EYE MOVEMENT ANALYSIS IN VISUAL INSPECTION OF GEOSPATIAL DATA

Eugene Levin, Assistant Professor
William S. Helton, Assistant Professor
Robert Liimakka, Assistant Professor
School of Technology
Michigan Technological University
Houghton, MI 49931
elevin@mtu.edu
ralimakk@mtu.edu
wshelton@mtu.edu

Gennady Gienko, Senior Lecturer
School of Geography
The University of the South Pacific,
Suva, Fiji Islands
gennady.gienko@usp.ac.fj

ABSTRACT

Modern geospatial data acquisition systems deliver vast amounts of multi-domain remotely sensed data such as multi/hyper spectral imagery and LIDAR point-clouds. Unfortunately geospatial products automatically derived from source geospatial data are burdened with residual errors and artifacts which should be manually inspected, cleaned and corrected. These tasks become critical in many large-scale projects that require real-time processing of immense amounts of visual information and usually require manual post-processing or visual inspection of the source and/or derived data. The process of visual inspection could be divided in two general phases: perception and reaction. Scene perception comprises several steps such as visual search, feature selection and identification. Reaction reflects a decision made by an operator and usually involves other types of modalities (e.g. physical action such as mouse movements or typing). Human analysts perceive visual data through intensive movements of eyes which subconsciously select the most distinctive features in an image in order to reduce our overall ambiguity about the observed scene. A sequence of eye movements may then be understood within a framework of sequential accumulation of information. Whether or not all informative points are detected depends on both the observer’s current knowledge of the stimulus and the particular task. Dynamics of this information accrual process can be documented and quantified using analysis of eye movements during the process of natural, active visual perception.

This paper presents theoretical and practical investigations that have been made to illustrate the feasibility of 3D gaze tracking as a disambiguation tool in the process of visual search of a target in high resolution imagery.

INTRODUCTION

Currently, geospatial data acquisition systems provide data in multiple domains. This data is acquired in different spectral domains including various spatial and spectral resolutions, geometry (vertical, oblique and panoramic) and sensor models. Surveillance and reconnaissance involving geospatial imaging missions place a high demand on processing of multitemporal imagery. Data acquisition with periodicity varies from several minutes to several days. Photogrammetric and interpretative processing of this multi-domain data requires fusion and conflation. Numerous attempts in the development of methods and technology for multi-domain geospatial data fusion have occurred. The complexity and variety of these data sets and the processing of various types of terrain requires and in the future will require multiple human computer interactions. Furthermore, results of automated processing usually are inspected by human analysts for blunder detection and quality assurance purposes. Thus all these processes can be defined as a human-in-the-loop. Therefore, study of human geospatial analyst’s mental workflow may be helpful in research and development efforts resulting in acceleration of semi-automated and foundational automated image fusion systems. This paper presents a human-centric technology approach which efficiently deploys eye-tracking for derivation of
important information from eye movements of the human operator involved in visual geospatial data inspection.

Spatial and temporal data derived from eye movements, compiled while the operator observes the geospatial imagery, retain meaningful information that could be successfully utilized in geospatial image processing and interpretation. When the human analyst perceives geospatial data, intensive movements of the eyes subconsciously select the most distinctive features in an image in order to reduce the overall ambiguity occurring in the observed scene. It is possible to analyze features that attract the attention of the analyst and to measure the analyst’s reaction time by deploying eye-tracking technology. Specifically eye fixations identified in eye-tracking protocols could be interpreted as coordinates of the featured points of an object being observed before the final decision making process (such as mouse clicked measurement) was performed. This paper presents some experimental results in the study of human-analysts performance in geospatial data inspection.

GEOSPATIAL DATA PERCEPTION AND EYE MOVEMENTS

We may define the technical approach of interactive human geospatial data inspection as Eye-grammetry, which is the technology based on the principles of tracking the human eye movements while perceiving the visual scene (Gienko, Levin, 2005, Gienko, Levin, 2007). The virtual scene, imagined in the brain, is inherently related to neuro-physiological features of the human visual system and differs from the actual real world image or scene. The brain processes visual input by concentrating on specific components of the entire sensory area so that the interesting features of a scene may be examined with greater attention to detail than only peripheral stimuli. Visual attention, responsible for regulating sensory information to sensory channels of limited capacity serves as a “selective filter”, interrupting the continuous process of ocular observations by visual fixations. That is, human vision is a piecemeal process relying on the perceptual integration of small regions to construct a coherent representation of the whole (Duchowski, 2007, Mishkin et.al., 1983).

