

Subtle Gaze Manipulation for Improved Mammography Training

Srinivas Sridharan*
Reynold Bailey
Rochester Institute of Technology

Ann McNamara
Texas A&M University

Cindy Grimm
Washington University in St. Louis

Abstract

We use the Subtle Gaze Direction technique (SGD) to guide novices as they try to find abnormalities in mammograms. SGD works by performing image-space modulations on specific regions of the peripheral vision to attract attention. Gaze is monitored and modulations are terminated before they are scrutinized with high-acuity foveal vision. This approach is preferred to overt techniques which permanently alter images to highlight areas of interest. SGD is used to guide novices along the scanpath of an expert radiologist. We hypothesized that this would increase the likelihood of novices correctly identifying irregularities. Results reveal that novices who were guided in this manner performed significantly better than the control group (no gaze manipulation). Furthermore, a short-term post-training lingering effect was observed among subjects guided using SGD. They continued to perform better than the control group once the training was complete and gaze manipulation was disabled.

CR Categories: I.3.3 [Computer Graphics]: Picture/Image Generation—Display algorithms;

Keywords: gaze manipulation, radiological training, mammography, image-space modulations

1 Introduction

The American Cancer Society estimates that 250,000 new cases of breast cancer were diagnosed in the US in 2009-2010 [American Cancer Society]. The statistics also show that breast cancer is the leading form of cancer among women. Early diagnosis and treatment of breast cancer are essential to maintaining the patients long-term health. In addition to breast self-exams and clinical breast exams, mammograms play an important role in the early detection of breast cancer. Radiologists undergo extensive training to become proficient at reading mammograms. In the US, an expert radiologist typically completes four years of undergraduate study, four years of medical school, one year of internship, four years of residency training and one or two years of additional fellowship training. Despite advances in technology, radiological training still uses the conventional approach of having a trainee work alongside an expert radiologist.

In this paper we present a novel training technique that uses gaze manipulation to guide novices along the recorded scanpath of an expert radiologist. We accomplish this using the Subtle Gaze Direction technique (SGD) [Bailey et al. 2009] which combines real-time eye-tracking with subtle image-space modulation. It has minimal impact on the viewing experience as it does not change the overall



Figure 1: Photograph of experiment setup. The eye-tracking hardware is fixed to the bottom of the screen. Participants were asked to identify irregular regions in a sequence of mammogram images. Gaze manipulation techniques were used to influence their sequence of fixations during the experiment.

appearance of the image being viewed. Subtlety is achieved by presenting the modulations only to the low-acuity peripheral regions of the field of view so that the viewer is never allowed to scrutinize the modulations. The technique has been shown to be quite fast and accurate: viewers typically attend to target regions within 0.5 seconds of the onset of the modulation and the resulting fixations are typically within a single perceptual span of the target. While this shows that the technique is successful in directing gaze, its usefulness for training applications has not yet been established. This paper explores the application of Subtle Gaze Direction within the context of digital mammography training.

We hypothesized that guiding a novice along the scanpath of an expert radiologist would improve the likelihood that the novice correctly identifies the abnormalities in the mammograms. To test this hypothesis, we designed an experiment to explore whether Subtle Gaze Direction is capable of improving the efficiency of digital mammography training. During a training session we noticed that actively guiding participants along the expert's scanpath or directing them to the regions marked by the expert significantly improved their accuracy compared to a control group who viewed the mammograms without gaze manipulation. We also conducted two follow-up sessions without the use of gaze manipulation, a short-term follow-up session (5 minutes later) and a long-term follow-up session (1 week later), to determine if the novice participants had become sensitized to the strategies used during the training session. In the short-term follow-up session, we observed that the groups of participants who were guided using SGD continued to perform better than the control group. This effect was not observed in the long-term follow-up session.

The remainder of this paper is organized as follows: background and related work is presented in section 2, the experiment design is presented in section 3, analysis and discussion of the experimental results are presented in section 4 and the paper concludes in section 5 with a summary of the contributions and potential avenues of future research.

*e-mail: sxs9716@rit.edu

Copyright © 2012 by the Association for Computing Machinery, Inc. Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, to republish, to post on servers, or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from Permissions Dept, ACM Inc., fax +1 (212) 869-0481 or e-mail permissions@acm.org.

ETRA 2012, Santa Barbara, CA, March 28 – 30, 2012.
© 2012 ACM 978-1-4503-1225-7/12/0003 \$10.00

2 Background

Understanding the diagnostic process used by radiologists when searching for cancerous regions in mammograms is an interesting and active area of research. Studies have been conducted to analyze the fixation pattern of expert radiologists when viewing mammogram images. Kundel et al. [2008] recorded and analyzed the fixations of expert radiologists from three independent institutions as they studied cancerous regions on mammograms. Their work provides useful information about the duration of expert viewing and identifies several regions of interest in digital mammograms. Mello-Thoms et al. [2006] also analyzed dwell time and number of fixations by experts during scans for breast cancer in mammograms using both head mounted and remote eye tracking devices. Krupinski et al. [1996] studied decision making among mammographers and radiology residents with gaze duration as the key parameter.

The study of expert gaze patterns within the medical field is not limited only to breast cancer research and mammograms. Krupinski [2000] summarizes work in this area and discusses the importance of perception research in medical imaging. Carmody et al. compare fixations between instructors and radiology residents for lung scan images [Carmody et al. 1980] and chest scans [Carmody et al. 1984]. Sowden et al. [2000] used eye tracking to study how perceptual learning affects a subject's ability to detect features in general X-ray images. Litchfield et al. [2010] [2011] showed that viewing an expert's eye movements can help to improve identification of pulmonary nodules in chest x-rays and aids problem solving.

