Eye Tracking in Geoscience: Data Registration and Visualization*

Bo Hu[1], Tommy P. Keane[1], John A. Tarduno[2], Brandon B. May[1], Nathan D. Cahill[1,3], Robert A. Jacobs[4], and Jeff B. Pelz[1]

Search and Discovery Article #120143 (2014)
Posted March 25, 2014

*Adapted from extended abstract prepared in conjunction with oral presentation at AAPG Hedberg Conference, 3D-Structural Geologic Interpretation: Earth, Mind and Machine, June 23-27, 2013, Reno, Nevada, AAPG©2013

[1]Center for Imaging Science, Rochester Institute of Technology, Rochester, NY, USA (bhu@rit.edu)
[2]Department of Earth and Environmental Sciences, University of Rochester, Rochester, NY, USA
[3]Center for Applied and Computational Mathematics, Rochester Institute of Technology, Rochester, NY, USA
[4]Department of Brain and Cognitive Sciences, University of Rochester, Rochester, NY, USA

Abstract

Geologists routinely infer geodynamic history by viewing landscapes, but the expertise for correct interpretation requires years to develop. Eye movement data, such as gaze timings, area of interests (AOI), and order of interpretation, provide insight in understanding viewing strategies developed in experts that may be valuable in the education of novices. Toward this goal, we are studying the eye movements of novice and expert geologists during an annual class fieldtrip at numerous sites in central California and Nevada using advanced portable mobile eye-trackers and high-resolution imaging systems.

In this article, we describe algorithms for mapping videos from multiple eye-trackers onto a single high-resolution panoramic image of a scene. The algorithms are solely image-based, comparing the outward-pointing videos captured from the portable eye-trackers with the high-resolution panoramic image using standard ideas from feature estimation and multiple view geometry, without requiring any extrinsic method to relate the positions of the eye-trackers.

The high-resolution scene images, from which the panoramas are created, can be also used in creating immersive “virtual field trips”. We have found that simple rectilinear projections of the panoramas do not provide the correct perspective and sense of scale, and therefore have limited value in research and education. We describe a software framework that displays the images with correct viewing parameters, mimicking faithfully the view of the original scenes. In addition, the software tools allow intuitive annotation of AOIs by geologists, facilitating analyses of eye-tracking data.

Eye movement data provide valuable information in understanding viewing patterns of novice and experts. However, data collection in natural geological settings is demanding. Here we discuss the analysis and visualization challenges. We present an approach involving algorithms and software tools to register eye movement data in a common reference frame, and to effectively visualize and annotate high-resolution imagery. Application to our growing expert/novice data set will be discussed.

Data Acquisition

During each field trip, we simultaneously record data from 11-13 observers. The observers are free to move around the outdoor environment as their eye movements are monitored. Once the observers are equipped and the eye-trackers are calibrated, field trip participants walk to an area from which they can view geologically significant features. The class professor asks the group to visually inspect a scene for evidence of geological activity. It is during this period of active and goal-directed visual search that we have been focusing our analyses. After about 90 seconds of silent viewing, the students discuss the features they noticed, and the professor gives a lecture on the scene.

A video camera attached to the headgear captures a view of what each person is observing. A high-resolution panorama is captured at a location close to the center of the observers. We use a customized system using the Sky-Watcher® Merlin mount, based on the principles of the GigaPan® EPIC 100 system designed by Carnegie Mellon University and NASA Ames Intelligent Systems.

The scene videos are low resolution and have poor color fidelity, but they have a wider than typical field of view because of the use of a wide-angle lens. Due to the nature of the wide dynamic range typical of outdoor scenes and the use of automatic gain control, it can be particularly difficult to match these scene videos to the high-resolution panorama.

Data Registration

In order to bring eye-tracking data from the observers into a single frame of reference, we register the eye-tracking scene videos to the high-resolution panorama (Figure 1). We have designed and implemented a suite of algorithms to overcome two main difficulties in doing this registration. First, every scene video and the panorama itself are captured from different points of view, and second, the image quality of the scene videos is vastly inferior to the image quality of the panorama.

