IJESRT

INTERNATIONAL JOURNAL OF ENGINEERING SCIENCES & RESEARCH TECHNOLOGY

A Two Tier Approach For Content Based Microscopic Image Retrieval System

Packia lakshmi.T[*1], R.Karthikeyan[2], M.Rameshkumar[3]

[*1]M.E- Student, Department of CSE, Vel Tech Multitech Dr.Rangarajan Dr.Sakunthala Engineering College, India

[2, 3] Asst. Professor, Department of CSE, Vel Tech Multitech Dr.Rangarajan Dr.Sakunthala Engineering College, India

abishenba@gmail.com

Abstract

Image retrieval is an active research area in image processing, pattern recognition, and computer vision. In this project, a multi-tiered CBIR system for microscopic images utilizing a reference database that contains images of more than one disease is described. This system enables both multi image query and slide level image retrieval in order to protect the semantic consistency among the retrieved images. To improve the performance, the proposed system uses z-score normalization and non- negative matrix factorization (Sub-space projection) methods. It achieves high accuracy and reduces the processing time greatly. It provides better performance for all type of images that have inter- and intra- reading semantic variations, intra-slide semantic variations, and inter-subtype visual similarities.

Keywords: Content-based image retrieval (CBIR), Support Vector Machine (SVM), Image Retrieval (IR), Microscopic images, weighting scores.

Introduction

MULTIMEDIA contents are growing explosively and the need for multimedia retrieval is occurring more and more frequently in our daily life. Due to the complexity of multimedia contents, image understanding is a difficult but interesting issue in this field. Extracting valuable knowledge from a large-scale multimedia repository, so-called multimedia mining, has been recently studied by some researchers. Typically, in the development of an image requisition system, semantic image retrieval relies heavily on the related captions, e.g., file-names, categories, annotated keywords, and other manual descriptions. Unfortunately, this kind of textual-based image retrieval always suffers from two problems: high-priced manual annotation and inappropriate automated annotation. On one hand, high-priced manual annotation cost is prohibitive in coping with a large-scale data set.

As a result, a number of powerful image retrieval algorithms have been proposed to deal with such problems over the past few years. Content-Based Image Retrieval (CBIR) is the mainstay of current image retrieval systems. In general, the purpose of CBIR is to present an image conceptually, with a set of low-level visual features such as color, texture, and shape. These conventional approaches for image retrieval are based on the computation of the similarity between the user's query and images via a query by example (QBE) system [21]. Content-based image retrieval (CBIR) systems [2]-[7] for medical images are important to deliver a stable platform to catalog, search, and retrieve images based on their content.

Although several CBIR projects exist for radiology [8]–[10] and several other projects are underway, there is an acute need for a comprehensive and flexible CBIR system for microscopic images with direct implications for the field of pathology and cancer research. Microscopic images present novel challenges because they 1) are large in size 2) demonstrate high degree of visual variation due to large variation in preparation (e.g., staining, thickness), and 3) show huge biological variation. Therefore, a well-designed CBIR system for microscopic images can be extremely useful resource for cancer research, diagnosis, prognosis, treatment, and teaching. In other words, such a system can 1) assist pathologists in their diagnosis and prognosis, 2) potentially help to reduce inter- and intra-reader variability in clinical practice for the diseases, especially those with complicated classification, 3) help cancer researchers in better understanding of cancer development, treatment monitoring, and clinical trials, and 4) train future generation of

researchers by providing consistent, relevant and always available support and assistance. In this paper, we describe the design and the development of a multi-tiered CBIR system for microscopic images from a reference database that contains more than one disease.

To provide a motivating example and to test the ideas developed in this paper, images in our reference database include sample regions cropped from digitized hematoxylin and eosin (H&E) stained whole slides. Neuroblastoma (NB) and follicular lymphoma (FL) tissue images have been collected as part of our ongoing projects for both diseases. FL cases are stratified to three histological grades from low- to high-risk category as follows: Grade I, Grade II, and Grade III. According to the International Neuroblastoma Classification System, NB tissues are mainly divided into two subtypes such as stroma rich (SR) or stroma poor (SP) based on the degree of Schwannian stroma development [13]. Additionally, SP tissue has three subtypes such as undifferentiated (UD), poorly differentiated (PD), and differentiating (D).

Annotation of microscopic images, e.g., H&E-stained pathology slides, with subtypes of the main disease needs an expert pathologist to select pathology-bearing regions or regions of interests from the whole slide. Then each selected region is annotated semantically by giving a score according to its visual qualitative characteristics. The final decision on the grade or subtype of the disease for the whole slide is given after considering the annotations of all sample regions, i.e., the average subtype-related score over all sample regions is assigned as the final score of that whole slide. Considering the extremely large sizes of microscopic images, it is obvious that manual annotation of these images is a time-consuming process and those annotated images may not be easily available for clinical use. Therefore, one of the aims of this study is to organize the annotated microscopic images in a database and utilize these images for the training of a CBIR system for microscopic images with different disease types and with their subtypes.

