A "SIGHT-SPEED" HUMAN-COMPUTER INTERACTION FOR AUGMENTED GEOSPATIAL DATA ACQUISITION AND PROCESSING SYSTEMS

Gennady Gienko [a*], Eugene Levin [b]
[a] The University of the South Pacific, Fiji Islands - gennady.gienko@usp.ac.fj
[b] Michigan Technological University, Houghton MI, USA – elevin@mtu.edu

KEY WORDS: eye-tracking, human-computer symbiosis, photogrammetry, 3D modelling, knowledge elicitation

ABSTRACT

Many real-time tasks in geospatial data analysis are based on matching of visual data, i.e. finding similarity and/or disparity in geoimages either in the remotely sensed source data or in geospatial vector and raster products. When human eyes scrutinize a scene, the brain performs matching of visual streams, acquired by eyes and transmitted via all chains of human visual system. As the result the brain creates a comfortable visual model of the scene and alarms, in case some distinct disparity in visual perception is found. While observing a scene, the optical axes of both eyes are naturally directed to the same area of the object, which is particularly true for visual perception of stereoscopic images on a computer screen. If eye movements are recorded while observing the virtual stereoscopic model generated in the brain, it is possible to detect such regions of interest using fixations, identified in eye-tracking protocols. These fixations can be considered as coordinates of feature points of an object being observed (regions of interest) and can be used to reconstruct corresponding 3D geometric models by applying classical stereo photogrammetric procedures. This novel way of utilizing eye-tracking leads to establishment of "eye-grammetry" - a new approach which melds human visual abilities and the computational power of modern computers to provide "sight-speed" interaction between the human operator and the computer software in augmented geospatial data acquisition and processing systems. This paper reviews theoretical and practical aspects of eye-tracking for real-time visual data processing and outlines two particular fields of applications where the proposed technology could be useful: gaze-tracking based 3D modeling and geospatial knowledge elicitation.

1. INTRODUCTION

Many real-time tasks in geospatial data analysis are based on matching of visual data, i.e. finding similarity and/or disparity in geoimages either in the remotely sensed source data or in geospatial vector and raster products. Image fusion, change detection, 3D surface reconstruction, geospatial data conflation, – these are the only few examples of tasks that employ visual matching. Humans can instantaneously "sense" visual disparity due to the fundamental capabilities of the human visual system to perform matching. When human eyes observe a scene, the brain performs matching of visual streams, acquired by eyes and transmitted via all chains of the human visual system.

The brain creates a comfortable visual model of the scene and alarms, if some distinct disparity in visual perception has been found. Human-computer symbiosis, in augmented geospatial data acquisition and processing systems is based on eye-tracking techniques that makes it possible to arrange a "sight-speed" loop for interaction of human operator and computer software.

The virtual scene, imagined in the brain, is inherently related to neuro-physiological features of the human visual system and differs from the real world. 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 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 of interest (ROI) to construct a coherent representation of the whole.

While observing a scene, the optical axes of both human eyes are naturally directed to the same area of the object, which is particularly true for visual perception of stereoscopic images on a computer screen. Human eyes, under subconscious control, move very rapidly to scan images and the result of this scanning is sent to the brain.

If eye movements are recorded while observing the virtual stereoscopic model, generated in the brain, it is possible to detect such regions of interest using fixations, identified in eye-tracking protocols. These fixations can be considered as coordinates of feature points of an object being observed (regions of interest) and can be used to reconstruct corresponding 3D geometric model, applying classical stereo photogrammetric procedures. This novel way of utilizing eye-tracking data leads to establishment of "eye-grammetry" - a new branch of photogrammetry, which synthesizes human visual abilities and fundamentals of classic stereometry for real-time 3D measurements.

