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*

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.

[☆] 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

Email addresses: malay@isical.ac.im (Malay Kumar Kundu), st.manishc@gmail.com (Manish Chowdhury), rotabulo@fbk.eu (Samuel Rota Bulò)

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

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, 2, 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, 5, 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, 8, 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 etc., 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, 22, 23, 24].

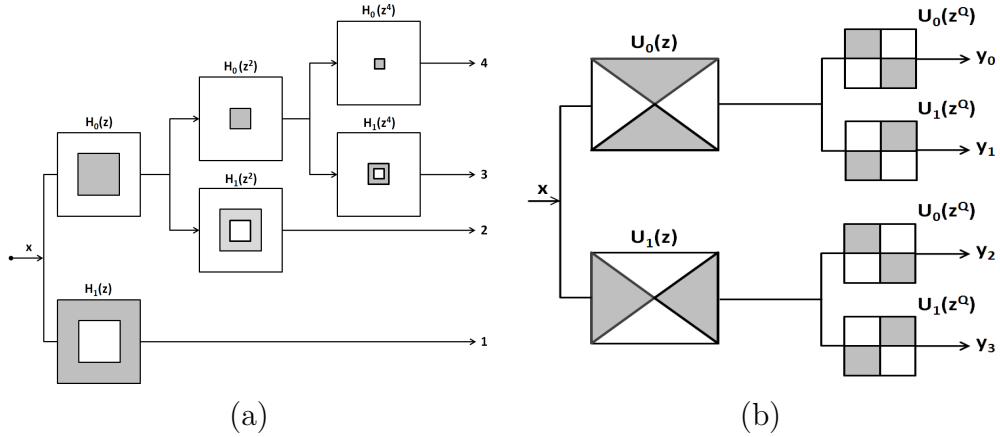
Besides extracting good feature representations, a further important task is feature selection, which aims to seek optimal subsets of the extracted features that preserve most of the information carried by the collected data [25]. The main goal is to facilitate future analysis in the presence of high-dimensional data by improving the query performance of the CBIR system and by reducing the storage requirement. In addition, the reduction in dimensionality favours denoising, 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, 29, 30, 31, 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, 37, 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 labelled 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, 43, 44]. Other approaches define a generative model that uses the unlabelled 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 Section 3 and Sec-



(a) (b)

Figure 1: (a) Non-subsampled Pyramid Filter Bank: Three-stage decomposition. (b) Four-channel NSDFB constructed with two-channel fan filter banks.

tion 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 upsamplers 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 subband decomposition that resembles the Laplacian pyramid, which ensures the multi-scale property of the NSCT. As shown in Figure 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

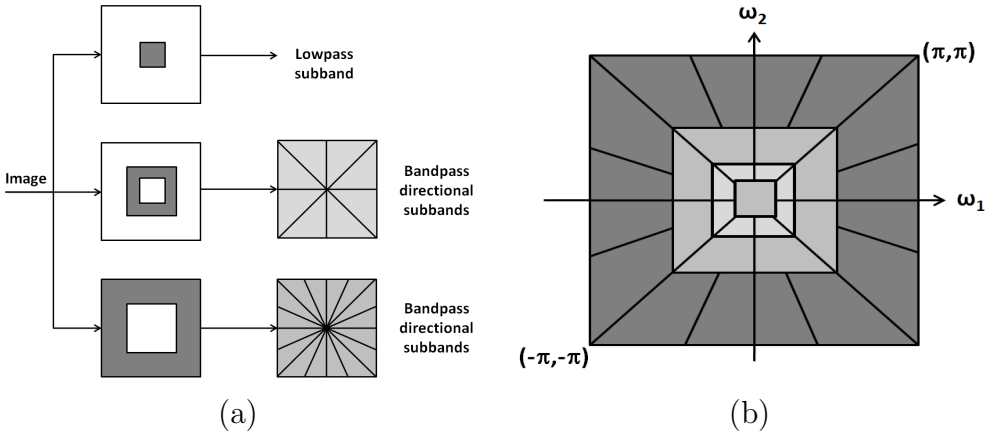


Figure 2: Non-subsampled contourlet transform (a) NSFB structure that implements the NSCT. (b) Idealized frequency partitioning obtained with the proposed structure.

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. Figure 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 Figure 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 Figure 2(a). The resulting filtering structure approximates the ideal partition of the frequency plane displayed in Figure 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 levels in the NSDFB at the jth scale. We refer to [20] for further details about NSCT.

Among the different MGA tools NSCT has better frequency selectivity and regularity, as well as it is a flexible multi-scale, multi-directional, and shift-invariant image decomposition method [47]. NSCT coefficients are capable of capturing the fine details present in the image, which is essential in CBIR, as natural images are full of subtle image information. Moreover,

it is widely accepted that the human visual system bases its perception on multiple channels that are tuned to different ranges of spatial frequencies and orientations. Measurements of the receptive fields of simple cells in the primary visual cortex revealed that these channels exhibit approximately a dyadic structure [48]. This behavior is well matched by NSCT decomposition. These facts motivate the use of NSCT to construct the image signatures in the proposed CBIR system.

3. NSCT based Image Representation

An efficient representation of images usually leads to improvements in terms of storage, computational complexity and performance of image processing algorithms. One way to obtain such representations is by means of image transformation methods. Conventional transformation methods such as Fourier transform and wavelet transform suffer from discontinuities in images like the one introduced by edges. To overcome this problem, we adopt the NSCT method described in the previous section, which exhibits a larger degree of shift-invariance and anisotropy as well as better directional decomposition and frequency localization.

Before extracting the NSCT-based features, we map the RGB images to the CIE Lab color space in order to deal with a color metric that better matches the human perception [49]. This is important as human will play a key role during the retrieval phase. Moreover, by decoupling the brightness intensity and the chromatic channels we ensure an independence between textural and color characterizations. Indeed, the NSCT decomposition over the intensity channel (L) captures the texture information, while the same decomposition over the chromatic coordinates (a and b) characterizes color information. Texture and color information are extracted by using NSCT on the Lab channels with a 4 level (1, 2, 4, 4) decomposition. This decomposition configuration provides for each image in the database 11 ($= 1 + 2 + 4 + 4$) subbands per channel (in total 33 subbands per image). We consider a 4-level NSCT decomposition because the Peak Signal to Noise Ratio (PSNR), which measures the quality of the reconstructed image, does not improve significantly when the level of decomposition is greater than 4. Consequently, the designed level provides a good compromise between having a low loss of information and a compact feature-representation. Note that this empirical evidence is not new, as it was already mentioned in [50]. Each subband $\mathtt{S}^j$, $0 \leq j \leq 33$ is then summarized in terms of its mean (f^j_{mean}), standard

deviation (f^j_{std}) and energy (f^j_{energy}) as follows:

$$f^j_{mean} = \frac{1}{\mathsf{hw}} \sum_{m=1}^{\mathsf{h}} \sum_{n=1}^{\mathsf{w}} \mathtt{S}^j_{mn} \,, \tag{1}$$

