Exploring use of Images in Clinical Articles for Decision Support in Evidence-Based Medicine

Sameer Antani, Dina Demner-Fushman, Jiang Li[a], Balaji V. Srinivasan[b], George R. Thoma

National Library of Medicine, National Institutes of Health, Bethesda, MD 20894
[a] Department of Computer Science and Engineering, University at Buffalo, The State University of New York, Buffalo, NY 14260
[b] Department of Electrical and Computer Engineering, University of Maryland, College Park, MD 20742

Abstract

Essential information is often conveyed pictorially (images, illustrations, graphs, charts, etc.) in biomedical publications. A clinician's decision to access the full text when searching for evidence in support of clinical decision is frequently based solely on a short bibliographic reference. We seek to automatically augment these references with images from the article that may assist in finding evidence.

In a previous study, the feasibility of automatically classifying images by usefulness (utility) in finding evidence was explored using supervised machine learning and achieved 84.3% accuracy using image captions for modality and 76.6% accuracy combining captions and image data for utility on 743 images from articles over 2 years from a clinical journal. Our results indicated that automatic augmentation of bibliographic references with relevant images was feasible. Other research in this area has determined improved user experience by showing images in addition to the short bibliographic reference. Multi-panel images used in our study had to be manually pre-processed for image analysis, however. Additionally, all image-text on figures was ignored.

In this article, we report on developed methods for automatic multi-panel image segmentation using not only image features, but also clues from text analysis applied to figure captions. In initial experiments on 516 figure images we obtained 95.54% accuracy in correctly identifying and segmenting the sub-images. The errors were flagged as disagreements with automatic parsing of figure caption text allowing for supervised segmentation. For localizing text and symbols, on a randomly selected test set of 100 single panel images our methods reported, on the average, precision and recall of 78.42% and 89.38%, respectively, with an accuracy of 72.02%.

1. Introduction and Background

Clinicians and medical researchers routinely use online databases such as MEDLINE® (http://www.pubmed.gov) to search for bibliographic citations that are relevant to a clinical situation. They can fairly accurately form an opinion about the relevance of a publication to the clinical situation based on its title alone; however the title is not always sufficient in determining the evidence-based practice usefulness (henceforth evidence-based utility or clinical utility) of a publication [1]. There are indications that user experience can be improved by augmenting short bibliographic references with images [2].

The successes in automatic image annotation [3], content-based image retrieval (CBIR) [4], and text classification based on image captions have motivated integration of image data for biomedical text categorization [5]. Other efforts have explored biomedical article retrieval based on image content [6,7], and use of textual and image features for image classification in biomedical articles [8]. While preliminary studies on image-only retrieval have resulted in mediocre results, classification of bioscience images into six generic categories achieved 73.66% average F-score [9].

Given that images, medical illustrations, graphs or charts often convey essential information in compact form in clinical articles, we seek to automatically identify figures that could help clinicians evaluate potential usefulness of a publication in a clinical situation at hand. We hypothesize that in many cases a short outcome statement that we currently automatically extract to augment the title of a MEDLINE citation [9] could be rendered more useful if accompanied by one or more extracted images. For example, given a title "Clinical management and microscopic

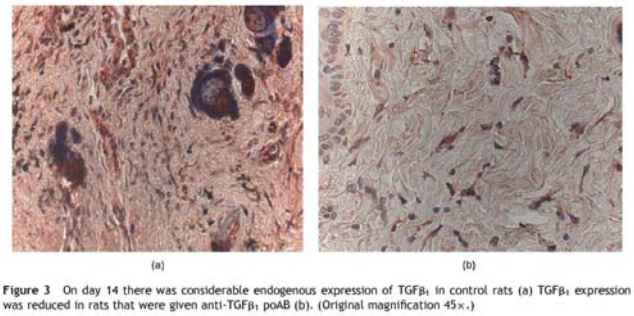
characterization of fatigue-induced failure of a dental implant" a clinician might not know if the article applies to the case for which evidence is sought. The automatically extracted outcome statement in Figure 1 will clarify that the "fatigue-induced failure" is a fracture, and the adjacent x-ray will illustrate what is meant by the term "typical signs" of a fracture.