While it is generally agreed upon that fixations correspond to the image measurements, it is less clear exactly when fixations start and when they end. Common analysis metrics include fixation or gaze durations, saccadic velocities, saccadic amplitudes, and various transition-based parameters between fixations and/or regions of interest. The analysis of fixations and saccades requires some form of fixation identification (or simply identification) - that is, the translation from raw eye-movement data points to fixation

* Corresponding author

locations (and implicitly the saccades between them) on the visual display. Fixation identification is an inherently statistical description of observed eye movement behaviors. Comparative study of fixation identification algorithms (Salvucci and Goldberg, 2002) suggests dispersion-threshold method as a fast and robust mechanism for identification of fixations. This method is also quite reliable in applications requiring real time data analysis, which is a critical aspect in real-time photogrammetric applications.

The previous research outline theoretical and practical aspects of combining the human capability of matching visual images with a computer's capability of fast calculation, data capturing and storage to create a robust system for real-time visual data processing (Gienko and Chekalin, 2004). Theoretical research has been done to investigate neuro-physical features of human visual system and analyze technical parameters of eye-tracking systems. Research work was aimed on designing of a prototype of the system, developing algorithms for stereoscopic eye-tracking and investigation of accuracy issues of visual perception and stereoscopic measurement of geospatial information, - aerial photographs in particular (Gienko and Chekalin, 2004, Gienko and Levin, 2005).

The present paper describes two fields of applications where the proposed technology could be useful: 1) fast generation of 3D models based on eye movement measurements during observation by human operators of stereoscopic models; 2) knowledge elicitation: automated eye-tracking allows establishment of a protocol of an expert's conscious and subconscious processes during visual image interpretation sessions, that enables extraction and formulation of knowledge which, being asked, experts are usually unable to articulate.

2. VISUAL PERCEPTION: SEEING AND MATCHING

Typical tasks in geospatial data visual analysis include, but 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 required to perform these tasks in the real-time 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

Stage	Agent	
General matching of observed scenes	brain	
Tuned area matching	brain	computer
Disparity evaluation	brain	computer
Finding spot correspondence	brain	
Object recognition	brain	
Measuring (un)matched objects	brain	computer
Measurements registration		computer
Statistics		computer
Analysis	brain	computer

This 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 geospatial visual data in the most efficient way. As humans perceive and process vast amount of information about the environment through their visual system at extremely high speed, it is seems reasonable to combine this human's ability and computational power of computers to build a Human-Computer Symbiosis 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, 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.

3. VISUAL PERCEPTION AND EYE MOVEMENTS

The virtual scene, imagined in the brain, is inherently related to neuro-physiological features of human visual system and differs from the real world. 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 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.

Neurophysiological literature on the human visual system suggests the field of view is observed through brief fixations over small regions of interest (ROIs) (Just and Carpenter, 1984). 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.

It has been found (Mishkin et.al.,1983) 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, but this is the task of the what-system. The two systems work independently of each other and never converge to one common representation (Goldman-Rakic,1993). Physiologically, they are separated throughout the entire cortical process of visual analysis.

When the brain processes a visual scene, some of the elements of the scene are put in focus by various attention mechanisms (Posner et.al.,1990). 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. The characteristic properties of saccadic eye movements (or saccades) have been well studied (Carpenter, 1988).

Saccades are naturally linked with fixations – relatively stable positions of the eye during a certain time. Varieties of researches prove that visual and cognitive processing do occur 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).

4. FROM EYE FIXATIONS TO IMAGE MEASUREMENTS

In continuous movements eyes can be relatively stable only limited time, in most cases 200 to 800 msec. These fixations in eye positions occur in and correspond to certain regions of interest where the eyes perceive featured objects of the scene or part thereof. Projection of a certain fixation into the object's plane corresponds to a gaze position which in case of an image displayed on computer screen corresponds to certain area of an image matrix which allows as consider these gaze positions as image measurements.

While it is generally agreed upon that fixations (through their projected coordinates into the object's plane) correspond to coordinates of points in image, it is less clear exactly when fixations start and when they end. Common analysis metrics include fixation or gaze durations, saccadic velocities, saccadic amplitudes, and various transition-based parameters between fixations and/or regions of interest (Salvucci and Goldberg, 2000). The analysis of fixations and saccades requires some form of fixation identification (or simply identification)—that is, the translation from raw eye-movement data points to fixation locations (and implicitly the saccades between them) on the visual display. Fixation identification is an inherently statistical description of observed eye movement behaviors.

