Image and Vision Computing 31 (2013) 704–713

Contents lists available at ScienceDirect

Image and Vision Computing

journal homepage: www.elsevier.com/locate/imavis

An improved distance-based relevance feedback strategy for image retrieval☆

Miguel Arevalillo-Herráez *, Francesc J. Ferri

Department of Computer Science, University of Valencia, Avda. de la Universidad s/n, 46100 Burjasot, Spain

ARTICLE INFO

Article history:
Received 18 December 2012
Received in revised form 13 May 2013
Accepted 18 July 2013

Keywords:
CBIR
Image retrieval
Relevance feedback
Nearest neighbor

ABSTRACT

Most CBIR (content based image retrieval) systems use relevance feedback as a mechanism to improve retrieval results. NN (nearest neighbor) approaches provide an efficient method to compute relevance scores, by using estimated densities of relevant and non-relevant samples in a particular feature space. In this paper, particularities of the CBIR problem are exploited to propose an improved relevance feedback algorithm based on the NN approach. The resulting method has been tested in a number of different situations and compared to the standard NN approach and other existing relevance feedback mechanisms. Experimental results evidence significant improvements in most cases.

1. Introduction

Content based image retrieval (CBIR) refers to the application of techniques to retrieve digital images from large databases, by analyzing the actual content of the image rather than the metadata associated with it.

In general, a CBIR system represents each image in the repository as a set of features (usually related to color, texture and shape), and uses a set of distance functions defined over this feature space to estimate similarity between pictures.

A query can be understood as the intention of a user to retrieve a certain kind of images, and it is usually materialized as one or more sample pictures. The goal of a CBIR system is to retrieve a set of images that is best suited to the user's intention. Obviously, the potential results of such a system will strongly depend not only on the particular features of the representation space but also on the implicit or explicit distance functions used to measure similarity between pictures [1–3].

This way of assessing similarity comes along with the implicit assumption that image resemblance is related to a distance defined over a particular feature space. This leads to the so-called semantic gap, between the semantics induced from the low level features and the real high level meaningful user interpretation of the image. To reduce this gap, relevance feedback has been adopted by most recent CBIR systems [4]. When relevance feedback is used, the search is considered an iterative process in which the original query is refined interactively, to progressively obtain a more accurate result. At each iteration, the system retrieves a series of images according to a predefined similarity measure, and requires user interaction to mark the relevant and non-relevant retrievals. This data is used to modify some system parameters and produce a new set of results, repeating the process until a satisfying enough result is obtained. In this context, the relationship between any image in the database and the user's desire is usually expressed in terms of a relevance value. This value is aimed at directly reflecting the interest that the user may have in the image and is to be refined at each iteration.

Most relevance feedback algorithms use the user's selection to search for global properties which are shared by the relevant samples available at each iteration [4]. From a Pattern Recognition viewpoint, this can be seen as obtaining an appropriate estimate of the probability of (subjective) relevance. Many different approaches exist to model and progressively refine these estimates. But relevance feedback faces a small sample problem whose models cannot be reliably established because of the semantic gap. In this context, nonparametric distance-based methods using neighbors are particularly appealing [5–8]. The aim of these methods is to assess relevance of a given image by using distances to relevant and non-relevant neighbors. In particular, an image is considered as much relevant as its distance from the nearest relevant image is small compared to the distance of its nearest non-relevant image.

In the present work, all these considerations about distance-based CBIR approaches are taken into account to derive a novel way of reliably estimating relevance from distances. The algorithm is then evaluated exhaustively in three databases and in a variety of contexts, including both query by example and refinement of a textual search.

The paper is organized as follows. The next section explains the model used; outlines the assumptions made; presents the name

☆ This paper has been recommended for acceptance by Nicu Sebe.
* Corresponding author. Tel.: +34 96 3543962.
E-mail addresses: miguel.arevalillo@uv.es (M. Arevalillo-Herráez), francesc.ferri@uv.es (F.J. Ferri).

0262-8856/$ – see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.imavis.2013.07.004

conventions used throughout the article; and describes the plain nearest neighbor approach. Section 6 exposes several key facts about the way relevance is estimated and introduces a novel alternative. In the experimental section the proposed algorithm is compared against the original one [5,6], some extensions [7,8], and other representative relevance feedback methods. Finally, the main conclusions of the proposed approach are outlined along with some work in progress.

2. Related work

Relevance feedback in CBIR has been an active topic of research for the last two decades. In general, the performance of CBIR algorithms depends critically on the (dis)similarity measure used to rank the images in the repository. This measure is commonly built/adapted at each iteration, by using the information made available by the user. In this section, we summarize some previous work in relevance feedback, with the intention of contextualizing the method presented in the paper. For a more comprehensive survey, the reader is referred to recent reviews on the topic [1,9].

First approaches were aimed at progressively adapting the similarity measure and/or move the query point so that it becomes closer to positive samples and farther from negative ones. Query point movement and axis re-weighting methods fall within this category of techniques [10–12]. In general, these approaches model the query as a point in a (possibly deformed) feature space, and retrieve results according to their distance to the query. A major advantage of these techniques is that they are relatively fast, and scale reasonably well with the size of the repository. On the negative side, they usually ignore dependencies between features [13], treat features globally [14] and are only effective if the query concept consists of a convex region in the feature space.

