Published in final edited form as:

Inf Process Manag. 2011 September 1; 47(5): 676–691. doi:10.1016/j.ipm.2010.12.001.

A Query Expansion Framework in Image Retrieval Domain Based on Local and Global Analysis

M. M. Rahman¹, S. K. Antani, and G. R. Thoma

U.S. National Library of Medicine, National Institutes of Health, Bethesda, MB, USA

Abstract

We present an image retrieval framework based on automatic query expansion in a concept feature space by generalizing the vector space model of information retrieval. In this framework, images are represented by vectors of weighted concepts similar to the keyword-based representation used in text retrieval. To generate the concept vocabularies, a statistical model is built by utilizing Support Vector Machine (SVM)-based classification techniques. The images are represented as “bag of concepts” that comprise perceptually and/or semantically distinguishable color and texture patches from local image regions in a multi-dimensional feature space. To explore the correlation between the concepts and overcome the assumption of feature independence in this model, we propose query expansion techniques in the image domain from a new perspective based on both local and global analysis. For the local analysis, the correlations between the concepts based on the co-occurrence pattern, and the metrical constraints based on the neighborhood proximity between the concepts in encoded images, are analyzed by considering local feedback information. We also analyze the concept similarities in the collection as a whole in the form of a similarity thesaurus and propose an efficient query expansion based on the global analysis. The experimental results on a photographic collection of natural scenes and a biomedical database of different imaging modalities demonstrate the effectiveness of the proposed framework in terms of precision and recall.

Keywords

Image Retrieval; Vector Space Model; Support Vector Machine; Relevance Feedback; Query Expansion

1 Introduction

To reduce the ambiguity due to the word mismatch problem in information retrieval and to thereby increase retrieval effectiveness, a variety of query modification and reformulation strategies have been investigated [1–6]. There are three main approaches for query reformulation [7]:

1. Interactive techniques based on feedback information from the user, commonly known as *relevance feedback* (RF);

¹ Corresponding author. Tel.: 1 301 435 3262, rahmanmm@mail.nih.gov.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

2. Automated techniques based on the global information derived from the entire collection or corpus, commonly known as *thesaurus-based query expansion*, and
3. Automated (might be interactive also in some cases) techniques based on local information from the top retrieved results, commonly known as *local feedback*.

The efficiency of image relevance computation has accelerated the development and use of RF in these three approaches [8–10] in content-based image retrieval (CBIR) applications [11–14]. In CBIR, the access to any information is performed at a perceptual level based on automatically extracted visual features, such as color, texture, shape, and location [17]. There is a great interest and a wealth of promise in CBIR as an emerging technology. Comprehensive surveys exist on both early low-level and recent semantic-level image retrieval techniques [17–19]. However, even after almost two decades of intensive research, CBIR systems still lag behind the best text-based search engines of today. One of the leading causes for this is the well known “*semantic gap*” problem [17], which is the mismatch between human understanding of image content consisting of semantic concepts (high level) and image representation through fairly rudimentary features (low level). To narrow the semantic gap, techniques have been developed over the last few years for semantic image classification and retrieval, which utilize various off-line and on-line machine learning-based methods [18–20,11,12].

Among these, a number of RF-based approaches have been proposed in the CBIR domain, such as: query point movement, feature re-weighting, active learning, etc. [11–14]. The majority of these approaches estimate the ideal query parameters from the low-level image features at a global level. However, for the high-level concepts or semantics that cannot be adequately represented by low-level features, these systems often do not return many relevant results even with a large number of feedback iterations. To address this limitation, some recent systems incorporate the RF approach on both low-level content and high-level context (keyword) based feature spaces for Web image retrieval [15,16]. Although the RF algorithms provide a performance boost, one of the major drawbacks is that the users do not always provide enough feedback information to make the methods perform effectively. It has often been found to be a complex mechanism with different levels of feedback, such as the relevance levels in the MARS [12] and the goodness scores in the Mindreader [14] systems.

The automatic query reformulation based on term co-occurrence or term similarity also has been widely investigated in the text retrieval domain with varying degrees of success [2–6]. These approaches have a significant advantage over interactive RF as they require no effort on the part of the user. The techniques that have so far been investigated can be described as being based on either global or local analysis [2,3]. In a global analysis approach, all the documents in the collection are analyzed to determine a global thesaurus-like structure that defines certain term relationships. This structure is then used to select additional terms for the query expansion. In local analysis, the top retrieved documents for a query are examined (usually without any assistance from the user) at query time to determine the terms for the query expansion [3,4].

Due to the nature of the low-level continuous feature representation, it is difficult to perform query reformulation or expansion based on term co-occurrence or the term similarity in the CBIR domain. The obvious choice for terms in text documents are the words that are directly related to the semantic contents. However, the images contain information in a very dense and continuous form where the units of an image at a low level are pixel values having no direct semantic meaning. In order to perform an automatic query expansion based on term-term correlations, the images should have a similar feature type and distribution as the ones from the text. Therefore, the number of possible image features should be high, and

the number of image features per image should be low. Recently, the ongoing success of the “*bag-of-words*” methodology in the text retrieval domain have motivated researchers to explore analogous techniques mainly for scene classification and retrieval [20]. The models are applied to images by using a visual analogue of a word by automatically extracting different predominant color or texture patches or semantic patches, such as, water, sand, sky, cloud, etc. in natural photographic images. This intermediate semantic level representation is introduced as a first step to deal with the semantic gap between low-level features and high-level concepts.

Recent works have shown that local features represented by “*bags-of-words*” are suitable for scene classification, yielding impressive levels of performance [22–26]. For example, a framework to automatically generate the visual terms (“keyblock”), proposed in [22], applies a vector quantization or clustering technique. It represents images similar to the “*bags-of-words*” representation in correlation-enhanced feature spaces based on the n-gram and bi-gram models [7]. For the reliable identification of image elements, the work in [23] manually identifies visual patches (“visual keywords”) from the sample images. In [24], a compact and sparse representation of images is proposed based on using a region codebook generated by a clustering technique. A semantic modeling approach, investigated in [25] for a small collection of images, is based on the binary classification of semantic patches of local image regions. Similarly, in [26] a binary support vector machine (SVM) is employed for the model generation of 23 selected concepts. At the testing stage, unlabeled regions are fed into all the models, and the concept from the model giving the highest positive result is associated with the region. A learning-based unified image retrieval framework to represent images in local visual and semantic concept-based feature spaces is presented in our previous work [27]. In this framework, a visual concept vocabulary (codebook) is automatically constructed by utilizing a self-organizing map (SOM), and statistical models are built for local semantic concepts using a probabilistic multi-class support vector machine (SVM). A detailed review and accuracy comparison of some of the most commonly used scene classification approaches has recently been presented in [20]. Experimental evaluation has shown that approaches using intermediate semantic concepts are more appropriate to reduce the gap between low and high-level.

