

# Biomedical Article Retrieval Using Multimodal Features and Image Annotations in Region-based CBIR

Daekeun You<sup>a</sup>, Sameer Antani, Dina Demner-Fushman, Md Mahmudur Rahman,  
Venu Govindaraju<sup>a</sup>, George R. Thoma

National Library of Medicine, National Institutes of Health, Bethesda, MD 20894

<sup>a</sup>CUBS, Dept. of Computer Science and Engineering, SUNY at Buffalo, Buffalo, NY 14260

## ABSTRACT

Biomedical images are invaluable in establishing diagnosis, acquiring technical skills, and implementing best practices in many areas of medicine. At present, images needed for instructional purposes or in support of clinical decisions appear in specialized databases and in biomedical articles, and are often not easily accessible to retrieval tools. Our goal is to automatically annotate images extracted from scientific publications with respect to their usefulness for clinical decision support and instructional purposes, and project the annotations onto images stored in databases by linking images through content-based image similarity.

Authors often use text labels and pointers overlaid on figures and illustrations in the articles to highlight regions of interest (ROI). These annotations are then referenced in the caption text or figure citations in the article text. In previous research we have developed two methods (a heuristic and dynamic time warping-based methods) for localizing and recognizing such pointers on biomedical images. In this work, we add robustness to our previous efforts by using a machine learning based approach to localizing and recognizing the pointers. Identifying these can assist in extracting relevant image content at regions within the image that are likely to be highly relevant to the discussion in the article text. Image regions can then be annotated using biomedical concepts from extracted snippets of text pertaining to images in scientific biomedical articles that are identified using National Library of Medicine's Unified Medical Language System<sup>®</sup> (UMLS) Metathesaurus. The resulting regional annotation and extracted image content are then used as indices for biomedical article retrieval using the multimodal features and region-based content-based image retrieval (CBIR) techniques. The hypothesis that such an approach would improve biomedical document retrieval is validated through experiments on an expert-marked biomedical article dataset.

**Keywords:** Biomedical image analysis, biomedical article retrieval, content-based image retrieval, image overlay extraction, pointer symbol extraction, figure caption text analysis

## 1. INTRODUCTION

Clinicians and medical researchers routinely use online databases such as MEDLINE<sup>®</sup> to search for bibliographic citations that are relevant to a clinical situation. The biomedical evidence they seek is available through clinical decision support systems (CDSS) that use text-based retrieval enhanced with biomedical concepts. However text information is sometimes insufficient in determining the usefulness of a publication. Authors of biomedical publications frequently use images to illustrate the medical concepts or to highlight special cases. These images often convey essential information and can be very valuable for improved clinical decision support (CDS) and education. The text-based retrieval of the images has, so far, been limited mostly to caption and/or citation information. To be of greater value, images in scientific publications need to be first annotated (preferably, automatically) with respect to their usefulness for CDS to help determine relevance to a clinical query or to queries for special cases important in educational settings [1-3].

This article discusses a method for multimodal image annotation that utilizes (i) image analysis techniques for

localization and recognition of author provided overlays on the images; (ii) image feature extraction methods for content-based image retrieval (CBIR); (iii) natural language processing techniques for identifying key terms in the title, abstract, figure caption, and figure citation (mention) in the article; and (iv) use of structured vocabularies, such as the National Library of Medicine’s Unified Medical Language System (UMLS<sup>®</sup>), for identifying the biomedical concepts in the text. As discussed in earlier works [4,5], these steps can be used to associate the biomedical concepts in the text to specific regions in the image. The relevance to a clinical query is aided by this addition of semantic information to extracted image features for improved CBIR. Traditionally, CBIR tends to be limited to use of visual features in identifying similarity among a collection of images. This has spurred discussion on the “semantic gap” [6] that is introduced when high-level concepts are represented through low-level visual features such as image color, and texture (for example). Such a semantic gap can be minimized through annotation by biomedical concepts that are extracted from the article text and applied to relevant regions within an image.

Conventional approaches for biomedical journal article retrieval have been text-based with little attention devoted to the use of images in the articles. Figure 1 tries to encapsulate this conventional approach and place it in context with the proposed multimodal method. Text-based retrieval is shown in the upper part of the Figure 1. The middle shaded blocks show our proposed retrieval approach modeled as a cascaded application of text and image feature matching methods. The hypothesis is that text retrieval can be very valuable in identifying relevant articles. Image regions of interest (ROI) are extracted by localizing and recognizing image annotations (such as pointers, symbols) overlaid by authors to indicate the ROIs and mentioned in image captions and discussions. The ROIs can then be used to rank the relevant articles identified by the text retrieval by measuring region-based image similarity.

General content-based image retrieval (CBIR) also could be improved by the proposed approach in a similar manner as text-based retrieval is improved. In this case no text information is available, but only visual features are used. The CBIR identifies relevant articles as text-based retrieval does in the multimodal method. Annotations and ROIs in retrieved images can be identified by the annotation recognizer and then be used to re-rank the results. The bottom part of the Figure 1 shows a general CBIR approach followed by the proposed annotation analysis.

