K-NEAREST NEIGHBORS DIRECTED SYNTHETIC IMAGES INJECTION

Luca Piras, Giorgio Giacinto

Dept. of Electrical and Electronic Eng.
University of Cagliari
09123 Piazza D'armi, Cagliari, Italy

ABSTRACT

It is widely acknowledged that good performances of content-based image retrieval systems can be attained by adopting relevance feedback mechanisms. One of the main difficulties in exploiting relevance information is the availability of few relevant images, as users typically label a few dozen of images, the majority of them often being non-relevant to user's needs. In order to boost the learning capabilities of relevance feedback techniques, this paper proposes the creation of points in the feature space which can be considered as representation of relevant images. The new points are generated taking into account not only the available relevant points in the feature space, but also the relative positions of non-relevant ones. This approach has been tested on a relevance feedback technique, based on the Nearest-Neighbor classification paradigm. Reported experiments show the effectiveness of the proposed technique relatively to precision and recall.

1. INTRODUCTION

Search engines are becoming increasingly popular in a large number of application domains. As far as the amount of information is available in digital form, there is an increasing need of smart tools that allows searching into multimedia collections. Such tools are aimed at retrieving data not only according to the associated metadata used to describe them, but also in term of their content [1]. Since it is very difficult to capture the complex semantics of an image, the vast majority of techniques for Content Based Image Retrieval (CBIR) relies on the representation of images by low-level features, e.g., color, texture, shape, etc. [2, 1]. As a consequence, content-based queries are typically expressed by visual examples in order to retrieve from the database all images that are similar to the examples.

As retrieval results may fit users' needs only partly, relevance feedback technique are employed to capture the subjectivity of them and to refine the search. The user is asked to label the images as being relevant or not, and the similarity measure is modified accordingly. Recently, relevance feedback has been formulated in terms of a classification problem [3]. This formulation requires a careful design, since the number of training samples is typically small (the user is asked to mark as being relevant or not a number of images in the order of few dozens), whereas the number of features used to represent image content can be large.

In this paper, we address the small sample-size problem by resorting to techniques known in the literature as *noise injection*. Noise injection has been proposed to regularize statistical classifiers when training data is affected by small sample-size problem. In particular, the effectiveness of these techniques has been studied w.r.t. neural network training, linear and density-based classifiers [4, 5]. Basically, new training patterns are generated *at random* by assuming Gaussian spherical noise around the available training patterns. This approach may be effective if we assume that all the features used to represent pattern are equally important, and that patterns of the class of interest can be distributed in all directions. However in a number of applications, patterns usually lie on a lower dimensional manifold of the original feature space. This is the case of image retrieval, where images are presented in very high dimensional feature spaces made up of low level features, while images of interest to the user may lie in a low-dimensional manifold [6]. As a consequence, the generation of artificial new patterns by random Gaussian spherical noise is prone to errors, as it may generate patterns outside the relevant manifold.

As a solution, we propose to create new random artificial patterns by exploiting nearest-neighbor information, and by constraining these patterns in a region of the feature space containing relevant images. The new patterns are created as a linear combination of nearest neighbors relevant images taking into account the low dimensional space where relevant images are assumed to lie. Recently, some papers on the content-based image retrieval addressed the small sample-size problem by proposing the generation of artificial patterns [7, 8]. While in this paper we propose the generation of new patterns in the feature space for improving the performances of learning mechanisms, other papers propose the generation of synthetic images for the sake of receiving informative feedback from the user. Our proposal, on the contrary, does not involve the interaction with the user, and it can be considered a "trick" to improve the learning process.

This paper is organized as follows. The proposed technique for creating random new patterns is illustrated in Section 2. Section 3 shows the integration of the proposed technique in the framework of relevance feedback mechanisms. Experimental results on an image dataset are reported in Section 4. Reported results show that the proposed method allows improving the performances of a relevance feedback mechanisms based on Nearest-neighbor. Conclusions are drawn in Section 5.

