Automatic segmentation of subfigure image panels for multimodal biomedical document retrieval

Beibei Cheng[a], Sameer Antani, R. Joe Stanley[a], Dina Demner-Fushman, George R. Thoma

National Library of Medicine, National Institutes of Health, Bethesda, MD 20894
[a]Dept. of Electrical and Computer Engineering, Missouri University of Science and Technology, Rolla, MO 65409

ABSTRACT

Biomedical images are often referenced for clinical decision support (CDS), educational purposes, and research. They appear in specialized databases or in biomedical publications and are not meaningfully retrievable using primarily text-based retrieval systems. The task of automatically finding the images in an article that are most useful for the purpose of determining relevance to a clinical situation is quite challenging. This task can be done by automatically annotating images extracted from scientific publications with respect to their usefulness for CDS. As an important step toward achieving this goal, we proposed figure image analysis for content-based image retrieval (CBIR) techniques. Extracted image features from the entire image and relevant local image regions can then be associated with identified biomedical concepts extracted from the meta-text in figure captions and discussion in the full text for improved hybrid (text and image) retrieval of biomedical articles.

A challenge toward this goal is separating individual panels from a multi-panel figure that is often found as a single image in the biomedical article. 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 from articles over 2 years from a clinical journal. However, the figures images in this study had to be manually segmented into individual panels. In this work we present methods that add make robust our previous efforts reported here that, though successful, were limited in their scope and were unable to meet the challenges of segmenting figure illustrations, graphs, and charts. For the latter, we present a novel particle swarm optimization (PSO) clustering algorithm to locate related figure components

Results from preliminary evaluation are very promising with the area under the ROC curve at 94.9% for regular (non-illustration) figure images and 92.1% accuracy for illustration images. More intensive tests are in progress to evaluate impact of automatic figure panel segmentation and use of ROI in image annotation and retrieval.

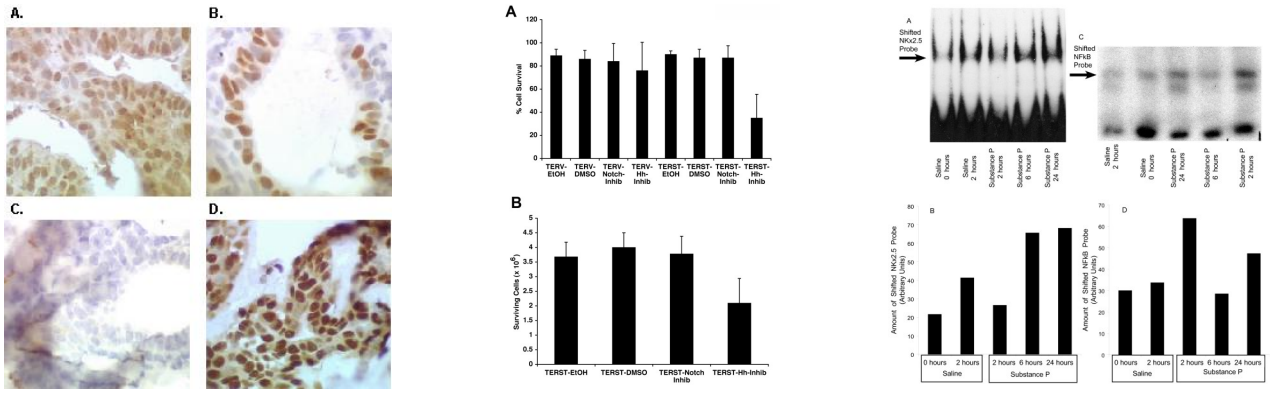
1. INTRODUCTION

Biomedical images are frequently used in publications to illustrate the medical concepts or to highlight special cases. They are invaluable in establishing diagnosis, acquiring technical skills, and implementing best practices in many areas of medicine. Conventional approaches for biomedical journal article retrieval have been text-based with little attention devoted to the use of images in the articles. Text-based retrieval uses text information automatically extracted from title, abstract, figure captions, and discussions (mention). It provides fairly good results; however, the relevance quality sometimes is not satisfactory. Content-based image retrieval (CBIR) also has been applied to biomedical image retrieval. However, the retrieval performance is far behind the text-based retrieval due to several gaps [1]. Low level features such as color, textual, and shape used in CBIR are insufficient to represent medical concepts or meaningful diagnostic information in the images effectively.

