Visual Interaction in Diagnostic Radiology

E. Rogers , R.C. Arkin , M. Baron

Abstract

The concept of Visual Interaction is introduced as the process which links perception and problem-solving such that problem-solving is affected by what is seen, and conversely, what is seen and perceived is affected by the current state of the problem-solving process. This paper describes the development of a cognitively-based model of the visual interaction process in diagnostic radiology, and shows how aspects of this model are being incorporated into the design and implementation of an intelligent computer-based radiological assistant. In order to achieve this, it is necessary to extract information about the nature and type of knowledge involved in this process, and then to determine how that knowledge is used to accomplish the task of radiological diagnosis.

1 Introduction

With the advent of powerful new medical imaging technologies the effective use of computer capabilities in a clinical setting is becoming increasingly important. Traditionally, the study of radiological diagnosis and efforts to provide computer-aided assistance in this task, have focussed on two primary areas of consideration, namely, 1) The visual representation of information about the human body that resides in data acquired with current imaging technologies, and 2) The nature of diagnostic problem-solving with respect to goals, strategies, effective use of information and evidence, reasoning techniques, etc.

However, there is a third, equally important consideration which is the relationship between perception of the visual representation and problem-solving. Knowledge about this relationship provides information about the occurrence of perceptual events in the course of problem-solving activities, and suggests that perceptual assistance in the form of image enhancements may be a useful supplement to the radiologist's own abilities. The term Visual Interaction is used to denote the process which links perception and problem-solving, and it is this process which is the object of the cognitive modeling effort.

Historically, a great deal of artificial intelligence (AI) research in computer vision and image understanding, as well as problem-solving paradigms, has been devoted to trying to have a machine demonstrate aspects of human-like reasoning capabilities with little or no human intervention. A different approach, which has been adopted for this work, is to maintain the human "in the loop". That is, the purpose of the research is to design a system which facilitates and stimulates human reasoning capabilities by providing "intelligent" assistance. Such intelligence consists of knowing what type of assistance is needed, and when it may be cognitively optimal to afford it, and it is from the cognitive model that this information is obtained.

2 Related Work

A brief overview of related research provides a context for our work. Current radiological workstations have impressive capabilities to process images, and to aid decision-making, and with present trends towards the digitization of radiology departments (e.g., [14]), it is becoming increasingly important to develop clinically viable computer systems. Some examples of systems under development include: ICON [17] - a computer-based expert system which prompts for radiological findings, and then discusses evidence for diagnoses with the radiologist; AXON [2], a prototype knowledge base of strategies to guide the intelligent retrieval of radiologic studies; a visual feedback system to improve lesion detection, based on eye fixations [11], and various biomedical image processing and analysis software packages (e.g., [5]). However the diversity of these examples also
highlights a gap in current technology. There is clearly a need for a system which integrates image processing capabilities and decision-making assistance. In addition, particularly in the area of image processing and analysis, there are further drawbacks. First, the radiologist must know about the effects of various image enhancement techniques, in order to choose correctly or effectively. Second, many of these systems are time-consuming to use, since effective use is often a matter of trial-and-error. These two points alone are enough to preclude the use of many of these systems in a clinical setting where time and efficiency are crucial. However, a third point is that techniques selected by radiologists to improve the appearance of images do not necessarily lead to improved diagnostic performance.

Research in related areas such as image enhancement and automatic detection (e.g., [16]), models of visual image understanding (e.g., [8]), cognitive models of diagnosis (e.g., [1]), and expert/system differences (e.g., [9]) provides a foundation for the work described in this paper. However, the novelty of our approach lies in the attempt to bring together both perceptual and cognitive information in a common display medium in order to improve diagnostic performance in a range of experience levels.

3 Project Review

The collection and analysis of the visual protocol data is linked with a joint research project between Georgia Institute of Technology and Emory University School of Medicine. It has been described to explore the nature of visual radiological diagnosis and has as its main objectives, 1) the development of an understanding of how radiologists interpret medical images in the process of solving diagnostic problems, and 2) the design and implementation of knowledge-based visualization techniques in order to improve diagnostic accuracy and throughput. The first two phases of the project, called the Observation Phase and the Intrusive Phase, respectively, are of direct interest to the work under discussion, since they concern the collection and analysis of data on how radiologists interact with images, both perceptually and cognitively.

The Observation Phase, reported in [13], involved both informal observations of radiologists during their routine evaluations in the chest x-ray reading room, and a more formal set of experiments. In the latter case, visual protocol data was collected from ten radiologists (ranging in experience from first year resident to expert), in the task of examining chest x-rays displayed on film and on a computer screen (CRT). The results of this work provided fundamental guidelines for the Intrusive Phase data collection, where the subjects were asked to provide extensive “talking aloud” reports while examining computer-displayed chest x-ray images. The collection and analysis of those extensive verbal reports constitute the basis for the development of the cognitive model, and are described in greater detail in the following sections.

