**Biomedical image representation
approach using visualness and spatial
information in a concept feature
space for interactive region-of-
interest-based retrieval**

Md. Mahmudur Rahman
Sameer K. Antani
Dina Demner-Fushman
George R. Thoma

Biomedical image representation approach using visualness and spatial information in a concept feature space for interactive region-of-interest-based retrieval

Md. Mahmudur Rahman,^{a,*} Sameer K. Antani,^b Dina Demner-Fushman,^b and George R. Thoma^b

^aMorgan State University, Computer Science Department, Calloway 308, 1700 E Cold Spring Lane, Baltimore, Maryland 21251, United States

^bU.S. National Library of Medicine, National Institutes of Health, 8600 Rockville Pike, Bethesda, Maryland 20894, United States

Abstract. This article presents an approach to biomedical image retrieval by mapping image regions to local concepts where images are represented in a weighted entropy-based concept feature space. The term “concept” refers to perceptually distinguishable visual patches that are identified locally in image regions and can be mapped to a glossary of imaging terms. Further, the visual significance (e.g., visualness) of concepts is measured as the Shannon entropy of pixel values in image patches and is used to refine the feature vector. Moreover, the system can assist the user in interactively selecting a region-of-interest (ROI) and searching for similar image ROIs. Further, a spatial verification step is used as a postprocessing step to improve retrieval results based on location information. The hypothesis that such approaches would improve biomedical image retrieval is validated through experiments on two different data sets, which are collected from open access biomedical literature. © 2015 Society of Photo-Optical Instrumentation Engineers (SPIE) [DOI: [10.1117/1.JMI.2.4.046502](https://doi.org/10.1117/1.JMI.2.4.046502)]

Keywords: medical image representation; content-based image retrieval; interactive retrieval; region-based search; image concepts; biomedical literature.

Paper 15003PR received Jan. 7, 2015; accepted for publication Nov. 23, 2015; published online Dec. 30, 2015.

1 Introduction

Literature-based informatics techniques are essential for managing the rapidly increasing volume of information in the biomedical domain. Medical researchers and clinicians routinely use online databases such as MEDLINE¹ via PubMed² search engine to enable easier access of relevant biomedical bibliographic citations based on keyword search on different fields such as the main text, author, and date. Given the rapid growth of scientific discovery, it is increasingly difficult to locate information at right place and right time within large volumes of literature. Literature search refers to the process in which a wide variety of users, including patients and their families, researchers, general practitioners and clinicians, and practitioners with specific expertise such as radiologists, use tools to search for literature relevant to their individual needs. Many of these existing tools focus on the full-text of the article, its categories, or keywords, where text-based approaches to retrieval of biomedical literature have been well researched.³

Scientific articles convey information using multiple and distinct modalities, including text and images. For example, authors of journal articles frequently use illustrations (e.g., charts, graphs, maps, diagrams, color, or gray-scale photographs) to elucidate the text, or to highlight special cases as region-of-interests (ROIs). These images often convey essential information in context and can help uncover hidden visual patterns not contained in the article’s text. For example, an author might report on the size and shape of some nodules depicted in

a patient’s lung computed tomography (CT) scan as illustration, using the visual characteristics of these regions as a basis for a possible cancer diagnosis. A picture is worth a thousand words, and even more so in medical domain as images of diverse modalities constitute an important source of anatomical and functional information for the diagnosis of diseases, research, education.^{4,5} Overall, biomedical literature incorporates approximately 100 million figures, whereas the biomedical open access literature of PubMed Central (PMC) of National Library of Medicine (NLM) alone contained almost two million images in 2014.^{6,7} Due to the continued growth of the health care industry, it poses significant technical challenges to transform of massive volumes of image and text data from biomedical articles into useful information and actionable knowledge in the form of effective and efficient search process.^{8,9} Until now, little attention is devoted to the use of images in the articles as the meaning of images cannot be understood by analyzing their content alone. However, only text information is sometimes insufficient in determining the usefulness of a publication. The importance of illustrations in scientific publications is well established. In a survey of information needs of researchers and educators, Sandusky and Tenopir¹⁰ found that scientific journal-article components, such as tables and figures, are often among the first parts of an article scanned or read by researchers.

Some retrieval systems in the medical domain, such as BioText, Goldminer, and Yale Image Finder (YIF), search for images within a collection of biomedical articles commonly represented and retrieve them according to their collateral text, such as captions, title, abstract, etc. For example, the BioText¹¹ search

*Address all correspondence to: Md Mahmudur Rahman, E-mail: md.rahman@morgan.edu

engine searches over 300 open access journals and retrieves figures as well as text. BioText uses the Lucene¹² text search engine to search full-text or abstracts of journal articles, as well as image and table captions. Yottalook¹³ allows multilingual searching to retrieve information (text or medical images) from the Web and journal articles. The goal of the search engine is to provide information to clinicians at the point of care. Other related work includes the Goldminer search engine that retrieves images by searching figure captions in peer-reviewed journal articles appearing in the Radiological Society of North America journals *RadioGraphics* and *Radiology*. It maps keywords in figure captions to concepts from the Unified Medical Language System (UMLS)¹⁴ Metathesaurus. The YIF¹⁵ searches text within images, captions, abstracts, and titles to retrieve images from biomedical journal articles. YIF uses optical character recognition to recognize text in images in both landscape and portrait modes. However, the above systems do not use image features to find similar images for biomedical information retrieval (IR).