To improve the relevance quality of conventional retrieval approaches, we have proposed an approach using hybrid (text and image) features [2]. Information retrieval (IR) techniques are used to identify key textual features in the title, abstract, figure caption, and figure citation (mention) in the article. Structured vocabularies, such as the NLM's Unified Medical Language System (UMLS ®) are used as well to identify the biomedical concepts in these. Unlike conventional CBIR that just uses image features from the entire image our proposed approach uses a combination of features computed over the entire image and those computed from specific image regions of interest (ROIs). In computing image features, however, it is necessary that the image should be unimodal and individual. Often, in biomedical articles author put related images from different modalities (CT, MR, or Ultrasound, for example) as different subfigures in a single figure, or put slices from a CT or MR study as subfigures. For CBIR to be effective it is necessary to separate individual subfigure panels.

This article presents our efforts to improve our prior work [2] on subfigure localization and segmentation toward improved biomedical article retrieval. In [2], 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 from articles over 2 years from a clinical journal. However, the figures images in this study had to be manually segmented into individual panels. In this work we present methods that add make robust our previous efforts [3] that, though successful, were limited in their scope and were unable to meet the challenges of segmenting figure illustrations, graphs, and charts. For the latter, we present a novel particle swarm optimization (PSO) clustering algorithm to locate related figure component. Results from preliminary evaluation are very promising with the area under the ROC curve at 94.9% for regular (non-illustration) figure images and 92.1% accuracy for illustration images. More intensive tests are in progress to evaluate impact of automatic figure panel segmentation and use of ROI in image annotation and retrieval.

This article is organized as follows. Section 2 describes prior work on the topic. Section 3 describes the methods. Section 4 presents the experiments and we conclude with Section 5.



(a) (b) (c)

Figure 1. Examples of different types of images. (a) Regular image. (b) Illustration image. (c) Mixed image.

2. PRIOR WORK

In our prior work a heuristic two-phase algorithm was developed and applied for detection and decomposition of multi-panel images. Information from figure caption was used to obtain an estimate of number of panels [3]. The algorithm looked for strong white or black lines or a sharp transition between image panels. If any panels were found then the image was segmented along identified boundaries and recursively applied to segmented panels until no further segmentations are found. 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. 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 that evaluation. Overall result combining these scores was 95.54% detection and decomposition accuracy. 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. A further limitation of the method is that it was heuristic and could not adapt to variations in the figure layouts.

3. METHODS

For this research, we use 1237 medical image annotated by modality (radiological, photo, etc.) selected from 5 BioMedCentral[1] journals (Cancer, AFPS, Urology, Surgery and Cardiovascular Ultrasound). The set comprises of 756 regular images and 481 illustration images. The collection was then divided into three catalogs: (i) *regular images* such as color or X-ray image (as shown in Figure 1(a)), (ii) *illustration image*s (as shown in Figure 1(b)), and (iii) *mixed image*s (color mixed with illustration image or X-ray mixed with illustration image) (as shown in Figure 1(c)). We are

[1] http://www.biomedcentral.com/

presenting results from our work in segmenting figure panels in regular and illustration images, respectively. Segmenting mixed figure images remains a goal for us.