Image processing and computer vision techniques can be used to automate the process of abnormality detection in mammograms and other medical images. Neural network techniques for example, have been used to detect solid breast tumors on ultrasonic images [Chen et al. 1999]. Other techniques use first- and second-order histogram statistics on segmented regions and local textures to diagnose micro-calcifications in malignant breast tumors [Dhawan et al. 1996]. Markov's Random Field segmentation can be used to detect suspicious regions and distinguish them from normal regions using a fuzzy binary decision tree [Li et al. 1995]. While these techniques automate the process of abnormality extraction in medical images, it is still necessary for the results to be validated by an expert radiologist.

Currently most radiological training programs still use the conventional approach of having a trainee work alongside an expert radiologist. We propose a novel computer-based training technique that uses gaze manipulation to guide novices along the recorded scanpath of an expert radiologist. Computer-based training systems and workstations have become more popular in the medical field with the emergence of digital medical images and improved computer graphics and visualization techniques [Gay et al. 1997; Dugas et al. 2001; Sharples et al. 2000; Muniyandi et al. 2003; Dessler 2007; Chentanez et al. 2009; Relan et al. 2009].

Strategies for enhancing performance on visual search tasks range from guiding attention to previously unattended regions [Qvarfordt et al. 2010] to guiding attention directly to the relevant regions in a scene [Grant and Spivey 2003; Thomas and Lleras 2007; Wang and Spelke 2002]. Such techniques have been shown to improve reaction time and enhance problem solving capabilities. In the experiment presented in this paper, we recruit subjects having no prior knowledge about mammogram images and use the Subtle Gaze Direction technique to guide them along the scanpath of an experienced radiologist as they search for irregularities. In our experiment we also analyze the post-training effects on the subjects guided using Subtle Gaze Direction compared to the subjects in the control group who were not gaze directed.

3 Experiment Design

This section describes an experiment conducted to investigate if subtle gaze manipulation is capable of improving the efficiency of digital mammography training. During a training session, participants viewed a randomized sequence of mammogram images and were asked to identify what they considered to be irregularities. There were four groups of participants. One group viewed the images without the use of gaze manipulation. They served as the control group for the experiment. For the other groups, the Subtle Gaze Direction technique was used to guide the participants in various ways. Two follow-up sessions were completed without gaze manipulation to determine if the participants became sensitized to the training method used.

3.1 Stimuli

The database of mammogram images used in this study was provided by the Mammographic Image Analysis Society [Suckling et al. 1994]. The database contains pairs of mediolateral images from 161 patients along with a separate text file containing information collected from an expert radiologist such as the x,y image coordinates of the center of abnormalities and the approximate radius in pixels of a circle enclosing the abnormality. What is missing however, is the expert's scanpath when trying to locate these abnormalities. We hired our own expert radiologist to view a subset of 65 images from the database and to mark any abnormalities present by drawing a circle enclosing the abnormality. The expert's scanpath was recorded during this process and later used to guide one group of the novice participants.

Stimuli for the experiment were presented on a 22 inch widescreen monitor, operating at 60 Hz with a resolution of 1680 x 1050. The stimuli consisted of the 65 pairs of mammogram images that were viewed by the expert. Five of these images were used for an initial tutorial session to familiarize the participants with the software interface. Twenty images were used for a training session, twenty images were used for a short-term follow-up session, and twenty images were used for a long-term follow-up session. For all sessions, the participants were instructed to identify what they considered to be irregularities in the images. For all sessions, images were displayed for 10 seconds before input from the user was accepted. For the gaze-directed groups, the Subtle Gaze Direction technique was applied during this time frame (but not during the subsequent marking stage). This helped to minimize the impact of viewing duration on the results and also allowed enough time to guide the viewer's gaze about the images. Ten seconds was chosen based on the expert's average viewing time before making the first selection (9.71 seconds).

The pairs of images for each patient were arranged so that the image of the left breast was placed on the right and the image of the right breast was placed on the left to mimic the preferred arrangement of the expert. Figure 1 shows a participant viewing one of these images on a monitor connected to an eye-tracking device. The eye-tracker used in this study is a SensoMotoric Instruments iView X Remote Eye-Tracking Device operating at 250 Hz with gaze position accuracy $< 0.5^\circ$.

3.2 Participants

20 novice participants (7 females, 13 males), between the ages of 18 and 30 volunteered to participate in this study. All participants reported normal or corrected-to-normal vision with no color vision abnormalities. Participants were randomly assigned to one of four groups:

- **Static group:** 5 participants were presented with a randomized sequence of the 20 training images without the use of gaze manipulation. This group served as the control group for the experiment.
- **Gaze-directed group using expert scanpath :** 5 participants were presented with a randomized sequence of the 20 training images with gaze manipulation used to guide them to follow a similar scanpath (sequence of fixations) as the expert.
- **Gaze-directed group using expert selections :** 5 participants were presented with a randomized sequence of the 20 training images with gaze manipulation used to guide them only to the regions marked by the expert as irregularities. The overall scanpath of the expert was not used.
- **Gaze-directed group using adversarial scanpath:** 5 participants were presented with a randomized sequence of the 20 training images with gaze manipulation used to guide them along a scanpath that was chosen by the researchers to follow different directions and locations than that of the expert.

We hypothesized that using gaze manipulation to guide a novice along a similar scanpath as the expert or directly to the locations marked by the expert would improve the likelihood in correctly identifying irregular regions on the mammograms. Similarly, we hypothesized that using gaze manipulation to guide the novice's focus along an adversarial scanpath would reduce the likelihood that the novice correctly identifies the irregular regions on the mammograms.