This article presents our new method for localizing and recognizing such image annotation overlays, such as pointers, (shown as MRF recognizer in Figure 1) and develops all components of the proposed framework as a pilot to the development of an extended retrieval system. The retrieval of images and linked articles will be demonstrated and improvement in their retrieval will be measured against text-only approaches. The remainder of this article is organized as follows. Section 2 provides a brief background of our previous work on the proposed retrieval system. Section 3 describes the method and challenges in this research. Section 4 presents evaluation results of pointer recognition and retrieval test and section 5 discusses conclusions and future work.

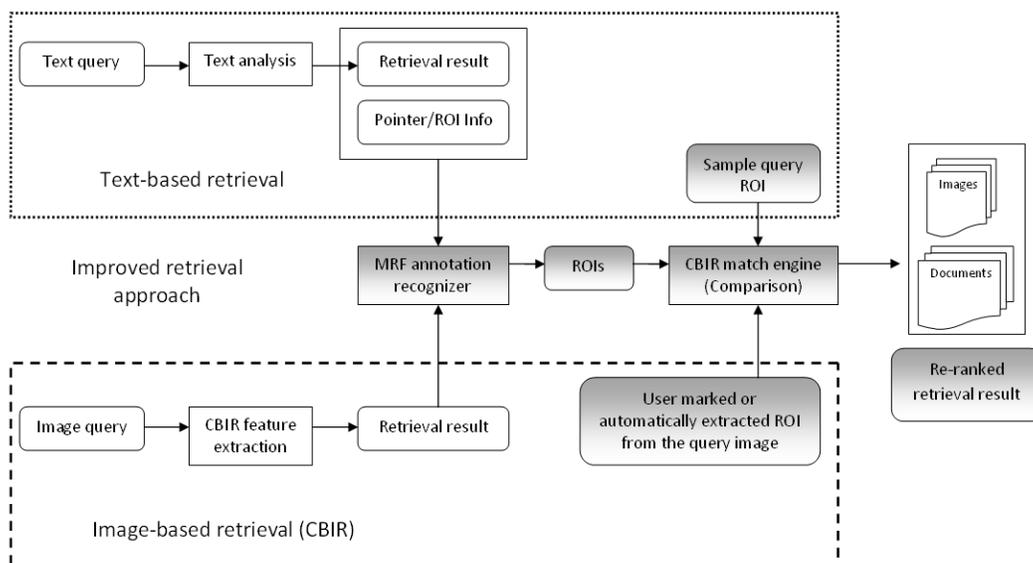


Figure 1. Overview of biomedical journal article retrieval system

## 2. BACKGROUND

In our earlier work on coarse automatic indexing of images by modality (color image, gray-scale image, graph, graphic illustration, etc.) and image utility (suggested by the Evidence Based Medicine paradigm’s six elements of a clinical scenario that an image might illustrate), we combined image and textual features in a supervised machine learning approach. Textual features were obtained from the captions to the images and paragraphs of text containing discussion (“mentions”) of these images. In addition to image captions and mentions, more image-related information may appear in the title, abstract, and MeSH terms (assigned by expert indexers to describe the publication and provided in MEDLINE citations) [2]. Texture features were computed as a 3-level discrete 2-D Daubechies wavelet transform. The four most dominant colors were computed in the perceptually uniform CIE LUV color space and proved most effective. At this coarse level of granularity, a multi-class SVM classifier trained on a bag-of-words representation of image captions performed better in determining *image modality* ( $84.3\% \pm 2.6\%$  accuracy) than when trained on a combination of textual and image features or features reduced to the domain specific vocabulary. For *image utility*, however, the combination of image and textual features was better than any single-source feature set achieving  $76.6\% \pm 4.2\%$  accuracy [1].

Figure 2 shows image annotation and retrieval processes [3]. We combined our tools and those publicly available in a pipeline that starts with text and image pre-processing and ends with retrieving images that are ranked by relevance to a given information need or annotated as relevant (Tools developed by the authors are shown in double-bordered boxes). In this work, we compare the efficacy of these two approaches in finding clinically relevant images.

Our previous approach to locating and recognizing pointers in biomedical images proposed a two-step method: (i) pointer segmentation, and (ii) pointer classification [4]. The pointer segmentation step applied edge detection to an input image and then binarized the edge image. Chain coding scheme followed by line segment extraction from the chain coded boundary was introduced to extract the pointer boundary. Then several boundary pre-processing methods were applied to handle the affine transformations. Dynamic time warping (DTW) based line segment matching was used in the pointer classification step. Two line segments, one from the input boundary and the other from the pointer template, were compared and similarity score of two boundaries was obtained from the DTW distance table. The method was evaluated using commonly used measures: missed detection and false hit. Main causes of the missed detection were identified as weak edges and errors in the rotation axis detection. The line segment approximation and the DTW matching algorithm were primarily responsible for the false hits. Overall, the method resulted in an average precision of 92.3% and 75.3% average recall.

