Using Relevant Regions in Image Search and Query Refinement for Medical CBIR

Edward Kim[1], Sameer Antani[2], Xiaolei Huang[1], L.Rodney Long[2], Dina Demner-Fushman[2]

[1]Department of Computer Science and Engineering, Lehigh University, Bethlehem, PA;
[2]Communications Engineering Branch, National Library of Medicine, Bethesda, MD

ABSTRACT

In clinical decision processes, relevant scientific publications and their associated medical images can provide valuable and insightful information. However, effectively searching through both text and image data is a difficult and arduous task. More specifically in the area of image search, finding similar images (or regions within images) poses another significant hurdle for effective knowledge dissemination. Thus, we propose a method using local regions within images to perform and refine medical image retrieval. In our first example, we define and extract large, characteristic regions within an image, and then show how to use these regions to match a query image to similar content. In our second example, we enable the formulation of a mixed query based upon text, image, and region information, to better represent the end user's search intentions. Given our new framework for region-based queries, we present an improved set of similar search results.

Keywords: Content-based iamge retrieval (CBIR), picture archiving and communication systems (PACS), scalable vector graphics (SVG), information retrieval, region-based image retrieval

1. INTRODUCTION

The effective retrieval of relevant scientific publications and images can greatly assist in an educational or clinical setting. Especially in clinical specialties such as radiology, dermatology, and trauma surgery, image search and retrieval are suitable for diagnosis or clinical decision support; however, few efforts exploring this direction exist.[1] Further, progress in the area of content-based image retrieval (CBIR) is notoriously difficult due to the various gaps[2,3] that exist between human and machine understanding. For example, a *content* gap describes the discontinuity between human level semantics and image data. A *feature* gap could refer to the extraction and representation of these features. A *usability* gap could refer to the ease of use of a system in routine applications. To overcome several of these gaps, especially the content, feature, and usability gaps, we present our system for CBIR based upon relevant region-based image retrieval.

We build upon the ITSE system developed by the Image and Text Integration (ITI) group at the National Library of Medicine (NLM).[4] Over 70,000 images exist in this database with associated text, e.g. title, abstract, mention, image caption, and outcome. This system performs a combinational query based upon the relevant text as well as image features. From this system, we randomly sample 1000 images and associated metadata from which our framework can efficiently perform region-based image retrieval. Region based image retrieval has been performed in the past;[5] however, automatically and efficiently utilizing regions for retrieval still remains an open question. In our proposal, we explore two applications that can benefit from using regions; the first application is a global image matching problem using regions, whereas the second application addresses the more specific problem of selecting and matching medically relevant regions.

2. METHODOLGY

We utilize an image data representation that revolves around a hierarchical scalable vector graphics (SVG) abstraction,[6] i.e. a visually extensible and interactable framework. The SVG abstraction allows us to store multiple levels of segmentation, region features, and also allows us to define the interactivity with the segmentation regions. As a pre-processing step for our retrieval system, each image in our database is hierarchically segmented into a number of regions. Any automatic segmentation method can be used e.g. watershed, mean shift,[7] normalized cuts,[8] but for our application we use the gPb-owt-ucm[9] method. From the segmentation,

image features for each region are extracted and stored, including area, centroid position, 32 bin $L^*a^*b^*$ color histograms, eccentricity, orientation, major axis length, and minor axis length. Region bounding box information in the form of x,y position and width, height is also encoded. We also encode several more complex region features including the GIST[10] descriptor and PHOG[11] (a spatial histogram of oriented gradients pyramid) descriptor. The GIST descriptor attempts to capture the general properties of an image (or in our case region) using a weighted combination of multiscale-oriented filters. In our application, we compute the descriptor over 4x4 grid on 3 scales and 4 orientations per scale. The total GIST vector size is 192. For the PHOG descriptor, we compute the gradient response on using an 8 bin histogram over 3 levels. Thus, the total vector size of our PHOG descriptor for each region is 168. In order to capture a global feature representation of the image, we also compute all of the above features for the whole image.