Because the relative depth of objects in the scene is small compared to the distance from the camera to the scene, we model the geometric transformations between video frames and the panorama using planar homographies. Before matching video frames to the panorama, the scene video is analyzed to determine the frame-to-frame homographies. This allows for the calculation of temporally consistent and smooth results when matching individual video frames to the panorama. To do this, SIFT (Scale-Invariant Feature Transform) features are computed and matched between each frame and its neighbor and a homography is estimated via a robust statistical procedure called RANSAC (RANdom SAmple Consensus). A record is kept of the most robust matches in order to generate feature tracks across frames. The videos and initial SIFT data are then put through a second pass that uses only those robust SIFT features that last through a minimum specified number of frames to recompute new frame-to-frame homographies, which are used to register scene video frames to panorama. A naïve way is to match a single keyframe to the panorama and transform other frames using inter-frame homographies. Though ensuring smoothness of the motion, this method leads to significant drift caused by accumulations of small errors in the frame-to-frame homographies summed over a large number of video frames.

To counteract drift, multiple video keyframes are selected to anchor the spatial transformation between the scene video and the panorama. We determine a frame in a given time interval whose transformation to the panorama is closest to an affine transformation as a keyframe. Once the set of keyframes has been selected, each of them needs to be matched to the panorama with a homography transformation. This could be done in the same way as computing the frame-to-frame transformations in the video sequence. In practice, it will fail due to the large differences in image quality between the video frames and the panorama. Instead we go through ever increasing levels per octave of computation when detecting SIFT features. Thus, the first computation compares SIFT features detected at three levels per octave to see if a valid homography can be determined using RANSAC. If not, the procedure is repeated at four levels per octave all the way up until twelve levels per octave if necessary (Figure 1b). Once a valid homography is found, the process repeats for the next keyframe. Together with our criteria of selecting keyframes, this procedure maximizes numerical stability and gives robust registration results.

Once the keyframes to panorama homographies have been calculated, the frame-to-panorama homographies for all other video frames are computed by composing the frame-to-frame homographies with the neighboring keyframe to panorama homographies. Thus, if   represents the previous keyframe,   represents the next keyframe and the current frame is frame, there will be two frame to panorama homographies, given by:

$$H^1_{j,p} = H_{k_1,p} \prod_{\ell=j}^{k_1+1} H_{\ell,\ell-1} \text{ and } H^2_{j,p} = H_{k_2,p} \prod_{\ell=j}^{k_2-1} H_{\ell,\ell+1}$$

To ensure a smooth transition between keyframes and to counter any errors due to drift, the homography at each frame is computed as a linear combination of the transformations computed using its surrounding keyframes, weighted by the distance to each keyframe; i.e.,

$$H_{j,p} = \alpha_j H^1_{j,p} + \left(1 - \alpha_j\right) H^2_{j,p}, \text{ where } \alpha_j = \frac{k_2 - j}{k_2 - k_1}.$$

Data Visualization and Annotation

As another part of our goal towards providing our data for research purposes and general investigation, we are working to combine data collected in the field into a unified, interactive “virtual field trip” framework. Utilizing the open source language Python, and the open source libraries of OpenGL, we have created an intuitive and interactive visualization tool for the high-resolution imagery, GigaPan® imagery, eye-tracking videos and gaze projections, lecture and observer audio files, and GPS location measurements.

