AN INFORMATION GATHERING SYSTEM FOR MEDICAL IMAGE INSPECTION

Young-Jin Lee and Peter Bajcsy

National Center for Supercomputing Applications, University of Illinois at Urbana-Champaign

ABSTRACT

We present an information gathering system for medical image inspection that consists of software tools for capturing computer-centric and human-centric information. Computer-centric information includes (1) static annotations, such as (a) image drawings enclosing any selected area, a set of areas with similar colors, a set of salient points, and (b) textual descriptions associated with either image drawings or links between pairs of image drawings, and (2) dynamic (or temporal) information, such as mouse movements, zoom level changes, image panning and frame selections from an image stack. Human-centric information is represented by video and audio signals that are acquired by computer-mounted cameras and microphones. The short-term goal of the presented system is to facilitate learning of medical novices from medical experts, while the long-term goal is to data mine all information about image inspection for assisting in making diagnoses.

In this work, we built basic software functionality for gathering computer-centric and human-centric information of the aforementioned variables. Next, we developed the information playback capabilities of all gathered information for educational purposes. Finally, we prototyped text-based and image template-based search engines to retrieve information from recorded annotations, for example, (a) find all annotations containing the word "blood vessels", or (b) search for similar areas to a selected image area. The information gathering system for medical image inspection reported here has been tested with images from the Histology Atlas[1,2] database.

Keywords: Expert knowledge, computer-centric information, human-centric information, visual annotations, audio/video signal processing, semantic gap, content-based medical image retrieval system

1. INTRODUCTION

Although medical image inspection is one of the most frequently used clinical techniques, it has been considered to be difficult for medical novices, such as medical students, to master. One of the many reasons making it difficult to be learned is the abstract nature of medical image inspection processes. Traditionally, medical image inspection has been taught within a didactic framework[3] where an expert doctor demonstrates how he or she examines an image to his or her students. The expert doctors, however, may not clearly explain why and how they reach their final diagnosis while inspecting images. Many medical image inspection techniques are based on tacit knowledge of experts, and hence it is difficult for novices to learn image inspection processes. Naturally, one would like to answer the following two questions. First, could we learn from medical experts by gathering all information about medical image inspection? Second, if we would gather all information about experts during image inspection, could we develop a computer-based assistant for medical image inspection? Our work is motivated by the first question and approached by collecting as much information about medical inspection process as possible.

In this paper, we present an information gathering system for medical inspection processes. The information captured in the present system can be grouped into two categories including computer-centric information and human-centric information. Computer-centric information refers to information that the present system produced while users are inspecting their images. Typical examples of computer-centric information are image overlay drawings, textual descriptions attached to interesting image sub-areas, and speed of users' mouse movements. Human-centric information, on the other hand, is information created solely by users, such as voice and facial expressions, while they are examining their images. Next sections explain how the developed system captures computer-centric and human-centric information in more detail.

2. COMPUTER-CENTRIC INFORMATION

In this work, computer-centric information includes various forms of information during image inspection by users of the developed information gathering system. We categorize computer-centric information into two classes, (a) static computer-centric information and (b) dynamic computer-centric information. Examples of static computer-centric information are image drawings and textual annotation describing meaningful image sub-areas. On the other hand, examples of dynamic computer-centric information are speed and acceleration of the mouse movements of users, zoom level changes, and image panning and frame selections from an image stack. One of the ways to distinguish dynamic computer-centric information from static computer-centric information is to consider how conscious users are while producing information. If users expect explicit outputs from their actions, e.g., creating image and/or textual annotations on an image sub-area, such actions will be captured as static computer-centric information. On the other hand, if they do not expect explicit outcomes from their actions, e.g., moving the mouse pointer slower or faster during image inspection, such actions will be captured as dynamic computer-centric information. Next subsections describe how the present system collects computer-centric information in greater depth.

2.1. Static Computer-Centric Information

2.1.1. Image Annotations

At the initial phase of this research, we have consulted histologists and pathologists to find out useful image annotations that could facilitate their image inspection processes. The rationale of supporting image inspections with image annotations is to enable medical experts to easily create meta-data. The meta-data are representations of expert's knowledge applied to an image sub-area during image inspection. Such meta-data files can help other people, especially medical novices, better understand what the experts had discovered in an inspected image.

