DOI: 10.2352/ISSN.2470-1173.2016.16HVEI-127

# Using Eye Tracking Metrics and Visual Saliency Maps to Assess Image Utility

*Laura E. Matzen, Michael J. Haass, Jonathan Tran, Laura A. McNamara; Sandia National Laboratories; Albuquerque, New Mexico*

## Abstract

*In this study, eye tracking metrics and visual saliency maps were used to assess analysts' interactions with synthetic aperture radar (SAR) imagery. Participants with varying levels of experience with SAR imagery completed a target detection task while their eye movements and behavioral responses were recorded. The resulting gaze maps were compared with maps of bottom-up visual saliency and with maps of automatically detected image features. The results showed striking differences between professional SAR analysts and novices in terms of how their visual search patterns related to the visual saliency of features in the imagery. They also revealed patterns that reflect the utility of various features in the images for the professional analysts. These findings have implications for system design and for the design and use of automatic feature classification algorithms.*

## Introduction

Human visual processing is guided by two parallel processes: bottom-up and top-down visual attention, also known as stimulus-driven and goal-oriented attention [1]. Bottom-up visual attention is captured automatically by the physical properties of a stimulus (e.g. contrast, color, motion) while top-down visual attention is allocated voluntarily and is driven by the viewer's goals and expectations (e.g. what information the person is looking for and past experience with where to find that information [2]). The cognitive processing underlying visual search is thought to have two main processes. In the first stage, which happens very rapidly when a person first sees an image, the visual cortex of the brain pre-attentively filters the stimulus, identifying the most visually salient regions (the regions with high bottom-up saliency). The information obtained at this stage of processing is then used to guide top-down visual attention, in which the viewer processes information serially by moving his or her eyes from one region of interest to another [3]. Regions with high bottom-up saliency may or may not be relevant to the viewer's task and goals, so there is a constant interplay between the two neural systems that guide visual attention and eye movements [4].

Since the brain is so highly attuned to processing visual information, most human-computer interfaces rely heavily on the capabilities of the human visual system. A great deal of effort is devoted to finding ways to visualize information so that humans can understand and make sense of it. This is particularly challenging when the information is multidimensional, such as in visualizations with a temporal component. Once a visualization has been developed, assessing its utility for a human analyst can prove to be even more challenging than developing the visualization itself. Ideally, a visualization should draw the viewer's attention to the information that is most useful to the viewer's task. In other words, there should be overlap between the features that are visually salient and those that are most important from a top-down, goal-oriented perspective.

In this paper, we describe a study in which we assessed the utility of images by comparing viewers' eye movements to maps of visual saliency and image features. The project focused on Synthetic Aperture Radar (SAR) and Coherent Change Detection (CCD) imagery. SAR is used in a variety of surveillance and mapping applications and the radar data is converted into a two-dimensional image (see Figure 1) for use by human analysts [5].

*Figure 1. Synthetic Aperture Radar (SAR) image of a baseball diamond. Image courtesy of Sandia National Laboratories, Airborne ISR.*

CCD images (Figure 2) are created by co-registering SAR images of the same scene and measuring changes in coherence that can reveal changes that have taken place in the scene over time [6].

*Figure 2. Coherent Change Detection (CCD) image highlighting several changes between images taken of the same scene at two different times. Image courtesy of Sandia National Laboratories, Airborne ISR.*

DOI: 10.2352/ISSN.2470-1173.2016.16HVEI-127

## Applied Studies of Imagery Analytic Workflows

The work described in this paper is part of an interdisciplinary family of research activities, in which Sandia National Laboratories researchers are examining how computational technologies influence the performance of professional imagery analysts. In this context, *imagery analysis* describes the perceptual and cognitive work of evaluating features of interest captured in two-dimensional images generated from remotely sensed data.

Visual inspection of imagery is an important component of work in a wide range of domains, from medical diagnostics to tactical military planning. However, the technologies used in imagery analysis have changed dramatically over the past couple of decades. Even as recently as the 1990s, "hardcopy" imagery and light tables comprised the major tools of imagery analysts. Importantly, the standards that express nominal thresholds for the detectability of feature classes in image products are rooted in psychophysical studies with imagery analysts using the hardcopy tool suite [7].

These days, however, computational or "softcopy" platforms are the main tools of imagery analysis. In many government workplaces, for example, light tables have disappeared as organizations have wholeheartedly embraced desktop computing systems and imagery analytic software. In a complementary fashion, computers have facilitated the development of image processing algorithms that can highlight or emphasize different features in a scene; for example, by exploiting changes in waveform characteristics to reveal ground changes in a scene-something that CCD imagery does very well. In short, the entire technological suite of imagery analysis has evolved dramatically over the past twenty years, with a wide array of electronic platforms and new image products available to support analytic workflows.

The imagery analytic revolution has raised questions about the functional equivalence of hardcopy vs. softcopy imagery for human visual detection tasks. A related issue is assessing the degree to which emerging image products might be used to support particular analytic workflows or feature detection goals. Finally, the rapid evolution of softcopy imagery also creates opportunities to examine how people interact with various types of image products as they are performing the visual cognitive work of professional imagery analysis. Of particular importance is the acquisition of perceptual skills, as people learn to "read" different types of imagery. We are particularly interested in understanding how imagery analysts learn to focus on the most valuable regions of an image product in relation to top-down analytic goals; and how these top down goals interact with bottom-up sensory and perceptual events driven by qualities of a given image product. Understanding these micro-processes is critical if we are to understand how people interact with imagery to establish a plausible narrative about the meaning of events captured in an image - for example, the import of footprints and tire tracks indicative of human activity in a rural area.