3.3 Procedure

The expert radiologist hired to participate in this study helped to guide the user interface design of our training platform so that it closely mimics that of standard digital mammography systems. The expert also described external factors such as preferred lighting conditions for studying mammograms which we incorporated into the design of our experiment. The experiment was divided into four stages — a tutorial session, a training session, and two follow-up sessions to determine if the novice users had become sensitized to the training method used.

3.3.1 Tutorial Session

All participants took part in the tutorial session. They were each presented with the same five images in a randomized order and asked to identify what they considered to be irregularities in the images. The images were displayed for 10 seconds before user input was accepted. A left mouse click-and-drag motion was used to draw a circular region that enclosed the abnormality. The Shift key was locked in software to ensure that the region drawn was always circular. Multiple regions could be selected in this manner. A double-click moved on to the next image. The sole purpose of the tutorial session was to get users familiar with the user interface and method of region selection to be used in the remaining stages of the experiment. No feedback about the accuracy of their selected regions was given.

3.3.2 Training Session

All participants took part in the training session. They were presented with the same twenty images in a randomized order and asked to identify what they considered to be irregularities in the images. For participants in the gaze-directed groups, the Subtle Gaze Direction technique was used to influence the order and location of their fixations during the first 10 seconds of viewing. Subtle Gaze Direction was implemented as described in [Bailey et al. 2009], i.e.

viewer gaze was monitored in real time to ensure that modulations were only presented to the peripheral regions of the field of view and modulations were immediately terminated as the viewer's focus approached the modulated regions. We found that no single modulation intensity worked well in all cases because of the wide variation in mammogram images from black to white. Some required more intense modulation than others to attract attention in the peripheral vision. To overcome this problem, the intensity was set to increase gradually until it captured the user's attention as evidenced by a saccade toward the modulated region.

To determine the location and order of the modulations for the expert-scanpath gaze-directed group, the expert's scanpath was replayed in slow motion and the researchers manually selected what they considered to be the center of each cluster of fixations in sequence. This manual approach worked well since there were only 20 expert videos to consider. Of course, clustering algorithms [Xu and Wunsch 2005] or connected components analysis [Gonzalez and Woods 2006] could be used to automate this process. Using these simplified expert scanpaths as a guide, the researchers also manually selected the order and locations of the modulations for the adversarial scanpaths. These were chosen to follow different directions and locations than that of the expert. Figure 2 shows an example of the expert's raw scanpath and the resulting simplified scanpath and adversarial scanpath. It also shows an overlay of both the simplified scanpath and the adversarial scanpath to illustrate how different they are. For the expert-selection gaze-directed group, the modulations corresponded to the centres of the regions marked by the expert as containing irregularities. The modulations were presented in the order they were marked.

3.3.3 Short-Term Follow-up Session

Following the training session, the participants were given a five minute break. A follow-up session was run using twenty fresh images. No gaze manipulation was used in this session for any of the participants. The purpose of this follow-up session was to determine if the novice participants had become sensitized (short-term) to the strategies used during the training session.

3.3.4 Long-Term Follow-up Session

The participants were asked to return one week later for a second follow-up study. Similar to the first follow-up session, this session was run using twenty fresh images and no gaze manipulation was used for any of the participants. The purpose of this follow-up session was to determine if the novice participants had become sensitized (long-term) to the strategies used during the training session.

4 Results

To facilitate analysis of the novices' performance, we define a weight-based accuracy measure as follows:

$$Accuracy = \left(\frac{1 + (w_h * h + w_c * c + w_m * m)}{1 + h + c + m} \right) * 100 \quad (1)$$

where h is the number of hits, c is the number of close matches, and m is the number of misses. w_h , w_c , and w_m are corresponding weights. Figure 3 illustrates the concepts of hits, close matches, and misses which are defined in terms of the circles drawn by the expert and the circles drawn by the novice. A *hit* occurs if the distance between the centers is less than either of the radii as shown in the top row of Figure 3. A *close match* occurs if the distance between the centers of the circles is less than the sum of the radii as shown in the bottom left of Figure 3. Finally, a *miss* occurs if the circles

