Large-Scale Overlays and Trends: Visually Mining, Panning and Zooming the Observable Universe

Timothy Basil Luciani, *Student Member, IEEE*, Brian Cherinka, Daniel Oliphant, Sean Myers, W. Michael Wood-Vasey, Alexandros Labrinidis, and G. Elisabeta Marai, *Member, IEEE*

Abstract—We introduce a web-based computing infrastructure to assist the visual integration, mining and interactive navigation of large-scale astronomy observations. Following an analysis of the application domain, we design a client-server architecture to fetch distributed image data and to partition local data into a spatial index structure that allows prefix-matching of spatial objects. In conjunction with hardware-accelerated pixel-based overlays and an online cross-registration pipeline, this approach allows the fetching, displaying, panning and zooming of gigabit panoramas of the sky in real time. To further facilitate the integration and mining of spatial and non-spatial data, we introduce interactive trend images—compact visual representations for identifying outlier objects and for studying trends within large collections of spatial objects of a given class. In a demonstration, images from three sky surveys (SDSS, FIRST and simulated LSST results) are cross-registered and integrated as overlays, allowing cross-spectrum analysis of astronomy observations. Trend images are interactively generated from catalog data and used to visually mine astronomy observations of similar type. The front-end of the infrastructure uses the web technologies WebGL and HTML5 to enable cross-platform, web-based functionality. Our approach attains interactive rendering framerates; its power and flexibility enables it to serve the needs of the astronomy community. Evaluation on three case studies, as well as feedback from domain experts emphasize the benefits of this visual approach to the observational astronomy field; and its potential benefits to large scale geospatial visualization in general.

Index Terms—Data fusion and integration, scalability issues, geographic/geospatial visualization

1 Introduction

Advances in data acquisition technology enable astronomers to amass large collections of complementary data, ranging from large scale, gigabit images to spectroscopic measurements. With the insight gained by these observations, researchers can better understand the happenings in our own galaxy by studying similar events in distant ones.

However, due to the increasing scale and variety of data sources, astronomical workflows are becoming cumbersome. To gather the data needed for a particular study, astronomers query multiple surveys for images, cross-correlate complementary images of the same object or set of objects, search multiple catalogs for potential additional details, then flip back and forth between these spatial and non-spatial details and the image context. This process is tedious, as well as challenging, and can often take hours to complete. As a result, time that could be spent analyzing data is instead spent mining it.

Inspired by an analysis of observational astronomy workflows, we propose a web-based visual infrastructure for the interactive navigation and mining of large-scale, distributed, multi-layer geospatial data. We introduce an automated pipeline for cross-correlating image data from complementary surveys, and we enable the visual mining of catalog information in conjunction with the large scale image data (Fig. 1). A spatially indexed, hardware-accelerated, client-server backbone allows fetching, displaying, panning and zooming of gigabit panoramas in real time.

The contributions of this work (extended from our Best Paper Runner-Up Award [1] at the IEEE Large Data Analysis and Visualization Symposium 2012) are as follows: 1) a formal analysis of the data and tasks specific to the observational astronomy domain; 2) the design of a client-server architecture for the interactive navigation of large scale, complementary astronomy observations; 3) two compact, scalable visual abstractions—hardware-accelerated pixel-based overlays and trend images—to enable the interactive mining, panning and zooming of these data; 4) a web-based, cross-platform implementation of this approach; and 5) the application of this approach to observational astronomy data through three case studies.

2 Related Work

Multiple attempts have been made to facilitate the observational astronomy workflows. However, none integrate large

- T.B. Luciani, S. Myers, A. Labrinidis and G.E. Marai are with the Department of Computer Science, University of Pittsburgh, Pittsburgh, PA 15206.
  E-mail: seanmyers0608@gmail.com, {tbl8, labrinidis, marai}cs.pitt.edu.
- B. Cherinka and W.M. Wood-Vasey are with the Department of Physics and Astronomy, University of Pittsburgh, Pittsburgh, PA 15206.
  E-mail: wmwv@pitt.edu, havok2063@gmail.com.
- D. Oliphant is with Google, Pittsburgh, PA 15217.
  E-mail: tbl8, marai@cs.pitt.edu.