### *Current Research*

The objective of this project was to identify which features in SAR and CCD imagery drew the attention of experienced and novice analysts during a visual search and decision making task. Our aim was to inform system design by identifying differences in search patterns between groups with varying levels of experience and relating those patterns to features in the imagery and their visual saliency.

SAR imagery is well-suited for this type of investigation for several reasons. First, SAR and CCD images are superficially similar to optical imagery, but extensive training is required for analysts to learn to interpret SAR phenomenology correctly. This creates unique advantages for studying the influence of experience and top-down visual attention on visual search behavior. Professional imagery analysts who work with SAR perform visual search tasks using SAR and CCD images on a daily basis, developing extensive expertise and efficient visual search and decision making strategies. At the same time, there are many true novices who have never seen SAR or CCD images, yet the similarity between SAR imagery and optical images enables novices to complete visual search tasks despite their lack of domain-specific experience. Second, several feature detection algorithms have been developed for SAR and CCD images. These algorithms can identify specific terrain features and image regions that are particularly useful (or not useful) to the visual search task. This allows us to map the participants' gaze patterns against image features with high or low importance from the perspective of top-down attention. Finally, prior research has shown that visual saliency maps designed for optical imagery, such as the tool developed by Itti and Koch [8], are also applicable to SAR and CCD images because of their scene-like properties [9]. This allows us to contrast the participants' gaze maps with maps of the bottom-up visual saliency of the images. All of these characteristics make SAR a particularly useful domain for studying differences in visual search between experienced and inexperienced viewers, and how those differences relate to properties of the images.

In the study, we collected behavioral and eye tracking data from three groups of participants with varying levels of experience with SAR imagery, ranging from true novices to professional SAR imagery analysts. The participants completed a visual search and decision making task in which they were asked to search SAR and CCD images for targets. The targets were specific types of changes within the scenes. The gaze maps collected from the three groups of participants were then contrasted with visual saliency maps and with maps of automatically segmented terrain features. We also conducted an exploratory analysis in which the gaze maps were compared to a metric of change susceptibility within the scenes, described in more detail below.

We hypothesized that in situations where the decision-relevant information was not the most visually salient information, novice viewers would be more likely to get distracted. In contrast, experienced analysts are likely to have developed strategies to discount salient but irrelevant visual features. We predicted that the experienced analysts would focus on the most task-relevant regions of the images, regardless of their visual saliency. Comparing the performance and eye movements of groups with varying levels of experience allowed us to investigate the influence of top-down visual attention on task performance and to explore the interplay between expertise and image utility.

## Eye Tracking Study

### *Method*

#### Participants

Twenty-four participants completed a target detection task using SAR images while their eye movements were recorded at 60 Hz using the FaceLab 5 Standard system and EyeWorks software. Eight of the participants were professional SAR analysts who conduct visual search tasks using SAR imagery on a daily basis. Eight were non-analysts who work with SAR images regularly,

©2016 Society for Imaging Science and Technology
DOI: 10.2352/ISSN.2470-1173.2016.16HVEI-127

typically on a weekly basis. They had extensive knowledge of the domain, but do not typically engage in visual search tasks using the imagery. Most of the participants in this group were radar engineers who design and test SAR systems. We refer to this group as the "experienced non-analysts." The remaining eight participants were novices with no prior exposure to SAR imagery. All participants gave their written informed consent before participating in the study.

### Materials

Participants completed a target detection task using 20 pairs of images. Each pair consisted of a SAR image and a CCD image of the same scene. The CCD image was created by co-registering SAR images of the same scene over time and measuring changes in coherence that can reveal temporal changes [6]. Essentially, the SAR image provided viewers with contextual information about the scene and the CCD image provided viewers with information about the presence or absence of targets in the scene.

Half of the 20 image pairs contained a target and half did not. The targets were the same types of targets that the professional SAR analysts look for in their daily work. The experienced non-analysts were also familiar with the nature of the targets and view them frequently, although not in the context of a visual search task. The novices were not familiar with the domain, so they were shown examples of targets before beginning the experiment. They received instructions about what to look for to determine whether or not a target was present in the scene.

### Procedure

The participants completed a battery of general cognitive and visual search tasks in addition to the target detection task using SAR imagery [10]. In the target detection task, they were asked to stare at a fixation cross in the center of the computer screen. The cross remained on the screen for one second, and then one of the image pairs appeared on the screen. The SAR image was shown to the left of the fixation cross and the CCD image of the same scene was shown to the right of the fixation cross.

Participants were instructed to search the images for targets and to use a 1-4 scale to record their assessment of whether or not each scene contained a target. A response of "1" indicated that they were sure that there was not a target in the scene. A response of "2" indicated that they thought there was no target, but they were unsure. A response of "3" indicated that they thought there was a target present, but were unsure. A response of "4" indicated that they were sure that there was a target present. The SAR and CCD images remained on the screen until the participants responded or until 45 seconds had elapsed. The participants did not receive feedback about their answers until after the experiment was completed.