The fractured implant showed the typical signs of a fatigue-induced fracture in the coronal portion of the implant together with numerous micro-fractures in the apical one …

Figure 1. A fractured implant and relevant fragment of the automatically extracted outcome statement (reproduced with author's permission from [11])

Encouraged by success achieved in various informatics applications through combining textual and image data, we explored a new area of biomedical image annotation using textual and image data – that of classifying images in biomedical articles with respect to their utility for clinical decision support and if such images could be reliably extracted from the original articles [10]. We selected 2004 -- 2005 issues of the *British Journal of Oral and Maxillofacial Surgery*, manually annotating 743 images by utility and modality (radiological, photo, etc.) Image data, figure captions, and paragraphs surrounding figure mentions in text were used in classification. Automatic image classification achieved 84.3% accuracy using image captions for modality and 76.6% accuracy combining captions and image data for utility [10].

We observed during our study that articles included multi-panel images in figures as a single composite image. These multi-panel images are often identified using subfigure labels, e.g., "Figure 1(a), … Figure 1(b), …", in the figure captions or mentions, as shown in Figure 2. Occasionally, a multi-panel image would have no caption or can be considered as a composite image (of CT slices, for example) requiring segmentation for successful annotation. These multi-panel images were hand segmented in our previous study. In general, this step would need to be automated for purposes of image analysis and annotation.



(a) (b)

Figure 3 On day 14 there was considerable endogenous expression of TGFβ1 in control rats (a) TGFβ1 expression was reduced in rats that were given anti-TGFβ1 poAB (b). (Original magnification 45×.)

Figure 2. Example of a multi-panel image with 2 subfigures (reproduced in reduced form [12])

Further, we observed that there were several images that had text overlays to indicate particular regions in an image. Examples of these are shown in Figure 3. Localizing and recognizing symbols (such as arrows) and text using OCR methods in such images would be valuable in retrieval of articles. In this article we present initial results from our methods designed to (a) identify and segment composite multi-panel images, and (b) localize text and symbols on these image panels. The novelties in our segmentation method are (i) capability to process composite images in a variety of layouts with or without boundaries separating images or illustrations, (ii) use of suggestions (probabilities) from our automated figure caption text localizing and parsing methods [10], and (iii) reporting disagreements in number of panels detected with those estimated by caption parsing allowing for supervised post-processing of such cases. We have achieved 95.54% accuracy in detection and segmentation of (single and) multi-panel images.

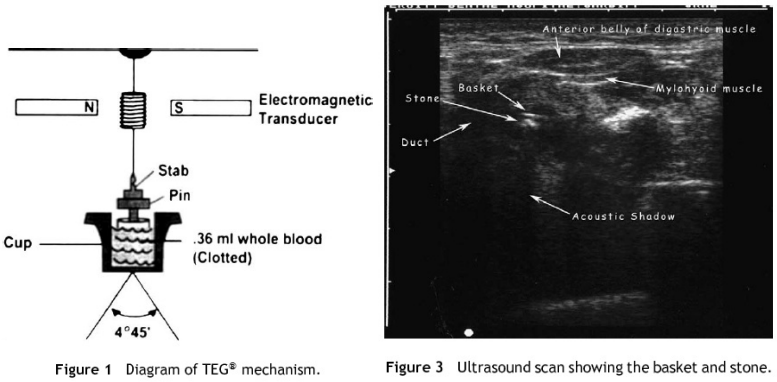


Figure 1 Diagram of TEG® mechanism. Figure 3 Ultrasound scan showing the basket and stone.

Figure 3. Examples of figures with text on them (reproduced in reduced form [13,14])

Localization of text in images has long been studied in the literature [15-25]. Most methods have been developed for text in video and are not readily applicable to our data. On the observation that different image modalities exhibit relatively different text/symbol characteristics, our method includes an automated preprocessing step for detecting image modality that sets parameters for various methods. We merged modalities listed in Table 1 into three broad categories: color image, illustration image, or X-ray image, shown in the "Tested As" column.

Table 1. Image modality categories typically found in biomedical journals. "Tested As" column identifies the image class for our text and symbol localization.