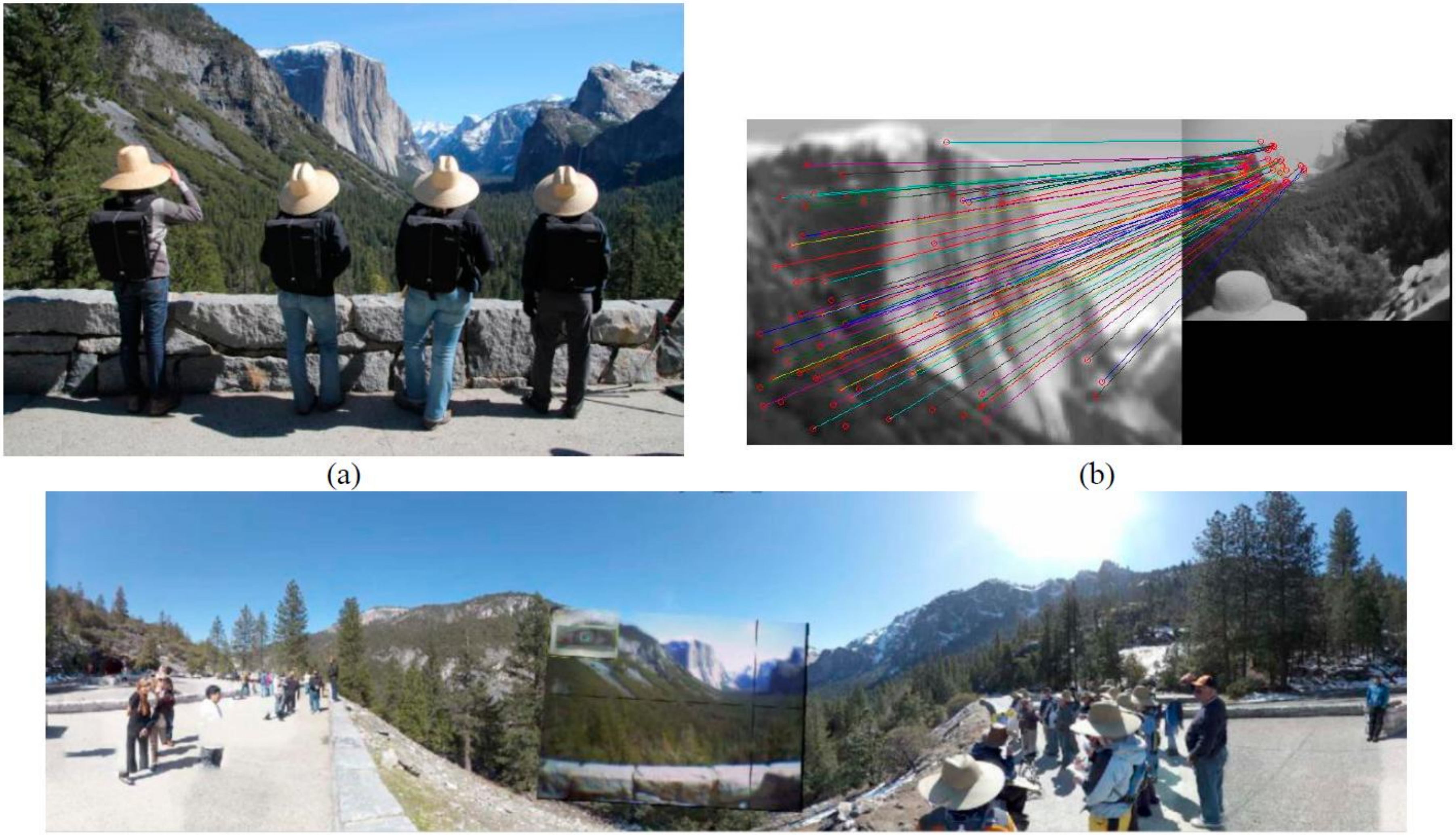
The rectilinear mappings of the GigaPan® outputs, while being mathematically robust, result in a loss of proper sense of scale and positioning of the geographical features and hence highlights of landscapes. For annotation, visualization, and understanding purposes, this ambiguity (if not actually a misleading modification, Figure 2a) of the view and structure of a scene poses a major hurdle. To address this issue, we have moved away from the commercial GigaPan® stitching and mapping software, and developed a software framework to actually position hundreds of scene images in camera-based, mathematically rigorous orientations in a virtual 3-dimensional space. Using a 2-D graphical user interface, a user can navigate this space (Figure 2). Since we impose rigorous camera and view modeling parameters, the user can achieve very close approximation to the original view of the scene, as if they stood where the camera was at that site. This makes research and visualization intrinsically linked.

