

# ATTENTION DRIVEN APPROACH FOR GEOSPATIAL DATA UPDATE AND FUSION

**Eugene Levin**, Surveying Engineering Program Chair and Associate Professor,  
School of Technology, Michigan Technological University, Houghton, MI 49931  
[elevin@mtu.edu](mailto:elevin@mtu.edu)

**Anna Banaszek**, Associate Professor, Department of Real Estate Resources,  
University of Warmia and Mazury, Olsztyn, Poland  
[anna.banaszek@uwm.edu.pl](mailto:anna.banaszek@uwm.edu.pl)

**Jessica L. McCarty**, Research Scientist II, Michigan Tech Research Institute,  
Ann Arbor, MI 48105  
[jmccarty@mtu.edu](mailto:jmccarty@mtu.edu)

**Aleksander Zarnowski**, Professor, Department of Geoinformatics,  
Koszalin University of Technology, Koszalin, Poland  
[aleksander.zarnowski@uwm.edu.pl](mailto:aleksander.zarnowski@uwm.edu.pl)

## ABSTRACT

Updating maps is a process where the main goal is to bring the content of spatial databases in their electronic and hardcopy versions up to the most current state. The three main tasks undertaken in updating maps are change detection, image interpretation, and metadata and geographic correction (i.e., georeferencing). However, an analysis of map update technology shows that the most time consuming and difficult in the process of detecting changes is the map itself. When raw aerial or orthophoto imagery is deployed the main method is photointerpretation that uses photogrammetric instruments and tools, as well as a variety of thematic maps and data from geodetic measurements. In the last decade, a number of studies have developed methods for automatic and semi-automatic identification of terrain changes. This paper describes a research effort in establishing a novel approach for geospatial data actualization. We performed a series of experiments where multiple geospatial data sets were superimposed in real-time in a dynamic window driven by the analyst's attention as detected by an eye-tracking system. Specifically, edge detection and edge matching tasks were performed. Area-based or feature-based image matching (I2i) can be performed within specific windows (i.e., areas of interest) rather than globally. This paper describes the experiments in more detail and contains an initial comparison of this method's accuracy and productivity to other state-of-the-art updating techniques.

## INTRODUCTION

The analysis of the map update technology, the aim of which is to perform three main stages: detection of changes, their identification (photo-interpretation) and correction maps, shows that the most time-consuming and difficult is the process of detecting changes in the contents of the map itself (Armenakis et al, 2003). When using aerial images (digital or analog) or digital orthophoto, the main method is photointerpretation that deploys photogrammetric instruments and tools, as well as a variety of thematic maps and data from geodetic measurements. Nowadays, the commercial market has quite a selection of satellite digital images with high resolution and ground resolution of 0.5 m and less. It is possible to deploy digital images from IKONOS, WorldView, GeoEye, QuickBird, OrbView-3, KOMPSAT, and other satellite systems. Digital images obtained from satellites are characterized by high geometric and radiometric accuracy, and are used in the preparation of maps and their updates (Holland et al., 2006), the study of the environment, and provide valuable input to GIS.

Taking into account the interests of the high-resolution space-based systems vendors and their desire to use their products for updates of a range of maps and plans, we note that the spatial resolution is one of the most important factors determining the extent of updated maps. Therefore, aerial photography (analog and digital) can have a distinct advantage in the updating process. Modern analog and digital cameras (for example, ADS Leica 40, RC-30, DIMAC, UltraCam D and others) produce images with a resolution (GSD) of 5-12 cm. Modern systems of GPS / IMU can provide accurate delineation of linear elements of exterior orientation aerial photograph of approximately 10-15 cm. And the experimental work performed on the use of unmanned aerial vehicles (UAVs) shows that one can obtain digital images with a pixel size ranging from 1-3 cm.

In the last decade, a number of studies focused on the development of methods for automatic and semi-

automatic identification of terrain changes and included the following approaches:

- A comparison of two images taken at different times (image-image comparison). This operation is performed for each pixel of the registered sensor in the space of the electromagnetic radiation. The result will be obtained by the new image, which will be assigned pixel values related to the observed changes. The literature describes three main methods of image analysis based on the Boolean transformation of images and image classification.