$$f^j_{std} = \sqrt{\frac{1}{\mathsf{hw}} \sum_{m=1}^{\mathsf{h}} \sum_{n=1}^{\mathsf{w}} (\mathtt{S}^j_{mn} - f^j_{mean})^2} \,, \tag{2}$$

$$f^j_{energy} = \frac{1}{\mathsf{hw}} \sum_{m=1}^{\mathsf{h}} \sum_{n=1}^{\mathsf{w}} (\mathtt{S}^j_{mn})^2 \,, \tag{3}$$

where, $\mathtt{S}^j_{mn}$ represents the NSCT coefficient at the spatial location (m, n) of the subband $\mathtt{S}^j$, begin a $\mathsf{h} \times \mathsf{w}$-dimensional matrix. Finally, each image I in the database is abstracted in terms of a feature vector comprising the aforementioned statistics for each subband. This yields the following 99-dimensional vector:

$$\boldsymbol{f}^I = \left(f^1_{\text{mean}}, ..., f^{33}_{\text{mean}}, f^1_{\text{std}}, ..., f^{33}_{\text{std}}, f^1_{\text{energy}}, ..., f^{33}_{\text{energy}}\right) \,. \tag{4}$$

3.1. Features selection

The success of the image recognition task depends on the feature representation adopted. Large feature vectors are prone to contain helpful information, but on the other side an excess of poorly informative features can interfere with the positive outcome of the learning algorithm. Moreover, from a computational perspective, large feature representations increase the recognition complexity. For this reason, a good practice consists in trying to reduce the feature space in a way to ideally preserve only the most informative features (*feature selection*).

In our work, we retain the relevant NSCT features by adopting a feature selection algorithm based on the Maximal Information Compression Index (MICI) [28]. This unsupervised method does not rely on searching techniques, thus avoiding over-consumption of calculation time, but it removes redundancies among the features by evaluating their mutual dissimilarity in terms of MICI. In details, given two random variables x and y, let Σ be their covariance matrix with entries Σ_{xx}, Σ_{yy} and $\Sigma_{xy} = \Sigma_{yx}$. The value of MICI for x and y is defined as the smallest eigenvalue $\lambda_2(x, y)$ of the covariance matrix Σ, which is given by

$$\lambda_2(x, y) = \frac{1}{2} \left[\Sigma_{xx} + \Sigma_{yy} - \sqrt{(\Sigma_{xx} - \Sigma_{yy})^2 + 4\Sigma^2_{xy}}\right] \,. \tag{5}$$

The value of $\lambda_2(x, y)$ is zero when the features represented by x and y are linearly dependent (hence, similar) and it increases as the degree of dependency decreases. The algorithm clusters the features into homogeneous groups by exploiting this MICI-based feature dissimilarity and retains a representative feature for each cluster. The set of representatives forms the reduced set of features and its size is controlled by a user-defined parameter (see [28] for further details).

Hereafter, we denote by $\overline{\boldsymbol{f}}^I$ the reduced NSCT feature vector associated to image I, obtained after the application of the aforementioned feature selection procedure.

4. Graph based image re-ranking

Human perception of image similarity is subjective, semantic- and task-dependent. Hence, a CBIR system cannot deliver satisfactory results in general by relying merely on features extracted in an unsupervised way, but the user has to be incorporated in the retrieval loop as a source of feedback. By doing so, the system can iteratively adapt the query results to meet user-specific requirements.

The solution that we adopt has been proposed in [42] and consists in a random-walk-based re-ranking algorithm that relies on a graph abstraction of the images dataset. Let $\mathcal{V} = \{0, \ldots, \mathtt{n}\}$ and let $\mathcal{I} = \{I_u\}_{u \in \mathcal{V}}$ be a set of $\mathtt{n} + 1$ images, I_0 being the query image. We represent our image database in terms of a *(edge-weighted) graph* $G = (\mathcal{V}, \mathcal{E}, \mathtt{W})$, where $\mathcal{V}$ is the set *vertices* corresponding to image indices and $\mathcal{E} \subseteq \mathcal{V} \times \mathcal{V}$ is a set of *edges*, each one being augmented with a nonnegative weight. The *weights* are provided by the $\mathtt{n} \times \mathtt{n}$ matrix $\mathtt{W} = (\mathtt{W}_{ij})$, where each entry $\mathtt{W}_{ij}$ holds the weight of edge $(i, j) \in \mathcal{E}$ or zero if $(i, j) \notin \mathcal{E}$. In our specific case, we use a Gaussian kernel function to determine similarities among pairs of images abstracted in terms of the reduced NSCT features, *i.e.*,

$$\mathtt{W}_{ij} = \exp\left(\frac{\left\|\overline{\boldsymbol{f}}^{I_i} - \overline{\boldsymbol{f}}^{I_j}\right\|^2}{\sigma^2}\right),$$

where the variance σ^2 determines the similarity scale level. The *Laplacian matrix* $\mathtt{L}$ of graph G is defined as $\mathtt{L} = \mathtt{D} - \mathtt{W}$, where $\mathtt{D}$ is a diagonal matrix having diagonal entries $\mathtt{D}_{ii} = \sum_{j \in \mathcal{V}} \mathtt{W}_{ij}$.

The user, who issues the query, can determine whether any image in the database is relevant or not. We denote by $\varphi_i \in \{0, 1\}$, $i \in \mathcal{V}$, the feedback that the user would give to the ith image, namely $\varphi_i = 1$ if the image is relevant and $\varphi_i = 0$ if it is non-relevant. Clearly, we always assume the query image itself to be relevant, *i.e.*, $\varphi_0 = 1$. At each feedback iteration $t > 0$, the user is given the opportunity to indicate which images among the provided ones are relevant for the query. By doing so, the system gathers more and more information about the relevancy of the images in the database. We denote by $\mathcal{L}^{(t)} \subseteq \mathcal{V}$ the indices of images that have been labelled as relevant/non-relevant by the user up to the tth feedback round. Initially, only the query image is labelled, *i.e.*, $\mathcal{L}^{(0)} = \{0\}$, and the set of labelled images can never shrink as the number of feedback rounds increases, *i.e.*, $\mathcal{L}^{(t-1)} \subseteq \mathcal{L}^{(t)}$ for all $t > 0$.