Based on the suggestions from the histologists and pathologists we have designed the system capable of creating basic annotations and compound annotations. Basic annotations are designed to create image overlay drawings over a single image sub-area of interest while compound annotations are designated to represent a relationship between two different image sub-areas. The developed system allows users to create six different types of basic image annotation and one compound image annotation. The basic image annotation types supported are rectangle, circle, points, threshold, segment and cluster types. For example, shows (1) a circle type image annotation that can be effectively used for an oval-shaped structure, such as an artery, and (2) a point-type image annotation that can be used to precisely select a set of pixels of interest and group them together as a single image annotation instance.
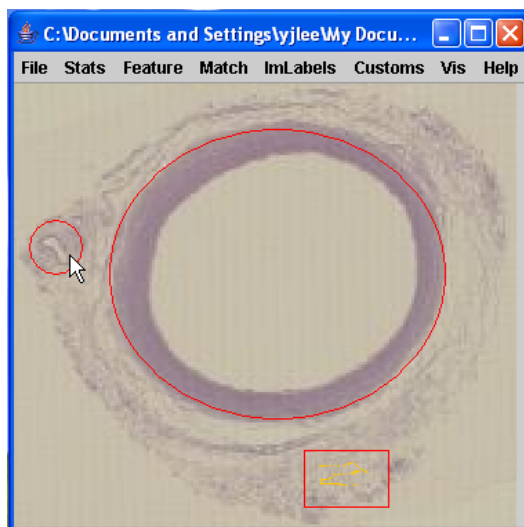

Figure 1: Example annotations created with the developed system. The large circular type annotation indicates an image of mixed artery, and the small circular type annotation indicates vasa vasorum. The points-type annotation shown as a yellow stippling with the box indicates loose connective tissue of the adventitia. Double-click on the annotations will show the textual descriptions associated with them.

Although a point type image annotation allows users to create an image annotation of irregular shape, it may not be efficient with a large irregular image sub-area because users have to manually click all image pixels forming an irregular region. In order to facilitate image overlay drawings for regions of irregular shape, the developed system provides threshold, segment, and cluster annotation types by utilizing the I2K toolkit[4]. Each annotation type is suitable for different cases of irregularly shaped regions.

A threshold type image annotation is shown in Figure 2. It creates an overlay drawing by thresholding selected image sub-area, tracing the boundary of a user chosen blob, and filling up the extracted boundary with a specified color. The threshold type image annotation is best suited when a region of interest contains holes due to heterogeneous colors and the holes should be part of the drawing (see Figure 2). If holes inside of an irregular region are meaningful, then segment annotation type can be used. The developed system takes advantage of the region growing segmentation algorithm of the I2K toolkit[4] to create a segment type image annotation. During sub-area segmentation, users specify initial spatial locations (seeds) of desired image segments (contiguous regions) by a mouse click, and select minimum similarity of interior pixels for pixels forming a contiguous segment. If the objective is to annotate a set of pixels that have similar colors, then cluster annotation type should be chosen. The developed system uses the Isodata clustering algorithm implemented in the I2K toolkit[4], which represents an advanced K-means clustering algorithm[5]. The difference between segment and cluster image annotation is that the segment image annotation will form an irregularly shaped region that is an aggregation of spatially contiguous image pixels while the cluster image annotation will contain a set of any image pixels satisfying intensity similarity criterion.

