HUMAN CENTRIC APPROACH TO INHOMOGENIOUS GEOSPATIAL DATA FUSION AND ACTUALIZATION

Eugene Levin, Assistant Professor,
School of Technology
Michigan Technological University
Houghton, MI 49931
elevin@mtu.edu

Alexander Zarnowski, Professor,
Department of Photogrammetry and Remote Sensing
The University of Warmia and Mazury in Olsztyn, Poland
aleksander.zarnowski@uwm.edu.pl

Cheryl A. Cohen, Visiting Assistant Professor,
Cognitive and Learning Sciences
Michigan Technological University
Houghton, MI 49931
cacohen@mtu.edu

Robert Liimakka, Assistant Professor
School of Technology
Michigan Technological University
Houghton, MI 49931
raliimakk@mtu.edu

ABSTRACT

This paper describes a research which attempts to combine the advantages of human analysts and computer automated processing for efficient human computer symbiosis in geospatial data fusion. Specifically, experiments performed were related to the analysis of the potential use of inhomogeneous (composed of different sources) stereo pairs for mapping dataset actualization. Inhomogeneous stereo pairs were combined with images of the map to be updated along with actual aerial images of the same territory. The anaglyphic product obtained after image processing of such stereo pairs was demonstrated to human analysts (subjects) and stereo perception of such stereo pairs was achieved. The most interesting finding of this experiment is the fact that some objects existing only on the aerial photo appeared in the inhomogeneous stereo pairs as 3D. This effect is caused by phenomena within the human eye-brain system known as human stereopsis which is widely deployed in photogrammetry. For the quantities measurements of the effect obtained, an eye-tracking system was deployed. Analysis of human eye-movements (driven by conscious and subconscious brain processes) while perceiving an inhomogeneous stereo dataset -provides a unique opportunity for the human computer symbiosed geospatial systems. There are two potential outcomes of such approach: a) interpretative – analysts' gaze-fixation zones can help to localize the areas where mapping dataset should be updated b) quantitative – processing of eye fixations geometry during stereo model perception allows to transform the virtual 3D model to a geometrical one based on binocular summation measurements deploying eye-tracking.

INTRODUCTION

Nowadays vast amounts of remotely sensed geospatial data including real-time and near real-time updated sources make the problem of geospatial data fusion and mapping/GIS dataset actualization very important. Timely updating of mapping products in an automatic way is complicated by the fact that multi-sensor geospatial data is burdened by multiple errors. Therefore, geospatial analysts' involvement in mapping product actualization is still necessary in geospatial production environments. The fusion of geospatial data from multi-sensory and multi-temporal imaging and mapping sources has the potential to improve robust intelligence systems in significant ways. Potential applications for robust intelligence systems that use fused geospatial data include: terrain and planetary visualization and exploration; environmental GIS from battlefield surveillance; 3D multisensor non-destructive inspection; global guidance navigation of manned and unmanned robotic platforms; forensic; and homeland security related scenes analysis. Geospatial processes that support such proposed intelligence systems include: 1) registration (the spatial alignment of geospatial data sources in one coordinate system); 2) mosaicking (the creation of composite data sets; and 3) 3D measurements of the shapes and spatial dimensions of specific objects. State-of-the-art multi-sensor geospatial data fusion processes (NGA GGIS),(USGS Antarctica), (Abidi & Gonzales, 1993) use sophisticated and automated workflows based on the following processes: pixel-level fusion (Gonzalez et al,1994), (Delanoy et al, 1991); feature level fusion (Nikolov et al,2003),(Piela,2003); discrete and adaptive model matching (Hamilton and Kipp 1993, Petrovic and Cootes 2007]; decision level fusion (Silk et al 1991),(Singh et al 2006); and multiple-level fusion (Chu et al 1992), (Roy et al.2008).

Aerial image interpretation and fusion has a longer history than the history of computerized technologies because analysts interpreted these images before the era of computers. Specifically, described in this paper research was inspirited by analog technology known as Interpretoscope of Carl Zeiss Jena used for aerial photo interpretation and maps update in the early 80s. The name of this device came from Latin word interpreter – explain and Greek word skopéo – observe. The device allowed for the interpreting of stereo pairs of images of the same scale and different scales (up to 1:7.5) and performing of simple measurements and cartographic functions. Some modifications of this device were produced with a pair of stereoscopes for conversations between analysts. The modern geospatial image analyst operates with softcopies of source images and deploys special imaging workstations such as ENVI, ERMapper, ERDAS Imagine etc. In most other ways, the operational workflow of a modern image analyst is very much similar to hardcopy technology realized by the Interpretoscope (Zoom-In, Zoom-out, Spatial Associations Analysis, Make decision, Label Object).