Consider now a generic feedback round $t > 0$. The idea behind the re-ranking algorithm is to determine for each image I_j, $j \in \mathcal{V}$, the probability π_j that a random walk on graph G starting in j will reach a relevant image before hitting a non-relevant one. The ranking can then be obtained by sorting the images in decreasing order of π_j. If we assume the transition probability between two vertices $(i, j) \in \mathcal{E}$ of the graph G to be proportional to the edge weight $\mathtt{W}_{ij}$, the (column) vector of probabilities $\boldsymbol{\pi} = (\pi_0, \ldots, \pi_{\mathsf{n}})^\top$ can be determined as the solution of the following *convex*, quadratic optimization problem [42]:

$$\begin{aligned} \text{minimize} \quad & \boldsymbol{\pi}^\top \mathtt{L}\, \boldsymbol{\pi} \\ \text{s.t.} \quad & \pi_i = \varphi_i \quad \forall i \in \mathcal{L}^{(t)} \\ & 0 \leq \pi_i \leq 1 \quad \forall i \in \mathcal{V}\,. \end{aligned} \tag{6}$$

Without loss of generality, assume the labelled images to have the smallest indices in $\mathcal{V}$, *i.e.*, $\mathcal{L} = \mathcal{L}^{(t)} = \{0, \ldots, \mathsf{m}\}$, in a way to allow the following decompositions for $\mathtt{L}$ and $\boldsymbol{\pi}$ into parts involving labelled (L) and unlabelled (U) images:

$$\mathtt{L} = \begin{bmatrix} \mathtt{L}_{LL} & \mathtt{L}_{LU} \\ \mathtt{L}_{UL} & \mathtt{L}_{UU} \end{bmatrix} \qquad \boldsymbol{\pi} = \begin{pmatrix} \boldsymbol{\pi}_L \\ \boldsymbol{\pi}_U\,. \end{pmatrix}$$

Given this reordering of the indices, the solution $\boldsymbol{\pi}^*$ of (6) can be easily found by solving the following system of linear equations:

$$\mathtt{L}_{UU} \boldsymbol{\pi}_U = -\mathtt{L}_{UL} \boldsymbol{\pi}_L\,. \tag{7}$$

The system is nonsingular if the graph is connected or if every connected component contains a labeled image [51]. By applying some sparsifying strategy

on graph G, the system can be efficiently solved also in case of large-scale image datasets. Once $\boldsymbol{\pi}_U$ and thus $\boldsymbol{\pi}$ is computed, we can sort the images in decreasing order of π_i to get the most relevant images ranked first.

Interestingly, the minimization problem in (6) is a combinatorial formulation of the *Dirichlet problem* [52], *i.e.*, the problem of finding an harmonic function $\pi(\cdot)$ that satisfies the Laplace equation $\nabla^2\pi = 0$ subject to some boundary conditions $\pi(x) = \varphi(x)$ for all $x \in \Omega$ and some fixed function φ. The harmonic function that satisfies the boundary conditions minimizes the *Dirichlet integral* $D(\pi) = \frac{1}{2}\int_\Omega |\nabla u|^2 d\Omega$, for the Euler-Lagrange equation for $D(\cdot)$ gives the Laplace equation. Translated into a graph setting, we have that the function $\pi : \mathcal{V} \to \mathbb{R}$ assigns values to vertices of the graph (so π can be regarded as a vector $\boldsymbol{\pi}$) and the boundary conditions are fixed values of $\pi_i = \varphi_i$ on a subset of vertices $i \in \Omega$. In our case Ω is the set of labelled images $\mathcal{L}^{(t)}$ and φ_i is the feedback for the ith image. Under this discretized setting, the Dirichlet integral is given by $\boldsymbol{\pi}^\top \mathtt{L}\boldsymbol{\pi}$, if we take the Laplacian matrix $\mathtt{L}$ as the combinatorial Laplacian operator. Accordingly, its minimization under the boundary constraints, *i.e.*, the solution to (6), gives the solution to the (combinatorial) Dirichlet problem.[1] More details about the connection to the combinatorial Dirichlet problem and other connections to discrete potential theory are discussed in [52].

5. Summary of the algorithm

In this section we summarize the algorithm underlying our CBIR system with relevance feedback. We start by preprocessing the database of images as follows. We convert each image I in the database into a NSCT feature $\boldsymbol{f}^I$ as detailed in Section 2. By applying the feature selection criterion described in Subsection 3.1, we determine a reduced set of NSCT features, which allows us to lower the dimensionality of the data. As a result, each image I will be associated to a reduced feature vector $\overline{\boldsymbol{f}}^I$. Next, we pre-construct a graph $G_0 = (\mathcal{V} \setminus \{0\}, \mathcal{E}, \mathtt{W})$ having images as vertices and edge-weights representing the similarity of two images computed in terms of the distance between their reduced NSCT feature vectors. The graph construction procedure follows the description in Section 4 with the only exception that we do not consider the query image yet (indeed we excluded 0 from the vertex set).

[1] Note that the conditions $0 \leq \pi_i \leq 1$ in (6) are actually superfluous, for they are always satisfied as long as $0 \leq \varphi_i \leq 1$.

When the user performs a query by presenting an image I_0, we convert it into a reduced NSCT feature vector and we connect it to graph G_0, obtaining graph $G = (\mathcal{V}, \mathcal{E}, \mathtt{W})$ defined in Section 4. The initial query result consist of s images being the most similar to the query image. Here, $\mathtt{s}$ is a user-defined parameter that determines the size of the query result set. The user is then given the opportunity of refining the query result by providing a feedback, which consists in marking the relevant images. The feedback is used by the system to setup the optimization problem in (6). Once the solution $\boldsymbol{\pi}$ is computed by simply solving the linear system of equalities in (7), the images are re-ranked in decreasing order of π_i and the $\mathtt{s}$ best ranked images are returned to the user. This process is reiterated until the user is satisfied with the query result.

6. Experimental Results

We conduct extensive experiments to evaluate the performance of the proposed CBIR system with relevance feedback. We compare the performance of our system with several state-of-the-art CBIR systems. This section is organized as follows. We present the experimental setup in Subsection 6.1 and we motivate in Subsection 6.2 our preference for the CIELab color space as opposed to RGB and YCbCr. In Subsection 6.3, we determine the optimal number of NSCT features that should be retained by the feature selection algorithm presented in Subsection 3.1. In Subsection 6.4 and 6.5 we report qualitative and quantitative experimental results, respectively, obtained by our system against other approaches on three different datasets. Finally, in Subsection 6.6 we compare the execution time of the evaluated methods.

6.1. Experimental setup

We evaluate the proposed CBIR system on images taken from three publicly available image datasets. They are organized into semantic categories in a way to reflect the human perception of image similarity. Details of the datasets are given below:

- **ImgDb1.** This is the SIMPLIcity dataset [53]. It consists of 1000 images uniformly divided into 10 categories.
- **ImgDb2.** This is the Oliva dataset [54], which consists of 2600 images organized into 8 categories with a number of images per category ranging from 260 to 409.

- **ImgDb3.** This is a subset of the Caltech 256 dataset [55], obtained by randomly selecting 100 categories and 25 images per category. In addition, we created 10 copies of each image corrupted with a different type of noise: 5 images are rotated copies (with degree in $\{-10, -5, 0, 5, 10\}$), the remaining 5 instances are respectively a smoothed copy with a circular averaging filter (radius 5 pixels), a blurred copy with camera motion blur (length 9 pixels), a copy corrupted with salt-and-pepper noise (probability 0.05), a copy corrupted by white noise (variance 0.01) and finally a copy corrupted by speckle noise (variance 0.04). In total, we have 25000 images.