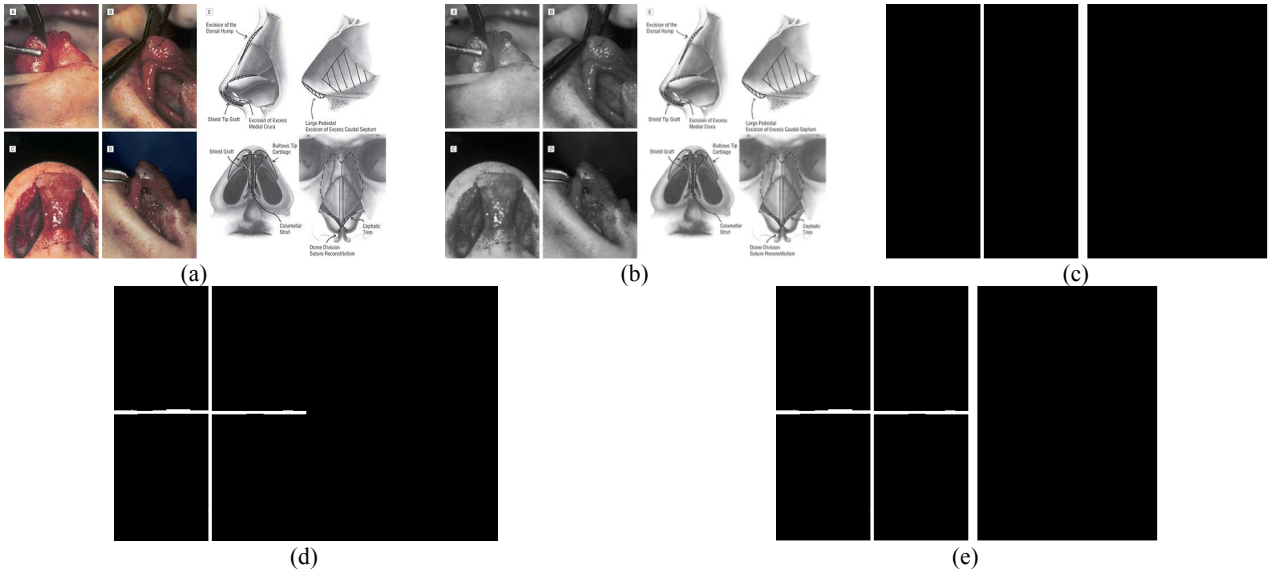


Figure 2. Image processing algorithm example. (a) Original image. (b) Gray image. (c) Output by using low variance line finding technique. (d) Output by using gray drop technique. (e) Output by Or ((c), (d)).

3.1 Segmenting multi-panel regular images

A two-phase algorithm is developed and applied for decomposition of *regular images*. The first phase is to find the inter-panel boundary line by using image-processing techniques, which are as follows:

1) Convert RGB images into gray images.
2) Calculate the variance of each vertical line and horizontal line across the image. The line with low variance is marked as the boundary. Call this *image_1*
3) Get the boundary edge with gray horizontal and vertical dynamic range between the minimum and maximum gray drop values of 10 and 25. A boundary edge is defined as a set of pixels whose horizontal neighbor difference is less than the minimum gray drop value while its vertical gray drop value is greater than the maximum gray drop value, or vice versa.
4) A logical OR or image outputs from steps 2) and 3). Call this *image_2*.

Figure 2 presents an image example of the image processing steps for the original image to generate the boundary output. From *image_1* and *image_2*, we compute the number of subfigure panels and assign them to variables: *panel_1* and *panel_2,* respectively. In addition, an estimate of number of panels is obtained using Natural Language Processing (NLP) techniques from the figure caption [3], named *panel_3*. Next, four features are generated, which are (i) *panel_3* / *panel_1*, (ii) *panel_3* / *panel_2*, (iii) standard deviation of the subpanels size from *image_1*, and (iv) the standard deviation of the subpanels size from *image_2*.

The second phase is using above features as the input to train and test the neural network. The performance is computed by comparing the values of *panel_1, panel_2,* and *panel_3* with a manually determined *real_panel_number.* If the *real_panel_number* = *panel_1*, then the output is set to 1; else if *real_panel_number* = *panel_2*, then the output is set to 2; else if *real_panel_number* = *panel_3*, then the output is set to 3; else, it is set to 4.

3.2 Segmenting multi-panel illustration images

Unlike regular images illustration images do not have a clear inter-panel boundary. This makes segmenting the panels more challenging. Often the separation is obvious to the human observer. Some cues regarding illustration placement

and size were used to develop an algorithm for segmenting the drawn groups. An overview of the algorithm investigated is shown in Figure 3. Each flowchart component is discussed in turn.

