should be maximized.

When a problem has multiple objectives, several optimal solutions may co-exist. These are all possible non-dominated solutions to the problem. A solution is said to be non-dominated if there is no other solution which simultaneously satisfies all the objectives better. In the absence of any further information, these cannot be said to be worse than any other. The set of all non-dominated solutions to a problem is commonly referred to as the Pareto optimal set.

The calculation of the Pareto optimal set may yield a large number of non-dominated solutions, specially for large numbers of relevant selections. In general, this implies that it would not be possible to show the entire set to the user. For this reason, we chose conveniently scattered representative samples from the Pareto optimal set. In addition, trade-off solutions which are closer to a negative selection than to a positive one are removed, according to the principles of a nearest neighbor classifier. The remaining ones are treated as seeds for potential regions of interest. Then, a ranking is produced for each seed. These are computed by sorting all images in the repository by their similarity to the seed, according to the function s. The rankings are finally combined iteratively, by taking one element from each ranking at each round.

2.2 Implementation

Determining the Pareto optimal set in a discrete solution space is a simple but also a time consuming operation. Every solution has to be compared against the rest and, in the worst case, it takes $O(P \cdot M^2)$, with P representing the number of objectives (positive selections) and M the number of solutions evaluated (the number of images in the database). With usual values of M in CBIR systems, this cost becomes prohibitive.

An alternative is to assume a continuous search space and use a MOEA (Multi-Objective Evolutionary Algorithm) to determine a spread of solutions along a set which is close to the true Pareto optimal front. The algorithm NSGA-II [17] has been chosen for this purpose. Despite that this algorithm does not perform best with a large number of objectives [18], it provides a spread set of solutions which is sufficient for our purpose.

In addition, the use of this approach provides two major advantages. In the one hand, the parameterless diversity preservation mechanism of NSGA-II provides a representative spread set of solutions directly, with as many elements as the population size used. This means that the optimal Pareto set does not need to be post-processed to obtain the desired representative set of spread solutions. In the other hand, the stopping criteria can be decided so that the response time is kept within reasonable limits.

In our implementation, the feature vectors of the positive solutions are provided as an input (these are known to be part of the optimal Pareto set), and the genetic algorithm generates a set of feature vectors that represent the optimal Pareto set. Negative selections in previous iterations of the same search are accumulated and a restriction is imposed on potential solutions to the problem. They have to be closer to a positive than to a negative selection, as determined by the similarity function s. If this is not the case, the potential solution is discarded. To avoid inconsistent solutions or solutions outside the border of the multi-dimensional search space, all feature vectors produced in the process are repaired so that their features are all in range and histogram descriptors add to the appropriate amount. To this end, out of range values are replaced by the nearest valid value and histograms are linearly scaled so that all components add to one.

Once the representative set of spread solutions has been determined, each of its members is used as a seed to drive the search process. Separate rankings are built for each feature vector in the set. To build the final ordering, these are visited iteratively until no elements are left. At each iteration, the top element from each ranking is extracted, and added to the final ordering if it is not already present. Observe that, for the simplest case when a single picture is selected as relevant, the optimal Pareto set would be the image itself, and all pictures in the repository would be ranked according to their similarity to this image.

3 Evaluation

To evaluate the results, an experimental set-up similar to those reported in [9] and [19] has been implemented. These systems use classified databases and simulate user judgment according to the class information available. In our case, a fixed number of images were chosen at random from each class, avoiding repetitions, and these were submitted as targets to the system. At each iteration, the system made automatic judgments on the first 50 images returned by the algorithm. Images which belong to the same class as the target were considered relevant and any other non relevant.

We compare the results obtained with this algorithm to those obtained by using other existing techniques, namely a) a classical feature weighting and query movement approach, implemented as presented in [5]; and b) an engine that uses similar principles to those used in the PicSOM system [20]. From now on, these algorithms will be referred to as the Query movement and the SOM-based approaches respectively. The SOM-based approach uses 64x64 SOMs for the first repository and 16x16 SOMs for the second. Because of the relatively small size of the repositories, standard SOMs have replaced the hierarchical SOMs used in the original publication. To allow for a fair comparison, these two algorithm have been adapted to work with the same feature sets. Note that although the results obtained with SOM approach may not be generalizable (the performance depends on the size of the maps and the low pass filter applied), they provide an indicative baseline for comparison purposes.