The link type image annotation is the only compound annotation currently supported. A user can connect two image sub-areas of interest using the compound annotation. This type of annotation could be useful in describing a cause-and-effect relationship between two image sub-areas, e.g., a cancel cell is a cause of deformation in other area. We plan to extend the link type annotation to connect any types of two basic annotations in the future.
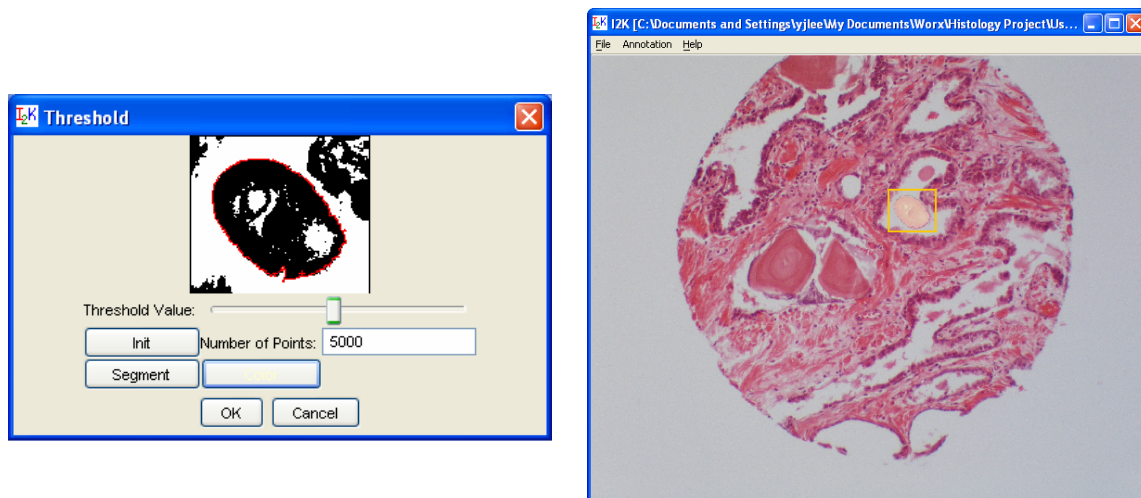

Figure 2: Extracting an irregular boundary on the thresholded image sub-area (left) and create a threshold type image annotation on the image (right).

2.1.2. Textual Annotations

Although the image annotation discussed in the previous section enables users to specify important image sub-areas, it is still not easy for other users to understand what these image annotations actually mean. This difficulty is well known as a "semantic gap," which refers to the loss of information from an actual image to a representation by low level image features[6]. For instance, as shown in Figure 1, even though a medical expert created a point type annotation to indicate that these points represent "loose connective tissue of the adventita" in that area, it may not be clear to medical novices what these points actually represent. In order to bridge this semantic gap, the developed system allows users to use textual annotation to describe the image sub-area of interest, and associate it with the corresponding image sub-area. When an image annotation is being created, the developed system provides a dialog box where users can type in textual descriptions of the selected image drawing (see Figure 3). Once a textual description is attached to an image sub-area, users can easily edit or delete it later. For instance, a user can double-click on an image annotation to bring up the

Annotation Dialog with a textual description, and change or delete the text in the Annotation Dialog. Although users are required to type free text to describe the area of interest in the current implementation, it is planned to provide pre-determined keywords from the ontology of the domain knowledge in the future.
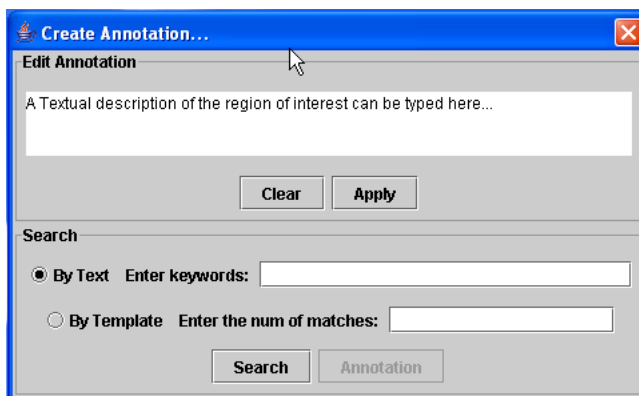

Figure 3: An Annotation Dialog box to associate a textual description with an image sub area.

