Recurrence Quantification Analysis reveals Eye-Movement Behavior Differences between Experts and Novices

Preethi Vaidyanathan*
Rochester Institute of Technology

Jeff B. Pelz†
Cecilia O. Alm‡
Rochester Institute of Technology

Pengcheng Shi§
Anne R. Haake¶
Rochester Institute of Technology

Abstract

Understanding and characterizing perceptual expertise is a major bottleneck in developing intelligent systems. In knowledge-rich domains such as dermatology, perceptual expertise influences the diagnostic inferences made based on the visual input. This study uses eye movement data from 12 dermatology experts and 12 undergraduate novices while they inspected 34 dermatological images. This work investigates the differences in global and local temporal fixation patterns between the two groups using recurrence quantification analysis (RQA). The RQA measures reveal significant differences in both global and local temporal patterns between the two groups. Results show that experts tended to refixate previously inspected areas less often than did novices, and their refixations were more widely separated in time. Experts were also less likely to follow extended scan paths repeatedly than were novices. These results suggest the potential value of RQA measures in characterizing perceptual expertise. We also discuss potential use of the RQA method in understanding the interactions between experts' visual and linguistic behavior.

CR Categories: H.1.2 [User/Machine Systems]: Human Information Processing— [I.2.10]: Vision and Scene Understanding—Perceptual Reasoning

Keywords: recurrence quantification analysis, perceptual expertise, eye-tracking with medical images

1 Introduction

Experts have the ability to identify and discriminate a specific category of stimuli and through perceptual expertise can form efficient mental representations of information [Tanaka et al. 2005]. In knowledge-rich domains such as dermatology, perceptual expertise has significant influence on an expert's viewing behavior and on the diagnostic inferences made based on the visual input. Studies show that experts' eye movements are primarily influenced by their domain knowledge whereas eye movements of novices are primarily influenced by features in the visual environment itself [Tanaka et al. 2005]. Experts' ability to identify and evaluate important information with relevance to specific contexts is what makes them unique [Hoffman and Fiore 2007]. For example, in addition to perceiving the presence of certain cues, they also perceive the absence of other critical cues, performing meaningful integration of that information. Characterizing experts' perceptual and conceptual expertise by capturing their viewing behavior and spoken description will improve our understanding of how experts perform such complex cognitive tasks using their domain knowledge and this will benefit image informatics systems [Vaidyanathan et al. 2011; Li et al. 2012]. This work focuses on how the RQA method [Anderson et al. 2013] can be extended to understand the effect of perceptual expertise on eye movement patterns and its potential use in investigating the interactions between experts' eye movements and spoken description.

It is difficult for experts to express the cognitive processes they go through in a purely didactic manner, demanding new and effective means by which to elicit and codify perceptual expertise. One solution is to have the experts manually annotate the important areas in the image. While successful, this method is laborious and fails to obtain crucial information that has helped the expert arrive at the decision of which areas are important. To remedy this, we focus on eliciting perceptual expertise via analyzing eye movement behaviors.

Human perception results from a combination of bottom-up image feature-driven processes and top-down cognitive processes. Together, saccades and intervening fixations can be used to characterize an expert's eye movement behavior when viewing an image. Therefore it is critical to first understand whether and how experts' eye movement behavior differs from that of novices. Empirical studies demonstrate specific and predictable differences between novices' and experts' perceptual behavior [Li et al. 2012]. Research has shown that experts perceive and process more information in less time than the novices in their domain of expertise [Krupinski 2000]. However, perceptual skills exhibited by radiologists when searching medical images did not transfer to a more general task [Krupinski 2000]. Because the expert-novice distinction apparently applies within a domain, it is important to investigate expertise-related differences domain-internally.

This work investigates whether eye movement behavior of expert dermatologists differs from that of novices with no training and attempts to quantify them by analyzing the global and local temporal eye movement patterns using RQA [Anderson et al. 2013].

