Young-Jin Lee, yjlee@ncsa.uiuc.edu
Peter Bajcsy, pbajcsy@ncsa.uiuc.edu

Automated Learning Group
National Center for Supercomputing Applications
605 East Springfield Avenue, Champaign, IL 61820

Software Tools for Recording Image Inspection Processes

Abstract
We present an information gathering system for medical image inspection that can record and replay computer- and human-centric information. Computer-centric information consists of static measurements about image drawings and associated textual annotations, as well as, dynamic measurements about image frame selection, zoom level, image panning, sub-area cropping, and mouse movement. Human-centric information includes audio and video signals during the image inspection process. Both types of information are recorded for educational and analysis purposes. All software tools were tested with images from the Histology Atlas database (www.histo.net and www.histo.org).

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 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 where an expert doctor demonstrates how he or she examines an image to medical students. The expert doctors, however, may not be able to clearly explain why they reach the final diagnosis because many of medical image inspection techniques are based on tacit knowledge, making it difficult for novices to learn the image inspection processes.

The developed system addresses this problem by recording computer-centric and human-centric information while medical experts are inspecting images. By replaying the recorder information, medical novices can go over the image inspection procedures as performed by medical experts. Computer-centric information is about medical expert's interaction with digital images during image inspection by using human computer interfaces (HCI), for example, a mouse and a keyboard. Computer-centric information can be divided as (a) static measurements, such as image drawings or textual descriptions attached to interesting image sub-areas, or (b) dynamic measurements such as speed of mouse movements and zoom level changes. Human-centric information, on the other hand, is information created solely by users, such as voice and facial expressions, during image inspection. In the next sections, we describe the developed tools for recording and replaying computer-centric and human-centric information in more detail.

2   Computer-Centric Information

2.1   Static Measurements

The prototype software consists of two components, the Image Viewer and the Control Panel as shown in Figure 1. The Image Viewer allows users to view images in various formats and create annotations without changing the underlying image while the Control Panel allows users to control the type of image annotations in the Image Viewer.
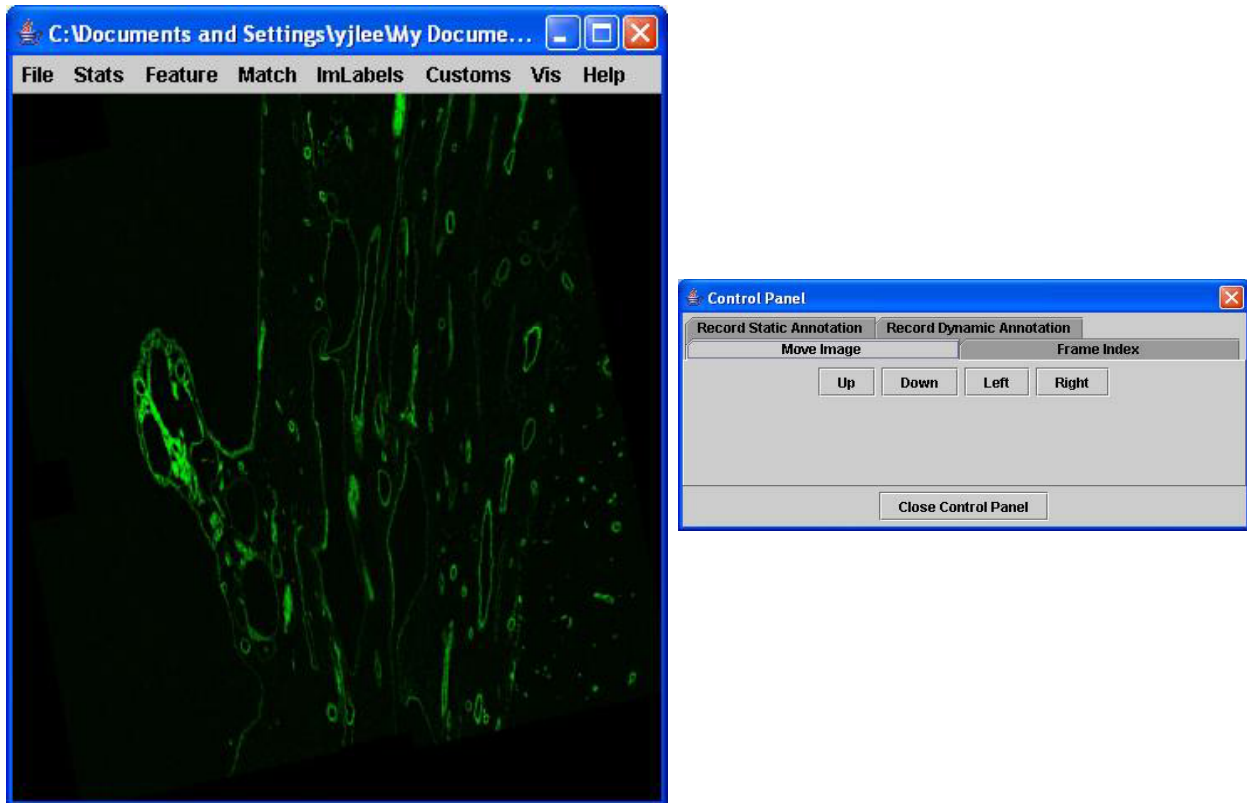
Figure 1: The Image Viewer (left) and the prototype Control Panel dialog (right).