4 Data Collection

In this phase, the investigator assumed a more active role in eliciting information from the subjects while they looked at images, and this information consisted of a variety of visual protocol data obtained during the execution of the radiological diagnostic task. A total of eight subjects from the original group of ten, described for the earlier Observation Phase in [13], took part in the experiment. From our collection of digitized chest x-rays, seven new cases were selected, representing a number of abnormalities in a range of difficulty, including: lung lesion due to bronchogenic carcinoma, hilar adenopathy, tuberculosis, lung mass with appearance of elevated diaphragm, and mitral valve disease. One normal chest x-ray was also included. Three of the seven images had a brief case history associated with them and the images were presented on the same Sun/Parc computer equipment described in [13], in the same unenhanced format.

For each subject, there were three parts to the experiment. First, each case was examined and the subject was instructed to “talk aloud” during this examination, and then to dictate findings and impressions as would normally be done for an x-ray case. The second part of the experiment required the investigator to pose specific questions after each case that clarified or extended comments made by the subject during the talking aloud segment. The importance of significant image features, landmarks, image quality and other characteristics mentioned by the subject were questioned, as well as details of size estimates. Then the subject was asked to rate the case with respect to three scales: image quality, confidence in diagnosis, and case difficulty. Finally, after all seven images for the session were read, another set of questions was presented, which prompted
5.1 Diagnosis

5.1.1 Encoding

followed by the development of a formal encoding scheme to capture the concepts of the verbal reports. Next, the *Statement Analysis* involves the application of the encoding scheme to the actual talking aloud reports produced by the subjects, and finally, under the heading of *Contextual Analysis*, the encoded data is re-examined in order to identify larger units of behavior over time, incorporating both perceptual and problem-solving components.

5 Data Analysis

The raw data for this part of the study consists of videotapes and transcripts of each subject's experimental session, and a three-part methodology for analyzing the data has been adopted. The first, called *Task Analysis*, consists of a functional study of the task of radiological diagnosis, followed by the development of a formal encoding scheme to capture the concepts of the verbal reports. Next, the *Statement Analysis* involves the application of the encoding scheme to the actual talking aloud reports produced by the subjects, and finally, under the heading of *Contextual Analysis*, the encoded data is re-examined in order to identify larger units of behavior over time, incorporating both perceptual and problem-solving components.

5.1 Task Analysis

The purpose of this stage is to demonstrate a characterization of the task of radiological diagnosis that can serve as a framework for the formal encoding of the verbal protocols. This is done by combining results of experimental observations (both our own and from the literature), with specific information extracted from a subset of the the verbal protocols themselves, in keeping with the recommendations of Ericsson and Simon [5].

From the observational data gathered in the x-ray reading room, a simple preliminary description of the actual task of radiological diagnosis can be expressed in information-processing terms: *input → visual interaction → output*, where *input* includes x-ray images, case history and viewing tools, while *output* includes findings, impression, and diagnosis. An initial decomposition of functions of the visual interaction “black box”, taken from Lengold et al [8], provides the following functions: *detection* of film abnormality, *description* of abnormality feature, *localization* within anatomy, *medical explanation*, and *resolution* of overall case. This initial taxonomy is then further refined in light of specific results from the observational phase research [13], and more general problem-solving results from the related literature (e.g., [4], [12], [15]), to produce the expanded taxonomy of Figure 1. The first three categories can be said to be more perceptually-oriented functions since they are related to information obtained from the stimulus, while the last two categories involve primarily problem-solving activities.

This taxonomy provides an outline of the types of general concepts that are needed to encode the verbal protocols. However, examination of a subset of the actual protocols reveals that a greater level of detail is needed in order to account for the different types of statements made by the subjects. Furthermore, in order to develop a concept categorization that is relatively free of specific medical terminology, a preliminary encoding of the medical concepts is needed, together with a lexicon of the terms used. This initial class of concepts contains three categories: *Anatomy, Finding, and Diagnosis*. These categories allow the prior extraction of medical terminology from the protocol statements in such a way that someone performing the actual encoding does not need to have a medical background in order to apply the concepts. An example

<table>
<thead>
<tr>
<th>Perceptual Examples</th>
<th>Problem-Solving Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>DETECTION</td>
<td>MED. EXPLANATION</td>
</tr>
<tr>
<td>DESCRIPTION</td>
<td>RESOLUTION</td>
</tr>
<tr>
<td>LOCALIZATION</td>
<td></td>
</tr>
<tr>
<td>Spatial</td>
<td>Hypothesis</td>
</tr>
<tr>
<td>Abnormal Objects</td>
<td>Formulation</td>
</tr>
<tr>
<td>Properties</td>
<td>Hypothesis</td>
</tr>
<tr>
<td>Feature</td>
<td>Testing</td>
</tr>
<tr>
<td>Properties</td>
<td></td>
</tr>
<tr>
<td>Classifications</td>
<td>Induction</td>
</tr>
<tr>
<td>Out-of-Plane</td>
<td>Rationalization</td>
</tr>
<tr>
<td>Relationships</td>
<td></td>
</tr>
</tbody>
</table>

Figure 1: Extended Taxonomy of Radiological Task.
of how these categories are utilized is presented in the following section.