Using these image features, we can now define how to compute the region similarity between regions i and j. For normalized scalar features between two regions, f_i and f_j, the region similarity is compared using a weighted sum of squared differences method (SSD),

$$SSD(i,j) = \sum_{n=1}^{N} w_n (f_i - f_j)^2 \tag{1}$$

where N is the total number of scalar features and w_n are feature weights. For The histogram comparison between two histograms, $h_i(k)$ and $h_j(k)$, are performed using a χ^2 measure,

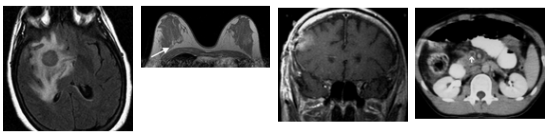
$$\chi^2(i,j) = \frac{1}{2} \sum_{k=1}^{K} \frac{[h_i(k) - h_j(k)]^2}{h_i(k) + h_j(k)} \tag{2}$$

where K is equal to 32 bins for the χ^2_{Lab} features, 192 bins for the χ^2_{GIST} features, or 168 for the χ^2_{PHOG} features. From this we compute the similarity score or distance as a weighted sum,

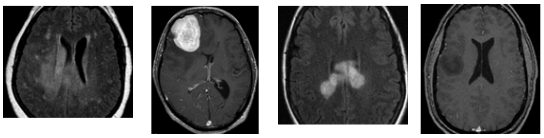
$$Dist_{sim}(i,j) = SSD(i,j) + \lambda_1 \sum_{m=1}^{3} \chi^2_m(i,j) + \lambda_2 \chi^2_{GIST}(i,j) + \lambda_3 \chi^2_{PHOG}(i,j) \tag{3}$$

where m represents the color channels and λ is a weighting constant. This region similarity score uses equal weights, e.g. $w_n = \frac{1}{N}$, $\lambda = \frac{1}{3}$, and $\lambda_2 = \lambda_3 = 1$ for all features.

2.1 Characteristic Regions for Global Image Similarity

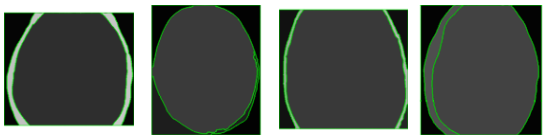


(a) Various medical images.



(c) MR images of the brain in axial view.

(b) Associated SVG large characteristic objects automatically segmented.



(d) Associated SVG large characteristic objects automatically segmented.

Figure 1. Large object characteristic regions used as an image similarity filter. Typically, similar images have similar large objects automatically segmented and stored by our method. We can further refine global image similarity by utilizing these characteristic regions.

In many medical retrieval applications, the primary objective is to find similar images to a given query image. For these systems, a logical approach would be to utilize the global features extracted from the image

in an image similarity computation. For example, if we assume that the i and j elements in equation 2 are images, I and J, then the χ^2_{Lab}, χ^2_{GIST}, and χ^2_{PHOG} computations would compute the distance between images. We notice that finding similar images through global image features or patch features provides an effective way of narrowing down the number of image candidates, but may produce erroneous results where the global characteristics cannot differentiate between two very different images. Given our encoded representation at multiple levels of segmentation, we can perform a region based image similarity step to refine the search results. At the top level of our automatic segmentation, we have very large, well defined regions that are characteristic of their particular type of image. We define these characteristic regions as having an area greater than 30% of the image, > 0.3. For example, here in Figure 1, we see several well defined regions of different types of images (a) and the characteristic regions created of a MR image scan of the brain (c) in axial view. We can incorporate our characteristic regions in a global image similarity metric. If we compare images with their global features using χ^2_{Lab}, χ^2_{GIST}, or χ^2_{PHOG}, we can utilize our characteristic regions in combination with a GIST feature as follows,

$$Dist_{GIST+R}(I, J) = Dist_{sim}(C_i, C_j) + \gamma\chi^2_{GIST}(I, J) \tag{4}$$

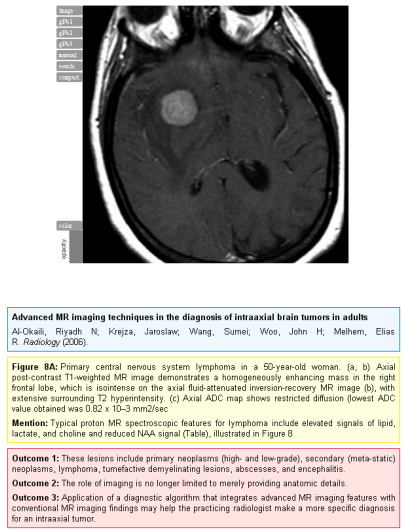
Alternatively, we can utilize the PHOG global features with characteristic regions,

$$Dist_{PHOG+R}(I, J) = Dist_{sim}(C_i, C_j) + \gamma\chi^2_{PHOG}(I, J) \tag{5}$$