The Control Panel consists of four tabs, "Move Image," "Frame Index," "Record Static Annotation," and "Record Dynamic Annotation." With the "Move Image" tab selected, users can change the viewing area of an image under investigation by clicking "Up," "Down," "Left," or "Right" button. For multiple frame images, such as confocal microscope images, the slider in the "Frame Index" tab can be used to load an appropriate image frame (see Figure 2).
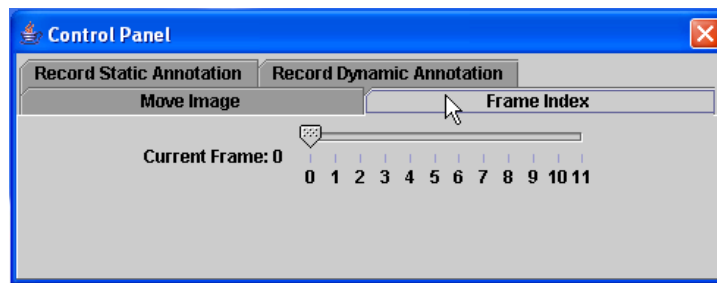

Figure 2: The "Frame Index" tab of the Control Panel.

2.1.1 Image Drawings

Users can select an appropriate type of image annotations with the buttons under the "Record Static Annotation" tab of the Control Panel as shown in Figure 3. The "Rect" and "Circle" button can be used to create a rectangular and oval shape image annotation, and "Points" button enables users to create an arbitrary-shaped image annotation by allowing them to precisely select pixels of interests. The "Link" button creates a link-type image annotation,

which can be used to describe a relationship between two image sub-areas. In addition, the developed system enables creating an image annotation using the image contents rather than using predefined shapes, which will be explained in the next subsection. All image annotations are visible by default in the Image Viewer when they are created. However, users can control the visibility of image annotations by selecting either "Show Annotations" or "Hide Annotations" radio button in the Control Panel (see Figure 3).
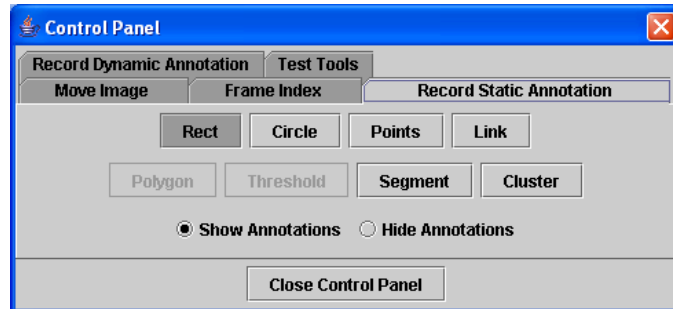

Figure 3: Dialog for choosing the type of image drawing.