The remaining two classes, called Cognitive Concepts and Descriptive Concepts, respectively, contain the concept categories that constitute the formal encoding scheme for the protocols. Both of these classes capture perceptual and problem-solving elements that are needed to describe the task of radiological diagnosis, and are illustrated in Figure 2.

5.2 Statement Analysis

The methodology for the application of the encoding scheme to the protocol data is now presented. The first part of the data consists of the talking aloud reporting provided by the subject while examining the chest x-ray image. Each protocol is parsed into simple phrases or statements, and then the words that can be classified as either Anatomy, Finding, or Diagnosis are extracted and listed in the appropriate columns next to the statement. Figure 3 shows an example of this preliminary procedure.

In order to obtain some measure of reliability of the encoding process, an independent observer was recruited from the Psychology Department at Georgia Tech to encode a randomly selected test protocol case for each of the eight subjects. After an initial 90-minute training period, the observer and the investigator proceeded to independently encode the protocols at the same time. Both sets of protocols were then compared and scored according to whether the same concepts appeared in each category for each statement. The results of this procedure showed an overall similarity between the encodings of 89%, with a low score of 83% for the first subject’s protocol and a high score of 95%. In a breakdown of individual categories, the greatest discrepancies occurred under the Mental concepts category, and further analysis revealed that stricter rules governing the way this category is encoded must be enforced. However, the overall results were encouraging enough to proceed with the encoding of the remaining protocols, and an example of such encoding is presented in Figure 3.

The remaining protocol analysis is organized according to the type of abnormality present, and the first type of case to be considered is that of the malignant tumor. The corresponding talking aloud reports and dictations have been encoded for all eight subjects, and this information is then used as input to the subsequent Contextual Analysis stage, described in the following section.
5.3 Contextual Analysis

This stage of analysis is where a certain amount of context, e.g., task-related, time-related and/or experience-related is introduced into the data analysis. It is the next level of abstraction away from the original data, and provides some insights into the type of larger units of activity that may occur during execution of the task.

The first type of context to be considered is task-related, and three main classes of activities are identified: 1) those related to the primary abnormality, 2) those related to secondary abnormalities, and 3) those related to remaining anatomical objects and general comments. Under these three headings, each subject’s actions are listed in terms of the concept codes, in the same order as they appear in the verbal report. An example of this is demonstrated in Figure 4.

This re-organization of the concepts indicates that there may be some definite patterns that occur in the process of radiological diagnosis. This is not entirely surprising since part of the early training of radiologists includes suggestions for developing a methodological way of examining x-rays. What is interesting to note is that, as a result of the encoding, the composition of these patterns starts to emerge, particularly with respect to perceptual activities, such as looking at and describing anatomical objects, compared to problem-solving activities, such as hypothesis formation. This leads to a number of further questions: Are hypotheses generated and then tested by obtaining perceptual evidence, or is evidence gathered first, followed by a hypothesis? How often is prior medical knowledge or reasoning from memory used to make a decision, compared to utilization of perceptual information from the image? Some of these questions may also be experience-related, or abnormality-related, and as the contextual analysis proceeds with the remaining cases, it is expected that some answers will become evident. However, a certain amount of preliminary information is already available from this single tumor case.

6 Preliminary Results

The findings that are described in this section have to do with concepts that arise directly from the encoding of the verbal protocol reports. There are no inferences made at this time about cognitive activities that may be occurring, which are not reported by the subjects. What is of interest is that nonetheless, certain trends emerge across the range of expertise.

One example of such a trend is the two-concept combination of LOOK-AT(primary finding) followed by DESCRIBE(primary finding), which occurs early in the examination of the case. In fact, for six of the eight subjects, this combination is the very first activity that occurs. The remaining two subjects had this combination as the second activity: one subject looked briefly at a normal object first, while the other commented on image quality first. This result suggests that for this type of abnormality (i.e., lung mass), the attention of the radiologist is initially captured by the finding itself. Further, an attempt is made to describe this object. Size, shape, location, edges and texture appear to be the most common descriptive concepts for this type of abnormality. Since different subjects utilize different combinations of descriptive features, further study is needed to determine if all of these are needed to make an accurate diagnosis, or if a subset is sufficient.
However, from this preliminary evidence arise two predictions for this type of abnormality: 1) Image enhancements should perhaps be implemented at the early stages of this type of case, as soon as a mass is identified, and 2) Algorithms which enhance size, shape, location, edge and texture assessments may be the most useful at this stage.