1.1 Motivations and Contributions

Besides the loss of all ordering structure, each word (term) is considered independent of all the words in the *bag-of-words* model. Although this simple assumption has proved to be effective in the text retrieval domain, it is found that the words are related and their similarity of occurrence in documents can reflect the underlying semantic relations between them. Similar to the text retrieval, we might find several correlated or co-occurring concepts for a particular one in image domain. For example, in an outdoor image, there might be a higher probability of occurrence of a blue sky around the sun. Whereas, a yellow colored flower has more probability to co-occur with green leaves in an image of a garden. Hence, there is indeed a need to expand the concepts in an original query image by exploiting the information on correlation patterns to improve the retrieval effectiveness.

Due to the similarity of the concept-based feature representation schemes to the keyword-based representation of documents and to address the limitations of both the low-level and concept-level feature representations, we explore analogous query expansion techniques in the image domain from a novel perspective. We propose automatic query expansion approaches in a concept feature space based on both the local and global analysis. At first, to generate the vocabulary of the local concepts and represent images with the concept frequencies, a statistical model is built by utilizing the SVM-based multi-class classification technique. For the automatic query expansion based on a local analysis, we exploit the concept-concept correlations by analyzing a local clustering method that takes into account

the co-occurrence information and the metrical constraints based on the neighborhood proximity between the concepts in encoded images. We also analyze the concept similarities in the collection as a whole in the form of a global similarity thesaurus and explore an efficient query expansion technique based on the global analysis. This work is an extended version with global analysis of our previous work [21] of query expansion in image domain.

The organization of the paper is as follows: Section 2 reviews existing query reformulation and modification approaches in the CBIR. In Section 3, the semantic concept-based image representation approach is described. Section 4 presents a query expansion approach based on the local analysis and Section 5 presents a query expansion technique based on the global information analysis of the entire collection. Exhaustive experiments and analysis of the results are presented in Sections 6 and 7 that are based on a photographic image collection and a medical image collection. Finally Section 8 provides our conclusions.

2 Related Work

The idea of query expansion or reformulation techniques based on the term co-occurrence or the term similarity is relatively new in the CBIR domain due to the nature of the low-level continuous feature representation. Until recently only a few techniques have been investigated along this direction [28–31]. For example, a context expansion approach has been explored recently in [28] to take advantages of the correlations between key regions (representing the main semantic contents) and environmental regions (representing the contexts) by expanding the key regions of the queries using highly correlated environmental regions according to an image thesaurus. The thesaurus is constructed through a data-driven approach, serves as both a mapping function between image low-level features and concepts and a store of the statistical correlations between different concepts. Due to its data driven approach, it does not consider user perceived semantics through feedback iterations. In [29], a keyword similarity matrix is learned by a statistical algorithm through users feedback information, and a soft query expansion approach based on this similarity matrix is adopted to infer keywords which is somewhat related to the users interest. For example, when “Ford” and “Toyota” is used as query, the system will automatically infer “car” through the query expansion approach. A cross-modality query expansion approach is proposed in [30] to further improve image retrieval performance. When a query is given in the form of keywords, it is expanded by a set of keywords selected from a semantic keyword network constructed through keyword propagation, and the low-level features of the corresponding images are then incorporated into that of the query. In [31], query expansion is applied in its relevance model based on a set of training images to automatically annotate and retrieve images. An image is represented as a set of blobs resulted from clustering on image features as well as a set of keywords. It is assumed that for each image there exists a relevance model containing all possible blobs that could appear in the image and all the keywords that could appear in the caption of the image. The probability of drawing a keyword or a blob from the model of an image is given by a linear combination of the probability of this keyword/blob appearing in this image and that in the training set. This linear combination procedure is fundamentally a smoothing scheme, which is different from our approach.

The approaches mentioned above are either data dependent on the entire collection [28] or relied on keywords from image captions [29–31]. Our proposed approach of query expansion is fundamentally different from the above approaches. It relies solely on a concept-based image representation where images are modeled as “*bag of concepts*”. Here, the concept co-occurrence or the concept similarity are analyzed by considering both the global collection and the local feedback level.

3 Image Representation on Local Concept Space

A major component of the proposed query expansion framework is the image representation in a local concept space. By the term “*local concept*”, we refer to the semantically distinguishable local patches or regions in the individual images. For example, in an image collection of natural scenery, the specific local patches such as, water, sand, grass, sky, snow, and so on can be identified with different semantic interpretations. In a heterogeneous collection of medical images, it is possible to identify specific local patches that are perceptually and/or semantically distinguishable, such as homogeneous texture patterns in grey level radiological images, differential color and texture structures in microscopic pathology and dermoscopic images. However, semantic similarity does not always correspond to perceptual or visual similarity. For example, Figure 1 shows several patches of sky (top row) and water (bottom row) that are cropped from a fixed partition-based image segmentation. Although the images here depict only two different concepts (e.g., sky and water), we can notice a large variability in the appearances in terms of the color and texture features.

Although we can easily distinguish such concepts with variations due to our long accumulated knowledge, it is always difficult for a machine to do the same. The variation in these local patches can be effectively modeled by using supervised learning based classification techniques, such as the SVM [32]. In this context, the SVM will create statistical models from the training data, where an instance (e.g., local concept) in the training set is represented by a feature vector and contains category specific labels.

3.1 Multi-class SVM

In its basic formulation, the SVM is a binary classification method that constructs a decision surface and maximizing the inter-class boundary between the samples. Stated mathematically,

$$f(x) = \text{sign} \left(\sum_{i=1}^N \alpha_i y_i K(x_i, x) + b \right) \quad (1)$$

where $\mathbf{x} \in \mathbb{R}^d$ is an input vector, $\mathbf{x}_i$ is a training sample vector along with its label $y_i \in \{+1, -1\}^N$, b is a bias and K is a kernel function which maps the vectors into a higher dimensional space by the non-linear mapping function $\phi: \mathbb{R}^d \rightarrow \mathbb{R}^l$, where $l > d$ or l could even be infinity.

A number of methods have been proposed for the extension of the SVM to multi-class classification problems [33]. We utilize one such a method by combining all pairwise comparisons of binary SVM classifiers, known as *one-against-one* or pairwise coupling (PWC) [34]. The first step of this approach is to construct a training set of local semantic patches from individual image regions. We consider a semi-automatic approach for the semantic patch generation. In this approach, the training images are at first fixed partitioned equally into an even grid of non-overlapping regions. Due to the fixed partitioning scheme, some regions or patches would contain multiple local concepts. Hence, instead of considering all the patches from fixed partitioning, a subset of these are manually selected based on the criteria that the area of each region should correspond to a particular concept by at least 70-80%.