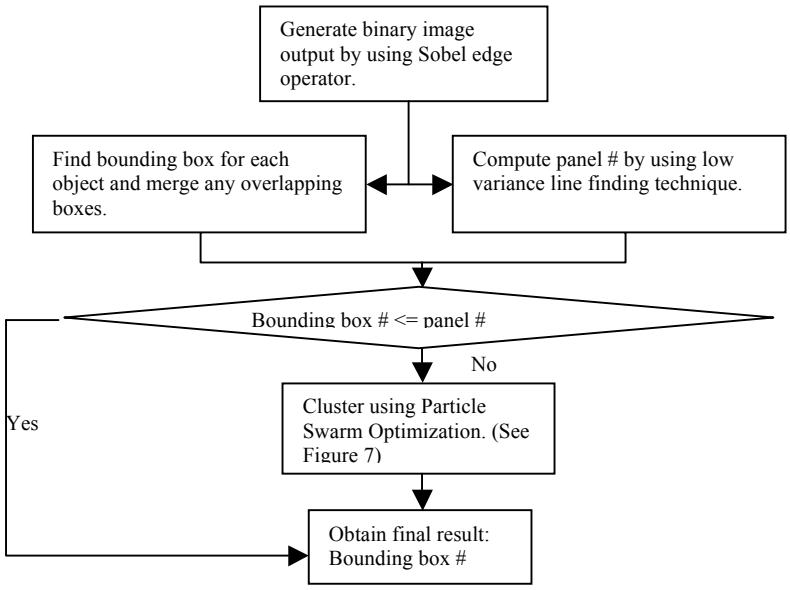


Figure 3. Overview of multi-panel images decomposition process.

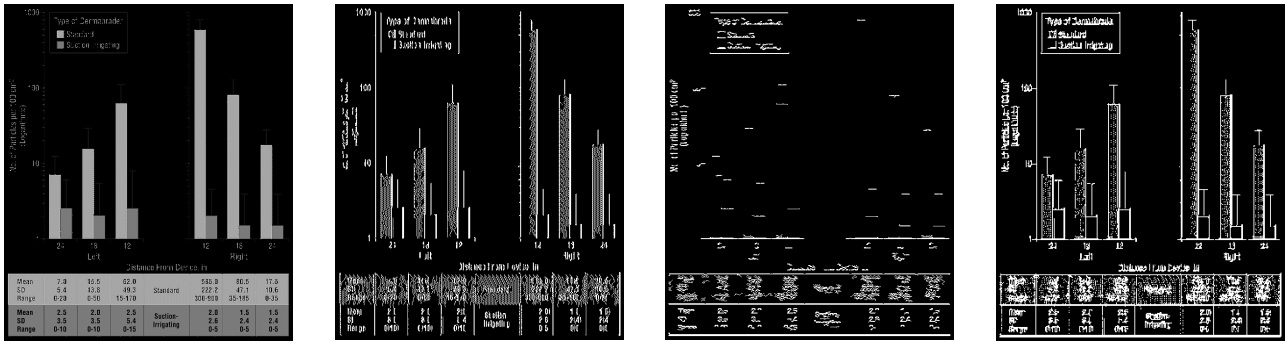
1. **Sobel edge operator:** Edges in the image are computed in the vertical and horizontal direction using the Sobel edge detector [4] with kernel values shown in Equations (1) and (2).

$$dx = \begin{matrix} -0.05 & 0 & 0.05 \\ -0.1 & 0 & 0.1 \\ -0.05 & 0 & 0.05 \end{matrix} \quad (1) \qquad dy = dx' \quad (2)$$

   The input image is convoluted with *dx* and *dy* and then Otsu's filter [5] is applied to the output. The outputs are logically OR-ed to generate the image for use in the next step. Example images from this step are shown in Figure 4.

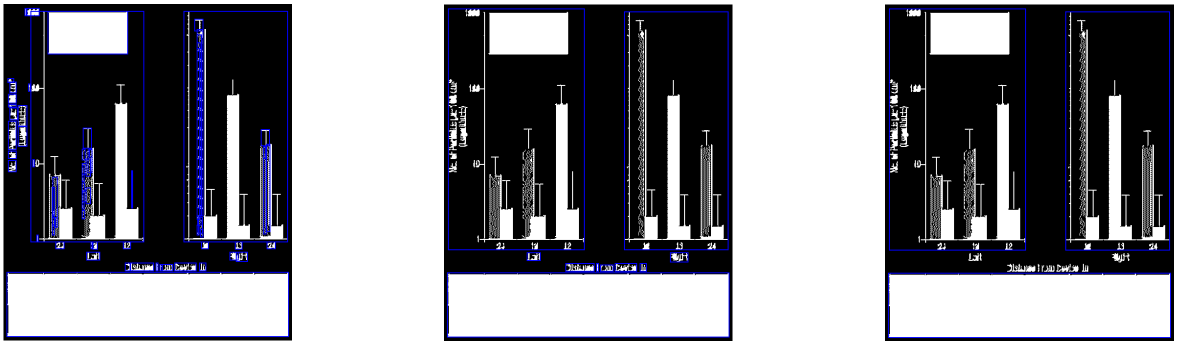
2. **Form bounding boxes:** The image output from above step is processed to fill holes inside each foreground object. Small noise objects are removed by morphological processing. Bounding box can be drawn around each object and any overlapping bounding boxes are merged. Small bounding boxes may be removed. Example images from this step are shown in Figure 5.

