Understanding Search and Errors in Comparative Visual Search: Insights from Eye-Gaze Studies

M. S. Atkins, A. Moise,

School of Computing Science

R. Rohling

School of Engineering

Abstract

We demonstrate that the search for a target during a comparative visual search task of two side-by-side images is undertaken in two stages. First a regular scan path search phase for a likely target is made, followed by a confirmation phase of several fixations on the target in each side-by-side image. The horizontal saccades occurring during the confirmation phase and fixations between left and right side images are necessary because of the limitations of the visual working memory; the subjects prefer to make multiple saccades rather than perform cognitive effort to remember features from one side image to another.

Errors can be classified into different categories based on cumulative fixation durations on the targets in both left and right images. Analysis of our false negative errors showed that recognition errors arose when less than three fixations and horizontal saccades were made on the missed targets in the right and left images. For cumulative fixations >1200 msecs on missed targets in both images the errors can be considered as decision errors rather than recognition errors.

These results show that eyegaze tracking during a comparative visual search task yields readily observed insights for search strategy and decision-making.

CR Categories: H.1.2 [**Information Systems**]: Models and Principles -User/Machine Systems – *Human Factors; Human information processing; software psychology design*; I.4.8 [**Computing Methodologies**]: Image Processing and Computer Vision- Scene Analysis; Tracking

Keywords: eyegaze tracking, comparative visual search, cognition

1   Introduction

Radiologists perform diagnostic tasks by interpreting details on patient images. The diagnostic task has stringent requirements of accuracy to prevent errors, yet it is often repetitive, and speed is also necessary, to complete the workload in a timely way. Our research focuses on designing new interaction techniques to improve radiology workstations. Classic methods for image manipulation involve selection of images from iconic thumbnails, which takes the user's perception away from the visual task. Different interfaces to provide automated image organization and display may help speed up the interpretation process [Moise et al. 2004b] which may lead to shorter completion times due to fewer disruptions of the visual search. To evaluate different interface designs, we developed an artificial radiology look-alike task consisting of a comparative visual search task for artificial targets, as shown in Figure 1. We justified using artificial targets instead of radiology images, for the sake of gaining experimental control. The controlled, abstract conditions permit accurate estimation of the time and errors when performing the tasks with each interaction technique.

We wished to analyse if and how disruption of the visual search affected the search and error performance.
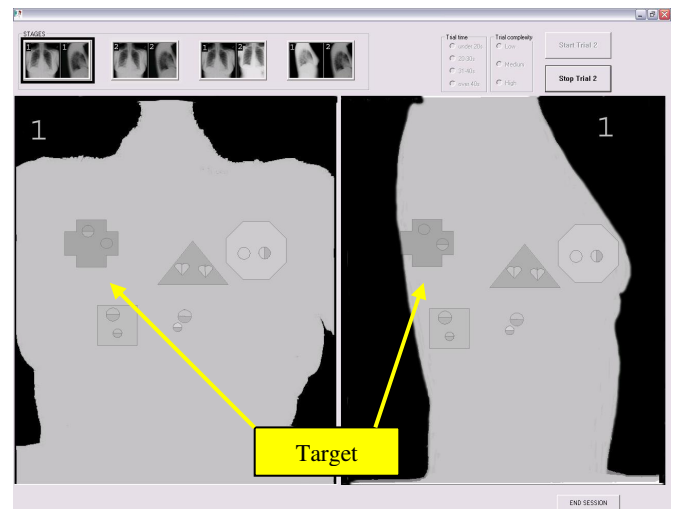
Figure 1. Screen shot of our look-alike comparative visual search task: find the target in a pair of images. The target is in the cross at the top left of each image.

2   Background

In earlier work we showed how sophisticated hanging protocols called Stages, which automate the image organization and display, can help speed up the radiology interpretation process [Moise 2002; Moise et al. 2004a]. The major difficulty was how to evaluate different interface designs, as radiologists' time is very expensive and difficult to obtain. We hypothesised that novices could be used as subjects to evaluate different interaction techniques, given an artificial look-alike radiology task. The novices' behaviour would be similar to the experts' behaviour. We developed an artificial radiology look-alike task (a

comparative visual search task) to test-drive two different workstation interaction techniques. Without eyegaze tracking, we performed evaluations with 20 lay subjects doing the task and with 4 radiologists subjects doing the same task. We showed that naïve subjects could be used instead of experts [Moise et al. 2005]; both groups of subjects had similar performance and error rates.