2.1.3. Example Use of Static Computer-Centric Information

2.1.3.1. Text-Based Search

Once image drawings and text annotations are created, they can be used for search purposes. If users want to find all existing image drawings associated with a particular textual description, they can use the "search by keywords" functionality provided in the Annotation Dialog. After specifying a target text in the "Enter keywords" text field (see Figure 3), the developed system returns a list of annotation objects whose textual descriptions include the target keywords provided (see Figure 4). This keyword-based search becomes useful in those situations when the number of annotations is large.

We should mention that the problem of storing meta-data including image drawings, textual descriptions and original images has been solved by using an open source Hierarchical Data Format (HDF[7]) file format There is a need to search through many input images and many annotations, and the HDF format is a convenient way to package and search all recorded information. For example, if one deals with images of Tissue MicroArrays (TMA) acquired at multiple optical resolutions, a file format like HDF becomes a good solution. Figure 4 shows the results of a keyword-based search executed against multiple TMA images packaged into one HDF file. The tree structure in the left panel of Figure 4 represents all matching textual annotations, and the image in the right panel is updated with the image of a sub-area associated with the textual annotation selected by a mouse click.
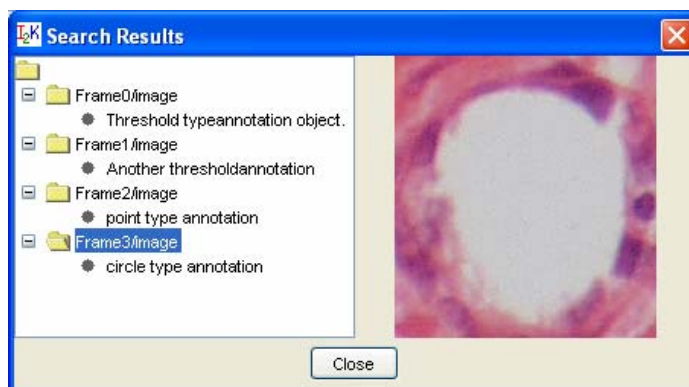

Figure 4: The keyword-based search results. The tree structure shown in the left lists all matching annotation objects and the panel in the right displays the image sub-area currently selected.

2.1.3.2. Image Template-Based Search

In addition to the text-based search, the developed system allows users to search for similar image sub-areas to a selected image template. This functionality can be used for identifying similar regions in multiple images and propagate annotations from the already inspected images. The propagation of textual annotation can be accomplished with the image template-based search function according to the following description. First, users double-click on any existing annotation object that would be used as a template for finding other similar image sub-areas. Then, specify the number of similar image sub-areas that they would like to find in the "Enter the number of matches" text field in the Annotation Dialog box (see Figure 3) and choose the "By Template" radio button. Then, upon pressing the "Search" button, the developed system tries to find similar image sub-areas by calculating normalized correlations between the template and other parts of the image. When the template search finishes, all matching image sub areas are marked as green rectangles with a cross as shown in Figure 5. Although the current implementation only uses a normalized correlation to find similar image parts, other image matching techniques, such as color, shape, texture, or wavelet coefficient, could be incorporated in the future to enhance the accuracy of the image template-based search capability.
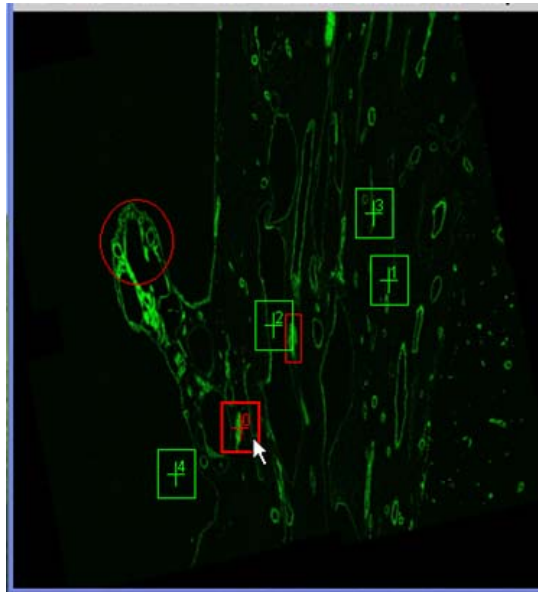

Figure 5: An outcome of a template-based search. The rectangles are areas that a template-based search found similar to the template used in image characteristics.

