Computer Vision and Image Understanding 151 (2016) 138–152

Contents lists available at ScienceDirect

Computer Vision and Image Understanding

journal homepage: www.elsevier.com/locate/cviu

Modeling eye movement patterns to characterize perceptual skill in image-based diagnostic reasoning processes

Rui Li*, Pengcheng Shi, Jeff Pelz, Cecilia O. Alm, Anne R. Haake

Golisano College of Computing and Information Science, Rochester Institute of Technology, 1 Lomb Memorial Drive Rochester, NY 14623, USA

ARTICLE INFO

Article history:
Received 24 February 2015
Accepted 1 March 2016

Keywords:
Visual attention
Multi-modal data
Diagnostic reasoning
Probabilistic modeling
Nonparametric Bayesian method
Markov chain Monte Carlo
Combinatorial stochastic processes

ABSTRACT

Experts have a remarkable capability of locating, perceptually organizing, identifying, and categorizing objects in images specific to their domains of expertise. In this article, we present a hierarchical probabilistic framework to discover the stereotypical and idiosyncratic viewing behaviors exhibited with expertise-specific groups. Through these patterned eye movement behaviors we are able to elicit the domain-specific knowledge and perceptual skills from the subjects whose eye movements are recorded during diagnostic reasoning processes on medical images. Analyzing experts' eye movement patterns provides us insight into cognitive strategies exploited to solve complex perceptual reasoning tasks. An experiment was conducted to collect both eye movement and verbal narrative data from three groups of subjects with different levels or no medical training (eleven board-certified dermatologists, four dermatologists in training and thirteen undergraduates) while they were examining and describing 50 photographic dermatological images. We use a hidden Markov model to describe each subject's eye movement sequence combined with hierarchical stochastic processes to capture and differentiate the discovered eye movement patterns shared by multiple subjects within and among the three groups. Independent experts' annotations of diagnostic conceptual units of thought in the transcribed verbal narratives are time-aligned with discovered eye movement patterns to help interpret the patterns' meanings. By mapping eye movement patterns to thought units, we uncover the relationships between visual and linguistic elements of their reasoning and perceptual processes, and show the manner in which these subjects varied their behaviors while parsing the images. We also show that inferred eye movement patterns characterize groups of similar temporal and spatial properties, and specify a subset of distinctive eye movement patterns which are commonly exhibited across multiple images. Based on the combinations of the occurrences of these eye movement patterns, we are able to categorize the images from the perspective of experts' viewing strategies in a novel way. In each category, images share similar lesion distributions and configurations. Our results show that modeling with multi-modal data, representative of physicians' diagnostic viewing behaviors and thought processes, is feasible and informative to gain insights into physicians' cognitive strategies, as well as medical image understanding.

1. Introduction

Solely behavioral variables from task manipulations, such as response time or accuracy, are insufficient to determine whether a particular cognitive process is engaged or whether a particular cognitive architecture theory is correct. Since visual attention, as a selective dynamic cognitive process, is dominated by knowledge, interest, and expectations of the scene [7,23], it is possible to acquire insight into some aspects of subjects' interests or cognitive strategies by analyzing their eye movement sequences while they are pursuing certain tasks in domains of expertise where perceptual skills are paramount. One key step to manifest perceptual skill and uncover underlying cognitive processes is to discover expertise-specific perceptual viewing behaviors and differentiate the stereotypical and idiosyncratic behavioral patterns that characterize a group of subjects at the same training level. Addressing this problem requires segmenting an eye movement sequence into a set of time intervals that have a useful interpretation, as well as summarizing the commonality of eye movement patterns shared within and between expertise-specific groups. Furthermore, these meaningful patterns enable us to uncover time-evolving properties of subjects' perceptual reasoning processes and to understand images at a domain-knowledge level.

* Corresponding author.
E-mail addresses: lr8032@gmail.com, rxlics@rit.edu (R. Li), pengcheng.shi@rit.edu (P. Shi), jeff.pelz@rit.edu (J. Pelz), coagla@rit.edu (C.O. Alm), anne.haake@rit.edu (A.R. Haake).

http://dx.doi.org/10.1016/j.cviu.2016.03.001
1077-3142/

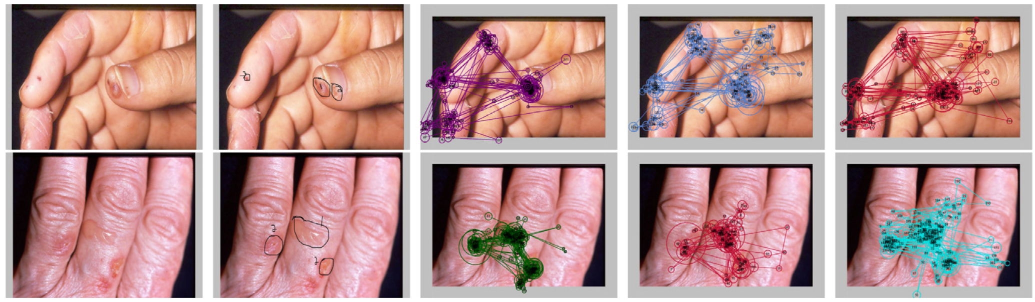

Fig. 1. Two example dermatological images examined by the subjects. The images from left to right are the original images, the primary and secondary abnormalities marked and numbered by an experienced dermatologist and three subjects' complete eye movement sequences acquired during the inspection process super-imposed onto the image, respectively. To visualize eye movement sequences, each circle center represents a fixation location and the radius is proportional to the duration time on that particular fixation. A line connecting two fixations represents a saccade.