The novel aspects of our multi-tiered approach are: 1) it retrieves the most similar disease types in the slide level rather than in the image level by enabling multi-image queries in order to ensure the consistency among the retrieved images, and 2) slide-level scores are weighted in a sophisticated way by modifying the term *frequency* (*tf*) – inverse document *frequency* (*idf*) weighting concepts of information *retrieval* (*IR*) theory [14] to decrease the sensitivity of the proposed CBIR system to erroneously annotated sample images in the database. Since in real medical applications, especially for microscopic images at high magnifications, the query object is more likely to be a set of sample images extracted from a whole-slide image rather than being a single image, the multi-image query model suits perfectly for our case. It has been also proved that query by multi-images leads to more scalable and satisfactory query performances by overcoming the limitation on the specification of image content of single-image queries [15], [16].

In CBIR systems, images are typically represented with feature vectors extracted using low-level image processing techniques [8], [9], [17]. However, similarities in feature vector level do not always guarantee the semantic similarity (i.e., interpretations of images according to their predefined categories) between query image and retrieved images. This is known as the semantic gap problem [18], [19]. In this paper, we will explore the effect of slide-level retrieval system with multiple query images in order to increase the semantic relevance of query image set and retrieved images.

A general flowchart of the proposed CBIR system is illustrated in Fig. 1. It shows the main steps of the CBIR algorithm, e.g., feature extraction, major disease-type classification (first tier), image retrieval according to the subtypes of the diseases (second tier).

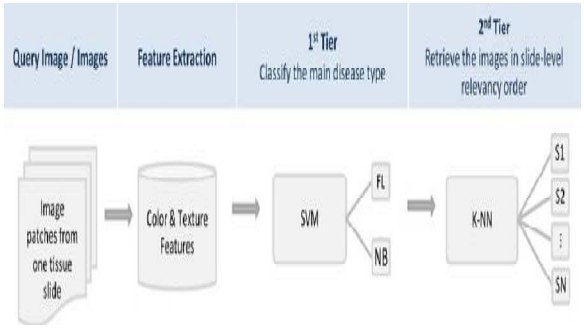

Figure 1: General flowchart for the CBIR system for a given query image or images.

Related Work

Most of the commercial search engines (e.g., Google, Yahoo!, Bing Image Search) are built around a semantic search, i.e., the user needs to type in a series of keywords and the images in those databases are also annotated using keywords; the match is accomplished primarily through these keywords. CBIR systems have been developed in the recent years to organize and utilize the valuable image sources effectively and efficiently for diverse

collections of images. Most of the recent CBIR systems in biomedicine [5], [8], [9], are designed to classify and retrieve images according to the anatomical categories of their content, i.e., head or chest X-ray images or abdominal CT images. For example, the Automatic Search and Selection Engine with Retrieval Tools (ASSERT) system [5] was designed for high-resolution CT images of the lung, where each set of feature was extracted from the pathology-bearing regions. Previously, a prefiltering approach [9] was proposed to reduce the search space of query images by categorizing the images using multiclass support vector machines (SVMs) and fuzzy c-mean clustering. The retrieval after prefiltering was done according to main disease categories only, which is similar to the first tier of our two-tiered approach.

In another study [20], expectation-maximization algorithm was used to generate clusters of block-based low-level features extracted from radiographic images. Then, the similarity between two clusters was estimated as a function of the similarity of both their structures and the measure components. Pourghassem and Ghassemian [21] proposed a two-level hierarchical medical image classification method. The first level was used to classify the images into the merged and nonmerged classes. They tested their algorithm on medical X-ray images of 40 classes. Although this is a two-level hierarchical classification, it is different from our approach because only the merged classes were evaluated in the second level to be classified with multilayer perceptron (MLP) classifiers into 1 of 40 classes.

Traditional indexing and search strategies used in radiological systems are not directly applicable in the context of digital microscopy since it is not obvious how to define a primary key or major anatomical structure for such images. To complicate things further, most known structures (e.g., cells, its components, tissue, etc.) are much more complex and require more detailed analysis than that would be needed at the higher resolutions and scale of radiological images. The feature extraction from microscopic images is also challenging because these images are composed of varying textures, overlapping structures, and different cell constituents even for the same disease types.
In the last decade, a few CBIR systems for the microscopic images have been developed for clinical use [6], [7], [17], [22], [23]. Mehta *et al.* designed a region-specific retrieval system based on sub-image query search on whole-slide images by extracting scale invariant features on the detected points of interests and 80% of match was achieved with the manual search for prostate H&E images [23] in the top five searches. Zheng *et al.* [6] proposed a CBIR system based on the weighted similarities of four feature types such as color histogram, image texture, Fourier coefficients, and wavelet coefficients. The retrieval performance of their system was tested using agglomerative cluster analysis for different pathology image categories and the best retrieval performance was observed for prostate query images.

Recently, Yang *et al.* [7] developed a Web-based system called *PathMiner*, which includes automatic segmentation, CBIR, and classification modules to assist diagnostics in pathology. They evaluated the classification performance of their system on five different blood cells such as chronic lymphocytic leukemia, mantle cell lymphoma, follicular center cell lymphoma, and acute lymphocytic leukemia and acute myelogenous leukemia by using SVM classifiers with texton histogram features and 87.27% of classification accuracy was achieved on an open set with large variations in staining characters.