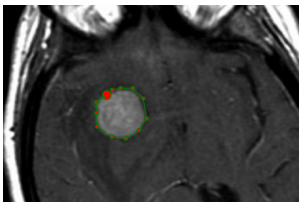
Where C_i and C_j are characteristic regions of images I and J, and the γ parameter is a normalizing constant. The $Dist_{sim}$ computation is derived using the extracted region features such as size, location, orientation, color, eccentricity, GIST, and PHOG in equation 3. Figure 1(d) illustrates that the large characteristic regions of a brain in axial view look very similar to one another, despite the detailed differences within the actual image.

2.2 Relevant Region-based Retrieval

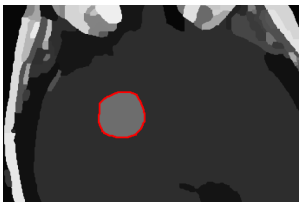
Beyond using image regions for global image similarity, we can search for a specific region within an image. For this search to be effective, there are two necessary components. The first component is a user interface that enables the user to select a region of interest in a query image. The second component is a method to determine what regions within our database are medically relevant and to compare these regions with the query region. For the first component, we built an interactive system around our SVG abstraction layer that minimizes the



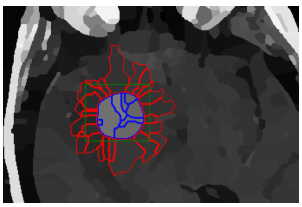
(a) Image and abstraction overlay with associated text.



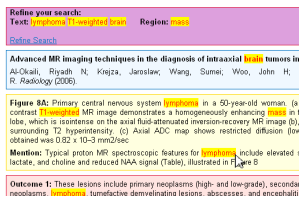
(b) Manual selection of ROI.



(c) Using segmentation regions to select the ROI.



(d) Using our bounding box method to compute a ROI.

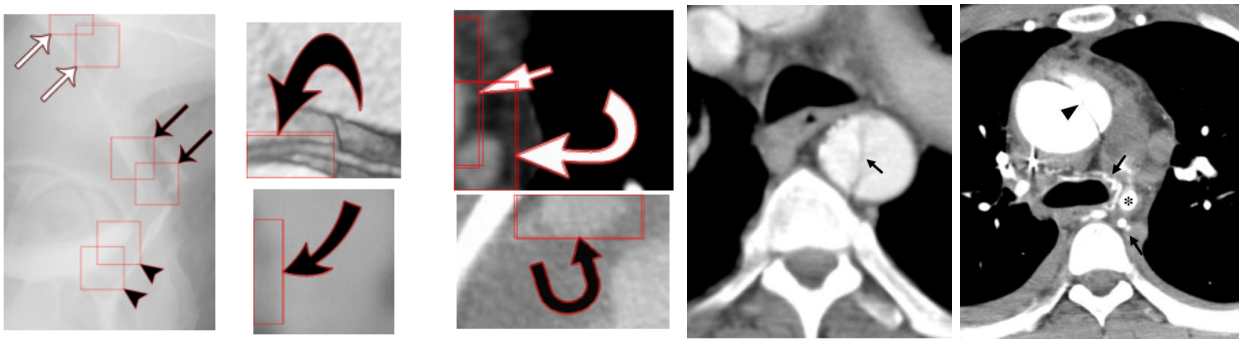


(e) Refining the query using text.

Figure 2. Example of the query refinement interface and sample of different methods enabled by our SVG abstraction to refine the user search query.

usability gap. As our SVG abstraction is web-based, our system can be used within any major web browser. Also, we can define the interaction with our SVG layer via Javascript and the XML DOM (Document Object Model). Using these different web standards, we developed and encoded several methods to perform query refinement. First, we allow for the manual selection of region by clicking points around a region of interest, Figure 2(b). Second, we can use the encoded segmentation results to select regions or combine regions to form our query, Figure 2(c). Finally, we developed a bounding box method that calculates the regions within the box and also the regions that intersect the bounding box, Figure 2(d). From this we can create a model of the background and foreground and classify each internal region as either belonging to the foreground or more closely resembling the background. In terms of text refinement, we are given several text cues related to the image (caption, outcome, title, etc.) We can select keywords from this text and perform a binary match to other associated image text to again further narrow down our search results, see Figure 2(e).