Our system uses a 4-level NSCT decomposition as described in Section 3 using the CIELab color space. Our choice for the Lab color space has been motivated by the analysis conducted in Subsection 6.2. The feature selection phase has been tuned according to the procedure presented in Subsection 6.3. The performance on each database is evaluated by considering each image as the query image and by measuring the average classification precision (P) and recall (R) defined as:

$$P = \frac{N_{RIR}}{N_{RIR} + N_{IRIR}} \qquad R = \frac{N_{RIR}}{T_{RID}}$$

where, N_{RIR} is the number of relevant images retrieved, N_{IRIR} is the number of irrelevant images retrieved and T_{RID} is the total number of relevant images in the database.

6.2. Effect of color space transformation

In this experiment, we evaluate the quality of the NSCT-based low-level features computed using different color spaces, namely RGB, YCbCr and CIELab. Figure 3(a)-(c) shows the retrieval performance obtained with the different color spaces on the datasets mentioned before in terms of average precision/recall at varying scope sizes, *i.e.*, the number of images returned by the query, namely 20, 40, 60 and 100. The results show that the CIELab color space outperforms the other ones on all the databases. This is probably due to its perceptual uniformity property, while RGB and YCbCr suffer from color closeness which is not supported by the human visual system [56]. The advatage of adopting the CIELab color space will be evident also in the quantitative experiments of CBIR with relevance feedback that we conduct in Subsection 6.5.

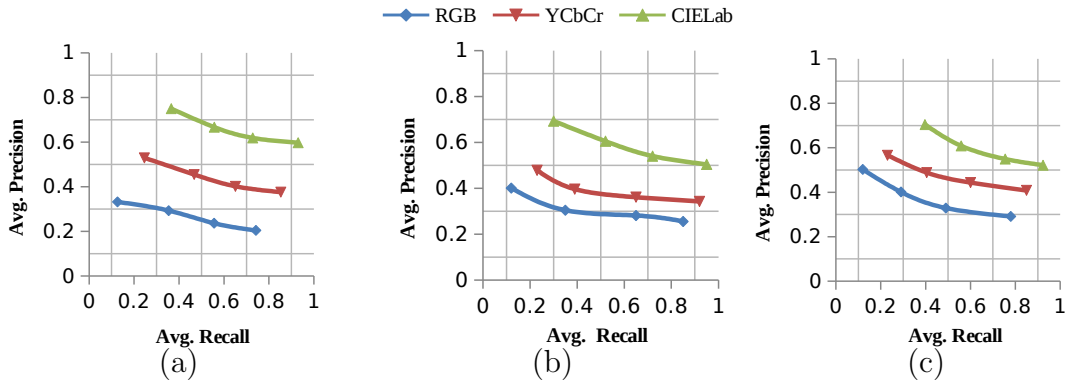


Figure 3: Average Precision Vs Average Recall Graph for Different Color Model (RGB, YCbCr and CIELab) (a) ImgDb1 (b) ImgDb2 (c) ImgDb3

6.3. Feature Evaluation

In this section, we tune for each dataset the number of features that should be selected by the algorithm presented in Subsection 6.3. The idea is to select the number of features that allows for the best category discrimination. To this end, we train a multi-class Least-Squares Support Vector Machine (LS-SVM) classifier using a one-versus-one strategy with a 10-fold cross-validation on the parametrization. A share of 60% labelled samples from the datasets forms the training set, while the remaining 40% is reserved for the test set. We evaluate the performance of the classifier using different shares of selected features (20%, 40%, 60% and 100%) in terms of misclassification error on each dataset (ImgDb1-3). According to the results reported in Figure 4, we can clearly see that the feature selection step has an impact on the performance. The best results are achieved with a share of 60%, 40% and 20% features on ImgDb1, ImgDb2 and ImgDb3, respectively. Hence, the feature selection algorithm will be tuned for the subsequent experiments in a way to return the optimal shares of features that we found for each dataset.

6.4. Qualitative Results

In Figures 5-7 we provide a sample query from each dataset before and after some feedback steps. Figure 5(a) shows the initial retrieval results obtained on ImgDb1 dataset. Here, the top-left image is the query image. As it turns out, 17/20 images are from the same class “flower” and have an appearance similar to the query flower. The remaining ones are misclassified images from the class “food”, which mostly arise due to the prominent role

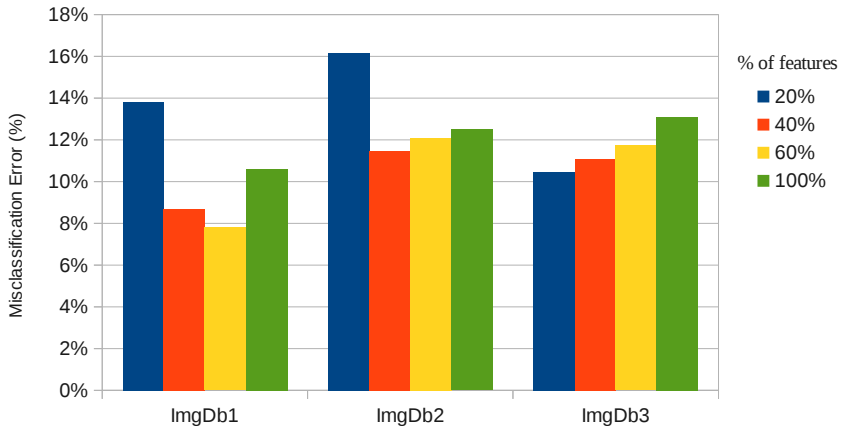


Figure 4: Results obtained on three datasets with a LS-SVM trained considering different share of features.

of the color feature in both the flower and food classes. Nevertheless, also the shape component is partially causing the misclassification, for some erroneous images do have a flower like shape (see,*e.g.*, bottom-right image in (a)). After 1 feedback iteration only 1 error remains, while we achieve 100% accuracy within the 8th feedback iteration.

Figure 5: Retrieval results on ImgDb1 with the query image in the top-left corner. (a) Initial result (85% accuracy) (b) 1st iteration (95%) (c) 8th iteration (100%)

Figure 6(a) shows the initial retrieval results obtained on ImgDb2 dataset. Here, 15/20 images are similar to the query image (top-left), while 5 images

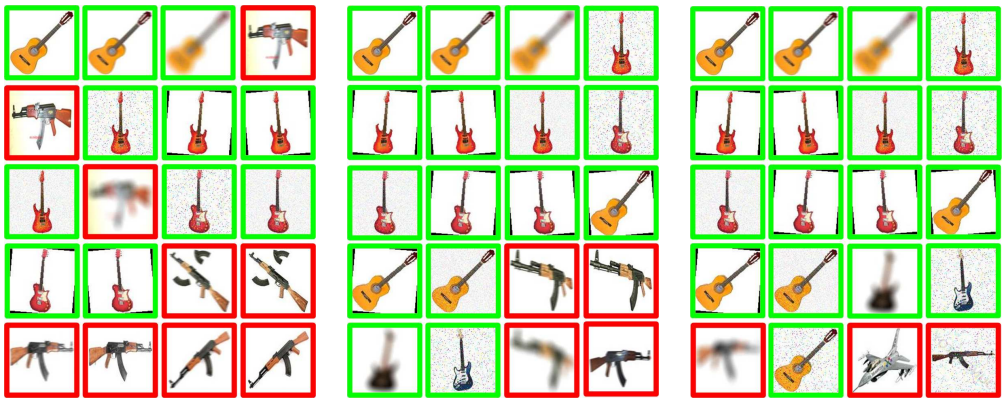
are from different classes. Also in this case, the color cue plays an important role (*e.g.*, color of road and sky), After the first human intervention (see Figure6(b)), the results are improved by reducing the errors to just 3. We achieve 100% accuracy within 4 iterations as shown in Figure 6(c).