Given that images are such a crucial source of information within the biomedical domain, using visual image features (e.g., color, texture, shape, edge, etc.) in content-based image retrieval (CBIR) has been gaining popularity.^{16,17} The goal of CBIR is to obtain from a possibly very large image database those images that are similar in content (e.g., color, texture, shape, edge, etc.) to an image of interest (the query image). Datta et al.¹⁶ and Müller et al.¹⁷ provide an exhaustive overview of existing image classification and retrieval approaches in general and medical domains, respectively. During the last decade, several medical CBIR prototypes have been proposed and developed.^{17,18,19} The majority of these systems are developed around a specific imaging modality and retrieval methods in these systems are task-specific. For instance, the ASSERT system¹⁸ is designed for high-resolution computed tomography images of the lung and the SPIRS system¹⁹ for digitized x-rays of the spine. There are only a few systems that have the goal of creating CBIR systems for heterogeneous image collections.^{20–22}

The importance of medical illustrations in journal articles has also motivated active research in image retrieval within the yearly ImageCLEF medical image retrieval tasks²³ and by individual researchers. However, after almost a decade of intensive research in medical image retrieval, progress has been slow due to the inability of image-processing algorithms to automatically identify the content of images in the manner that IR and extraction systems have been able to do so with text. Biomedical image collections present unique challenges, where subtle differences determine retrieval accuracy between otherwise highly similar images. For example, a posteroanterior (PA) chest x-ray of a tuberculosis patient appears overall very similar to a PA chest x-ray of a patient with interstitial lung disease, but retrieving both these images is incorrect for any request more specific than for a chest x-ray with pathology. The challenge is to find images that are semantically similar, and not merely similar in appearance. This gap, often referred to as the “semantic gap,” is a significant hindrance to the practical use of CBIR systems.

Since this problem remains unsolved, research in image retrieval focuses on new methods to characterize the image content with higher level semantics, closer to that familiar to the user and potentially more useful in retrieving similar-appearing images.²⁴ Although several efforts have been conducted to integrate semantics into image retrieval applications,^{25–27} most of the proposed approaches are dedicated to automatic annotation

of the images with semantic terms, and they do not focus on the potential relations among the semantic terms during the retrieval step. For example, in an effort to minimize the semantic gap of in CBIR, the bag-of-visual words-based feature representation scheme become popular.²⁸ In this approach, generally the low-level visual features of local regions of points, such as color, texture, and so on, are vector quantized to generate the visual words. The bag-of-words (BoW) models have been successfully used in natural language processing²⁹ to capture a summary of the semantics of text based on word content. Although it has proved to be more effective for image representation similar to document representation, the unsupervised clustering to generate the words or dictionary largely neglects the semantic contexts of the local features. As a result, commonly generated visual words are still not as expressive as keywords in text documents. Furthermore, since images are more complex than text documents for the task of representation, only considering the visual word frequency at both image and collection level (such as, “TF-IDF” weighting) might not be sufficient to determine its importance for image representation.

Authors of biomedical articles often use arrows, pointers, and other annotations such as text labels overlaid on figures and illustrations to highlight the ROIs as special cases. These annotations are then often referenced and correlated with concepts in the caption text or figure citations in the article text. For example, Fig. 1 (and caption) shows different arrows/arrow-heads pointing to honeycomb, cystic, and bronchial patterns in the article³⁰ “Lymphangioleiomyomatosis: Pulmonary and Abdominal Findings with Pathologic Correlation” in the journal *RadioGraphics*. These annotation markers are often a good source of information to locate specific visual patterns as an ROI in the image. Users of an image retrieval system of these biomedical articles are often more concerned with finding images that contain similar patterns of any highlighted ROIs instead of overall similarity of entire images. Hence, retrieval systems should be interactive where a user can select an ROI with a specific pattern to look for and should have the option for region-based search by returning images with a specific pattern as well as in a specific location.

To minimize the limitations of stand-alone (e.g., without any user interaction) CBIR systems based on low-level and BoW-based feature representation schemes that result in the semantic

Fig. 1 Example chest computed tomography (CT) image with associated caption and different region-of-interests (ROIs) pointed by arrows.

gap and motivated by the successful use of machine learning in IR, we present an interactive concept and ROI-based image retrieval approach. 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 gray-level radiological images, differential color, and texture structures in microscopic pathology and dermoscopic images.²⁵ The variation in these local patches can be effectively modeled as “concepts” by using supervised learning technique, such as support vector machine (SVM).³¹ Here, concepts are more expressive semantically compared to the visual words generated by employing unsupervised clustering techniques. The images are represented using concepts that comprise of color and texture patches from local image regions in a multidimensional feature space.