Modality	Definition	Tested As
Chart / Graph	A geometric diagram consisting of dots, lines, and bars.	Illustration
Drawing	A hand drawn illustration.	Illustration
Flowchart	A symbolic representation of sequence of activities.	Illustration
Histology	An image of cells and tissue on the microscopic level.	Color image
Photograph	Picture obtained from a camera (visible light spectrum)	Color image
Radiology	A 2D view of an internal organ or structure (includes X-ray, CT, PET, MRI, ultrasound)	X-ray
Table	Data arranged in a grid	Ignored
Mixed	Images combining modalities (e.g., drawings over an x-ray)	Mixed

In this article we present initial text and symbol localization results from our methods. On a randomly selected test set of 100 single panel images our methods reported, on the average, precision and recall values of 78.42% and 89.38%, respectively, with an accuracy average of 72.02%. We are in the process of conducting OCR evaluation experiments and aim to include them in this article upon their conclusion.

The article is organized as follows. In Section 2 we describe the multi-panel detection and segmentation method and the text and symbol localization method. In Section 3, we discuss initial results. With Section 4 we present our conclusions, ongoing research, and longer-term future work.

2. Methods

2.1 Detecting and decomposing multi-panel images

A two-phase algorithm is developed and applied for detection and decomposition of multi-panel images. Before algorithm initiation, however, an estimate of number of panels is obtained using Natural Language Processing (NLP) techniques from the figure caption [10]. It is possible that the figure has no caption depriving the method of any cues. Conversely, it is also possible that the estimate is incorrect due to text analysis errors.

The first phase of the algorithm is a coarse segmentation method that looks for strong white or black lines between image panels. First the image is binarized using Otsu's method [26]. Horizontal and vertical profiles are computed

and searched for evidence of continuous (image width or height, as applicable) white or black lines less than or equal to 5% of total image width or height. If any panels are found then the image is segmented along identified boundaries and recursively applied to segmented panels until no further segmentations are found. However, this method does not always give a good split of the panels because these identified pixels need not be a border separating two panels and can be just a part of the single panel image. If the number of detected panels matches with the estimate from the caption, the image coordinates corresponding to this split is returned and the method terminates.

If there is mismatch between the estimate of the number of panels and the computed number from the first phase, the number of panels is estimated by the second phase. This method is based on an assumption that all multi-panel images have a sharp transition at the border separating the panels. Again an Otsu [26] binarized image is used and horizontal and vertical profiles are computed. These profiles are then analyzed for evidence of sharp boundaries. If any are found then the image is segmented and the algorithm is recursively applied. If the final number of panels detected using this technique matches with the estimate from the caption, then image coordinates corresponding to the splits are returned and the algorithm terminates. If there continues to be a mismatch between the estimate of the number of panels from text analysis and the number of detected panels then panel segmentation with the least difference with estimate is chosen and an entry is logged for later review and possible supervised segmentation.

2.2 Localizing text and symbols on images

After studying the literature we developed methods for localizing text from figure images in biomedical journals. In localizing text/symbols we encounter the following challenges: size, fonts, stroke color, image noise, interference with background objects, arbitrary locations, and text block sizes varying from single character to few words spanning multiple lines. In general no single method was found to be directly applicable to the problem. Each image modality had peculiarities that made its identification important. A color image is one with different distributions in the 3 color components of native color space. An illustration image, either color or gray scale image, has a relatively uniform background and simple histogram distribution. An X-ray image is a gray scale image with continuous histogram distribution and a black background. Examples of these image types are shown in Figure 4. All processing is done in gray scale following a 2D adaptive noise-removal using Wiener filters [27] with a neighborhood of size 3x3 pixels.

For a color image, we observe that even though the image may have a complex background, the texts/symbols on it are machine overlays and usually have a uniform color stroke that is very often either white or black. For detecting text in such images, shown in Figure 4(a), we follow the following steps:

1. Calculate the mean of each color plane (Red, Green, Blue, for example) of the image.
2. Select the plane with highest mean as gray scale image for detecting black text and symbols (opposite for white characters).
3. Examine the histogram of given gray scale component image, find the first valley that is most against to the background and with certain accumulated size, as threshold to binarize the gray scale image.
4. Use morphological method to eliminate binarization noise, using a diamond elemental structure of size 5.
5. Find the bounding box for the text in the binarized image, together with applied constraint for aspect ratio, size, etc.
6. Merge boxes that are horizontally near each other.