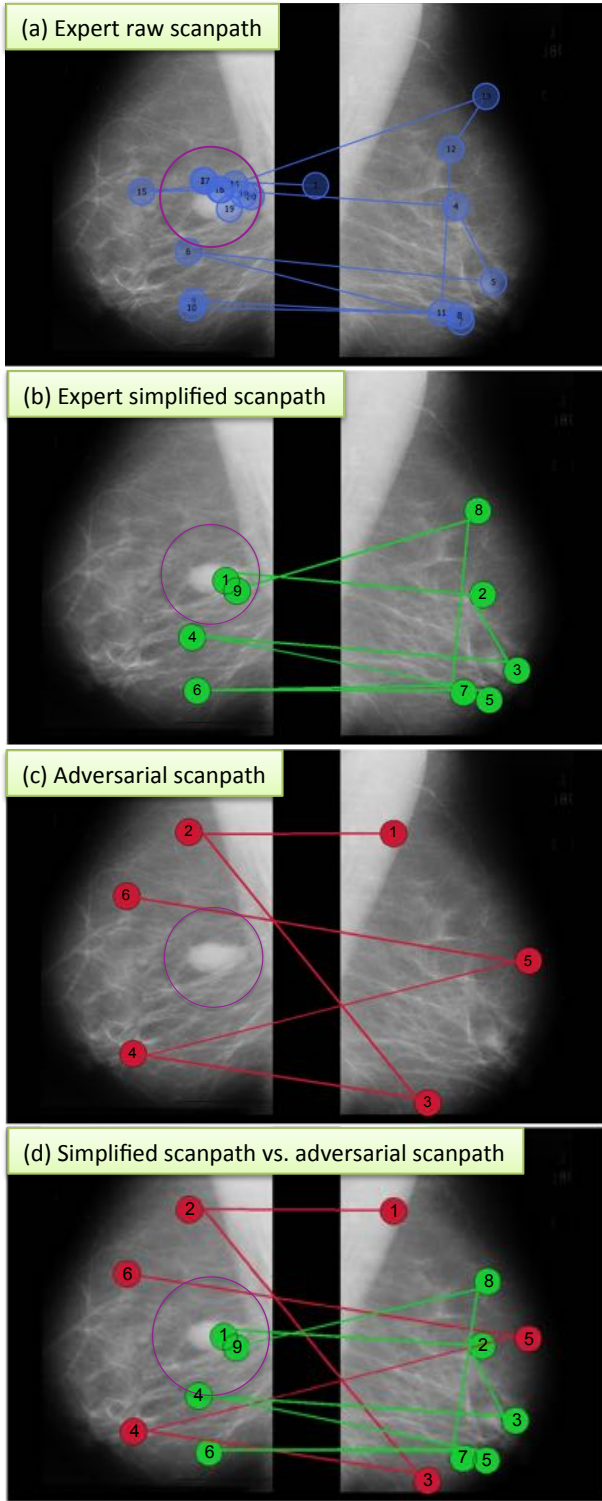


Figure 2: Expert’s raw scanpath (a), simplified expert scanpath (b), adversarial scanpath (c), and comparison of simplified expert scanpath and adversarial scanpath (d) for one image from the dataset. The large purple circle shows the area marked by the expert as an irregularity. The numbers in the smaller circles of (a) indicate the order of fixations of the expert. The numbers in the smaller circles of (b) and (c) indicate the order of modulations used to guide participants in the corresponding groups.

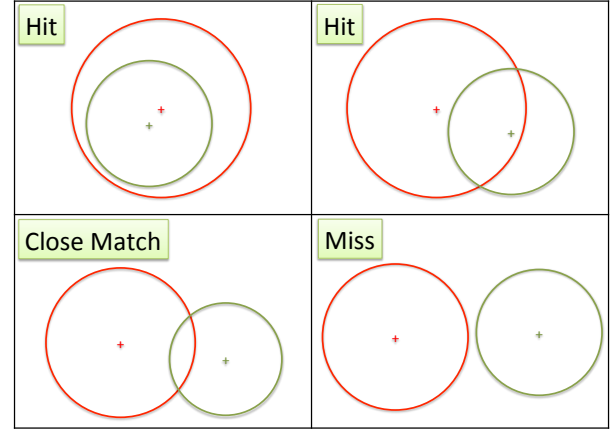


Figure 3: Illustration of hits, close matches, and misses. The circles represent selections by the expert and novice. The center of the circles are represented by the crosses.

do not overlap as shown in the bottom right of Figure 3. If multiple novice selections result in hits to a single region selected by the expert, only one is considered to be a hit and the others are ignored. Likewise if multiple novice selections result in close matches to a region selected by the expert, only one is considered to be a close match and the others are ignored. In the case where there is a mix of hits and close matches to a single region selected by the expert, only one hit is considered and the others are ignored. We assign the following weights: $w_h = 1$, $w_c = 0.75$, and $w_m = 0$. Note that the weight for a close match is biased more towards a hit than a miss. The reason for this is that the selection is sufficiently close to an abnormality to warrant further investigation instead of labeling it simply as a random selection.

Alternatively, binary classification statistics [Agresti 1989] [Fawcett 2004] can be used to establish measures of accuracy as well as sensitivity and specificity. To calculate these properties it is necessary to categorize the test outcomes as *true positives*, *true negatives*, *false positives*, and *false negatives*. Table 1 shows the mapping of test outcomes to these conditions.

True positive (TP)	Abnormal regions exist, and are correctly identified by the subject
False positive (FP)	No abnormal regions exist, yet subject marked regions as being abnormal
True negative (TN)	No abnormal regions exist, and no regions were marked by subject as being abnormal
False negative (FN)	Abnormal regions exist, but were not marked by the subject

Table 1: Mapping of test outcomes to true positives, true negatives, false positives, and false negatives.

For this study *sensitivity* refers to the ability of the subject to correctly identify existing abnormalities in the mammograms. Sensitivity is computed as follows:

$$Sensitivity = \frac{(\#ofTP)}{(\#ofTP + \#ofFN)} * 100 \quad (2)$$

On the other hand, *Specificity* refers to the ability of the subject to correctly identify mammograms containing no abnormalities.

Specificity is defined as follows:

$$Specificity = \frac{(\#ofTN)}{(\#ofTN + \#ofFP)} * 100 \quad (3)$$

The sensitivity and specificity values can then be combined to produce a binary classification based measure of accuracy as follows:

$$Accuracy = \frac{(\#ofTP + \#ofTN)}{(\#ofTP + \#ofTN + \#ofFN + \#ofFP)} * 100 \quad (4)$$

4.1 Training Session Results

Figure 4 shows the average weight-based accuracy (equation 1) for the various groups of participants during the training session. The static group averaged 54.2%, the group that was guided by the expert scanpath averaged 64.5%, the group that was guided by the expert selections averaged 68.8%, and the group that was guided by the adversarial scanpath averaged 55.5%. These values were obtained by averaging the accuracy for all participants in a group over all images used in the training session.