2 Method

Eye-Tracking Session: Eye movement data was collected for two groups of subjects recruited from the same geographical area with different domain-expertise levels, while they inspected dermatological images. The first group included 16 dermatologists (12 board-certified 'attendings' and 4 'residents') and the second group consisted of 13 undergraduate novices with no dermatology training. Set of 42 dermatological images was used for data collection. Images were presented Figure 1 to the subjects on a 1680x1050 resolution monitor attached to a 50 Hz SensoMotoric Instruments (SMI) eye tracking apparatus with reported accuracy of 0.5 degree for the collection of eye movement data. A nine-point calibration followed by a four-point validation was performed at the start of each trial and repeated every ten images. We used a double computer set-up

*e-mail: pxv1621@rit.edu
†e-mail: pelz@cis.rit.edu
‡email:coagla@rit.edu
§e-mail:spcast@rit.edu
¶e-mail:arhics@rit.edu

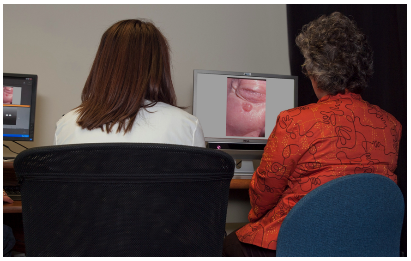

Figure 1: *Data-Collection Setup: an expert dermatologist (right) sitting in front of the eye-tracker. A physician assistant student (left) sitting next to the expert playing the role of apprentice for the master (expert) in the Master-Apprentice model.*

wherein one of the computers was used to present the image and the other was used to run SMI iViewX 2.4.19 and Experiment Center 2.3 software. The expert dermatologist group was instructed to "examine and describe each image verbally as if teaching the trainee to make a diagnosis based on the image." Physician assistant students were recruited to serve as 'trainees' in order to motivate the expert dermatologists through the modified Master-Apprentice model [Beyer and Holtzblatt 1997; Vaidyanathan et al. 2011]. We also recorded expert's spoken description. To motivate a similar underlying cognitive process in the undergraduate novice group, they were instructed to "examine and describe each image as if you are describing it over the phone to a dermatologist who cannot see the image but has to diagnose it."

Data Quality: Data from 4 experts and 1 novice were removed due to calibration errors and data file loss. Due to partial data loss for other observers the final image set consisted of 34 images that had been viewed by all participants. Hence, we report on results for 12 experts and 12 novices on 34 images.

2.1 Recurrence Quantification Analysis

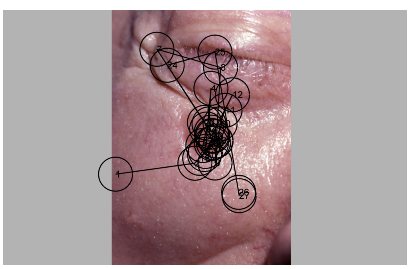

Figure 2: *Fixation sequence of a subject overlaid on the image. Numbers represent fixation order; circles represent a radius of 64 pixels.*

Recurrence analysis is a technique to investigate the time evolution of data series widely used in describing complex dynamic systems [Webber and Zbilut 1994]. Recently cross-recurrence analysis has been used to investigate the coupling between speakers' and listeners' eye movements [Richardson and Dale 2005]. [Anderson et al. 2013] describe and demonstrate how recurrence quantification analysis (RQA) and the associated measures can be used to differentiate eye movement behavior during different viewing conditions and image type finding significant differences. Compared to viewing conditions and tasks, differentiating eye movement behavior between populations is more complex. This work uses the RQA method and measures described in [Anderson et al. 2013] to investigate the differences in the spatial and temporal characteristics of

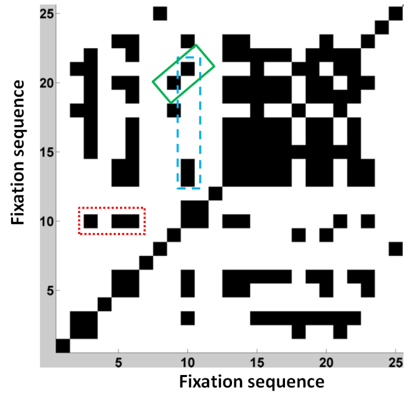

Figure 3: *Recurrence plot for the scanpath shown in Figure 2. The black squares represent recurring fixations which means they were within 64 pixel radius of each other. Examples of diagonal line for determinism (solid green box) and of horizontal and vertical lines for laminarity (dotted red and dashed blue boxes) are indicated.*

expert and novice eye movement behavior. The method allows us to analyze both global and local temporal fixation sequences and hence will be beneficial for understanding the eye movement behavior of a subject over a period of time. A brief description of the RQA method that takes fixation duration into account is provided below, and more details can be found in [Anderson et al. 2013].