In order to perform the multi-class SVMs training based on the local concept categories, a set of L labels are assigned as $C = \{c_1, \dots, c_i, \dots, c_L\}$, where each $c_i \in C$ characterizes a local

concept category. The training set of the local patches is annotated manually with the concept labels in a mutually exclusive way. Hence, each patch is labeled with only one local concept category and is represented by a combination of color and texture moment-based features. The color feature is composed of the mean and the standard deviation of each channel in the *RGB* color space and the texture feature is composed of second order texture moments, such as energy, maximum probability, entropy, contrast, and inverse difference moment from a grey level co-occurrence matrix (GLCM) [36]. For SVM training, the initial input to the system is the feature vector set of the patches along with their manually assigned corresponding concept labels.

3.2 Image Encoding and Feature Representation

Images in the data set are annotated with local concept labels by partitioning each image I_j into an equivalent $r \times r$ grid of l vectors as $\{\mathbf{x}_{1j}, \dots, \mathbf{x}_{kj}, \dots, \mathbf{x}_{lj}\}$, where each $\mathbf{x}_{kj} \in \mathbb{R}^d$ is a d -dimensional combined color and texture feature vector. For each $\mathbf{x}_{kj}$, the local concept category probabilities are determined by the prediction of the multi-class SVMs as [34]

$$p_{ikj} = P(y=i | \mathbf{x}_{kj}), \quad 1 \leq i \leq L. \quad (2)$$

For example, Figure 2 shows a particular region in a segmented image and its probabilistic membership scores to different local concept categories. The probability or confidence scores of all categories now forms an L -dimensional vector for a region $\mathbf{x}_{kj}$ of I_j as

$$\mathbf{p}_{kj} = [p_{1kj} \cdots p_{ikj} \cdots p_{Lkj}]^T \quad (3)$$

Here, each p_{ikj} , $1 \leq i \leq L$, denotes the probability that a region $\mathbf{x}_{kj}$ belongs to the category $c_i \in C$. Based on the probability scores, the category label of $\mathbf{x}_{kj}$ is determined by $m = \arg \max_{1 \leq m \leq L} [p_{mk}]$ that is the label of the category c_m with the maximum probability score. Hence, the region $\mathbf{x}_{kj}$ is annotated with the label m and the entire image is thus represented as a two-dimensional index linked to the concept or localized semantic labels assigned for each region. Based on this encoding scheme, an image I_j is represented as a vector in a local semantic concept space as

$$\mathbf{f}_j = [f_{1j}, \dots, f_{ij}, \dots, f_{Lj}]^T \quad (4)$$

where each f_{ij} corresponds to the normalized frequency of a concept c_i , $1 \leq i \leq L$ in image I_j .

However, this representation scheme captures only a coarse distribution of the concepts and is analogous to the distribution of quantized color in a global color histogram. In the image domain, the concepts contain information from both perceptual and semantical perspectives. Some concepts are very likely to appear in all the images in a collection, whereas other concepts occur often in few images but relatively rare in the entire collection of images. For example, the semantic concept “*sky*” would occur in many outdoor images whereas the concept “*snow*” might be found only in images of a snowy mountain category in an image collection of natural scenery. The rationale is that the concepts that occur frequently in the entire collection have low information content. However, if a single image contains many occurrences of a concept then it is probably desirable to distinguish that image from other images in a collection. To consider this effect into account, a concept is given a weight

depending on its information content similar to the *tf-idf* weighting in IR [7,8]. In this context, for the *tf-idf* based scheme, the local weight is denoted as $L_{i,j} = f_{i,j} / \max_l f_{l,j}$, where $f_{i,j}$ is the raw frequency of occurrence of a local visual concept c_i or a local semantic concept sc_i in image I_j and the maximum is computed over all the concepts which are presented in I_j . The global weight G_i is denoted as inverse image frequency (e.g., inverse document frequency in text) as $G_i = \log_e (M/M_i) + 1$, $1 \leq i \leq L$, where M_i is the number of images in which concept c_i is found [8]. Finally, an element w_{ij} is expressed as the product of local and global weights. Based on the above weighting scheme, an image I_j can be represented as a vector in a concept space as

$$\mathbf{f}_j^{\text{Concept}} = [w_{1j} w_{2j} \cdots w_{ij} \cdots w_{Lj}]^T \quad (5)$$

where w_{ij} is the weight of the concept c_i . Here, the vector dimension equals to the number of local concept categories.

4 Query Expansion Based on Local Analysis

Query expansion based on local feedback and cluster analysis (commonly known as local analysis) has been one of the most effective methods for expanding queries in text retrieval domain [5,3,2]. Generally, the techniques based on local analysis expand a query based on the information from the top retrieved items (without any assistance from the user) for that query. In most expansion methods that uses the local analysis, there are five key stages. First, the original query is used to rank an initial set of documents. This set is then retrieved from disk and all the terms are extracted from those documents. The correlated terms are identified and ranked in order of their potential contribution to the query. The top ranked terms are re-weighted and appended to the query, and finally the reformulated query is re-issued and a final set of documents is ranked [2]. Before presenting our query expansion method in image domain, some basic terminologies are defined as follows:

Definition 1

For a given query image I_q , the set S_l of images retrieved is called the local image set. Also, the set $C_l \subseteq C$ of all distinct concepts $c_i \in C_l$ in the local image set S_l is called the local vocabulary of concepts.

Since, the correlated terms for an expansion are those present in the local cluster, we first need to generate such a cluster from S_l and thereafter from C_l . To generate the cluster, we rely on a local correlation matrix that is built based on the co-occurrence of concepts inside images.

Definition 2

Let $\mathbf{A}_l = [a_{uv}]$ be a $|C_l| \times |C_l|$ matrix in which the rows and columns are associated with the concepts in C_l . Each entry a_{uv} expresses a normalized correlation factor between concepts c_u and c_v as

$$a_{uv} = \frac{n_{uv}}{(n_u + n_v - n_{uv})} \quad (6)$$

where n_u be the number of images in S_l that contain the concept c_u , n_v be the number of images that contain the concept c_v , and n_{uv} be the number of the top retrieved images in S_l

that contain both concepts. The entry a_{uv} measures the ratio between the number of images where both c_u and c_v appear and the total number of images in S_l where either c_u or c_v appear and its value ranges to $0 \leq a_{uv} \leq 1$. If c_u and c_v have many co-occurrences in images, the value of a_{uv} increases and the images are considered to be more correlated.

The global version of this matrix, which is termed as the *connection matrix*, is utilized in a fuzzy information retrieval approach in [35]. The local correlation (connection) matrix $\mathbf{A}_l$ is created based on the frequency of co-occurrence of pairs of concepts in images and it does not take into account where the concepts are located in an encoded image. Since, the two concepts that occur relatively close in a neighborhood seem more correlated than the two concepts which occur far apart in an image. We provided such an example in the previous section for the concepts of sun and sky for outdoor images. Hence, it would be worthwhile to factor the distance between two concepts in the computation of their correlation factor.