## *Results*

### Behavioral Results

The behavioral results showed that the professional imagery analysts were able to detect the targets more accurately than the novices and faster than both the novices and the experienced non-analysts. The analysts responded correctly to 74.4% of the trials, on average, with an average reaction time of 9.5 seconds. The experienced non-analysts responded correctly to 70.0% of the trials with an average reaction time of 14.5 seconds. The novice participants responded correctly to 56.9% of the trials with an average reaction time of 22.4 seconds.

One-way ANOVAs showed that the groups differed significantly in both their average accuracy ($F(2,21) = 4.62$, $p < 0.03$) and their average reaction times ($F(2,21) = 11.98$, $p < 0.001$). Post-hoc t-tests showed that the analysts had significantly higher accuracy ($t(14) = 2.95$, $p < 0.01$) and faster reaction times ($t(14) = 4.34$, $p < 0.001$) than the novices. The experienced non-analysts also had significantly higher accuracy ($t(14) = 2.14$, $p < 0.03$) and reaction times ($t(14) = 2.57$, $p < 0.02$) than the novices. The accuracy of the analysts and experienced non-analysts did not differ significantly ($t(14) = 0.73$), but the analysts had significantly faster reaction times ($t(14) = 2.93$, $p < 0.01$).

### Eye Tracking Results

Two participants, one from the novice group and one from the experienced group, were excluded from the eye tracking data analysis due to noisy data. A region of interest (ROI) was

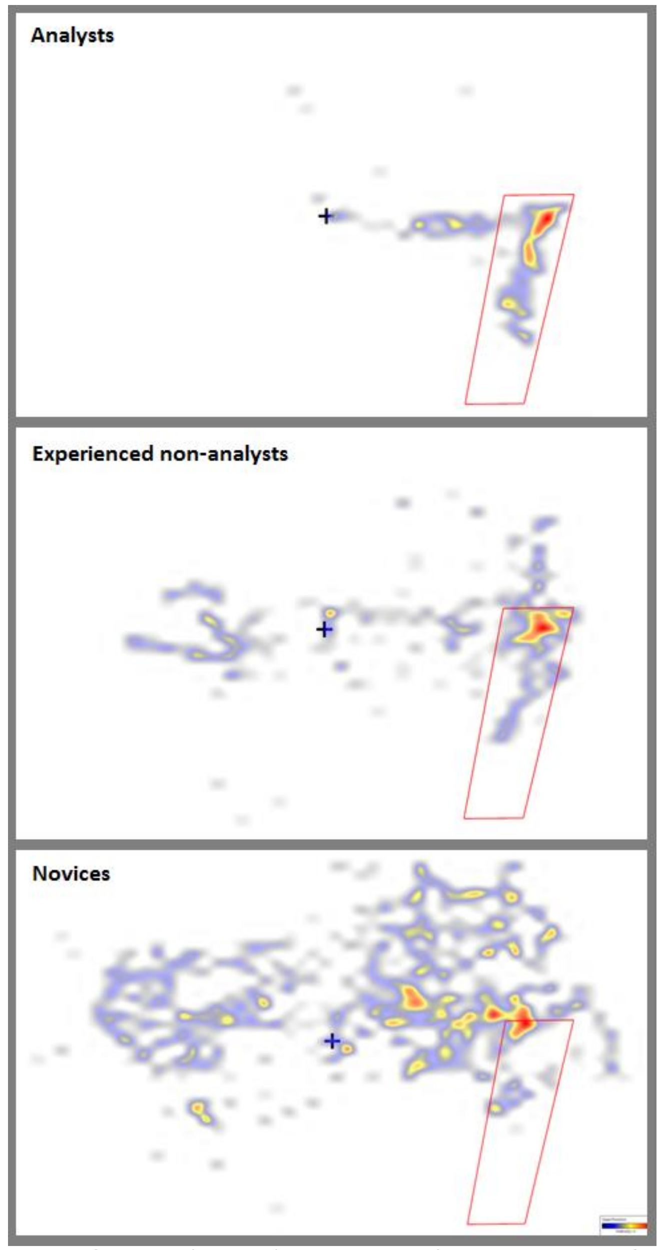


*Figure 3. Gaze maps for each of the three groups of participants with the ROI indicated in red.*

©2016 Society for Imaging Science and Technology
DOI: 10.2352/ISSN.2470-1173.2016.16HVEI-127

demarcated around each target that contained the target itself plus a buffer intended to represent a person's useful field of view (approximately 90 pixels on each side of the target).

The time to first fixation in the ROI was calculated for each trial in which a target was present. The average time to the first fixation in the ROI was 5.3 seconds for novices, 3.0 seconds for experienced non-analysts, and 2.1 seconds for analysts. The difference between groups was significant ($F(2,19) = 9.21$, $p < 0.01$). Post-hoc t-tests showed that the experienced non-analysts and the analysts were both significantly faster than the novices ($t(12) = 2.41$, $p < 0.02$ and $t(13) = 4.36$, $p < 0.001$, respectively). However, the experienced non-analysts and the analysts did not differ significantly from one another ($t(13) = 1.53$, $p = 0.08$).