Recently, eyegaze tracking systems have appeared as a valuable tool for eye-movement-based analysis. They have been used for medical image perception studies [Kundel 2004], for laparoscopic surgery training [Nicolaou et al. 2004; Law et al. 2004] and for level of detail control for real-time computer graphics and virtual reality [O'Sullivan et al. 2003]. Yang et al. [2002] proposed using eyegaze tracking to solve many applications problems in visual search tasks. We showed that eyegaze tracking is a useful tool to help design and assess workstation interaction techniques: different interfaces affect response time performance through disruption of the visual search process [Atkins et al. 2006]. In this paper we discuss how eyegaze tracking yields insights into the search and error processes. We use eyegaze patterns to examine the use of visual memory in visual perception during the search for a matching object (a target) in side-by-side images, and examine the impact of this on the search pattern and the errors.

We hypothesise that the dynamic eyegaze fixations during the search for a target will occur in two phases as others have reported [Pomplun et al. 2001; Inamdar and Pumplun 2003, Kundel 2004]. The first phase, called the search phase, will occur while the subject is searching the whole image for a suspicious object, or a likely target. The second phase, called the recognition phase, will occur when multiple fixations are used to confirm the target. We hypothesise that in a comparative visual search task these two phases are readily separated, using the fixations and saccade patterns within and across hemifields.

We also hypothesise that eyegaze tracking can be used to explain the decision processes during errors. For false negative errors when the target is missed, we expect that the duration of eyegaze fixations on the missed targets can be used to place these errors into three classes proposed by others [Nodine and Kundel 1990; Nodine et al. 1996; Kundel 2004] based on increasing gaze fixation time: search error, recognition (perception) error, and decision error. Others have studied the boundary between recognition errors and decision errors, where, in single images, an individual fixation of longer than about 1000 msecs on the missed target is considered to be a faulty decision error [Berbaum et al. 2001; Kundel and Nodine, 1978; Nodine and Kundel 1987; Kundel et al. 1987]. The advantage of a comparative visual search task is that these decision processes can be readily visualised.

We also expect to observe similar effects reported by others: that the eyegaze fixation time on false negative targets will be longer than on true positive targets [Kundel 2004; Krupinski 2000].

3 Methods and Experimental Design

3.1 Targets

Our artificial look-alike radiology task consists of a search for an artificial target in a pair of images. Targets are two circular disks of the same size split in half in the same direction, either

horizontally or vertically, as shown in the top row of Figure 2. Images also contain distracters such as unequally sized disks, hearts, or octagonal-sided disks. The targets are occluded into two side-by-side images by the addition of a "wild card" to represent the disk divider. A target is incompletely presented in a single image. The viewer can only discriminate a target from a distracter by integrating the information from two related images displayed on a left and a right viewport i.e. there is partial occlusion in each image. The occlusion is simulated by adding a "wild card" disk with a uniform fill to represent the disk divider, as shown in the middle row of Figure 2. Note that a wild card can also instantiate in a disk with incorrect divider orientation, so it is a distracter, as shown in the bottom row of Figure 2. Note a potential target always has a wild-card. Consequently, for every potential target containing a wild card a subject has to register complementary information from the two images of the same study. Hence the comparative visual search task for a target involves a match of several features, including the shape in which the target appears, the size of the target, and the orientation of a dividing line within the target.
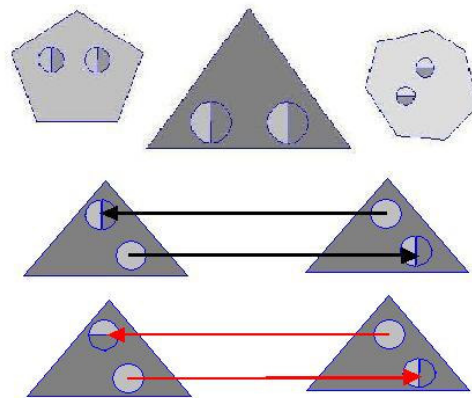
Figure 2. Targets and distracters. Top row: typical targets of 2 circular disks of the same size split in half in the same direction, either vertically or horizontally. Middle row: A target seen in the comparative search over two images: the target is incompletely presented on each image, instantiated with the wildcard shown as the source of the black arrow. Bottom row: The wild card instantiates in a disk with incorrect divider orientation, so it is a distracter

3.2 Simulation of Time

To simulate the radiologist's follow-up on a radiographic examination, we introduced a time dimension by presenting to our subjects two instances of the same scene, corresponding to different time moments. Hence we asked our subjects to detect the target from the two images shown first in study 1, and then track the evolution of the target in time, shown as study 2. In each trial, a target (if present) had to be located in the pair of images in study 1, and its evolution noted in a pair of images in study 2. If the target were present in study 1, it would be in the same position in study 2.