Most of the CBIR approaches designed for microscopic images have their own specific application area, specific feature extraction technique, or a specific similarity measure for the evaluation. For example, disease-specific CBIR systems [17], [22], [23] have been developed for clinical decision support of specific diseases, while some of the CBIR systems were designed for the classification of different types of pathology images, i.e., liver tissue, prostate tissue, breast tissue, lymph node, and so on [6].

Although many promising CBIR approaches were developed for medical applications, there are still gaps in terms of image content, retrieval methodology, performance evaluations, and their application areas [18], [19], which make this research area an open problem for further studies. Particularly, the majority of the retrieval methodology of the published CBIR techniques focused on image-level retrieval either by choosing or defining an appropriate distance metric to compare the feature vectors from the query and database images [8], [17], [24]. However, multi-image query based retrieval is more suitable for challenging medical CBIR applications. Especially, microscopic images at high magnifications require multi-image queries in order to specify the query images more efficiently. Therefore, our CBIR method will focus on defining a retrieval methodology for multi-image queries, which can be also applicable for any type of multi-image query and retrieval application.

In summary, our approach focuses on one modality, which is the digital bright field microscopic image of tissue slides. It does not aim to provide a way to search and index generic medical image collections. It differs from the existing microscopic CBIR methods mainly in two aspects. First, two different diseases (FL and NB) are processed within a CBIR system with their highlevel semantic annotations. The framework can also be extended to several other diseases. Second, our approach enables multi-image queries instead of one image query and provides a slidelevel retrieval by keeping the slide-level consistency among retrieved images by using weighting scores depending on the image-level rank order and distributions of the subtypes over the reference dataset.

Proposed Approach

Feature Extraction

In this section, we will explain the feature extraction techniques that we employed to the images in our database.

- **Low-Level Feature Extraction**

There are many factors affecting the performance and accuracy of CBIR systems, such as choosing more discriminative features, similarity measurement criteria, query formulation, and so on. In order to design an effective CBIR system, the initial step in our study is to extract discriminative features from the images in the reference database. These features will also be calculated for query images. One of the most discriminating characteristics of microscopic images is color, especially when compared to most common radiological images, which are mostly gray level. Due to the high resolution of microscopic images, subtle changes in characteristics of cells, combinations of cells, structures, and tissues can also be differentiated from each other by texture characteristics. Therefore, for our CBIR design, we heavily make use of color and texture characteristics and extract these features using low-level image feature extraction techniques.

1) Color features*:* H&E images have considerably limited color spectrum, i.e., there are few dominant colors (hues of blue and pink), as shown on the sample images. Therefore, in order to better represent the limited color information in more detail, we used two more color spaces in addition to red–green–blue (*RGB*) color space. These additional color spaces are CIE*Lab* (*Lab*) and Modifed hue–saturation–value (MHSV) color spaces. In the *Lab* color space, *L* corresponds to illumination, and *a* and *b* channels correspond to color opponents. Thus, features extracted from the *Lab* space characterize the intensity and color information of images separately [25]. On the other hand, the *HSV* color space is known with similarity to the human conceptual understanding of colors. Besides this, *HSV* space can separate the chromatic and achromatic components, i.e., hue (*H*) channel distinguishes colors, saturation channel (*S*) represents the percentage of white light added to a pure color space, and value (*V*) refers to intensity of perceived light [25]. For each channel of a given color space, mean value and standard deviation are computed as first- and second-order statistics features. In total, 18 (2 features × 3 channels × 3 color spaces) color features are extracted from each image. Additionally, mean value, standard deviation, skewness, kurtosis, maximum and minimum values, energy, and entropy values are computed for gray-level intensity image. In summary, 26 color and gray-scale features are extracted using three different color spaces for a given image.

2) Texture features: Microscopic images with different disease types and subtypes can be distinguished via their homogeneity or texture characteristics. To capture the discriminative texture information, we investigated several texture feature extraction methods in the literature

[26]. Co-occurrence histograms are the most frequently used method for texture feature extraction [4]. They can be defined as a sample of a joint probability density of intensity levels of two pixels separated by a given displacement. The distribution in the histograms depends on the rotation angle and distance relationship between pixels. Once the co-occurrence histogram is computed, various features can be extracted related to texture characteristics, lower and higher order statistics, information-theory-related features, and correlation measure. As a consequence, we extracted the following features: mean, standard deviation, contrast, correlation, energy, entropy, and homogeneity from the normalized co-occurrence histograms for each *RGB* and *Lab* color channels and gray-level images. In addition, mean value, homogeneity, and entropy values are extracted from the difference histograms of the normalized co-occurrence matrix. For a given image, a total of 80 texture-based features are extracted using RGB, MHSV color spaces, and gray-level intensities. It should be noted that average of the co-occurrence histograms for eight different directions, i.e., $0^\circ, \pm 45^\circ, \pm 90^\circ, \pm 135^\circ, 180^\circ$, are calculated in order to obtain rotation invariant features. It should be noted that the images are at the same magnification level; therefore, no scaling of the features is needed.

- **Feature Vector Representation**

Once all color and texture features are extracted, they are concatenated to form a 106-dimensional feature vector. After feature extraction, a Z-score

normalization is applied to each extracted feature in the feature vector by subtracting the mean of that feature followed by dividing to the standard deviation of that feature computed over the reference dataset. This normalization step converts all extracted features to a common scale with an average of zero and standard deviation of one. Then normalized feature vectors (*NF*) are stored for further CBIR processes. For each query image set, the system will employ the same feature extraction and normalization procedure to the query images.