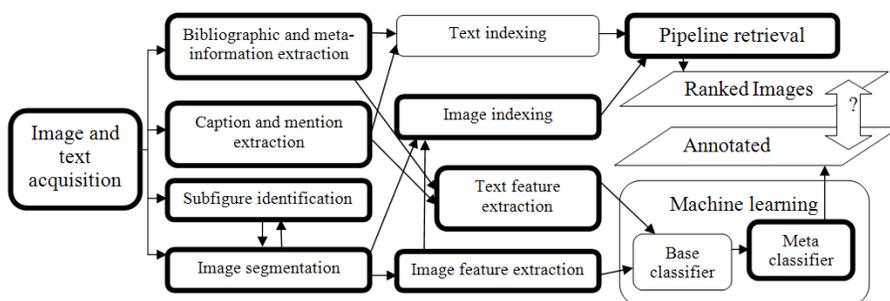


Figure 2. Image annotation and retrieval processes

## 3. METHODS

### 3.1 Challenges

We have recently developed a template matching-based pointer recognition method as mentioned in section 2. Our prior work showed fairly good performance on some types of pointers such as (straight) arrows and arrowheads; however, it showed some limitations as well on many curved arrows that are frequently used in biomedical images. We identified two main challenges in developing the pointer recognition algorithm based on our prior work and research to date.

- i) Arbitrary pointing direction: One of the most difficult tasks is how to handle the rotation problem. Pointers may indicate arbitrary directions and finding the pointing direction was an important step in our prior work. We failed to recognize a pointer unless we found its rotation axis, and rotated the pointer correctly to match with the templates. Many model-based object detection and recognition approaches consider the rotation problem. Solutions include defining rotation-invariant features and the range of rotations that object models can handle [7].
- ii) Variety in pointer shape: From our text analysis results we found that three terms viz., (straight) arrow (SA), curved arrow (CA), and arrowheads (AH), are frequently used in figure captions and text discussion to call the pointers used in biomedical images. Besides those arrow type pointers, symbols such as asterisks are used frequently. Figure 3 shows various curved arrows. We may find more curved arrows not shown in the figure and the number may increase.



Figure 3. Various curved arrows

### 3.2 Pointer recognition method

We propose a Markov random field (MRF)-based method to address the challenges. MRFs have been widely used in image analysis and successful for some applications such as texture analysis [8], image segmentation [9], image restoration [10], and binarization [11]. Most of the successful applications, however, are tasks of low-level image analysis, where the features are derived from image pixel intensity. Relatively less research has been done in high-level vision tasks such as object matching and recognition [12]. Here, one deals with more abstract features such as critical points, line segments, and surface patches/regions.

In our MRF framework, we model a pointer boundary as a random field which consists of a number of line segments. A pointer boundary is approximated by a set of line segments and there exist some contextual dependencies among neighboring line segments. Instead of defining pointer models and modeling the contextual dependency among all line segments in a pointer similarly as other methods did, we identified several boundary parts consisting of three consecutive line segments (part unit) and found frequently from the pointers, and created labels for them. Also contextual dependency between local segments and their labels are established and used as prior knowledge in our MRF model. A line segment will be labeled by i) identifying several similar boundary parts (i.e., candidates) in the label set with the line segment and its neighboring segments, ii) examining its relationship with neighboring segments and their labels, and iii) finding an optimal label (as part of an optimal labeling configuration) through the MRF model.

Since we label a line segment based on its neighborhood and do not use any pointer model, an unknown boundary cannot be identified directly from the MRF labeling result, i.e., labeling configuration, no matter how well the line segments are labeled. Hence a method such as classifying a labeling configuration into a proper object class is needed. A hidden Markov model (HMM)-based labeling configuration classifier is implemented in our method [13]. We observed that labeling configurations from a same pointer class consist of similar labels and their sequence is very similar, but quite different from those of other classes. The labeling configuration is used as a data sequence of HMM, and we found that HMM is suitable for our purpose.

Our combined MRF-HMM labeling-classification method has some benefits compared to the other MRF-based object detection/recognition methods [14, 15]. Defining proper pointer classes in our application may not be easy since the variation in shape is quite large as shown in Figure 3. Even if some of them can be grouped and handled by the same class, deciding which pointers belong to the same class is not a simple task either. Moreover it is almost impossible to find and include all pointers used in biomedical images in our training set. For these reasons pointer model-based approaches could not be the best solution even though they have been widely used in many other applications. Also, as mentioned before, arrow type pointers share many parts of their boundary shape with other shapes. This means feature extraction and classification using boundary parts may be much easier than doing the same thing with the whole boundary.

For the optimization in our MRF labeling, belief propagation (BF) is applied to find an optimal labeling configuration. Eq. 1 shows the belief  $b_i(f_i)$  and message update rule.

$$\begin{aligned}
b_i(f_i) &= Z_i^{-1} r_i(f_i) \prod_{j \in N_i} m_{i,j}(f_i) \\
m_{i,j}(f_j) &\leftarrow \alpha_{i,j} \max_{f_i} r_i(f_i) r_{i,j}(f_i, f_j) \prod_{k \in N_i - \{j\}} m_{k,i}(f_i)
\end{aligned} \tag{1}$$

where  $Z_i^{-1}$  and  $\alpha_{i,j}$  are normalizing constants.  $f_i$  is the label of segment  $i$  and  $r_i(f_i)$  and  $r_{i,j}(f_i, f_j)$  are unary and binary compatibility functions, respectively. All the terms and related issues will be discussed below.

### Pre-processing

The pre-processing methods used in our prior work [4] are used with our new pointer recognition method. A polygon approximation function in Intel Open Source Computer Vision Library (OpenCV) [16] was used to extract line segments from the chain coded boundary.