For our second component, we developed a method to identify and filter relevant regions within our database. This filter is necessary because simply matching against all the possible segmented regions in our database produces too many non-relevant region matches. To accomplish this task, we build upon the work performed by You *et al.*[12] In this work, You *et al.*[12] noticed that images often contained pointers overlaid on figures and illustrations to highlight regions of interest. These annotations were also referenced in the captions or figure citations in the text. Thus, the authors built an arrow detector using a Markov random field - hidden markov model (MRF-HMM). From the detected arrow positions and orientations, the authors hypothesized on the general region that the arrow points to, see Figure 3 (a).



(a) Results from from You *et al.*[12] (b) Sample Image (c) Sample Image

Figure 3. (a) Pointer recognition results and bounding boxes from You *et al.*[12] Caption for Image (b) Contrast-enhanced CT scan shows an intimal flap (arrow) in the descending aorta. Caption for image (c) Contrast-enhanced CT scan acquired 1 week later because the patient reported persistent pain shows aortic dilatation and dissection of the lumen (arrow).

In our proposal, we look to specifically identify the region that is referenced by these pointers. To accurately isolate the region of interest referenced by the arrows, we utilize the segmentation results stored in our SVG abstraction and our database of unlabeled images and their associated captions. Although, our images do not have any regions labeled, if we were to cluster image captions, we could collect a set of images that have a high probability of containing similar regions of interest. In Figure 3 (b)(c), we see two images that are globally disparate, but still have similar regions as described in their captions.

To measure the similarity between text captions, we utilize the natural language full text matching function in MySQL. For an image in our database, I_0, we find the top 25 similar captions and their associated images, $I_{1..25}$, where all images are represented by our SVG abstraction (they are segmented into regions). We find the centroid points of every region, and the centroid of the bounding box specified by the arrow detector. Next, we compute the distance between the region center and the arrow detector center. Any region that lies within a specified threshold distance (we use 50 pixels), is added to the candidate region set, R. Next, we perform a k-means clustering on all of the PHOG features from the candidate regions, R, where our k = 5. If the most dominant cluster center is d, and the candidate regions in I_0 are $c_1..c_n$, then we compute the distance between the two histograms using $\chi^2_{PHOG}(d, c_1..c_n)$. Finally, we choose the region whose feature distance is minimum to the top cluster center as the most likely region pointed to by the arrow.

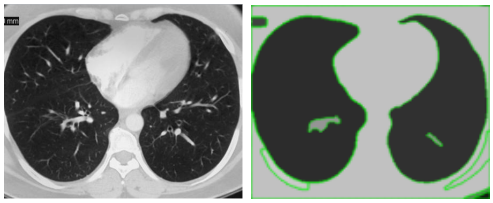
3. EXPERIMENTS AND RESULTS

For our results, we store the image features in a MySQL database and generate the SVG content on the fly. We select 1000 random images from the NLM's ITSE system database from which we apply our region-based framework. This subset of images contains multiple modalities of medical images including, CT, MR, Radiographs, Photographs, Ultrasound, Photomicrographs, and Diagrams. We perform two experiments, a global image similarity experiment using our characteristic regions, and a specific region retrieval.

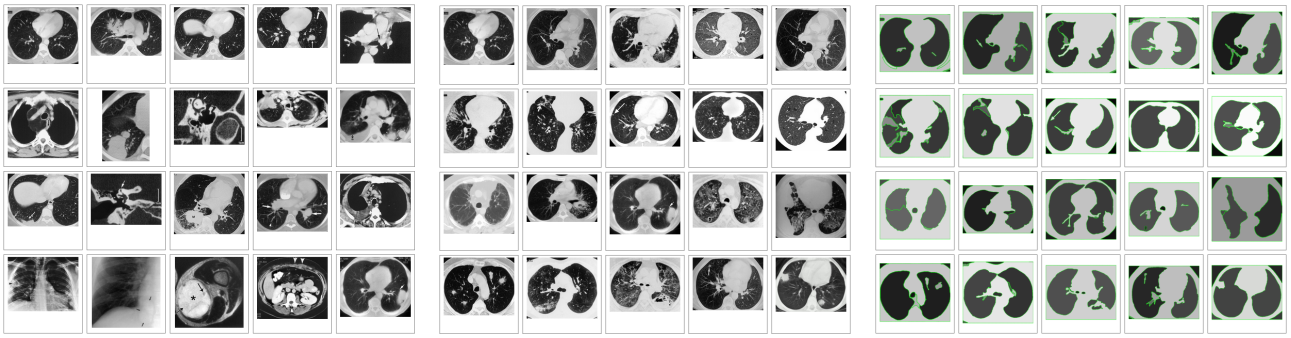
3.1 Image Retrieval using Characteristic Regions Results