To test the approach for different database systems two different collections have been used:

– The first repository is composed of 30 000 pictures from the Corel database. These were manually classified into 71 themes and used for evaluation purposes in [21]. This collection has been the largest found for which class information is available and can be obtained from the KDD-UCI repository (`http://kdd.ics.uci.edu/databases/CorelFeatures`), together with a set of 4 descriptors, namely: (a) a nine component vector with the mean, standard deviation and skewness for each hue, saturation and value in the HSV color space; (b) a 16 component vector with the second angular moment, the contrast, the inverse difference moment and the entropy for the co-ocurrence in the horizontal, vertical, and two diagonal directions; (c) a 32 component vector representing the 4 x 2 color HS histograms for each of the resulting sub-images after one horizontal and one vertical split; and (d) a 32 component vector with the 8 x 4 color HS histogram for the entire picture.

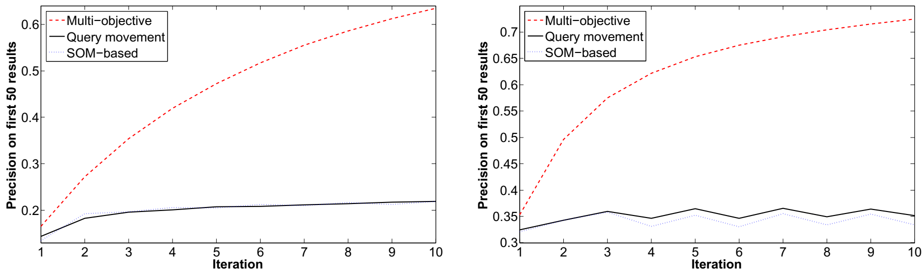
– A second smaller repository composed of 1 508 pictures, classified into a total of 29 categories. Some of these were extracted from the Web and some others were taken by the authors. The features used in this case were: (a) a 30 component vector with the 10×3 HS histogram (b) two 10 component vectors with the granulometries[22], calculated using a horizontal and a vertical segment as the structuring elements (each for a different feature vector).

In both repositories, (dis)similarity between features is estimated by using the histogram intersection on the color histogram vectors and the Euclidean distance on the rest. Results have been measured in terms of precision at a cutoff value, and precision vs recall curves, the most common methods to present results in the context of CBIR [1]. Precision is defined as the percentage of relevant images in the set of pictures retrieved, and it is usually expressed as a value in the range $[0, 1]$. Recall represents the percentage of the relevant images that are retrieved. When measuring precision at a cutoff value n, the precision is measured over the set composed of the first n images retrieved. We have chosen $n = 50$ as the area of interest. In table 1, the results for the multi-objective technique and those for the query movement algorithm and the SOM-based approaches are presented. To facilitate the comparison, this same data is also shown in Figure 1. The numbers shown are the average over a large number of searches. In particular, a total of 1 420 searches were performed on the first repository (20 queries for each class), and 1 022 on the second (50 per class, except for those classes containing less than 50 images). To diminish the possible variabilities introduced by the random selection of targets and by potentially unguided searches when no relevant images are selected, we have forced that there is at least one relevant sample in between the first 50 images in the initial order of pictures, and all techniques have been evaluated using the same list of targets and the same initial orderings.

Figure 2 shows the precision vs recall graphs at each iteration, properly scaled to the areas of interest, for the two databases considered. The two plots in figures 1 and 2 evidence the robustness of the method as a relevance feedback mechanism. In both repositories, the number of relevant results in between the first 50 retrievals significantly increases at each iteration. Unlike the query movement approach, the algorithm is able to maintain several concurrent search areas and discover new regions of interest as the search progresses.