### Label set

Figure 4 shows 43 labels defined in our method. The numbers used as labels have no specific meaning other than the first digit is same for the line segments from the same part of the pointer shape, e.g., 3xx for line segments in arrow tail. A label is assigned to the middle segment (shown as a solid line) of a three adjacent segments in consideration. The layout of the neighboring segments (shown in dashed line), in relation to the middle segment determines the label.

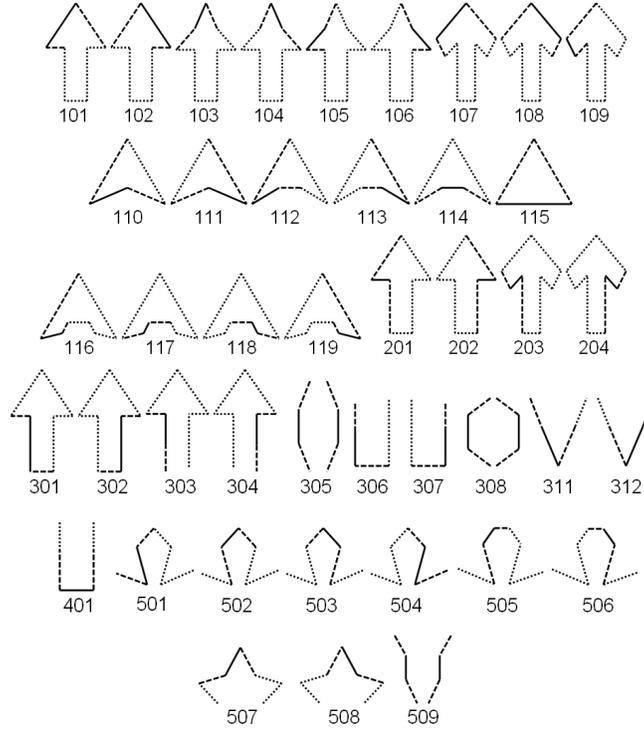


Figure 4. Label set defined from pointers

### Neighborhood system

Neighborhood system  $N_i$  includes two immediate neighboring segments of segment  $i$ . We obtain an ordered sequence of line segments from a boundary and only two segments, i.e. segments of indices  $i-1$  and  $i+1$ , are considered in the neighborhood system. Considering two direct neighboring segments, however, causes some problems in MRF labeling. If a necessary segment is not extracted or a noise segment is added to the neighborhood of a segment the segment cannot be correctly labeled. Figure 5 illustrates these problems. (a) shows a missing segment and (b) shows some extra noise

segments. We solved this problem by considering two indices (segments) from previous and next of index  $i$ . That is four pairs,  $(i-1, i+1)$ ,  $(i-1, i+2)$ ,  $(i-2, i+1)$ , and  $(i-2, i+2)$ , are considered as  $N_i$  and when index  $i-2$  or  $i+2$  is included in the  $N_i$  the  $i-1$  or  $i+1$  segment, respectively, is assumed as noise or assumed not being extracted. With this extended neighborhood system we can use smaller threshold for noise removal and save many small necessary segments which were removed before (see case (a)). The neighborhood system, however, may increase noise segments which were removed with larger threshold (center pointer in (b)). But small noise segments do not cause labeling problem in the extended neighborhood system unless both neighboring segments, i.e.  $i-1$  and  $i-2$  or  $i+1$  and  $i+2$ , are noise segments, which is a rare case. The labeling results of the two pointers in (b) were exactly the same and all noise segments in extended neighborhood were removed from the final configuration.

We identified all pairs of possible neighboring labels of each label and used them in computing compatibility functions and messages in BF. Among all possible labels of a certain label  $l$ , some labels cannot appear at the same time in the labeling configuration and those pairs should be identified and considered in labeling algorithm. A set of label pairs of label  $l$  is denoted by  $L_l$ . Figure 6 shows all possible neighboring labels of label 103 and  $L_{103}$ . Note that for example 102 and 104 cannot appear in the neighborhood of a segment if the segment is labeled 103.

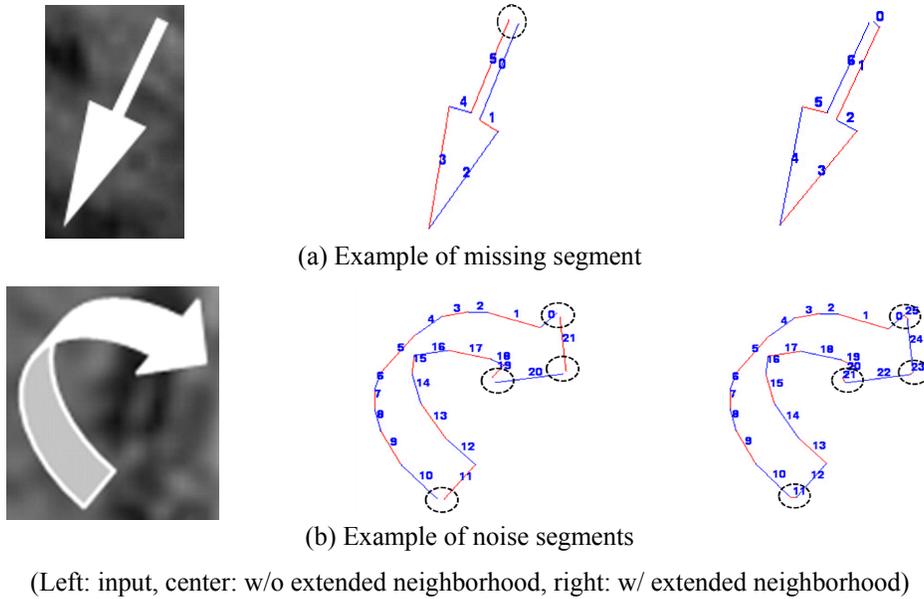


Figure 5. Problem cases due to small segments

Possible neighboring labels of 103  
 $L_{103}$  100, 102, 104, 105  
 (102, 105), (104, 105), (102, 100), (104, 100)

