Visual Queries: The Foundation of Visual Thinking

Colin Ware

Data Visualization Research Lab, Center for Coastal and Ocean Mapping,
University of New Hampshire, Durham, NH 03924, USA
colinw@cisunix.unh.edu

Abstract. There is no visual model of the world in our heads. Over the past few years the phenomena of change blindness and inattentional blindness as well as studies of the capacity of visual working memory all point to the fact that we do not retain much about the world from one fixation to the next. The impression we have of a detailed visual environment comes from our ability to make rapid eye movements and sample the environment at will. What we see at any given instant in time is determined by what we are trying to accomplish. We see what we need to see. If we need to find a path through a crowd we see the openings. If we are trying to find a friend we see the faces. We can think of this process of seeing as the execution of a continuous stream of visual queries on the environment. Depending on the task at hand the brain constructs a visual query and we execute a visual search to satisfy that query. Making visual queries a central concept opens the door to a theory of how we think visually with interactive displays. The process can be thought of as constructing and executing queries on displays. Problem components are formulated into questions (or hypotheses) that can be answered (or tested) by means of pattern discovery. These are formulated into visual queries having the form of search patterns. Visual eye-movement scanning strategies are used to search the display. Within each fixation, active attention determines which patterns are pulled from visual cortex subsystems that do pattern analysis. Patterns and objects are formed as transitory object files from a proto-pattern space. Elementary visual queries can be executed at a rate of 40 msec per simple pattern. Links to non-visual propositional information are activated by icons or familiar patterns, bringing visual information simultaneously into verbal working memory.

1 Introduction

How is it that we have the compelling illusion that we are aware of the visual complexity of the world? Make no mistake, this is an illusion, any numbers of experiments have shown it to be so. The studies of change blindness suggest that we hold only about three objects in our visual working memories from one second to the next (Vogel, Woodman & Luck, 2001). Studies of inattentional blindness show that we simply do not register things even though we may be looking right at them, if we are attending to some other visual patterns (Rock & Gutman, 1981). Studies using unsuspecting subjects in the real world show that people are unaware of 99% or more of what is in the visual field (Simons & Levin, 1998).

S.-O. Tergan and T. Keller (Eds.): Knowledge and Information Visualization, LNCS 3426, pp. 27 – 35, 2005.

So we have an apparent contradiction. On the one hand we subjectively think we see everything; on the other, it seems, we see almost nothing.

The resolution is that in the words of Kevin O'Regan "the world is its own memory" (O'Regan, 1992). We do not need to have the world in our heads because it is out there in all its glorious detail and by making rapid eye movement, as soon as we think we would like to know about something, we have it. It is subjectively instantaneous, although it actually takes about one tenth of a second. The brain is a slow machine, at least compared to modern computers, and so the time-to-execute the query is not noticed. Although we can fully process only very little information, we have sufficiently rapid access to any part of the visual field that we feel we are instantaneously aware of all of it. What we actually do see is determined by attention and the task at hand. Seeing can thus be thought of as a series of visual queries on the world. We are not consciously aware that our eyes are darting to and fro, gathering information, but they are. Most of the visual queries we make of the world seem literally effortless, so much so that we are not even aware that we are making them.

Understanding the process of seeing as a series of visual queries on the world provides a basis for a cognitive systems approach to visual thinking. Visual queries on displays can be faster and more effective than queries to access data in the brain and this is the reason why we think best with the aid of cognitive tools (Hutchins, 1995). Visualizations can be powerful tools. A visual query is executed through a search for a pattern and it is the pattern finding capacity of the visual system that makes visual displays so powerful. In many cases, to perceive a pattern is to solve a problem, and the human visual system is an extraordinarily flexible and adaptive pattern finding system.

This paper outlines the nature of visual thinking with the idea of the visual query as a core concept. First, we briefly review the evidence that there is no model of the world in the head. Next the core functional cognitive components are described and the visual thinking process is outlined.