2. PATTERNS INJECTION

One of the most severe problems in exploiting relevance feedback in image retrieval is the small number of images that the user considers as being "relevant" compared to the number of non-relevant images. This behavior can be observed especially during the first iterations of relevance feedback. There are two main reasons that motivate such a behavior: i) even in the case of very "cooperative" users, it is not feasible to display more than a few dozens of images to label; ii) if the database at hand is very large, then the number of images that are relevant to the query can easily be very small compared to the size of the database. On the other hand, the number of features used to represent the image content may be very large, in order to "cover" with a large number of low-level feature, a wide spectrum of high-level concepts. From the point of view of statistical learning theory, a small number of training patterns represented in a high-dimensional feature space makes the learning process a difficult task [9, 10].

We propose to address this problem by artificially increasing the number of relevant patterns in the feature space. We will refer to the proposed technique as *K-Nearest Neighbors Directed Pattern Injection* (K-NN DPI). The basic idea underlying K-NN DPI consists in adding some artificial patterns generated in the feature space by taking into account the K relevant images nearest to a reference point in the feature space. The way used to generate those patterns aims to increase the generalization capability of the classifier, as the risk of classifier over-training on few patterns is mitigated. Let us denote with $X^{(i)}$ a pattern of a feature space F belonging to a class π_i. We will create a number of artificial relevant patterns $U^{(i)}$ related to $X^{(i)}$ as in [5]:

$$U^{(i)} = X^{(i)} + Z^{(i)}. \quad (1)$$

where the *directional vector* $Z^{(i)}$ is evaluated as:

$$Z^{(i)} = \lambda \times \frac{1}{K} \sum_{k=1}^{K} \xi_k \left(X^{(i)} - Q_k^{(i)} \right), \quad (2)$$

where $\xi_k \sim N(0, 1)$, λ is a scale factor and $Q_1^{(i)}, \ldots, Q_K^{(i)}$ are $X^{(i)}$'s K nearest patterns belonging to the same class π_i. It is easy to see that for each $X^{(i)}$ and its K nearest patterns $Q_1^{(i)}, \ldots, Q_K^{(i)}$ it is possible to generate an infinite number of directional vectors $Z^{(i)}$ by varying the coefficients $\xi_1, \ldots, \xi_K$. The value of K, i.e. the number of nearest-neighbors considered for the creation of artificial patterns, should not be large to avoid taking into account pattern that are actually far from the reference point. We fixed K to the value of 2 as this value was found to provide good performances in some preliminary experiments. Thus, new patterns are created according to a linear combination of directions depending on the distribution of K-nearest neighbors of a reference image. The coefficient of the linear combination are chosen randomly from a normal distribution. It is easy to see that the new patterns lie in the neighborhood of relevant images. On the other hand, the creation of artificial patterns by assuming that other relevant patterns may be normally distributed around each relevant pattern (a.k.a. spherical noise injection) may create spurious patterns. This problem can be avoided by suitably choosing the values of the parameters which the K-NN DPI technique depends on, i.e.: the image $X^{(i)}$ used as a reference, the number of nearest-neighbors considered, the number of directional vectors that are generated, and the scale factor λ.

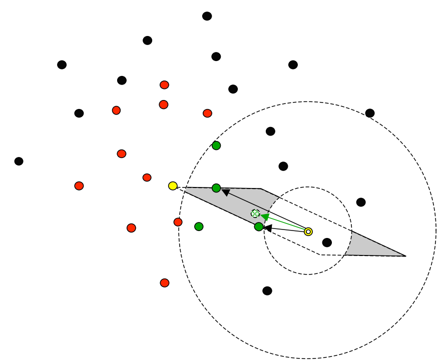

Fig. 1. Example of K-NN "synthetic" pattern generation in a two-dimensional feature space, where $K = 2$, $\xi_1 = \xi_2 = 0.5$ and $\lambda = 1$.