- A comparison of the image to a map (image-map comparison) to quantify changes. This method differs from the method of comparing two images. The map is a cartographic model of reality in an orthogonal projection, has a fixed threshold to generalize object classes, and a system of symbols and names. Aerial imagery has a central projection and a certain resolution. The best known two approaches are:

- pre-processing the image (post-extraction change detection methods), for which the important and necessary procedures are radiometric and possibly geometric correction and image classification;
- deploying the map (map-guided method) to obtain quantitative data concerning the objects (e.g., their location, size, shape, etc.) that allow the user to reduce the search space and minimize the possibility of error.

In spite of the abovementioned efforts in the automation of geospatial imagery interpretation and understanding, there are still significant amounts of time and effort spent by human analysts on imagery interpretation. Thus, nowadays operational workflows of geospatial imagery processing and interpretation can be identified as a “human-in-the-loop” approach. Our research explores the potential use of novel human computer interaction (HCI) technology such as eye-tracking for the optimization of geospatial imagery processing in terms of productivity and accuracy.

## WHY 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 potentially 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. The task-specific use of gaze is best understood for reading text (O’Regan 1990) where the eyes fixate on almost every word, sometimes skipping over small function words. In addition, it is known that saccade size during reading is modulated according to the specific nature of the pattern recognition task at hand (Kowler and Anton 1987). Tasks requiring same/different judgments of complex patterns also elicit characteristic saccades (Dupont et al. 2014). The role of gaze has been studied by Land and Furneaux (1997) in a variety of other vision-motor tasks such as driving, music reading, and playing ping pong. In each case, gaze was found to play a central *functional* role, closely linked to the ongoing task demands. In summary, these studies strongly suggest that gaze control and saccadic eye movements play a crucial role in mediating visual cognition, in addition to compensating for peripheral acuity limitations. It is well known from visual attention theory that the correlation between perception and eye-movement is eye-fixation (Yarbus 1967).

This paper is devoted to the research of developing a potential eye-driven image interpretation human-computer system through the performance of a simple task that an image analyst performs every day: manipulating a cursor towards target objects. Baseline control for comparison of eye-tracking-based cursor movement was the regular mouse control of the cursor to the same set of targets.

## RESEARCH EXPERIMENT AND RESULTS

### Theory behind man-machine interaction efficiency numerical estimation

Paul Fitts (Fitts 1954) proposed a metric to quantify the difficulty of a target selection task which nowadays is used by cognitive scientists as a law to model human psychomotor behavior. This metric was based on an information analogy, where the distance to the target (D) is like a signal and the tolerance or width of the target (W) is like noise. The metric is Fitts’s index of difficulty (ID, in bits):

$$ID = \log_2 \left( \frac{2D}{W} \right) \quad (1)$$

Applying Fitts’s Law for the mouse or eye-driven targeting, the time to move and point to a target of width W at a distance A is a logarithmic function of the spatial relative error (A/W), according to MacKenzie and Buxton (1992):