For spatial characteristics, (Salvucci and Goldberg, 2000) identify three criteria that distinguish three primary types of algorithms: velocity-based, dispersion-based, and area-based. Velocity-based algorithms emphasize the velocity information in the eye-tracking protocols, taking advantage of the fact that fixation points have low velocities and saccade points have high velocities. Dispersion-based algorithms emphasize the dispersion (i.e., spread distance) of fixation points, under the assumption that fixation points generally occur near one another. Area-based algorithms identify points within given areas of interest (AOIs) that represent relevant visual targets. These algorithms provide both lower-level identification and higher-level assignment of fixations to AOIs. Because fixations can also be used as inputs to AOI algorithms, these can also represent higher levels of attentional focus on a display (Scott and Findlay, 1993). These dwell times can be considered 'macro-fixations', in that they organize fixations into a larger picture. For temporal characteristics, (Salvucci and Goldberg, 2000) include two criteria: whether the algorithm uses duration information, and whether the algorithm is locally adaptive. The use of duration information is guided by the fact that fixations are rarely less than 100 ms and often in the range of 200-400 ms. The incorporation of local adaptivity allows the interpretation of a given data point to be influenced by the interpretation of temporally adjacent points; this is useful, for instance, to compensate for differences between 'steady-eyed' individuals and those who show large and frequent eye movements.

Comparative study of fixation identification algorithms (Salvucci and Goldberg, 2002) suggests dispersion-threshold method as a fast and robust mechanism for identification of fixations. This method is also quite reliable in applications, requiring real time data analysis, which is a critical aspect in real-time photogrammetric applications (Gienko and Chekalin, 2004).

5. EYE-GRAMMETRY

Spatial and temporal data derived from eye movements, compiled while the operator observes the geospatial imagery, retain meaningful information that can be successfully utilized in image analysis and augmented photogrammetry. We call this technology Eye-grammetry - a new approach to a 'sight-speed' acquisition and processing geospatial visual data using real-time eye-tracking technologies. Eye-grammetry is derived from words "eye" and "grammetry" (measure) and stands for obtaining reliable information about physical objects and the environment by detection and analysis of human eye movements observing these objects or their visual representations in images.

In general, the word grammetry refers to non-contact measurements of the object from images. Nowadays we use a number of "grammetric" techniques, aimed on precise measurement of the object, pictured in images. To acquire these images, some very advanced technologies are used – in different spectral zones and data presentations. Every new technological break-through, resulting in appearance of a new sensor, introduces a new definition – radar-grammetry, sonar-grammetry, x-ray-grammetry, etc. Sometimes looking at some of such images it is hardly to say that it is an image in the sense, that it was used early last century for conventional photographs.

Several attempts have been made to introduce the broader definitions such as “iconoactinometry” to describe new methods of registration and visual representation of the imaged objects using modern techniques, but the term “grammetry” is still well known and widely accepted within the professional community. So, to keep the traditions, we name our method Eye-grammetry – a new technology for measuring and interpretation the images.
In general, eye-grammetry could be defined as a technology of obtaining quantitative information about physical objects and the environment. This is done through measuring and interpreting images, acquired by different terrestrial, airborne or satellite sensors. In contrast to traditional principles of creation of photogrammetric terms, the first word component introduces spectral characteristics of registered radiation (photo, radar, x-ray), the word eye in our definition is interpreted as a "tool" and grammetry is widened for "image measurements". Therefore, eye-grammetry means measuring of objects in images by the eyes.

Technically, eye-grammetry is a technology based on principles of tracking the human eye movements while perceiving the visual scene. Spatial and temporal data derived from eye movements, compiled while the operator observes geospatial imagery, retain meaningful information that was successfully utilized in image analysis and augmented photogrammetry. This challenge is achievable based on human stereopsis principles.

