Visual Interaction: A Link Between Perception and Problem Solving

Erika Rogers and Ronald C. Arkin

Abstract—The approach taken in this research is to develop a cognitive model of how a human observer extracts information from a visual display and then uses this perceptual information in a decision-making task. 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 is a useful supplement to the user's own abilities. This knowledge is then to be embedded in an intelligent computer-aided assistant which is designed to facilitate and stimulate the human problem-solving process.

I. INTRODUCTION

With the advent of powerful new technologies for displaying multi-dimensional scientific data, the development of new strategies for efficient use of these capabilities is of foremost concern. It is becoming more and more evident that in order to perform complex tasks, man and machine can no longer be treated as separate entities, but must be considered together as a unified decision-making system. The powerful computational resources of computer technology must be coupled with powerful human perceptual and problem-solving capabilities in order to achieve this goal.

Efforts to apply artificial intelligence techniques in pursuit of this goal have resulted in a number of different approaches. One direction has been the development of automated reasoning systems, where conceptually, the “faulty” aspects of human decision-making are replaced with more mathematically precise components. At one time, it was thought that such systems might one day replace humans altogether. However, more recently, there has been a shift in emphasis, especially in light of the fact that these autonomous intelligent systems have not really been as successful as hoped in complex task domains. Rubin et al [17] note that “knowledge-based computer control systems are unable to effectively handle degraded information or novel situations with ambiguous conditions in which previous guidance of an explicit nature is not available”. The more contemporary trend is to consider how to use automation to enhance the effectiveness of the human performing a complex task, to build “tools instead of prostheses” [3,5]. Holtzman [6] states that “effective decision systems must concentrate on assisting the decision-maker to gain insight into the decision problem at hand rather than on merely supplying a somehow ‘right’ answer”. It is also stressed that the way in which this is to be achieved must involve an understanding of the cognitive abilities of the human user in order to implement a “cognitive cooperation” [3] between the human and the computer system. With respect to these stated philosophies, the research presented in this paper is in complete agreement.

However, in practice, the implementations of these ideas have concentrated primarily on the development of expert-like systems, both in medicine and industry. Such systems usually achieve their cognitive plausibility by extracting domain-specific knowledge from one or more experts in an area, and then reformulating this knowledge in a rule-based format. There are three main drawbacks of such systems: 1) the knowledge extracted tends to be shallow; 2) experts can also be fallible, and therefore a range of experience would be more desirable for our modeling effort; and 3) expert systems tend to rely on verbal representations of knowledge, and few are designed to accommodate “image” or “spatial reasoning” that is needed for task domains where the decision-making process must rely on interpretation of a visual image. Therefore, our approach is to first develop a detailed cognitive model of the human capabilities in our area of interest. This model is then used to predict the type of assistance that will be most useful to the user at various stages of the problem solving or decision-making process.

Systems which do address image reasoning issues may be found primarily in domains that contain sensor-derived data. Examples of such systems include AXON - a computer-based intelligent assistant for retrieval of radiographic studies [2], Intelligent Atlas - an expert system for neuroanatomy diagnosis, which presents the inference process with both verbal expressions and image presentations from a pictorial database [12], and research by Kraiss on an intelligent dual-screen workstation for feature extraction and interpretation of sonar data [7]. It is interesting to note that although these systems present
perceptual material in the form of images, and perhaps even allow perceptual enhancements in the form of image processing menu choices, the primary "intelligence" still focuses on reasoning capabilities. The perceptual capabilities of the user have not been addressed, particularly with respect to effects on performance. The choices of image enhancements are left up to the user, and yet there is no guarantee that improvement in appearance will lead to improvement in performance. Our own approach to this problem is best expressed in the following working hypotheses:

1. The application of cognitive science methods leads to an understanding of the close interaction between the user's perceptual processes and his problem solving capabilities in domains where computer-displayed images form an integral part of the problem solving process.