$$MT = a + b * \log_2 \left( \frac{2A}{W} + c \right) \quad (2)$$

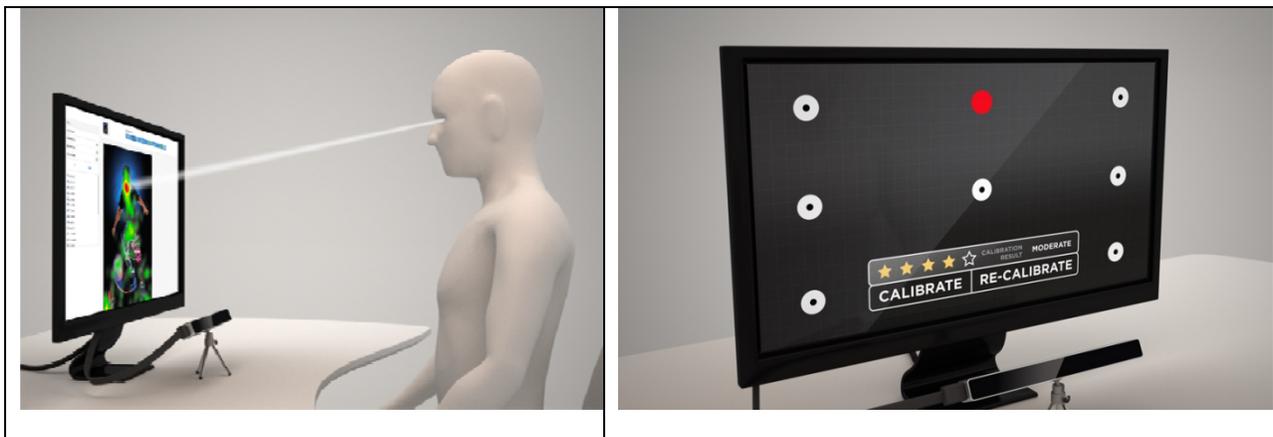
where

- $MT$  is the movement time
- $a$  and  $b$  are empirically determined constants, that are device dependent
- $c$  is a constant of 0, 0.5 or 1
- $A$  is the distance (or amplitude) of movement from start to target center
- $W$  is the width of the target, which corresponds to accuracy

The term  $\log_2(2A/W + c)$  is called the index of difficulty (ID). It describes the difficulty of the motor tasks.  $1/b$  is also called the index of performance (IP), and measures the information capacity of the human motor system. Thus comparative verification of numerical performance of the mouse versus eye targeting may give us an initial idea of estimating eye-driven man-machine interfaces efficiency.

### Open-sourced eye-tracking system

One of the most interesting trends in eye-tracking technology is a fact that this technology made an evolution from the exceptionally expensive systems deployed in medical field to the inexpensive ubiquitous systems that are widely applied, for example in controlling a computer and/or communication aids by people with disabilities (COGAIN 2015). Specifically, we deployed for our research an open source eye tracker, The Eye Tribe, which is available for under USD \$100 (The Eye Tribe 2015). The Eye Tribe Tracker is an eye tracking system that can calculate the location of a person's gaze by means of information extracted from the face and eyes. The eye gaze coordinates are calculated with respect to a screen the person is looking at, and are represented by a pair of  $(x, y)$  coordinates given on the screen coordinate system. A typical scenario is represented in Figure 1(a).



**Figure-1.** a) The Eye Tribe System b) Calibration process screen (The Eye Tribe).

To compute  $(x,y)$  coordinates on the screen and transform from those coordinates to displayed image coordinates, a calibration process is typically performed as it is depicted on Figure 1(b). Any computer equipped with an eye tracker enables users to use their eye gaze as an input modality that can be combined with other input devices like a mouse, keyboard, touch and gestures, which is referred to as active applications. We used the eye tracker as a mouse manipulator in the frame of the Fitts's law research. The details of our research are described in the following sections.

### Experimental study description

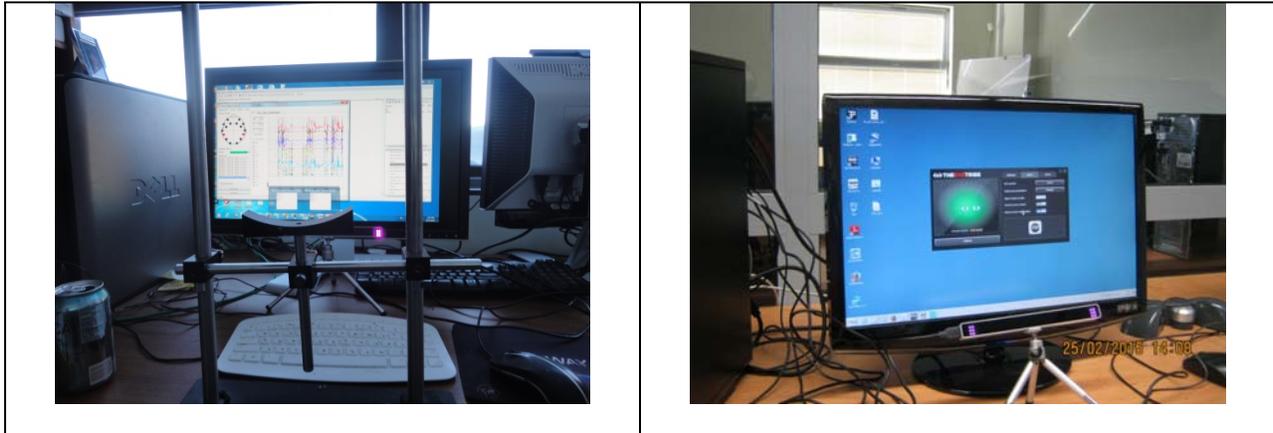
An experimental study was performed at both Michigan Technological University (USA; [www.mtu.edu](http://www.mtu.edu)) and Koszalin University of Technology (Poland; [www.tu.koszalin.pl/](http://www.tu.koszalin.pl/)). Total of 10 participants included:

- 5 students majoring in Surveying Engineering at Michigan Tech
- 5 students majoring in Geodesy and Cartography at Koszalin U

For the experiments were used:

- 21" Displays with 1600x2000 c resolution;
- PCs with USB 3.0 port;
- The Eye Tribe Tracker.

Figure 2(a) depicts the experimental setup in the US and Figure 2(b) in Poland, respectively. The US experimental setup was also equipped with chest holder to stabilize results of the eye-tracker calibration, as shown in Figure 2(a).



**Figure-2.** Experimental setup: a) at Michigan Tech University and b) at Koszalin University of Technology.

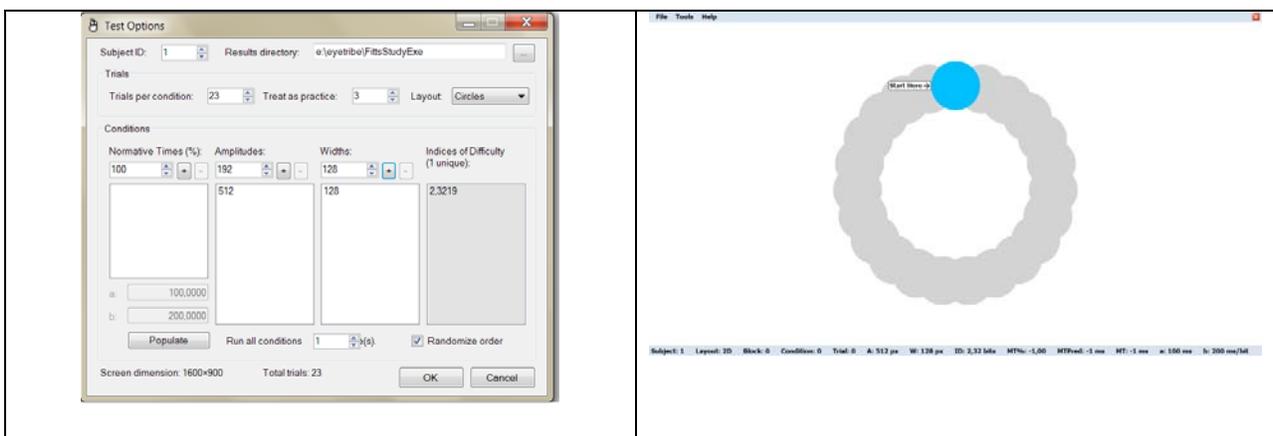
For experimental software we used:

- The Eye Tribe SDK-0.9.41-x86 to calibrate the system tracking the eye movements for all experiment participants;
- FittsStudy research software developed by Central Washington University (Wobbrock et al 2011) for screen test-objects generation and analysis of the results with ISO 9241-9 standard compliance.

Each participant in the experiment carried out three successive operations:

- Calibration with post-calibration tests;
- Measurements of test-objects with cursor control by standard mouse;
- Measurements of test-objects with cursor controlled by eye-tracker.

The test objects were generated by FittsStudy accordingly to Test Options as is shown in Figure 3(a). Test parameters were consistently the same for all experiment participants. Figure 3(b) depicts a sample of circular test objects with 128 pixels diameter demonstrated randomly within 512 pixels radius circular test-field.



**Figure-3.** a) the experimental parameters setup and b) the test-object sample.

Measurements results along with a timeline were recorded in XML format where screen coordinates are in display system units and time is in milliseconds. Sample of the raw experimental data are given in Table 1 below.