1.1 The Evidence

There are three major lines of evidence supporting the idea that we do not have a visual model of the world in out heads. First, we are only sensitive to detail in the center of the visual field. At any instance we cannot be aware of detail anywhere except for where we are fixating. Second, there is no evidence that we store more than a minimal amount from one fixation to the next. Indeed the best estimates are that the most we can store are three very simple colored shapes (Vogel et al., 2001). Third, the inattentional blindness studies of Mack and Rock (1998) show that people focusing on a task are generally unaware of visual events that are not relevant to that task, even though these events can be occurring right next to the point of fixation. In the absence of a detailed visual model of the environment in our heads, the most plausible explanation for our conscious feeling of awareness of visual detail is that eye movements are rapid enough that we only have to think we need something and we have it, seemingly without delay. This process can be thought of as the execution of a task-related visual query on the world.

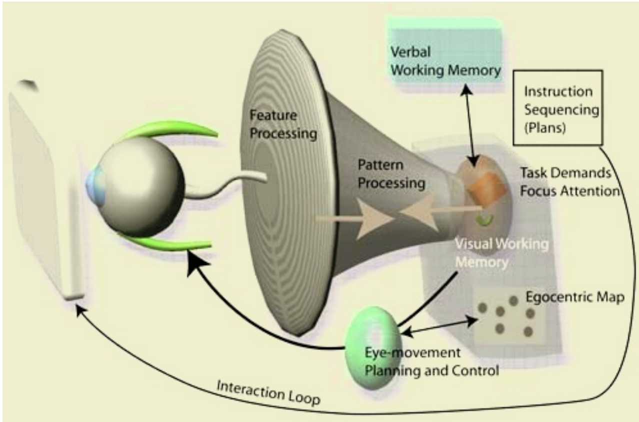

Fig. 1. The major system components involved in visual thinking

1.2 The Power of a Visualization

A visualization consists of both a visual structure and a set of symbols. Structures are embedded in maps, and various types of node-link diagrams. The symbols can be of various types: words, symbolic shapes, icons, glyphs. If the symbols are already familiar they automatically excite the corresponding concepts and cause them to be loaded into verbal working memory. Kahneman, Triesman and Gibbs (1992) coined the term "object file" to describe a short-term linking device that can hold together visual structures in visual working memory together with concepts in verbal working memory. Data structures expressed through effective layout and graphical design can make relationships between concepts readily accessible.

The power of a visualization comes from the fact that it is possible to have a far more complex concept structure represented externally in a visual display than can be held in visual and verbal working memories. People with cognitive tools are far more effective thinkers than people without cognitive tools and computer-based tools with visual interfaces may be the most powerful and flexible cognitive systems. Combining a computer-based information system with flexible human cognitive capabilities, such as pattern finding, and using a visualization as the interface between the two is far more powerful than an unaided human cognitive process.

The remainder of this chapter presents a high-level overview of visual query construction. The three component model illustrated in Fig. 1 is a useful simplification to illustrate the different visual subsystems involved. At the lowest level information is processed through massively parallel feature finding mechanisms. Pattern finding occurs in the mid level; patterns are constructed from low level features according to the top down demands of attention operating in the context of a temporary store called *visual working memory*. Queries are executed by means of eye movements and a focusing of attention on task-relevant patterns so that patterns are held briefly in working memory. In the following sections we consider the critical cognitive subsystems in more detail beginning with the central role of pattern finding.

2 Visual Query Patterns

Visual thinking is, to a large extent, synonymous with pattern finding. In many cases to perceive a pattern is to have a solution to a problem and this is done through visual queries tuned to specific patterns. Visual attention acts as a filter, influencing the middle layer pattern forming mechanisms of vision, so that only the current search pattern, if present, is brought into working memory (Baddeley & Logie, 1999). This process can be thought of as a mechanism whereby competing bottom-up forces are modified by the top down task driven attentional processes. The net result is that we only see what, for the most part, we need to see.