2.1.1.1 <u>Predefined Drawings</u>

The developed system provides four predefined image annotation types, "Rect," "Circle," "Points," and "Link". This subsection explains how to create four annotation types. The first step to create either a rect-type or circle-type annotation is to specify a region of interest in the Image Viewer component where the annotation object will be placed after creation. Users can specify a region of interest by drawing a rectangle in the Image Viewer component. Then, double-click on inside of a rectangle and enter textual description into a displayed Annotation Dialog (see Figure 4). When the "Apply" button of the Annotation Dialog box is clicked, a new either rect-type or circle-type image annotation is created. The type of annotation (Rectangle or Circle) depends on the selection in the "Record Static Annotation" tab of the Control Panel shown in Figure 3.
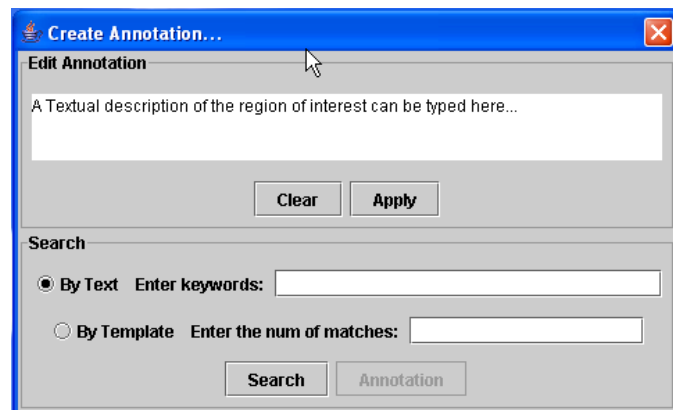

Figure 4: The Annotation Dialog

The points-type annotation can be created in a similar way. The only difference is that users need to select all image pixels in the Image Viewer component that they want to associate with textual descriptions in the Annotation Dialog. In this mode, the mouse cursor changes its appearance to a finger shape, and operates like a free-hand paintbrush tool. Figure 5 illustrates examples of using (a) the circular type annotation for delineating a mixed artery (large circle) or

vasa vasorum (small circle) and (b) the point type annotation (yellow stippling within the box) for indicating loose connective tissue of the adventitia.
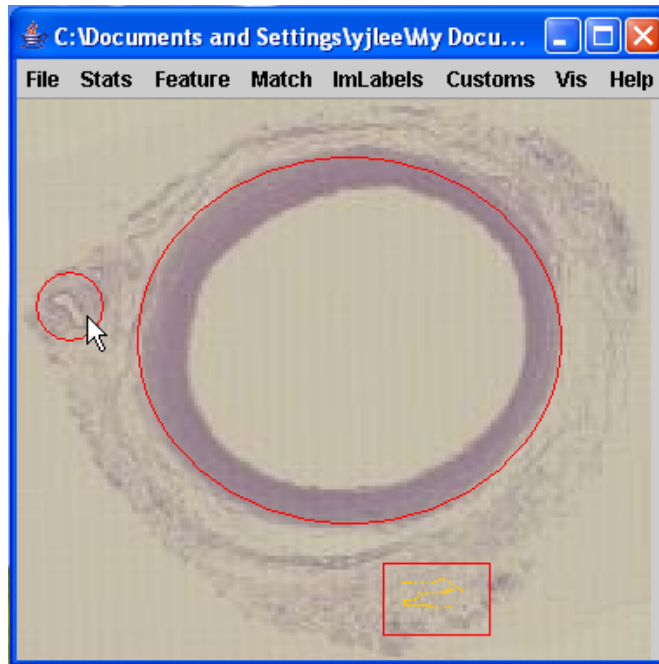

Figure 5: Example annotations created with the developed software. The large circular type annotation indicates an image of mixed artery, and the small circular type annotation indicates vasa vasorum. The yellow stippling with the box indicates loose connective tissue of the adventitia. Double-click on the annotations will show the textual descriptions associated with them.

The current implementation of the prototype software supports also a link-type annotation for annotating a relationship between two image sub-areas. In order to create a link-type image annotation, users drag the mouse cursor from one area of interest to another one. In the case of a rect- or circle-type image annotation, double-clicking on the link will bring up the Annotation Dialog for typing in the corresponding textual annotation. Finally, a link type image annotation is created when the "Apply" button is clicked while the selection of the "Link" button in the "Record Static Annotation" tab of the Control Panel.

Regardless of the type of annotations, any image annotation can be easily viewed, updated or delete later with a simple double-click on it. Double-clicking will display the Annotation Dialog containing the associated textual description that can be modified and saved again.