**Table 1. Cursor registration and time in microsecond raw data sample.**

<move index="0"	point=	X=793	Y=217	Time=130
<move index="1"	point=	X=789	Y=232	Time=350
<move index="2"	point=	X=786	Y=254	Time=360
...	...	...	...	...

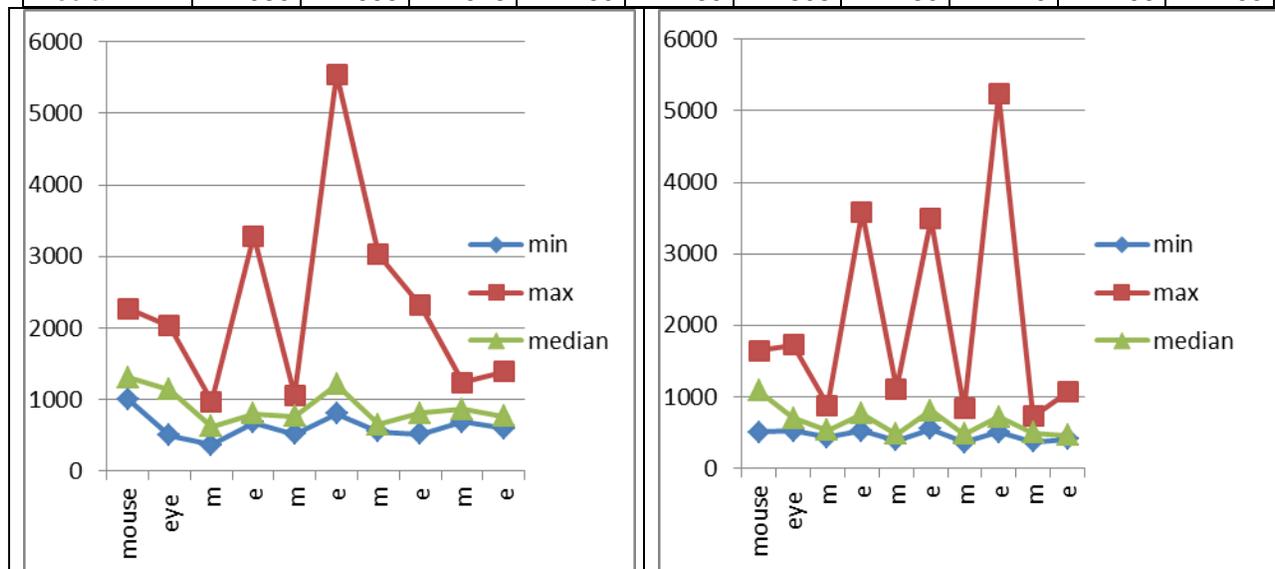
The purpose of our experiment was to determine the time to move the cursor to a specified object with the mouse and by gaze control with the eye-tracker. We have to recognize that when using the mouse, test participants were already exposed to this process and input modality with significant prior experience. Even for non-active computer users, during the day a mouse is engaged hundreds of times, and for active users this is likely thousands of times. However, the object measurements via the eye tracker were the first time that all experiment participants used this technology and without prior training. Therefore, it can be assumed that the direct use of these experiments does not allow a fully holistic evaluation of system cursor control by gaze.

**Results analysis**

For each test we performed statistical processing, including minimal, maximal and a median time for measurement of test object by cursor. The results of the experimental data processing are shown in tabular (Table 2) and graphical (Figure 4) forms. Specifically Figure 4(a) depicts statistics for experiments at Michigan Tech University and Figure 4(b) for Koszalin University of Technology, respectively.

**Table 2.** Statistical processing results comparing time for manipulating mouse cursor via manual control (Mouse) and the eye tracker (Eye) for subjects in Poland.