To resolve this issue, the concept of a metric cluster was introduced in the text retrieval domain [5]. In this cluster generation process, the correlation factors are computed by considering the inverse of a distance between two terms. A document can be viewed as a one dimensional ordered list of terms where the distance between two terms is calculated by counting the number of words between them. For the image domain, the concepts are organized in a two-dimensional grid in encoded images. To measure the distance between two concepts, we need to rely on some sort of a neighboring relationship to count how many block units (maximally) they are apart by considering both the horizontal and vertical directions. Such a neighborhood relationship is defined as follows:

Definition 3

Let a two dimensional encoded image I_j be represented as a grid of P columns and Q rows of block units. Each block $r(x, y)$, $0 \leq x \leq P - 1$, $0 \leq y \leq Q - 1$, has a topological d -neighborhood and is mapped into a concept $c_i \in C_l$ of the local codebook as shown in Figure 3. The value of d can go up to a maximum of $P - 1$ or $Q - 1$ for horizontal or vertical directions. If we consider two blocks $r(x, y)$ and $r(u, v)$ for $0 \leq x, u \leq P - 1$, $0 \leq y, v \leq Q - 1$ are mapped to concepts c_i and c_j respectively. The distance $\text{dis}(c_i, c_j)$ between these two concepts is measured as

$$\text{dis}(c_i, c_j) = \max\{|u - x|, |v - y|\} \quad (7)$$

Hence, if c_i and c_j are 1-neighborhood apart then $\text{dis}(c_i, c_j) = 1$ and if c_i and c_j are in distinct images then, $\text{dis}(c_i, c_j) = \infty$.

For example, Figure 3 shows the distance between concepts c_i and c_j as $\text{dis}(c_i, c_j) = 3$ based on the definition above. Now, a local *concept-concept* metric correlation matrix $\mathbf{M}_l$ is defined as follows [5]:

Definition 4

$\mathbf{M}_l = [m_{uv}]$ is a $|C_l| \times |C_l|$ matrix. Each entry m_{uv} of $\mathbf{M}_l$ expresses a normalized metric correlation factor between concepts c_u and c_v as

$$m_{uv} = \frac{c_{uv}}{(|S(c_u)| \times |S(c_v)|)} \quad (8)$$

where $S(c_u)$ and $S(c_v)$ indicate the sets of the concepts c_u and c_v based on the coordinate positions that are mapped at different regions in encoded images of the local image set S_l . The metric correlation c_{uv} is calculated as

$$c_{uv} = \sum_{c_u \in S(c_u)} \sum_{c_v \in S(c_v)} \frac{1}{dis(c_u, c_v)} \quad (9)$$

Given the above definitions of the local connection matrix $\mathbf{A}_l$ and the metric matrix $\mathbf{M}_l$, we can use them to build local association and metric clusters respectively for the query expansion. Let $f_i(n)$ be a function that takes the i -th row and return the ordered set of n largest values a_{ij} from $\mathbf{A}_l$ or m_{ij} from $\mathbf{M}_l$, where j varies over the set of local concepts and $i \neq j$. Then $f_i(n)$ defines a local association or metric cluster around the concept c_i (as a blue star as shown in Figure 4). Here, the concept c_j (e.g., red star) is located within a neighborhood $f_i(n)$ associated with the concept c_i .

Now, the concepts that belong to clusters associated with the query concepts can be used to expand the original query. Often these neighbor concepts are correlated by the current query context [7]. The steps of the query expansion process based on the metric cluster are given in Algorithm 1. Similar steps can also be applied for the query expansion based on the correlation cluster where we just need to replace $\mathbf{M}_l$ with $\mathbf{A}_l$.

Algorithm 1 Query Expansion through Metric Clusters

- 1: Initialize a temporary expanded query vector that need to be added to the original vector of query image I_q , as $\mathbf{f}_q^e = [\hat{w}_{1_q} \hat{w}_{2_q} \cdots \hat{w}_{i_q} \cdots \hat{w}_{L_q}]^T$ where each $\hat{w}_{i_q} = 0$.
- 2: For an original query vector $\mathbf{f}_q^o = [w_{1_q} w_{2_q} \cdots w_{i_q} \cdots w_{L_q}]^T$ of I_q , perform initial retrieval by applying a similarity (e.g. cosine) matching function.
- 3: Consider, the top ranked K images as the local image set S_l and determine the local codebook C_l .
- 4: Construct, the local metric matrix $\mathbf{M}_l$ based on equations (8) and (9) at query time.
- 5: **for** $i = 1$ to L **do**
- 6: **if** $w_{i_q} > 0$ **then**
- 7: Consider the i -th row in the metric matrix $\mathbf{M}_l$ for concept c_i .
- 8: Return $f_i(n)$, the ordered set of n largest values m_{ij} , where $i \neq j$, therefore $c_j \in C_l - \{c_i\}$.
- 9: **for each** c_j **do**
- 10: Add and re-weight the corresponding element in query vector as $\hat{w}_{j_q} = w_{j_q} - ((w_{i_q} - 0.1) \times k/n)$, where k is the position of c_j in the rank order.
- 11: **end for**
- 12: **end if**
- 13: **end for**
- 14: Obtain the re-formulated or modified query vector as $\mathbf{f}_q^m = \mathbf{f}_q^e + \mathbf{f}_q^o$.
- 15: Perform the retrieval with the modified query vector $\mathbf{f}_q^m$.
- 16: Continue the process, i.e., steps 3 to 15 until no more changes are noticed.

Based on Step 10 of the Algorithm 1, weights are assigned in such a way that a top ranked concept in a ordered set gets the largest weight value than the next one in the set. After this expansion process, new concepts may have been added to the original query based on Step

10 and the weight of an original query concept may have been modified had the concept belonged to the top ranked concepts based on Step 14 of the algorithm.

The main drawback of any local analysis based query expansion approaches is that it is computationally expensive to construct the local matrix at query time. As a result, the query modification process might somehow reduce the interactive nature of a system. However, this query expansion approach would be of great assistance for the application domain where effectiveness is the major issue then efficiency, such as the medical image collections.

5 Query Expansion Based on Global Analysis

This section presents another query expansion approach in the concept space by generalizing upon an expansion model by Qiu and Frei [6] based on global analysis in the text retrieval domain. In global analysis, the query is expanded using information from the entire collection [4,6]. Here, a thesaurus (matrix) is constructed based on a *term-term (concept-concept)* correlation or similarity relationship in the entire collection instead of a local set in IR. The global analysis techniques are computationally intensive but the computations are done off-line once per database. The only component done on a per query basis is the actual query expansion itself. The query expansion methods tend to add a term (concept) from a thesaurus when it is strongly related in terms of similarities or correlations to one of the query terms. However, unlike the query expansion approach in previous section, it needs not to be limited to the terms that cluster together with the query terms.

A query expansion model for the text retrieval using a similarity thesaurus is presented in [6]. They obtain a relatively large improvement (15-30%) in performance by adding terms that have the largest similarity to the entire query concept, rather than individual terms. To determine the ranking of each of the selected terms, the approach in [6] relies on a term-term similarity matrix. It is based on how the terms in the collection are indexed by documents, i.e., for each term there is a document vector space. The query expansion terms are weighted according to the similarity between these terms and the query.