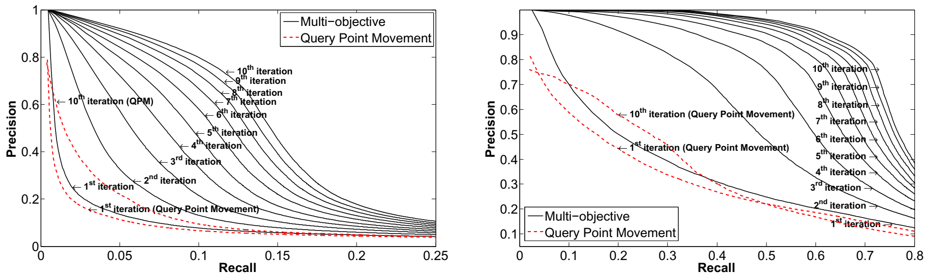
Table 1. Precision obtained at a cutoff value of 50 for each of the algorithms considered in the cases of the Large and Small repositories

	Algorithm	Iteration 1	2	3	4	5	6	7	8	9	10
Large	Multi-objective	0.1656	0.2716	0.3540	0.4193	0.4721	0.5174	0.5551	0.5861	0.6119	0.6345
	Query-movement	0.1434	0.1826	0.1958	0.2007	0.2074	0.2082	0.2118	0.2139	0.2171	0.2193
	SOM-based	0.1346	0.1928	0.1965	0.2055	0.2048	0.2125	0.2098	0.2167	0.2126	0.2190
Small	Multi-objective	0.3531	0.4963	0.5749	0.6216	0.6534	0.6754	0.6914	0.7046	0.7151	0.7244
	Query-movement	0.3247	0.3425	0.3596	0.3461	0.3646	0.3461	0.3656	0.3491	0.3643	0.3513
	SOM-based	0.3208	0.3428	0.3580	0.3308	0.3522	0.3307	0.3548	0.3339	0.3544	0.3341



(a) (b)

Fig. 1. Precision at a cutoff value of 50 for the three algorithms compared. (a) in the large repository; (b) in the small repository.



(a) (b)

Fig. 2. Precision vs recall graphs at each iteration. (a) in the large repository; (b) in the small repository.

4 Conclusions

A relevance feedback method based on formulating the CBIR problem from a multi-objective optimization perspective has been presented in this paper. The main advantage of the method is that it is able to simultaneously explore regions around the relevant selections and others which are in between them. This allows the method to recover images in regions that other methods would not explore. Results show that the method performs reasonably well on two manually classified repositories of different characteristics.

One major drawback of the technique is the relatively high computational time involved in the calculation of the pareto optimal sets. Despite that the number of iterations may be adjusted to keep the response time under reasonable limits, the query movement and SOM-based approaches are considerably faster. As an illustrative figure, and fixing the response time as 1 second for the method proposed, running time becomes two to three order of magnitude higher than for the other two methods in the comparison. The study of alternative methods to compute the pareto optimal front is still an issue under investigation, and may yield important improvements in retrieval performance and/or execution time. Currently, further work is directed towards a more detailed characterization of the way this method behaves in order to integrate this and other more powerful strategies into a combined scheme.

Acknowledgments. We would like to thank Dr. G. Giacinto for his help by facilitating the manually performed classification of the 30 000 images repository. This work has been partially funded by FEDER and Spanish Government through projects TIN2011-29221-C03-02, TIN2009-14205-C04-03, and Consolider Ingenio 2010 CSD07-00018.