For each trial, we calculated the percentage of total fixations that occurred within the ROI. On average, 17.4% of the novice's fixations were in the ROI, compared to 25.3% for the experienced non-analysts and 38.9% for the analysts. The difference between groups was significant ($F(2, 19) = 8.08$, $p < 0.01$). Post-hoc t-test showed that the experienced non-analysts had a significantly higher percentage of fixations in the ROI than the novices ($t(12) = 2.47$, $p < 0.02$) and the analysts had a significantly higher percentage of fixations in the ROI than the experienced non-analysts ($t(13) = 2.13$, $p < 0.03$).

### *Discussion*

Working within their domain of expertise, the SAR imagery analysts and experienced non-analysts were both more accurate in their responses than the novices, who had not viewed SAR imagery before taking part in the experiment. In addition to their high accuracy, the analysts were faster than experienced non-analysts and novices, both in terms of overall task reaction time and in terms of the time to first fixation in the ROI. The analysts were highly efficient in their ability to identify the ROI, typically fixating in the ROI within two seconds of stimulus onset. They devoted a higher proportion of fixations to the ROI than either of the other groups.

The efficiency of the analysts indicates that their visual search performance is driven by top-down visual processing. The analysts were able to rapidly triage the information in the imagery, zeroing in on the task-relevant information in the ROIs. In the analyses described below, we contrasted the gaze maps of the analysts and novices with other information about the content of the scenes, including bottom-up visual saliency and automatically detected terrain features. These analyses allowed us to further tease apart the contributions of bottom-up and top-down visual processing to the participants' visual search performance.

## Comparison of Gaze Maps to Saliency Maps

In order to compare the visual search patterns of the participant groups to visual properties of the imagery, gaze maps were created for each stimulus using each group's tracking data. Following the approach of Wooding [11], the gaze maps were constructed by pooling the raw eye tracker samples over all subjects in each group (i.e. analysts, experienced non-analysts and novices) and accumulating a two dimensional Gaussian function at each point. The standard deviation of the Gaussian function was defined to equal a two degree field of view (90 pixels) at the average viewing distance.

Visual saliency maps for each stimulus where created using the Itti and Koch model [12] as implemented in Harel's Graph Based Visual Saliency Toolbox [13]. The Itti and Koch model decomposes images into three feature sets that are based on processes in the human visual cortex: color, orientation and intensity. These feature sets are constructed at multiple scales using Gaussian pyramids. Areas of the image with the greatest differences in features across scales are assigned larger saliency values while areas with smaller differences in features across scales are assigned lower saliency values. In this study, participants were viewing two images placed side by side on the screen. Because the two image products have different mean intensity levels, we calculated the saliency maps separately for each image product to avoid saliency artifacts at the image product boundary.

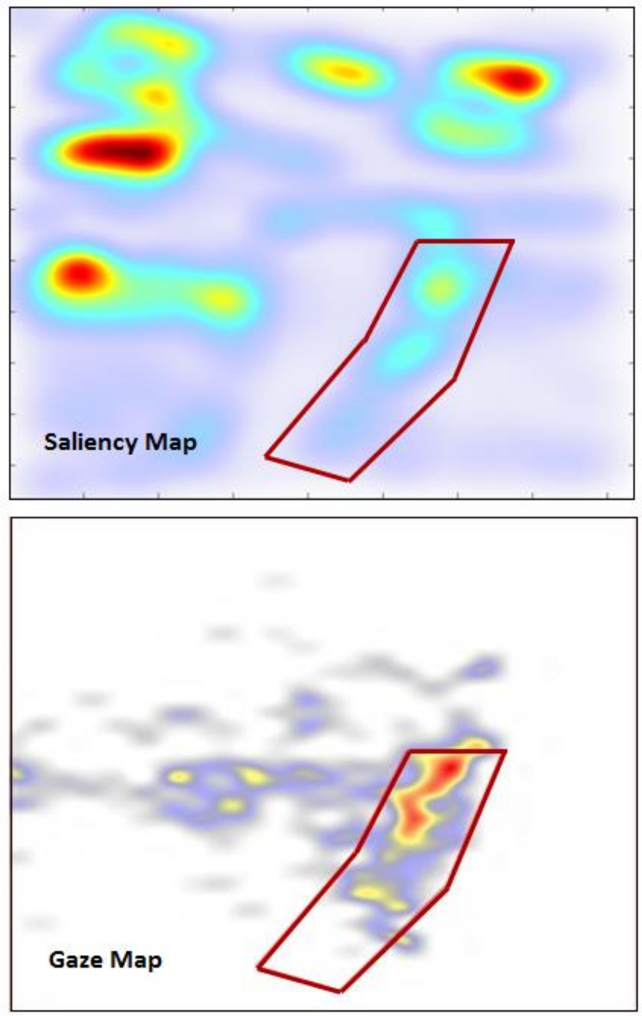


*Figure 4. The top panel shows the saliency map for one of the CCD stimuli used in the study and the bottom panel shows the analysts' gaze map for the same stimulus. The ROI is indicated in red.*

### *Results*

For each of the 10 stimuli in the eye tracking study that contained a target, we calculated the percentage of the overall visual saliency that fell within the ROI around the target. Then, for each group of participants, we calculated the percentage of gaze observations that fell within the ROI for that stimulus. For all of the target-containing stimuli, an average of 17% of the total visual saliency fell within the ROIs. For the professional analysts, an

©2016 Society for Imaging Science and Technology
DOI: 10.2352/ISSN.2470-1173.2016.16HVEI-127

average of 57% of the gaze observations fell within the ROIs, consistent with the behavioral finding that the analysts were very efficient in identifying the ROIs. The experienced non-analysts and novices had lower percentages of gaze observations in the ROIs, with 42% for the experienced non-analysts and 27% for the novices.