Therefore each trial consisted of two studies, where each study has two images. The two images of study 1 were presented first, and the two images of study 2 had to be viewed next, to detect the evolution in size of any target seen in study 1. We used the following notation convention for trial outcome: '0' means no

target in the study, '1' means a target was present. We therefore distinguish the following five trial outcome conditions: "00" meaning there is no target in both studies, "01" meaning there is no target in the first study but there is a target in study 2, "10" meaning there is a target in the first study but not in the second, "11 same" and "11 diff" meaning there is a target in both studies, which may be the same or different sizes.

3.3 Experimental Protocol

Four radiologist fellows (2 males, 2 females) took part in the experiment. All subjects had normal or corrected vision. There were 15 trials of two studies each, for each of two interaction techniques, so each subject performed 30 trials. Each study had two images containing complementary visual information. A trial had at most one target in study 1 and at most one target in study 2 (in the same position).

An ASL 504 eyegaze tracker was used to record the eyegaze coordinates [ASL 2002]. To display the stimuli images we used a 17" Samsung 770 TFT LCD monitor, with a resolution of 1280x1024, brightness of 220cd/m², contrast ratio of 400:1, and viewing angle of 160/160 (Horizontal/Vertical). The subjects sat 55cm from the screen with their chin in a chin rest to prevent excessive head movement. Each eye gaze sample was averaged over 4 fields using the ASL control software to smooth out small-magnitude eye gaze position jitter. The eye tracker control software ran on an IBM 390E laptop. A Focus Enhancements TView Gold 300 scan converter was used to create a composite video frame of eye gaze location overlaid on the scene (i.e., the frames captured of the experimental task). The experimental task was implemented as a multi-threaded Visual C++ application, to separate the user interaction thread from the data recording thread. Recorded data include time, x and y eye position coordinates and pupil diameter, with a frequency of 60Hz, an accuracy of 0.5° visual angle, and a resolution of 0.25° visual angle. Fixations are assumed to be of at least 100 msecs duration, and are calculated from the points of gaze using a dispersion threshold algorithm based on an algorithm developed by [Salvucci 2000]. For our analysis we defined a fixation cluster to subtend an angle of 3° at the fovea, which is about the size of our image features such as targets and distracters. Although this is smaller than the 5° used by Kundel [Kundel et al. 1978] we found this was large enough to detect all the fixations, and gave the fixation centroid on appropriate image features. Each subject was calibrated three times during the experiment.

The screen layout for the task, shown in Figure 1, consists of left and right viewports containing the stimuli images, the controls used for image selection, and the controls used to start/stop the current trial. Subjects were asked to find an abstract target on a grey background. Each subject performed two consecutive blocks of 15 trials, one for each of the two different interaction techniques. In each trial, a target (if present) had to be located in the first study set of 2 images, and its evolution noted in a second study set of 2 images. The subjects were asked to identify where they found a target by pointing with the mouse and saying "here is the target". This was recorded on camera and used to decide which trials were correct and which trials had an error. Each subject performed the same 30 trials; two started with stimuli set A and the other two started with set B. Instructions about the task were given using several training steps presented on the computer screen. Each training step was followed by a short practice session, where the subjects' understanding of the recently learned concepts was tested, before the trials were started.

4 Results and Discussion—Search Patterns

Figure 3 shows a typical search pattern in a trial with no target. The numbered dots represent fixations in temporal order, and the lines between the dots represent the saccades between the fixations. The search is made over all the potential targets. Note that the label "1" in the top corner of these images indicates that this is study 1 of a trial. The images of study 2 of a trial are indicated by a "2" in the top corner of the image.
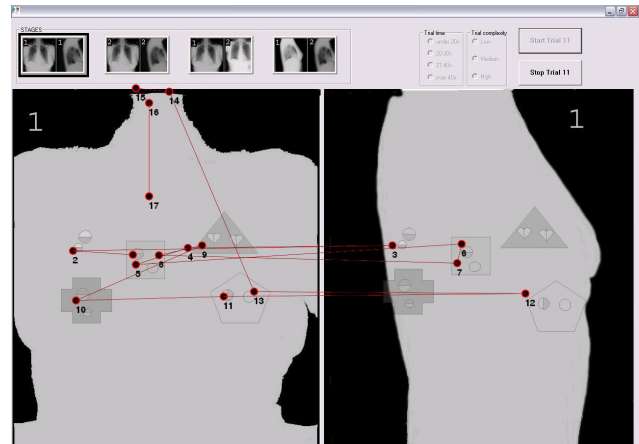