We test our system performance on global image retrieval utilizing our characteristic regions. We utilize two methods from equation 4 and equation 5, which we refer to as GIST+R and PHOG+R respectively. We compare our method against three other image similarity representations, the basic $L^*a^*b^*$, GIST, and PHOG features.

For example, in Figure 4, we present the top 20 results from an $L^*a^*b^*$ feature in (c). We show our results by matching large region objects using PHOG+R as shown in Figure 4(e). Qualitatively, it can be seen that our results more closely resemble the query image in layout and content in Figure 4(d).



(a) Image. (b) Large Region.

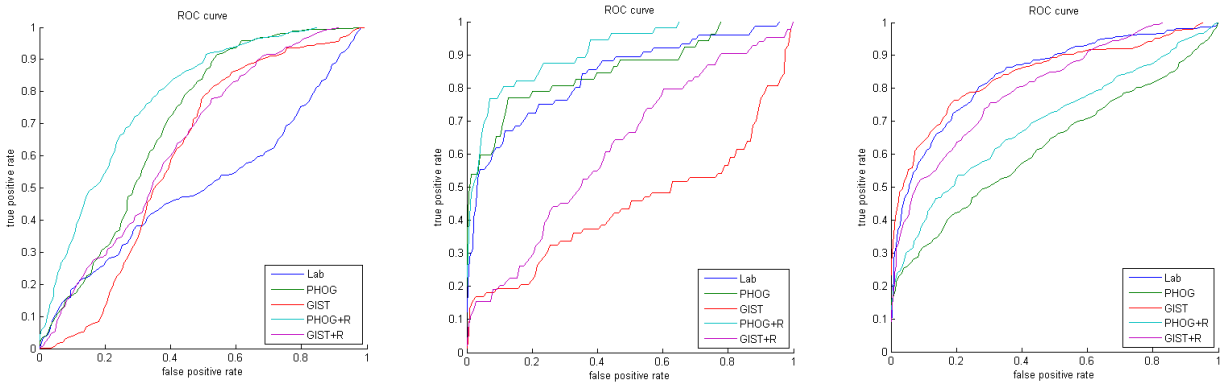


(c) Results using $L^*a^*b^*$ features. (d) Results using PHOG+R. (e) SVG abstraction showing the large regions.

Figure 4. The image similarity results from a query formed by the image (a) and large image regions shown in (b). The top 20 ranked images using global $L^*a^*b^*$ features (c) versus matching with PHOG+R and large object regions (d)&(e).

Quantitatively, we test how well we retrieve relevant images from a database. For this experiment, we define a relevant image as an image that is of the same modality as the query image. We compute an ROC curve (true positive rate vs. false positive rate) for three different modalities, CT, Radiograph, and Photomicrograph. For each modality, we average the results from 20 different images randomly selected from the database. Our results can be seen in Figure 5.

From our results, we conclude that our characteristic regions generally improve search results over the basic global image features. We also verify that different image features work well on different image modalities. For example, the GIST feature works extremely well with the Radiograph images, but poorly on CT and Photomicrographs.



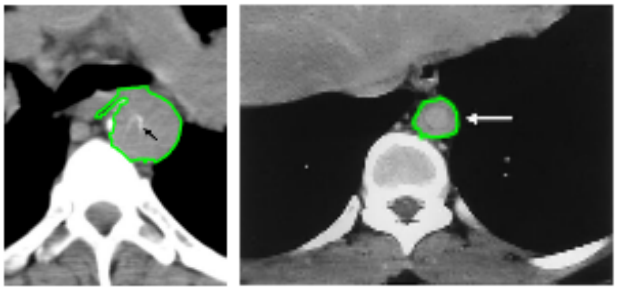
(a) CT modality (b) Photomicrograph modality (c) Radiograph modality

Figure 5. ROC curves for different modalities in our dataset. (a) CT modality, (b) Photomicrograph modality, and (c) Radiograph modality.