Correlations were calculated between the percentage of visual saliency in the ROI and the percentage of gaze observations in the ROI for each stimulus within each group of participants. The results showed that the correlation was significant for the novices ($R^2 = 0.71$, $p < 0.01$) and for the experienced non-analysts ($R^2 = 0.52$, $p = 0.01$). However, for the professional analysts, there was not a significant correlation between the percentage of saliency in the ROIs and the percentage of gaze observations in the ROIs ($R^2 = 0.02$).

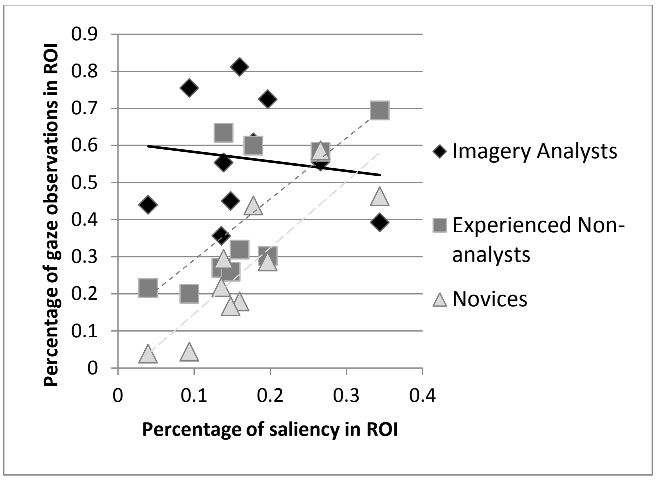


*Figure 5.The percentage of gaze in the ROI versus the percentage of saliency in the ROI for each participant group for every stimulus that contained a target.*

As discussed above, we hypothesized that professional analysts would rely on their past experience and on top-down visual attention to focus on the most task-relevant information, regardless of whether or not it was salient from a bottom-up perspective. The results of the eye tracking study and our comparisons between the gaze maps and saliency maps supported this hypothesis. To further explore the relationships between terrain features, visual saliency, and visual search, we compared the participants' gaze maps to automatically generated maps of image features. We chose to investigate two specific types of terrain features: SAR shadows and regions categorized as supporting change detection through a method called Index for Surface Coherence (ISC). These analyses and the preliminary results are described in the sections below.

## Comparison of Gaze Maps and Terrain Features

SAR imagery has unique properties that support a variety of methods for automatic feature detection. For example, specific terrain features can be detected and labeled by automated image processing algorithms such as superpixel segmentation and classification [14, 15]. Superpixel segmentation groups pixels by capturing image redundancy [16, 17]. A new method known as ISC extends this capability by identifying image regions in which the terrain features are more or less conducive to change detection [18].

We chose to focus our analyses on two types of automatically detected terrain features. First, we contrasted the gaze maps with maps of SAR shadows. The shadows in SAR images have relatively low importance in target detection tasks, but have high visual saliency. We predicted that experienced analysts would ignore shadow regions while novices would be more likely to be distracted by their high visual saliency. Second, in an exploratory analysis, we contrasted the gaze maps with ISC maps representing regions of the images that were most supportive of change detection. We predicted that the analysts would devote more attention to the regions that were most likely to support change detection, particularly since they were being asked to complete a target detection task in which the targets were changes to the scene. In contrast, we predicted that novices would not have the experience needed to determine which regions were most valuable to completing the task, making them less sensitive to this metric.

### *Modulating Saliency Maps Using Terrain Features*

In order to test the analysts' and novices' ability to ignore the highly salient but low value shadows, we calculated the overlap between the participants' gaze maps and the saliency maps with and without the shadows. First, algorithms were used to segment [14] the stimuli used in the eye tracking study into superpixels and to classify [15] the shadow superpixels.

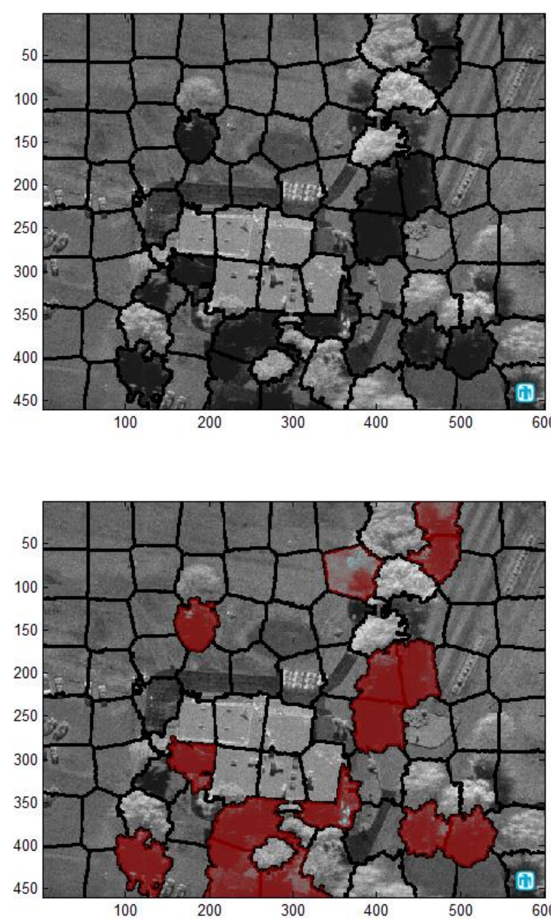


*Figure 6. The top panel shows a superpixel segmentation of a scene and the bottom panel shows superpixels classified as shadow regions in red.*

DOI: 10.2352/ISSN.2470-1173.2016.16HVEI-127

Next, modified saliency maps were created in which the superpixels identified as shadow regions were masked out, as shown in Figure 7.