Time (MicroSec)	Subject1 Mouse	Subject1 Eye	Subject2 Mouse	Subject2 Eye	Subject3 Mouse	Subject3 Eye	Subject4 Mouse	Subject4 Eye	Subject5 Mouse	Subject5 Eye
Min	504	520	432	520	384	552	360	504	368	408
Max	1648	1720	877	3584	1109	3488	829	5232	726	1064
Median	1080	696	528	760	480	808	480	720	496	455

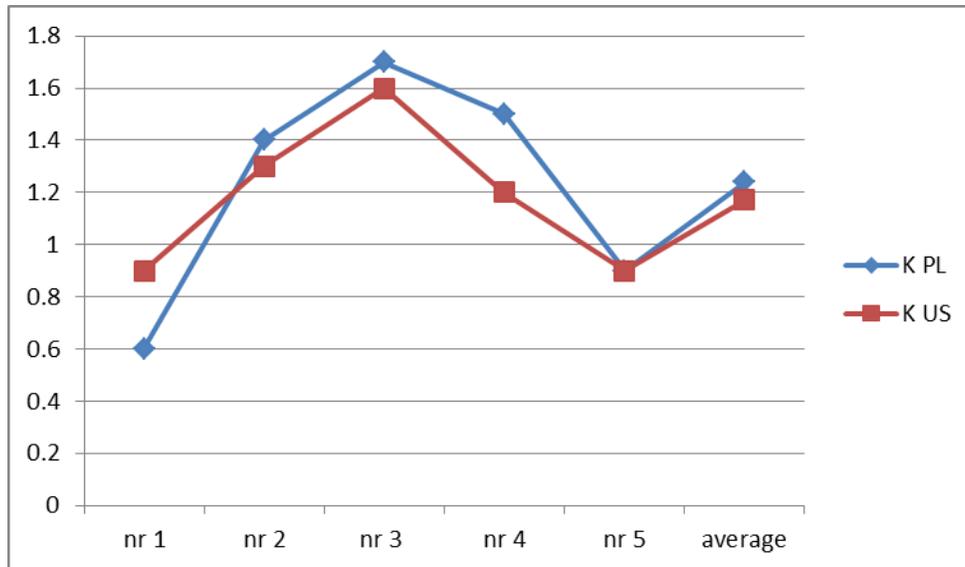


**Figure-4.** Statistical results of experimental participants for a) Michigan Tech University and b) Koszalin University of Technology using the manual mouse (m) and eye tracker (e) cursors.

Time ratio between mouse and eye tracker modes is an average of 1.21. Details are given for both US and Polish test subjects in Table 3 and in Figure 5.

**Table 3. Time ratio for cursor setup by location; coefficients K are computed as (TimeMouse/TimeEyeTracker);-**

	Subject 1	Subject 2	Subject 3	Subject 4	Subject 5	Average
K for Poland site	0.6	1.4	1.7	1.5	0.9	1.24
K for US site	0.9	1.3	1.6	1.2	0.9	1.17

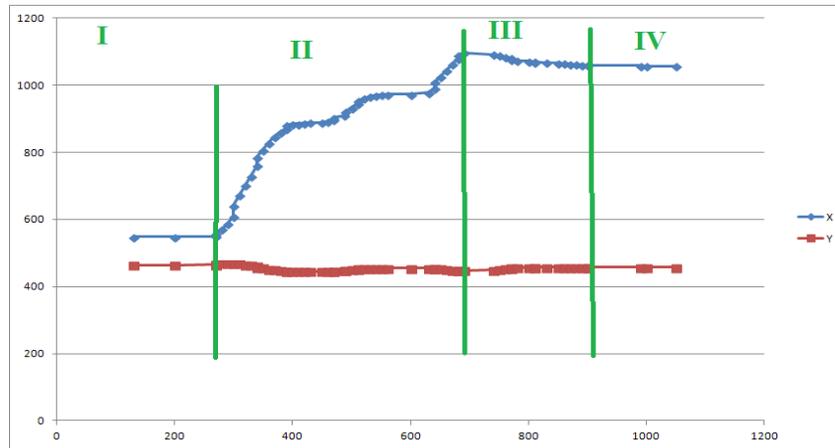


**Figure- 5.** Graphical representation of mouse/cursor time ratio; nr# is the participant ID; coefficients K are computed as (TimeMouse/TimeEyeTracker); PL = Poland, US = USA.

It is visible from Figure 5 that experiment results are practically the same in the US and Poland groups. Approximately 4 participants at each site were working with the eye tracker faster than with the mouse. It is obvious that the eye tracker results depend on:

- Eye resolution (ophthalmology factor)
- Overall reaction ability of subject
- Sizes and forms of test objects (rectangle, strip, cross)
- Distance between eyes and display
- Subject motivations
- Training and practical use of the gaze-control method

For the partial elimination of the abovementioned factors, we will analyze test results of cursor movements in a mouse mode. Sample of cursor trajectory for the statistically averaged experiment is shown in Figure 6.