Our ultimate goal is to develop a retrieval system that finds images contain patterns (concepts) similar to an interactively marked ROI. In order to achieve this goal, we developed a local concept-based feature representation and retrieval methods. With an on-line concept classification scheme, our methods are capable of automatically mapping the appearance of visual entities within selected image ROI to a limited set of concepts (vocabulary), and then use these concepts to search images indexed with similar concepts and contains concepts in similar location to the query ROI. In addition, to overcome the limitation of the TF-IDF-based weighting in commonly used BoW-based image representation, the importance or visualness of concepts is measured as Shannon entropy³² based on pixel values of image patches and used to refine the feature vector.

Finally, to consider the localization information of concept distribution in images, we propose a semiglobal concept-based representation scheme. Each image is segmented into five overlapping regions based on the use of a grid of cells superimposed on the encoded images and local concept feature vectors are generated from those regions to finally form a combined multidimensional vector. The spatial information is introduced as a postprocessing step to rerank the retrieved images via a spatial verification based on similar overlapping region matching of query image ROI and database images. The relevance to a clinical query is aided by this addition of semantic information to extracted image features for improved retrieval. The conceptual difference between this work and majority of the aforementioned approaches is the use of domain knowledge as supervised learning for concept modeling and the interactive nature of the query formulation process for region-specific search with spatial verification in image space.

The rest of this article is organized as follows: Sec. 2 describes our concept-based image representation approach. Section 3 details our ROI-based image search method in the concept feature space. The experiments and analysis of the results are presented in Secs. 4 and 5 and finally, Sec. 6 provides the conclusions.

2 “Bag-of-Concepts”-Based Image Representation

A major component of the proposed retrieval approach is the image representation in a visual concept feature space. By the term concept, we refer to the perceptually distinguishable image patches that are identified locally in image regions. For example, if we consider a CT image of lung or chest, which appears in many radiographic or medical-related journals, we observe several image regions with texturally different visual

patterns that are semantically distinguishable from each other as shown in Fig. 4. For example, an area (ROI) of a CT scan having a slightly bright and hazy appearance can be mapped to the pattern “ground-glass opacity.”

When the domain knowledge is available for these kinds of images, it would be effective to exploit it by utilizing any supervised learning-based techniques. A supervised learner can create a model of concepts to capture the variabilities of the local patches with sufficient training samples. In this context, an instance (e.g., local patch) in the training set can be represented by a feature vector along with its local concept- or category-specific labels.

Therefore, in order to perform concept-based image representation, the first step is to generate a set of concepts from distinguished local image patches by employing a supervised learning technique, such as SVM.³¹ To perform learning, 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 visual concept. The training set comprised of the local patches is generated by a fixed-partition-based approach and represented by a combination of color and texture moment and edge frequency-related features.

A multiclass SVM³¹-based classification method is used by combining all pairwise comparisons of binary SVM classifiers, known as “one-against-one” or pairwise coupling (PWC).³³ PWC constructs binary SVMs between all possible pairs of classes. Hence, for L classes, this method uses $L * (L - 1)/2$ binary classifiers that individually compute a partial decision for classifying a data point (image). During the testing of a feature $\mathbf{x}$, each of the $L * (L - 1)/2$ classifier votes for one class. The winning class is the one with the largest number of accumulated votes.

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. 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 feature vector.

For each $\mathbf{x}_{kj}$, the local concept category probabilities are determined by the prediction of the multiclass SVMs as

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

Finally, the category label of $\mathbf{x}_{kj}$ is determined as c_m , which is the label of the category with the maximum probability score. Hence, the entire image is thus represented as a two-dimensional index linked to the concept labels. Based on this encoding scheme, an image I_j can be represented as a vector in a local semantic concept space as

$$\mathbf{f}_j^{\text{Concept}} = [f_{1j}, \dots, f_{ij}, \dots, f_{Lj}]^T, \quad (2)$$

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.

2.1 Entropy Weighted Concept Vector

Measuring only the concept frequency in individual images might not be sufficient to determine their importance for

image representation. Additional information is necessary. In general, the popular TF-IDF term-weighting scheme is used, where the vector elements (feature attributes) are expressed as the product of local and global weights. This weighting scheme amplifies the influence of terms, which occur often in a document (e.g., *tf* factor), but relatively rare in the whole collection of documents (e.g., *idf* factor).³⁴ However, images are more complex than text documents and image patches contain more information about visual patterns of particular concepts than single keywords in text documents. To know which concept has visually discriminative power is important for image representation as the “bag of concepts.”

The visualness of a concept is generally defined as to what extent a concept has visual characteristics. To know which concept has visually discriminative power is important for image representation, especially refining the bag-of-concepts-based representation, since not all concepts are related to visual contents. An entropy of an image patch (e.g., concept) can be valuable to determine its importance of visualness for image representation.³⁵ Entropy can be considered as the average number of bits one needs to represent a symbol in a stationary system, where the limited source symbols have fixed probabilities of occurrence.³² In other words, an entropy is a quantity that is used to describe the degree of randomness of an image. Low entropy image patches will have very little contrast and a relatively uniform color information, such as those that appear in the background area of medical images, as shown in Fig. 5 for the concepts black and gray background. On the other hand, high entropy image patches such as the foreground image region have a great deal of contrast from one pixel to the next and are more important in capturing image content.