3. **Particle Swarm Optimization (PSO) clustering:** As shown in Figure 6, some images may not result in bounding boxes surrounding each subfigure. To overcome this difficulty, we compare the number of bounding boxes with the number of subpanel identified in the text caption using low variance line finding technique. If the number of bounding boxes is less equal to the number of subpanels then it is deemed correct. However if it is greater than the number of subpanels, an over-segmentation may be assumed and the particle swarm optimization (PSO) [6] algorithm is used to correct any error. Using shape, size, and position constraints, we devise a fitness function for the PSO algorithm (shown in Figure 7) where subfigure objects of similar shape and size that are near each other are considered as belonging to one region.



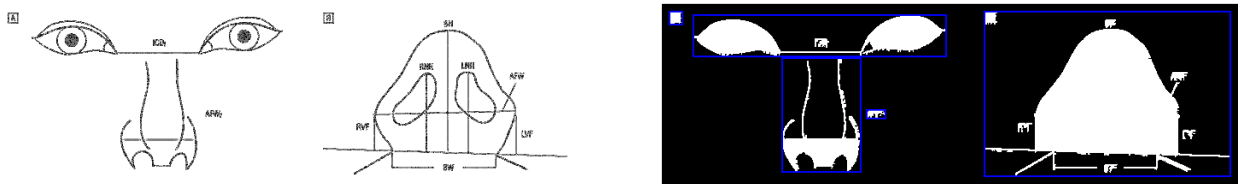
(a) (b) (c) (d)

Figure 4. Sobel edge operator processing. (a) Original image. (b) Image output after convoluting the input with ***dx*** and using Ostu's method. (c) Image output after convoluting the input with ***dy*** and using Ostu's method. (d) Shows the image after logically OR-ing images in (b) and (c).



(a) (b) (c)

Figure 5. Form bounding boxes. (a) Bounding box for each object. (b) Bounding box with overlap removed. (c) Bounding box after removing small noise bounding boxes.



(a) (b)

Figure 6. Incorrect Bounding Box Segmentation. (a) Original image. (b) Bounding box result.

We select five features for computing the fitness function of size and fourteen features for shape. These characteristic features and their explanations are shown as follows, the first five are size features and others are shape features.

- **Width**: The width of bounding box.
- **Height**: The height of bounding box.
- **Major Axis Length**: major axis length of ellipse with same normalized 2nd central moments of region.
- **Minor Axis Length**: minor axis length of ellipse with same normalized 2nd central moments of region.
- **Axis Ratio**: ratio of MajorAxisLength with MinorAxislength.
- **Solidity**: area of the region divided by the area within the Convex Hull.
- **Extent**: ratio of area to bounding box area.
- **Weighted Density Distribution features**: Twelve features are extracted by correlating the shape samples of the object inside the bounding box with weighted density distribution functions (WDD) [7,8]. The samples are computed by dividing the height of the bounding box by 20. This value is

empirically determined from prior work [8]. The length of the section of the object at each of these 20 horizontal segments is used as a shape sample for convolution and computation of the WDD features.