**Figure-6.** Average trajectory of cursor movement by mouse; X and Y are directional displacements (in Pixels); horizontal graph axe is a Time in microseconds; green Roman numerals and bars indicate tasks

It is visible from Figure 6 that the time taken to put the cursor on a target can be represented in 4 components:

$$t = t_1 + t_2 + t_3 + t_4 \quad (3)$$

where

$t$  – total time;

$t_1$  – target search time;

$t_2$  – time of cursor movement to target which depends on cursor current position and selected target;

$t_3$  – time for correction of cursor position on target and decision making; and

$t_4$  – time to click the mouse.

Per our experimental average results, these time components are:

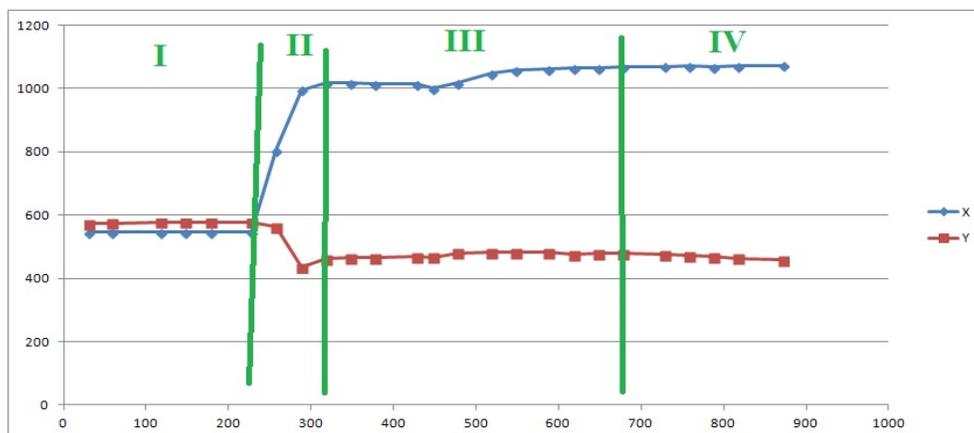
$t_1 = 250-300$  ms;

$t_2 = 350$  ms for the 510 pixels in average distance between current and targeted cursor position;

$t_3 = 300-350$  ms; and

$t_4 = 150-200$  ms.

Similar analysis can be performed also for the eye tracking mode of our experiment and is depicted in Figure 7.



**Figure-7.** Average trajectory of cursor movement for eye tracking method; X and Y are directional displacements (in Pixels); horizontal graph axe is a Time in microseconds; green Roman numerals and bars indicate tasks.

Analogously, for the mouse mode we can also decompose a common time to perform a task into 4 stages:

$$t' = t'_1 + t'_2 + t'_3 + t'_4 \quad (4)$$

where

$t'$  - common time;

$t'_1$  - target search and eye inertia time;

$t'_2$  - time for movement to target which does not depend on distance between current cursor position and selected target;

$t'_3$  - time for correction of cursor position on target and decision making; and

$t'_4$  - time to click a target using the mouse .

In average for the eye tracking mode this time components are:

$t'_1 = 200-250$  ms;

$t'_2 = 60-100$  ms - for the 510 pixels in average distance between current and targeted cursor position;

$t'_3 = 350-400$  ms; and

$t'_4 = 150-200$  ms.

### Gaze-driven geospatial data actualization system prototype

The previous experiments show the potential for the acceleration of map analysis and revision via gaze-driven HCI. We developed a working prototype of such a system by scripting its functionality on MV-TEC Halcon environment (Halcon). Operational workflow of such an experiment encompasses:

- Co-registration of mapping and imagery dataset
- Image fusion of both datasets with possibility to adjust opacity
- Eye-tracker calibration
- Passing of control of system cursor to the eye-tracker with stabilization

Each time the user decide a area of themap needs to revised and updates, the system performs image segmentation and superimposes this segmentation results on the same work view. User can then edit that data and decide if the data quality is sufficient for the database or map to be updated. Screen-short of HCI prototype is given in Figure 8.