(a) (b) (c)

Figure 6: Retrieval results on ImgDb2 with the query image in the top-left corner. (a) Initial result (75% accuracy) (b) 1st iteration (85%) (c) 4th iteration (100%)

Figure 7(a)-(c) shows the retrieval results obtained on ImgDb3 using a guitar query image. The proposed algorithm retrieves both images that represent similar guitars as well as images of the same guitar that have been corrupted (*e.g.*, blurred and rotated), thus demonstrating the robustness of the method to different types of noise. In this case, the distinctive feature is shape and not just color. We experience a high misclassification rate in the initial retrieval result (45%), which decreases to 20% after 1 iteration and to 15% after 4 feedback rounds.

Finally, Figures 8, 9 and 10 show the results of a sample query on ImgDb1 obtained by our approach, by the Ego Similarity Measure (ESM) [57] and by the Semi-Supervised Learning (SSL) approach of Yang *et al.* [58]. Both ESM and SSL are recent approaches for CBIR with relevance-feedback: SSL is a state-of-the-art, graph-based ranking mechanism that is closer in spirit to our method thus being a direct competitor, whereas ESM is among the most recent approaches based on a weight-updating scheme. As we can see from the plots, our approach achieves the best accuracy (90%) already after the first iteration, followed by SSL with 85% and ESM with 80%. Also in terms of ranking quality, our approach outperforms the competitors, for the first error



(a) (b) (c)

Figure 7: Retrieval results on ImgDb3 with the query image in the top-left corner. (a) Initial result (55% accuracy) (b) 1st iteration (80%) (c) 4th iteration (85%)

appears at the 17th position compared to the 4th and 9th position of ESM and SSL, respectively. After 4 feedback iterations, our approach achieves the highest accuracy and the best ranking quality compared to ESM and SSL, which in turn obtain an accuracy of 85% and 90%. It is interesting to noice that most of the errors are in favor of the "African people" and "building" class, probably due to the presence of color combinations or edge distributions, that resemble the ones of the typical bus images.

6.5. Quantitative results

We conduct in this section some quantitative experiments to show the effectiveness of our approach.

In the first place we do not involve the user in the retrieval loop and, therefore, we simply measure the quality of the initial retrieved set of images. To the sake of comparison, we evaluate our approach on ImgDb1 against the following approaches: [53], [59], [60], [61] and [62]. Note that we limit the analysis under this setting to ImgDb1 because no score was reported for the other two datasets in the competitors' papers. We report the average precision obtained on 10 random queries per class (in total 80 queries) with a scope size of 20 images. As we can see in Figure 11, our approach gives an overall satisfactory performance compared to the other state-of-the-art approaches. We outperform the competitors in 6/10 classes and obtained comparable results on the "Dinosaurs" class. As for the remaining classes, we never fall below 50% and we exhibit on average a good performance.

(a) (b) (c)

Figure 8: Retrieval results on ImgDb1 obtained by our approach with the query image in the top-left corner. (a) Initial result (80% accuracy) (b) 1st iteration (90%) (c) 4th iteration (100%)

(a) (b) (c)

Figure 9: Retrieval results on ImgDb1 obtained by ESM with the query image in the top-left corner. (a) Initial result (80% accuracy) (b) 1st iteration (80%) (c) 4th iteration (85%)

(a) (b) (c)

Figure 10: Retrieval results on ImgDb1 obtained by SSL with the query image in the top-left corner. (a) Initial result (80% accuracy) (b) 1st iteration (85%) (c) 4th iteration (90%)

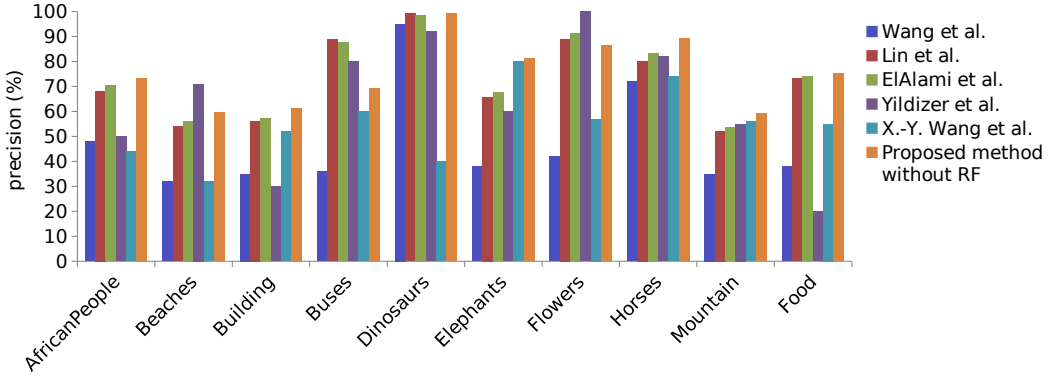


Figure 11: Comparisons with other existing CBIR systems in terms of average precision on ImgDb1

We conduct a second experiment on all datasets involving also the relevance feedback. We compared our approach against the ESM and SSL relevance feedback methods mentioned in the previous section, considering a scope size of 20 images. In Figures 12(a)-(c) we report the results obtained by our approach, ESM and SSL in terms of average precision with respect to 10 random queries per image category. Moreover, we compare the performance of the three approaches by varying also the starting color spaces, namely CIELab, RGB and YCbCr. All approaches start from the same baseline in terms of the initial query result, which is obtained from the NSCT-based representation proposed in this paper. This gives indeed a good starting point when we consider the CIELab color space with an average precision above 60%, whereas the starting performance is lower with RGB and YCbCr. This gap in performance between the color spaces is consistent with the preliminary experiment that we conducted in Subsection 6.2. Our relevance feedback mechanism outperforms the competitors on all the experiments conducted. In general, the results on small datasets involving less categories like ImgDb1 are better in absolute terms that the one obtained on larger ones like, *e.g.*, ImgDb3. Nevertheless, the improvements deriving from the user's feedback in our method are around 25% for all datasets when we consider the CIELab color spaces. The performances are get worse when we move from CIELab to YCbCr and RGB. As for the competitors, we have that SSL outperforms ESM, thus highlighting the superiority of graph-based ranking mechanisms over the simpler weight-updating schemes.