2.2. Dynamic Computer-Centric Information

We hypothesize that dynamic computer-centric information, such as speed of mouse movements, zoom-level changes, or the order of frame inspection, can be used in estimating complexity and significance of information in the image areas. The rationale behind this assertion is that the time to inspect a given image sub-area would be proportional to the amount of information presented in that area. In other words, we can relate the significance or usefulness of an image sub-area to the amount of time people spend to examine that particular area. The developed system is able to monitor users' image inspection activities, and produce dynamic computer-centric information. Upon clicking the "Start Recording" button in the Dynamic Annotation Dialog, all image inspection activities become recorded until the "Stop Recording" button is clicked, and the "Save Record" button saves the recorded information into a text file (see Figure 6). Furthermore, with the "Play Record" button users can playback any previously saved user activities and watch how experts reach their final diagnosis. This playback functionality could help novices to go over the medical image inspection procedure as performed by the medical experts.
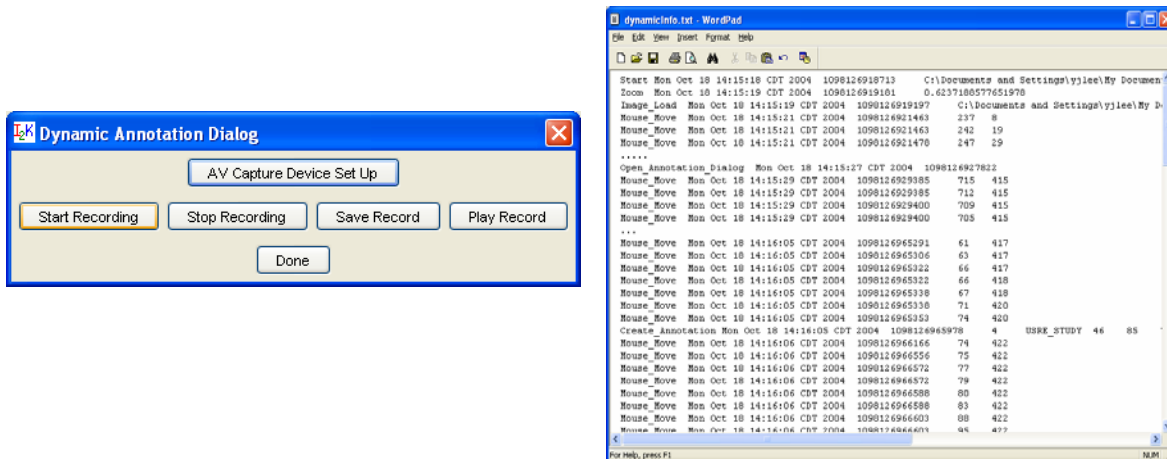
Figure 6: The Dynamic Annotation Dialog (left) and the dynamic computer-centric information produced (right).