2.1.1.2 Drawings Driven By Image-Content

In addition to the predefined shape drawings, users can create image drawings that have irregular shapes by using "Threshold[1]," "Segment," and "Cluster" tools available in the "Record Static Annotation" tab of the Control Panel. The "Threshold" tool extracts a boundary inside of an image sub-area that corresponds to a blob (connected set of pixels) with intensities above or below a certain intensity threshold. The irregularly shaped drawing is formed by filling up the extracted boundary with a specified color. Although threshold, segment, and cluster image

---

[1] The Threshold tool is available only to the Annotation Tool based on the I2K version 2. The Annotation Tool based on the I2K version 2 can be checked out from the i2kv2 branch of the CVS branch.

annotations are all applicable to an image sub-area of irregular shape, they are designed for different cases. For instance, the threshold type image annotation is best suited when the image sub-area has an irregular shape, has empty holes in it and the holes inside are not of interests because the resulting threshold image annotation will fill up the holes inside. On the other hand, if holes inside an irregular are meaningful, either segment or cluster type image annotations can be created. The developed system takes advantage of the region growing segmentation algorithm of the I2K toolkit [1] to create a segment type image annotation as illustrated in Figure 6. Users can precisely control the segmentation process by specifying initial spatial locations (seeds) of desired image segments (regions) by a mouse click and selecting minimum similarity of interior pixels forming a contiguous segment. Finally, the present system uses the Isodata clustering algorithm from the I2K library, which implements an extended K-means clustering algorithm for the cluster type image annotation. The difference between segment and cluster image annotation is that the segment image annotation will represent an irregular image shape that is spatially contiguous aggregations of image pixels while the cluster image annotation will represent aggregations of any image pixels satisfying intensity similarity criterion.
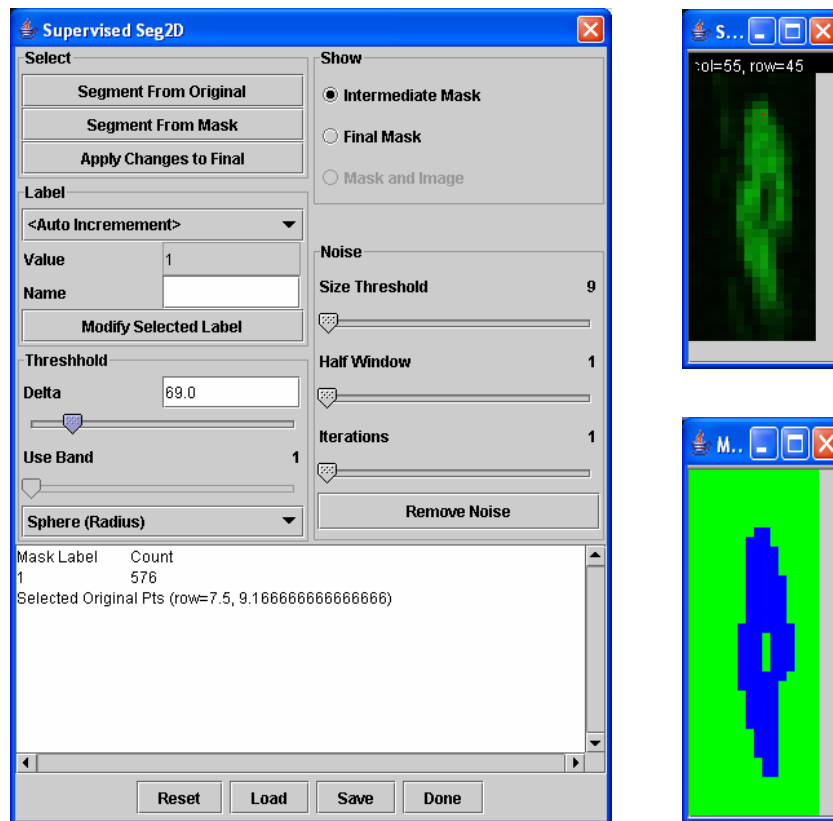

Figure 6: Users can precisely control the process of creating a segment-type image annotation by appropriately changing various variables (left). The image in the top right is the image sub-area selected and the image in the bottom right is the resulting segment-type image annotation.