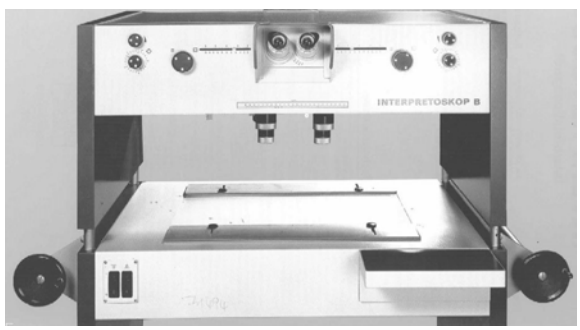


Figure 1. Carl Zeiss Jena analog Interpretoscope (Gupta,2003);

In terms of Interpretoscope it is attempt to realize in a digital way technology when it was possible to update map during steroscopic fusion of this map with newer aerial image and perform cartometric measurements on this dataset. These kinds of processes were state-of-the-art at times when Interpretoscope was deployed in research and production environments. Described here approach can be termed as "technology fusion" encompassing photogrammetry, cognitive neurosciences and computer vision technologies.

HUMAN-STEREOPSIS IN PERCEPTION OF BI-SENSOR/BI-TEMPORAL DATA

Described in this paper approach makes use of well-known properties of cyclopean vision, the brain's ability to merge separate retinal projections into a single visual representation. In normal human vision, each eye receives a different two-dimensional retinal projection of an object. These two retinal projections are offset by the lateral distance between the two eyes, a phenomenon known as *binocular disparity*. The visual system calculates the relative depth of an object with respect to the object that the eyes are fixated on. The perception of three-dimensional depth from two disparate retinal projections is called stereoscopic depth perception, or *stereopsis* (Qian, 1997). Normal human stereopsis requires that projections of viewed object be present in both left and right retinal images. When monocular projections to the retina are similar or identical, a unified binocular representation is formed. This process is called binocular fusion (Liu, Tyler, & Schor, 1992). It is also possible for humans to perceive three-dimensional depth information about an object when a retinal projection of the object is present in one eye only. In our pilot experiments, analysts viewed two images (a stereopair) through anaglyph glasses, which separated the visual information accessible to each eye. A target object (for example, a white box, as shown in Figure 2) was presented in the view for one eye, but absent in the view of the other eye.

Figure 2. Example of bi-temporal stereo-pair.

Despite the fact that the target object is present in one image only, the cyclopean percept of the stereopair includes a three-dimensional representation of the target image. In this unified percept, one of the stereopair images appears to be semi-transparent and overlaid on top of the other image (Figure 2).

This percept can be explained by the phenomenon of *binocular summation* (Liu, Tyler, & Schor, 1992). Anaglyph glasses present images dichoptically (a separate image shown for each eye). When pair of low contrast images is shown dichoptically, a *dichoptic plaid* is created consisting of the images overlaid on each other. The perception of a dichoptic plaid results from a process of binocular summation. In this process, dissimilar monocular images reach consciousness at the same time, and their effects are added up, point by point. Each image contributes to a unified perception while retaining its own distinct properties. (Liu, Tyler, & Schor, 1992). In our pilot experiments, participants also report a percept of a dichoptic plaid while viewing dissimilar stereopairs of bisensory images (images representing the same spatial location obtained by different sensors). Figure 4.5c shows a stereopair of bisensory images and a resulting dichoptic plaid. The goal of our research is to measure the 3D properties of objects perceived during dichoptic viewing of (Liu, Tyler, & Schor, 1992) bitemporal and biosensor stereopairs. We used eyetracking methodology to compute the fixation point of each eye as the analyst views a pair of dissimilar images through analglyph glasses. Eyetracking is a non-invasive methodology that records the position of the fovea during viewing. The fundamental assumptions of eyetracking methodology are that the eye fixates on objects of interest during visual attention (Duchoswski, 2003),(Just & Carpenter, 1976).