7 Preliminary Modelling

The construction of the preliminary model of visual interaction in radiological diagnosis proceeds from a reconsideration of the four key issues of focus-of-attention, dimensionality, context, and expectation [13], in light of the preliminary results from the data analysis. It appears that for this one type of abnormality, there is an immediate perceptual impression that gives rise to further examination of the abnormality with respect to certain dimensions. This immediate perceptual impression may be said to trigger a context that is affiliated with this type of abnormality. This context may include information about the different types of descriptive concepts that may be relevant to the primary finding (i.e., dimensionality), types of secondary findings that need to be considered, and also expectations about possible hypotheses of what this abnormality could be. It seems likely that affiliated with each hypothesis is some list of features and findings that most contribute to the confidence about that hypothesis. Thus Figure 5 gives an example of what such a context might look like and shows how a hypothesis of cancer, for example, may require a certain size and edge description, together with the presence of bone lesions in order to be activated.

Focus-of-attention can be considered the mechanism by which further perceptual information is obtained. The control of this focus-of-attention is dependent upon the type of strategy that is used to solve the problem. One strategy is to gather perceptual information first, and then to consider appropriate hypotheses according to this evidence. There is some evidence for this "bottom-up" type of strategy in our example case. Another possible strategy is to choose one or more likely hypotheses first, and then only gather the evidence that is required to make decisions about those hypotheses. One of the questions to be answered is whether this latter "top-down" type of strategy can be identified in some of the other cases.

A further important consideration is the correlation between certain kinds of perceptual information and certain hypotheses. This type of information can assist in determining whether providing certain kinds of perceptual assistance will lead to better performance in choosing the most likely hypotheses. Present theories of high-level vision (e.g., [7]) and problem-solving (e.g., [10]) appear to be consistent with our modelling approach, and as the model is refined, it will be re-evaluated with respect to current research. The next step of the modelling process is to examine the reports of further cases, and to refine the concepts presented here in light of those results.

8 Design of Prototype System

The initial design of the prototype computer system that will allow us to test and evaluate the predictive capabilities of the model is currently underway. Issues to be considered include: 1) Design and implementation of knowledge control and representation for a number of different abnormalities, 2) Development of an easy-to-use interactive mouse-driven user interface, and 3) Appropriate choice of image enhancement algorithms.

The particular diagnostic problem-solving activities indicated by the model may be represented by frame-like structures, which, for the purposes of this example, are called "problem-solving scripts". For instance, if the radiologist's initial impression is that there's a lung mass, our preliminary results indicate that there are certain dimensions that must be considered, such as size,
Figure 6: Example of Mass Script.

Texture, edges, etc., in order to make a diagnosis. Figure 6 shows an example of what a “Lung Mass Script” might look like. It is triggered if, for example, the radiologist has indicated a density with well-defined edges or a round shape. Statements labelled “Request” indicate dialogue between the system and the radiologist, where information is obtained from the radiologist, and is then posted to a session summary area. Statements such as “Add-Ruler-To-Image”, or “Apply-Texture-Enhancement” indicate automatic enhancements activated by the system during the course of this script.

The user interface is planned to utilize the two screens of the Sun/Pixar computer system, such that images and icons appear on the Pixar monitor, while dialogue and session summary information are presented on the Sun monitor. When an icon is activated by clicking with a mouse, the corresponding dialogue and script are activated. The script then takes over and tries to obtain relevant descriptive information, while automatically applying appropriate enhancements to the image. A number of specialized image processing algorithms will be utilized to perform the enhancements dictated by the scripts.

9 Conclusion

This work is potentially significant in a number of areas. It may provide new directions for clinically useful interactive radiological systems. It is also seen as a useful radiological teaching tool, providing “hands-on” experience with a clinical aid, and further, it may prove to be an effective tool for studying the radiological process itself. More generally, the methodology employed for developing the cognitively-based model of visual interaction may be applicable to a number of different domains, where visual perception is an integral part of the problem-solving process.
Acknowledgements

This research was supported by the Emory/Georgia Tech Biotechnology Research Center, and by a Georgia Tech Interdisciplinary Seed Grant. In addition, the authors would like to thank the following people for their assistance to this project: Dr. Norberto Esquerra, Office of Interdisciplinary Programs, Georgia Tech; Dr. Ernest Garcia, Dept. of Radiology, Emory University Hospital; Dr. John Pari, Dept. of Psychology, Emory University, Ma.; Debbie Mitchell, Dept. of Psychology, Georgia Tech.

References