Knowledge-Based Systems 73 (2015) 254–264

Contents lists available at ScienceDirect

Knowledge-Based Systems

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

A graph-based relevance feedback mechanism in content-based image retrieval ☆

Malay Kumar Kundu [a], Manish Chowdhury [a,*], Samuel Rota Bulò [b]

[a] Machine Intelligence Unit, Indian Statistical Institute, 203 B.T. Road, Kolkata 108, India
[b] FBK-irst, Via Sommarive, 18, I-38123 Trento, Italy

ARTICLE INFO

Article history:
Received 27 March 2014
Received in revised form 18 July 2014
Accepted 9 October 2014
Available online 16 October 2014

Keywords:
NSCT
Content based image retrieval
Re-ranking
Relevance feedback
Feature evaluation index

ABSTRACT

Content-Based Image Retrieval (CBIR) is an important problem in the domain of digital data management. There is indeed a growing availability of images, but unfortunately the traditional metadata-based search systems are unable to properly exploit their visual information content. In this article we introduce a novel CBIR scheme that abstracts each image in the database in terms of statistical features computed using the Multi-scale Geometric Analysis (MGA) of Non-subsampled Contourlet Transform (NSCT). Noise resilience is one of the main advantages of this feature representation. To improve the retrieval performance and reduce the semantic gap, our system incorporates a Relevance Feedback (RF) mechanism that uses a graph-theoretic approach to rank the images in accordance with the user's feedback. First, a graph of images is constructed with edges reflecting the similarity of pairs of images with respect to the proposed feature representation. Then, images are ranked at each feedback round in terms of the probability that a random walk on this graph reaches an image tagged as relevant by the user before hitting a non-relevant one. Experimental analyses on three different databases show the effectiveness of our algorithm compared to state-of-the-art approaches in particular when the images are corrupted with different types of noise.

1. Introduction

A Content Based Image Retrieval (CBIR) system enables a user to organize and retrieve images in a database by analyzing the characteristics of the visual content. The whole process is usually done by presenting a visual query to the system and by extracting a set of images from the database that have highest resemblance to the query image [1–3]. This query-by-example procedure compares the visual content of images in terms of low level features by computing a distance between the features of the query image and the possible target images in the database [4–6].

A modern interactive CBIR system consists of the following main parts: feature extraction, feature reduction, ranking and relevance feedback. The first two phases allow to obtain abstract, compact representations for the query and database images, which possibly summarize their most distinctive features. The ranking phase consists in sorting the database images based on their relevancy to the query image. Finally, the relevance feedback phase involves the user intervention to tag the images in the result set as relevant or irrelevant. This triggers a re-ranking of the database images which accounts for the new feedback information. Multiple feedback rounds can follow until user satisfaction is achieved.

Various feature extraction and feature reduction schemes have been used in the literature to find the low-dimensional salient and significant features, which can be effectively used to represent the underlying image's characteristics [7–9]. It has been found that feature extraction techniques working in the frequency domain are more effective in representing the significant and subtle details of the image than the conventional spatial domain schemes [10,11]. Among the various frequency domain methods, Wavelet Transform (WT) and its variants (like M-band wavelet, complex wavelet, wavelet packets etc.) have been extensively used in CBIR systems [12,13,10,14]. Low level features based on WT provide a unique representation of the image and they are highly suitable to characterize textures of the image [12,15,16]. However, the main problem of WT-based features is the inherent lack of support to directionality and anisotropy. To overcome these limitations, a recent theory called Multi-scale Geometric Analysis (MGA) for high-dimensional signals has been introduced and several MGA tools have been developed like Ripplet, Curvelet and Contourlet,

☆ This work is supported by internal academic project fund of Machine Intelligence Unit, Indian Statistical Institute, 203 B.T. Road, Kolkata 108.
* Corresponding author. Tel.: +91 33 2575 3100; fax: +91 33 2578 3357.
E-mail addresses: malay@isical.ac.in (M.K. Kundu), st.manishc@gmail.com (M. Chowdhury), rotabulo@fbk.eu (S. Rota Bulò).

http://dx.doi.org/10.1016/j.knosys.2014.10.009
0950-7051/© 2014 Elsevier B.V. All rights reserved.

with application to different problem domains [17,18]. In general, CBIR systems based on these MGA tools turn out to be more effective than WT-based traditional CBIR schemes [18,19]. An improvement of the Contourlet Transform (CNT) has been proposed in [20] to mitigate the shift-sensitivity and the aliasing problems of CNT in both space and frequency domains. Their solution known as Non-Subsampled Contourlet Transform (NSCT) combines both Non-Subsampled Pyramid (NSP) and Non-Subsampled Directional Filter Bank (NSDFB). In the literature, different NSCT-based CBIR systems have appeared [21–24].

Besides extracting good feature representations, a further important task is feature selection, which aims to seek optimal subsets of the extracted features that preserve most of the information carried by the collected data [25]. The main goal is to facilitate future analysis in the presence of high-dimensional data by improving the query performance of the CBIR system and by reducing the storage requirement. In addition, the reduction in dimensionality favors de-noising, for noise is typically concentrated in the excluded dimensions [26]. Conventional methods of feature selection involve evaluating different feature subsets using some index and selecting the best among them [27]. We refer to [28–32] for a detailed discussion about various unsupervised feature selection algorithms. Some approaches exploit also supervision by employing a labeled training set and mutual-information-based criteria to find class-specific relevant and non-redundant features (see, *e.g.*, [33]).