Instead of analyzing the contribution of extracted features based on selected color spaces or texture features by using feature selection algorithms, we preferred to use subspace projection method in order to represent the feature vectors more sparsely by decreasing the correlation among the features. In the literature, subspace projection methods have been widely used for dimensionality reduction and feature extraction. They are popular to analyze structures where large amount of correlated numerical data is available. Nonnegative matrix factorization (NMF) [32] is one of the data-driven subspace projection methods, which aims to factorize a data matrix into basis vectors and their combiner coefficients. They perform better for features extracted from partially represented data [33]. In our case, features from different color spaces and texture features can be assumed to be features of a partially represented data. Using a training dataset, *FDS* with size *lxT*, the *m* basis vectors, columns of *W*, are obtained as follows:

$$FDS \approx WH$$

where *l* is the length of the feature vector, *T* is the number of samples in the dataset, and *m* ($m < l$) is the size of NMF features. In the factorization in (1), the columns of the *lxm* matrix *W* stand for the basis vectors and the columns of the *mxT* matrix *H* determine how the basis vectors are activated to reconstruct the feature matrix *FDS*. The columns of *H* represent the NMF-based feature vectors of the corresponding data. The classification of a test feature vector *FQ* is based on its NMF features given by $h = W+FQ$. The number of columns *m* in the (basis) matrix *W* was determined for each disease type empirically during training stage.

- **Two Tier Approach for Multi-image queries**

Our CBIR system operates at two tiers. In the first tier, the designed classifier categorizes the query image/images into one of the major disease types such as FL and NB. Once the disease category of the image is determined, the search for the query image can be carried out among the category relevant subtypes in the subsequent tier. For example, when the query image belongs to NB disease, database images in the first tier will be filtered according to the NB disease category. Then the subsequent search will be only performed on the NB category subset to retrieve the images from the correct category of the query images.

In the second tier, we will use our proposed multi-image query and retrieval methodology to retrieve the images from the reference database in the order of their image-level visual similarities by preserving the slide-level semantic similarity.

A. First Tier: Classification of Disease Type with SVM

An SVM-type classifier was employed to categorize the query image into one of the major disease type such as NB or FL using the extracted features. SVM classifiers are well founded in statistical learning theory and have been successfully used for various classification tasks in computer vision. Their purpose is to find a decision hyperplane for a binary classification problem by maximizing the margin, which is the distance between the hyperplane and the closest data points of each class in the training set that are called support vectors. The hyperplane is chosen among all the possible hyperplanes through a complex combinatorial problem optimization so that it maximizes the distance (called the margin) between each class and the hyperplane itself. As SVMs are restricted to binary classification, several strategies are developed to adapt them for multiclass classification problems such as one-against-all classification and pair-wise classification.

B. Second Tier: Slide-Level Image Retrieval

In this part of the CBIR algorithm, we proposed a two-level retrieval system; in the first level, the search is performed similar to traditional CBIR systems such that the images are retrieved based on their image-level similarities. In the second level, the images will be retrieved according to their similarities in the slide level. Once the category of the query image is detected in the first tier, further search is performed on the pre-filtered database, which includes only the sample images of the detected disease category. As we described in Section V-A, each disease has higher level semantic annotations based on their histological grades such as Grade-I, Grade-II, and Grade-III in FL disease or D levels such as SR, UD, PD, and D in NB disease. Therefore, it is necessary to retrieve images related to their higher level semantic characteristics in order to provide more accurate results to the user of the CBIR system. Algorithm 1 summarizes the image-level search and Fig. 2 illustrates a sample nearest-neighbor search scheme for a given query mage set in imagelevel.

Algorithm 1 Image-level Retrieval

for for a given query image-set Q, with K retrieval **do**
  $Image\_Score(1 \dots T) - [0 \dots 0]$
  // Initially score of each image in the corresponding // dataset equals to zero.
  for n = 1 to N **do**
    for t = 1 to T **do**
      $distance(n,t) = Dist(F_{Q^n}, F_{DS^t})$
    end for
    $ind = sort(distance)$ in descending order,
    Retrieve and display the K-closest images to the user
    and
    $Image\_Score(ind[1:K]) = Image\_Score(ind[1:K]) + 1$
    // Add 1 to the score of K-nearest images which has // the smallest distance from the corresponding dataset to // the query image
  end for
  // Scores are accumulated if $N > 1$
end for

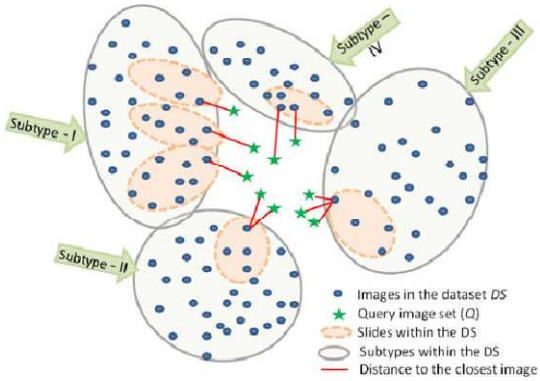

Fig 2: Sample image-level nearest neighbor search scheme for a given query image set.