Figure 3. Typical search pattern in a trial with no target. The numbered dots represent fixations in temporal order, and the lines between the dots represent the saccades between the fixations.

Figure 4 shows a typical search pattern in a trial with a target. Two phases can be observed. Fixations 1-18 show saccades during the initial search phase, and fixations 19-25 show saccades and clusters of fixations on the target in both hemifields during the recognition phase.
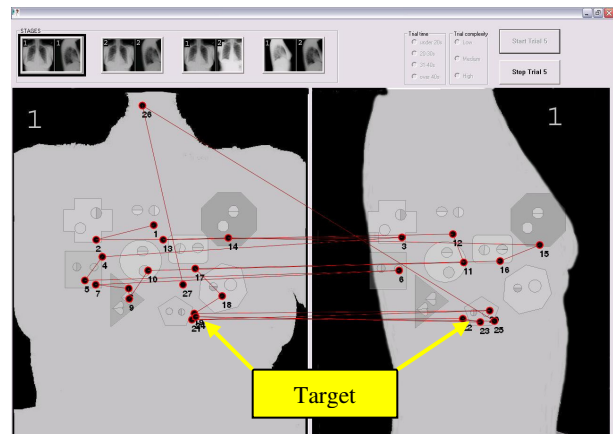
Figure 4. Typical search pattern in a trial with a target (in the pentagon): two phases corresponding to search and recognition.

A "short cut" for target search is possible for study 2, as seen in Figure 5, which is the second study of the same trial shown in Figure 4. When a target has been found in Study 1, the subject can take a short cut and look immediately at the same position for a target in study 2.
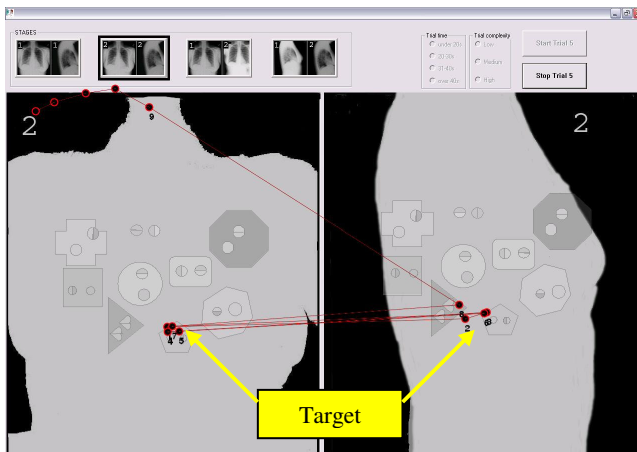
Figure 5. Eyegaze pattern for target search in study 2, where the target from study 1 is immediately fixated.

It is seen that the search pattern in study 2 depends on outcome from study 1 (target, or no target).

Our hypothesis that the search will occur in two phases (search, then recognition) is true. It is readily observed in our comparative visual search task because of the number of horizontal saccades across the hemifield images to compare the targets in each image during the recognition phase. The average fixation duration is around 268 msecs, so the cumulated fixations on the true target often exceed 2000 msecs.

Insights from search patterns reveals that this comparative visual search task for a target involves a match of several features including the shape in which the target appears, the size of the target, the orientation of a dividing line within a target, and the location in the image. The human visual system has a limited short-term memory, for only 4-5 simple features. Therefore to confirm each feature, horizontal saccades are performed, in preference to overloading the cognitive memory; usually at least 3 fixations on each side are made, and 5 horizontal cross-field saccades.

Because of the complex nature of our targets, the limited capacity of the human visual memory, and the need for target comparisons between hemifield images, the cumulated dwell time on our true targets is often much higher than the 1000 msecs reported by others [Kundel 2004].