Fig. 2 illustrates how we can focus attention to pull out different parts of a pattern. In this figure three sequences of symbols are encoded: one in fine lines, the second in blurred lines and the third in transparent yellow lines. If you try to read the blurred symbols you will find that you tune out the other patterns. Alternatively you can attend to the thin lines or the broad yellow lines and tune out the others.

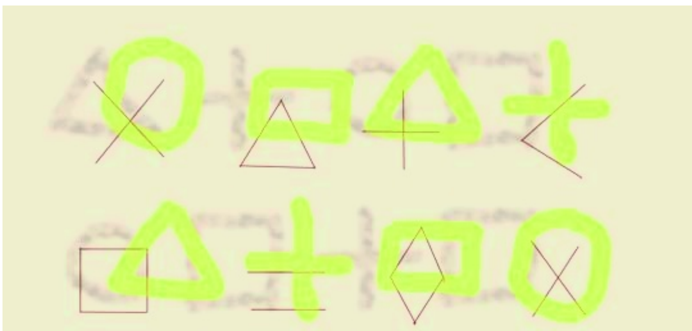

Fig. 2. Three symbol sequences are encoded in different ways. Through attention we can tune our mid-level pattern finding machinery allowing us to read off either the fine-line symbol set, the yellow symbol set, or the fuzzy-grey symbol set

The patterns we can easily tune for are by no means universal. Indeed, understanding which patterns are easily discriminated and which are not is an invaluable tool for the designer. The rules for easy-to-see patterns are complex and difficult to summarize (see Ware, 2004 for an overview). However we can say that the major function of the pattern finding mechanism is to segment the visual world into regions based on some combination of contour, color, motion and texture. The extraction of contours and the connections between objects is critical. Organizing information by regions and contours is, unsurprisingly, critical in display design.

The patterns that can be formed as queries are infinitely diverse: a major highway on a map winding though a number of towns, the pattern of notes on a musical score that characterizes an arpeggio, or the spiral shape of a developing hurricane. For a display to be effective such data patterns must be mapped into visual patterns in such a way that they are visually distinct.

The studies of Triesman (1985) and others showed that we process simple visual patterns serially at a rate of about one every 40-50 msec. Since each fixation typically will last for 100-300 msec. this means that our visual systems process 2-6 objects within each fixation, before we move our eyes to visually attend to some other region.

3 Visual and Verbal Working Memories

The most critical cognitive resource involved in visual thinking is visual working memory. Theorists disagree on details of exactly how visual working memory operates but there is broad agreement on basic functionality and capacity, enough to provide a solid foundation for a theory of visual thinking. Visual working memory can be roughly defined as the visual information retained from one fixation to the next. A list of some key properties follows.

- There is not a single working memory supporting cognition; rather there are several limited-capacity systems for processing auditory, visual and haptic information (Baddeley & Logie, 1999; Baddeley & Hitch, 1974) and there may be additional stores for sequences of cognitive instructions.
- Visual working memory capacity is limited to a small number of simple visual objects and patterns, perhaps 3-5 simple objects (Rensink, O'Reagan & Clark, 1997).
- Kahneman et al. (Kahneman et al., 1992 coined the term object file to describe the temporary grouping of a collection of visual features together with other links to verbal-propositional information. They hypothesized that an object file would contain the neural equivalent of pointers reaching into the part of the brain where visual features are processed as well as pointers to verbal working memory structures and to stored motor memories needed to generate an appropriate response. Rensink (2000, 2002; Rensink et al., 1997) coined the term nexus to describe this instantaneous grouping of information by attentional processing. The semantic meaning or gist of an object or scene (related more to verbal working memory) can be activated in about 100 msec.
- Positions of objects are stored in an egocentric map. This stores some information about approximately nine locations (Postma & De Haan, 1996); three of these may contain links to object files, while the remaining locations code that there is something at a particular region in space, but very little more.
- Deeper semantic coding is needed for items to be processed into long term memory.