Here we used the term of image set in order to represent multiple images in one query. Note that image set may include only one image or several images cropped from one tissue slide. The distance between each image of query *Q* and the individual images in the dataset are computed using the correlation distance measure, as shown in the following:

$$\begin{aligned} \mathrm{Dist}(F_{Q^n}, F_{DS^t}) &= 1 - \{\mathrm{Correlation}\{F_{Q^n}, F_{DS^t}\}\} \\ &= 1 - \left\{ \frac{\langle (F_{Q^n}), (F_{DS^t}) \rangle}{\|F_{Q^n}\| \|F_{DS^t}\|} \right\}, \\ t = 1, \dots, T,\ T = |F_{DS}| \text{ and } n = 1, \dots, N,\ N = |Q| \end{aligned} \quad (2)$$

where *N* is the number of individual query entities in the given query image set *Q*, *T* is the number of images in the reference dataset *DS*, and *FQn* represents the feature vector of the n^{th} query image, *FDSt* represents the feature vector of the t^{th} image of the given dataset, and $|\cdot|$ is the cardinality.

Algorithm 1 provides us the frequency of similar images per image in the dataset to a given query image set or a slide in terms of scores. Scores are computed by summing the number of occurrences of each image in the dataset for a *k*-nearest neighbor (KNN) search of that query image set. The output of this algorithm is the traditional image-level-based retrieving of most similar images from the given dataset and their image-level scores.

Algorithm 2 Slide-level Retrieval

for for a given query image-set Q, with K image-level retrieval **do**
  Perform Algorithm 1
  // Do not display the retrieved images to the user
end for
for c = 1 to C **do**
  for s = 1 to S^c **do**
    $Score\_itf^{(c,s)} = \sum_{i=1}^{I^{(c,s)}} Image\_Score_i^{c,s} / I^{(c,s)}$
  end for
end for
// Sort $Score\_itf$ in descending order and select top $K2$ to // compute $Rank\_weight$
for c = 1 to C **do**
  $Rank\_weight^{(c)} = \sum_{k=1}^{K2} Sorted\_Score\_itf^{(c,k)}$
end for
for c = 1 to C **do**
  for s = 1 to S^c **do**
    $Weighted\_Score^{(c,s)} = Score\_itf^{(c,s)} * Subtype\_isf^c * Rank\_weight^{(c)}$
  end for
end for
// Sort the weighted scores in descending order
// Display the user n-highest scored slides

In our alternative approach to image-level retrieval, we propose to retrieve similar images from the database by keeping the slide-level semantic grade among the retrieved images. For this purpose, we introduced a slide-level retrieval methodology, which is summarized in Algorithm 2.

In our proposed approach, the first step is to scale the score of each slide by assigning different weight parameters based on subtype frequencies over the reference database. Therefore, our algorithm assigns higher weights to the slides of FL Grade-I since its frequency is lower than FL Grade III. Similarly, the number of images per slide is varying among the slides. In order to make a comprehensive and intelligent relevance ranking system, it is necessary to take into account those statistical variations among slides and subtypes. Assigning weights to each slide and to each subtype based on the distribution (or frequency) of images per slide and distribution of slides per subtype is motivated by similar approaches in IR theory. In information theory, "*tf*" refers to the frequency of an index term in a reference document and "*idf*" is inversely proportional to the number of documents containing that index term and they are

used to assign weights for each term of the documents before computing similarity.

However, in our case, we do not have definite terms (i.e. words in documents), but we have scores representing the unweighted similarities between the query image set and the reference slides. Therefore, we adapted these concepts to assign weights to normalize the similarity scores of each slide and each subtype depending on the slide- and image-level statistics of the dataset (e.g., the number of images per slide or the number of slides per subtype). In our slide-level retrieval system, we redefined scores in terms of image term frequency (*itf*), which corresponds to normalized number of image count of a particular slide for a given query set.
Additionally, inverse slide frequency (*isf*) is inversely proportional to the number of slides per subtype, and it gives lower weights to the slides occurring in a larger set of subtype.

Algorithm2 summarizes the proposed weighting score approach. We assigned a weighting term, called *Rank weight*, to each subtype. First, *Score itf* values were sorted in descending order and top *K*2 of the sorted scores are summed according to their subtypes. These summed scores represent the *Rank weight* term for each subtype. Basically, *Rank weight* term corresponds to the proportion of summed *itf* scores within the top *K*2 *itf* scores per subtype. The purpose of *Rank weight* is to increase the likelihood of retrieving the subtype of the highest scored slides by assigning higher weights to the slide scores of that subtype