By generalizing their approach in our image framework based on the concept model, each concept c_i , $i \in \{1, \dots, L\}$ can be associated with a vector $\mathbf{c}_i = \langle w_{i1}, \dots, w_{ij}, \dots, w_{iM} \rangle$ where M is the number of images in the collection. The element w_{ij} is the weight for the concept c_i in image I_j , which is computed in a rather distinct form as [6]:

$$\bar{w}_{ij} = \frac{\left(\frac{f_{ij}}{\max_j(f_{ij})} \right) ikf_j}{\sqrt{\sum_{l=1}^M \left(\frac{f_{il}}{\max_l(f_{il})} \right)^2 ikf_l^2}} \quad (10)$$

where f_{ij} be the frequency of occurrence of the concept c_i in the image I_j and $\max_j(f_{ij})$ computes the maximum frequency of c_i under all images in the collection. Further, the inverse concept frequency ikf_j for I_j , (e.g., analogous to the inverse image (document)

frequency), is computed as $ikf_j = \log \frac{L}{k_j}$, where k_j be the number of distinct concepts in the I_j . After generating the concept vectors, a similarity matrix $\mathbf{S}_{L \times L} = [s_{u,v}]$ is built through the computation of each element $s_{u,v}$ as the normalized cosine relationship or dot product between two concept vectors $\mathbf{c}_u$ and $\mathbf{c}_v$ as

$$s_{u,v} = c_u \cdot c_v = \sum_{j=1}^M \bar{w}_{uj} * \bar{w}_{vj} \quad (11)$$

Although, the construction of the matrix $\mathbf{S}$ is prohibitively difficult for large collections, it needs to be computed only once and can be computed off-line. The only component done on a per query basis is the utilizing the matrix elements for the query modification.

Given the global similarity matrix $\mathbf{S}$, the query expansion is performed in the following four main steps [6]:

- We map an original query vector $f_q^o = [\bar{w}_{1_q} \cdots \bar{w}_{i_q} \cdots \bar{w}_{L_q}]^T$ from the concept space to a space as the virtual query concept vector c_q^* which is computed as

$$c_q^* = \sum_{c_u \in I_q} \bar{w}_{uq} c_u \quad (12)$$

where $\bar{w}_{uq}$ is the weight of the concept c_u in I_q or in other words, $\bar{w}_{uq} c_u$ expresses the importance of the concept vector c_u for the query.

A temporary expanded query vector is also initialized with each element has a zero value as $f_q^e = [\bar{w}_{1_q} \bar{w}_{2_q} \cdots \bar{w}_{i_q} \cdots \bar{w}_{L_q}]^T$, that needs to be added to the original vector f_q^o .

- Based on the similarity matrix $\mathbf{S}$, a similarity between each concept $c_v \in C$ and the query image I_q is computed as

$$\text{sim}(I_q, c_v) = c_q^{*T} \cdot c_v = \left(\sum_{c_u \in I_q} \bar{w}_{uq} \cdot c_u \right)^T \cdot c_v = \sum_{c_u \in I_q} \bar{w}_{uq} \cdot (c_u^T c_v) \quad (13)$$

where $(c_u^T c_v) = s_{uv}$ is the similarity between the two concepts c_u and c_v that is pre-computed as defined in (11).

- Based on the $\text{sim}(I_q, c_v)$ value, each concept $c_v \in C$ is ranked and the top n ranked concepts are selected for query expansion. After selecting the additional search concepts, we need to assign weight to them so that they can be added to the original query vector. The weight of each concept c_v is calculated as [6]:

$$\text{weight}(I_q, c_v) = \text{sim}(I_q, c_v) / \left(\sum_{c_u \in I_q} \bar{w}_{uq} \right) \quad (14)$$

where $0 \leq \text{weight}(I_q, c_v) \leq 1$.

- Finally, we obtain the re-formulated or modified query vector as $f_q^m = (f_q^e + f_q^o)$. For $f_q^e = [\bar{w}_{1_q} \bar{w}_{2_q} \cdots \bar{w}_{v_q} \cdots \bar{w}_{L_q}]^T$,

$$\hat{w}_{vq} = \begin{cases} \text{weight}(I_q, c_v), & \text{if } c_v \text{ belongs to top } n \text{ ranked concepts;} \\ 0, & \text{otherwise;} \end{cases}$$

Finally, f_q^m is used as the reformulated query vector to retrieve images in the next iteration.

With the above steps, the query and the concepts most similar to the virtual query concept are classified in the same cluster. After this expansion process, new concepts may have been added to the original query and the weight of an original query concept may have been modified had the concept belonged to the top ranked concept based on (14).

To illustrate the above process, Figure 5 shows a virtual query concept vector c_q^* (e.g., the red star) and its relationships with other close concept vectors in an Euclidean space. The closeness is defined by the similarity calculation based on (13) and the closer two linked concepts (based on solid or dashed line) are to each other, the more similar they are. From Figure 5, it is clear that concepts c_c and c_d are closer to the individual query concepts $c_u, c_v \in I_q$ (e.g., the gray stars) respectively. However, in terms of the whole query concept (i.e., the virtual query vector c_q^*), concepts c_d and c_b are qualified as the closer one, if we consider only two concepts out of all other concepts in the vocabulary. It implies that, the concepts selected here for the query expansion might be distinct from the ordinary global or local analysis method.

6 Experiments

To evaluate the effectiveness of the proposed retrieval framework based on query expansion, exhaustive experiments were performed in two different image collections. Since these collections have greatly different characteristics, together they offer a diverse set of data with complementary nature. The first collection comprises a varying cross-section of 4000 still natural images taken from locations around the world. It is basically a subset of the COREL dataset and the IAPR TC-12 benchmark that was created under the Technical Committee 12 (TC-12) of the International Association of Pattern Recognition (IAPR) [38]. The benchmark is currently used for ad-hoc photographic retrieval task in ImageCLEF 2. The domain of the images in this collection is very generic in that it covers a wide range of daily life situations with many different images of similar visual content but with varying illumination, viewing angle and background. This makes it more challenging for the successful application of image retrieval techniques. To evaluate retrieval performance, we have manually categorized 4000 images from this collection into three levels and 26 different disjoint categories at a global level as shown in Figure 6 with gray-level leaf nodes. The categories include group photos in different settings and closeup of individual persons; indoor and different outdoor natural landscapes or manmade objects; activities, such as different sports and so on. This category information is served as ground truths for the experimental purposes.