2.1.2 Textual Descriptions

As described in the previous subsection, double-clicking on the image drawing (region of interest) brings up the Annotation Dialog (see Figure 4) where textual descriptions can be entered, updated and deleted. The textual descriptions typed in the Annotation Dialog will be

associated with the region of interest in the Image Viewer when the "Apply" button is clicked in the Annotation Dialog. One of the possible uses of the textual descriptions is to bridge a semantic gap, the loss of information from an actual image to a representation by low level image features [2]. The present system could help people address this semantic gap issue by providing a means to use high-level textual description, text associated with an image annotation, when they need to refer to an interesting image sub-area in the image under inspection.

2.1.3 Searching Through Static Measurements

2.1.3.1 <u>Search Based on Image Content</u>

In general, image analyses include a broad variety of scientific challenges, such as, image calibration, registration, fusion, segmentation, clustering, feature extraction, feature selection, or feature classification. As an analysis example, the prototype software allows users to create new annotations from an already existing one, which is called "propagation of annotation". The propagation of annotation can be accomplished with the following steps. First, users double-click on any existing annotation that would be used as a template for finding other similar image areas. Next, a user specifies the number of similar image areas he or she would like to find. The number is entered in the "Enter the number of matches" text field with selection of the "By Template" search method in the Annotation Dialog, as shown in Figure 4. After pressing the "Search" button, the developed system finds similar image areas by calculating normalized correlations between the template (the image annotation selected in the Image Viewer component) and other parts of the image. When the search is done, the Image Viewer component displays all matching areas found (see Figure 7) and activates the "Annotation" button in the Annotation Dialog. Finally, when a user accepts the search results with a click on the "Annotation" button in the Annotation Dialog, new annotations are created with the same textual descriptions as the template annotation object.
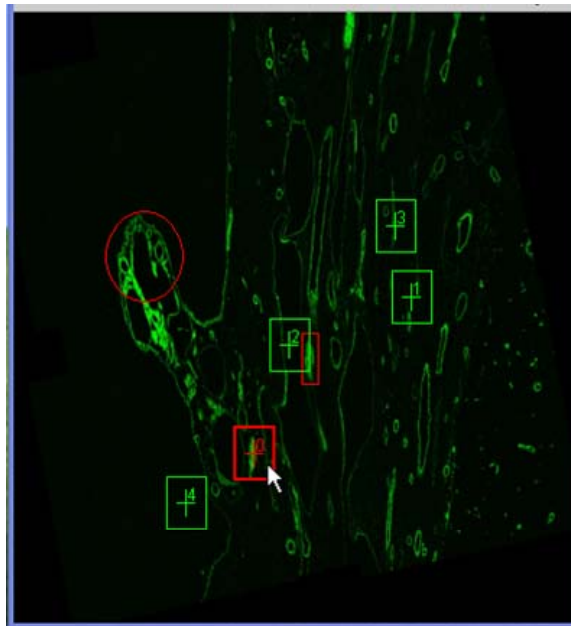

Figure 7: An outcome of a template search. The rectangles are areas that a template-based search found similar to the template used in image characteristics.

The current implementation of the prototype software only uses normalized correlation to find similar image parts. However, other image matching techniques could be prototyped based on different matching approaches. For example, the prototype software is already capable of extracting a diverse set of image features including color, shape, edge, texture, and wavelet coefficients that could be used for any needed image analysis.