Neurophysiological literature on the human visual system suggests the field of view is observed through brief fixations over small regions of interest (ROIs). This allows perception of detail through the fovea. When visual attention is directed to a new area, fast eye movements (saccades) reposition the fovea. Foveal vision allows fine scrutiny of approximately 3% of the field of view but takes approximately 90% of viewing time, when subtending 5 deg of visual angle occurs. A common goal of eye movement analysis is the detection of fixations in the eye movement signal over the given stimulus or within stimulus ROIs (Carpenter, 1988, Mishkin et al., 1983, Gienko, Chekalin, 2004).

It has been found that humans and higher animals represent visual information in at least two important subsystems: the where- and the what systems. The where-system only processes the location of the object in the scene. It does not represent the kind of object; this is the task of the what-system. The two systems work independently of each other and never converge to one common representation. Physiologically, they are separated throughout the entire cortical process of visual analysis. (Mishkin et al., 1983, Goldman-Rakic, 1993).

When the brain processes a visual scene, some of the elements of the scene are put in focus by various attention mechanisms. When the brain analyses a visual scene, it must combine the representations obtained from different domains. Since information about the form and other features of particular objects can be obtained only when the object is foveated, different objects can be attended to only through saccadic movements of the eye – the rapid eye movements, which are made at the rate of about three per second, orienting the high-acuity foveal region of the eye over targets of interest in a visual scene (Posner, 1990). The characteristic properties of saccadic eye movements (or saccades) have been well studied (Duchowski, 2007, Carpenter, 1988). Saccades are naturally linked to fixations – relatively stable positions of the eye during a certain time. Varieties of research prove that visual and cognitive processing occurs during fixations (Just and Carpenter, 1984). The process of fixation identification is an inherently statistical description of observed eye movement behaviors and separating and labeling fixations and saccades in eye-tracking protocols is an essential part of eye-movement data analysis (Salvucci and Goldberg, 2000).

Typical tasks in visual geospatial data analysis include, but are not limited to retrieval of information, image interpretation, change detection, 3D surface reconstruction and updating of derived geospatial data such as GIS vector layers. In many application scenarios such as risk management or military targeting etc. it is necessary to perform these tasks in the realtime mode. Specifically all these tasks require visual data matching and fusing performed by a human analyst, who at the same time can be a Subject Matter Expert (SME) and under certain circumstances act as a Decision Maker. Thus, the solutions described below constitute some useful technology
empowering certain types of decision support systems, which in terms of Computer Sciences can be defined as a Human-Computer Symbiosis (HCS) in visual data analysis. Table 1 outlines main stages of a typical image analysis process which usually involves certain human intellectual and computerized recourses, employed simultaneously or concurrently, whichever is the most effective for a particular task:

Table 1. Human and computers in image analysis

<table>
<thead>
<tr>
<th>Stage</th>
<th>Agent</th>
</tr>
</thead>
<tbody>
<tr>
<td>General matching of observed scenes</td>
<td>brain</td>
</tr>
<tr>
<td>Tuned area matching</td>
<td>brain</td>
</tr>
<tr>
<td></td>
<td>computer</td>
</tr>
<tr>
<td>Disparity evaluation</td>
<td>brain</td>
</tr>
<tr>
<td></td>
<td>computer</td>
</tr>
<tr>
<td>Finding spot correspondence</td>
<td>brain</td>
</tr>
<tr>
<td>Object recognition</td>
<td>brain</td>
</tr>
<tr>
<td>Measuring (un)matched objects</td>
<td>brain</td>
</tr>
<tr>
<td></td>
<td>computer</td>
</tr>
<tr>
<td>Measurements registration</td>
<td>computer</td>
</tr>
<tr>
<td>Statistics</td>
<td>computer</td>
</tr>
<tr>
<td>Analysis</td>
<td>Brain</td>
</tr>
<tr>
<td></td>
<td>computer</td>
</tr>
</tbody>
</table>

The authors’ point of view on comparative effectiveness of human analysts and automated computer programs at a particular stage of image analysis prompts us to develop a human-in-the-loop technology for processing of visual geospatial data in the most efficient way. As humans perceive and process vast amounts of information about the environment through their visual system at extremely high speed, it is seems reasonable to combine the human ability and computational power of computers to build a Human-Computer Symbiosis (HCS) platform for processing of visual geospatial data. Such HCS can be based on registering of visual activity of an operator using techniques of real-time eye-tracking. While the human brain performs searches and analysis of visual data, the operator’s eyes subconsciously scan the visual scene. Such eye movements are driven by and indirectly represent results of internal processes of visual searching and matching, performed by the whole human visual system. Tracking and analyzing eye movements allows us to arrange a ‘sight-speed’ loop with the computer which should perform the rest of the tasks where computations and data storage are predominant.