4 Eye-Movement Strategies

In a visual search task the eye moves rapidly from fixation to fixation. The dwell period is generally between 200 and 600 msec and the saccade takes between 20 and 100 msec. A simple heuristic strategy appears to be employed by the brain to plan a sequence of eye movements (Wolfe & Gancarz, 1996). The egocentric map is weighted according to the current task. For example, if we are scanning a supermarket to look for oranges, regions of space with the color orange will be set up for searching. Next, eye-movements are executed in sequence, visiting the strongest possible target first, and proceeding to the weakest. Once each area has been processed it is cognitively flagged as visited.

5 Problem Solving with Visualizations

We are now in a position to discuss how thinking can be augmented with visualizations of data. Fig. 5 provides an overview of the various components. This borrows a great deal from Rensink (2000; Rensink et al., 1997) as well as earlier theorists (Baddeley & Logie, 1999; Jonides, 1981; Triesman, 1985). The whole process can be thought of as a set of embedded procedures:

1. Problem components that have potential solutions based on pattern discovery are identified. These are formulated into visual queries consisting of simple patterns.
2. Visual eye-movement scanning strategies are used to search the display for patterns.
3. Within each fixation, the query weights attention and determines which patterns are pulled from the pattern analysis subsystems to answer the queries.
   a. Patterns and objects are formed as transitory object files from a proto-object and proto-pattern space.
   b. Only a small number of objects or pattern components are retained from one fixation to the next. These object files also provide links to verbal propositional information in verbal working memory.
   c. A small number of cognitive markers are placed in a spatial map of the problem space to hold partial solutions where necessary. Fixation and deeper processing is necessary for these markers to be constructed.
4. Links to verbal/logical complex information are activated by icons or familiar patterns, bringing in other kinds of information.

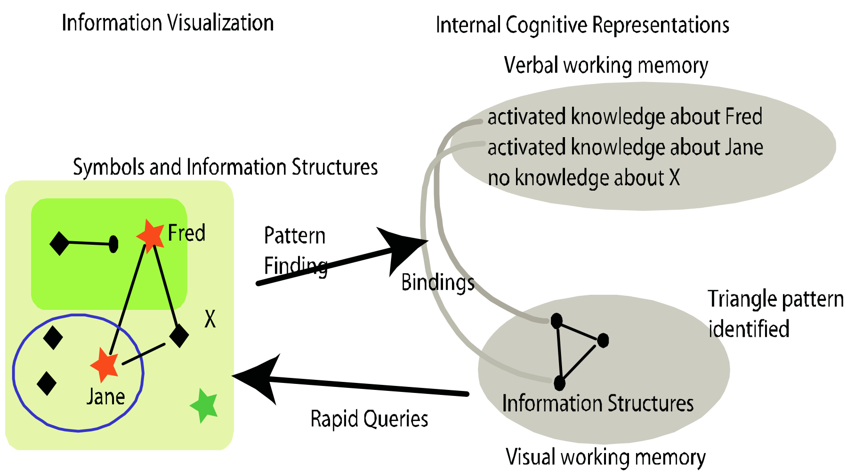

Fig. 3. A snapshot of a cognitive system in action

Fig. 3 illustrates a snapshot of the cognitive system in operation. The cognitive tool is a visualization representing various people in an organization. Visual pattern elements show subgroups and connecting lines show various kinds of working relationships. Visual queries result in different sets of relationship loaded into visual working memory at the same time as corresponding knowledge structures are activated in verbal working memory. For example, the high-level query “Who works with Fred?” might result in a series of visual-pattern queries; one of them focused on the green sub region enclosing Fred and other workers, another focused on the triangular structure. One answer is that Fred works with Jane. Of course this is a very simple example constructed for the purpose of illustration. The power of visualizations, especially if they are interactive front ends to databases, is that they can provide rapid querying on very complex structures.