Figure 6. An example of  $L_l$

**Feature extraction and compatibility functions**

To handle the rotation problem we are using only angle features between line segments. Figure 7 shows the angle measures. The segment  $i$  and two neighboring segments in  $N_i$  form a part unit and two angle measurements (left and right angles) are obtained from each unit. The angles are measured in range of  $0^\circ \sim 359^\circ$ . We collected 2-tuple feature values  $(L_i, R_i)$  from part units corresponding to each label from training samples and computed probability density function (pdf) of the angle feature of each label.

The compatibility function  $r_i(f_i)$  in Eq. 1 is then computed by Eq. 2.

$$r_i(f_i) = \max_{f_{i-1}, f_{i+1}} [pdf(d_{i-1} | f_{i-1}) + pdf(d_i | f_i) + pdf(d_{i+1} | f_{i+1})] \quad (2)$$

where  $pdf(\cdot)$  is the probability density function of the angle feature given the label and  $d$  is the angle measure ( $L_i, R_i$ ).

In Eq. 2, by considering  $pdf(\cdot)$  of neighboring segments we can obtain more precise  $r_i(f_i)$ . Some labels have similar angle features and can be identified from others only by the neighborhood.

The binary compatibility function  $r_{i,j}(f_i, f_j)$  is defined by the Eq. 3.

$$r_{i,j}(f_i, f_j) = \max_{f_k} m_{h,k}(f_k) \quad (3)$$

where  $h \in \{N_k - i\}$ ,  $k \in N_i \neq j$ ,  $i \in N_j$ ,  $(f_k, f_j) \in L_{f_i}$ , and  $m$  is the message in BF.

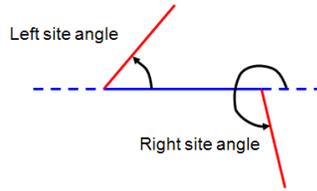


Figure 7. Angle measurement

### Post-processing

Upon the termination of the belief propagation, the label with largest belief can be selected as label of each line segment. In our application, however, the simple way does not guarantee an optimal labeling configuration. Some local labels are reasonable for the corresponding boundary part but they may not be optimal labels of the part from the view point of global optimal solution. This incorrect labeling configuration can be assumed as a local optimal solution. To find the global optimal solution, a dynamic programming (DP) technique is applied to the labels corresponding to the top- $n$  beliefs of all line segments. The best path, i.e. global optimal solution, can be obtained by a simple backtracking algorithm from the label generating the largest DP cost at the last segment.

### Hidden Markov model (HMM) classification

A hidden Markov model (HMM)-based labeling configuration classifier is implemented to classify a pointer boundary into three classes viz., (straight) arrow (SA), curved arrow (CA), and arrowheads (AH). Two 7-state left-right models and one 3-state left-right model are created to solve the 3-class classification problem. Figure 8 shows our HMM models.

To use HMM in our problem we search for the head segments such as labels 101 or 103 (both cannot appear together) in a configuration and use them as starting segments. Then every configuration can be reordered so that it starts from the same point, traces the configuration in the same direction, and ends at the same point. The reordered configuration is provided as a data sequence to the HMM classifier and a classification score of each class is obtained. The class id generating largest HMM score is selected as final pointer recognition result. Our HMM classifier provides classification scores less than 0 and larger score is better one.

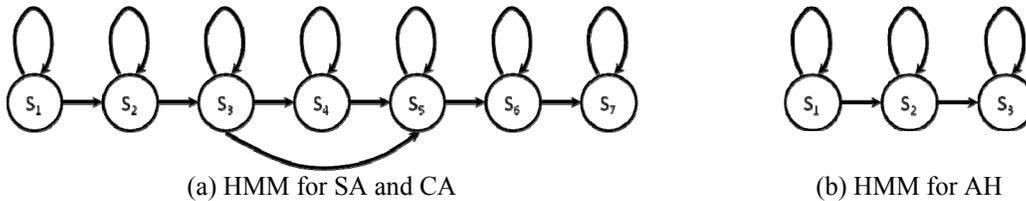


Figure 8. Structures of HMMs

### 3.3 Image Annotation, Feature Extraction, and Retrieval

A unique feature of the pointer recognition method described above is its capability of identifying the region in the image pointed to by the overlay. This feature is not necessary, however, for other symbols such as the asterisk or single character overlays ('A', 'B', etc.) The ability to isolate the image region of interest is valuable in associating the biomedical concept extracted from text analysis of the caption and figure mention text as described in Section 2. UMLS concept ids are associated with a 200×200 image region of interest (ROI). The size of the ROI is arbitrarily selected so as to be meaningful as proof-of-concept for the pilot. It is desirable to detect the ROI as a single homogeneous region and efforts toward this goal are in progress, but beyond the scope of this article.