RESEARCH EXPERIMENT DESCRIPTION

Computing depth from binocular disparity is of interest to the geospatial information sciences because it can be used to automate quick depth calculations. Our research experiment uses dichoptic devices (separate images presented to the left and right eyes) to study foveal fixation during underlying stereopsis A common method of presenting dichoptic images is through anaglyph imagery. In this process, the analyst views a pair of specially prepared images through spectacles with lenses of two different colors, usually red and cyan. Each individual anaglyph image is composed of two superimposed images, which represent the same object at two views slightly offset from each other. Each layer of an analglyph image is printed in a different color. The analyst views the merged image through the colored lens absorbing light from one view of offset image, so that images received by the analyst's left and right eyes are offset with binocular disparity. Alternative technologies for separating the two views in a dichoptic pair are: polarization; temporal separation by shutter-glasses: chroma-stereoscopy; and auto stereoscopic displays utilizing spatial separation by lens or silts. Eyetracking technology can compute the depth of a fixation point from the binocular disparity and express the disparity numerically. While observing a scene, the optical axes of both human eyes are naturally directed to the same point on the object. This is particularly true for visual perception of stereoscopic images on a computer screen. Human eyes move very rapidly while scanning images and the result of this scan is sent to the brain. Our previous research (Levin et al, 2008a),(Levin et al, 2008b) recorded participants' eye movements while they observed the virtual stereoscopic models and used eyetracking protocols to identify fixations. The fixations are assumed to be feature points of the viewed objects and are used to reconstruct 3D geometric models by applying classical stereo photogrammetric procedures. This eyegrammetric approach is useful for 3D models restoration measured by stereopsis. As show in Figure 4.1 this method is not applicable for binocular summation estimation because one of the inhomogeneous stereopairs images does not contain the same features.

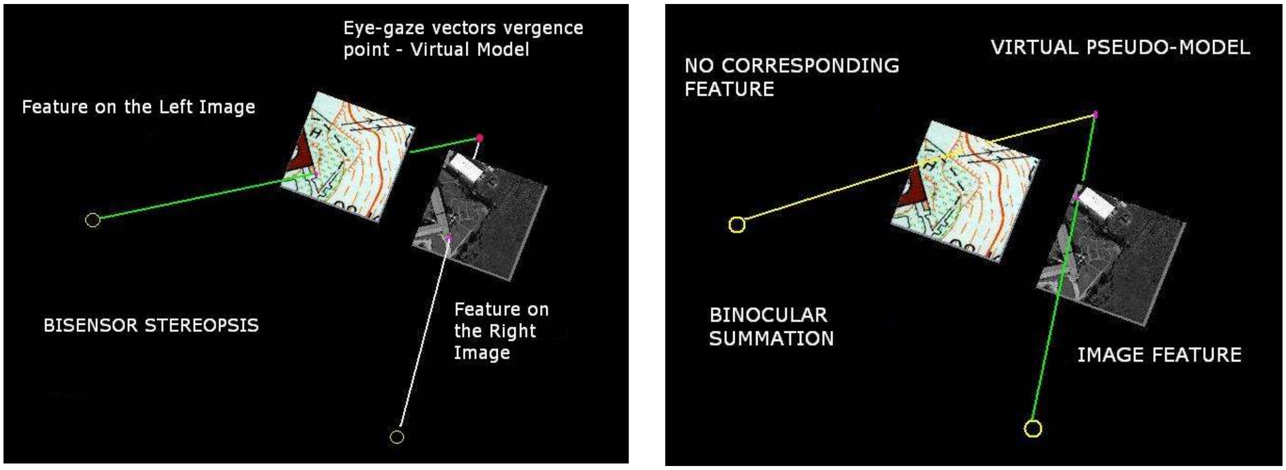


Figure 3. Difference between a) stereopsis and b) binocular summation image features based points matching .

During binocular summation, the human brain makes "guesses" of where the corresponding feature should be. 3DGICS binocular summation reconstruction consists of two problems : *1)* how to select from the left and right eye fixation cloud points (depicted in Figure 4-2a) the optimal combination of gazes from the left and right eyes; *2)* Determining accuracy of the 3D estimates, expressed in eye vergence angles. The following approach is proposed to solve the first problem:

1. Image feature pixel is selected by of corner detection algorithms (Harris, 1988) and applied to the center fixation or cursor position if a cursor-click measurement was performed by the analyst to measure the point.
2. Selecting image feature where exist by corner detection algorithm and then search of corresponding eye-gaze vector as the most satisfying for classical co-planarity photogrammetry equation.
3. Selecting corresponding left and right gaze pair by least-square optimization.
4. 3D coordinates of points are computed from the stereo-pair system transformed to 3D geographical coordinates (latitude, longitude and altitude).