The second image collection consists of 5000 biomedical images in 30 manually assigned disjoint global categories, which is a subset of a larger collection of six different data sets used for the medical image retrieval task in ImageCLEFmed 2007 [39]. In this collection, the images are classified into three levels as shown in Figure 7. At the first level, images are categorized according to modality (e.g., X-ray, CT, MRI, etc.). At the next level, each of the

²<http://www.imageclef.org>

modalities is further classified according to the examined body parts (e.g., head, chest, etc.) and finally it is further classified by orientation (e.g., frontal, sagittal, etc.) or distinct visual observation (e.g., CT liver images with large blood vessels). The disjoint categories are selected only from the leaf nodes (gray in color) to create the ground-truth data set. The categories are selected based on analyzing the visual and some mixed-mode query topics during the last three years (2005, 2006, and 2007) of ImageCLEFmed [39] retrieval campaign. The contents of this collection represent a broad and significant body of medical knowledge and as the entire collection contains a variety of imaging modalities, image sizes, and resolutions, similarity searching based on current CBIR techniques is difficult.

6.1 Training for the SVM

The selection of the training image set that represent all the categories in the collections well is critical for the modeling of the local concepts. The training sets used to generate the SVM model files for the concepts consist of only 5% of the images from each collection with all the categories in equal portions, resulting in a total of 400 images for the photographic collection and 500 images for the medical collection. The remaining images (e.g., 90% of the collections) were used as ground-truth for all other testing and evaluation purposes. To generate the local patches, each image in the training set is first partitioned into an 8×8 grid generating 64 non-overlapping regions. Only the regions that conform to at least 70-80% of a particular concept category out of the pre-defined categories are selected and labeled with the corresponding category label. We manually assigned 22 local concept classes, such as blue sky, cloud, sand, rock, snow, and so on for the photographic collection. The concepts are selected in such a way that images in the same category share many common concepts, such as, rock, sky, grass, etc. As well as images from different categories also contain common concepts, such as blue sky in most of sub-categories under the outdoor category. For the medical image collection, local concepts are selected as the ones that exhibit some meanings to the physicians with distinct visual appearances. We defined 30 concept categories from the patches, such as X-ray lung tissue and bone, normal and abnormal skin, microscopic images of different color and texture patterns, and so on. Table 1 and 2 show the statistics of each local concept with the number of regions to present them in the training sets for the photographic and the medical collection respectively.

For the SVM training, we utilized both the radial basis function (RBF) and the polynomial kernels. There are two tunable parameters while using RBF kernels: C and γ . It is not known beforehand which values of C and γ are the best for the classification problem at hand. Hence, a 10-fold cross-validation (CV) is conducted. Basically pairs of (C, γ) are used and the one with the best CV accuracy is picked. We also experimented with the polynomial kernel of degree 1 and 2 with $C = 100$. However, the best accuracies are achieved by the RBF kernel in both the collections as shown in Table 3. Hence, after finding the best values of parameters C and γ of the RBF kernels, they are utilized for the final training to generate the SVM model files for the local concepts. We utilized the *LIBSVM* software package [37] for the implementation of the multi-class SVM classifiers.

7 Results

For a quantitative evaluation of the retrieval results, we selected all the images in the individual collections as query images and used “*query-by-example*” as the search method, where a query is specified by providing an example image to the system. A retrieved image is considered to be a correct match if it is in the same category (based on the ground truth) as the query image. Precision (percentage of retrieved images that are also relevant) and recall (percentage of relevant images that are retrieved) are used as the basic evaluation measure of retrieval performances [7]. The average precision and recall are calculated over all the queries to generate the precision-recall (PR) curves in different settings.

We at first evaluated the effectiveness of the concept-based feature representation scheme compared to the two well known low-level feature representation schemes, the MPEG-7 based Color Layout Descriptor (CLD) and Edge Histogram Descriptor (EHD) [40]. The CLD represents the spatial layout of the color images in a very compact form and the EHD represents spatial distribution of edges as a global shape feature. Fig. 8 shows the PR curves based on the Euclidean similarity matching in the different feature spaces. By analyzing the Fig. 8, we can observe that proposed concept-based feature representation schemes performed much better for both the collections compared to the CLD and EHD based features in terms of precision at each recall level. The better performances are expected as the concept features are more semantically oriented that exploits the domain knowledge of the collections at a local level.

Fig. 9 shows the PR graphs of the automatic query expansion approaches based on the local analysis (as described in Section 4) in the concept feature space after one iteration of feedback information. For the experimental evaluation purpose, we considered the top $K = 20$ retrieved images from the previous iteration as the positive local feedback or relevant images for the next iteration. Most online search engines like Google, Yahoo, and Altavista display 20 images by default when we search for images. For query expansion, three additional concepts from the local clusters are selected for each query concept. It is clear from Fig. 9 that the performances were improved for both the collections for the local analysis based query expansion approaches as compared to the case when no query expansion is performed on the concept feature space. The performance of the query expansion based on the local analysis of metric cluster (e.g., “QE Local (Metric)”) and correlation cluster (e.g., “QE-Local (Conn)”) are almost similar with no significant differences at the PR curves for both the collection. In general, we achieved around 5-15% performance improvement in precision at each recall level for the medical collection and similar improvement in the photographic collection at the early (e.g., before 0.6) recall levels. Overall, from the PR graphs of the two different image collections, we can conjecture that the query expansion based on local analysis indeed increases the performances in image domain and the results are consistent to the performances in the text domain as reported previously [5].

To check the consistency of the query expansion approaches based on the local analysis by employing both correlation and metric cluster, we also considered 5 iterations of feedback and compared the performance by considering the average precision within the top 20 (e.g., $P(20)$) retrieved images as shown in Figure 10. From Fig. 10, we observe one common trend that the precision increases during the first 2-3 iterations for both approaches under the medical and photographic collections. After that, the improvement subsides to the previous level or even lower due to the system convergence. Hence, it concludes that the performance of the query expansion approach is consistent from one iteration to another.

Finally, Fig. 11 shows the PR graphs of the automatic query expansion approach based on the global analysis (as described in Section 5) in the concept feature space. For each query, five additional concepts are selected for the expansion. Here, we only need to exploit the information from the global cluster and do not require any feedback information. The query expansion approach based on the global analysis (e.g., “QE Global (Sim)”) shows improved precisions in terms recalls at different levels when compared to the precisions without using any expansion (e.g., “Concept”). It also performed almost equally well when compared to the performances based on the local analysis based query expansion. Although the performances for the query expansion based on local analysis is slightly better due to the query-specific nature of the expansion algorithm as mentioned before. The performance is also compared with a query expansion approach based on performing similar algorithm of local query expansion in a global correlation cluster (e.g., “QE Global (Conn)”). The slightly

improved performances for “QE Global (Sim)” in both the collection validates the initial assumption that instead of expanding the terms (concepts) based on each query term, it is effective to select expanded terms based on a single virtual query vector only.

The results were also evaluated by comparing the overall average precisions among the average precision on top 20, 30, 50, and 100 retrieved images with global query expansion (“QE Global (Sim)”) and without any expansion with different number of added concepts. From Table 4 and Table 5, it can be observed that the improvement by the expanded queries increases when the number of additional concepts increases. But after adding certain number of concepts, the performance started to decrease and at a point it even became lower than the original query. This could be explained by the fact that fewer selected concepts contain better information and when the number gets larger, the expansion approach might add the concepts that are not related to the relevant images.