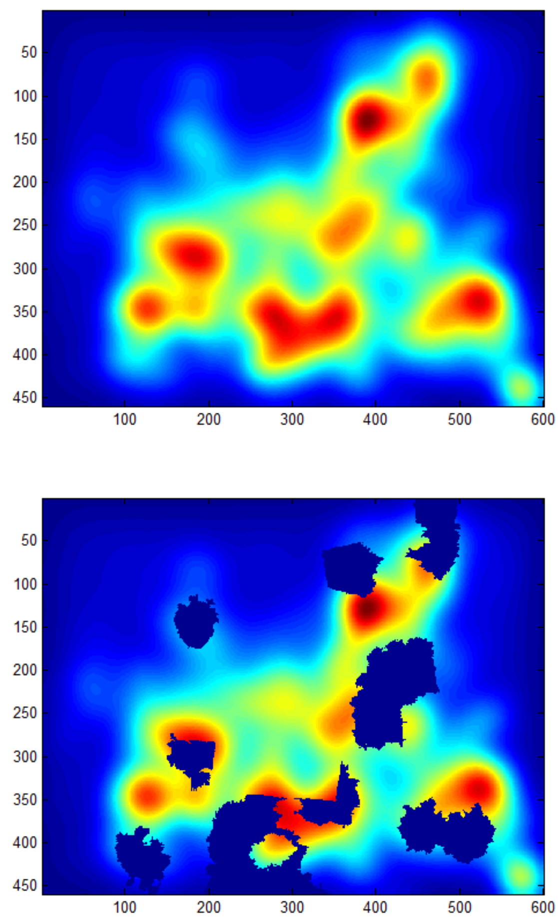


*Figure 7. The top panel shows the visual saliency map created from the SAR image in Figure 6. The bottom panel shows the masking of the superpixels classified as shadow regions.*

The gaze maps were compared to the original and masked saliency maps using the linear correlation coefficient (CC) metric. CC has been used in prior studies to measure performance of saliency estimation algorithms by comparing saliency maps to human gaze maps [19]. CC is a measure of the strength of a linear relationship between a gaze map (G) and a saliency map (S)

$$CC(G,S) = \frac{cov(G,S)}{\sigma_G \sigma_S} \quad . \qquad (1)$$

When CC is close to ±1, there is almost a perfectly linear relationship between the human gaze map and the predicted saliency map.

A subset of the eye tracking data (three analysts and three novices) was used to test the effects of masking shadows out of the saliency maps. For the analysts, masking the shadow regions improved CC agreement between saliency and gaze maps by a factor of 3.3 times. For the novices, masking the shadow regions *reduced* CC agreement by only 0.95 times.

These results provide further evidence to support our finding that professional analysts successfully relied on top-down visual attention, largely ignoring regions that were not relevant to the target detection task even if they were highly visually salient. The approach developed here could be applied for any other terrain features, allowing system designers to conduct a detailed analysis of how much experienced and novice users rely on each feature when completing a particular task. This could be a powerful method for assessing image quality by testing the relative contributions of each image feature to both the visual saliency of the scene and to the users' task performance.

### Comparing Gaze Maps to the Index of Surface Coherence

As discussed above, CCD images provide a method for observing changes in a scene that would otherwise be undetectable to the human eye [20]. By using multiple SAR collects, the magnitude and phase difference between each collect can be utilized to detect changes in a SAR image. However, the method used to calculate this change product is agnostic to the underlying terrain on which the calculation is made. Some features (such as walls) are stationary and not susceptible to change, appearing as areas that cohere perfectly in the CCD images. Other features, such as vegetation, have low coherence due to their random geometries and continuously show up as changes in the CCD product. Both types of features can be distracting to an analyst or algorithm looking for changes of interest (i.e. areas of low coherence in the scene that typically have high coherence). Discerning changes of interest in natural scenes requires training for human analysts and a better understanding of the underlying terrain for algorithms.

A new method to address this issue creates maps of the Index of Surface Coherence (ISC) for SAR images. These maps can be used to mask a CCD product and eliminate the areas that do not support detection of changes of interest. To create these maps, a long-term observation of an area is utilized to acquire the underlying nature of the terrain. With many observations of the same area over a period of time, a stack of images can be created. By registering all of the images and taking the median of each pixel in the stack, a stable representation of the area is observed. Using a median radar cross section (RCS) and median CCD product, the terrain in the area can be classified according to its coherence properties. The median RCS (MRCS) and median CCDs (MCCD) images are segmented into superpixels using the SLIC superpixel segmentation, which allows a user to define how compact the superpixel appears and the number of superpixels in the image. This allows a user to create a nearly uniform grid of pixel groups [14, 17]. A truly uniform segmentation would provide pixel groups and reduce the computing complexity, but the pixels in those groups would be visually and statistically very dissimilar.