Image features are computed on the whole image as well as the ROI for CBIR. The features used include MPEG7 color and texture descriptors, LIRE (Lucene Image Retrieval Engine) descriptors, Gabor filters, dominant colors, discrete wavelet transform, among other features. Image similarity metrics include standard measures such as L2 distance and also include visual concepts defined on image patches [17].

As shown in Figure 1, the retrieval framework will process queries in two modalities: text and image. A text query can be matched against the figure caption, article title, abstract, and if available, the figure mention in the full text. This approach is similar to BioText Search Engine [18], and Yale ImageFinder [19]. The proposed framework will also use biomedical concepts for query expansion, a feature available as a part of NLM's Essie search engine [20]. The results of the text search will be limited to images linked to articles. If the query was hybrid, i.e. included an image component, or was image only, the image features from the query image (or user marked ROI) will be compared with those resulting from the text search. In the case of an image only search, the features will be compared directly with the indexed image features.

The pointer and overlay finding results is expected to improve the specificity in image retrieval. If a pointer has been detected in an image through the use of text clues indicating their presence or directly by the MRF pointer recognizer, the image region pointed to can provide greater specificity on the image content. We expect this to significantly improve retrieval quality.

## 4. EXPERIMENTS

### 4.1 Pointer recognition

#### Data set

In our pointer recognition test, we assumed that results from the pre-processing step are satisfactory and have few problems. To create a test image set satisfying the assumption we cropped some pointers from images sampled from the ImageCLEF08 data set used in our prior work. Figure 9 shows some sample pointer images. Each pointer image has at least one pointer and the overall image quality is good. By using the pointer images, some critical failure factors from the pre-processing step such as broken or double boundaries can be minimized resulting in a more meaningful evaluation of the recognition algorithm.



Figure 9. Sample pointer images

Table 1. Number of sample images used in training and testing

	Straight arrow (SA)	Curved arrow (CA)	Arrowhead (AH)	Total
Training	114	74	57	245
Test	569	253	38	860

Table 1 shows the number of pointer images used for training and testing for each pointer class. Examples of a straight arrow (SA) and a curved arrow (CA) are shown as the first two images, respectively, in Figure 9. Training images were used to define the label set, computing the probability density functions, and for HMM training.

### Evaluation results

We applied our proposed algorithm to the pointer images and examined the results mainly in three points; MRF labeling, HMM classification, and line segment approximation result. Total 2,313 pointer boundaries were extracted from 860 test images. 82.0% out of the total boundaries were correctly labeled and classified. Figure 10 shows some pointer images successfully recognized and ROI extracted. On the other hand, 18.0% had some errors in MRF labeling or HMM classification and were not recognized correctly. We categorized main failure cases and the causes into four types.

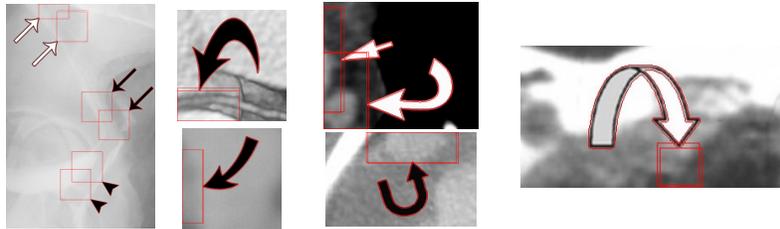


Figure 10. Pointer recognition results and ROIs (rectangles)

#### Type 1: MRF labeling error

Type 1 errors are pure MRF labeling errors. In many Type 1 cases, BF results were good enough and correct labels were included in the top-n beliefs, but the DP algorithm could not find them. Figure 11 (a) shows an example of Type 1 error. Incorrect labeling occurred in the circle.

#### Type 2: Unknown part unit

Type 2 errors occurred in pointers whose part units were not defined in the label set. Figure 11 (b) shows an example of Type 2 cases. The labels 202 and 106 in the circle may be the best choices for the line segments but their neighborhood is not similar with the true neighborhood of label 202 and 106. This caused the incorrect labeling in the local sites, and as a result, incorrect labeling configuration was obtained.

#### Type 3: Line segment approximation error

Boundary over- or under-segmentation occurred in the line segment approximation and it caused MRF labeling errors. Figure 11 (c) shows two over-segmented line segments which were supposed to be extracted as two single line segments.

#### Type 4: HMM classification error

MRF labeling error is one obvious cause of HMM classification error. But classification errors can occur with some pointers whose line segments and labeling configuration are perfect. Figure 11 (d) shows Type 4 error samples. The MRF labeling and HMM classifier are mainly responsible for the left and right error cases, respectively. The segments with NULL (0) labels in left pointer should be labeled correctly first. The right sample was trained as SA but classified as CA. The result, however, is considered reasonable.

## 4.2 Biomedical image retrieval test

### Test setup

Two topics from ImageCLEFmed were selected for this test: i) *Topic 11: show me abdominal CT images showing liver blood vessels*, and ii) *Topic 21: Show me photographs of tumors*. Text-based retrieval retrieved 1,000 and 191 images for the topic 11 and 21 text queries, respectively. Pointers and ROIs are extracted automatically from the retrieved images. Only one ROI with highest similarity score is used to decide the ranking if an image has multiple ROIs. Three sample ROIs for each topic are extracted from three images relevant to each topic, and they are compared with the ROIs extracted from the retrieved images (see Figure 1). Re-ranked results are compared with the text-based retrieval results to evaluate the performance of our proposed approach.