5 Results and Discussion—Errors

5.1 Number of Errors

Our subjects were instructed to be as accurate as possible, so a correct diagnosis is their primary task. Completing each trial in the shortest possible time interval was a secondary requirement. Thus we traded time for accuracy. There were 17 errors in total, out of 120 trials (each subject performed 30 trials). Subject 1 made 12 errors, subject 2 made 1 error, subject 3 made 4 errors, and subject 4 made 0 errors. There were 16 false negative trials, where a target was missed, and 1 false positive error where a distracter was taken as a target (subject 2's only error). 7 false negative errors were in trials with outcome "01" where there was no target in study 1, but there was a target in study 2.

Table 1 shows a summary of the erroneous false negative trials grouped according to the cumulated fixation time spent on the missed target.

Table 1. False Negative errors: range of cumulated duration of fixations on missed target (msecs), total number of fixations on missed target, number of cross-hemifield transitions on the missed target, and the number of erroneous trials where these conditions hold.

| Duration of fixations on missed target	Total no. of fixations on missed target	No. of cross-hemifield transitions on target	No. of trials
<100	0	0	1
100-1000	2-5	<4	7
1000-1250	2	2	1
1250-1500	6	5	1
1500-5000	>7	>5	6

We expect to observe 3 kinds of false negative errors based on increasing duration of fixations on the missed target: faulty search error, faulty recognition error, and faulty decision error. The first category, faulty search error, when there are no fixations on the target, is easy to define and we had one such occurrence. The difficulty arises in finding the duration threshold separating the pattern recognition errors from the faulty decision errors. There are several possible demarcation points between recognition and decision errors, so further analysis of the eyegaze patterns is required.

We chose to group the trials with cumulated duration between 100 and 1000 msecs into one row (row 2 of Table 1), because these trials likely corresponded to a recognition error as others have found. We also chose to group the trials with much longer fixation dwell times, greater than 1500 msecs together, in the last row of Table 1. In these 6 trials with a total fixation on the missed target of more than 1500 msecs, at least 6 fixations are placed over the missed target in each hemifield, indicating a recognition error. We placed the remaining two trials, where the cumulated dwell time on the missed target is in-between 1000 and 1500 msecs, in individual rows, in order to separate recognition from decision errors.

5.2 Faulty Search Error

Figure 6 shows the single example from our trials, summarized in the top row of Table 1, where the subject never fixated on the missed target. This means the cumulated dwell time on the target was less than 100 msecs.
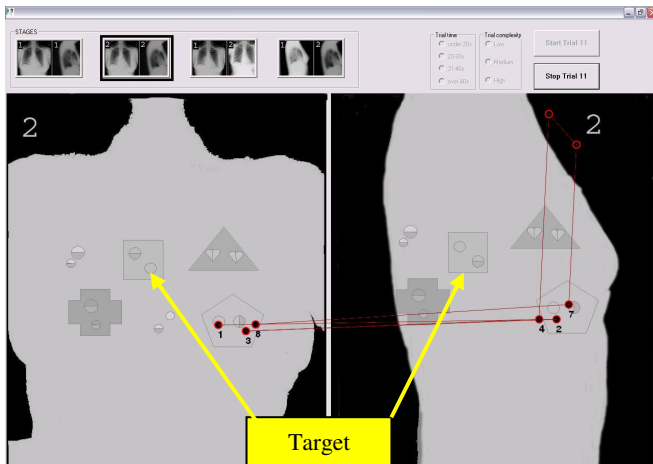
Figure 6. Example of a faulty search error in trial with no target in study 1, but with a target in study 2. There are no fixations on the target (in the rectangle).
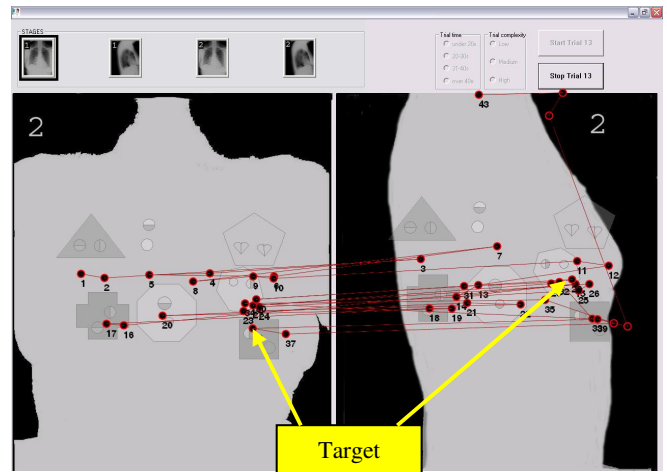
5.3 Faulty Recognition Error