2. The resultant cognitive models can be embedded in interactive, cooperative computer systems, which are designed to provide intelligent assistance to the human agent. Such intelligence consists of knowing what type of assistance is needed, and when it may be cognitively effective to afford it.

3. By providing appropriate image enhancements at key nodes in the problem-solving process, the human user's own perceptual abilities are enhanced, and overall performance in the task is improved.

The emphasis of this paper is on the exploration of the first of these hypotheses and presents the components of the cognitive model which has been developed through experiment and analysis.

II. PROJECT BACKGROUND

Our preliminary work [15] has introduced the concept of Visual Interaction 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. Furthermore, the early phases of development of a cognitively-based model of the visual interaction process have been described for the domain of diagnostic radiology. This development domain was chosen because the specialists are trained to extract information from data that already has a known visual representation. That is, the numerical values produced by the various imaging technologies are presented in the form of images of human anatomy which are particularly meaningful to medically-trained personnel. Investigation of visual interaction in the context of such "visual familiarity" facilitates our task considerably. Furthermore, by studying an environment in which the natural layout entails a static stimulus (i.e., the image) and a relatively stationary observer, we constrain our problem, while still obtaining realistic data. That is, we circumvent the accusation, endemic to much AI research, of creating completely artificial laboratory conditions that have no connection with the real world.

The collection of both observational data and concurrent protocol data was organized in order to develop a thorough understanding of the task environment, and the task requirements. Details of the experimental data collection and preliminary results are reported in [15,16]. An overview of the Data Analysis Methodology developed to extract perceptual and problem solving concepts from the verbal protocols is presented in Figure 1. The details of the three main stages, namely, Task Analysis, Full Protocol Analysis, and Contextual Analysis are presented in [15]. In this paper, we present more concrete results from the Contextual Analysis stage, which have had a direct impact on our modelling process.

III. RESULTS OF ANALYSIS

By partitioning the protocol statements according to task-related categories, and by maintaining the temporal order of these statements, we see clusters of activity which are common to the majority of subjects for each case. Many of these clusters can be classified as perceptual or problem solving events, and they provide clues to the understanding of the visual interaction process.

![Figure 1: Data Analysis Methodology](image-url)
The first type of visual interaction behavior that we noticed is what we have termed "immediate visual capture". That is, certain types of findings in the image (e.g., lung mass, hilar adenopathy) seem to attract immediate attention, and, for most of the subjects, are mentioned as soon as the verbal reporting begins. These findings are then described with respect to features such as size, shape, location, texture, and edges. The findings themselves vary according to different levels of specificity, and the amount of description associated with each level also varies. These kinds of perceptual events occur very early in certain cases, and then trigger a set of one or more candidate diagnosis hypotheses.

A different type of behavior expressed by the majority of subjects in virtually all of the cases is called "deliberate landmark search". Here the subjects examine landmarks in the chest, and classify them according to whether they appear "normal" or not. The classification is fairly rapid, and appears to be almost a "check-list" type of activity. Although the order itself varies from individual to individual, the main anatomical landmarks mentioned are common to most of the subjects. If there is an immediate visual capture of an abnormality, then the deliberate landmark search takes place later in the protocol. It is sometimes interrupted by changes in focus of attention, and then resumed at a later stage. If the case presents a normal chest, or one in which there is no immediately evident abnormality, then the subjects usually begin with the deliberate landmark search, and continue until either an abnormality is detected, or they are satisfied that there is none.

Further results show that often secondary evidence and/or case history information is needed to disambiguate between diagnostic hypotheses. Sometimes incorrect anatomical localization (e.g., mediastinal tumor vs. lung tumor) can lead to an incorrect diagnostic hypothesis. However, there is also evidence that sometimes perceptual evidence in the form of combinations of critical features (e.g., size and shape) can override the incorrect diagnostic hypothesis, and lead to a correct one.