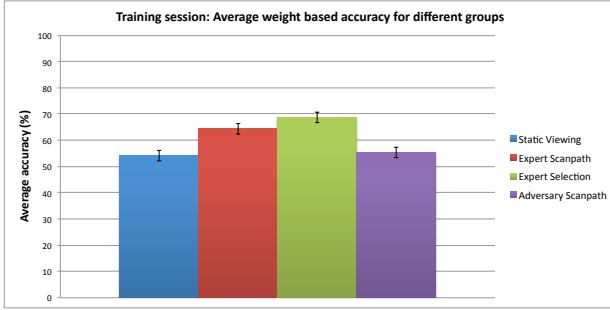


Figure 4: Average weight-based accuracy (equation 1) for different groups during the training session. The error bars represent standard error.

The averages show that the participants from both expert guided groups were more accurate than the participants from the static group. Independent-samples t-tests reveal that this effect was significant in both cases and not due to chance:

$$t(198) = 3.227; p < 0.05 \text{ (expert scanpath vs. static)}$$

$$t(198) = 4.530; p < 0.05 \text{ (expert selection vs. static)}$$

Figure 5 shows the binary classification based accuracy (equation 4) for the various groups of participants during the training session. The static group averaged 53.6%, the group that was guided by the expert scanpath averaged 68.3%, the group that was guided by the expert selections averaged 64.7%, and the group that was guided by the adversarial scanpath averaged 52.0%. When using this accuracy measure we again observe that both expert-guided groups perform better than the static group.

These observations are not surprising as previous studies have already established that guiding attention to the relevant regions of a scene facilitates task completion (see section 2). With Subtle Gaze Direction however, there is the added benefit that the cues used to attract the viewer's attention have minimal impact on the viewing experience as they occur only in the viewer's peripheral vision and do not permanently alter the overall appearance of the image being viewed.

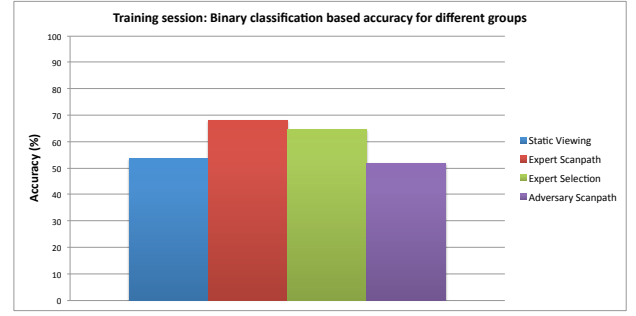


Figure 5: Binary classification based accuracy (equation 4) for different groups during the training session.

4.2 Short-Term Follow-up Session Results

Figure 6 shows the average weight-based accuracy (equation 1) for the various groups of participants during the short-term follow-up session. The static group averaged 51.3%, the group that was guided by the expert scanpath during the training session averaged 59.5%, the group that was guided by the expert selections during the training session averaged 63.1%, and the group that was guided by the adversarial scanpath during the training session averaged 53.9%. These values were obtained by averaging the accuracy for all participants in a group over all images used in the short-term follow-up session.

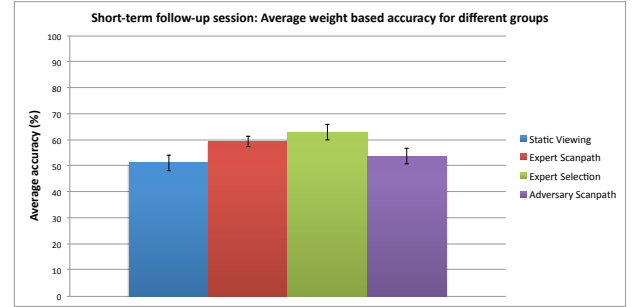


Figure 6: Average weight-based accuracy (equation 1) for different groups during the short-term follow-up session. The error bars represent standard error.

The averages show that the participants who were trained using the expert-guided approaches performed better than participants who were not gaze directed during training. This indicates that there are short-term lingering effects related to the use of gaze manipulation and that the participants are becoming sensitized to the method of training used. Independent-samples t-tests confirm that this effect was significant in both cases and not due to chance:

$$t(198) = 2.311; p < 0.05 \text{ (expert scanpath vs. static)}$$

$$t(198) = 3.189; p < 0.05 \text{ (expert selection vs. static)}$$

Figure 7 shows the binary classification based accuracy (equation 4) for the various groups of participants during the short-term follow-up session. The static group averaged 26.9%, the group that was guided by the expert scanpath averaged 45.8%, the group that was guided by the expert selections averaged 41.6%, and the group that was guided by the adversarial scanpath averaged 27.5%. When using this accuracy measure we again observe that both expert-guided groups continued to perform better than the static group in the short-term follow-up session.