After the median MRCS and median MCCD images are segmented, a training process is used in which terrain types that support change detection are identified and a subset of superpixels capturing each terrain type is chosen. In this study, approximately 20 superpixels consisting of 500 pixels for each terrain type were selected. For each data type, a distribution curve is generated for both the MRCS and MCCD products. The distribution curve is generated by fitting common distribution types (Gamma, Beta, Log-Normal, Exponential, and Gaussian) to the each data type's scaled histogram data. The distribution type, distribution parameters, and scaling are saved to represent each terrain type.

©2016 Society for Imaging Science and Technology
DOI: 10.2352/ISSN.2470-1173.2016.16HVEI-127

With the training finished, new images can be evaluated by segmenting the image into superpixels and comparing each superpixel in the image to the previously trained data. For each superpixel in the image, its pixels are scaled and fit with the distribution according to each terrain types training data. The distribution curve of the superpixel is then compared to the terrain type's distribution curve using Kullback-Leibler (KL) Divergence to get a similarity score. Using probabilistic fusion [21, 22], the KL scores of the MRCS and MCCD images are translated into p-scores which can then be added despite the KL scores being statistically different. These added scores can then be used to form a heat map to indicate where an image is most likely to support change detection.

We conducted a proof-of-concept analysis in which an ISC map of one of the CCD images from the eye tracking study was compared to participants' gaze maps. To compare the image p-scores to the human gaze maps, we first created a set of 20 thresholded images (P) using the original p-score image and thresholding each pixel for thresholds 1,2,3,…20. We then calculated the CC metric for each thresholded image, $P_i$, compared to the gaze map from either the IAs or the novices.

$$CC(P_i, S_j) = \frac{cov(P_i, S_j)}{\sigma_{P_i}\sigma_{S_j}}$$

$$\textit{Where } i = 1,2,\ldots 20;\ j = 1(\textit{analysts}),\ 2(\textit{novices}) \qquad (2)$$

At the lower thresholds, the maps show only regions that never change, while at higher thresholds the maps show regions with increasing susceptibility to change. This analysis showed that the CC metric peaked for novices at a p-score threshold of 2 while peaking for experts at a p-score threshold of 7. Although exploratory, these results indicate that the gaze maps of the novices were relatively insensitive to the likelihood that a particular region would support change detection. They devoted their attention to terrain features that did not provide much support for change detection and therefore had low p-scores in the ISC map. In contrast, the analysts devoted more attention to regions that had higher p-scores and were likely to support change detection.

## Discussion

The results of this experiment revealed distinct differences between the visual search patterns of the participants in the three experience groups. Professional SAR imagery analysts were faster and more accurate in finding targets in a visual search task using SAR and CCD images. The results of the eye tracking study showed that the analysts were rapidly able to identify the ROI in the scenes containing targets and spent a significantly higher proportion of their time inspecting the ROI than the other groups of participants. The viewers with less experience, including non-analysts and true novices, spent more time viewing other regions of the images, which had a negative impact on their speed and accuracy.

To explore the relationships between the participants' gaze maps and the visual features of the imagery, we compared the gaze maps to bottom-up saliency maps and to maps of image features that were either irrelevant (shadows) or relevant (regions supporting change detection) to the task. While the gaze maps of the novices and experienced non-analysts were correlated with the bottom-up saliency of the images, the gaze maps of the professional analysts showed no such correlation. These results indicate that the less experienced groups were at least somewhat distracted by visual features that had high visual saliency but little relevance to the task. In contrast, the analysts focused their attention on task-relevant features, whether they were highly visually salient or not. In other words, the analysts' visual search processes appear to be driven primarily by top-down, goal-directed visual attention, while the less experienced participants were influenced more by bottom-up visual saliency.

The comparisons of the participants' gaze maps to automatically detected image features also supported this interpretation of the eye tracking data. We chose SAR shadows as an example of a visual feature that was highly salient but had little relevance to the task. When superpixels from shadow regions were masked out of the visual saliency maps, the match between the saliency maps and the analysts' gaze maps improved substantially. When the same masking was done for the novices, the match between the saliency maps and gaze maps was reduced. The comparison between the gaze maps and the ISC maps had a similar result. The highest match between the novices' gaze maps and the ISC maps was at a very low threshold, where the ISC map showed areas with little susceptibility to change. These areas are not very informative in a change detection task, but novice participants spent quite a bit of time looking at them. The analysts ignored those regions, focusing their attention on regions that were supportive of change detection and were therefore task-relevant.

The results of this study revealed information about what types of SAR and CCD image features are used by people with different levels of experience. By studying the professional analysts' approach to the visual search task and identifying the features and regions that they focus on, we were able to identify which features are most relevant to their real-world visual search tasks. This information can be used to inform system design and the design of new image products and image processing algorithms to support the analysts in their daily work. By comparing the professional analysts to experienced non-analysts and novices, we were also able to identify image features that might be distracting to less experienced viewers. This information can inform the training of new analysts. It can also help to validate new image processing algorithms. For example, the comparison between the participants' gaze maps and the ISC maps provided valuable feedback about the value of the ISC method for identifying regions that are relevant to the end users of the imagery. The threshold cutoffs identified by the gaze map comparisons can be used when deploying the algorithm to help analysts filter out potential false alarms.