3. RELEVANCE FEEDBACK WITH K-NN DPI

The use of the K-NN DPI technique in the framework of any Relevance Feedback technique based on learning paradigms is very simple:

i) In the first iteration, L images are retrieved and they are labeled by the user as being relevant or not.

ii) If there the user cannot find any relevant image, the relevance feedback technique can be called again, without generating new patterns.

iii) If at least one image is labelled as relevant, and the non-relevant images outnumber the relevant images according to some threshold, the directional vectors are computed according to equation (2), and the related artificial patterns (equation (1)) are generated. These patterns are labelled as relevant, and added to the set of relevant images

iv) A new set of L images are retrieved from the image database, and are shown to the user and the retrieval procedure can be either stopped or it can be continued by labeling the new images as being relevant or not.

Let us examinate some choices we have investigated for the reference point and recall that the directions used to generate new patterns are those defined by the reference point and the related relevant nearest neighbors.

The reference point can be the **pattern associated to the query image**, the **mean vector of all the relevant images**, **each relevant image as a reference image**, and **a point computed according to some query movement approach**. The last option has been implemented by resorting to a query movement approach proposed by some of the authors, namely the Bayes Query Shifting approach (BQS) [11].

Let us now address the problem of the choice of the number of artificial patterns to be created. In fact, while in principle it is possible to generate an infinite number of patterns, there is the risk to overfit the new patterns. In addition, a large number of artificial relevant patterns w.r.t. the number of non-relevant pattern in the region of the space explored by the search, may distort the actual distribution of relevant and non-relevant images in that region. For these reasons, we constrained the generation of new patterns so that at each iteration the ratio between the relevant and non-relevant images was constantly of $1 : n$, with $n = 2, \ldots, 5$. We furthermore propose to constrain the creation of new patterns so that artificial pattern are considered only if they are between the nearest and the farthest relevant image w.r.t. the reference image used in equation (1) to generate new patterns.

Figure 1 shows an example of how the proposed technique works, where the yellow dot represents the query, while the red, green, and black dots respectively represent non-relevant, relevant and unlabeled images. The yellow ring represents the point computed according to Q_{BQS}. The green dots pointed by the black arrows are the two relevant images nearest to Q_{BQS}, and the parallelogram delimitates the grey area where artificial patterns can be created. The two circumferences drawn with broken lines bound the area between the nearest and the farthest relevant images. The green dot drawn with a broken line and pointed by the green arrow is the injected "synthetic" pattern.

4. EXPERIMENTAL RESULTS

In this section we describe the experimental results that have been obtained applying the technique proposed in Sections 2 and 3. The Caltech-256 dataset obtained from the California Institute of Technology repository[1] has been used. It consists of 30607 images subdivided into 257 semantic classes [12].The *Tamura* representation has been used as it is one of the most common feature vectors used to represent images. Each image is represented by a 18-dimensional feature vector [13]. In order to test the performances of the proposed mechanism, 500 images have been randomly extracted from all of the 257 classes, and used as query.

The images are scored using a relevance feedback techniques: the 2^{nd}-Nearest Neighbor (2NN) [14]. The 2NN technique assigns a relevance score to each image according to its distances from the second nearest relevant image, and the second nearest non-relevant image. This score is further combined to a score related to the distance of the image from the region of relevant images, the coefficient of the combination being dynamically computed according to the proportion of relevant images available w.r.t. the number of non-relevant images. Performances are evaluated in terms of retrieval precision, recall and the F-measure = $((2 \cdot prec)^{-1} + (2 \cdot recall)^{-1})^{-1}$ [15]. Experiments have been performed using different values of the parameters discussed in Section 3.

Reported experiments are aimed to show the performances of artificial pattern generation with different choices of the reference point. We used the normalization parameter $\lambda = \left(\xi_1^2 + \xi_2^2\right)^{-2}$, and fixed the ratio between relevant and non-relevant patterns to $1 : 2$.