**Figure-8.** a) an eye-controlled cursor is moving over fused map plus image view and b) when decision to update data is made, the system performs feature detection on imagery layer and superimposes the result on fused view.

A preliminary analysis of the system shows that the user may need to perform multiple recalibrations of the eye-tracker when lock on the pupils is lost by the eye-tracking cameras. In order for the HCI system to be convenient, this technology limitation will need to be solved.

## CONCLUSION AND FUTURE RESEARCH

1. Gaze controlled cursor movement to the target is almost independent from the distance to the on-screen target and occurs in 50-100 milliseconds. This speed is on average 3 to 6 times faster than the mouse controlled cursor movement.
2. When a subject is trained in the use of an eye tracker, then time  $t_3$  is significantly reduced. Subject also may gain experience and confidence in cursor positioning by gaze and thus become an advanced user of the eye-controlled environments.
3. The command to fix the cursor position on screen, for example by a double-blink or EEG command, could significantly reduce the  $t_4$  time.

4. Prototyping of the gaze-controlled geospatial data actualization indicates feasibility of the novel technology we are researching.

Our future research efforts will be devoted to the development of the novel human-in-the-loop geospatial imaging environments which will increase the productivity of humans in visual analysis operations and could be especially useful in data fusion environments. This approach of a gaze-controlled geospatial environment can be potentially expanded to video domain.

## REFERENCES

(Armenakis et al,2003) Armenakis, C., Leduc, F., Cyr, I., Savopol, I., Cavayas, F., 2003. A comparative analysis of scanned maps and imagery for mapping applications. *ISPRS Journal of Photogrammetry and Remote Sensing* 57 (5-6), p. 304-314.

(COGAIN) [http://wiki.cogain.org/index.php/Eye\\_Trackers#Eye\\_Trackers\\_for\\_Assistive\\_Technology\\_and\\_AAC](http://wiki.cogain.org/index.php/Eye_Trackers#Eye_Trackers_for_Assistive_Technology_and_AAC)  
Last accessed April 2015

(Dupont et al. 2014) Dupont, L., Antrop, M., & Van Eetvelde, V. (2014). Eye-tracking analysis in landscape perception research: Influence of photograph properties and landscape characteristics. *Landscape Research*, 39(4), 417-432.

(Fitts 1954) Fitts, P. M. (1954). The information capacity of the human motor system in controlling the amplitude of movement. *Journal of experimental psychology*, 47(6), 381.

(Halcon) <http://www.halcon.com/>, last time accessed Apr 2015

(Holland et al,2006) Holland, D.A., Boyd, D.S., Marshall, P., 2006. Updating topographic mapping in Great Britain using imagery from high-resolution satellite sensors. *ISPRS Journal of Photogrammetry and Remote Sensing* 60 (3), p. 212-223.

(Kowler and Anton 1987) Kowler E. and S. Anton. Reading twisted text: Implications for the role of saccades. *Vision Research*, 27:45–60, 1987.

(Land and Furneaux 1997) Land, M. F., & Furneaux, S. (1997). The knowledge base of the oculomotor system. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 352(1358), 1231-1239.

(MacKenzie and Buxton 1992) MacKenzie, I.S., and Buxton, W. (1992). Extending Fitts' law to two dimensional tasks. *Proceedings of the CHI '92 Conference on Human Factors in Computing Systems*. New York: ACM.

(O'Regan 1990) O'Regan, J.K. Eye movements and reading. In E. Kowler, editor, *Eye Movements and Their Role in Visual and Cognitive Processes*, pages 455–477. New York: Elsevier, 1990.

(TheEyeTribe) <https://theyetribe.com/> (last accessed on Apr 15 2015)

(Wobbrock et al 2011) Wobbrock, J.O., Shinohara, K. and Jansen, A. The effects of task dimensionality, endpoint deviation, throughput calculation, and experiment design on pointing measures and models. *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI '11)*. Vancouver, British Columbia (May 7-12, 2011). New York: ACM Press, pp. 1639-1648.

(Yarbus 1967) Yarbus, A. *Eye Movements and Vision*. Plenum Press, New York, 1967.