Egyptian Informatics Journal (2016) xxx, xxx–xxx

Cairo University

Egyptian Informatics Journal

www.elsevier.com/locate/eij
www.sciencedirect.com

FULL-LENGTH ARTICLE

Distance selection based on relevance feedback in the context of CBIR using the SFS meta-heuristic with one round

Mawloud Mosbah *, **Bachir Boucheham**

University 20 Août 1955 of Skikda, Algeria

Received 20 August 2015; revised 24 June 2016; accepted 1 September 2016

KEYWORDS

CBIR;
Distance-selection;
Relevance feedback;
SFS meta-heuristic;
Precision and Recall

Abstract In this paper, we address the selection in the context of Content Based-Image Retrieval (CBIR). Instead of addressing *features' selection* issue, we deal here with *distance selection* as a novel paradigm poorly addressed within CBIR field. Whereas distance concept is a very precise and sharp mathematical tool, we extend the study to weak distances: Similarity, quasi-distance, and divergence. Therefore, as many as eighteen (18) such measures as considered: distances: {Euclidian, ...}, similarities{Ruzika, ...}, quasi-distances: {Neyman-X^2, ...} and divergences: {Jeffrey, ...}. We specifically propose a hybrid system based on the Sequential Forward Selector (SFS) meta-heuristic with one round and relevance feedback. The experiments conducted on the *Wang database* (*Corel-1K*) using color moments as a signature show that our system yields promising results in terms of effectiveness.

1. Introduction

As any information retrieval system, a Content Based-Image Retrieval (CBIR) system aims at satisfying the user need through extracting, from the image database, a subset of images deemed as similar to the submitted query, let alone relevant to the user expectations. For doing so, a CBIR system utilizes some low-level features such as color, e.g. [1], texture, e.g. [2,3] and shape, e.g. [4]. A comparative study of some CBIR works is reported in [5]. Unfortunately, users are still usually unsatisfied with results answered by actual CBIR systems, owing to the semantic gap problem. Indeed, there is a gap between the relevance notion from the user viewpoint and the automatic relevance of the system. For improving results given by a CBIR system, one must, therefore then, reduce the gap between the two previous cited kinds of relevance. The relevance from the user perspective is related to what he/she has in his/her mind about his/her needs, whereas relevance from the system viewpoint is related to the query.

* Corresponding author.
E-mail addresses: mos_nasa@hotmail.fr, m.mosbah@univ-skikda.dz (M. Mosbah).
Peer review under responsibility of Faculty of Computers and Information, Cairo University.

Production and hosting by Elsevier

http://dx.doi.org/10.1016/j.eij.2016.09.001
1110-8665 © 2016 Production and hosting by Elsevier B.V. on behalf of Faculty of Computers and Information, Cairo University.

Please cite this article in press as: Mosbah M, Boucheham B, Distance selection based on relevance feedback in the context of CBIR using the SFS meta-heuristic with one round, Egyptian Informatics J (2016), http://dx.doi.org/10.1016/j.eij.2016.09.001

Among suggested solutions to the semantic gap, some authors used multiple query techniques, e.g. [6].

A basic key affecting the system relevance, and in consequence its accuracy, is the matching measure to work with. Review of literature shows that there are many matching measures ranging from distances and similarities to quasi-distances and divergences. To the best of our knowledge, few works have addressed the matching measure as a point of interest in the context of CBIR, e.g. [7–10]. The question to ask then is what matching measure should one use when building a CBIR system. Similarly, what matching measure should be used with respect to a specific query? This question leads consequently to the legitimate issue of *matching measure selection*.

A natural answer to the question of matching measure selection is the learning process. Review of literature shows that there are two manners for implementing the matching measure selection in the CBIR field: utilizing selection problem tools or learning through relevance feedback.

The proposed work falls within both aforementioned areas of CBIR field: selection paradigm and relevance feedback. These notions are explained in the following subsections.