The particular patterns of activity found in our data, (such as deliberate landmark search, anatomical localization, gathering of secondary evidence, etc.), imply a certain amount of goal-oriented behavior. The use of such plans may be driven by a strategy of collecting general feature information about a finding, without any commitment to a particular hypothesis, or it may reflect a strategy of collecting particular evidence for a particular hypothesis. Furthermore these strategies may vary depending on the experience of the subject, and the nature of the case under consideration. It is also possible for a particular subject to use a number of strategies during the course of a single case. This implies that, although there may be a plan-like structure, it must be flexible enough to allow changes in strategy to occur.

These results have a number of important implications for the cognitive model and its subsequent effects on the computerized assistant. To summarize:

Figure 2: Perception-Problem Solving Interaction

1. Important descriptive features are linked to findings in the image. This implies that guiding the user to look at these features will be a useful enhancement.
2. Findings in the image are linked to a set of candidate diagnostic hypotheses. This suggests that maintaining and displaying hypothesis information will be helpful to the user by relieving some of the cognitive loading on short term memory.
3. Hypotheses are linked to particular kinds of evidence. Therefore it will be useful to let some expectation guide the search for further perceptual evidence.
4. Feature combinations and relevant secondary findings can override incorrect hypotheses. This suggests that providing enhancements of the features of image findings and anatomical landmarks should improve the user's assessment of these features and therefore lead to improved diagnostic performance.
5. Goal-oriented behavior occurs throughout the diagnostic process. Therefore plan-like structures will be useful to guide the diagnostic strategies.

IV. DESCRIPTION OF MODEL

In the course of normal human activities, it can be shown that there must be a relationship between perception and problem solving such that perception "delivers" information about the environment to the problem solving process, and, conversely, the problem solving process communicates "directions" to the perceptual process (e.g., I need this type of information rather than that type). Moreover, we are interested in the class of problems that requires extensive interaction between perception and problem solving, where the problem input is in visual format and the task is to interpret this input in a meaningful way.

Based on current models of perception (e.g., [1,9]) and problem solving (e.g., [8,11]), the mechanisms for such two-way communication are already potentially in place, and are conceptually illustrated in Figure 2. Constraints placed on the components labelled Model and Plans are
critical to our understanding of this interaction. For example, the Model should be able to accommodate knowledge from both sides: visual information delivered by the perceptual process (e.g., percepts that describe findings in the image), and decision-related knowledge based on the current state of the problem-solving process (e.g., what hypotheses are active, what kinds of information do they need for evidence, etc.). Therefore we need a way to reconcile and combine these different types of information in the Model. In addition, there should be a way to account for different levels of Plans. For example, a plan to pursue hypothesis-directed search vs. data-driven search is at a different level of abstraction than the detailed plan for gathering the specific perceptual evidence required by a particular hypothesis. Therefore we need a mechanism that coordinates these different levels, and ensures that the plans are executed, modified or abandoned according to both current perceptual information, and the current state of the decision-making process.

We therefore hypothesize that there is a mediating process, which we call the Visual Interaction Process, which oversees the transfer of information and instructions between the perceptual cycle and the problem solver (which will henceforth be called the Perceptual Process and the Problem Solving Process, respectively). This mediating process is responsible for maintaining a Mental Model depending on current information from both of the other two processes, and also for managing the transition from high-level plans to detailed low-level plans. In addition to the primary processes, we also take into account the need for both Long Term Memory and Working Memory as parts of the cognitive architecture, and describe the components of our model in the following sections.

A. Long Term Memory

In addition to general knowledge about the world, the Long Term Memory contains domain-specific knowledge that is relevant to the problem-solving task, and persists over long periods of time (e.g., the career of an active radiologist). The data which we have collected indicates that this knowledge falls into a number of categories, which include the following:

- Different levels of findings (e.g., object findings such as "masses", landmark-related findings such as "adenopathy", etc.);
- Features associated with the findings (e.g., size, shape, texture);
- Solution hypotheses (complete or partial, e.g., differential disease diagnoses);
- Previous theoretical knowledge (e.g., medical training in anatomy and signs/symptoms of disease);
- Previous experiences (e.g., previously seen x-ray cases).