We observe that in medical images, various regions in the foreground (such as the concepts honeycomb, bronchi, etc.) appear as rough or fractal. So the entropy of image patches is helpful for their separation from the uniform and smoothly varying background. The concepts in the part of background usually appear more frequently in images (similar to “stop-words” in documents) than those in the part of foreground; hence, they are less effective in representing images for classification or retrieval. Thus, we use the entropy as a measure of concept importance similar to the keyblock entropy in Ref. 36, which is the Shannon entropy of the patches based on pixel values. However, we only consider the intensity values of the gray-scale image patches instead of considering all three color channels in RGB space. The entropy of an image (patch) I is expressed as

$$E = - \sum_{I=0}^{255} P(I) \log_2 P(I), \quad (3)$$

where $P(I)$ is probability of occurrence of the intensity value (gray value) I appears in a patch.

Now, to measure the entropy of each concept categories $c_i \in C$, we select all the training patches in each category, sum up their entropy values based on Eq. (4), and obtain the average values based on the number of training sample in each category as

$$e_i = \frac{\sum_{j=0}^{N_{c_i}} E_j}{N_{c_i}}, \quad (4)$$

where E_j is the entropy of a patch $\mathbf{x}_j$ and N_{c_i} is the number of training samples in concept category c_i .

Finally, by using the entropy values of each concept categories, an image I_j is represented as a weighted-concept vector

$$\mathbf{f}_j^{\text{wConcept}} = [\hat{f}_{1_j} \cdots \hat{f}_{i_j} \cdots \hat{f}_{L_j}]^T, \quad (5)$$

where each $\hat{f}_{i_j}$ is weighted as $f_{i_j} * e_i$ to take into effect of entropy-based weight of each concept c_i .

3 Region-of-Interest-Based Image Search in Concept Space

The representation scheme described in the previous section captures only a coarse distribution of the concepts similar to a global color histogram. To overcome this drawback, each image is segmented into five overlapping regions based on the use of a grid of cells superimposed on the encoded images. These cells are obtained by first dividing the entire image space into 16 nonoverlapping subimages and four connected subimages are grouped to generate five different larger overlapping subimages (1 to 5) as shown in the left side of Fig. 2. Each region is now considered as an independent image where a weighted-concept vector $\mathbf{f}_r^{\text{wConcept}}$ is computed, and finally, feature vectors of the five regions are combined to generate a multi-dimensional semiglobal vector $\mathbf{f}^{\text{SG-Concept}}$.

In CBIR, the term “global” feature vector is generally used when we compute feature from entire image space without considering localization information. In this work, we divided each image into five overlapping subimages based on their five different locations in image space and extracted concept features separately from each subimage for future comparison with query image ROI or comparing with the entire image concept vector by concatenating feature vector of subimages to a combined vector of dimension five times larger than subimage feature vector. Hence, the term “semiglobal” is used here to distinguish it from global feature vector.

Based on this representation scheme, an interactive search mechanism is also implemented where a user can select an ROI as a bounding rectangle and have the option for region-based search by returning images with similar pattern in any one of the five clusters. As shown in Fig. 2, each overlapping region (1 to 5) is encoded with 16 (4×4) concept labels. So, when an ROI (red rectangle) is selected by user, the corresponding concept labels beneath it (e.g., c_5 and c_8) and location of the ROI in overlapping regions (e.g., subimages 4 and 5 in this case) are identified first. Next, based on the concept label statistics of the ROI, an entropy-weighted feature vector is generated.

The images are indexed in an inverted index with the concept labels, where each concept maintains a list of images where it is occurred at least once. An inverted index is a very common indexing structure for BoW-based document representation scheme, which is actually a hashmap-like data structure that directs from a word to a document.³⁴ So, for the ROI-based search process, we only consider those images in list of corresponding ROI concept labels. The concept vector of the query image ROI is now matched with the five overlapping region (subimage) vectors of selected images to perform an approximate region matching. The images are finally ranked in descending order according to the similarity matching scores of query ROI and any image regions under that image. In addition, we can filter the retrieved images further through a spatial verification based on the location of query ROI and similar overlapping

Fig. 2 An example image with five overlapping subimages (1 to 5) and the schematic representation of inverted indexing.

region in the same subimage(s) (e.g., subimages 4 and 5). The step-by-step approach of an ROI-based spatial (similar location) image search process is presented in Algorithm 1. The relevance to a clinical query is aided by this addition of location information to extracted image features for improved retrieval.

4 Experiments

To evaluate the effectiveness of the proposed image representation approach, exhaustive experiments were performed in two different image collections. The first image collection consists of 5000 multimodal 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'2007.³⁷ In this collection, the images are classified into three levels as shown in Fig. 3. At the first level, images are categorized according to modality (e.g., x-ray, CT, MRI, etc.). At the next level, each of the modalities is further classified according to the examined body parts (e.g., head, chest, etc.) and finally, 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 generally used in ImageCLEFmed²³ 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. Around 80% of the images are gray-level (e.g., x-ray, CT, and MRI) and 20% are color images (e.g., microscopic pathology, histology, and dermatology) with varying resolutions.