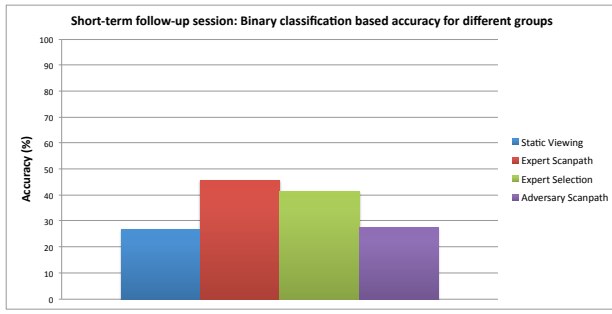


Figure 7: Binary classification based accuracy (equation 4) for different groups during the short-term follow-up session

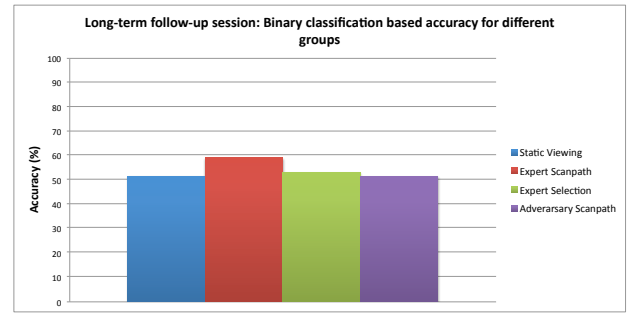


Figure 9: Binary classification based accuracy (equation 4) for different groups during the long-term follow-up session

4.3 Long-Term Follow-up Session Results

Figure 8 shows the average weight-based accuracy (equation 1) for the various groups of participants during the long-term follow-up session. The static group averaged 64.2%, the expert scanpath group averaged 64.2%, the expert selection group averaged 65.4%, and the adversarial scanpath group averaged 65.2%. These values were obtained by averaging the accuracy for all participants in a group over all images used in the long-term follow-up session. The averages show that there is little difference between the groups and independent-samples t-tests confirm that the differences are not significant. This suggests that there are no long-term lingering effects related to the use of gaze manipulation.

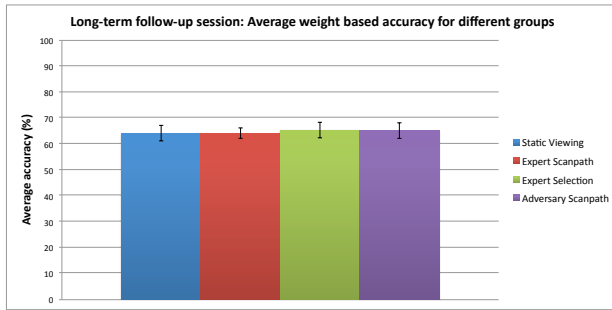


Figure 8: Average weight-based accuracy (equation 1) for different groups during the long-term follow-up session. The error bars represent standard error.

Figure 9 shows the binary classification based accuracy (equation 4) for the various groups of participants during the long-term follow-up session. The static group averaged 51.6%, the group that was guided by the expert scanpath averaged 59.1%, the group that was guided by the expert selections averaged 53.0%, and the group that was guided by the adversarial scanpath averaged 51.5%. When using this accuracy measure we again observe that there is little variation between the groups with the exception of the expert scanpath group which appears to be higher. Further analysis using the receiver operating characteristics (ROC) space (described below) shows however, that all of the groups are actually performing at the accuracy level equivalent to a random guess.

4.4 Receiver Operating Characteristic (ROC) Results

Figure 10 shows the sensitivity and specificity for each group of participants across all sessions. Using these values we can generate receiver operating characteristic (ROC) plots for each session as shown in Figure 11. ROC plots (i.e. sensitivity vs. 1-specificity)

	Static Viewing		Expert Scanpath		Expert Selection		Adversarial	
	Sensitivity	Specificity	Sensitivity	Specificity	Sensitivity	Specificity	Sensitivity	Specificity
Training	63.19	21.15	72.92	30.30	66.31	59.26	64.12	27.66
Short term	50.15	7.07	63.13	19.23	46.13	33.33	46.75	11.65
Long term	53.33	47.06	68.87	33.33	56.21	45.71	55.56	41.18

Figure 10: Sensitivity and specificity results for training, short-term follow-up, and long-term follow-up sessions.

help us assess the performance of the training techniques used. The blue dotted line at 45° represents the line of random guess. Any data above this line indicates a better-than-random classification and below this line a worse-than-random classification.

The ROC plot for the training session shows that the group guided by expert selection had a higher specificity compared to the other groups. The higher specificity may be attributed to the fact that the subjects associated the presence of the subtle modulations with the existence of abnormalities. Consequently when no modulations were present they were less likely to mark the image as being irregular. The ROC plot for the training session also shows that both expert-guided groups were more sensitive than the static and adversarial groups. This supports the accuracy findings described above.

The ROC plot for the short-term follow-up session shows that the expert-guided groups perform better than the control group and the adversarial group. Note however that both the sensitivity and specificity for all groups is lower than in the training session. This suggests that there was some common phenomenon affecting the subjects of all groups. We believe that this could be due to fatigue or boredom (recall that the participants had already viewed 25 images lasting a minimum of 10 seconds each - not including the time it took for them to respond to each image).

The ROC plot for the long-term follow-up session shows that the performance of all the groups lie on the random guess line indicating that there are no difference between the groups and that there are no long-term lingering effects associated with the training methods used. Notice also that the phenomenon that caused the lower sensitivity and specificity observed during the short-term follow-up session is no longer present after a week-long break.