References

1. Thomee, B., Lew, M.S.: Interactive search in image retrieval: a survey. International Journal of Multimedia Information Retrieval 1(2), 71–86 (2012)
2. Datta, R., Joshi, D., Li, J., Wang, J.Z.: Image retrieval: Ideas, influences, and trends of the new age. ACM Computing Surveys 40(2), 1–60 (2008)
3. Ishikawa, Y., Subramanya, R., Faloutsos, C.: Mindreader: Querying databases through multiple examples. In: Proc. 24th Int. Conf. Very Large Data Bases, VLDB, New York, USA, pp. 433–438 (1998)
4. Rui, Y., Huang, S., Ortega, M., Mehrotra, S.: Relevance feeback: a power tool for interactive content-based image retrieval. IEEE Transaction on Circuits and Video Technology 8(5), 644–655 (1998)
5. Ciocca, G., Schettini, R.: A relevance feedback mechanism for content-based image retrieval. Information Processing and Management 35(1), 605–632 (1999)
6. de Freitas, N., Brochu, E., Barnard, K., Duygulu, P., Forsyth, D.: Bayesian models for massive multimedia databases: a new frontier. Technical Report TR-2003-5, Department of Computer Science, University of British Columbia (2003)
7. Vasconcelos, N., Lippman, A.: Learning from user feedback in image retrieval systems. In: Proceedings of Advances in Neural Information Processing Systems (NIPS 1999), Denver, Colo, USA, pp. 977–986 (November-December 1999)
8. Arevalillo-Herráez, M., Ferri, F.J., Domingo, J.: A naive relevance feedback model for content-based image retrieval using multiple similarity measures. Pattern Recognition 43(3), 619–629 (2010)
9. Giacinto, G.: A nearest-neighbor approach to relevance feedback in content based image retrieval. In: Proceedings of the 6th ACM International Conference on Image and Video Retrieval (CIVR 2007), Amsterdam, The Netherlands, pp. 456–463. ACM Press (1993)
10. Arevalillo-Herraez, M., Ferri, F.J.: Interactive image retrieval using smoothed nearest neighbor estimates. In: Hancock, E.R., Wilson, R.C., Windeatt, T., Ulusoy, I., Escolano, F. (eds.) SSPR&SPR 2010. LNCS, vol. 6218, pp. 708–717. Springer, Heidelberg (2010)
11. Zhou, X.S., Huang, T.S.: Small sample learning during multimedia retrieval using biasmap. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 11–17 (2001)
12. Chen, Y., Zhou, X.S., Huang, T.S.: One-class svm for learning in image retrieval. In: Proceedings of the IEEE International Conference on Image Processing, pp. 34–37 (2001)
13. Tao, D., Tang, X., Li, X.: Which components are important for interactive image searching? IEEE Transactions on Circuits and Systems for Video Technology 18(1), 3–11 (2008)
14. Arevalillo-Herráez, M., Zacarés, M., Benavent, X., de Ves, E.: A relevance feedback CBIR algorithm based on fuzzy sets. Signal Processing: Image Communication 23(7), 490–504 (2008)

15. Koskela, M., Laaksonen, J., Oja, E.: Use of image subset features in image retrieval with self-organizing maps. In: Enser, P.G.B., Kompatsiaris, Y., O'Connor, N.E., Smeaton, A.F., Smeulders, A.W.M. (eds.) CIVR 2004. LNCS, vol. 3115, pp. 508–516. Springer, Heidelberg (2004)
16. Arevalillo-Herráez, M., Ferri, F.J., Moreno-Picot, S.: Distance-based relevance feedback using a hybrid interactive genetic algorithm for image retrieval. Applied Soft Computing 11(2), 1782–1791 (2011)
17. Deb, K., Pratap, A., Agarwal, S., Meyarivan, T.: A fast and elitist multiobjective genetic algorithm: Nsga. Transactions on Evolutionary Computation 6(2) (April 2002)
18. Pighetti, R., Pallez, D., Precioso, F.: Hybdrid content based image retrieval combining multi-objective interactive genetic algorithm and svm. In: 2012 21st International Conference on Pattern Recognition (ICPR), pp. 2849–2852 (2012)
19. Müller, H., Müller, W., Squirre, D.M.: Automated benchmarking in content based image retrieval. In: IEEE International Conference on Multimedia and Expo, pp. 321–324 (2001)
20. Laaksonen, J., Koskela, M., Oja, E.: Picsom-self-organizing image retrieval with mpeg-7 content descriptors. IEEE Transactions on Neural Networks 13(4), 841–853 (2002)
21. Giacinto, G., Roli, F.: Nearest-prototype relevance feedback for content based image retrieval. In: ICPR 2004: Proceedings of the 17th International Conference on Pattern Recognition, vol. 2, pp. 989–992. IEEE Computer Society, Washington, DC (2004)
22. Soille, P.: Morphological Image Analysis: Principles and Applications. Springer, Berlin (2003)