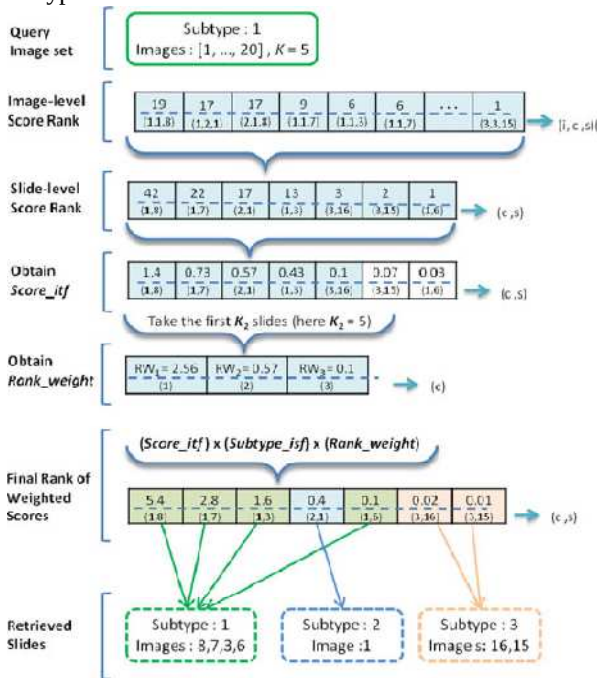

Fig. 3. Computational model representing the transition from image-level scores to slide-level retrieval, where *i* = image number, *c* = subtype number, and *s* = slide number. Here, the query *Q* is an image set with 20 images belonging to subtype 1. Image-level scores, slide-level scores, *Score itf* , *Rank weight*, relevancy rank of slides with weighted scores are computed, respectively, for the given sample query.

Experimental Results

Annotated Microscopic Image Dataset

The number of cropped images per slide is between 11 and 30 for FL cases and between 7 and 35 for NB cases. For FL slides, a team of experienced hemato-pathologists selected about 10 random microscopic high power fields (HPF) to interpret the disease grade in terms of the average number of centroblasts per HPF. Note that, for both FL and NB, we use internationally accepted and clinically practiced standards. For FL, our collaborating pathologists use the World Health Organization grading system. The consensus of pathologists is used to stratify cases into their histological grades. The sizes of the cropped images are 1353 × 2168 pixels for FL cases and 1024 × 1024 and 1712 × 952 pixels and for NB cases.

TABLE II
AVERAGE CLASSIFICATION RESULTS (%) FOR THE FIRST TIER OVER 50 TEST REPETITIONS

	Fl. - image	NB - image	Fl. - slide	NB - slide
FL	99.4	0.6	100	0
NB	5.3	94.7	2.3	97.7

For NB slides, pathologists pick the representative regions (images) from the whole slide and examine those regions at higher magnifications. The final decision for the differentiation grade of the entire slide is based on the grades of the sample images selected from that slide. Due to this differentiation grades, NB disease is differentiated to two subcategories such as SR and SP. SP subtype has three more subtypes such as D, PD, and UD. In total, NB disease has four subtypes. Fig. 5 illustrates sample images cropped from different slides with different differentiation grades of NB to give an idea about their visual appearances.
Because of the heterogeneous characteristic of these tumors, all image-level annotations may not match with the annotation of the entire slide, which causes intraslide variations. Additionally, there may be variations among inter-and intrareadings of pathologists because of which FL Grade-I and FL Grade-II subtypes and NB-PD and NB-UD subtypes [26], are the most confused subtypes of the FL and NB diseases, respectively.

Results of the First Tier

The organization of the test set and training set is performed in patient (slide) independent manner. In other words, none of the images of a test slide is included in the training set in order to obtain more realistic results both in the first -and second-tier experiments.

In the experiments of the first-tier evaluation, we randomly selected five FL and five NB slides for each test set and the remaining slides were used for the training of the SVM classifier. In total, 10 slides were randomly chosen for the test set and 91 slides were used for training. In order to comprehensively test and train all NB and FL sample slides with different test sets, we repeated the testing scheme until all the slides were used as a test slide, an approach similar to the leave-one-out testing.

These results were obtained with SVM classifiers trained with normalized features. The classification accuracies were evaluated in two different ways. One way is to evaluate the results at the image level such that each image is classified independent from the other images of that test slide. The other way is to interpret the results at the slide level by combining the decisions on all images of a test slide using decision fusion rules. Here, the majority rule is employed to the assigned classes of the test images to determine the slide-level classification of that image set. In other words, the majority of the assigned classes for each test image are chosen as the representative class for that given slide. It is noticed that all images of that NB slide were also misclassified at the image level. This misclassified NB slide was used with both NB and FL slides in order to evaluate the retrieval accuracy in the second tier of the algorithm in case of a misclassified slide.

Results of the Second Tier

After determining the classes of query slides in the first tier, the next step is to retrieve the most relevant images from the database according to the main disease type of the query image set. Leave-one-slide-out cross-validation testing scheme was employed for each disease type separately such that at each round one tissue slide with all corresponding images was used as a query image set and the images of the remaining slides were used as the reference dataset for that query.

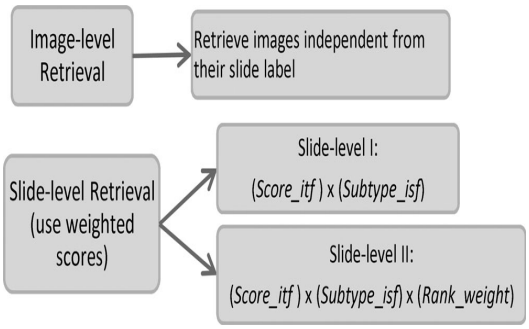

Fig 4: Second-tier experimental scheme.