5 Summary and Conclusion

In our experiment we explored whether subtle gaze manipulation is capable of improving the efficiency of mammography training. Using a weight-based accuracy measure we observed the following:

- For the training session, participants in both expert-guided groups performed significantly better than the participants in

Receiver Operating Characteristic (ROC) plot for each session

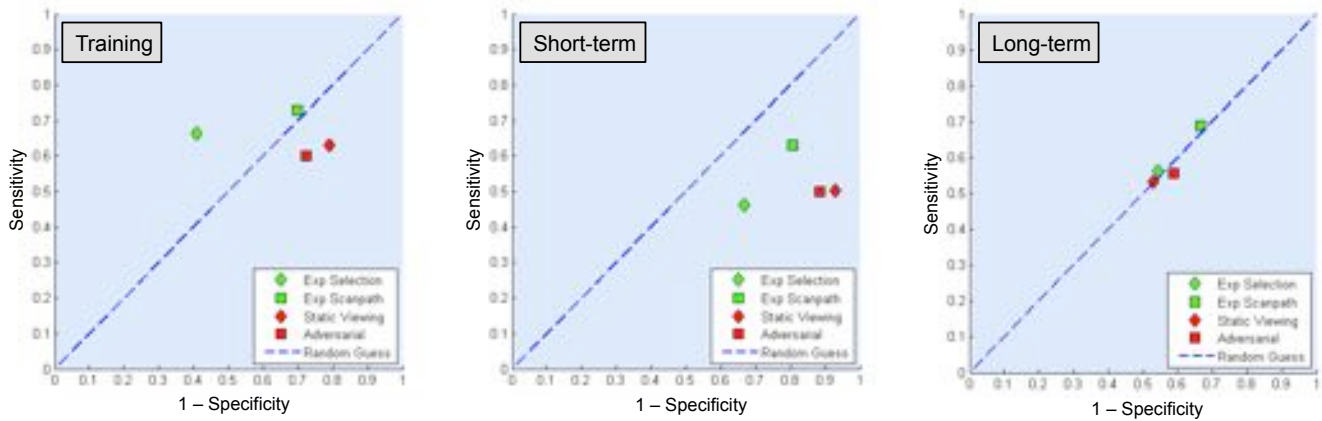


Figure 11: ROC plots for each session of the experiment for all groups. From left to right: training session, short-term follow-up session, and long-term follow-up session.

the static group.

- For the short-term follow-up session, participants in both expert-guided groups continued to perform significantly better than the participants in the static group.
- For the long-term follow-up session, no significant effects were observed.

Similar results were observed using the binary classification accuracy measure for all sessions. The general agreement between these two accuracy measures provides added assurance about the reliability of the results.

Based on the ROC plots we can conclude that actively guiding the subjects using the expert-based approaches results in better performance in terms of sensitivity and specificity. In the short-term follow-up session we noticed that both sensitivity and specificity of the group guided by the expert selections dropped dramatically compared to the group guided using the expert scanpath. For this reason, and for the fact that constantly modulating a single selection may become annoying, we conclude that it is better to guide the novices using the expert scanpath.

Overall, our findings indicate that using the Subtle Gaze Direction technique to actively guide novice participants along the scanpath of an expert significantly improves accuracy compared to static viewing. In addition to exhibiting great promise for digital mammography training as well as training on other medical image modalities such as X-rays, CT scans, PET scans, and MRIs, these observations have implications for a wide range of visual search and learning applications including:

- **Surgical Simulator Training:** Surgical simulation systems are becoming more common. Many of these systems are complex and require significant time and attention from the trainees. Gaze manipulation may facilitate more efficient training on these platforms.
- **Education and Learning:** Often the amount of information presented to students in educational settings, be it in the classroom or online, is overwhelming. Gaze manipulation can potentially be used to enhance the learning experience and benefit those who have difficulty paying attention during learning.
- **Aiding Symbolic Development in Children:** Children with early childhood developmental disorders often have difficulty

associating words to images and symbols. Subtle Gaze Direction can be used in a symbolic teaching system to assist these children in learning new symbols and images and associate them with the words they are hearing. The benefit of this approach is that it does not introduce any permanent visual cues which may complicate the learning process.

In addition to these applications, there are several other avenues for future work. We would like to explore if prolonged exposure (i.e. multiple training sessions using SGD as opposed to one) would have any long-term impact on the novices' performance. Further analysis of the eye-tracking data will be conducted to quantify how much deviation was present between the scanpaths of the novices and the expert. We also plan to repeat the entire experiment using participants with prior experience in analyzing mammograms (i.e. radiology trainees and residents). Finally, by combining computer vision techniques with the scanpath data of the expert that we previously recorded, we plan to explore the feasibility of an automatic scanpath predictor algorithm for mammograms.

Acknowledgements

This material is based on work supported by the National Science Foundation under Award No. IIS-0952631. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

References

- AGRESTI, A. 1989. A survey of models for repeated ordered categorical response data. *Stat Med* 8, 10 (Oct), 1209–24.
- AMERICAN CANCER SOCIETY. Cancer Facts & Figures. <http://www.cancer.org/Research/CancerFactsFigures/index>.
- BAILEY, R., MCNAMARA, A., SUDARSANAM, N., AND GRIMM, C. 2009. Subtle gaze direction. *ACM Trans. Graph.* 28 (September), 100:1–100:14.
- CARMODY, D. P., NODINE, C. F., AND KUNDEL, H. L. 1980. An analysis of perceptual and cognitive factors in radiographic interpretation. *Perception* 9, 339–344.