Text information in an illustration image is usually black text on uniform background. The lines in most drawings are thin and may be removed due to morphological operations. We take the following steps for detecting and localizing text in illustration images:

1. Binarize image using Otsu's method [26].
2. Use a vertical line elemental structure to apply morphological methods on image.
3. Use a horizontal elemental structure to connect components that are resulting from earlier operation.
4. Find and merge resulting text boxes in the binarized image

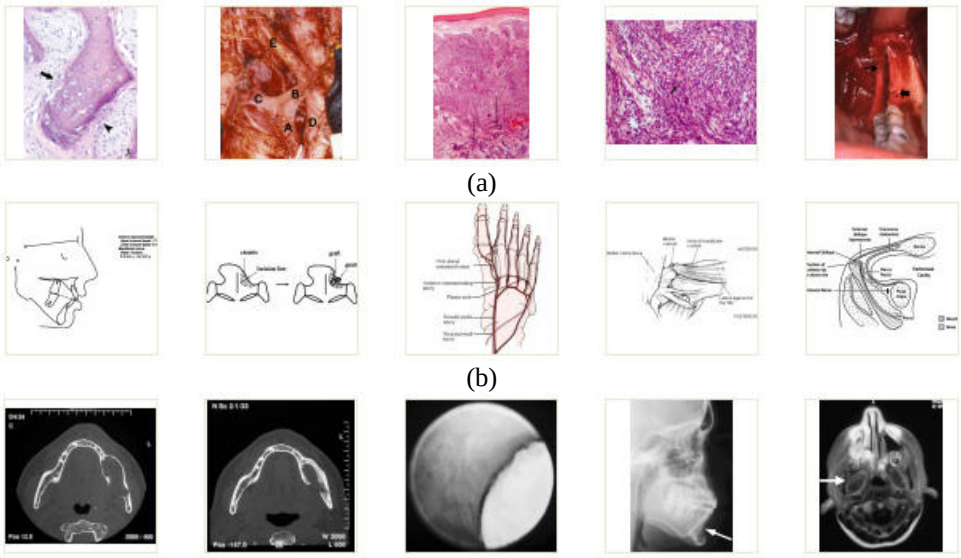


Figure 4 (a) Examples of color images, (b) examples of illustrations, and (c), examples of X-ray images found in biomedical journal articles with text or symbols on them.

Text and symbols in X-ray images are derived using iterative refinement. The following steps are taken:

1. The image is recursively decomposed using the Quadtree technique as a coarse top-down approach with low threshold to find largely homogeneous area that contains a foreground object and text or is classified as background. The method is based on the following assumptions: (a) the text areas usually have high changes in intensity caused by the character/symbol stroke against its surrounding areas, and (b) the background area in X-ray images typically has relatively uniform gradient.
2. Candidate text regions in the image derived from the first step are subdivided into 8x8 blocks and Discrete Cosine Transform (DCT) features are computed on them for identifying text areas on the image. Most of the signal information tends to be concentrated in a few low-frequency components of the DCT, however, high frequency perturbations caused by character strokes can be detected using this method.
3. Finally, connected component analysis is used to refine the results. 8x8 blocks that are deemed as candidate blocks following DCT analysis are locally binarized using Otsu's method [26]. For each binarized block we compute statistics along 4 directions to determine character stroke strength. These values are then compared against a threshold to assert text areas. The connected components for the text are then used to fit bounding boxes.

Table 2. Results of detection of multi-panel images. Note that "Panel# Mismatch" implies that a multi-panel image was detected but with a disagreement with caption text analysis.

Ground Truth Category	**Detected Category**		
	Single Panel	**Multi-Panel**	**Panel# Mismatch**
Single Panel	409 (**95.78%**)	18 (4.22%)	---
Multi-Panel	5 (5.62%)	84 (**94.38%**)	6 of 84 (7.14%)

3. Results and Discussion

3.1 Detecting and decomposing multi-panel images

Detection and decomposition of multi-panel images was tested on 516 figure images extracted from 2 years (2004 – 2005) issues of the *British Journal of Oral and Maxillofacial Surgery*. In this set, 427 images were single panel images and 89 were multi-panel. Not all images had figure captions. The confusion matrix resulting from our method is presented in Table 2.

Overall 409 or 95.78% of the single panels and 84 or 94.38% of the multi-panel images were correctly identified. In case of multi-panel images, 6 of 84 had been correctly identified as multi-panel images having a disagreement with the caption analysis. This disagreement was usually minor (±1 panel). These images are deemed as correct detection of a multi-panel image for purposes of this evaluation. Overall result combining these scores is 95.54% detection and decomposition accuracy.