The organization of the performed experiments for the second tier is shown in Fig. 4. For the slide-level retrieval, we used the proposed weighted scores to rank the slides according to their relevancy to a given query slide. In order to assess the performance of *Rank weight* on the retrieval system, we evaluated the experiments both with (slide-level II) and without (slide-level I) using this weighting term.

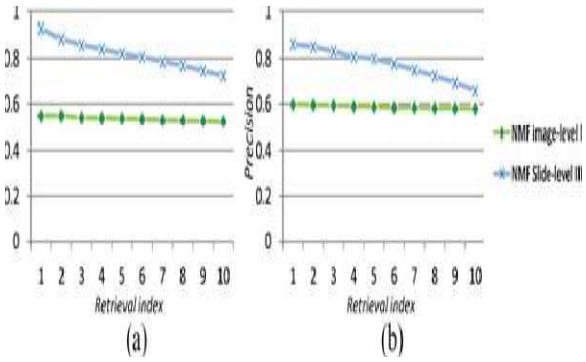

Fig. 5. Comparison of average precision values for slide-level II and image-level retrieval algorithms for FL and NB diseases. (a) FL disease. (b)NB disease.

Conclusion

In this paper, we have presented a novel content-based microscopic image/slide retrieval algorithm. We have demonstrated that by using the proposed weighting scheme inspired by IR theory, the slide-level retrieval performance of the CBIR system is considerably better than the traditional image-level retrieval accuracy for all seven subtypes of two challenging diseases, which have inter- and intrareading semantic variations, intraslide semantic variations, and intersubtype visual similarities. In the first tier, only one slide among 44 NB slides is misclassified, and in the second tier, about 26 percentage points of improvement was achieved on the classification accuracy at the first rank retrieval over all diseases by using the proposed score weighting strategy. This CBIR system can enable the user, e.g., a pathologist, to select multiple HPF regions from a suspected tissue and submit those

images as a query to the CBIR system and retrieve the most relevant slides with their semantic annotations with higher accuracies. The results, achieved under those challenging conditions, are also promising for automatic and unsupervised selected query images based on their HPF regions. Application of the proposed weighting strategy, inspired by the IR theory, is not limited to microscopic images only, and can be also useful for any type of multiquery search and content-based retrieval systems.

In our future work, we will 1) investigate more effective texture and color feature extraction methods, 2) improve the robustness of the system by increasing the number of patients/slides in the database, 3) enhance the diversity of the database by including microscopic images from different disease types, 4) evaluate the performance of the system on automatically selected HPF regions for the query, and finally 5) develop a multipurpose Web-based tool for training future generations of researchers by providing consistent, relevant, and always available support and assistance for the challenging diseases, and finally help cancer researchers in better understanding of cancer development, treatment monitoring, and clinical trials.

References

[1] R. H. Choplin, J. M. Boehme, and C. D. Maynard, "Picture archiving and communication systems: An overview," *Radiographics*, vol. 12, no. 1, pp. 127–129, 1992.

[2] H. Muller, N. Michoux, D. Bandon, and A. Geissbuhler, "A review of content-based image retrieval systems in medical applications clinical benefits and future directions," *Int. J. Med. Informat.*, vol. 73, no. 1, pp. 1–23, 2004.

[3] W. Hsu, L. R. Long, and S. Antani, "Spirs: A framework for content-based image retrieval from large biomedical databases," *Stud. Health Technol. Informat.*, vol. 129, no. Pt. 1, pp. 188–192, 2007.

[4] H. L. Tang, R. Hanka, and H. H. S. Ip, "Histological image retrieval based on semantic content analysis," *IEEE Trans. Inf. Technol. Biomed.*, vol. 7, no. 1, pp. 26–36, Mar. 2003.

[5] C.-R. Shyu, C. E. Brodley, A. C. Kak, A. Kosaka, A. M. Aisen, and L. S. Broderick, "Assert: A physician-in-the-loop content-based retrieval system for HRCT image databases," *Comput. Vis. Image Understand.*, vol. 75, no. 1–2, pp. 111–132, 1999.

[6] L. Zheng, A. Wetzel, J. Gilbertson, and M. Becich, "Design and analysis of a content-based pathology image retrieval system," *IEEE Trans. Inf. Technol. Biomed.*, vol. 7, no. 4, pp. 249–255, Dec. 2003.

[7] L. Yang, O. Tuzel,W. Chen, P.Meer, G. Salaru, L. Goodell, and D. Foran, "Pathminer:Aweb-based tool for computer-assisted diagnostics in pathology," *IEEE Trans. Inf. Technol. Biomed.*, vol. 13, no. 3, pp. 291–299, May 2009.

[8] S. A. Napel, C. F. Beaulieu, C. Rodriguez, J. Cui, J. Xu, A. Gupta, D. Korenblum, H. Greenspan, Y. Ma, and D. L. Rubin, "Automated retrieval of ct images of liver lesions on the basis of image similarity: Method and preliminary results.," *Radiology*, vol. 256, no. 1, pp. 243–252, 2010.

[9] M. M. Rahman, P. Bhattacharya, and B. C. Desai, "A framework for medical image retrieval using machine learning and statistical similarity matching techniques with relevance feedback," *IEEE Trans. Inf. Technol. Biomed.*, vol. 11, no. 1, pp. 58–69, Jan. 2007.