Human stereopsis declares that while observing a scene, optical axis of the both human eyes are naturally directed to the same area of the object, which is particularly true for visual perceiving of stereoscopic images on a computer screen. Processing recorded movements of eyes, subconsciously scanning scene or image, it is quite possible to identify centers of gravity of fixation areas, which correspond to (and coincide with) identical points of the object on the left and right images of a stereopair (Figure 1).

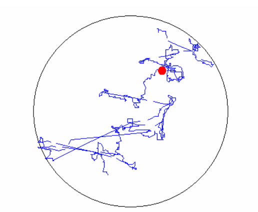

Figure 1. Eye movement trajectory in fixation area

Projected gaze directions of the operator’s eyes for corresponding fixations can be interpreted as coordinates of the featured points of an object being observed in stereo image (Figure 2). Thus, the main challenge in eye-grammetry is identification of fixations in eye-tracking protocols and calculation corresponding gaze directions to define coordinates of points in observed images, which then can be treated as conventional photogrammetric measurements.

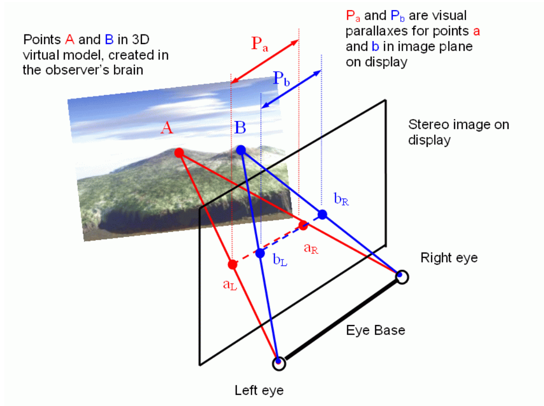

Figure 2. Stereoscopy in eye-grammetry

6. EYE TRACKING FOR 3D GEOSPATIAL MODELING

Several techniques are used to track eye movements. The electro-oculography technique is based on electric skin potential, and uses the electrostatic field that rotates along with the eye. By recording quite small differences in the skin potential around the eye, the position of the eye can be detected (Mowrer et al, 1936, Gips et al, 1993). If the users wear a special contact lens, it is possible to make quite accurate recordings of the direction of gaze. By engraving one or more plane mirror surfaces on the lens, rejections of light beams can be used to calculate the position of the eye (Ward, 2001). The most common techniques of eye tracking are based on rejected light. These techniques employ limbus tracking, pupil tracking and corneal reflection (Mauller et al., 1993, Ward, 2001). The highest spatial and temporal resolution could be achieved using the dual-Purkinje eye-trackers (Cornsweet and Crane, 1973).

Design of an eye-tracking system optimal for the geospatial data processing was an initial effort of the current research (Gienko and Chekalin, 2004, Gienko and Levin, 2005). Figures 3 and 4 demonstrate the principal design and working prototype of the system demonstrated in 2004 at XXth ISPRS congress in Istanbul (Geoiconics 2004, Intelligaze 2007).

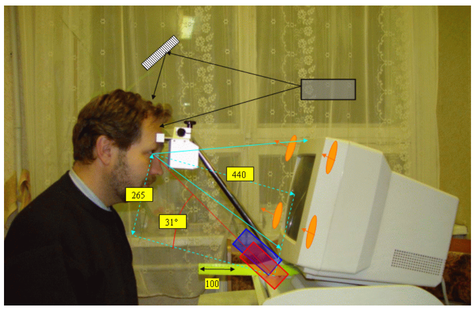

Figure 3. Principal design of eye-tracking system for geospatial visual data processing

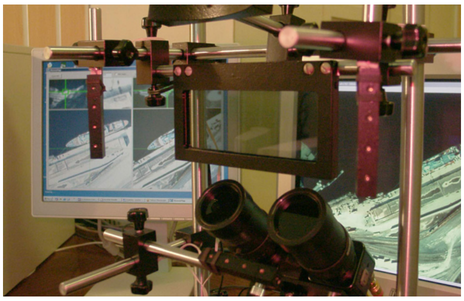

Figure 4. Working prototype of eye-tracker for processing of geospatial visual data

Calibration of precise eye-tracking systems is a bottleneck in augmented photogrammetric systems. Depending on chosen technique and hardware process of calibration involves the following major steps: 1) photometric calibration video cameras; 2) estimation positions of IR LEDs in order to estimate center of cornea; 3) resolving the geometric properties of a monitor; 4) determining relationship between video cameras and the screen to transform camera and monitor coordinate systems; and 5) determining the angle between visual-optical axis.