6.6. Running Time

The experiments have been conducted with Matlab 2012b on a Windows 7 machine equipped with an Intel i7-2600 CPU 3.40 GHz and 4Gb of RAM. In this section we report the query time, which is an important component since the relevance feedback algorithm will be run several times between user interactions.

In Figure 13, we report the average execution time in seconds of each approach on the different datasets with the NSCT-based features as a function of each relevance feedback round. Our approach is slightly slower than ESM, but faster than SSL. It is worth mentioning, however, that our current implementation employs a rather dense graph representation for the image datasets. In [42], it is indeed shown that sparsifying the graph allows to have a performance boost both in terms of execution time and accuracy.

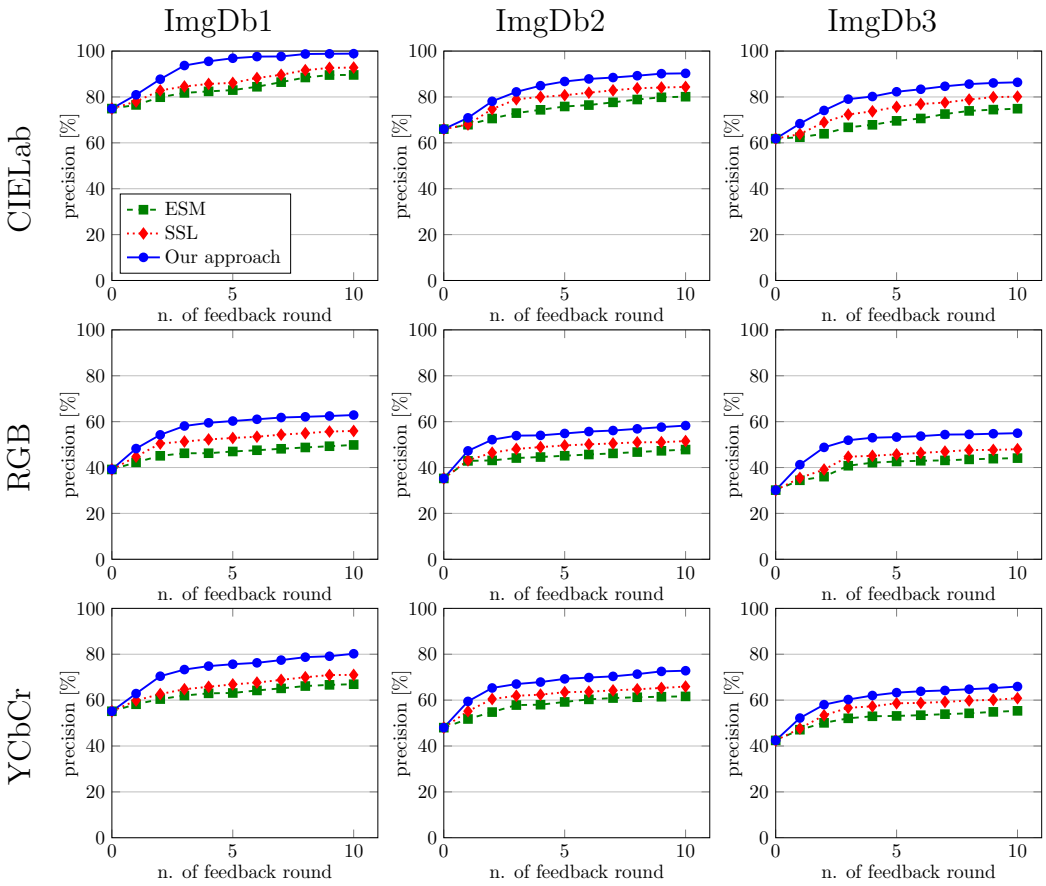


Figure 12: Plots of the average precision of different CBIR with RF approaches using CIELab/RGB/YCbCr color spaces and scope size 20 on datasets ImgDb1, ImgDb2 and ImgDb3.

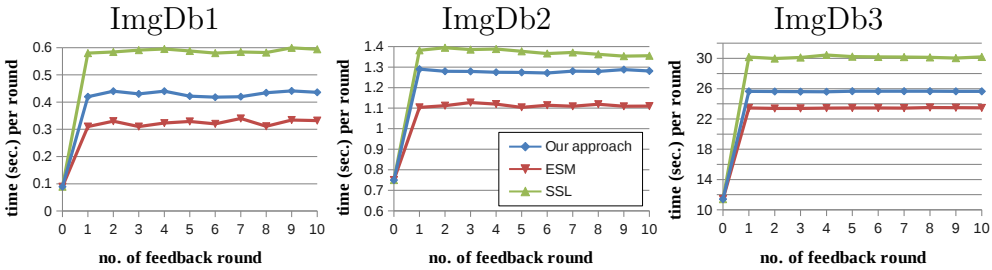


Figure 13: Plots of the average running time in seconds at the different feedback rounds on dataset ImgDb1, ImgDb2 and ImgDb3, using CIELab color space.

7. Conclusions

In this paper, we presented an efficient interactive CBIR system with relevance feedback, which is based on NSCT features that allow to capture geometrical information about the images. To enhance the retrieval accuracy, we performed feature selection using an unsupervised algorithm. The new reduced feature abstractions are used to retrieve similar images from the dataset, given a query image. In addition, we employed a random-walk-based method to incorporate a relevance feedback mechanism and thus better fit the retrieval results to the user's needs.

Experimental evaluations of the proposed framework have shown that our approach is effective and improves the retrieval performance of CBIR systems significantly compared to other state-of-the-art methods. In more details, the adoption of a feature representation, which derives from a flexible multi-scale, multi-directional and shift-invariant image decomposition method like NSCT, endows our system with the ability of capturing similarities within images in either shape or color cues also in the presence of changes in scale, rotation and translation. The experiment on CBIR without user intervention that we have reported in Figure 11 showed how this feature abstraction choice allows to achieve good performances on classes of images like,*e.g.*, horses and elephants, where both color and shape cues play an important role. At the same time, it also showed encouraging discriminant capabilities on classes of images that exhibit a larger variability in either cues like, *e.g.*, buildings and food. Part of the efficacy of the method is also due to the feature selection step, which in our case was based on MICI. As one can read from the experiment we conducted in Figure 4 on a supervised classification task, this step contributes by reducing the dimensionality of

the feature representation in a way to keep the positive signal and remove redundancies that might negatively affect the subsequent retrieval process in terms of quality of the result and execution time. Finally, a graph-based ranking system has been proposed to include the user in the retrieval loop in order to reduce the semantic gap. The quantitative experiments on CBIR with relevance feedback that we have conducted on three different datasets, and reported in Figure 12, have shown that the proposed ranking mechanism effectively exploits the user's feedback to improve the quality of the retrieved set of images on all settings we considered.