A different way to approach the definition of an adequate similarity function is from a pattern recognition perspective. Relevance feedback can be considered as a classical machine learning problem, in which the user feedback is used as an input to a learning algorithm to address the classification of images as relevant images [9]. This opens the scope for the application of a large diversity of popular algorithms in this context. For example, labeled samples can be used to build a projection into a subspace of a lower dimensionality where relevant samples appear close to each other [15,16]; or to learn a Mahalanobis metric based on pairwise (dis)similarity constraints [17] or small subsets of points that are known to belong to the same class [18]. Closely related, support vector machines (SVM) methods attempt to find the hyperplane which achieves a maximum separation between two classes [19–22]. One class, two classes and other extensions have been adapted to overcome some of the inherent limitations of standard SVMs, e.g., imbalanced training set, high computational burden [9]. A major problem with some of these methods refers to the high sensibility of the parameters required for fine tuning the algorithms [7]. Other related approaches include the use of neural networks, e.g., radial basic functions (RBF) networks [23], self-organizing maps (SOM) [14], fuzzy sets [24] or regression methods [13]. Despite the many efforts performed in this direction, many of the algorithms suffer from the small sample size problem, caused by the relatively scarce information provided at each relevance feedback iteration.

Other strategies also include Bayesian approaches. In this case, posterior probability distributions are estimated according to the data gathered through the relevance feedback process. In particular, the probability densities of the relevant and non-relevant classes are usually approximated by using different types of estimators [25–29]. Then, the probability of being relevant is used as a similarity measure, as in the case of using a soft classifier. In this way, relevance values are implicitly modeled as a probability distribution, rather than as a single point in the feature space. Nearest neighbor methods can be classified in this category, and used in this context to estimate the posterior probabilities of the relevant and non-relevant classes. In addition, they are compatible with other distance metric learning approaches and can also be used as a framework to determine relevance by using other distances learned from the user feedback. These methods have previously been applied in the CBIR field [5–8], showing a good comparative performance superior to other Bayesian frameworks [5] and classical SVM techniques [7]. In this paper, we build on some of these previous works, by proposing series of strategies to face some fundamental shortcomings of NN-based approaches.

3. Interactive content-based image retrieval

3.1. Query representation

Assume that there is a repository of images $\mathcal{X}=\{x_1,\cdots,x_m\}$ conveniently represented in a metric feature space, $\mathcal{F}$, whose associate distance measure is $d:\mathcal{F}ñ\mathcal{F}\rightarrow\mathbb{R}^{\geq 0}$. This particular representation space is assumed to be the D-dimensional space $\mathbb{R}^D$ in this work as in much other closely related works. In some cases and specially in the image retrieval context, the representation space may embrace multiple low level descriptors (e.g., color, texture or shape) and the distance d is constructed by a combination from simpler distance measures over each descriptor [1].

When a particular user is interested in retrieving images from $\mathcal{X}$, his/her intention can be thought of as a semantic concept which can be more or less objectively specified (as e.g., images of bicycles, domestic animals, etc.). Regardless of the scope and specificity of this semantic concept, it can be modeled in the feature space as a probability function over the corresponding repository, $P(relevant|x)$, which can be extended to the whole representation space.

This probability of relevance over the space can be conveniently simplified and tackled. For example, single point query approaches assume that this probability function can be appropriately represented by a single (ideal) point $c\in\mathcal{F}$ possibly along with a convenient axis or feature re-weighting [30]. This can be seen as equivalent to considering uncorrelated Gaussian distribution functions. This approach can be extended to use a set of representative points or mixtures of Gaussians instead [31].

Single point methods use a distance measure to rank images. Multiple point methods usually combine distances (or rankings) to each representative in a set of (ideal) points $\mathcal{C}$. In any case, all methods end up using a particular ranking as a final tool to retrieve images regardless of the way they model the query in the feature space. Fig. 1 graphically illustrates this situation.

3.2. Relevance feedback

In this paper we assume the most usual case in relevance feedback in which the user marks or labels some images as relevant or non-relevant. In this setup, the available information or user feedback is given by the

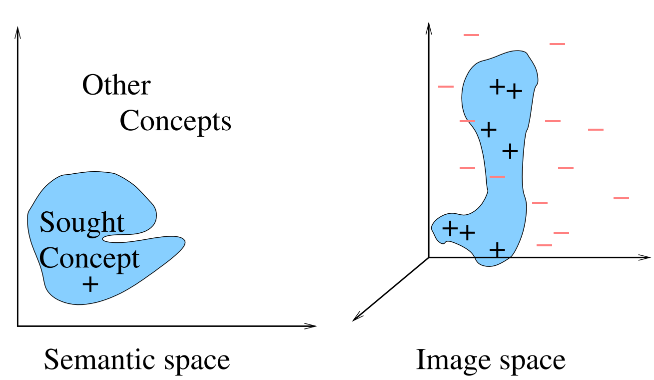

Fig. 1. An illustrative example of query representations. User intentions (semantic space) get translated into particular regions in the representation space where relevant images (marked as +) are more likely.