For concept model generation based on the SVM learning, 60 local concept categories are manually defined from a training set of more than 19,000 image patches.³⁸ The local concepts are selected as the ones that exhibit some meaning to the physicians with distinct visual appearances, such as different lung tissue patterns of x-ray and CT images, microscopic images of different color and texture patterns, and so on. To generate the local patches, each image in the training set (we used only 2% images of entire data set) is resized to 256×256 pixels and partitioned into an 8×8 grid generating 64 nonoverlapping regions of size 32×32 pixels. Only the regions that conform to at least 80% of a particular concept category are selected and labeled with the corresponding category label.

We chose to limit our preliminary retrieval experiment to thoracic CT scans only for ROI-based retrieval evaluation. Such images exhibit high regularity and account for a large portion of the images publicly available in ARRS GoldMiner,³⁹ OpenI,⁴⁰ and YIF.⁴¹ We generated a collection of 450 CT scans of six different categories as shown in Fig. 4. For concept model generation based on the SVM learning, 13 local concept categories are manually defined from 3200 image patches of

Algorithm 1 ROI-based spatial image search

```

1: Initially, for a query image  $I_q$ , let the user select a visual ROI  $I_{q_r}$  as
   a bounding rectangle.
2: Find the concept labels  $c_i$ ,  $1 \leq i \leq L$  beneath  $I_{q_r}$  from the encoded
   query image  $I_q$  in quantized concept space.
3: Represent  $I_{q_r}$  as an entropy-weighted vector  $\mathbf{f}_{q_r}^{w\text{Concept}}$  based on
   Eq. (5).
4: for each concept  $c_i$  in the query ROI do
5:   Retrieve the list of  $n$  images from inverted index that contain
   the concept, where  $(n \leq N)$  and  $N$  is the number of images in
   the database.
6: end for
7: for  $j = 1$  to  $n$  do
8:   for subimage  $reg = 1$  to  $5$  do
9:     if subimage ( $I_{q_r}$ ) == subimage ( $I_{reg}$ ) then
10:      Perform similarity matching  $S(I_{q_r}, I_{reg})$  as cosine similarity
      measure between feature vectors  $\mathbf{f}_{q_r}^{w\text{Concept}}$  and  $\mathbf{f}_{reg}^{w\text{Concept}}$ .
11:    end if
12:  end for
13: end for
14: Return the image of associated subimage with
    similarity scores in descending order for a rank-based retrieval.

```

size 64×64 pixels as shown in Fig. 5. Color moment, Tamura, autocorrelation, and edge frequency-based features are extracted and combined to form a 77-dimensional feature vector for all training patches in both collections.

For the SVM training, the best accuracy (77.53%) is achieved by using the RBF kernel in a 10-fold cross validation. Hence, after finding the best values of parameters $C = 100$ and $\gamma = 0.30$ of the RBF kernel, they are utilized for the final training to generate the model file for the concept learning. The LIBSVM software package⁴² is used for the implementation of multi-class SVM.

5 Results

For evaluation of the effectiveness of the proposed approach based on retrieval performance, the concepts of *Precision* and *Recall* from IR³⁴ were used. *Recall* is the ratio of the number

Fig. 4 Example of thoracic CT scans of six different categories (visual patterns).

Patch category	Visual example	# patches	Avg. Entropy
Background_black		169	3.554
Groundglass		257	6.401
Honeycomb		174	6.777
Cyst		270	6.500
Bronchi		263	6.750
Background_white		272	0.771
Nodule_micro		230	6.258
Nodule_macro		259	6.325
Corner_black_grey		287	5.158
Fat_lung_tissue		279	6.033
Fat_grey		283	3.500
Tissue_normal		275	4.989
Tissue_heart		182	4.671

Fig. 5 Example average entropies of different concept categories.

of relevant images returned to the total number of relevant images in the database. *Precision* is the ratio of the number of relevant images returned to the total number of images returned. So, when the top N (scope) images are considered and there are Q relevant images, the precision within top N

Fig. 3 Classification structure of the multimodal medical data set.

images is defined to be $Precision(N) = Q/N$, whereas recall will be $Recall(N) = Q/R$. Here, R is the number of all images that are relevant to the query image. A high *Precision* value means that there are few false alarms (i.e., the % of irrelevant images in the retrieval), while a high *Recall* value means that there are few false dismissals (the % of relevant images which failed to be retrieved). The average precision and recall are calculated over all the queries to generate the precision–recall (PR) curves in different settings. 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 a correct match if and only if it is the same semantic category as the query image.