The content-based image retrieval system can be adapted also for video retrieval. To this end, our future research efforts will be geared towards incorporating motion information in the proposed pipeline to allow for an effective exploitation of the temporal dimension.

Acknowledgements

The work is mainly funded by Machine Intelligence Unit, Indian Statistical Institute, Kolkata-108 (Internal Academic Project) for providing facilities to carry out this work. Malay K. Kundu acknowledges the Indian National Academy of Engineering (INAE) for their support through INAE Distinguished Professor fellowship. We would also like to thank Dr. Yi Yang (Carnegie Mellon University) for providing us with the SSL source code.

References

[1] A. Dong, B. Bhanu, Active concept learning in image databases, IEEE Trans. Syst. Man, Cybern. B, Cybern. 35 (3) (2005) 450–456.

[2] A. Adler, M. E. Schuckers, Comparing human and automatic face recognition performance, IEEE Trans. Syst. Man, Cybern. B, Cybern. 37 (5) (2007) 1248–1255.

[3] A. Perina, M. Cristani, V. Murino, Unsupervised learning of saliency concepts for natural image classification and retrieval, in: Progress in Pattern Recognition, Image Analysis and Applications, 13th Iberoamerican Congress on Pattern Recognition, CIARP, Vol. 5197 of Lecture Notes in Computer Science, Springer Berlin Heidelberg, 2008, pp. 169–177.

[4] S. Zhang, J. Huang, H. Li, D. N. Metaxas, Automatic image annotation and retrieval using group sparsity, IEEE Trans. Syst. Man, Cybern. B, Cybern. 42 (3) (2012) 838–849.

[5] A. Bishnu, B. B. Bhattacharya, M. K. Kundu, C. A. Murthy, T. Acharya, Euler vector for search and retrieval of gray-tone images, IEEE Trans. Syst. Man, Cybern. B, Cybern. 35 (4) (2005) 801–812.

[6] M. E. ElAlami, Supporting image retrieval framework with rule base system, Knowledge-Based Systems 24 (2011) 331–340.

[7] L. Zhang, L. Wang, W. Lin, Generalized biased discriminant analysis for content-based image retrieval, IEEE Trans. Syst. Man, Cybern. B, Cybern. 42 (1) (2012) 282–290.

[8] J. Yu, D. Liu, D. Tao, H. S. Seah, On combining multiple features for cartoon character retrieval and clip synthesis, IEEE Trans. Syst. Man, Cybern. B, Cybern. 42 (5) (2012) 1413–1427.

[9] R. Datta, D. Joshi, J. Li, J. Z. Wang, Image retrieval: Ideas, influences, and trends of the new age, ACM Comput. Surv. 40 (2) (2008) 1–60.

[10] G. Quellec, M. Lamard, G. Cazuguel, B. Cochener, C. Roux, Adaptive nonseparable wavelet transform via lifting and its application to content-based image retrieval, IEEE Trans. Image Process. 19 (1) (2010) 25–35.

[11] E. Yildizer, A. M. Balci, T. N. Jarada, R. Alhajj, Integrating wavelets with clustering and indexing for effective content-based image retrieval, Knowledge-Based Systems 31 (2012) 55–66.

[12] Q. Tian, N. Sebe, M. S. Lew, E. Loupias, T. Huang, Image retrieval using wavelet-based salient points, J. of Electronic Imaging 10 (4) (2001) 835–849.

[13] M. Kokare, P. K. Biswas, B. N. Chatterji, Rotation-invariant texture image retrieval using rotated complex wavelet filters, IEEE Trans. Syst. Man, Cybern. B, Cybern. 36 (6) (2006) 1273–1282.

[14] T. Celik, T. Tjahjadi, Multiscale texture classification and retrieval based on magnitude and phase features of complex wavelet subbands, Computers and Electrical Engg. 37 (5) (2011) 729–743.

[15] M. Banerjee, M. K. Kundu, Edge based features for content based image retrieval, Pattern Recognition 36 (11) (2003) 2649–2661.

[16] M. Acharyya, M. K. Kundu, Extraction of features using M-band wavelet packet frames and their neuro-fuzzy evaluation for multi-texture segmentation, IEEE Trans. Pattern Anal. Mach. Intell. 25 (12) (2003) 1639–1644.

[17] M. N. Do, M. Vetterli, The contourlet transform: an efficient directional multiresolution image representation, IEEE Trans. Image Process. 14 (12) (2005) 2091–2106.

[18] S. M. Youssef, ICTEDCT-CBIR: Integrating curvelet transform with enhanced dominant colors extraction and texture analysis for efficient content-based image retrieval, Computers and Electrical Engg. 38 (5) (2012) 1358–1376.

[19] M. K. Kundu, M. Chowdhury, M. Banerjee, Interactive image retrieval with wavelet features, Proc. of 4th Int. Conf. on Pattern recognition and Machine Intelligence (2011) 162–172.

[20] A. L. da Cunha, J. Zhou, M. N. Do, The nonsubsampled contourlet transform: Theory, design, and applications, IEEE Trans. Image Process. 15 (10) (2006) 3089–3101.

[21] Q. Zhang, L. Gao, Medical image retrieval based on nonsubsampled contourlet transform and fractal dimension, in: Proc. of 3rd Int. Conf. on Bioinformatics and Biomedical Engineering, 2009, pp. 1–4.

[22] X. Xu, D. Zhang, X. Zhang, An efficient method for human face recognition using nonsubsampled contourlet transform and support vector machine, Optica Applicata XXXIX, (3) (2009) 601–615.

[23] H. Ng-Duc, T. Le-Tien, T. Do-Honl, C. Bui-Thu, T. Ng-Xuan, Image retrieval using contourlet based interest points, Proc. of 10th Int. Conf. on Inf. Sci. Signal Process. and their Applications (2010) 93–96.

[24] M. Chowdhury, S. Das, M. K. Kundu, Compact image signature generation: An application in image retrieval, in: Proc. 5th Int. Conf. on Computer Science and Inf. technology, 2013, pp. 1–7.

[25] L. Xu, I. King, A pca approach for fast retrieval of structural patterns in attributed graphs, IEEE Trans. Syst. Man, Cybern. B, Cybern. 31 (5) (2001) 812–817.

[26] T. Xia, D. Tao, T. Mei, Y. Zhang, Multiview spectral embedding, IEEE Trans. Syst. Man, Cybern. B, Cybern. 40 (6) (2010) 1438–1446.

[27] D. T. B. Xie, Yang Mu, K. Huang, m-SNE: Multiview stochastic neighbor embedding, IEEE Trans. Syst. Man, Cybern. B, Cybern. 41 (4) (2011) 1088–1096.

[28] P. Mitra, C. A. Murthy, S. K. Pal, Unsupervised feature selection using feature similarity, IEEE Trans. Pattern Anal. Mach. Intell. 24 (4) (2002) 301–312.