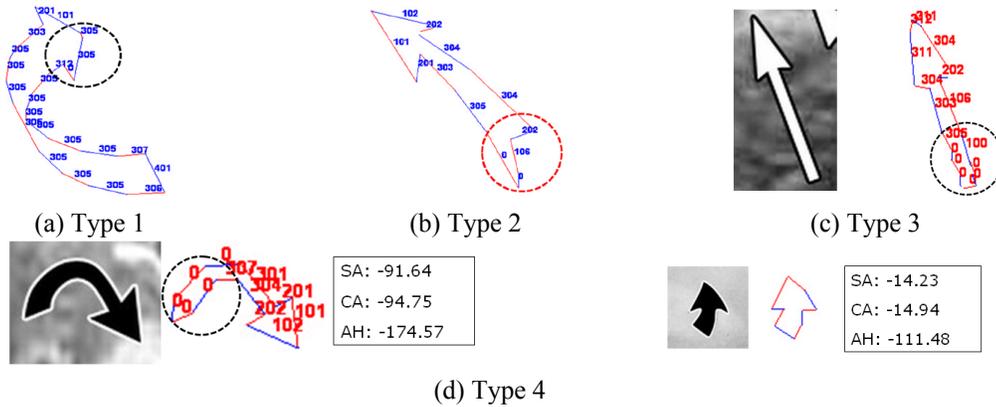


Figure 11. Sample images of each type of failure

### Evaluation results

Table 2 shows the evaluation results obtained by using TRECEVAL package [21]. Among the three query ROIs used in the retrieval test, only one showing the best performance is selected and shown in Table 2. Pointers and ROIs were extracted from 554 and 71 images out of the 1,000 and 191 images in topic 11 and 21, respectively. Other images, i.e. images without pointers, were ranked lower than the images that contain pointers.

Table 2. Retrieval results

Topic	Measurements	Text-based retrieval	Proposed retrieval
11	Number retrieved	1,000	1,000
	Number relevant	331	331
	Number relevant and retrieved	95	95
	<b>MAP</b>	<b>0.0898</b>	<b>0.0379</b>
21	Number retrieved	191	191
	Number relevant	334	334
	Number relevant and retrieved	25	25
	<b>MAP</b>	<b>0.0094</b>	<b>0.0168</b>

The preliminary retrieval test shows some promising results. The topic 21 retrieval result proves that text-based retrieval result could be improved by our proposed approach. In the text-based retrieval, only two relevant images were ranked within top-20 ranking. After re-ranking, six new images which were not in the top-20 of the text-based retrieval result were ranked within top-20, and the two images initially within top-20 were ranked at bottom where images without pointers were ranked.

Results of topic 11, however, show a lower *mean average precision* (MAP) score with the proposed approach. Initially 11 relevant images were ranked within top-20 in the text-based retrieval. After re-ranking only three new images were ranked within top-20, and the 11 images were ranked at lower than 20. We identified three main causes for relevant images being assigned a lower rank after re-ranking using our method: (i) images may not have any pointers; (ii) failures in pointer recognition; and (iii) inaccurate ROI extraction. Visual analysis revealed that of the poorly ranked relevant images: three images did not contain any pointers, one had a pointer that was not recognized, and seven had inaccurate ROI extraction. Figure 12 shows sample images where the ROI rectangles are inaccurate. Boundary of the desired ROI is marked by a dashed contour. The rectangle in Figure 12(a) is too large. It contains the entire true ROI but has too much unnecessary region included with it. In Figure 12(b) none of the ROI rectangles contain the entire true ROI but contain some part of them. From the results we conclude that improved ROI extraction techniques are needed for taking advantage of pointer localization and region-based multi-modal CBIR.

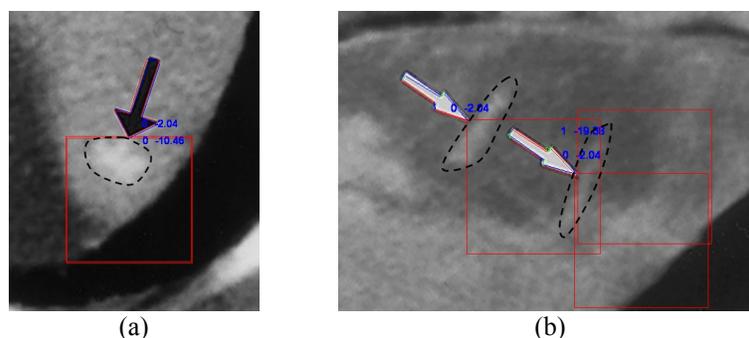


Figure 12. Inaccurate ROI extraction

## 5. CONCLUSION

Detecting arrows, pointers, and other annotations such as text labels, can be very beneficial in locating image regions of interest within figures in biomedical articles. Such annotations can be identified through clues available via relevant text snippet analysis (captions, mentions) and image analysis methods. Text concepts as well as image features can now be used to tag image content. We expect this to improve image indexing quality and consequently the indexing and retrieval of biomedical articles.