2.2.1. Example Analysis of Dynamic Computer-Centric Information

Among the variety of recorded variables as dynamic computer-centric information, we have focused on the analysis of (a) the speed of mouse movements and (b) the changes in image zoom-in/out levels. It is expected that mouse movements of users would become slower near the meaningful image sub-areas, and image zoom-in/out activities would be performed in the proximity of areas of interest. Figure 7 shows a preliminary analysis of the relationship between mouse speed and complexity of image areas under investigation. As illustrated in Figure 7, although many slow mouse movements are recorded around the image annotations created, as we expected, there are several image sub-areas that are also related to slow mouse movements without being associated with image annotations. Those areas can be considered as interesting (or similar to suspicious areas) from an expert's viewpoint, but are not annotated since they do not contain enough evidence (critical image characteristics) for expert pathologists to create annotations.
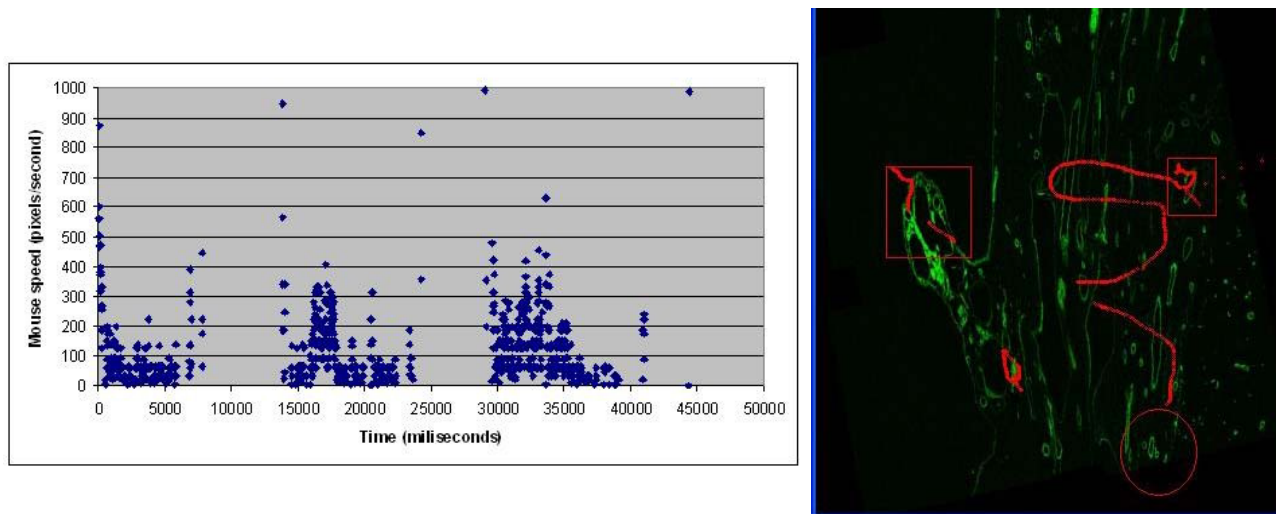

Figure 7: An illustration of a mouse movement speed as a function of inspection time (left). The speed significantly decreases in the proximity of interesting and annotated regions. One can visualize image areas where the mouse movement speed was smaller than 50 pixels per second (right).

3. HUMAN-CENTRIC INFORMATION

As briefly discussed earlier, human-centric information is created solely by human users while they are inspecting images. It is anticipated that human-centric information can provide additional information complementary to

computer-centric information. For instance, changes in the size of human eye pupils during an image inspection process can indicate that users have noticed something meaningful even though no image drawings or textual annotations are created[8] It is also possible to record user's voice during image inspection. One can associate annotations with words, pauses and audible voice changes indicating relevant and irrelevant information, and potentially expert's confidence in the created annotations.

The current version of the developed system captures two types of human-centric information, the voice and facial expressions of users, using the JMF (Java Media Framework) toolkit[9]. As illustrated in Figure 8, the "AV Capture Device Setup" button of the Dynamic Annotation Dialog allows users to capture the voice and facial expressions of users with any audio and video equipments supporting JMF. The captured voice and facial expressions can be saved in many different formats including AVI, QuickTime or MPEG format (see Figure 8).
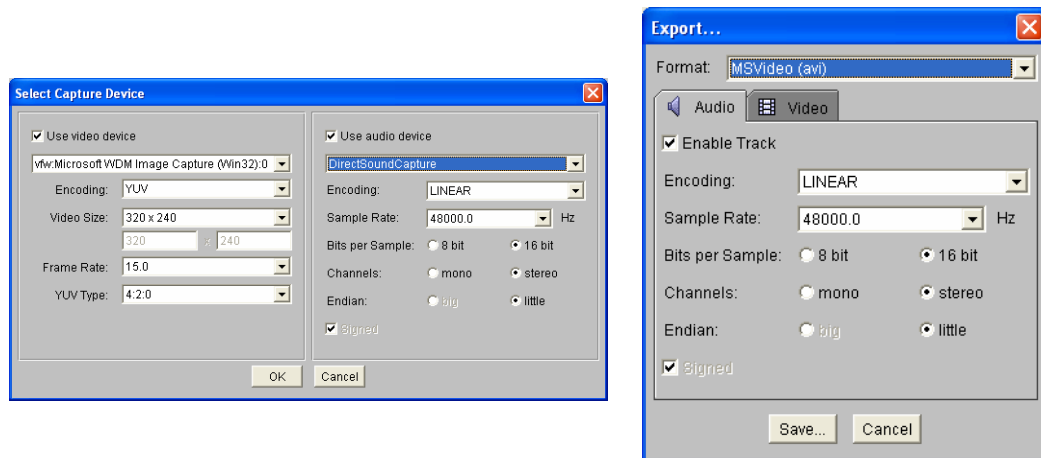