Inhomogeneous Stereopairs Preparation Technology

For the eye-tracking experiments we generated anaglyphic bi-sensor and bi-temporal stereopairs using terrestrial digital photographs and online photogrammetric and map actualization data and system developed by Photogrammetry and Remote Sensing department of Warmia and Mazury University in Olstyn Poland available online at http://www.kfit.uwm.edu.pl/zp/. Figure 4 depicts user interface of this inline mapping tool which we used for imitation of stereo-mode of map renovation of Interpretoscope. It is visible from Figure 4 that map components and aerial image components represents the same spatial extent. However, stereopair composed of this image sources is inhomogeneous and can be perceived stereoscopically only due to binocular summation human-vision effect. We generated several datasets composed of geospatial images of known geometry. Some examples are shown below in Figure 5. Our challenge was to compare 3D depth measured on standard stereopsis and binocular summation vision. This comparison can give a proof-of-feasibility of using inhomogeneous stereopairs in geospatial data processing. Specifically, depth measurements can be and indicator of the facts how eye-tracking derived binocular summation parallaxes are different from respective parallaxes derived for the same stereopairs in normal stereopsis mode. As it is known, stereo measurements by human operators were widely deployed in analog photogrammetry era. Definition of operator ability of 3D perception and "personal difference" were some biometric procedures which were commonly used at those times in photogrammetric production environments.

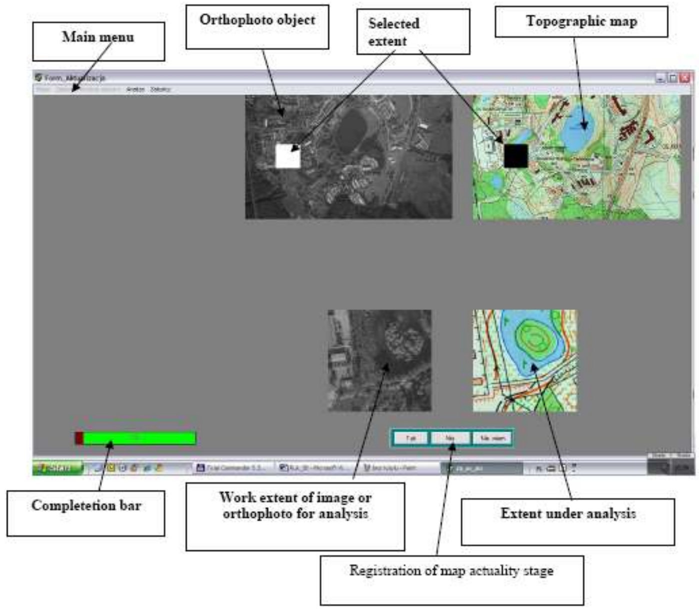


Figure 4. User interface of online map actualization system used to generate bi-sensor/bi-temporal stereopairs.

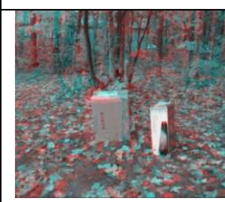
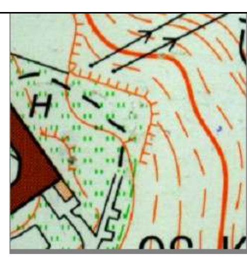
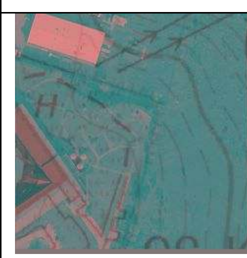
Experiment description	Left stereo-pair component image	Right stereo-pair component image	Stereo-pair recorded
Normal stereopair for stereopsis measurement			
Bitemporal stereopairs one of objects presented in one image only			
Bisensor stereo composed of aerial image and topographic map			

Figure 5 Images from research experiments, showing stereoscopic input and cyclopean percepts.