There is multiple ongoing research devoted to the problems of knowledge extraction to accelerate geospatial image analysts’ work. One of the most interesting investigations (Kovalerchuk et.al. 2005) describes the Imagery Virtual Expert System (IVES) for imagery analysis that creates tools to capture imagery specific information and knowledge for IVES, and creates tools to foster intelligent consultation with IVES. Knowledge extraction methods in this system are based on Interactive on-the-fly recording of expert’s actions during geospatial datasets processing. These actions are obtained by recording the click-button events when the image analyst has made a decision. A similar approach is constituted by the DARPA program “Rapid Knowledge Formation” (RKF, 1999) formulating new requirements that include parallel entry of knowledge by teams of 25-50 individuals (end users) for test tasks such as crisis management and battle space understanding. Geospatial situation understanding is a key element of such a common picture.

All the processes referred above deploy the result of the decision. However, the perception stage which precedes decisions is not taken to consideration here. This stage is in fact important since human knowledge overcomes computing capabilities in many cases as it follows from Table 1 and can serve as a component of knowledge for the extraction system. Eye-tracking is a mechanism that may bring this methodology into geospatial data inspection processes. How can it be theoretically performed? The idea of using eye-tracking for geospatial Subject Matter Expert (SME) knowledge is based on discovering and formalizing associations and correlations between the content of the image observed, the expert’s eye-movement trajectories and particular tasks given to an
expert – whether it is targeting of specific objects in a set of multi-sensor and multi-temporal images, pure image classification or some other task involving geospatial data such as maps and GIS layers or other visual information. The system tracks the expert’s gaze directions while he selects and labels objects, then calculates parameters of the selected objects and generates preliminary classification rules by applying a dedicated knowledge mining approach. The challenge of this approach is to improve data mining procedures by means of the rules extracted from human analyst deploying eye-tracking system. The technological scheme of the eye-tracking based visual knowledge extraction is depicted in Figure 1.

![Figure 1. Eye-tracking based knowledge extraction process.](image)

Once the expert finishes natural process of image recognition, the full set of extracted rules is verified by reapplying those rules by automatic classification of the same source image. All automatically extracted and classified objects are then matched against the results of the expert’s natural work. Unmatched objects indicate inadequacy of the extraction rules. The expert interactively reviews and verifies results of image interpretation to discover the reasons of inadequacy which then will be used to adjust algorithms and parameters for automated extraction of decision rules. It is an iterative and interactive process, so the results will be immediately applied to the source image and the expert will be able to evaluate effectiveness of the newly added or modified rule. Once finished, the system will apply a “reverse rule” verification to cluster extracted rules and rate them in order to select the minimum set of major rules and knowledge sufficient for robust classification of particular objects. Experimental research given in following section demonstrates practical investigation of the described methodology.

**EYE MOVEMENT ANALYSIS IN MULTI-SENSOR REAL-TIME GEOSPATIAL VISUAL DATA CONFLATION**

While a numerous research has been done to investigate eye movements and attention on casual scenes, there is a lack of information about special investigations on the expert’s eye movements and visual attention, specifically when an operator is engaged in real-time visual data fusion with geospatial data inspection and quality assurance. In this paper we investigate the applicability of eye-tracking technology for numerical assessment of efficiency of an operator in fusion of multi-sensor and multi-geometry visual data in real time multi-sensor geospatial imagery conflation tasks. The goal of imagery conflation is the correlation and fusion of two or more images or geospatial databases. “The process of transferring information (including more accurate coordinates) from one geospatial database to another is known as ‘conflation’” (FGDC, 2000). Typically, the result of the conflation is a combined image produced from two or more images with: (1) matched features from different images and (2) transformations.

**ASPRS 2008 Annual Conference**
**Portland, Oregon • April 28 - May 2, 2008**
that are needed to produce a single consistent image. To define such a transformation the operator (or subject in the experiment) was challenged to define as many possible pairs of identical (tie) points while observing the experimental dataset.

The experimental setup has been designed in the way to protocol the expert’s eye movements and visual attention while simultaneously observing several multisensor imagery pairs from the existing database - “Reference Image Database” (RIB). An experimental setup enabled gaze-tracking simultaneously with recording of the mouse-clicks (“decisions”) on the same computer (see Fig. 2).