Figure 8: Users can choose appropriate audio and/or video source to capture human-centric information (left) and save the captured information into various audio/video formats (right).

In our current work, human-centered information is not processed and analyzed after it has been collected. However, we anticipate that human-centered information could be processed in the following ways. First, audio information could be embedded into image annotations. In other words, as users create image annotations, users can record their voice describing the image annotation being created. It could be also possible to produce computer-generated voice reading the textual descriptions attached to an image annotation to which the mouse cursor is pointing as users browse the image under inspection. A voice-based search would be another example use of human-centric information. Second, human-centered information can be analyzed to support dynamic computer-centric information. Based on video signal, eye movement researchers have shown that fixations in eye movements are strongly linked to intense cognitive processing. According to Viviani[8], at least three processes occur during an eye fixation: encoding of visual stimuli, sampling of the peripheral field, and planning for the next eye movements, often called saccade. Therefore, we can verify our hypothesis that the mouse movements of users would be slower near important image sub-area by comparing where slow mouse movements occur and where eye fixations happen.

4. DISCUSSIONS

In this paper, we described an information gathering system capable of collecting various computer-centric information and human-centric information. Computer-centric information includes (a) static information, such as image drawings and associated textual annotations, and (b) dynamic information, such as speed of mouse movements, zoom level changes, and image panning and frame selections. Human-centric information is represented by voice and facial expressions of users while inspecting images. The developed system can be directly used as a computer-based learning environment for medical novices. Static and dynamic computer-centric information created by experts during image inspection will enable medical novices to go over the image inspection process as performed by highly knowledgeable people. The developed system will help medical novices replay the image inspection sequence to improve their understanding about how experts reached the final diagnosis.

In future, we would like extend the developed system into an image retrieval system that could bridge semantic gaps, e.g., the loss of information from an actual image to a representation by low level image features[6]. Another part of

future research will be to demonstrate that dynamic computer-centric information, such as mouse movements and zoom level changes during image inspections, are reliable for discriminating content-important image sub-areas from content-irrelevant and background image sub-areas.

REFERENCES

1. Internet Atlas of Histology, Retrieved November 15, 2004, from http://www.histo.net
2. Urbana Virtual Microscope, Retrieved November 15, 2004 from http://www.histo.org
3. R. L. Johnson et al., *Final Report of the FOPE II Education of the Pediatrician Workgroup*, Pediatrics, 106/5 Supplement, 1175-1198, 2000
4. I2K, Image To Knowledge, Retrieved November 15, 2004 from http://i2k.ncsa.uiuc.edu
5. J. T. Tou and R.C. Gonzales, Pattern Recognition Principles. Addison Wesley Publishing Company, pp. 95-107 (Isodata Clustering), 1974 .
6. A. W. M. Smeulders, M. Worring, S. Santini, A. Gupta, R. Jain, *Content-based image retrieval at the end of the early years*, IEEE Transactions on Pattern Analysis and Machine Intelligence, 22/12, 1349-1380, 2000.
7. HDF, Hierarchical Data Format, Retrieved November 15, 2004, from http://hdf.ncsa.uiuc.edu
8. P. Viviani, *Eye Movements and Their Role in Visual and Cognitive Processes*, Chapter 8, Elsevier Science, Amsterdam, 1990.
9. Java Media Framework API Guide, Retrieved November 15, 2004, from http://java.sun.com/products/java-media/jmf/2.1.1/guide/index.html