Eye-tracking Experiment Setup and Data Collection Procedure

We have performed a several research experiments following this methodology using Seeing Machines facelab 4.0 eyetracking system of (SM) , shown below:

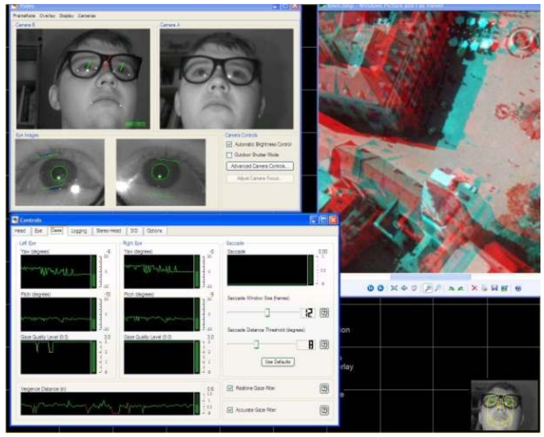
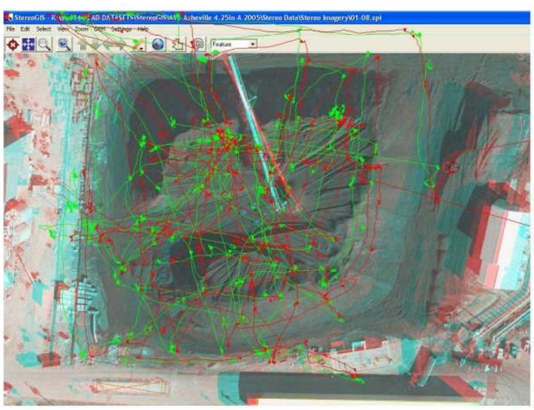


Figure 6. Eye-tracking experimental setup.

Experiments were collected with to following sequence:

- 2D and 3D(wearing anaglyphic glasses) calibration of eye-tracking system with subject;
- recording of eye-movements protocols simultaneously with 3D cursor fixation on virtual 3D model;
- detection of eye-fixations in eye-movement protocols corresponding to mouse-click events;
- calculation of relative depth based left and right eye fixations corresponding to measurements and calibration parameters;
- comparison of relative depth for normal stereopsis and binocular summation cases.

To perform 3D calibration we used control points in the geographical 3D space. Specifically, we used StereoGIS (SimWright) workstation provided by SimWright. Figure 7 depicts generated anaglyphically test image from StereoGIS and used for the measurements on eye-tracker.



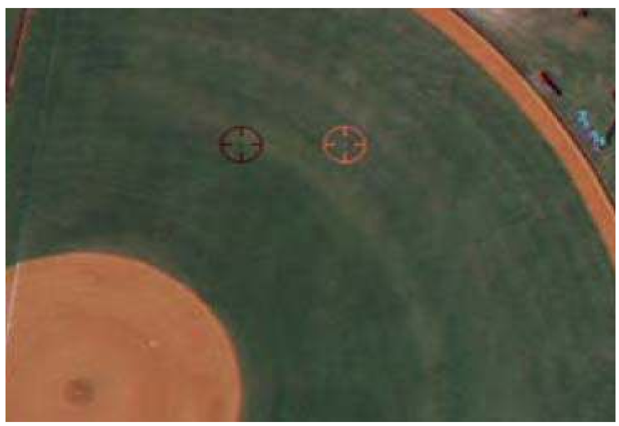


Figure 7. a) Stereopair processed on StereoGIS and used for 3D calibration b) stereoscopic cursor used for control synchronization of attention eye-fixations with manual control measurements procedures.

Graphical illustration of eye-movement trajectory and parallaxes for inhomogeneous case is presented on Figure 8

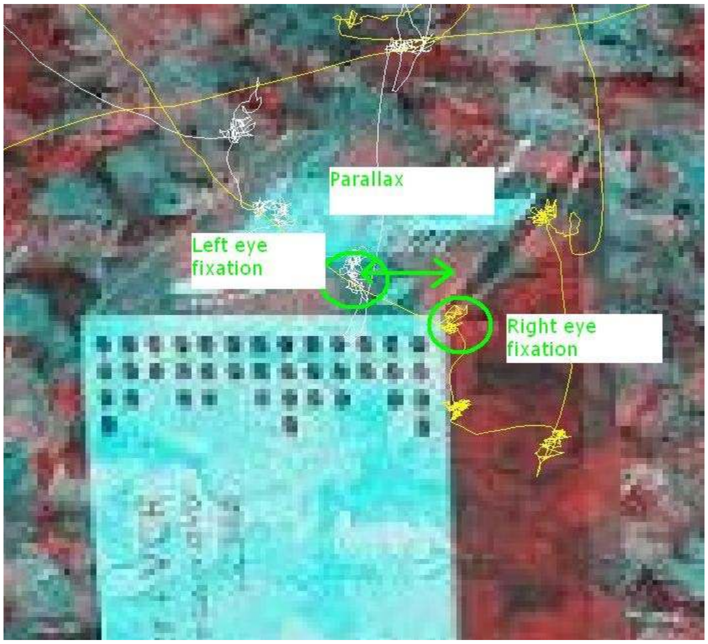


Figure 8 . Left (white) and Right (yellow) eye-movement trajectories, corresponding fixations and parallax definition principles.