For a qualitative evaluation, Fig. 12, and Fig. 13 show the snapshots of the retrieval results, in which the retrieved images are ranked and displayed in the descending order of their similarity scores from the top left to the bottom right (e.g., top most 15 similar images), with the image in the left panel being the query image. In this example, the query image belongs to the “Skin Lesion” category of the medical collection. From Figure 12, we observe that there are only 6 images returned which belong to the same category as the query image based on the concept feature space. Whereas, for the query expansion based on local analysis in correlation cluster, the system returned 14 images out of top 15 as shown in Fig. 13. Based on the above results, we can see that the query expansion approach provides better precision in top ranked images compared to the approach when no query expansion is used.

8 Conclusions

We have investigated automatic query expansion techniques in the CBIR domain, inspired by the ideas from the text retrieval domain in IR. The decision to rely on techniques from the text retrieval domain and generalize them to our concept feature-based image representation scheme has proven effective, as demonstrated empirically on two largely different image collections. The proposed expansion techniques exploit the correlations among the concepts based on both local and global analysis. Due to the nature of the concept-based image representation schemes, there always exists enough correlations among the concepts. Hence, exploiting this property reduces the concept-mismatch problem and removes the burden from the users to provide feedback information to a system. For the feature representation, we limited our approaches by modeling only the intermediate level local concepts. It would be more effective if specific objects can be identified in large collections irrespective of their variations and occlusions. However, the main focus in this work is to explore the query expansion and the consequent retrieval to prove its effectiveness in the image domain. As object recognition techniques mature to a certain level in the future, our approaches could be easily extended to a higher level concept-based representation.

Acknowledgments

This research is supported by the Intramural Research Program of the National Institutes of Health (NIH), National Library of Medicine (NLM), and Lister Hill National Center for Biomedical Communications (LHNCBC). The authors would like to thank the CLEF organizers 3 for making the databases available for the experiments.

³<http://www.clef-campaign.org/>

References

1. Salton G, Buckley C. Improving Retrieval Performance by Relevance Feedback. *J American Soc for Info Scie.* 1999; 41(4):288–297.
2. Xu J, Croft WB. Improving the effectiveness of information retrieval with local context analysis. *ACM Trans on Info Sys.* 2000; 18:79–112.
3. Xu, J.; Croft, WB. Query Expansion Using Local and Global Document Analysis. *Proc 19th Annual Int'l ACM SIGIR Conf on Research and Develop in Info Retrieval*; 1996. p. 4-11.
4. Crouch CJ. An approach to the automatic construction of global thesauri. *Info Process and Management.* 1990; 26(5):629–640.
5. Attar R, Fraenkel AS. Local feedback in full-text retrieval systems. *J ACM.* 1977; 24(3):397–417.
6. Qiu, Y.; Frei, HP. Concept Based Query Expansion. *Proc 16th Int ACM SIGIR Conf on R&D in Info Retrieval*; June; 1993. p. 160-169.
7. Yates, RB.; Neto, BR. *Modern Information Retrieval.* 1st. Addison Wesley; 1999.
8. Salton, G.; McGill, MJ. *Introduction to Modern Information Retrieval.* McGraw-Hill; 1983. Computer Science
9. Rocchio, JJ. Relevance feedback in information retrieval. In: Salton, G., editor. *SMART retrieval system -experiments in automatic document processing.* Prentice Hall; Englewood Cliffs, NJ: 1971. p. 313-323.
10. Ruthven I, Lalmas M. A survey on the use of relevance feedback for information access systems. *Knowledge Eng Review.* 2003; 18(2):95–145.
11. Zhou XS, Huang TS. Relevance feedback for image retrieval: a comprehensive review. *Multimedia Syst.* 2003; 8(6):536–544.
12. Rui, Y.; Huang, TS.; Mehrotra, S. Content-based image retrieval with relevance feedback in MARS. *Proc IEEE Int'l Conf Image Process*; 1997. p. 815-818.
13. Tong, S.; Chang, E. Support vector machine active learning for image retrieval. *Proc 9th ACM Int'l Conf Multimedia*; 2001. p. 107-118.
14. Ishikawa, Y.; Subramanya, R.; Faloutsos, C. MindReader: Querying Databases Through Multiple Examples. *Proc 24th Int'l Conf Very Large Databases (VLDB '98)*; New York. 1998. p. 24-27.
15. Zhang H, Chen Z, Li M, Su Z. Relevance Feedback and Learning in Content-Based Image Search. *World Wide Web: Internet and Web Information Systems.* 2003; 6:131–155.
16. Hu, YC.; Zhu, X.; Zhang, H.; Yang, Q. A unified framework for semantics and feature based relevance feedback in image retrieval systems. *Proc ACM Int'l Conf on Multimedia*; 2000. p. 31-37.
17. Smeulders A, Worring M, Santini S, Gupta A, Jain R. Content-Based Image Retrieval at the End of the Early Years. *IEEE Trans Pattern Anal Machine Intell.* 2000; 22(12):1349–1380.
18. Liua Y, Zhang D, Lu G, Ma WY. A survey of content-based image retrieval with high-level semantics. *Pattern Recognition.* 2007; 40:262–282.
19. Datta R, Joshi D, Li J, Wang JZ. Image Retrieval: Ideas, Influences, and Trends of the New Age. *ACM Computing Surveys.* 2008; 40(2):1–60.
20. Bosch A, Muoz X, Mart R. Which is the best way to organize/classify images by content? *Image and Vision Computing.* 2007; 25(6):778–791.
21. Rahman, MM.; Bhattacharya, P. Image Retrieval with Automatic Query Expansion Based on Local Analysis in a Semantical Concept Feature Space. *Proc ACM International Conference on Image and Video Retrieval 2009 (CIVR 2009)*; 8-10 July 2009; Greece. Article No. 20
22. Zhu L, Zhang A, Rao A, Srihari R. Theory of keyblock-based image retrieval. *ACM Trans Info Syst.* 2002; 20(2):224–257.
23. Lim JH. Building visual vocabulary for image indexation and query formulation. *Pattern Anal Applications (Special Issue on Image Indexation).* 2001; 4:125–139.
24. Jing F, Li M, Zhang HJ, Zhang B. An efficient and effective region-based image retrieval framework. *IEEE Trans on Image Process.* 2004; 13:699–709.
25. Vogel J, Schiele B. Semantic Modeling of Natural Scenes for Content-Based Image Retrieval. *Internat J Comput Vis.* 2007; 72(2):133–157.