Figure 3: Detailed Model of Visual Interaction

Long Term Memory also contains some "Mental Models" used in problem solving, or at least the components required to construct these mental models. Patel et al use the following expression: "...when a doctor is presented with a problem that he or she is familiar with, an appropriate prototype would be invoked to suggest possible findings [13]". The three primary processes of our model access this knowledge at various times in both bottom-up and top-down fashion. This is in keeping with both our own observations, and those of researchers such as Reiman and Chi [14].

B. Working Memory

On the other hand, Working Memory is characterized as a mental workspace that is easily accessible, holds current information about the task (in particular, the "prototype" or "Mental Model" acquired from Long Term Memory), and is quickly updated as more or new information becomes available. As with Long Term Memory, the Working Memory of our model is also accessed by all three basic processes, and is divided into a number of components which include Percepts, Mental Model, Hypothesis Report, High-Level Plan and Detailed Plan. The entire model is illustrated in Figure 3. In describing the requirements for a cognitive architecture, Newell et al [10] discuss the need for "interfaces that connect the sensory and motor devices to the symbol system". They state that "the external world and internal symbolic world proceed asynchronously and therefore there must be a buffering of information between the two in both directions". The structures in our Working Memory can be grouped together to reflect their relationship to the three primary processes. The Perceptual Process and the Visual Interaction Process communicate via the
Perceptual Buffer, and likewise, the Visual Interaction Process and the Problem Solving Process communicate through the Problem Solving Buffer. Each Buffer contains both “data” components (i.e., Percepts and Hypothesis Report, respectively) and “control” components (i.e., Detailed Plan and High Level Plan, respectively). The following is a description of each of these components together with that of the Mental Model.

C. Percepts

We recall from our data the evidence for “immediate visual capture”. This early information may be used as an initial cue to obtain relevant domain-specific information from Long Term Memory. This is supported by findings in the literature as well. For example, Reiman and Chi mention the use of such a cue to “trigger a particular [problem] schema” [14], and, more generally, Newell et al [10] point out the need for access to distal memory structures. Later, as more information is added, or different objects (such as findings or landmarks) capture attention, there is also a need for a structure which allows transfer of this information from the Perceptual Process to the Visual Interaction Process. We say that this structure consists of one or more Percepts which constitute the output of perceptual processing, and can be described as labelled objects with some number of features (e.g., size, shape) attached. The purpose of this structure is to supply the perceptual cue(s) used by the Visual Interaction Process to initialize and update the current information that is maintained in the Mental Model during the problem-solving task.

D. Mental Model

The next observation that we make from the data is that during the problem solving task, the types of knowledge available to the subjects include: a candidate list of descriptive features associated with the finding, candidate anatomical locations, candidate diagnostic hypotheses, candidate secondary findings, and remaining anatomical landmarks. Descriptive features (or combinations of features) and anatomical locations are utilized as evidence for particular diagnostic hypotheses. On the other hand, sometimes the hypotheses are directly used to control the acquisition of particular feature/finding information. This implies a tight coupling between the perceptual type of information and the candidate hypotheses. Our data shows that both bottom-up and top-down processing occur at various times in the subjects’ protocols, and furthermore, that precedence can alternate (e.g., sometimes combinations of features can override incorrect hypotheses). This is consistent with the results of Reiman and Chi [14] and Patel et al [13].

Therefore, the Mental Model maintained in working memory contains a combination of relevant finding and hypothesis information, and reflects the relationship between current finding/feature information, and the evidence obtained and/or needed to distinguish between current candidate hypotheses. It is the task of the Visual Interaction Process to initialize and update this Mental Model based on the Percepts delivered by the Perceptual Process, and also on information dictated by the Problem Solving Process.