[10] C. Akgul, D. Rubin, S. Napel, C. Beaulieu, H. Greenspan, and B. Acar, "Content-based image retrieval in radiology: Current status and future directions," *J. Digital Imag.*, vol. 24, no. 2, pp. 208–222, Apr. 2011.

[11] O. A. OConnor and J. M. Vose, "Indolent follicular lymphoma." Lymphoma Research Foundation, 2008.

[12] K. Belkacem-Boussaid, S. Samsi, G. Lozanski, and M. Gurcan, "Automatic detection of follicular regions in h&e images using iterative shape index," *Comput. Med. Imag. Graph*, vol. 35, pp. 592–602, 2011.

[13] H. Shimada, I. M. Ambros, L. P. Dehner, J.-I. Hata, V. V. Joshi, B. Roald, D. O. Stram, R. B. Gerbing, J. N. Lukens, K. K.Matthay, and R. P. Castleberry, "The international neuroblastoma pathology classification (the Shimada system)," *Cancer*, vol. 86, no. 2, pp. 364–372, 1999.

[14] R. Baeza-Yates and B. Ribeiro-Neto, *Modern Information Retrieval*, 1st ed. New York: Addison Wesley, May 1999.

[15] Q. Iqbal and J. K. Aggarwal, "Feature integration, multi-image queries and relevance feedback in image retrieval," in *Proc. 6th Int. Conf. Vis. Inf. Syst. DMS*, 2003, pp. 467–474.

[16] T. Bjoerge and E. Chang, "Why one example is not enough for an image query," in *Proc. IEEE Int. Conf. Multimedia Expo*, 2004, vol. 1, pp. 253– 256.

[17] J. Naik, S. Doyle, A. Basavanhally, S. Ganesan, M. D. Feldman, J. E. Tomaszewski, and A. Madabhushi "A boosted distance metric: Application to content based image retrieval and classification of digitized histopathology," in *SPIE Medical Imaging: Computer-Aided Diagnosis*, 2009.

[18] T. Deserno, S. Antani, and R. Long, "Ontology of gaps in content-based image retrieval," *J. Digital Imag.*, vol. 22, pp. 202–215, 2009.

[19] X. S. Zhou, S. Zillner, M. Moeller, M. Sintek, Y. Zhan, A. Krishnan, and A. Gupta, "Semantics and CBIR: A medical imaging perspective," in *Proc. 2008 Int. Conf. Content-Based Image Video Retrieval*, New York: ACM Press, 2008, pp. 571–580.

[20] D. Iakovidis, N. Pelekis, E. Kotsifakos, I. Kopanakis, H. Karanikas, and Y. Theodoridis, "A pattern similarity scheme formedical image retrieval," *IEEE Trans. Inf. Technol. Biomed.*, vol. 13, no. 4, pp. 442–450, Jul. 2009.

[21] H. Pourghassem and H. Ghassemian, "Content-based medical image classification using a new hierarchical merging scheme," *Comput. Med. Imag.Graph.*, vol. 32, no. 8, pp. 651–661, 2008.

[22] A. CruzRoa, J. Caicedo, and F. Gonzlez, "Visual pattern analysis in histopathology images using bag of features," in *Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications*, 2009, pp. 521–528.

[23] N. Mehta, R. S. Alomari, and V. Chaudhary, "Content based sub-image retrieval system for high resolution pathology images using salient interest points," *Int. Conf. Proc IEEE EMBS*, vol. 1, pp. 3719–3722, 2009.

[24] R. C. Gonzalez and R. E. Woods, *Digital Image Processing*, 2nd ed. Boston, MA: Addison-Wesley/Longman, 2001.

[25] Y. D. Chun, N. C. Kim, and I. H. Jang, "Content-based image retrieval using multiresolution color and texture features," *IEEE Trans. Multimedia*, vol. 10, no. 6, pp. 1073–1084, Oct. 2008.

[26] R. M. Haralick, K. Shanmugam, and I. Dinstein, "Textural features for image classification," *IEEE Trans. Syst., Man Cybern.*, vol. 3, no. 6, pp. 610–621, Nov. 1973.

[27] M. M. Rahman, S. K. Antani, and G. R. Thoma, "A classification-driven similarity matching framework for retrieval of biomedical images," in *Proc. Int. Conf. Multimedia Inf. Retrieval*, New York: ACM Press, 2010, pp. 147–154.

[28] M. Unser, "Sum and difference histograms for texture classification," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 8, no. 1, pp. 118–125, Jan. 1986.

[29] D. D. Lee and H. S. Seung, "Learning the parts of objects by non-negative matrix factorization," *Nature*, vol. 401, no. 6755, pp. 788–791, 1999.

[30] C.-J. Lin, "Projected gradient methods for non-negative matrix factorization," *Neural Comput.*, vol. 19, pp. 2756–2779, 2007.

[31] C.-W.Hsu andC.-J. Lin, "A comparison of methods for multi-class support vector machines," *IEEE Trans. Neural Netw.*, vol. 13, no. 2, pp. 415–425, Mar. 2002.

[32] C.-C. Chang and C.-J. Lin. (2001). *LIBSVM: A library for support vector machines*. Software available: http://www.csie.ntu.edu