![Image of Seeing Machines Eye-tracker](image_url)

**Figure 2.** Experimental set: Seeing Machines Eye-tracker.

During the series of experiments the human subject consequently demonstrated four RIB objects and was challenged to click on tie-points on the left and right images. A single time line recorded eye-movement protocols and “mouse button down”, “mouse button up” events. Thus for the analysis, it was possible to superimpose terms of time and position on the screen with eye-fixations preceding the decision making documented as mouse-click events. RIB comprised images of different modality and complexity. An example of RIB-1 image pair with ordinary vertical photos of a stereo pair is given on Figure 3. Here, yellow points are tie-points which operator decided to select, black circles are eye-fixations with number and duration, which are connected by gaze trace lines. Object RIB-2 is composed of satellite imagery and a top-view rotated Google Earth generated object. Figure 4 represents object RIB-3 which is composed of top-view satellite imagery and an oblique view generated by means of the Google Earth program. And finally Figure 5 represents RIB-4 dataset which comprises multi-spectral satellite imagery and DEM data. It is obvious from Figures 3-5 that complexity of RIB objects was different and was increased from dataset to dataset.

In this paper an investigation of applicability of commonly used eye-tracking methods to quantitatively assess the efficiency of the operator’s ability to identify such tie-in featured objects was made by employing attention zones and fixation identification algorithms to eye-tracking protocols recorded during experimental sessions. Furthermore, since the decision is known it is possible to estimate the time needed for decision making. This time from one end is the “price” of one tie-point perception. From another aspect, the end ratio of perceptual time to the total time gives the unique opportunity to estimate the potential acceleration of visual geospatial data inspection automation by means of eye-tracking derived information as it is described in the previous section.
Figure 3. Object RIB-1 Vertical Images, tie-points fixations and saccades identified from scan paths in eye-racking protocols.

Figure 4. Object RIB-3 Vertical and Oblique images, tie-points fixations and saccades identified from scan paths in eye-racking protocols.

Figure 5. Object RIB-3 DEM and Satellite imagery, tie-points fixations and saccades identified from scan paths in eye-racking protocols.

ASPRS 2008 Annual Conference
Portland, Oregon • April 28 - May 2, 2008
Final statistical analysis of the experiment results is summarized in Table-2

<table>
<thead>
<tr>
<th>Object</th>
<th>Click-fixture delay</th>
<th>Total Fixation time( seconds )</th>
<th>Percent time fixated</th>
<th>Number of Fixations</th>
<th>Number of Input events</th>
<th>Tie-points pairs</th>
<th>Fixation per pair</th>
</tr>
</thead>
<tbody>
<tr>
<td>RIB-1</td>
<td>0.753</td>
<td>22.700</td>
<td>69.9</td>
<td>47</td>
<td>33</td>
<td>8</td>
<td>5.9</td>
</tr>
<tr>
<td>RIB-2</td>
<td>0.889</td>
<td>15.498</td>
<td>26.6</td>
<td>41</td>
<td>33</td>
<td>8</td>
<td>5.1</td>
</tr>
<tr>
<td>RIB-3</td>
<td>0.634</td>
<td>40.308</td>
<td>65.9</td>
<td>99</td>
<td>45</td>
<td>11</td>
<td>9.0</td>
</tr>
<tr>
<td>RIB-4</td>
<td>0.393</td>
<td>21.376</td>
<td>59.3</td>
<td>56</td>
<td>21</td>
<td>5</td>
<td>11.2</td>
</tr>
</tbody>
</table>

It is visible from Table 2 that fixations time vary from 26% till 70% depending on dataset complexity. Furthermore, for more complicated object single pair of the tie-points “costs” more fixations. Thus, analyst needs to spend more time in order to make decision. Experimental results confirm the feasibility of the eye-tracking effective deploying in geospatial visual scene analysis.

CONCLUSION AND FUTURE RESEARCH

Eye-tracking unique technologies utilize a natural, selective way of visual geospatial data analysis and quality control (guided by the human mind), which allows implementation of highly intelligent processing of multi-source imagery and fuzzy visual data, rapid 3D model reconstruction (including haze-clouded stereo pairs), multi-sources image conflation, and human-in-the-loop operator training. These technologies may be deployed in multiple application scenarios such as: manned/unmanned platform guidance navigation and control by means of human operator decisions, targeting and target recognition, emergency situations assessments and response. This paper presents an attempt to investigate applicability of eye-tracking methods and techniques for numerical assessment of efficiency of human actions in visual matching tasks to provide multi-source geospatial image conflation. Practical results of the research described above will be developed until technological implementation stages can be applied to:

- Attention sensitive / cognitive GIS systems;
- Human sight driven geospatial data inspection and quality control systems;
- Decision support systems empowered by SME knowledge extraction;

The future research is aimed at in-depth analysis of spatio-temporal patterns of visual behavior of operators, engaged in time-critical visual geospatial data inspection tasks.

REFERENCES

Goldman-Rakic, R.S., 1993. Dissociation of object and spatial processing domains in primate prefrontal cortex,

ASPRS 2008 Annual Conference
Portland, Oregon • April 28 - May 2, 2008