6 Visual Query Costs

For non-interactive displays, such as maps, eye-movements are the main way of obtaining more information. We simply point our foveas (the high-resolution area in the center of vision) and tune for the required patterns. With computer-based visualization, interactive techniques can be used to increases the size of the information space that can be obtained by means of visual queries. It is useful to compare eye movements with other navigation techniques.

Eye movements allow us to acquire a new set of informative visual objects in 100-200 msec. Moreover, information acquired in this way will be readily integrated with other information that we have recently acquired from the same space. This suggests that the ideal visualization is one where all the information for visualization is available on a single high-resolution screen.

It is instructive to compare the cost of getting information through an eye movement with other methods for getting new information into the visual field. For example, walking across a room, opening a filing cabinet and extracting a document can take minutes, a hugely costly query. Clicking on a hypertext link involves a 1-2 sec guided hand movement and a mouse click to generate an entirely new screenful of information. However, the entire information context typically has changed and the new information may be presented using a different visual symbol set and different layout conventions. Thus several seconds of cognitive reorientation may be required.

There are rapid interactive techniques that bind the human and the computer into a tightly coupled system. Both *brushing* (Becker & Cleveland, 1987) and *dynamic queries* are techniques that allow information to be revealed on some data dimension by making a continuous mouse movement. Hover queries cause extra information to rapidly pop-up as the mouse is dragged over a series of data objects (Munzner, Guimbretière & Robertson, 1999). All three of these require a mouse movement to get started, typically taking about 2 seconds, but after this initial setup time every change in mouse position changes the information visually available resulting in a tight exploratory visual feedback. The data is continuously modified and this may enable an effective rate of several queries per second, similar to the rate for eye movements. However this rate is only possible for quite specific kinds of query trajectories; we cannot jump from point to point in the data space as we can by moving our eyes.

The MEGraph System of Ware and Bobrow (Ware & Bobrow, 2004) provides an example of how a highly interactive node-link visualization can provide views a very complex semantic network, far larger than can be displayed using a static map. MEGraph supports queries in allowing for rapid highlighting of subsets of the graph by setting them in motion. This made it possible for users to rapidly explore a node link diagram that was essentially illegible because of the large number of links. We have recently extended this system to show graphs having up to 3200 nodes. This is more than two orders of magnitude greater than the typical node-link diagram which usually has fewer than 20 nodes.

7 Conclusion

The purpose of this paper has been to present a model of visual thinking based on current theories of visual perception. The model describes how visual queries can be executed with a combination of eye movement scanning patterns and attentional processes within each fixation. Visual working memory retains a small amount of object and location information from one fixation to the next, and this, for many tasks, is a major factor limiting the effectiveness of the visualization. One of the implications of this theory is that the cognitive cost of navigation in visual data spaces will be critical in determining the effectiveness of a particular interactive visual display. In a cognitive systems approach, what matters is how quickly and easily information can be acquired. This will be particularly true when more complex patterns are being sought and where it is necessary to integrate information across several screens. Indeed, the theory suggests that large high-resolution screens should be very effective for complex tasks because they can be navigated by means of eye movements, thereby reducing the need for cognitively disruptive screen changes. When screens are small, or information spaces are very large, the various alternative navigation methods should be weighed in terms of their cognitive load and time requirements. In general, rapid fluid access to information is likely to win out over attractive but slow to navigate 3D spaces.

The model suggests a significant research agenda since it is far from complete. Most of the underlying theory has been developed in vision research laboratories, and not with the goal of understanding and improving information visualizations. Much of the research into visual working memory has focused on simple geometric objects, such as those used by Vogel et al. (2001). Research is needed to understand the visual and cognitive resources needed to support common visual queries on information displays. For example many visualizations consist of node-link diagrams of one form or another. Common queries are “Which nodes are connected?”, “What is the shortest path between two nodes?” and “Is there a path between two particular nodes?” If we can understand the cognitive processes involved in these queries then we can optimize for them. Research is also needed to improve our understanding of how complex queries can be decomposed into simpler ones.

