Evaluating Human Visual Search Performance by Monte Carlo methods and Heuristic model

Giacomo Veneri, Elena Pretegiani, Pamela Federighi, Francesca Rosini, Antonio Federico & Alessandra Rufa

***Abstract*— Visual search is an everyday activity that enables humans to explore the real world. Given the visual input, during a visual search, it's required to select some aspects of the input in order to move to the next location. Exploration is guided by two factors: saliency of image (bottom-up) and endogenous mechanism (top-down). These two mechanisms interact to perform an efficient visual search. We developed a stochastic model, the "break away from fixations" (BAF), to emulate the visual search on a high cognitively demanding task such as a trail making test (TMT). The paper reports a case study providing evidence that human exploration performs an efficient visual search based also on an internal model of regions already explored.**

I. Introduction

Cognition works with vision by enabling the brain to create, maintain and update a representation of what is important during the exploration of a visual scene (goal-driven). Parallel, visual perception is also affected by external events (stimulus-driven), such as a salient region of the visual scene, flash and attended/unattended events. These two factors modulate attention in order to guide next exploration. In this dichotomy visual exploration is under control of the cognitive process (top-down, goal directed), and sensory stimulation factors (bottom-up, stimulus-driven).

A. Related works

Various formal models have been proposed in the last decade in order to describe the attentional selection mechanism: Feature Integration Theories (FIT; [1]), Guided Search (GS; [2]), Theory of Visual Attention (TVA; [3]) and a purely top-down model (WTA; [4]). FIT and TVA are very interesting models explaining the role of human attention during visual search; WTA and GS are devoted to assign saliency to locations in the visual field. Saliency is a linear combination of activations from retinotopic maps representing primitive visual features such as colors or orientation.

B. Objectives

In our study we propose a mathematical stochastic model based on Monte Carlo methods able to simulate human visual search in a high cognitively demanding task. In order to evaluate the performance of model we tested ten healthy subjects on trial making test (TMT, Fig. 2) and we compared the results with the model proposed.

Eye-Tracking & Vision Applications Lab. g.veneri@unisi.it-Departement of Neurological, Neurosurgical and Behavioral Science. University of Siena, Viale Bracci 2, 53100, Siena, Italy

Etruria Innovazione SCpA, via Banchi di Sotto 55, 53100, Siena, Italy

II. Method

A. Experiment design

In our study we explored the ongoing visual search of ten healthy subjects on the TMT (Fig. 2). The TMT is a neuropsychological instrument when number and letters should be connected to each other in numeric and alphabetic order (1-A-2-B-3-C-4-D-5-E). For each target (letter and number) we defined a region of interest (ROI) and we evaluated how humans directed next exploration according to the latest fixations distribution.

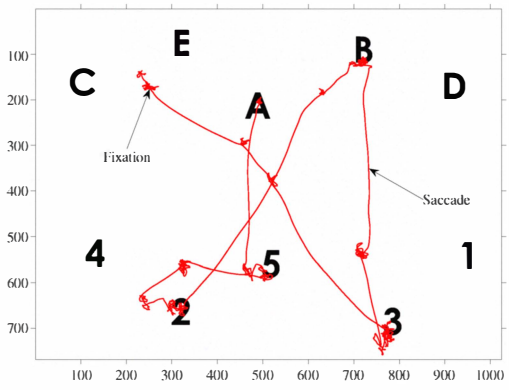

Fig. 1. Trail making: in the simplified form of the test, subjects were asked to trace a simple sequence such as 1-A-2-B-3-C-4-D-5-E with their gaze; peripheral vision was inhibited by a gaze contingent mechanism able to show the detailed information only around the current position of gaze. The numbers and letters appeared in a pseudo-random distribution (letters on top and numbers on the bottom) on a 1024x768px and 58x31cm screen for 30000 ms. Vision: human visual search is a process alternating the saccades and the fixations: during saccade ($duration \approx 40ms$ and $velocity \geq 500deg/sec$) the human moves eyes and during fixations ($duration \in (100, 1000)ms$) acquires information.

Subjects were asked to make a sequence by gaze; in order to avoid any influence of peripheral vision due to saliency map, we inhibited peripheral vision by a gaze contingent mechanism able to show detailed information only on an ellipse of 4 degree x 4 degree [5] and centered on the current point of gaze.

1) Preprocessing: Data were stored on a comma separated file imported into Matlab. A simple blink removal filter was applied; the filter substituted the blink values (pupil diameter equal to 0) or missing data (horizontal or vertical coordinates out of range) by linearly interpolated data. Large segments of missing data were marked (duration greater than 40ms) in order to avoid any analysis into these portions of data. Numbers and letters were sampled as pre-

978-1-4244-6561-3/10/$26.00 ©2010 IEEE