We found that in the 7 trials with a total fixation on the missed target of less than 1000 msecs, only one or two fixations are placed over the missed target in each hemifield, indicating a recognition error. Figure 7 shows an example such a recognition error, from the second row of Table 1, where the target was fixated briefly with two fixations on each hemifield image, but was not recognized. Cumulated fixation time on the missed target was 934 msecs.
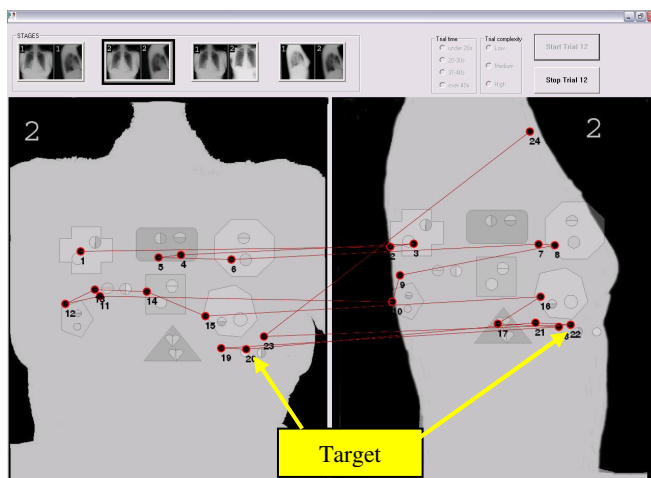
Figure 7. False negative result: Example of faulty recognition error where the target was fixated briefly for 934 msecs, but not recognized.

5.4 Faulty Decision Error

An example of a decision error is shown in Figure 8, where the target was fixated 13 times, but not identified. Cumulated fixation duration on the missed target was 4500 msecs.

Figure 8. False negative result: Example of faulty decision error where the target was fixated many times and for a long time, 4500 msecs, but not called.

Others used cumulated fixation duration of 1000 msecs to distinguish between recognition and decision errors [Kundel 2004]. We had two trials in the border, and we considered each separately. Figure 9 shows the trial with cumulated duration of 1183 msecs on the missed target (summarized in row 3 of Table 1).
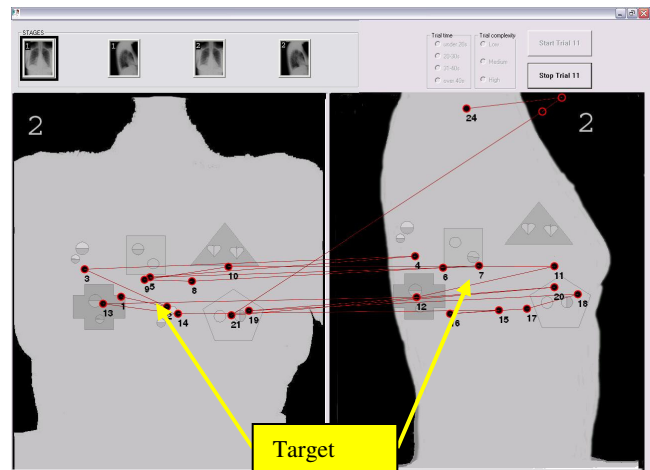
Figure 9. False Negative result: the entire image is covered with fixations, the target in the rectangle is fixated briefly, but no target was found. Cumulated time fixating on the missed target was 1183 msecs. This appears to be a recognition error.

This trial has only 2 fixations on the missed target in each hemifield; it appears that no particular attention was paid to this target. Therefore we considered it to be a recognition error.