1.1. Selection paradigm

To the best of our knowledge, the selection paradigm in the CBIR field has been so far restricted to features selection aspect only, e.g. [11]. Indeed, many authors have asked the following question: "which features are most suitable for a specific query?". Features selection methods search for the most relevant feature subset, belonging to the original feature space, according to a user defined criterion [12]. Features selection algorithms aim at choosing a reduced number of features that preserve the most relevant information of the dataset. Features selection is usually applied as a preprocessing step in data mining tasks by removing irrelevant or redundant features leading to more efficient and accurate classification, clustering and similarity searching processes [12]. There are three broad classes of features selection: filter methods, e.g. [13], wrapper methods, e.g. [14], and hybrid methods. Filter methods use general characteristics of the data independently of the classifier for the evaluation process. The evaluation process is classifier-dependent in wrapper methods. Finally, hybrid models use both filtering and wrapping methods for improving the performance of the selection process.

The problem with the selection tools is that the learning stage is expensive in terms of computing time. Therefore, it is done offline. In addition to that, tools, utilized in the learning stage, require evaluation according to a fitness measure. This poses other questions about the processed dataset devoted to learning. The retrieval problem, in the case of systems based on feature selection, can, therefore, be viewed as a classification problem. Evidently, in this case, the learning stage is crucial.

In this paper, we address *the matching measure selection paradigm* rather than *the feature selection paradigm*. More specifically, we aim to select, for each query, one matching measure that would be the best for a given query from the perspective of effectiveness. For doing so, we utilize the *Sequential Forward Selector* (SFS) algorithm [15,16] with one round. This choice has been motivated by the characteristic of the SFS-One-Round algorithm of being very efficient. In the following point, we explain briefly the SFS algorithm.

1.1.1. SFS algorithm

In this work, we use the SFS algorithm rather than other meta-heuristic algorithms such as Genetics Algorithm and Cuckoo Search Algorithm (CSA) owing to its simplicity. Other meta-heuristics than the SFS algorithm are of course of great interest as subject of study. Indeed, review of literature reveals that there exist many meta-heuristic algorithms applied in a variety of fields, e.g. [17–20]. However, choosing the best meta-heuristic algorithm for selecting the adequate matching measure goes beyond the scope of this paper.

Because we do not want to combine matching measures, we believe that one round is enough to answer the question: "which matching measure is the best?". The pseudo code of the SFS algorithm is presented in Fig. 1. In this pseudo code, the fitness value has trade-off with the sum of the ranks of images labeled as relevant by the user. This fitness is given by the following equation:

$$fitness = 1/n * \sum_{i=1}^{n} rank(i) \tag{1}$$

where n is the number of images labeled as relevant by the user.

The SFS algorithm with one round uses the pre-cited pseudo code from Step 1 to Step 4.

1.2. Relevance feedback

The relevance feedback concept, coming from documentary information retrieval [21,22], has received, in last few years, a lot of attention in the CBIR field, e.g. [23]. This scheme consists of receiving additional information from the user after visualizing the initial results. This additional information is simply the judgment of some visualized results by the user as relevant or non-relevant to his/her requirement. According to this judgment, the system proceeds to adjust its processing behavior for improving performances. The relevance feedback mechanism then is an additional tool for reducing the angle between the user relevance and system relevance by giving a

SFS Algorithm
Step 1: as initialization the algorithm starts with the following weighting (0, 0,..0) (no selected matching measure).
Step 2: each weight will be set 1 separately to generate many configurations.
Step 3: to evaluate each configuration based on the fitness.
Step 4: selecting the best configuration.
Step 5: comparing the actual selected configuration with the selected configuration of the previous iteration, if there is no improvement so go to the Step 8.
Step 6: set the other weights 0 except the weight of the selected matching measures is still 1.
Step 7: go to the Step 2.
Step 8: END.

Figure 1 Pseudo code of the SFS algorithm.