E. Hypothesis Report

The process of making a decision and completing the task frequently involves the evaluation of the current status of the candidate hypotheses. Frequently the subjects generate a list of candidate hypotheses, and rank them at various times in the problem solving task. Sometimes an initial ranking is expressed, which might later change according to the evidence gathered. At some point the subjects decide whether they have acquired enough information to make a final decision. This decision may be of the form of a definite, single diagnosis, a ranked list of differential diagnoses, an unranked list of differential diagnoses, or simply a list of findings. This type of reasoning does not seem to involve all the information currently gathered in the Mental Model, but rather focuses on the hypothesis-related components. This suggests that from the above Mental Model is extracted a substructure that contains a summary or report of the current status of the candidate hypotheses. This report reflects the strength of the evidence available at that stage, thus allowing the Problem Solving Process to produce a ranking, and to make a decision about whether to continue seeking further evidence, or to stop. We therefore call this the Hypothesis Report, which is delivered by the Visual Interaction Process to the Problem Solving Process.

F. High Level Plan

The Problem Solving Process, having received the Hypothesis Report from the Visual Interaction Process, must evaluate this information and determine whether there is enough to make a commitment to a decision. If not, then it must communicate directions to acquire whatever further information may be necessary to get closer to the goal. The Problem Solver therefore formulates a High-Level Plan that reflects the best strategy needed. For example, if the hypotheses can be ranked, then the plan may indicate that evidence for the first-ranked hypothesis should be acquired first, followed by that for the second, etc. On the other hand, if the hypotheses cannot be ranked due to insufficient information, the plan may indicate that more features of the primary finding should be collected, and may order some of these features according to their ability to distinguish between hypotheses. It is also possible that the Problem Solving Process may “discover” a new hypothesis (or set of hypotheses) that it wants to explore. This new requirement is reflected in the plan, together with instructions on whether to completely abandon the old hypotheses, or perhaps to maintain them in a lower priority.
G. Detailed Plan

When the Visual Interaction Process receives the High-Level Plan from the Problem Solving Process, it first uses the plan to reorganize the Mental Model to reflect the current priorities - either rankings of hypotheses, or features, or perhaps the addition or deletion of hypotheses. Then, in order to meet the goals of the High-Level Plan, detailed instructions must be passed on to the Perceptual Process. This is done in the form of a Detailed Plan, which is formulated by the Visual Interaction Process using the information in the Mental Model. This plan contains goals related to particular features of a finding (e.g., size, shape, texture, etc.), different types of findings (e.g., fluid, collapse, etc.) or different landmarks (e.g., heart, bones, etc.), and contains directions such as "LOOK-AT(lungs), "LOOK-FOR(fluid)", etc. This is in keeping with the earlier description of the perceptual cycle, which uses the model to formulate plans for action, which then interact with the outside world.

VI. CONCLUSION

The approach described in this paper has a number of advantages. It provides a methodology for developing a cognitive model of visual interaction, and then utilizes this knowledge to provide perceptual assistance in the course of the problem-solving process. Unlike automatic detection or automatic decision-making systems, the assistance is supplied at the interface between the two poles of perception and problem solving. The user must be guided through the process of assessing stimulus input, extracting relevant information, making a decision about the quality of that information, and perhaps returning again to the stimulus to obtain more evidence for the problem-solving process. Furthermore, this enables us to design a computer system which utilizes knowledge of how image processing affects perception of images to strategically implement physical image enhancements at key nodes in the user's problem-solving process. This latter work is currently in progress.

ACKNOWLEDGMENTS

The authors would like to thank the following for their assistance to this project: Dr. Murray Baron and Dr. Ernest Garcia, Dept of Radiology, Emory University Hospital; Dr. Norberto Esquerra, Office of Interdisciplinary Programs, Georgia Tech; Dr. John Pani, Dept. of Psychology, Emory University.

REFERENCES