This article presents an MRF-based pointer recognition algorithm. 43 labels are defined from boundary parts frequently seen in commonly used pointer shapes overlaid on biomedical images. MRF theory is applied to label the line segments extracted from pointer boundaries. HMM-based classifier following the MRF labeling is applied to classify a labeling configuration into three pointer classes. Our test result shows that the proposed method can recognize almost all arrow type pointers and is less affected by the large variation in pointer shape. Ongoing research effort on the pointer recognition aims to: (i) improve MRF labeling algorithm; (ii) train the recognizer with unknown part units; and (iii) improve boundary extraction algorithm to extract robust pointer boundary.

Biomedical image retrieval test and the results have been presented as well. We compared the proposed retrieval approach with conventional text-based retrieval results. The preliminary results partially prove that our hypothesis is correct and show effectiveness of the proposed retrieval approach. Future work includes: (i) improving ROI extraction; (ii) combining text analysis results with pointer recognition and ROI extraction algorithm; (iii) implementing some processes manually done in the retrieval test such as noise ROI removal; and (iv) more tests against mixed collections and other query topics.

## Acknowledgement

This research was supported by the Intramural Research Program of the Lister Hill National Center for Biomedical Communications, an R&D division of the National Library of Medicine, at the National Institutes of Health, U.S. Department of Health and Human Services.

## REFERENCES

1. Demner-Fushman D, Antani SK, Thoma GR, "Automatically Finding Images for Clinical Decision Support," Proceedings of Workshop on Data Mining in Medicine, 7<sup>th</sup> IEEE Intl Conf on Data Mining 2007:139-44.
2. Demner-Fushman D, Antani SK, Simpson M, Thoma GR, "Combining Medical Domain Ontological Knowledge and Low-level Image Features for Multimedia Indexing," Proc. 2nd International "Language Resources for Content-Based Image Retrieval" Workshop (OntoImage 2008), part of 6<sup>th</sup> Language Resources and Evaluation Conference (LREC 2008). 2008;CDROM Proceedings.

3. Demner-Fushman D, Antani SK, Simpson M, Thoma GR, "Annotation and retrieval of clinically relevant images," *International Journal of Medical Informatics* (Article In Press)
4. You D, Apostolova E, Antani SK, Demner-Fushman D, Thoma GR, "Figure content analysis for improved biomedical article retrieval," *Proc. SPIE-IS&T Electronic Imaging*. San Jose, CA. January 2009;7247:72470V(1-10).
5. Antani SK, Demner-Fushman D, Li J, Srinivasan BV, Thoma GR, "Exploring use of images in clinical articles for decision support in Evidence-Based Medicine," *Proc. SPIE-IS&T Electronic Imaging*. San Jose, CA. January 2008;6815:68150Q(1-10).
6. Deserno TM, Antani S, Long R, "Ontology of Gaps in Content-Based Image Retrieval," *Journal of Digital Imaging*. 22(2):202-15, April (2009).
7. Amit Y, [2D Object Detection and Recognition: Models, Algorithms, and Networks], The MIT Press, 2002.
8. Elfadel IM, Picard RW, "Gibbs random fields, co-occurrences, and texture modeling," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 16, pp. 24-37 (1994).
9. Dubes RC, Jain AK, Nadabar SG, Chen CC, "MRF model-based algorithms for image segmentation," *Proc. 10<sup>th</sup> International Conference on Pattern Recognition*, pp. 808-814 vol.1 (1990).
10. Zhang J, "The mean field theory in EM procedures for Markov random fields," *IEEE Transactions on Signal Processing*, vol. 40, pp. 2570-2583 (1992).
11. Cao H, Govindaraju V, "Handwritten Carbon Form Preprocessing Based on Markov Random Field," *IEEE Conference on Computer Vision and Pattern Recognition (CVPR '07)*, pp. 1-7 (2007).
12. Li SZ, [Markov Random Field Modeling in Image Analysis], Springer-Verlag New York, Inc., 2001.
13. Rabiner LR, "A tutorial on Hidden Markov Models and selected applications in speech recognition," *Proceedings of the IEEE*, 77 (2): 257-286 (1989).
14. Liu Y, Ikenaga T, Goto S, "An MRF model-based approach to the detection of rectangular shape objects in color images," *Signal Processing* 87(11): 2649-2658 (2007).
15. Li SZ, "Parameter Estimation for Optimal Object Recognition: Theory and Application," *International Journal of Computer Vision*, 21(3): 207-222 (1997).
16. Intel, Open source computer vision library, 2000.
17. Rahman MM, Antani SK, Thoma GR, "A Medical Image Retrieval Framework in Correlation Enhanced Visual Concept Feature Space," 22<sup>nd</sup> IEEE International Symposium on Computer-Based Medical Symposium (CBMS). August 2009. Albuquerque, NM.
18. Hearst MA, Divoli A, Guturu H, Ksikes A, Nakov P, Wooldridge MA, Ye J, "BioText Search Engine: beyond abstract search," *Bioinformatics*, 23(16):2196-2197 (2007).
19. Xu S, McCusker J, Krauthammer M, "Yale Image Finder (YIF): a new search engine for retrieving biomedical images," *Bioinformatics*, 24(17):1968-1970 (2008).
20. Ide NC, Loane RF, Demner-Fushman D, "Essie: a concept-based search engine for structured biomedical text," *J Am Med Inform Assoc*. 14(3):253-63 (2007).
21. <http://trec.nist.gov/>