Figures 6 and 7 show the PR curves of the concept-based search approaches in both global (entire image) and semiglobal (five overlapping regions) feature spaces with and without using entropy-based weighting as discussed in Sec. 2. The performances were also compared to the well-known color edge directional descriptor (CEDD) and fuzzy color texture histogram (FCTH) from the Lucene image retrieval (LIRE) library.⁴³ Due to the effectiveness of CEDD and FCTH features in ImageCLEF benchmark evaluation by several groups,^{23,44} we selected these for comparative evaluation with the proposed concept-based image representation approaches. For the low-level CEDD and FCTH feature vectors, we used the Euclidean similarity measure and for the concept feature vectors, the cosin” similarity measure is used as it is commonly used to calculate the similarity between different documents in BoW-based feature representation.

By analyzing Figs. 6 and 7, we can observe that the precision of concept-based features is much better at each recall level compared to the low-level CEDD and FCTH features for both collections. In addition, it improves the precision level further when the concept feature is entropy weighted. For example, we achieved further improvement in precision at lower recall levels (e.g., 0.1 to 0.5) when the search is preformed in the semiglobal-concept feature space in the lung CT collection as shown in Fig. 7. However, the overall search performance is comparatively much better for the first collection due to the lesser

Fig. 6 Precision-recall (PR) graph for search results with different image features in multimodal medical data set.

Fig. 7 PR graph for search results with different image features in lung CT database.

complexity of the images as organized by modality. It is always easier to find images of a specific modality compared to images with specific disease-related patterns.

It is relatively easy to perform quantitative evaluation of retrieval in entire image space as a batch process. However, it is really difficult to perform such a batch processing for all images in the database for an ROI-based search as described in Sec. 3. Instead, to perform a comparative evaluation, we manually selected 12 ROIs from different query images (e.g., a query ROI as a red rectangle is shown in a retrieval interface of Fig. 15) from the collection of 450 CT scans of six different categories. For each image category (out of the six categories), we selected 10 different query ROIs to search for. A retrieved image (e.g., associated with a retrieved ROI) was considered a match if it belonged to the same category as the query image (e.g., associated with the visual ROI of the query image) out of the six concept categories. Figure 8 shows the bar graphs of P5, P10, and P20 scores for searches by considering entire query images and query ROIs (manually selected) for 25 queries in each category. We counted the number of similar images retrieved for each categories by considering top 5 (P5), top 10 (P10), and top 20 (P20) retrieved images and measured the average precision for all queries. From Fig. 8, it is clear that precision is always improved when an ROI-based search is performed and overlapping regions are compared with query ROI (e.g., ROI-ROI) instead of comparing with entire images (e.g., ROI-entire).

Fig. 8 Precision over the top K (5, 10, 20) images for all six categories.

In addition, Figs. 9–14 show the bar graphs in a similar manner for individual categories. It is clear from the graphs that for some categories, such as “micronodules” and “fibrosis,” the performances are comparatively lower as the patterns (e.g., white speckles or dots) of these ROIs are often confused with the pattern in “bronchi” category and those categories have comparatively lower in numbers in the training sample for the model generation process.

For a qualitative evaluation of retrieval results, Fig. 15 shows a snapshot of the retrieval interface, where a user manually selected an ROI (red rectangle) from a query image in the left panel and the system returned top 10 most similar images based on performing the “cosine” similarity measure between query ROI and entire targeted images concept vectors. The concepts labels from the quantized concept image space beneath the query ROI are also shown below the right panel of the interface. The results show here that all the top five images are belonging to the same category as the query ROI image category (e.g., “bronchi”), which the user is intended to search for in this case. However, with this approach, the search is still performed in the entire image space at a global level similar to a global color histogram-based search. As a result, we can observe few misses in the bottom row of the search result.

Figure 16 shows a snapshot of the retrieval interface, where a user manually selected an ROI (yellow rectangle) from a query image in the left panel and the system returned images (shows only three images in the right panel for better viewing) with similar overlapping regions (red rectangles) to the query ROI. The concept labels from the quantized concept image space beneath the query ROI are also shown below the right panel of the

Fig. 9 Precision over the top K (5, 10, 20) images for the “bronchi” category.

Fig. 10 Precision over the top K (5, 10, 20) images for the “ground glass” category.

Fig. 11 Precision over the top K (5, 10, 20) images for the “honeycombing” category.

Fig. 12 Precision over the top K (5, 10, 20) images for the “micronodules” category.

Fig. 13 Precision over the top K (5, 10, 20) images for the “cyst” category.

Fig. 14 Precision over the top K (5, 10, 20) images for the “fibrosis” category.

Fig. 15 A snapshot of the retrieval interface where a query ROI is selected by user in the left panel and top matching images are retrieved based on the cosine similarity matching of concept vectors of query ROI and entire image space of targeted images.

Fig. 16 A snapshot of the retrieval interface where a query ROI is selected by user in the left panel and top matching images are retrieved based on the cosine similarity matching of concept vectors of query ROI and associated subimages (1 to 5) of targeted images.

interface. The search result correctly shows here that all three returned images contain associated similar subimages compared to the query image ROI in different locations and the search outcome better satisfies what a typical user is looking for based on his/her selected ROI in query images.