Summary

Eye movement data provide valuable information in understanding viewing patterns of novice and experts. However, data collection in natural geological settings is demanding; some of the practical issues are discussed in (Evans et al., 2012). Here we discuss the analysis and visualization challenges. We present an approach involving algorithms and software tools to register eye movement data in a common reference frame, and to effectively visualize and annotate high-resolution imagery. Application to our growing expert/novice data set will be discussed.

Reference Cited

Evans, K.M., R.A. Jacobs, J.A. Tarduno and J.B. Pelz, 2012, Collecting and analyzing eye-tracking data in outdoor environments: Journal of Eye Movement Research, v. 5/2-6, p. 1-19.



(a) (b) (c)

Figure 1. Registering eye-tracker scene videos to high-resolution panorama. (a) Each observer wears a mobile eye-tracker. What they see is captured by a video camera attached to the headgear. (b) A single scene video keyframe is registered to the panorama using robust SIFT (i.e., Scale-Invariant Feature Transform) matching. (c) The whole video of one observer is registered on the panorama frame by frame.

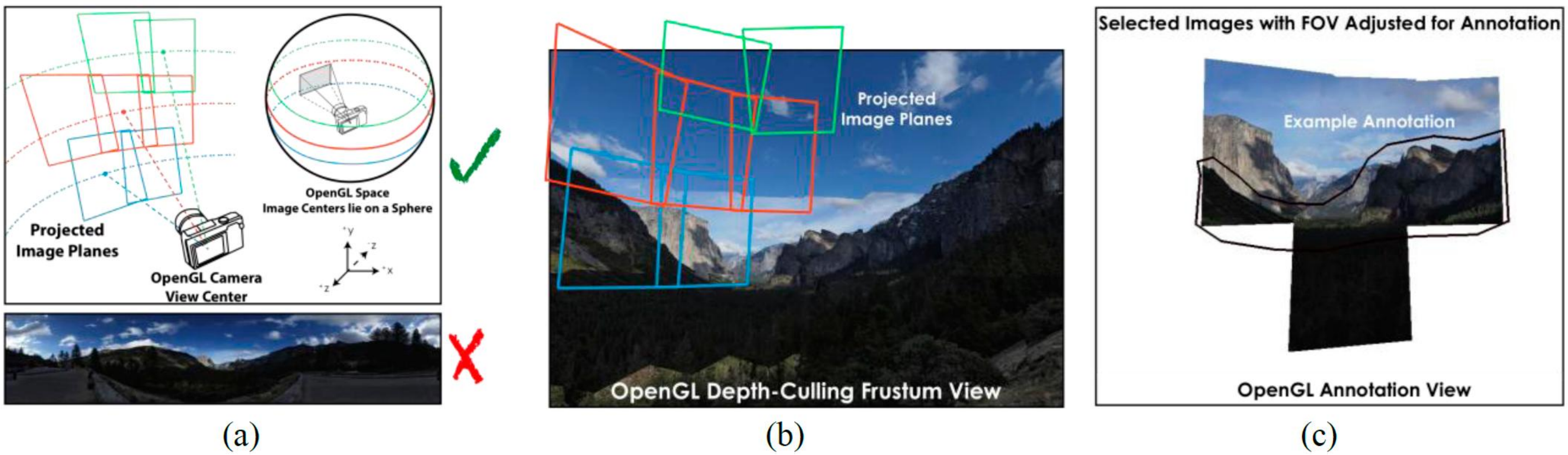


Figure 2. Intuitive viewing and annotating software. (a) The equirectangular-projection panorama is mathematically valid but confuses the geography and scale of the scene. Our 3D software view uses original imagery virtualized to positions by the rotation of the camera at the time of capture, (b) shows our geometrically corrected view (individual images not blended or color balanced to one another), and (c) exemplifies our software's ability to select individual frames to allow for an interactive polygonal annotation that is tied to each image, like a polygonal user-driven segmentation of the scene.