For a fixation sequence f_i and corresponding durations $t_i, i = 1, \ldots N$, two fixations (i, j) are recurrent if they are within certain distance of each other and a recurrence plot (visualization technique) is created by assigning the sum of the corresponding durations to the position i, j:

$$r_{ij} = \begin{cases} t_i + t_j, & d(f_i, f_j) \leq \rho. \\ 0, & \text{otherwise.} \end{cases} \tag{1}$$

where *d* is the distance metric and ρ is the radius, i.e., the maximum distance between two fixations to be considered recurrent. Distance can be defined in various ways and we use Euclidean distance with radius ρ = 64 pixels, approximately $1.5°$ visual angle for our experimental setup. This value approximates the size of the fovea and tracker error in our eye tracker. For calculations only the upper triangle is taken into account since the recurrence plot is symmetric and the diagonal does not provide additional information.

These plots provide useful visualization of the temporal behavior of a subjects' eye movements as shown in Figure 3. The four RQA measures used by Anderson et al. and explored in this work are: *recurrence*, *determinism*, *laminarity* and *center for recurrence mass*. The sum of recurrences in the upper triangle is defined as $R = \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} r_{ij}$, and $T = \sum_{i=1}^{N} t_i$ is the sum of the fixation durations used for normalization purposes. Each RQA measure quantifies a certain aspect of the fixation sequence and is defined as:

Recurrence (REC): This measure can be thought of as representing (in percent) how often a location is refixated.

$$REC = 100 \frac{R}{(N-1)T} \tag{2}$$

Determinism (DET): Determinism measures how often subjects repeat short subsequences in their overall fixation sequence. Recurrent points in the plot can form diagonal lines (D_L) that indicate repetition of short-subsequences. For example if a subject looks back and forth between two locations creating a repeated pattern, those fixations would constitute a diagonal line. The reported results were calculated using L = 2 (other line lengths showed similar results).

$$DET = \frac{100}{R} \sum_{(i,j)\in D_L} r_{ij} \tag{3}$$

Laminarity (LAM): Recurrent points can also form vertical (V_L) and horizontal (H_L) lines. Since the plot is symmetrical, vertical and horizontal lines in the upper half of the plot are the same as horizontal and vertical lines in the bottom half, respectively. A vertical line (upper half) indicates detailed rescanning of a location that was previously fixated with a single fixation. On the other hand, a horizontal line (upper half) shows brief refixation to a location that was previously scanned in detail with multiple fixations. Together the horizontal and vertical lines are used to calculate what is called laminarity (LAM) representing revisited locations in the scene.

$$LAM = \frac{100}{2R}\left(\sum_{(i,j)\in H_L} r_{ij} + \sum_{i,j\in V_L} r_{ij} \right) \tag{4}$$

Center of recurrence mass (CORM): This measure quantifies the temporal distribution of the recurrent points. A small CORM value would mean that most of the refixations occurred very close in time whereas a large CORM value shows that refixations were widely separated in time. The recurrence and CORM measures are more global temporal fixation sequences whereas local patterns are captured by determinism and laminarity.

$$CORM = 100\frac{\sum_{i=1}^{N-1}\sum_{j=i+1}^{N}(j-i)r_{ij}}{(N-1)^2 T} \tag{5}$$

3 Results and Discussion

Common metrics used in investigating perceptual differences between two groups are median fixation durations and saccade amplitudes. Figure 4 shows that for the novice group these two metrics were lower than for the expert group. A two-tailed Student's t-test indicated that these differences were significant.

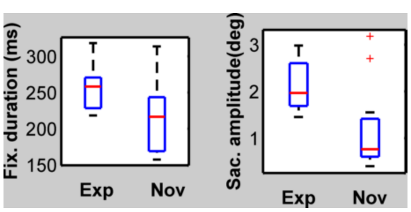

Figure 4: *Median fixation duration and saccade amplitude for the experts and novices along with the standard error. Notice that experts tend to have longer fixation durations and saccade amplitudes compared to novices.*

However, while these measures support a difference (Wilcoxon rank-sum test, $p < 0.05$), they do not take into account the crucial temporal order of the sequence. To better understand eye movement behavior we need to incorporate measures that are more descriptive. Figure 2 shows an example of a dermatology image overlaid with a subject's fixations and Figure 3 shows the corresponding recurrence plot. Using equations described in Section 2.1 we obtained $12(subjects) \times 34(images)$ recurrence plots for the two groups and the four RQA measures. Wilcoxon rank-sum test with $p = 0.05$ was used for significance testing to deal with the non-normal nature of the data.