- CARMODY, D. P., KUNDEL, H. L., AND TOTO, L. C. 1984. Comparison scans while reading chest images. Taught, but not practiced. *Invest Radiol* 19, 462–466.
- CHEN, D. R., CHANG, R. F., AND HUANG, Y. L. 1999. Computer-aided diagnosis applied to US of solid breast nodules by using neural networks. *Radiology* 213 (Nov), 407–412.
- CHENTANEZ, N., ALTEROVITZ, R., RITCHIE, D., CHO, L., HAUSER, K. K., GOLDBERG, K., SHEWCHUK, J. R., AND O'BRIEN, J. F. 2009. Interactive simulation of surgical needle insertion and steering. In *Proceedings of ACM SIGGRAPH 2009*, 88:1–10.
- DESSER, T. S. 2007. Simulation-based training: the next revolution in radiology education? *J Am Coll Radiol* 4, 11, 816–824.
- DHAWAN, A., CHITRE, Y., AND KAISER-BONASSO, C. 1996. Analysis of mammographic microcalcifications using gray-level image structure features. *Medical Imaging, IEEE Transactions on* 15, 3 (jun), 246–259.
- DUGAS, M., TRUMM, C., STÄBLER, A., PANDER, E., HUNDT, W., SCHEIDLER, J., BRÜNING, R., HELMBERGER, T., WAGERSHAUSER, T., MATZKO, M., AND REISER, M. 2001. Case-oriented computer-based-training in radiology: concept, implementation and evaluation. *BMC Med Educ* 1, 5.
- FAWCETT, T. 2004. Roc graphs: Notes and practical considerations for researchers. Tech. rep.
- GAY, S. B., SOBEL, A. H., YOUNG, L. Q., AND DWYER, S. J. 1997. Processes involved in reading imaging studies: workflow analysis and implications for workstation development. *J Digital Imaging* 10, 40–45.
- GONZALEZ, R. C., AND WOODS, R. E. 2006. *Digital Image Processing (3rd Edition)*. Prentice-Hall, Inc., Upper Saddle River, NJ, USA.
- GRANT, E., AND SPIVEY, M. J. 2003. Eye movements and problem solving: guiding attention guides thought. *Psychological Science* 14, 5, 462–466.
- KRUPINSKI, E. A. 1996. Visual scanning patterns of radiologists searching mammograms. *Acad Radiol* 3, 2, 137–144.
- KRUPINSKI, E. A. 2000. The importance of perception research in medical imaging. *Radiat Med* 18, 6, 329–34.
- KUNDEL, H. L., NODINE, C. F., KRUPINSKI, E. A., AND MELLO-THOMS, C. 2008. Using gaze-tracking data and mixture distribution analysis to support a holistic model for the detection of cancers on mammograms. *Academic Radiology* 15, 7, 881–886.
- LI, H., KALLERGI, M., CLARKE, L., JAIN, V., AND CLARK, R. 1995. Markov random field for tumor detection in digital mammography. *Medical Imaging, IEEE Transactions on* 14, 3 (sep), 565–576.
- LITCHFIELD, D., AND BALL, L. J. 2011. Using another's gaze as an explicit aid to insight problem solving. *Q J Exp Psychol (Hove)* 64 (Apr), 649–656.
- LITCHFIELD, D., BALL, L. J., DONOVAN, T., MANNING, D. J., AND CRAWFORD, T. 2010. Viewing another person's eye movements improves identification of pulmonary nodules in chest x-ray inspection. *J Exp Psychol Appl* 16 (Sep), 251–262.
- MELLO-THOMS, C., BRITTON, C., ABRAMS, G., HAKIM, C., SHAH, R., HARDESTY, L., MAITZ, G., AND GUR, D. 2006. Head-mounted versus remote eye tracking of radiologists searching for breast cancer: a comparison. *Acad Radiol* 13, 2, 203–209.
- MUNIYANDI, M., COTIN, S., SRINIVASAN, M., AND DAWSON, S. 2003. Real-time pc based x-ray simulation for interventional radiology training. *Stud Health Technol Inform* 94, 233–9.
- QVARFORDT, P., BIEHL, J. T., GOLOVCHINSKY, G., AND DUNNINGAN, T. 2010. Understanding the benefits of gaze enhanced visual search. In *Proceedings of the 2010 Symposium on Eye-Tracking Research & Applications*, ACM, New York, NY, USA, ETRA '10, 283–290.
- RELAN, J., SERMESANT, M., POP, M., DELINGETTE, H., SORINE, M., WRIGHT, G., AND AYACHE, N. 2009. Volumetric prediction of cardiac electrophysiology using a heart model personalised to surface data. In *MICCAI Workshop on Cardiovascular Interventional Imaging and Biophysical Modelling*, 19–27.
- SHARPLES, M., JEFFERY, N. P., DU BOULAY, B., TEATHER, B. A., TEATHER, D., AND DU BOULAY, G. H. 2000. Structured computer-based training in the interpretation of neuroradiological images. *International Journal of Medical Informatics* 60, 3, 263–280.
- SOWDEN, P. T., DAVIES, I. R., AND ROLING, P. 2000. Perceptual learning of the detection of features in x-ray images: a functional role for improvements in adults' visual sensitivity? *J Exp Psychol Hum Percept Perform* 26, 1 (Feb), 379–90.
- SUCKLING, J., PARKER, J., DANCE, D., ASTLEY, S., HUTT, I., BOGGIS, C., RICKETTS, I., STAMATAKIS, E., CERNEAZ, N., KOK, S., TAYLOR, P., BETAL, D., AND SAVAGE, J. 1994. The mammographic image analysis society digital mammogram database. In *International Congress Series*, vol. 1069, 375–378.
- THOMAS, L., AND LLERAS, A. 2007. Moving eyes and moving thought: on the spatial compatibility between eye movements and cognition. *Psychonomic bulletin and review* 14, 4 (August), 663–668.
- WANG, R. F., AND SPELKE, E. S. 2002. Human spatial representation: insights from animals. *Trends in Cognitive Sciences* 6, 9, 376–382.
- XU, R., AND WUNSCH, D. 2005. Survey of Clustering Algorithms. *IEEE Transactions on Neural Networks* 16, 3, 645–678.