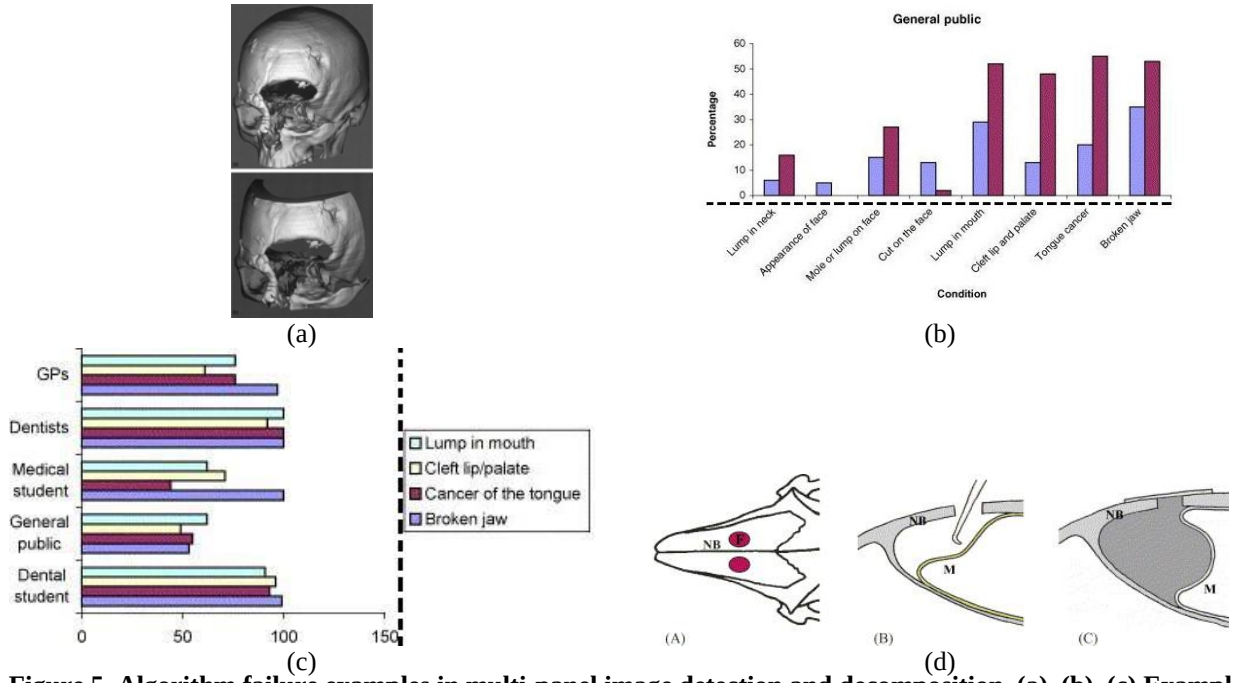


Figure 5. Algorithm failure examples in multi-panel image detection and decomposition. (a), (b), (c) Examples of single panel images identified as multi-panel. (d) Example of multi-panel images identified as single panel.

The method typically failed on cases where (i) inter-panel boundary width assumption exceeded our thresholds or (ii) there was a lack of a sharp transition between panels. Image examples for which the method made errors are shown in Figure 5. Figures 5(a), 5(b), 5(c) show single panel images that were identified to be multi-panel images. Separation boundaries for Figures 5(b) and 5(c) are shown as dashed lines on the images. Figure 5(d) is a multi-panel image that is misclassified due to lack of sharp transition between the image panels.