[29] P. W. Tommy W. S. Chow, E. W. M. Ma, A new feature selection scheme using a data distribution factor for unsupervised nominal data, IEEE Trans. Syst. Man, Cybern. B, Cybern. 38 (2) (2008) 499–509.

[30] K. Z. Mao, Identifying critical variables of principal components for unsupervised feature selection, IEEE Trans. Syst. Man, Cybern. B, Cybern. 35 (2) (2005) 339–344.

[31] Q. Huang, D. Tao, X. Li, L. Jin, G. Wei, Exploiting local coherent patterns for unsupervised feature ranking, IEEE Trans. Syst. Man, Cybern. B, Cybern. 41 (6) (2011) 1471–1482.

[32] E. Rashedi, H. Nezamabadi-pour, S. Saryazdi, A simultaneous feature adaptation and feature selection method for content-based image retrieval systems, Knowledge-Based Systems 39 (2013) 85–94.

[33] M. Banerjee, M. K. Kundu, P. Maji, Content-based image retrieval using visually significant point features, Fuzzy Set. and Syst. 160 (23) (2009) 3323–3341.

[34] L. Wang, K. L. Chan, P. Xue, A criterion for optimizing kernel parameters in kbda for image retrieval, IEEE Trans. Syst. Man, Cybern. B, Cybern. 35 (3) (2005) 556–562.

[35] E. Cheng, F. Jing, L. Zhang, A unified relevance feedback framework for web image retrieval, IEEE Trans. Image Process. 18 (6) (2009) 1350–1357.

[36] P. Y. Yin, B. Bhanu, K. C. Chang, A. Dong, Integrating relevance feedback techniques for image retrieval using reinforcement learning, IEEE Trans. Pattern Anal. Mach. Intell. 27 (10) (2005) 1536–1551.

[37] J. Han, K. N. Ngan, M. Li, H. J. Zhang, A memory learning framework for effective image retrieval, IEEE Trans. Image Process. 14 (4) (2005) 521–524.

[38] F. C. Chang, H. M. Hang, A relevance feedback image retrieval scheme using multi-instance and pseudo image concepts, IEICE Trans. Inf. System E89-D (2006) 521–524.

[39] J. He, M. Li, H. Zhang, H. Tong, C. Zhang, Manifold ranking based image retrieval, in: IEEE Int. Conf. on Mult. and Expo, 2004, pp. 9–16.

[40] K. Tieu, P. Viola, Boosting image retrieval, in: IEEE Conf. Computer Vision and Patt. Recogn., Vol. 1, 2000, pp. 228–235.

[41] L. Zhang, F. Lin, B. Zhang, Support vector machine learning for image retrieval, in: Int. Conf. Image Process., 2001, pp. 721–724.

[42] S. Rota Bulò, M. Rabbi, M. Pelillo, Content-based image retrieval with relevance feedback using random walks, Pattern Recognit. 44 (9) (2011) 2109–2122.

[43] H. Sahbi, P. Etyngier, J. Y. Audibert, R. Keriven, Manifold learning using robust graph laplacian for interactive image search, in: IEEE Conf. Computer Vision and Patt. Recogn., 2008, pp. 1–8.

[44] J. He, H. Tong, M. Li, W. Y. Ma, C. Zhang, Multiple random walk and its application in content-based image retrieval, in: Proc. of the 7th ACM SIGMM Int. Workshop on Mult. Inf. retrieval, ACM, New York, NY, USA, 2005, pp. 151–158.

[45] Y. Wu, Q. Tian, T. Huang, Discriminant-em algorithm with application to image retrieval, in: IEEE Conf. Computer Vision and Patt. Recogn., Vol. 1, 2000, pp. 222–227.

[46] M. J. Huiskes, M. S. Lew, Performance evaluation of relevance feedback methods, in: Proc. of the Int. Conf. on Content-based image and video retrieval, ACM, New York, NY, USA, 2008, pp. 239–248.

[47] S. Li, B. Yang, J. Hu, Performance comparison of different multi-resolution transforms for image fusion, Inf. Fusion 12 (2) (2011) 74–84.

[48] A. M. Garcia-Perez, The perceived image: Efficient modelling of visual inhomogeneity, Spatial Vision 6 (1992) 89–99.

[49] F. Lpez, J. Valiente, R. Baldrich, M. Vanrell, Fast surface grading using color statistics in the cie lab space, in: Pattern Recognition and Image Analysis, Vol. 3523 of Lecture Notes in Computer Science, Springer, Berlin Heidelberg, 2005, pp. 666–673.

[50] N. Doghmane, Z. Baarir, N. Terki, A. Ouafi, Study of effect of filters and decomposition level in wavelet image compression, Courrier du Savoir 3 (2003) 41–45.

[51] E. Mortensen, W. Barrett, Interactive segmentation with intelligent scissors, Graph. Mod. in Image Process. 60 (5) (1998) 349–384.

[52] L. Grady, Random walks for image segmentation, IEEE Trans. Pattern Anal. Mach. Intell. 28 (11) (2006) 1768–1783.

[53] J. Z. Wang, J. Li, G. Wiederhold, SIMPLIcity: Semantics-sensitive integrated matching for picture libraries, IEEE Trans. Pattern Anal. Mach. Intell. 23 (2001) 947–963.

[54] A. Oliva, A. Torralba, Modeling the shape of the scene: A holistic representation of the spatial envelope, Int. J. of Comp. Vision 42 (3) (2001) 145–175.

[55] A. H. G. Griffin, P. Perona, Caltech 256 object category dataset, Technical Report UCB/CSD-04-1366, California Institute of Technology.

[56] O. Ikeda, Segmentation of faces in video footage using controlled weights on hsv color, in: 13th Scandinavian Conf. on Image Analysis, Vol. 2749 of Lecture Notes in Computer Science, Springer, Berlin Heidelberg, 2003, pp. 163–170.

[57] C. H. Lee, M. F. Lin, Ego-similarity measurement for relevance feedback, Expert Syst. Appl. 37 (1) (2010) 871–877.

[58] Y. Yang, F. Nie, D. Xu, J. Luo, Y. Zhuang, Y. Pan, A multimedia retrieval framework based on semi-supervised ranking and relevance feedback, IEEE Trans. Pattern Anal. Mach. Intell. 34 (4) (2012) 723–742.

[59] C. H. Lin, R. T. Chen, Y. K. Chan, A smart content-based image retrieval system based on color and texture feature, J. Image Vision Comput. 27 (6) (2009) 658–665.

[60] M. E. ElAlami, A novel image retrieval model based on the most relevant features, Knowledge-Based Systems 24 (2011) 23–32.

[61] E. Yildizer, A. M. Balci, M. Hassan, R. Alhajj, Efficient content-based image retrieval using multiple support vector machines ensemble, Expert Syst. Appl. 39 (3) (2012) 2385–2396.

[62] X. Y. Wang, H. Y. Yang, Y. W. Li, F. Y. Yang, Robust color image retrieval using visual interest point features of significant bit-planes, Digital Signal Process. 23 (4) (2013) 1136–1153.