Acknowledgements. This research was supported from Grants from NSF 0081292 and NOAA. Discussions with Ron Rensink have helped refine some of the ideas.

References

Baddeley, A.D., & Logie, R.H. (1999). Working Memory: The Multiple-Component Model. In A. Miyake, & P. Shah (Eds.), *Models of Working Memory* (pp. 28-61). New York: Cambridge University Press.

Baddeley, A.D., & Hitch, G.J. (1974). Working memory. In G.H. Bower (Ed.), *The psychology of learning and motivation: Advances in research and theory*, Vol. 8 (pp. 647-667). Hillsdale NJ.: Erlbaum.

Becker, R.A., & Cleveland, W.S. (1987). Brushing scatterplots. *Technometrics*, *29*(2), 127-142.

Hutchins, E. (1995). *Distributed Cognition*. Cambridge, MA: MIT Press.

Jonides, J. (1981). Voluntary versus automatic control over the mind's eye. In J. Long, & A.D. Baddeley (Eds.), *Attention and Performance*, Vol. 9 (pp. 187-203). Hillsdale, NJ.: Erlbaum.

Kahneman, D., Triesman, A., & Gibbs, B.J. (1992). The reviewing of object files: Object-specific integration of information. *Cognitive Psychology*, *24*, 175-219.

Mack, A., & Rock, I. (1998). *Inattentional Blindness*. Cambridge, MA: MIT Press.

Munzner, T., Guimbretière, F., & Robertson, G. (1999). Constellation: A Visualization Tool for Linguistic Queries from Mindnet. In D. Keim, & G. Wills (Eds.), *Proceedings of the 1999 IEEE Symposium on Information Visualization* (pp. 132-135). Washington, DC: IEEE CS Press.

O'Regan, J.K. (1992). Solving the "real" mysteries of visual perception: The world as an outside memory. *Canadian Journal of Psychology*, *46*, 461-488.

Postma, A., & De Haan, E.H.F. (1996). What was where? Memory for Object locations. *Quarterly Journal of Experimental Psychology*, *49A*(1), 178-199.

Rensink, R.A. (2000). The dynamic representation of scenes. *Visual Cognition*, *7*, 17-42.

Rensink, R.A., O'Reagan, J.K., & Clark, J.J. (1997). To see or not to see: The need for attention to perceive changes in scenes. *Psychological Science*, *8*(5), 368-373.

Rensink, R.A. (2002). Change Detection. *Annual Review of Psychology*, *53*, 245-277.

Rock, I., & Gutman, D. (1981). The effect of inattention on form perception. *Journal of Experimental Psychology: Human Perception and Performance*, *7*(2), 275-285.

Simons, D.J., & Levin, D.T. (1998). Failure to detect changes to people during a real-world interaction. *Psychonomic Bulletin and Review*, *5*, 644-649.

Triesman, A. (1985). Preattentive processing in vision. *Computer Vision, Graphics and Image Processing*, *31*, 156-177.

Vogel, E.K., Woodman, G.F., & Luck, S.J. (2001). Storage of features, conjunctions and objects in visual working memory. *Journal of Experimental Psychology: Human Perception and Performance*, *27*(1), 92-114.

Ware, C. (2004). *Information Visualization: Perception for Design* (2nd Edition). San Francisco, CA: Morgan Kaufman.

Ware, C., & Bobrow, R. (2004). Motion to support rapid interactive queries on node-link diagrams. *ACM Transactions on Applied Perception*, *1*(1) 3-18.

Wolfe, J.M., & Gancarz, G.G. (1996). Guided Search 3.0. A model of visual search catches up with Jay Enoch 40 Years Later. In V. Lakshminarayanan (Ed.), *Basic and Clinical Applications of Vision Science* (pp. 189-192). Dordrecht, Netherlands: Kluwer Academic.