The ultimate goal of this work is to improve image retrieval capability with ROI-based search within the Open-I⁴⁰ multimodal retrieval system currently being developed by the U.S. NLM at National Institutes of Health (NIH). It enable users to search for retrieve citations that are enriched with relevant images and “bottom-line statements” extracted from a collection of 400,000 open access articles and nearly 1.3 million illustrations from the biomedical literature hosted at the NLM PMC repository. Currently, it supports text-based retrieval using UMLS concepts and allows visual retrieval using a unified multimodal feature vector.⁴⁵

6 Conclusion

This article introduces an approach for improving the accuracy of medical image retrieval by representing images in a weighted entropy-based “bag-of-concepts” feature space. An interactive ROI-based search is implemented based on the presented image representation scheme. Concept location information is also taken into account by dividing each image into five overlapping regions. A spatial verification step can further refine the retrieval results based on similar overlapping region matching of query image ROI and the overlapping regions of database images. The search approach realizes semantic abstraction via prior learning and user interaction and improves retrieval effectiveness as validated by experimental results in a collection of lung CT images. With these encouraging results, we hope to extend this approach to other modalities and eventually will be able to perform searches in larger collections.

Acknowledgments

This research was supported by the Intramural Research Program of the NIH, NLM, and Lister Hill National Center for Biomedical Communications. We thank the Image-CLEFmed²³ organizers for making the data set available for the experiments.

References

- MEDLINE®, <http://www.nlm.nih.gov/pubs/factsheets/medline.html> (10 January 2015).
- PubMed, <http://www.ncbi.nlm.nih.gov/pubmed> (10 January 2015).
- Z. Lu, “PubMed and beyond: a survey of web tools for searching biomedical literature,” *Database (Oxford)* **2011**, baq036 (2011).
- M. S. Simpson et al., “Literature-based biomedical image classification and retrieval,” *Comput. Med. Imaging Graphics* **39**, 3–13 (2015).
- J. Kalpathy-Cramer et al., “Evaluating performance of biomedical image retrieval systems—an overview of the medical image retrieval task at ImageCLEF 2004–2013,” *Comput. Med. Imaging Graphics* **39**, 55–61 (2015).
- M. S. Simpson, D. Demner-Fushman, and G. R. Thoma, “Evaluating the importance of image-related text for ad-hoc and case-based biomedical article retrieval,” *AMIA Annu. Symp. Proc.* **13**, 752–756 (2010).
- D. Demner-Fushman et al., “Combining text and visual features for biomedical information retrieval and ontologies,” Technical Report to the LHCBC Board of Scientific Counselors (2010).
- D. Demner-Fushman et al., “Annotation and retrieval of clinically relevant images,” *Int. J. Med. Informatics* **78**(12), e59–e67 (2009).
- M. S. Simpson et al., “Towards the creation of a visual ontology of biomedical imaging entities,” in *AMIA Symp. Proc.*, pp. 866–875 (2012).
- R. J. Sandusky and C. Tenopir, “Finding and using journal-article components: impacts of disaggregation on teaching and research practice,” *J. Am. Soc. Inf. Sci.* **59**(6), 970–982 (2008).
- M. A. Hearst et al., “Biotext search engine: beyond abstract search,” *Bioinformatics* **23**(16), 2196–2197 (2007).
- Apache lucene™ text search engine, <http://lucene.apache.org/core/> (15 February 2015).
- Yottalook™ radiology search, http://yottalook.com/index_web.php (15 February 2015).
- Unified medical language system® (UMLS®) <http://www.nlm.nih.gov/research/umls/> (15 February 2015).
- S. Xu, J. McCusker, and M. Krauthammer, “Yale image finder (YIF): a new search engine for retrieving biomedical images,” *Bioinformatics* **24**(17), 1968–1970 (2008).
- R. Datta et al., “Image retrieval: ideas, influences, and trends of the new age,” *ACM Comput. Surv.* **40**(2), 1–60 (2008).
- H. Müller et al., “A review of content-based image retrieval systems in medical applications clinical benefits and future directions,” *Int. J. Med. Informatics* **73**(1), 1–23 (2004).
- C. R. Shyu et al., “ASSERT: a physician-in-the-loop content-based image retrieval system for HRCT image databases,” *Comput. Vision Image Understanding* **75**, 111–132 (1999).
- W. Hsu et al., SPIRS: a web-based image retrieval system for large biomedical databases,” *Int. J. Med. Informatics* **78**, S13–S24 (2009).
- M. M. Rahman, P. Bhattacharya, and B. C. Desai, “A framework for medical image retrieval using machine learning & statistical similarity matching techniques with relevance feedback,” *IEEE Trans. Inf. Technol. Biomed.* **11**(1), 58–69 (2007).
- T. M. Lehmann et al., “Content-based image retrieval in medical applications—a novel multi-step approach,” *Proc. SPIE* **3972**, 312–320 (2000).
- H. Müller et al., “Integrating content-based visual access methods into a medical case database,” in *Proc. Med Inform Europe (MIE 2003)*, St Malo, France, pp. 480–485 (2003).
- H. Müller et al., Eds., “ImageCLEF: experimental evaluation in visual information retrieval,” in *Information Retrieval Series*, Vol. **32**, Springer, Heidelberg (2010).
- J. Deng et al., “ImageNet: a large-scale hierarchical image database,” in *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*, pp. 143–156 (2009).
- M. M. Rahman, S. K. Antani, and G. R. Thoma, “A medical image retrieval framework in correlation enhanced visual concept feature space,” in *22nd IEEE Int. Symp. on Computer-Based Medical Systems (CBMS)*, Albuquerque, New Mexico (2009).
- D. Zhang, M. M. Islam, and G. Lu, “A review on automatic image annotation techniques,” *Pattern Recognit.* **45**(1), 346–362 (2012).
- B. Andre et al., “Learning semantic and visual similarity for endomicroscopy video retrieval,” *IEEE Trans. Med. Imaging* **31**(6), 1276–1288 (2012).
- C. F. Tsai, “Bag-of-words representation in image annotation: a review,” *ISRN Artif. Intell.* **2012**, 376804 (2012).
- E. M. Voorhees, “Natural language processing and information retrieval,” *Lect. Notes Comput. Sci.* **1714**, 32–48 (1999).
- E. Pallisa et al., “Lymphangioleiomyomatosis: pulmonary and abdominal findings with pathologic correlation,” *RadioGraphics* **22**, S185–S198 (2002).
- V. Vapnik, *Statistical Learning Theory*, Wiley, New York (1998).
- C. E. Shannon and W. Weaver, *The Mathematical Theory of Communication*, University of Illinois Press, Urbana, Illinois (1949).
- T. F. Wu, C. J. Lin, and R. C. Weng, “Probability estimates for multi-class classification by pairwise coupling,” *J. Mach. Learn. Res.* **5**, 975–1005 (2004).
- R. B. Yates and B. R. Neto, *Modern Information Retrieval*, Addison Wesley, Boston (1999).
- J. W. Jeong, X. J. Wang, and D. H. Lee, “Towards measuring the visualness of a concept,” in *Proc. of the 21st ACM Int. Conf. on Information and Knowledge Management*, pp. 2415–2418 (2012).
- L. Zhu, C. Tang, and A. Zhang, “Using keyblock statistics to model image retrieval,” *Proc. LNCS* **2195**, 522–529 (2001).