3.2 Localizing recognizing text and symbols on images

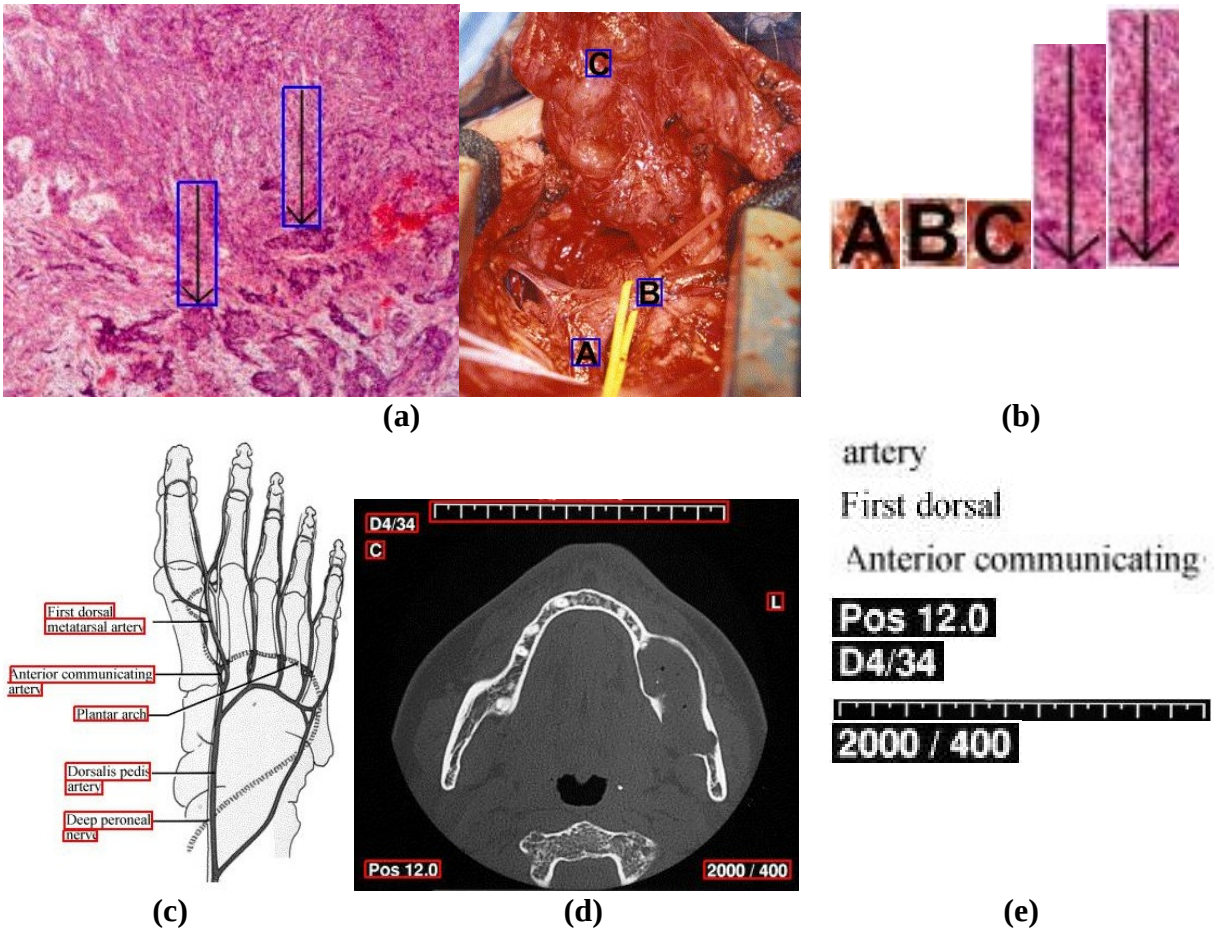
Text localization results are initially discussed for individual image modality followed by combined results. Sample image results for each image modality are shown in Figure 6.

Data Set: 100 single panel images were randomly selected for evaluation from the results of the multi-panel image segmentation method discussed earlier. In this set we found 47 color images, 12 illustrations, 23 X-ray images, and 18 images of mixed variety. We ignored the mixed images because they exhibit features from multiple image classes

and our methods, at present, are trained on individual class characteristics. In developing the ground truth on this set we found that 32 color images and 11 X-ray images had no text or symbols on them.

Evaluation Strategy: We compare detected text and symbol blocks with the ground truth. If a text or a symbol block is partially detected and would not be visually recognizable, when cropped, we mark it as false negative (missed-detection).

Color Images: Results from text and symbol localization on color images are shown in Table 4. The ground truth had 34 instances of text or symbols on 15 images. Our methods detected 37 text/symbol blocks with 7 false alarms and 4 missed-detections. Resulting precision and recall values are 81.08% and 88.24%, respectively. 32 images did not have text and 17 were accurately detected to have none resulting in 53.13% accuracy. While a fairly high percentage of text or symbol blocks are localized, the method suffers from a high false alarm rate as reflected in our average accuracy measure. One possible explanation is that uniform color on character stroke is an insufficient criterion. Adding stroke shape information in addition to the method should reduce false alarms. For color images with no text, the algorithm detected text shaped characters such as inner cavity in a nut. We expect that cues from text analysis may reduce such errors. We hope to present improved results in the final version of this article. OCR may also be used to further reduce the errors.