The methods developed for this study could be applied in other domains to assess image quality in terms of how well the images support the end user's top-down goals. By approaching the problem from the perspective of human cognition, we were able to learn a great deal about the features of the images that did or did not support the end users' cognitive needs.

## References

[1] Y. Pinto et al., "Bottom-up and top-down attention are independent," *Journal of Vision*, vol. 13, pp. 1-14, 2013.

[2] C. E. Connor, H. E. Egeth, and S. Yantis, "Visual attention: Bottom-up versus top-down," *Current Biology,* vol. 14, pp. R850-R852, 2004.

[3] J. M. Wolfe, "Guided Search 4.0: Current progress with a model of visual search," in *Integrated Models of Cognitive Systems,* W. Gray, Ed. New York: Oxford, 2007, pp. 99-119.

DOI: 10.2352/ISSN.2470-1173.2016.16HVEI-127

[4] T. Ogawa and H. Komatsu, “Target selection in area V4 during a multidimensional visual search task,” *Journal of Neuroscience*, vol. 24, pp. 6371- 6382, 2004.

[5] A. W. Doerry and F. M. Dickey, “Synthetic Aperture Radar,” *Optics and Photonics News*, vol 15, issue 11, pp 28-33, 2004.

[6] A. W. Doerry, “SAR data collection and processing requirements for high quality coherent change detection,” in *Proceedings of SPIE: Radar Sensor Technology XII, April 2008, Orlando, FL,* K. I. Ranney and A. W. Doerry, Eds. Society of Photo-Optical Instrumentation Engineers, 2008.

[7] J.M. Irvine, “National Imagery Interpretability and Rating Scales (NIIRS): Overview and Methodology,” *Proc. SPIE 3128, Airborne Reconnaissance XXI*, 93, November 21, 1997

[8] L. Itti and C. Koch, “Computational modelling of visual attention,” *Nature Reviews Neuroscience*, vol. 2, pp. 194-203, 2001.

[9] M. Haass et al., “Saliency estimation for advanced imaging scenes using pixel statistics,” presented at European Conference on Eye Movements 2015, Vienna, Austria, 2015.

[10] L. Matzen et al, “Effects of Professional Visual Search Experience on Domain-General and Domain-Specific Visual Cognition,” *Foundations of Augmented Cognition: Lecture Notes in Computer Science*, vol. 9183, pp. 481-491, 2015.

[11] D. S. Wooding, "Fixation maps: Quantifying eye-movement traces," in *Proceedings of ETRA 2002*. ACM, 2002, pp. 31-36.

[12] L. Itti, C. Koch and E. Niebur, “A model of saliency-based visual attention for rapid scene analysis,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 20, no. 11, pp. 1254-1259, Nov 1998.

[13] J. Harel, C. Koch, and P. Perona, "Graph-Based Visual Saliency,” in *Proceedings of Neural Information Processing Systems (NIPS)*, vol. 19, pp. 545-552, 2006.

[14] M. M. Moya, et al, “Superpixel segmentation using multiple SAR image products” in *SPIE Defense+ Security*. International Society for Optics and Photonics, 2014, pp. 90 770R.

[15] M.M. Moya, et al., “Superpixel Classification for Signature Search in Synthetic Aperture Radar Imagery,” presented at Conference on Data Analysis (CoDA), Santa Fe, NM, 2014.

[16] X. Ren and J. Malik, “Learning a classification model for segmentation,” *Computer Vision, 2003 Proceedings. Ninth IEEE International Conference on*. IEEE, 2003, pp. 10–17.

[17] R. Achanta et al., “Slic superpixels compared to state-of-the-art superpixel methods,” *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 34, no. 11, pp. 2274–2282, 2012.

[18] J. Tran, “Index for surface coherence (ISC): A method for calculating change susceptibility in SAR change products,” in *SPIE Defense + Security, Radar Sensor Technology XX*, Paper 9829-60 [Accepted], 2016.

[19] A. Borji et al, "Quantitative Analysis of Human-Model Agreement in Visual Saliency Modeling: A Comparative Study." *IEEE Transactions on Image Processing,* vol. 22, no. 1, pp. 55-69, 2013.

[20] C. V. Jakowatz et al., *Spotlight-Mode Synthetic Aperture Radar: A Signal Processing Approach*. Boston, MA: Kluwer Academic Publishers, 1996.

[21] K. M. Simonson. (1998, Aug.). Probabilistic Fusion of ATR Results,” Sandia National Laboratories, Albuquerque, NM, SAND Report SAND98-1699.

[22] D. R. Cox and D. V. Hinley, *Theoretical Statistics*, Chapman and Hall, 1974.

## Author Biography

*Laura Matzen received her BA in linguistics and cognitive science from Rice University (2003) and her PhD in cognitive psychology from the University of Illinois at Urbana-Champaign (2008). Since then she has worked in the Cognitive Science and Systems group at Sandia National Laboratories in Albuquerque, NM. Her work has focused on using cognitive neuroscience methods to understand how humans process and remember information while performing complex reasoning tasks.*