2.1.3.2 Search Based on Textual Descriptions

While the "By Template" search capability allows users to search similar image sub-areas and create new annotations based on image characteristics, the "By Text" search option[2] enables finding all existing textual annotations that match a given textual query. In the Annotation Dialog, a user can specify text string in the "Enter keywords" text field (see Figure 8). Then the developed software searches through textual descriptions of all existing annotation objects and returns the ones which textual description includes the requested query. The Image Viewer component then shows the matched annotation objects in yellow color (see Figure 8).
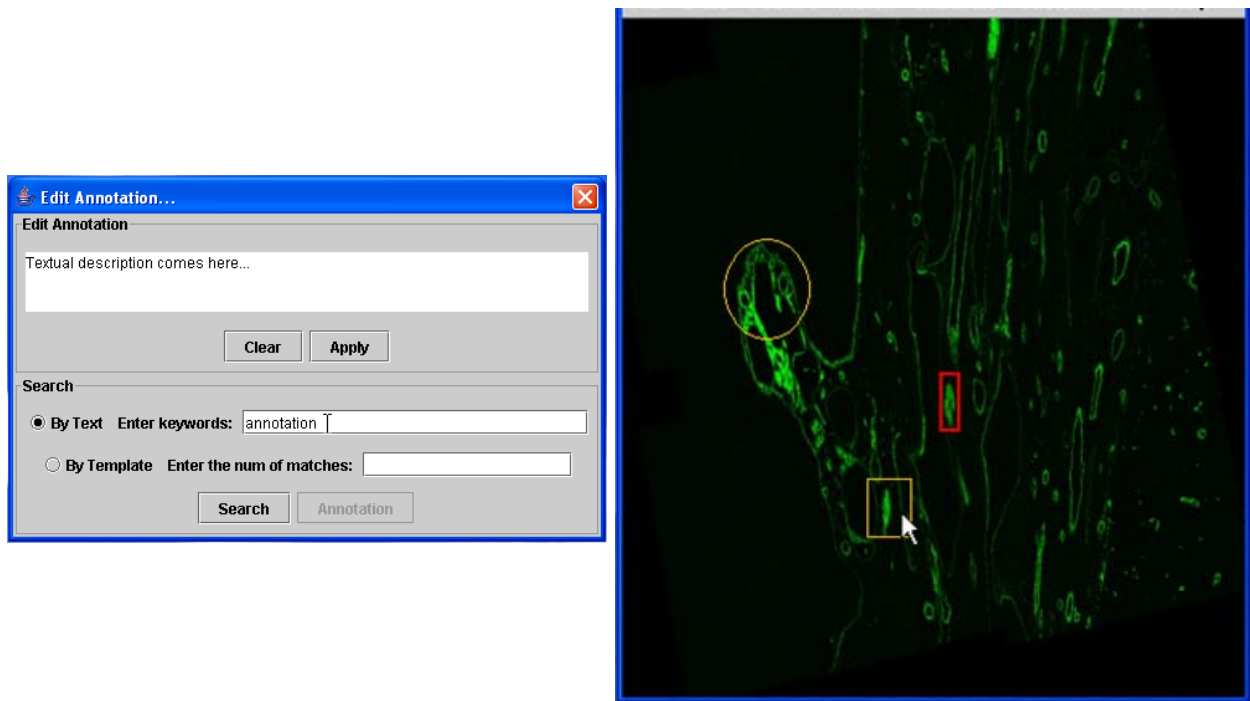

Figure 8: Requesting text-based search (left) and the outcome of text-based search.

2.2 Dynamic Measurements

We hypothesize that dynamic measurements, such as speed of mouse movements, zoom-level changes, or frame inspection orders, 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 image sub-area to the amount of time people spend to examine that area. The present system is able to monitor all image

---

[2] The Annotation Tool based on the I2K version 2, which is available at i2kv2 CVS branch, can perform the text-based search against a set of images packaged into a HDF file **Error! Reference source not found.**].

inspection activities of users, and produce dynamic computer-centric information. Upon clicking the "Start Recording" button in the "Record Dynamic Annotation" tab of the Control Panel, all image inspection activities become recorded until the "Stop Recording" is clicked, and the "Save Record" button creates a text file containing dynamic measurements (see Figure 9). Furthermore, with the "Play Record" button, users can playback any previously saved user activities. Thus, novices can watch how experts reach their final diagnosis, and hence better observe the medical image inspection procedures performed by experts.
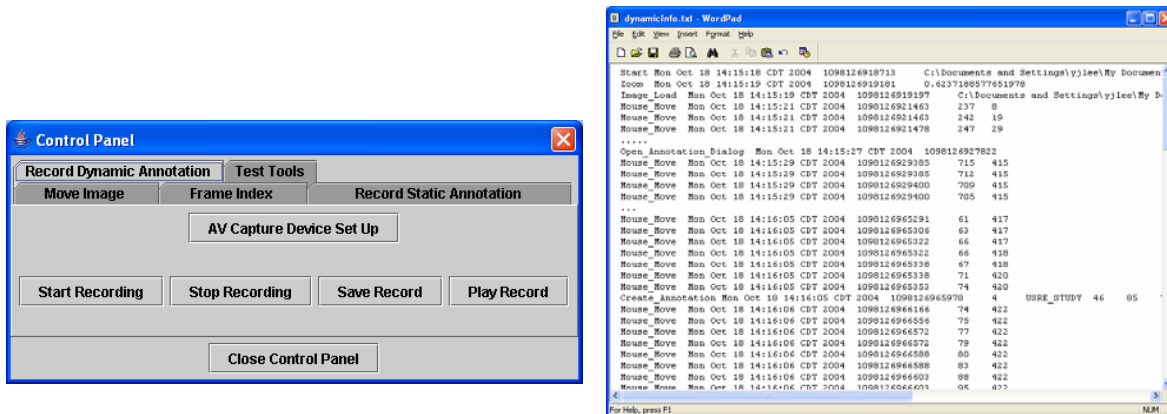