Once calibrated, the photogrammetric eye-tracking system can be used for two major applications, which involve matching of geospatial visual data: a) generation of 3D models based on eye-tracking; b) knowledge extraction based on eye-tracking protocols of the Subject Matter Experts (SME) and Decision Makers.

Figure 5 illustrates major stages of photogrammetric eye-tracking process for 3D modeling, assuming that eye-tracking system has been calibrated and the human analyst observes the scene stereoscopically under comfortable and stable stereoscopic conditions.

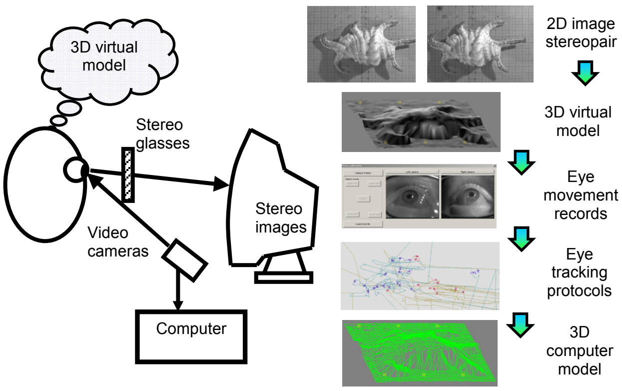

Figure 5. Principles of 3D scene restoration based on eye-movement measurements

The challenge in this technology is to extract a set of discrete and well-defined image measurements to reconstruct 3D model of a scene in the real-time. Point measurements are derived from fixations, which, in turn, statistically calculated and extracted from eye-movement protocols using individually set criteria, defined as a result of geometric calibration which contains personal data and parameters for each human analyst. Calibration process is personalized and sensitive to physiological parameters of eyes. Apart from projection parameters to calculate gaze position on the screen, the calibration involves compensation of head movements which is the second derivative and partially correlated with saccadic eye movements.

7. EYE-TRACKING AND GEOSPATIAL SME KNOWLEDGE EXTRACTION

The idea of using eye-tracking for geospatial Subject Matter Expert (SME) knowledge elicitation is based on discovering and formalizing associations and correlations between content of the image observed, expert's eye-movements 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. Technological scheme of the eye-tracking based visual knowledge extraction is depicted in Figure 6.

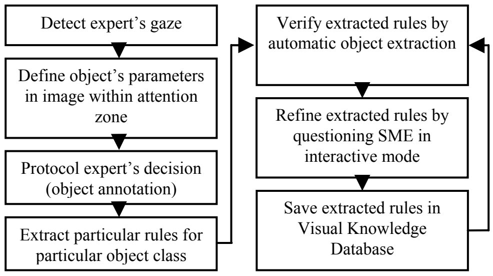

Figure 6. Eye-tracking based knowledge elicitation process

Once the expert finishes natural process of image recognition, the full set of extracted rules is verified by re-applying those rules by automatic classification of the same source image. All automatically extracted and classified objects 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 "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.

The system is designed to implement self-learning concept to accrue results of classification of the same image carried out by a number of experts with different levels of expertise. The system allows Subject Matter Experts (SMEs) to formalize and transfer their imagery knowledge into knowledge-based reasoning systems most efficiently, with minimal help of knowledge engineers. Conceptually this technology is based on research in neurophysiologic features of human visual system (HVS), particularly related to Gestaldt rules, and cognitive associations while perceiving meaningful visual information.