Unfortunately, low level features and distance metrics are not sufficient to reduce the semantic gap and thus rank the images according to the user's intentions. This motivates researchers to insert the user's feedback in the search loop and improve the ranking performance of the retrieved images by employing an interactive scheme [34]. Since mid-1990s, relevance feedback mechanisms have been adopted to retrieve images by exploiting the human visual impression as a feedback signal that is used to iteratively correct errors made by the CBIR system [35,9]. Such feedback process terminates when the user is satisfied with the retrieved images. The measure of distance between images is in general a function of the user, which encompasses her experience over the period of time from her early age. The user intervention becomes thus necessary if we aim at devising accurate, user-specific retrieval systems [36–38].

There are two main types of RF-based approaches for CBIR in the literature [39]: *inductive* and *transductive*. Inductive approaches use a classifier trained in a supervised way to discriminate between relevant and irrelevant images [40,41] and rank the images based on their relevancy. The major drawback of these methods is the limited number of examples that are actually labeled by the user, which prevents the classifiers from properly learning the true relevant/irrelevant separation boundaries. Transductive approaches mitigate this problem by exploiting also the distribution of the unlabeled data. Those approaches are typically based on manifold learning to propagate a ranking score or the class-posterior on the manifold of images [39,42–44]. Other approaches define a generative model that uses the unlabeled data to measure the relevance between query image and database images [45]. We refer to [9,46] for further references to RF-based CBIR schemes.

In this article, we propose a new CBIR system based on RF. Our system exploits feature representations for the images given in terms of first-order statistics computed from NSCT. This approach indeed guarantees a better preservation of the main cues of the images as NSCT is a flexible multi-scale, multi-directional and shift-invariant image decomposition method. After the feature extraction phase, an unsupervised approach based on the Maximal Information Compression Index (MICI) is adopted to select a subset of optimal features, which reduces the dimensionality of the data and implicitly suppresses part of the noise [28]. We then employ a transductive, graph-based ranking approach that exploits the RF information. It relies on a sparse graph representation in which the database images and the query image are nodes and edge weights are expressed in terms of the Euclidean distance between the image's feature abstractions. Following [42], ranking is initially carried out by a simple k-nearest-neighbor approach, while the subsequent rankings, which account for the user's feedback, are given according to the probability that a random walk starting from a node in the graph will reach a relevant image before hitting a non-relevant one. Extensive experiments using the proposed NSCT-based features and this graph theoretic RF mechanism on three different databases show the superiority of our proposed method over different recent approaches.

The rest of the manuscript is organized as follows: Section 2 introduces the non-subsampled contourlet transform, which will be adopted in the proposed approach. The detailed description of the feature representation used and the graph-theoretic RF mechanism are explained in Sections 3 and 4, respectively. The proposed CBIR system is described in Section 5. Section 6 is devoted to the experimental evaluation. Finally, in Section 7 we draw conclusions and discuss some future extensions of the work.

2. Non-Subsampled Contourlet Transform (NSCT)

In this section, we briefly describe the non-subsampled contourlet transform, which will be adopted in our system to devise a proper image representation.

NSCT is a fully shift-invariant, multi-scale, and multi-direction expansion with fast implementability [20]. As opposed to the contourlet transform, which is not shift-invariant due to the presence of down-samplers and up-samplers in both the Laplacian pyramid and Directional Filter Bank (DFB) stages, NSCT achieves the shift-invariance property by using non-subsampled pyramid filter banks and non-subsampled DFB.

2.1. Non-Subsampled Pyramid (NSP) filter bank

NSP is a shift-invariant filtering structure that leads to a sub-band decomposition that resembles the Laplacian pyramid, which ensures the multi-scale property of the NSCT. As shown in Fig. 1, it is constructed by using two-channel non-subsampled 2D filter banks, which produce a low-frequency and a high-frequency image at each NSP decomposition level. Filters at subsequent stages are obtained by upsampling the low-pass filters at the first stage. As a result, NSP can result in $\mathsf{k}+1$ sub-images, which consist of one low-frequency image and k high-frequency images whose sizes coincide with the source image, k being the number of decomposition levels. Fig. 1(a) gives the NSP decomposition with $\mathsf{k}=3$ levels.

2.2. Non-Subsampled Directional Filter Bank (NSDFB)

The NSDFB is constructed by eliminating the downsamplers and upsamplers of the DFB and by upsampling the filters accordingly [20]. This results in a tree composed of two-channel NSFB, described in Fig. 1(b) (4 channel decomposition). At each stage of the NSP, the NSDFB allows a decomposition into any number of 2^l directions, l being the number of levels in the NSDFB. This provides the NSCT with the multi-direction property and offers precise directional information. The combination between NSP and NSDFB is depicted in Fig. 2(a). The resulting filtering structure approximates the ideal partition of the frequency plane displayed in Fig. 2(b). Differently from the contourlet expansion, the NSCT has a redundancy given by $\mathsf{r}=1+\sum_{j=1}^{\mathsf{k}}2^{\ell_j}$, where ℓ_j is the number of