(a) **(b)** **(c)** **(d)** **(e)**

Figure 6 (a) Color image examples with mixed background with localized text and symbols. (b) Extracted blocks from color images. (c) Illustration with localized text regions. Arrow detection has been manually turned off for clarity. (d) X-ray image with localized text and symbols. (e) Samples of extracted blocks from illustration image and X-ray image. Note: All example images have been cropped or resized to fit in figure table. Some extracted text blocks have also been resized to fit in figure table.

Table 3. Results of text/symbol localization: Color Images

		Detected Blocks		
		Text	**No Text, No Symbols**	
Ground Truth	**Text / Symbols**	30	4	**Precision** = 81.08%
	No Text, No Symbols	7	--	**Recall** = 88.24%

Illustration Images: Results from text and symbol localization on illustration images are shown in Table 4. We found 12 illustration images in the test set. All these images had text and or symbols on them. For this image type only text blocks were considered, however, since symbols (arrows, in our case) usually have the same line thickness as other lines in the illustration. Of total of 79 ground-truth text blocks, 73 were correctly identified with 20 false alarm blocks and 6 missed-detections. The resulting precision and recall values are 78.49% and 92.40%, respectively. Most errors were due to very small or very thin characters.

Table 4. Results of text/symbol localization: Illustration

		Detected Blocks		
		Text	**No Text, No Symbols**	
Ground Truth	**Text / Symbols**	73	6	**Precision** = 78.49%
	No Text, No Symbols	20	--	**Recall** = 92.40%

X-ray images: Results from text and symbol localization on color images are shown in Table 5. Of the 23 X-ray images in the test set 11 images that had no text or symbols on them. Our methods accurately marked 10 as having no text resulting in an accuracy of 90.90%. Remaining 12 images had 64 text and symbol blocks marked in the ground truth. Our methods correctly identified 56 text blocks with 18 false alarms and 8 missed-detections. This results in precision and recall values of 75.68% and 87.50%, respectively.

Table 5. Results of text/symbol localization: X-ray images

		Detected Blocks		
		Text	**No Text, No Symbols**	
Ground Truth	**Text / Symbols**	56	8	**Precision** = 75.68%
	No Text, No Symbols	18	--	**Recall** = 87.50%

Overall results: On the average our methods reported a precision and recall of 78.42% and 89.38%, respectively, with an average accuracy of 72.02%.

4. Conclusions and Future Work

This article reports our current efforts in detection and segmentation of multi-panel images found as figures in electronic biomedical journals. Further, we develop and evaluate methods for detecting text on commonly found image modalities. Results of multi-panel image detection and decomposition and detection of text blocks on images from biomedical journals are quite promising and we hope to take advantage of this information in the integration of images in decision support for evidence-based medicine. We observe that for color and illustration type of images, most of the text/symbols are correctly localized. There is a need to reduce the false alarms, however. We expect that the inclusion of character/symbol stroke strength analysis, knowledge based (location/size/shape) analysis will improve results. For X-ray images, both false alarms and missed-detections are relatively high. Even though these images have a fairly uniform background, gray scale distributions perturbed by the imaged anatomy introduce these errors. In addition, characters touching foreground objects, strong stroke non-text foreground object, mixed background with lack of color for separation, and other interferences reduce the performance of our methods. As future work, we aim to do the following: (a) improve localization results using a better local adaptive binarization scheme for connected component analysis, and (b) Gabor feature with dynamic parameter set among other approaches.