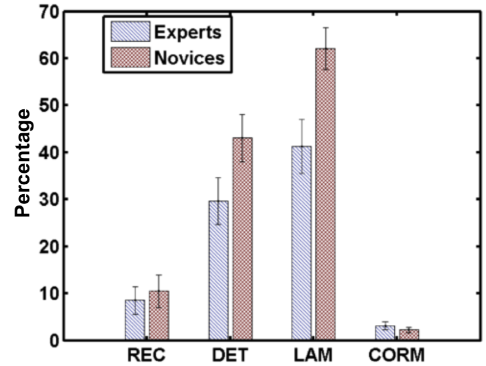

Figure 5: *Comparison of RQA measures between experts and novices: recurrence, determinism and laminarity are significantly lower for experts than novices; center of recurrence mass was higher for experts. These results indicate that experts refixate or repeat their scanpaths less often and that most of their refixations occur widely separated in time.*

Recurrence: We found significant difference between experts and novices with recurrence for experts being lower than recurrence for novices as shown in Figure 5. This shows that expert dermatologists tend to refixate previously inspected areas less often than novices suggesting that perceptual expertise probably helps experts to quickly obtain the required information pertaining to a region thereby requiring less rescanning.

Determinism: Our results suggest that experts repeat short sequences of fixations less often than novices. The rank-sum test indicates that the difference is significant and that determinism is higher among novices.

Laminarity: Results showed that experts were significantly lower in laminarity than were novices, indicating that they had fewer instances of repeated fixations within a relatively small region (defined by ρ).

Center of recurrence mass: CORM values among experts were significantly higher, indicating that experts refixated regions after longer intervals than did the novices. A probable reason is that experts fixate regions at the beginning of the trial and then revisit those regions towards the end when confirming their final diagnosis [Li et al. 2012].

Figure 5 shows that experts had lower recurrence, determinism, and laminarity. This suggests that experts are able to weigh a region's importance after a brief fixation, while novices exhibit multiple refixations. This could mean that experts are using their perceptual expertise to guide their gaze to maximize information intake. The high recurrence value along with higher number of fixations per second for the novice group suggests that novices are quickly scanning the scene with less strategy thereby having low fixation duration for individual fixations. On the other hand experts have longer fixation durations meaning they spend enough time on individual fixations to extract the useful information. This supports the low values of RQA measures except for CORM indicating involvement of different type of perceptual strategy than novices. When viewing dermatology images the high values of CORM could mean that experts initially inspect regions that are most informative or important, then fixate regions that might help them further in their diagnostic path followed by refixations to confirm their inferences. The sensitivity of the analysis to the parameters L and ρ were tested. The significance tests were unaffected by variations in line length.

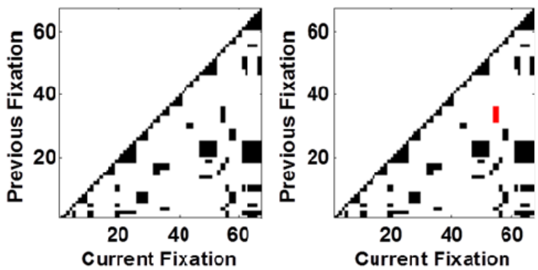

Figure 6: *(a) Recurrence plot with no verbal color coding. (b) Region highlighted in red shows subject was silent (SIL) during previous fixation (31 and 32) in a region and later uttered word related to the distribution of lesion (DIS) class while fixating (56) the same region.*

The number of fixations per second among the individuals can affect the RQA measures. In our work novices (μ = 4.5, SE = 0.02) had significantly higher number of fixations per second than experts (μ = 3.6, SE = 0.02).We used the bootstrap technique [Anderson et al. 2013] to test if the observed differences were purely by chance. All the RQA measures were significantly different for both the groups from those for random fixation sequences indicating the group behaviors were not random.

Motivated by these results we investigated if there were differences in the RQA measures between attending physicians and in-training residents. Due to unequal sample size an iterative test was conducted by comparing the 3 residents with 3 randomly selected attendings. Differences observed were not significant for all the iterations and depended on the individual attending. Larger sample size and statistically stronger tools are required to validate if differences exists between these two groups.