The trial with the next highest duration on the missed target, 1450 msecs, is shown in Figure 10 (row 4 of Table 1).
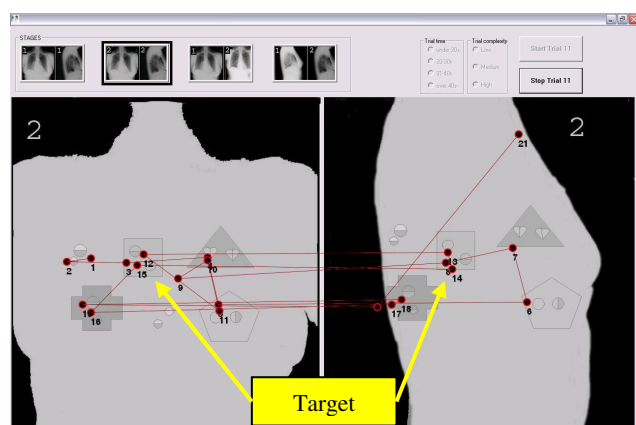
Figure 10. False Negative result: the missed target in the rectangle is considered carefully with 3 fixations in the left side, and 4 on the right side, yet the target is not called. Cumulated time fixating on the missed target was 1450 msecs. This appears to be a decision error.

This trial has several fixations on the missed target, and 4 cross-hemifield transitions, indicating that the subject recognized a potential target, but after examination, decided incorrectly that there was not a target.

5.5 Eyegaze Duration on false negative and true positive targets

We expected to observe similar effects reported by others [Krupinski 2000]: that the eyegaze fixation time on false negative targets will be longer than on true positive targets. This does not appear to be true for this task, as both the true positive targets and even the non target areas, often have an accumulated fixation duration of >4000 msecs. For example, in trials with outcome "10" where the target which was present in the first study is now missing in the second study, the non-target area in the second study is often fixated for several seconds with many cross-hemifield saccades. These fixation times overlap and even surpass the durations on the false negative fixations noted in Table 1. This arises because of the complex nature of our targets, and the limitations of the human visual system, which require several passes across the true targets, to confirm them. The implication of this negative hypothesis result is very important, in that it may not be possible to design a computer assisted diagnosis system to report the suggestion "re-examine" when a subject performs multiple saccades across hemifield images in a comparative visual search and does not call a target.

5.6 Eyegaze Duration on false positive targets

There was only 1 false positive call in the 120 trials, amounting to 0.83% of trials. This is lower than in most real radiology tasks likely because in real tasks, the consequences of a false positive are much less serious than the consequences of a false negative. The only false positive trial arose in the second study of a high complexity trial with outcome "00". In total there were 8 fixations on the false positive target totalling 2483 msecs,

showing that indeed the cumulated fixation duration on the false positive target was high, but not necessarily higher than on true targets. With only this one false positive result, we can neither support nor refute the hypothesis that false positive errors have longer duration fixations than true targets.

6 Conclusions and Future Work

In this study we have shown that eyegaze fixations provide insight into the search techniques used in comparative visual search tasks, by separating the search into two readily-discernable phases: initial search and confirmation. When a target is recognized, the eyegaze patterns change. First there is an initial systematic search phase for the target. After the target has been recognized, there are then several transition fixations on the target during the recognition phase. The subject usually makes 3-4 fixations on the target in each hemifield to confirm or not the presence of a target, and so the cumulated fixations on the true target usually exceed 2000 msecs. The human visual system has a limited short-term memory, for only 4-5 simple features. Therefore to confirm each feature, horizontal saccades are performed, in preference to overloading the cognitive memory; usually at least 3 fixations on each side are made, and 5 horizontal cross-field saccades. Because of the complex nature of our targets, the limited capacity of the human visual memory, and the need for target comparisons between hemifield images, the cumulated dwell time on our true targets is often much higher than the 1000 msecs reported by others for duration of fixations on targets in single images.

It also appears that the classification of false negative errors in this comparative visual search task can be made on the basis of cumulative fixations on the missed target and on the number of cross-hemifield saccades made. Cumulative fixations of <1250 msecs are recognition errors and cumulative fixations of > 1250 msecs should be treated as decision errors. Furthermore, the number of cross-hemifield saccades made can be used to differentiate recognition errors from decision errors; if less than 6 fixations and 4 transition saccades are made, it is a recognition error.

Future work will include studying the effects of fatigue and experience on performance. The next step is to evaluate radiologists performing real diagnosis tasks to observe in more detail whether eyegaze fixations and search paths can be used to highlight and possibly then eliminate false negative decisions.