Acknowledgement

This research was supported by the Lister Hill National Center for Biomedical Communications and intramural 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, Hauser S, Thoma G. The role of title, metadata and abstract in identifying clinically relevant journal articles. AMIA Annu Symp Proc; 2005::191-5.
2. Hearst MA, Divoli A, Ye J, Woolridge MA. Exploring the efficacy of caption search for bioscience journal search interfaces. Proc BioNLP 2007: Biological, translational, and clinical language processing. 73-80, 2007.
3. Lehmann TM, Güld MO, Thies C, Plodowski B, Keysers D, Ott B, Schubert H. IRMA – Content-based image retrieval in medical applications. MEDINFO 2004:842-8.
4. Antani S, Long LR, Thoma G. Content-based image retrieval for large biomedical image archives MEDINFO 2004:829-33
5. Shatkay H, Chen N, Blostein D. Integrating image data into biomedical text categorization. Bioinformatics. 2006 Jul 15;22(14):e446-53.
6. Deserno TM, Antani S, Long R. Exploring access to literature using content-based image retrieval. SPIE Medical Imaging 2007. vol. 6516.
7. Christiansen A, Lee D-J, Chang, Y. Finding relevant PDF medical journal articles by the content of their figures. SPIE Medical Imaging 2007. Vol. 6516
8. Rafkind B, Lee M, Chang SF, Yu H. Exploring text and image features to classify images in bioscience literature. Proceedings of the HLT-NAACL BioNLP Wksp on Linking Natural Language and Biology. 2006 Jun:73-80.
9. Demner-Fushman D, Lin J. Answering clinical questions with knowledge-based and statistical techniques. Computational Linguistics. 2007;33(1):63-104.
10. Demner-Fushman D, Antani SK, Thoma GR. Automatically Finding Images for Clinical Decision Support. Proceedings of Worshop on Data Mining in Medicine, 7th IEEE Intl Conf on Data Mining 2007:139-44.
11. Capodiferro S, Favia G, Scivetti M, De Frenza G, Grassi R. Clinical management and microscopic characterisation of fatigue-induced failure of a dental implant. Case report. Head Face Med. 2006 Jun 22;2:18.
12. Schultze-Mosgau S, Wherhan F, Rödel F, et al. Anti-TGFß$_1$ antibody for modulation of expression of endogenous transforming growth factor beta 1 to prevent fibrosis after plastic surgery in rats. British J Oral and Maxillofacial Surgery 42(1):112-19, 2004.
13. Mitchell DA, Gorton H. Thromboelastographic study of the effect of manipulation of central veins on coagulability of venous blood. British J Oral and Maxillofacial Surgery 42(1):112-19, 2004.
14. Drage NA, McAuliffe NJ. Ultrasound-guided basket retrieval of salivary stones: a new technique. British J Oral and Maxillofacial Surgery 43(3):246-48, 2005.
15. Jung K, Kim KI, Jain AK. Text information extraction in images and video: a survey. Pattern Recog. 37(5):977-97, 2004.
16. Zhong Y, Karu K, Jain AK. Locating text in complex color images. Pattern Recog. 28(10):1523-36, 1995.
17. Antani S, Crandall D, Kasturi R. Robust extraction of text in video. 15th Intl Conf Pattern Recog. 1:831-4, 2000.
18. Zhou J, Lopresti D. Extracting text from WWW images. Proc. 4th Intl Conf Doc Anal and Recog. 1:248-52,1997
19. Chuang L, Xiaoqing D, Youshou W. Automatic text location in natural scene images. 6th Intl Conf Doc Anal and Recog. 1069-73, 2001
20. Hao W. Automatic character location and segmentation in color scene images. 11th Intl Conf Image Anal and Proc. 2-7, 2001.
21. Jain AK, Bin Y. Automatic text location in images and video frames. 14th Intl Conf Pattern Recog. 2:1497-9, 1998.
22. Pyeoung-Kee K. Automatic text location in complex color images using local color quantization.TENCON, 1: 629-32, 1999.
23. Sobottka K, Bunke H, Kronenberg H. Identification of text on colored book and journal covers. Proc 5th Intl Conf Doc Anal and Recog. 57-62, 1999.

24. Young-Kyu L, Song-Ha C, Seong-Whan L. Text extraction in MPEG compressed video for content-based indexing. 15th Intl Conf Pattern Recog, 4:409-12, 2000.
25. Yu Z, Hongjiang Z, Jain AK. Automatic caption localization in compressed video. IEEE Trans Pattern Anal and Machine Intell. 22(4):385-92, 2000.
26. Otsu N. A Threshold Selection Method from Gray-Level Histograms. IEEE Trans Systems, Man, and Cybernetics. 9(1):62-6, 1979.
27. Lim JS. Two-Dimensional Signal and Image Processing. Prentice Hall, NJ, USA. Equations 9.44 – 9.46, p. 548, 1990.