26. Shi, R.; Feng, H.; Chua, TS.; Lee, CH. An adaptive image content representation and segmentation approach to automatic image annotation. *International Conference on Image and Video Retrieval (CIVR)*; 2004. p. 545-554.
27. Rahman MM, Bhattacharya P, Desai BC. A Unified Image Retrieval Framework on Local Visual and Semantic Concept-Based Feature Spaces. *Journal of Visual Communication and Image Representation*. 2009; 20:450–462.
28. Wang XJ, Ma WY, Li X. Exploring Statistical Correlations for Image Retrieval. *Multimedia Systems (MMSJ)*. 2006; 11(4):340–351.
29. Zhou XS, Huang TS. Unifying Keywords and Visual Contents in Image Retrieval. *IEEE Multimedia*. 2003; 9(2):23–33.
30. Zhang, HJ.; Su, Z. Improving CBIR by Semantic Propagation and Cross-Mode Query Expansion. *Proc International workshop on MultiMedia Content-Based Indexing and Retrieval (MMCBIR'01)*; 2001. p. 83-86.
31. Jeon, J.; Lavrenko, V.; Manmatha, R. Automatic Image Annotation and Retrieval using Cross-Media Relevance Models. *Proc 26th annual international ACM SIGIR conference on Research and development in informaion retrieval*; 2003. p. 119-126.
32. Burges JC. A Tutorial on Support Vector Machines for Pattern Recognition. *Data Mining and Knowledge Discovery*. 1998; 2:121–167.
33. Duan, K.; Keerthi, SS. Which is the best multiclass SVM method? An empirical study. *Proceedings of the LNCS. 3541 Springer, Sixth International Workshop on Multiple Classifier Systems*; Seaside, CA, USA. Jun. 13-15, 2005; 2005.
34. Wu TF, Lin CJ, Weng RC. Probability Estimates for Multi-class Classification by Pairwise Coupling. *J Machine Learn Res*. 2004; 5:975–1005.
35. Yasushi O, Tetsuya M, Kiyohiko K. A fuzzy document retrieval system using the keyword connection matrix and a learning method. *Fuzzy Sets and Systems*. 1991; 39(2):163–179.
36. Haralick RM, Shanmugam, Dinstein I. Textural features for image classification. *IEEE Trans Syst Man Cybernetics*. 1973; 3:610–21.
37. Chang, CC.; Lin, CJ. LIBSVM : a library for support vector machines. 2001. Software available at <http://www.csie.ntu.edu.tw/~cjlin/libsvm>
38. Grubinger, M.; Clough, P.; Müller, H.; Deseleers, T. The IAPR-TC12 benchmark: A new evaluation resource for visual information systems. *Proc Int'l Workshop on Language Resources for Content-Based Image Retrieval*; 2006. p. 1323
39. Müller, H.; Deselaers, T.; Kim, E.; Kalpathy, C.; Jayashree, D.; Thomas, M.; Clough, P.; Hersh, W. Overview of the ImageCLEFmed 2007 Medical Retrieval and Annotation Tasks. 8th Workshop of the Cross-Language Evaluation Forum (CLEF 2007), *Proceedings of LNCS*; 2008.
40. Chang SF, Sikora T, Puri A. Overview of the MPEG-7 standard. *IEEE Trans Circ Syst Video Technology*. 2001; 11:688–695.

Fig. 1. Sample patches to represent concepts “sky” and “water” in natural photographic images

Fig. 2.
Example of probabilistic membership or class confidence scores of a region x_k of a medical image for different concept categories.

Fig. 3.

An example of the distance between concepts c_i and c_j in an encoded image in the right side of the original image in the left based on the neighborhood relationship as defined in Definition 3.

Fig. 4. Concept c_j as a neighbor of the concept c_i based on a local cluster

Fig. 5. Relationships between local concepts and virtual query concept in an Euclidean space

Fig. 6.
Classification structure of the photographic collection.

Fig. 7.
Classification structure of the medical image dataset.

(a) Medical Collection

(b) Photographic collection

Fig. 8. Precision-Recall graphs for the retrieval results on different feature spaces in the medical and photographic collections

Fig. 9. Precision-Recall graphs for the retrieval results of query expansion based on the local analysis

Fig. 10. Accuracy comparison with different number of iterations by considering the average precision within the top 20 retrieved images

Fig. 11. Precision-Recall graphs for the retrieval results of query expansion based on the global analysis

Fig. 12. Top 15 most similar images for a query image of “skin lesion” category in the left panel based on performing similarity search with the concept feature

Fig. 13. Top 15 most similar images for a query image of “skin lesion” category in the left panel based on performing similarity search with query expansion on local correlation cluster

Table 1
Statistics of the Training Set for Local Semantic Concepts

Concept	# Regions	Concept	# Regions
Cloud	250	Blue Sky	250
Sea Water	300	Running Water	320
Sand	200	Soil	250
Rock	200	Brick	200
Grass	250	Green Leaves	200
Pavement	220	Court	270
Human Skin	30	Animal Skin	210
Crowd	260	Textured Cloth	200
Floral Red	245	Floral White	200
Floral Orange	290	Yellow Sun	280
White Snow	290	Black Background	280

Table 2
Statistics of the Training Set for Local Concepts

Concept	#	Concept	#
CT-Tissue-Brain	400	CT-Tissue-Abdomen	380
CT-Tissue-Lung	350	EC-Tissue-Gastro	300
CT-Bone-Whole	350	CT-Bone-Corner	300
Skin-Normal	420	Skin-Melanoma	370
Dental-White	330	Tissue-Mouth	310
MRI-Tissue-Leg	320	MRI-Tissue-Brain	350
Xray-Tissue-Lung	400	Xray-Hepatic-Veins	450
Xray-Chest-bone	260	Micro-Infection	200
US-Grey-Texture	445	Doppler-color	400
Micro-Blood-Blue	490	Histo-Blue	480
Micro-Pink-Bacteria	520	Micro-Tissue-Muscle	500
Photo-Tissue-Brain	520	Photo-Tissue-Cardiac	500
Backgrnd-White	500	Backgrnd-Grey	510
Backgrnd-Blue	350	Backgrnd-Brown	550
Backgrnd-Black	350	Line-Drawing	550

Table 3

Cross Validation Accuracy of local concept classification

Training Set	Kernel	C	γ	Degree	Accuracy
Photographic	RBF	200	0.3		77.68%
Photographic	Polynomial	100		1	73.02%
Photographic	Polynomial	100		2	73.14%
Medical	RBF	200	0.5		81.40%
Medical	Polynomial	100		1	78.76%
Medical	Polynomial	100		2	79.41%

Table 4
Improvement using expanded queries (Medical Collection)

# additional concepts	3	5	7	10
Without QE	0.3532	0.3532	0.3532	0.3532
With QE	0.3766	0.3845	0.3706	0.3478
Improvement	+ 6.62%	+ 8.86 %	+ 4.90 %	-1.50%

Table 5
Improvement using expanded queries (Photographic Collection)

# additional concepts	3	5	7	10
Without QE	0.2913	0.2913	0.2913	0.2913
With QE	0.3050	0.3092	0.2934	0.2908
Improvement	+ 4.71%	+ 6.14%	+ 0.72%	-0.17%