References

1. ASL, Applied Science Laboratories. 2000. ASL Eye Tracker Manual, www.asl.com
2. A. MOISE AND M.S. ATKINS. 2002. Workflow Oriented Hanging Protocols for Radiology Workstation. *Proc SPIE Medical Imaging* 4685, 189-199.
3. A. MOISE AND M.S. ATKINS. 2004a. Interaction techniques for radiology workstations: impact on users' productivity. *Proc SPIE Medical Imaging* 5371, 16-22.
4. A. Moise and M.S. Atkins. 2004b. Design Requirements for Radiology Workstations, *Journal of Digital Imaging* 17(2), 92-99, June 2004.
5. A. Moise and M.S. Atkins. 2005. Evaluating different radiology workstation interaction techniques with radiologists and laypersons. *Journal of Digital Imaging*, 18(2): 116-130, June 2005

6. M.S. Atkins, A. Moise and R. Rohling. 2006. An Application of Eyegaze Tracking for Designing Radiologists' Workstations: Insights for Comparative Visual Search Tasks. *ACM Transactions on Applied Perception:*3(2): 136-151, April 2006 .

7. BERBAUM, K. S., BRANDSER, E.A., FRANKEN, E.A., DORFMAN, D.D., CADWELL, R.T., AND KRUPINSKI, E.A. 2001. Gaze dwell time on acute trauma injuries missed because of satisfaction of search. *Academic Radiology* 8(4), 304-314.

8. INAMDAR, I. AND POMPLUN, M. 2003. Comparative Search Reveals the Tradeoff between Eye Movements and Working Memory Use in Visual Tasks. *Proceedings of the 25th Annual Meeting of the Cognitive Science Society, Boston*, 599-604.

9. KRUPINSKI, E.A. 2000. The importance of perception research in medical imaging. *Radiation Medicine* 18(6), 329-334.

10. KUNDEL, H.L., NODINE, C. F. AND CARMODY, D. 1978. Visual scanning, pattern recognition and decision-making in pulmonary nodule detection. *Investigative Radiology* 13(3), 175-181

11. KUNDEL, H.L.,NODINE, C.F.,THICKMAN, D., AND TOTO L.1987. Searching for lung nodules: a comparison of human performance with random and systematic scanning models. *Investigative Radiology* 22, 417-422

12. KUNDEL, H.L. 2004. Reader Error, Object Recognition, and Visual Search, *Keynote speech, In Proceedings of SPIE Medical Imaging 2004*, SPIE, 5372, 1-11

13. LAW, B., ATKINS, M.S., KIRKPATRICK, E.A., LOMAX, A., J., WILSON, J. 2003. Eye Gaze Patterns Differentiate Skill in a Virtual Laparoscopic Training Environment. *Proceedings of Eye Tracking Research and Applications, ETRA 2004*, pp. 41-47

14. NICOLAOU, M., JAMES, A., DARZI, A. AND YANG, G.-Z. 2004. A Study of saccade transition for attention segregation and task strategy in laparoscopic surgery. In *MICCAI* 2004, 97-104.

15. NODINE, C.F. AND KUNDEL, H.L. 1987. Using eye movements to study visual search and to improve tumor detection. *Radiographics* 7, 1241-1250.

16. NODINE, C. F. AND KUNDEL, H. L. 1990. A Visual Dwell Algorithm Can Aid Search and Recognition of Missed Lung Nodules in Chest Radiographs. In *Visual Search 2*. D. B. T. Francis. North-Holland, Amsterdam: 399-406.

17. NODINE, C.F., KUNDEL, H.L, LAUVER, S.C. AND TOTO, L.C. 1996. The nature of expertise in searching mammograms for breast masses. *Proc SPIE Medical Imaging* 2712, 89-94.

18. O'SULLIVAN, C., DINGLIANA, J. AND HOWLETT, S. 2003. Eye-movements and interactive graphics. In *The Mind's Eyes: Cognitive and Applied Aspects of Eye Movement Research,* R. RADACH, J. HYONA AND H. DEUBEL, Eds., Elsevier Science, Oxford, 555-571

19. POMPLUN, M., SICHELSCHMIDT, L., WAGNER, K., CLERMONT, T., RICKHEIT G. AND RITTER, H. 2001. Comparative Visual Search: a Difference that Makes a Difference. *Cognitive Science* 25(1), 3-36.

20. SALVUCCI, D. AND GOLDBERG, J. 2000. Identifying fixations and saccades in eye-tracking protocols. *Proceedings IEEE eye tracking research and applications conference, ETRA 2000*, 71-78.

21. YANG, G-Z., DEMPERE-MARCO, L., HU, X-P. AND ROWE, A. 2002. Visual Search: psychophysical models and practical applications. *Image and Vision Computing* 20, 273-287.