Reference points used in the experiments are related to the different choices discussed in Section 3. In particular, we used the following notation: **Query - DPI** (where the reference point is the original query point submitted by the user), **Mr - DPI** (where the reference point is the mean vector in the feature space of all the relevant images), **Ar - DPI** (where each relevant images generates a fraction of the artificial patterns that needs to be generated), **BQS - DPI** (where the reference point is computed according to the BQS query movement technique).

Retrieval performances attained using the proposed "pattern injection" technique have also been compared to those attained without "patterns injection". The precision of the 2NN relevance feedback with DPI is shown in Figure 2. It is easy to see that any choice of the reference pattern allows the DPI to improve the precision w.r.t. the 2NN technique without DPI. In particular the method that uses as reference point the mean vector of all relevant images (Mr-DPI) has an improvement of nearly 8%. If the recall and F-measure are taken into account (Figures 3 and 4), it can be easily seen that the differences between the "DPI techniques" are not so big. It can be argued that Mr-DPI allow retrieving at each iteration a higher number of relevant images, but not many of them have never been seen in previous iterations, as a consequence, the additional patterns created are close to known relevant images, and they are always ranked among the first. Nevertheless even if Mr-DPI does not have the highest value of recall improve both the recall and F-measure w.r.t. most of DPI methods and the 2NN technique without pattern injection.

5. CONCLUSIONS

In this paper we have addressed the small sample-size problem that is encountered when relevance feedback mechanisms are employed. In order to improve the learning process, we proposed to generate artificial patterns in the feature space by exploiting nearest-neighbor relationships between relevant images. Reported results showed that the proposed technique allows improving the performances of a relevance feedback mechanism based on Nearest-Neighbors.

[1] http://www.vision.caltech.edu/Image_Datasets/Caltech256/

6. REFERENCES

[1] M. S. Lew, N. Sebe, C. Djeraba, and R. Jain, "Content-based multimedia information retrieval: State of the art and challenges," *ACM Trans. Multimedia Comput. Commun. Appl.*, vol. 2, no. 1, pp. 1–19, 2006.

[2] A. Del Bimbo, *Visual information retrieval*, Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 1999.

[3] R. Datta, D. Joshi, J. Li, and J. Z. Wang, "Image retrieval: Ideas, influences, and trends of the new age," *ACM Computing Surveys*, vol. 40, no. 2, pp. 1–60, 2008.

[4] C. M. Bishop, *Neural Networks for Pattern Recognition*, Oxford University Press, Inc., New York, NY, USA, 1995.

[5] M. Skurichina, S. Raudys, and R. P. W. Duin, "K-nearest neighbors directed noise injection in multilayer perceptron training," *IEEE Trans. on Neural Networks*, vol. 11, no. 2, pp. 504–511, March 2000.

[6] Y. Lin, T. Liu, and H. Chen, "Semantic manifold learning for image retrieval," in *MULTIMEDIA '05: Proceedings of the 13th annual ACM international conference on Multimedia*, New York, NY, USA, 2005, pp. 249–258, ACM.

[7] B. Thomee, M. J. Huiskes, and E. Bakkerand M. S. Lew, "Visual information retrieval using synthesized imagery," in *CIVR '07: Proc. of the 6th ACM int. conf. on Im. and video retrieval*, New York, NY, USA, 2007, pp. 127–130, ACM.

[8] B. Thomee, M.J. Huiskes, E. Bakker, and M.S. Lew, "Using an artificial imagination for texture retrieval," in *Pattern Recognition, 2008. ICPR 2008.*, Dec. 2008, pp. 1–4.

[9] X. Zhou and Thomas S. Huang, "Small sample learning during multimedia retrieval using biasmap," in *Proc. IEEE Intl Conf. CVPR*, 2001, vol. 1, pp. 11–17.

[10] R. P. W. Duin, "Pattern recognition in almost empty spaces," Eindhoven, Netherlands, Jan. 2004.