Ongoing work: Eye movements can help us understand experts' perceptual processes and language can reveal conceptual elements of cognitively demanding tasks. Studies have found differences in linguistic behavior between experts and in-training residents [Womack et al. 2012]. We are investigating how RQA can be used to understand the interactions between annotated verbal data and eye movements. Selected words uttered by a subject are annotated as one of the 7 classes as defined by an expert dermatologist. The lower triangle of the recurrence plot is then color coded according to the thought units. For example Figure 6 shows that a subject previously fixated a location in the image while being silent therefore annotated as *SIL* (silence) and later fixated a location within 64 pixel radius of the former while uttering a word annotated as *DIS* (distribution). This point in the recurrence plot is assigned a specific color, red in this case to differentiate from others. This is useful in visualizing the relation between fixations and classes. RQA measures can be employed directly or after further investigation to quantify these relations. Additionally, stereotypical behavior among many physicians can be elicited using a scanpath clustering method that uses a dot plot (similar to recurrence plot) [Goldberg and Helfman 2010].

4 Conclusion

In conclusion, we used RQA to analyze differences in spatial and temporal fixation patterns between dermatology experts and undergraduate novices with no training in dermatology. The results show that dermatology experts behave significantly different from untrained undergraduate novices in their eye movement behavior. We observed that experts refixate less often and do so widely separated in time. This further supports the importance of eliciting perceptual expertise to understand how experts perform complex tasks with the prospect to aid image informatics systems, build decision-support systems, etc. Furthermore this work shows that RQA is a valuable tool that can be used to tease apart the temporal structure of eye movements. We are currently investigating how the RQA method could be extended to understand the interactions between experts perceptual behavior, their spoken description and perceptual regions of interest.

5 Acknowledgements

This work is supported in part by of NSF Grant IIS-0941452 and NIH Grant R21 LM01002901. We thank Logical Images Inc. for providing the images and Dr. Cara Calvelli for valuable input into the design and interpretation of the data-collection experiment.

References

ANDERSON, N. C., BISCHOF, W. F., LAIDLAW, K. E., RISKO, E. F., AND KINGSTONE, A. 2013. Recurrence quantification analysis of eye movements. *Behavior research methods*, 1–15.

BEYER, H., AND HOLTZBLATT, K. 1997. *Contextual design: defining customer-centered systems*. Morgan Kaufmann.

GOLDBERG, J. H., AND HELFMAN, J. I. 2010. Scanpath clustering and aggregation. In *Proceedings of the 2010 Symposium on Eye-Tracking Research & Applications*, ACM, 227–234.

HOFFMAN, R., AND FIORE, S. 2007. Perceptual (re) learning: A leverage point for human-centered computing. *Intelligent Systems, IEEE 22*, 3, 79–83.

KRUPINSKI, E. 2000. The importance of perception research in medical imaging. *Radiation Medicine 18*, 6, 329–334.

LI, R., PELZ, J., SHI, P., ALM, C. O., AND HAAKE, A. R. 2012. Learning eye movement patterns for characterization of perceptual expertise. In *Proceedings of the Symposium on Eye Tracking Research and Applications*, ACM, 393–396.

RICHARDSON, D. C., AND DALE, R. 2005. Looking to understand: The coupling between speakers' and listeners' eye movements and its relationship to discourse comprehension. *Cognitive Science 29*, 6, 1045–1060.

TANAKA, J., CURRAN, T., AND SHEINBERG, D. 2005. The training and transfer of real-world perceptual expertise. *Psychological Science 16*, 2, 145–151.

VAIDYANATHAN, P., PELZ, J., LI, R., MULPURU, S., WANG, D., SHI, P., CALVELLI, C., AND HAAKE, A. 2011. Using human experts' gaze data to evaluate image processing algorithms. In *IVMSP Workshop, 2011 IEEE 10th*, IEEE, 129–134.

WEBBER, C., AND ZBILUT, J. P. 1994. Dynamical assessment of physiological systems and states using recurrence plot strategies. *Journal of Applied Physiology 76*, 2, 965–973.

WOMACK, K., MCCOY, W., ALM, C. O., CALVELLI, C., PELZ, J. B., SHI, P., AND HAAKE, A. 2012. Disfluencies as extra-propositional indicators of cognitive processing. In *Proceedings of the Workshop on Extra-Propositional Aspects of Meaning in Computational Linguistics*, Association for Computational Linguistics, 1–9.