8. CONCLUSIONS AND OUTLOOK

Eye-grammetry is a very new direction in geospatial data acquisition, processing and analysis. Based on eye-tracking methods, eye-grammetry synthesizes human visual abilities and computational power of computers to build a new kind of Human-Computer Symbiosis, specifically designed to solve variety of tasks that involves extensive processing of geospatial visual data – from measurements to object recognition. Applications of eye-grammetry in geospatial technologies can be numerous – 3D modelling and eye-guided selective LIDAR data cleaning, DEM compilation and interactive geodatabase updating using visual data fusion, natural disaster assessment and decision making support in Geographic Expert Systems, education, training and Real-Time Expertise Transfer (RTET), air-traffic control and geo-monitoring and warning systems, homeland security and surveillance, etc.

Further theoretical and practical research and investigations should be carried out towards comprehensive analysis of neuro-physiological features of human visual system, particularly on study of optical and physical eye parameters for observations of 3D virtual models by viewing stereo images in photogrammetric applications; precision and accuracy issues - such researches should encompass an impact of digital image resolution, video frames frequency and visual a-synchronism of left/right eyes on the accuracy of identification of fixations. Developing rigorous mathematical models to link light-eye-camera-object parameters for precise geometric calibration is another niche for extensive investigations. Hardware and optical limitations, real-time hi-res video stream processing are the other challenges – some alternative programming languages and approaches have to be considered to ensure effectiveness of image measurements and data processing.

REFERENCES

Carpenter R.H.S., (1988). Movements of the Eyes. London: Pion.

Cornsweet, T. N., Crane, H. D., (1973). US Patent US3724932.

Geoiconics (2004). Corporate materials, http://www.geoiconics.com

Gienko, G., and Chekalin, V. (2004). Neurophysiological features of human visual system in augmented photogrammetric technologies. Proc. XX ISPRS Congress, Istanbul, 12-23 July, 2004.

Gienko, G., and Levin, E., (2005). Eye-tracking in augmented photogrammetric technologies. Proc. of ASPRS Int. conference, Baltimore, USA, March 2005

Gips, J. and P. Olivieri, and J. Tecce. (1993). Direct control of the computer through electrodes placed around the eyes. IEEE Transactions on Systems, Man, and Cybernetics, 20(4):630{635, 1993. In Fifth International Conference on Human-Computer Interaction, (HCI International '93).

Goldman-Rakic, R.S., (1993). "Dissociation of object and spatial processing domains in primate prefrontal cortex", Science, 260, 1955–1957.

IntelliGaze (2007). Corporate materials, http://www.intelligaze.com

Just, M.A., and Carpenter, P.A., (1984). Using eye fixations to study reading comprehension. In D. E. Kieras & M. A. Just (Eds.), New Methods in Reading Comprehension Research (pp. 151-182). Hillsdale, NJ: Erlbaum.

Mauller, P.U., D. Cavegn, G. d'Ydewalle, and R. Groner (1993). A comparison of a new limbus tracker, corneal reflection technique, purkinje eye tracking and electro-oculography. In G. d'Ydewalle and J. V. Rensbergen, editors, Perception and Cognition - Advances in Eye Movement Research, pages 393-401. Elsevier Science Publishers.

Mishkin M., Ungerleider, L. G. and Macko, K. A., (1983). Object vision and spatial vision: Two cortical pathways, Trends in Neuroscience, 6, 414–417.

Mowrer O.H., and T.C. Ruch, and N.E. Miller. The corneo-retinal potential difference as basis of the galvanometric method of recording eye movements. Am J Physiol, 114:423-428, 1936.

Posner, M.I., and Petersen, S.E., (1990). "The attention system of the human brain", Annual Review of Neuroscience, 13, 25–42.

Salvucci, D. D., and Goldberg, J. H., (2000). Identifying fixations and saccades in eye-tracking protocols. In Proceedings of the Eye Tracking Research and Applications Symposium (pp. 71-78). New York: ACM Press.

Scott, D. and J. M. Findlay, (1993). Visual search, eye movements and display units. Technical report, before 1993.

Ward, D. J. (2001). Adaptive Computer Interfaces. PhD thesis, Churchill College, Cambridge.