[11] G. Giacinto and F. Roli, "Bayesian relevance feedback for content-based image retrieval," *Pattern Recognition*, vol. 37, pp. 1499–1508, 2004.

[12] G. Griffin, A. Holub, and P. Perona, "Caltech-256 object category dataset," Tech. Rep. 7694, California Institute of Technology, 2007.

[13] H. Tamura, S. Mori, and T. Yamawaki, "Textural features corresponding to visual perception," *Systems, Man and Cybernetics, IEEE Trans. on*, vol. 8, no. 6, pp. 460–473, June 1978.

[14] G. Giacinto, "A nearest-neighbor approach to relevance feedback in content based image retrieval," in *CIVR '07: Proc. of the 6th ACM int. conf. on Image and video retrieval*, New York, NY, USA, 2007, pp. 456–463, ACM.

[15] P. H. Dionysius and S. Nicu, "How to complete performance graphs in content-based image retrieval: Add generality and normalize scope," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 2, pp. 245–251, 2005.

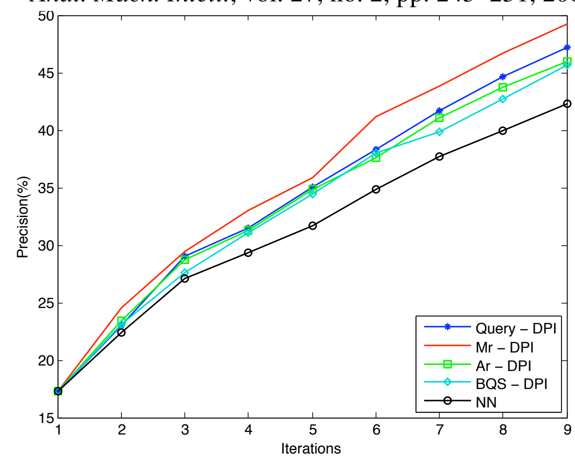

Fig. 2. Precision - Tamura

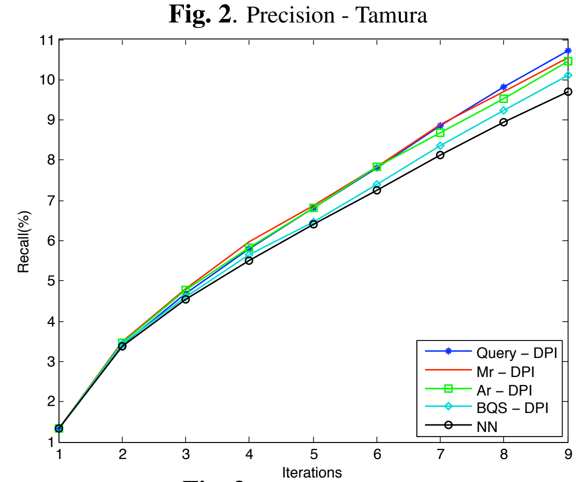

Fig. 3. Recall - Tamura

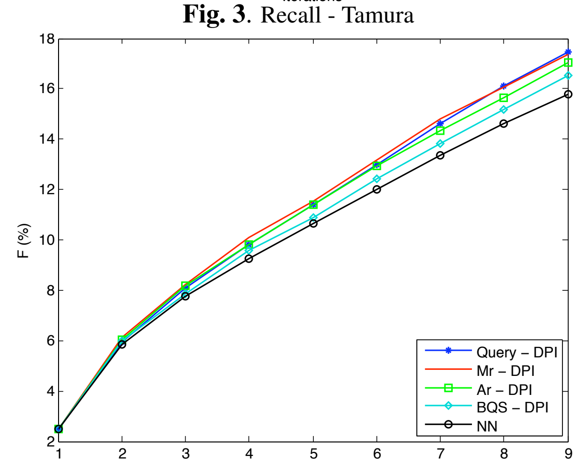

Fig. 4. F-measure - Tamura