Figure 9: The "Record Dynamic Annotation" tab of the Control Panel (left) and the dynamic measurements exported into a text file (right).

2.2.1 Searching Through Dynamic Measurements

It is planned to provide a search capability that can be used to find a specific instance of dynamic measurements, such as finding all image inspection activities performed in the zoom level of 50% or finding all time instances ±2 seconds when a mouse was in proximity of a particular row and column of the image under inspection. Currently we have been focusing on the use of speed of mouse movements and changes in zoom-in/out levels because it is expected that mouse movements of users would become slower and zoom-in/out activities would be performed near the meaningful image sub-areas. Figure 10 shows a preliminary analysis of the relationship between mouse speed and complexity of image areas under investigation. As illustrated in Figure 10, 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. Since those areas can be considered to contain enough of meaningful image characteristics to attract the attention of image inspectors, but are turned out to be non-significant, we could find critical image characteristics by comparing the image features of such areas to other image areas where expert pathologists do create annotations.
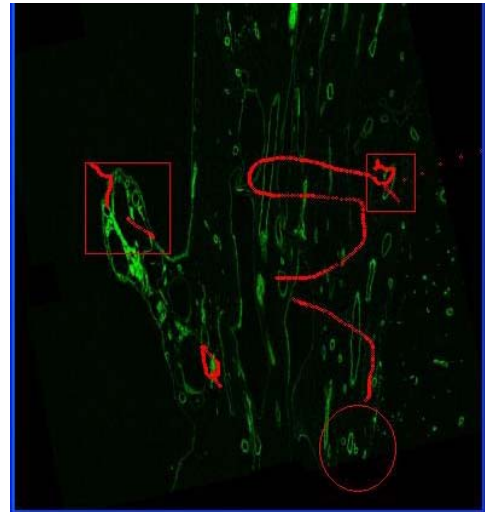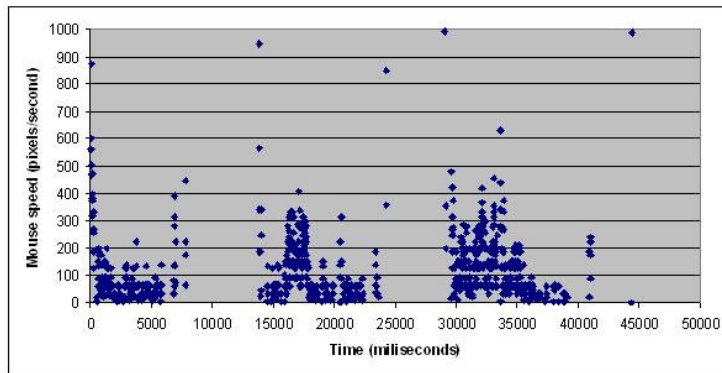
Figure 10: Temporal illustration of mouse movements (speed variable) during an image inspection (left). Spatial visualization of the mouse movements that had smaller velocity than 50 pixels per second (right – red color). The image sub-areas surrounding slow-motion mouse movements indicate image diagnostic application significance and complexity.