Experiment results

Results of 3D processing recorded images are summarized in Table-1 below

Experiment	Measured Dimension (m)	Stereopsis Result	Binocular Summation Result	Error % compare to Stereoppsis	Final Error Compare to Ground
Box#1 (bitemporal)	0.571	0.495	0.386	13.30998249	-32.39929947
Box#2 (bitemporal)	0.68	0.695	0.599	2.205882353	-11.91176471
Building (bisensor)	15.345	16.365	11.842	6.647116325	-22.82828283
Average					**-22.37978234**

Table-1 shows that binocular rival measurements provide about 70% of 3D depth measurement accuracy compare to ground dimension, and demonstrates the feasibility of the proposed system. Future research will concentrate on increasing participant numbers and experimental database size using various aerial and satellite geospatial images.

CONCLUSION AND OUTLOOK

Described approach establish significant elements necessary to establish a working prototype of a cognitive geospatial viewing platform that will address the limitations of current systems, such as noisy data, sensors of different modality, image scale, resolution, and pseudo-stereo. The goal of cognitive approach to human-centric geospatial data processing is to combine human stereo-perceptual ability with the computational power of computers to build a Human-Computer Symbiosis platform for robust intelligence in the domain of geospatial imaging.

FUTURE WORK

Future developments will be devoted to the integration of the developed approach with state-of-the-art geospatial imaging interpretation environment. The tasks to initiate stereoscopic viewing are eyetracker calibration, selecting images and rough registration to align the two images. In addition, eyetracker calibration is required periodically during the stereoscopic viewing. Consequently, we will use a dual monitor system: one monitor devoted to 2D viewing for calibration and image selection and another monitor devoted to stereoscopic viewing. The left and right images will be selected by database queries and selection from annotated thumbnail images. The initial registration will be accomplished by selecting corresponding points on the left and right image. At any time, the analyst can switch to stereoscopic viewing by glazing on the second stereoscopic monitor. Initial design of the future system's human-computer interface is depicted on Figure 9

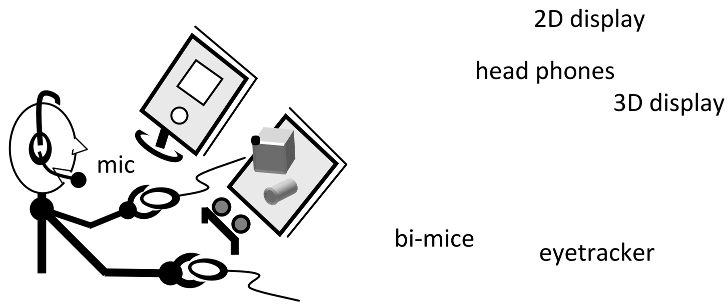


Figure 9. human-computer interactions initial design.

The tasks during stereoscopic viewing are fine registration of the images, visually navigating the image, recording and annotating portions of the views. During registration and navigation, the analyst will need to simultaneously scale, translate and rotate the two images individually and simultaneously. Voice commands interface potentially will make analysts' work more efficient.

Potential applications of the described research when will be developed till technology stage include: 1) expansion to motion imagery; 2) integration with state-of-the-art geospatial environments; and 3) expansion to other knowledge domains, such as medical radiology and geophysical data interpretation.

REFERENCES

(Abidi and Gonzales, 1993) M.A. Abidi and R.C. Gonzales, Eds., Data Fusion in Robotics and Machine Intelligence, Academic Press, Boston, 1993.

(Chau et al,1992) Chen-Chau Chu and J.K. Aggarwal, Image Interpretation Using Multiple Sensing Modalities, IEEE Trans. on Pattern Analysis and Machine Intelligence, August 1992, Vol. 14, No. 8, 840–847.