3.2 Region Retrieval Results

In our second experiment, we evaluate our framework in the specific region retrieval task. As previously shown in Figure 2, we provide the users a web-based platform to perform region selection. We pre-process our 1000 images from the database by running the gPb-owt-ucm[9] automatic segmentation algorithm on all of the images and extract 2 levels of segmentation. From this segmentation, we encode a total of 69,384 regions. Next, we utilize the arrow detection algorithm from You *et al.*,[12] which automatically detects 278 arrows from these images.

As our next task, we select a region within these images that correspond to the arrow's region of interest using the method described in Section 2.2. As a baseline, we build a basic region selector that simply selects the region whose centroid position is closest to the arrow detector's centroid position. We compare the baseline selector with our method that utilizes candidate regions selected by clustered text caption similarity. Because selecting the correct region is a very difficult task, we evaluate our method on how well it selects the general region of interest, and also how well it selects the specific region of interest as described in the image caption. The difference between the two is illustrated in Figure 6.



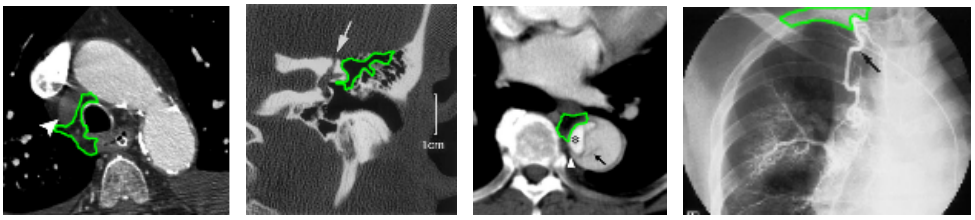
(a) General success, specific failure (b) General and specific success

Figure 6. (a) This is considered a general success as it has correctly selected the aorta which is the general area of interest. This is specific failure since the arrow is referring to the intimal calcifications within the aorta. (b) This is both a general and specific success. The caption refers to the well-defined 2-cm nodule pointed to by the arrow.

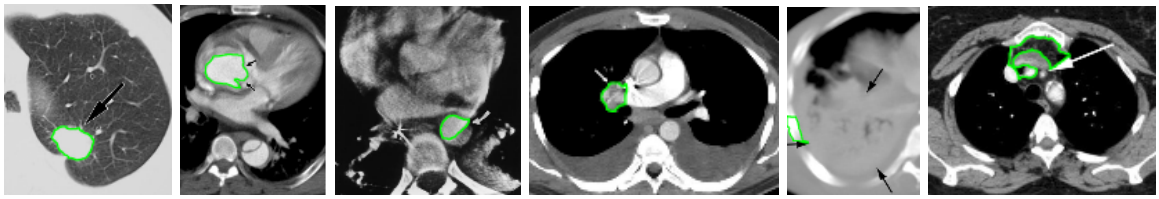
For our basic region selector, we obtain a 55.34% general region accuracy and a 28.30% specific region accuracy. When utilizing our method described in Section 2.2, we obtain a 66.38% general region accuracy and a 37.18% specific region accuracy. Figure 7 (a) illustrates some regions selected by our method, as compared to the errors made by a basic selector, Figure 7 (b). Some further examples are shown in Figure 7 (c)-(h).