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, the changes in the pupil size of users while they are inspecting images can indicate that users have noticed something meaningful even though no image annotations or textual annotations are created. Similarly, the voice of users could be used to estimate the difficulties they are experiencing during image inspections. 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 [4]. As illustrated in Figure 11, the "AV Capture Device Setup" button in the "Record Dynamic Annotation" tab of the Control Panel 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 11).

At the current implementation of the system, human-centered information is not used after it has been collected. However, we anticipate that human-centered information could be used in the following ways. First, audio information could be embedded into image annotations and textual 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-activated search would be another example use of human-centric information. Second, human-centered information can be used to verify the validity of dynamic computer-centric information in estimating significance or complexity of image sub-areas under inspection. Eye movement researchers have shown that fixations in eye movements are strongly linked to intense cognitive processing. According to Viviani [5], 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.
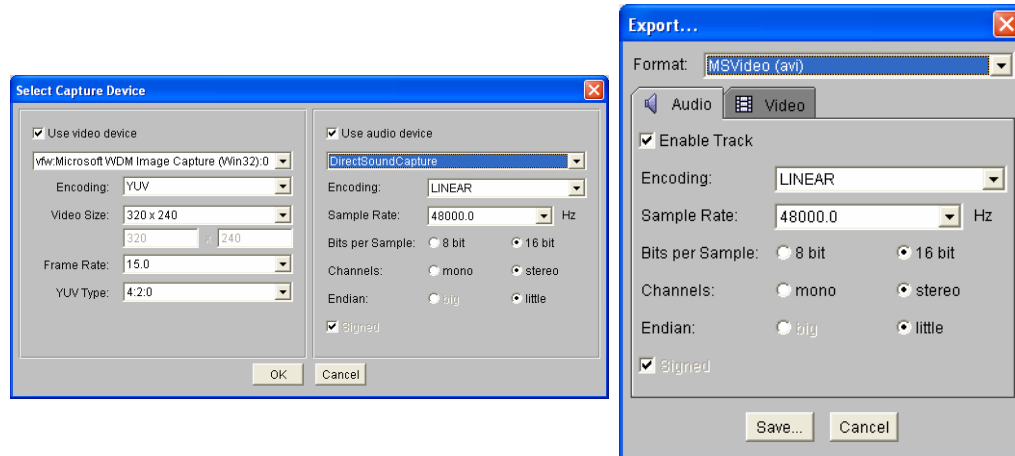

Figure 11: Users can choose appropriate audio and/or video source to capture human-centric information (left) and save the capture information into a various audio/video formats.

3.1   Searching Through Video and Audio Measurements

At the currently implementation, no search mechanism is provided for the captured video and audio measurement and it is left for future development.

4   Conclusion

This technical note describes an information gathering system capable of collecting various computer-centric information and human-centric information. Computer-centric information includes static information such as image annotations created on meaningful image sub-areas and textual annotations associated with those image annotations, and dynamic information such as speed of mouse movements, zoom level changes, and image panning and frame selections. Human-centric information is represented as voice and facial expressions of users while they are inspecting images. We envision that the present system can be used as a computer-based learning environment for medical novices such as medical students. Static computer-centric information that experts have created while they were examining the images will enable medical novices to easily recognize important image sub-areas that the medical experts have identified while dynamic computer-centric information will help the medical novices see how experts reach their final diagnosis. In addition to the educational uses, it is anticipated that the present system can be extended into an image retrieval system that can bridge semantic gaps, the loss of information from an actual image to a representation by low level image features, by allowing people to assign textual descriptions to images and to search images by the textual descriptions associated with them or low level image features. Another important assertion made in this work was that dynamic computer-centric information such as mouse movements and zoom level changes during image inspections can be used in discriminating import image sub-areas from background images. Further research with careful user studies will be conducted to verify this assertion in the future.

References

[1] Image to Knowledge, Retrieved November 15, 2004 from http://i2k.ncsa.uiuc.edu

[2] 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.

[3] Hierarchical Data Format, Retrieved November 15, 2004, from http://hdf.ncsa.uiuc.edu

[4] Java Media Framework API Guide, Retrieved November 15, 2004, from http://java.sun.com/products/java-media/jmf/2.1.1/guide/index.html

[5] P. Viviani, *Eye Movements and Their Role in Visual and Cognitive Processes*, Chapter 8, Elsevier Science, Amsterdam, 1990.