Perceptual skill is considered to be the crucial cognitive factor accounting for the advantage of highly trained experts [24]. Experts generate distinctively different perceptual representations when they view the same scene as novices [38,45]. Rather than passively "photocopying" the visual information directly from sensors into minds, visual perception actively interprets the information by altering perceptual representations of the images based on experience and goals. By analyzing the whole sequences of fixation and saccadic eye movements from groups with different expertise levels or no expertise, significant differences in visual search strategies between groups show that expertise plays a key role in medical image examination. In such knowledge-rich domains, perceptual expertise is particularly valuable but poses challenges to its extraction, representation and application. Analyzing medical image understanding via traditional knowledge acquisition methods such as experts' marking on images (as shown in Fig. 1), verbal reports, and annotations is not only labor intensive but also ineffective because of the variability and noise of experts' performance [21]. In contrast, experts' perceptual skill is a valuable yet effortless resource worth exploiting, particularly for training and designing decision support systems where knowledge regarding the basic diagnostic strategies and principles of diagnostic-reasoning are desired [9]. We propose that this subconscious knowledge can be acquired by extracting and representing experts' perceptual skill in a form that is ready to be applied.

In this article we describe human-centered experimental approaches, which actively engage humans in the experimental process, to observe and record their perceptual and conceptual processing while inspecting medical images such as in Fig. 1. We subsequently profile the shared time-evolving eye movement patterns among physicians through our novel computational model, and also time-align eye movement patterns with semantic labels annotated by independent experts based on other dermatologists' verbal descriptions. We are then able to integrate these multimodal data towards understanding diagnostic reasoning processes and the dermatological images as well.

1.1. Visual attention

Attention is a critical contribution to perception in that focus of attention determines the portion of the sensory input from the external environment that will be readily available to perceptual processes. Complex visual information available in real-world scenes or stimuli exceeds the processing capability of the human visual system. Consequently, human vision is an active dynamic process in which the viewer seeks out specific information to support ongoing cognitive and behavioral activity [23]. Since high visual acuity is limited to the foveal region and resolution fades dramatically in the periphery, we move our eyes to bring a portion of the visual field into high resolution at the center of gaze.

A series of fixations and saccades are used to describe such eye movements. Fixations occur when the gaze is held at a particular location, whereas saccades are rapid eye movements used to reposition the fovea to a new location. Both the number of fixations and their durations are commonly assumed to indicate the depth of information processing associated with the visual fields. Saccade amplitudes, which are rarely considered in the analysis of eye movement data, may also influence some conclusions drawn from the visual processing [7,39,51].

Studies have shown that visual attention is influenced by two main sources of input: bottom-up visual attention driven by low-level saliency features which are image properties that are distinctively different from their surroundings [27], and top-down cognitive processes, guided by the viewing task and scene context, influence visual attention [8,32,48]. Growing evidence suggests that top-down information dominates the active viewing process and the influence of low-level salience guidance is minimal [7]. It is acknowledged that covert visual attention can be dissociated from overt visual attention manifested by eye movements [22]. Nonetheless, studies have shown that overt and covert visual attention are tightly coupled in complex information processing tasks, such as reading and scene perception [40]. In particular, saccades which direct gaze to a new location usually follow a shift of covert attention to this location, leading to speculation that covert attention serves to plan saccades [26]. These theoretical findings provide us with the support to pursue the underlying engaged cognitive processing based on observed eye movements.

The concept of the saliency map [27] is based on the Feature Integration Theory [49]. A saliency map characterizes the bottom-up distinctiveness of a particular location relative to that of other locations in the scene through its conspicuousness. One derived computational model concerned with understanding people's visual attention deployments on natural images was developed [26]. The researchers built a computational model to evaluate the saliency level of an image based only on extracted low-level visual features such as intensity, color, and orientation. According to the computed saliency map, they attempted to predict people's visual attention allocation. The model has been tested over various image sets, and its performance is generally robust. Particularly in regards to man-made images, its performance is consistent with observations in humans. More recent research has moved beyond using only low-level visual features to compute the salient image areas, and has begun to investigate multiple cognitive factors that influence visual attention. The main additional factors include one's expectations about where to find information and one's current information need, as well [25]. To further formulate these cognitive factors, image saliency was redefined in terms of the combination of both