Finally, we can perform a region-based query using our framework. Given the regions delineated by the

(a)



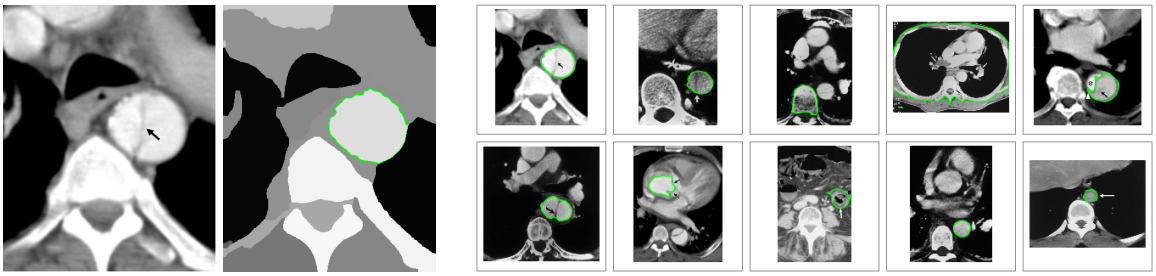
(b)



(c) (d) (e) (f) (g) Failure (h) Failure

Figure 7. (a) Region selection performed by our method. (b) Region selection performed by a basic region selector. Figures (c)-(h) display the arrow detection algorithm and the regions that are selected from our SVG abstraction. Failure cases can be seen (g)&(h) where the regions pointed to are incorrectly selected.

arrows and our SVG segmentation layer, we can search specifically for regions within images in our database that match the query region using equation 3 as seen in Figure 8.



(a) Query image. (b) Query region. (c) Top 10 regions that match.

Figure 8. Query refinement example and search for a specific region highlighted in green (b). The top 10 results of a region matching (c) with associated query text term, "aorta".

4. CONCLUSION

We present a framework for an effective region-based image retrieval based upon an SVG abstraction layer. Our framework can either perform a general image matching by using characteristic regions, or can perform a more specific region search. We show that utilizing regions in a global image similarity measure typically increases retrieval accuracy. The specific region search utilizes the interactivity of our abstraction layer and matches against relevant regions detected through arrows present in publication images. We develop and present results from our method that has improved region selection accuracy as compared to a basic region selector. Finally,

we show promising retrieval results and take a step towards minimizing the content, feature, and usability gap in medical CBIR.

REFERENCES

[1] Demner-Fushman, D., Antani, S., Simpson, M., and Thoma, G., "Annotation and retrieval of clinically relevant images," *international journal of medical informatics* **78**(12), e59–e67 (2009).

[2] Deserno, T., Antani, S., and Long, R., "Ontology of gaps in content-based image retrieval," *Journal of Digital Imaging* **22**(2), 202–215 (2009).

[3] Long, L., Antani, S., Deserno, T., and Thoma, G., "Content-based image retrieval in medicine: retrospective assessment, state of the art, and future directions," *International journal of healthcare information systems and informatics: official publication of the Information Resources Management Association* **4**(1), 1 (2009).

[4] Simpson, M., Rahman, M., Demner-Fushman, D., Antani, S., and Thoma, G., "Text-and Content-based Approaches to Image Retrieval for the ImageCLEF 2009 Medical Retrieval Track," *CLEF 2009 Working Notes. CLEF 2009 Workshop in cunjunction with ECDL2009* (2009).

[5] Carson, C., Thomas, M., Belongie, S., Hellerstein, J., and Malik, J., "Blobworld: A system for region-based image indexing and retrieval," in [*Visual Information and Information Systems*], 660–660 (1999).

[6] Kim, E., Huang, X., Tan, G., Long, L., and Antani, S., "A hierarchical SVG image abstraction layer for medical imaging," in [*Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*], **7628**, 7 (2010).

[7] Comaniciu, D. and Meer, P., "Mean shift: A robust approach toward feature space analysis," *IEEE Transactions on pattern analysis and machine intelligence* **24**(5), 603 (2002).

[8] Shi, J. and Malik, J., "Normalized cuts and image segmentation," *IEEE Transactions on pattern analysis and machine intelligence* **22**(8), 888–905 (2000).

[9] Arbeláez, P., Maire, M., Fowlkes, C., and Malik, J., "From contours to regions: An empirical evaluation," *CVPR* (2009).

[10] Oliva, A. and Torralba, A., "Building the gist of a scene: The role of global image features in recognition," *Progress in brain research* **155**, 23–36 (2006).

[11] Bosch, A., Zisserman, A., and Munoz, X., "Representing shape with a spatial pyramid kernel," *Proceedings of the 6th ACM international conference on Image and video retrieval* , 401–408 (2007).

[12] You, D., Antani, S., Demner-Fushman, D., and Thoma, G., "Biomedical article retrieval using multimodal features and image annotations in region-based CBIR," in [*Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*], **7534**, 30 (2010).