For n objects, each particle in the PSO is an *n*-dimensional vector where the dynamic range of each dimension *n* and objects with same value are assigned to the same cluster. The velocity is constrained to 2. The fitness function is defined as:

$$fitness_{total} = fitness_{size} + fitness_{shape}$$

where,

$$fitness_{size} = var(feature_{size})$$
$$fitness_{shape} = var(feature_{shape})$$

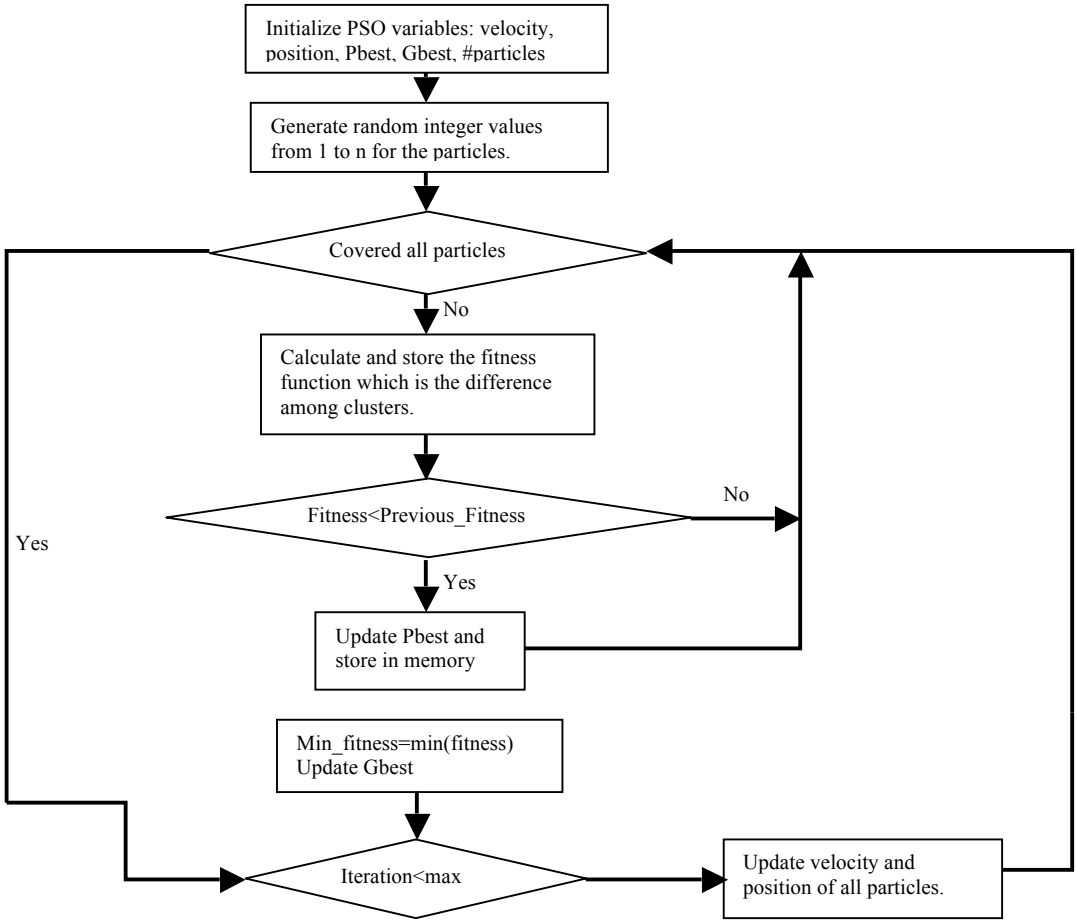


Figure 7. Steps for applying Particle Swarm Optimization algorithm.

4. EXPERIMENTS

The method for separating subfigure panels in regular images is evaluated by computing the area under ROC curve [9] by using a neural network with the architecture 5x5x1 and ten folds cross validation [10]. Preliminary results on the data set show the area to be 94.90%.

For illustration images we measure the performance by comparing the estimated number of panels with the manually generated ground truth. The method reports 92.10% accuracy in finding subfigures correctly. Finding subfigure panels in illustration images is more challenging than regular images. For success we use a combination of top-down and bottom-up approaches. The low variance line finding technique is like segmenting image in a top-down method, while, finding foreground objects and bounding boxes is akin to segmenting images in a bottom-up fashion.