37. H. Müller et al., "Overview of the ImageCLEFmed 2007 medical retrieval and annotation tasks," *Proc. LNCS* **5152**, 472–491 (2008).
38. M. M. Rahman et al., "Biomedical image representation and classification using an entropy weighted probabilistic concept feature space," *Proc. SPIE* **9039**, 903908 (2014).
39. ARRS goldminer[®], <http://goldminer.rrs.org> (15 February 2015).
40. Open-i biomedical image search engine, <http://openi.nlm.nih.gov/> (15 February 2015).
41. Yale image finder, <http://krauthammerlab.med.yale.edu/imagefinder/> (15 February 2015).
42. D. M. Hansell et al., "Fleischner society: glossary of terms for thoracic imaging," *Radiology* **246**(3), 697–722 (2008).
43. M. Grubinger et al., "Lire: lucene image retrieval: an extensible Java CBIR library," *Proc. of the 16th ACM Int. Conf. on Multimedia*, Vancouver, British Columbia, Canada, pp. 1085–1088 (2008).
44. H. Müller et al., "Overview of the CLEF 2010 medical image retrieval track," in *CLEF 2012 Evaluation Labs and Workshop, Online Working Notes*, Padua, Italy (2010).
45. D. Demner-Fushman et al., "Design and development of a multimodal biomedical information retrieval system," *J. Comput. Sci. Eng.* **6**(2), 168–177 (2012).

Md. Mahmudur Rahman received his PhD in computer science from Concordia University, Montreal, Canada (2008). He worked almost six years as a research scientist in the U.S. National Library of Medicine at NIH before joining (2014) as an assistant professor in the Computer Science Department at Morgan State University. He has good

expertise in the fields of computer vision, information retrieval, and machine learning, and their application to retrieval of biomedical images from large collections.

Sameer K. Antani is a staff scientist in the Lister Hill National Center for Biomedical Communications. He is a versatile researcher leading several scientific research projects. He applies his expertise in biomedical image processing, medical image informatics, visual information retrieval, and related topics in computer science and engineering technology, toward advancing the biomedical sciences, in global health applications, and the advanced methods for dissemination of scientific information.

Dina Demner-Fushman is a staff scientist in the Lister Hill National Center for Biomedical Communications of the U.S. National Library of Medicine at NIH. With her expertise in information retrieval and natural language processing, she leads several NLM projects in the areas of information extraction for clinical decision support, EMR database research and development, and image and text indexing for clinical decision support and education.

George R. Thoma is the chief of the Communications Engineering Branch of the Lister Hill National Center for Biomedical Communications of the U.S. National Library of Medicine at NIH. He directs R&D programs in image processing, document image storage on digital optical disks, automated document image delivery, digital x-ray archiving, and high speed image transmission. He is a fellow of the SPIE, the International Society for Optical Engineering.