(Delanoy et al, 1991) Richard Delanoy, Jacques Verly, and Dan Dudgeon, Pixel-Level Fusion Using "Interest" Images, Proc. 4th National Sensor Symp., August 1991, Vol. I, 29.

(Duchowski 2003)Duchowski, A. T. (2003). Eye tracking methodology: Theory and practice. New York: Springer. (Gonzales et al, 1994) Victor M. Gonzales and Paul K. Williams, Summary of Progress in FLIR/LADAR Fusion for Target Identification at Rockwell, *Proc. Image Understanding Workshop,* ARPA, November 1994, Vol. I, 495–499.

(Gupta 2003) Gupta, Ravi P, Remote Sensing Geology, 2nd ed., Springer 2003.

(Hamilton & Kipp 1993) M.K. Hamilton and T.A. Kipp, ARTM: Model-Based Mutisensor Fusion, *Proc. Joint NATO AC/243 Symp. on Multisensors and Sensor Data Fusion,* November 1993.

(Harris 1988) Harris C., Stephens, M. ``A combined corner and edge detector," in Alvey Vision Conf., 1988, pp. 147-151.

(Just & Carpenter, 1976) Just, M. A. & Carpenter, P. A. (1976). Eye fixations and cognitive processes. Cognitive Psychology, 8, 441-480.

(Levin et al 2008a) E Levin, W. Helton, R. Liimakka., G Gienko, EYE MOVEMENT ANALYSIS IN VISUAL INSPECTION OF GEOSPATIAL DATA, Proc. of ASPRS Annual. Conference, Portland, OR, 2008.

(E. Levin et al 2008b) E. Levin, G. Gienko, A Sergeev ,"HUMAN CENTRIC APPROACH FOR GEOSPATIAL DATA FUSION IN HOMELAND DEFENCE AND SECURITY APPLICATION SCENARIOUS", 2008 SPIE Defense and Security Symposium, Orlando FL ,March 2008.

(Liu Tyler, & Schor 1992) Liu, L., Tyler, C. W. & Schor, C. M. 1992 Failure of rivalry at low contrast: evidence of a suprathreshold binocular summation process. Vision Res. 32, 1471^1479.

(NGA) Global Geospace Information and Services (GGIS) available at : www. earth-info.nga.mil/gns/html/namefiles.htm

(Nikolov et al, 2003) Nikolov, S. G., Gilchrist, I. D., and Bull, D. R., Gaze-contingent multi-modality displays of multi-layered geographical maps, pp. 325 – 332. In Proceedings of the Fifth International Conference on Numerical Methods and Applications (NM&A02), Symposium on Numerical Methods for Sensor Data Processing, Borovetz, Bulgaria. LNCS 2542, Springer, Berlin, 2003.

(Qian 1997) Ning, Qian. Binocular disparity and the perception of depth. Neuron, 18, 359-368.

(Petrovic & Cootes,2007) Vladimir Petrovic and Tim Cootes , "Objectively adaptive image fusion", Information Fusion Volume 8, Issue 2, April 2007, pp. 168-17

(Piella 2003) G. Piella, "A general framework for multiresolution image fusion: from pixels to regions," *Information Fusion*, vol. 4, pp. 259–280, 2003.

(Roy et al 2008) D. P. Roy, J. Ju, P. Lewis, C. Schaaf, F. Gao,M. Hansen, and E. Lindquist, "Multi-temporal MODIS–Landsat data fusion for relative radiometric normalization, gap filling, and prediction of Landsat data," Remote Sens.Environ., vol. 112, no. 6, pp. 3112–3130, Jun. 2008.

(Silk et al 1991) James D. Silk, Jeffrey Nicholl, David Sparrow, Modeling the Performance of Fused Sensor ATRs, Proc. 4th Nat'l. Sensor Symp., August 1991, Vol. I, 323–335.

(Singh et al 2006) R. Singh, M. Vatsa, and A. Noore, Intelligent biometric information fusion using support vector machine, Soft Computing in Image Processing: Recent Advances by Springer edited by M. Nachtegael, D. Van der Weken, E. E. Kerre, and W.Philips Chapter 12, 2006 327-350.

(SimWright) StereoGIS product description available at : http://www.simwright.com/stereogis.htm

(SM) Seeing Machines, faceLAB; http://www.seeingmachines.com

(USGS Antarctica) Terrain view available at: http://terraweb.wr.usgs.gov/kids/Antarctica/nuvu/