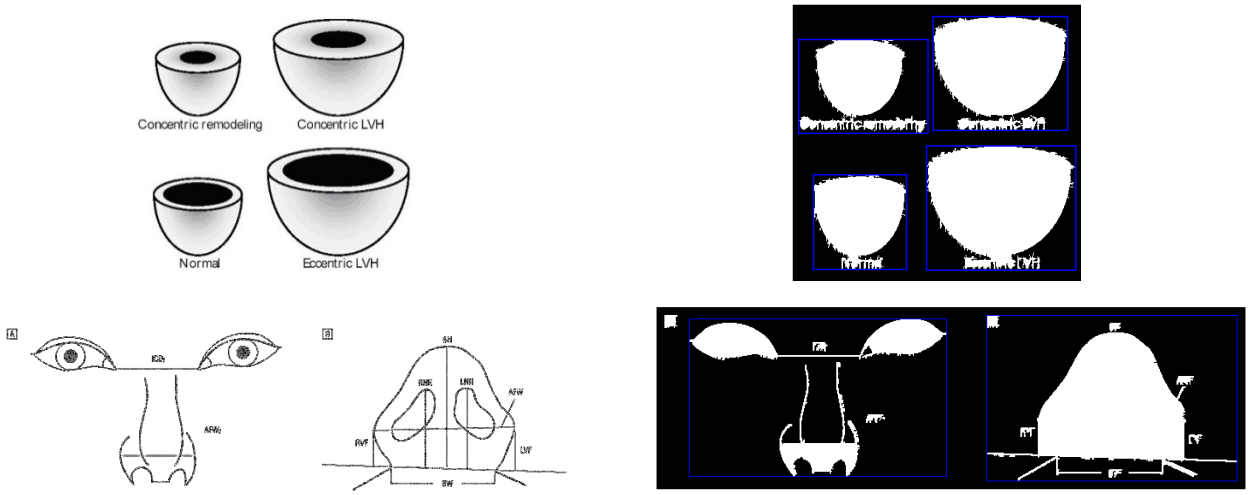


Figure 8. Sample results from Particle Swarm Optimization for finding subfigure panels.

5. CONCLUSIONS

Compared with our previous effort, this approach to locate subfigure panels is more robust. The fitness function of PSO includes not only the size but also the shape of each object, and is consequently able to identify similar shaped objects of different sizes. Furthermore, the fitness function can be easily extended and modified to support other characteristics. Experiments on a larger collection of images are ongoing and we expect to report on them in the final version of this article.

REFERENCES

[1] Deserno TM, Antani S, Long R, "Ontology of gaps in content-based image retrieval," Journal of Digital Imaging. 22(2):202-15, April (2009).

[2] Demner-Fushman D, Antani S, Simpson M, Thoma GR. Annotation and Retrieval of Clinically Relevant Images. International Journal of Medical Informatics. 2009;78(12):e59-e67.

[3] Antani S, 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. 2008, 6815:68150Q(1-10).

[4] Sobel I, Feldman G. A 3x3 Isotropic Gradient Operator for Image Processing. Presented at a talk at the Stanford Artificial Project in 1968, unpublished but often cited, orig. in Pattern Classfication and Scene Analysis, Duda R and Hart P, John Wiley and Sons, '73, pp. 271-272.

[5] Sezgin M, Sankur B. Survey Over Image Thresholding Techniques and Quantitative Performance Evaluation. Journal of Electronic Imaging, 2003, vol. 13, no. 1, pp. 146-165.

[6] Kennedy J, Eberhart R. Particle Swarm Optimization. Proceedings of the IEEE International Conference on Neural Networks, Piscataway, NJ, 1995, pp. 1942-1948.

[7] Piper J, Granum E. On Fully Automatic Feature Measurement for Banded Chromosome Classification. Cytometry, 1989, vol. 10, no. 3, pp. 242-255.

[8] Stanley RJ, Stoecker WV, Moss RH, Rabinovitz HS, Cognetta AB, Argenziano G, and Soyer HP. A Basis Function Feature-based Approach for Skin Lesion Discrimination in Dermatology Dermoscopy Images. Skin Research and Technology, 2008, vol. 14, no. 4, pp. 425-435.

[9] Fogarty J, Baker RS, and Hudson SE. Case Studies in the Use of ROC Curve Analysis for Sensor-based Estimates in Human Computer Interaction. Proc. of Graphics Interface, 2005, vol. 112, pp. 129-136.

[10] Kohavi R. A Study of Cross-validation and Bootstrap for Accuracy Estimation and Model selection. Proc. of the 14th International Joint Conference on Artificial Intelligence, 1995, vol.14, pp. 1137–1143.