Manuscript received 20 June 2013; revised 6 Feb. 2014; accepted 26 Feb. 2014. Date of publication 13 Mar. 2014; date of current version 29 May 2014.
Recommended for acceptance by G. Andrienko.
For information on obtaining reprints of this article, please send e-mail to: reprints@ieee.org, and reference the Digital Object Identifier below.
Digital Object Identifier no. 10.1109/TVCG.2014.2312008

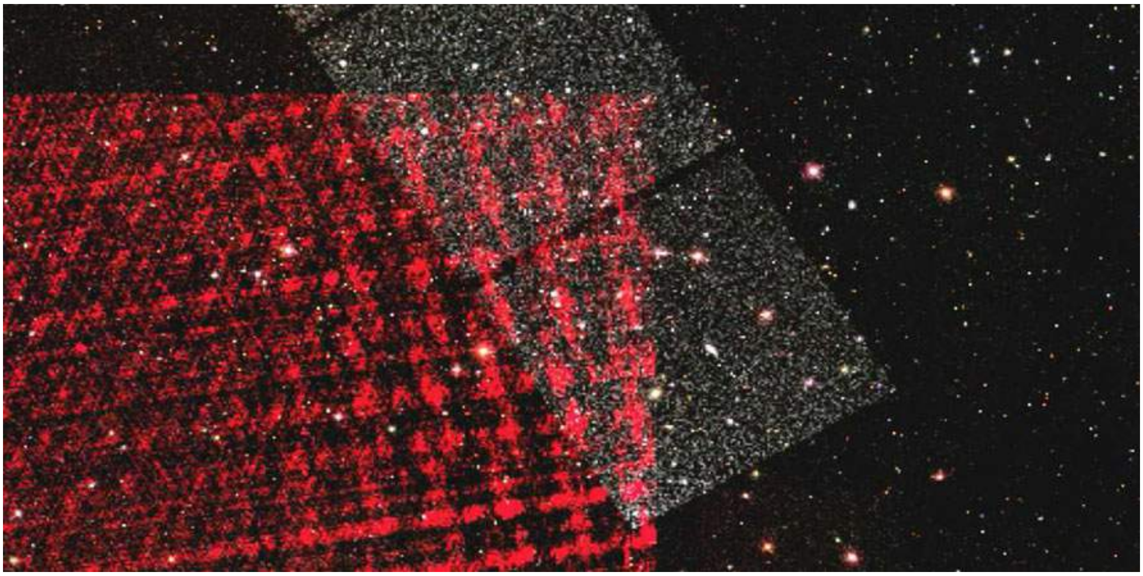

Fig. 1. Cross-correlated large-scale overlays of optical observations, radio-emission observations, and simulation results from the SDSS sky survey (color-on-black, full-coverage overlay), the FIRST sky survey (red overlay to the left), and the LSST data set (gray overlay, diagonal). Transparency can be interactively controlled for each overlay, enabling cross-spectrum analysis. Hardware-accelerated overlays coupled with a web-based client-server architecture allow panning and zooming of gigabit sky panoramas at interactive frame rates.

scale distributed astronomy research data while attaining interactive visual mining, panning and zooming framerates.

Google Sky (http://www.google.com/sky/) is a primarily educational, interactive, scalable view of the Sloan digital sky survey (SDSS). While it provides a friendly and clean interface, it also relies on local copies of the data to the exclusion of multiple surveys; which is a limiting factor for astronomy researchers. An additional drawback is its inability to integrate and share catalog data from multiple data sets. The National Virtual Observatory (http://www.us-vo.org/) (NVO) is a service designed primarily for aggregating and cross-matching information from multiple surveys. While it provides some form of catalog cross-registration, the NVO has a cumbersome interface which lacks a much-needed interactive visual component. The WorldWide Telescope (http://www.worldwidetelescope.org/) is a Microsoft Research, primarily educational project designed to allow users to view the Universe with a large, high resolution image of the sky. The ability to overlay multiple maps and visually cross-match objects is nonexistent. There is also a lack of connectivity with catalogs and other scientific data. Last, a variety of institutions have created web interfaces for accessing astronomical data, either for querying specific astronomy databases (http://www.sdss.org/dr7/, http://irsa.ipac.caltech.edu), or for aggregating data on many objects from multiple catalogs (http://ned.ipac.caltech.edu, http://simbad.u-strasbg.fr). These interfaces either lack a visual interface entirely or they provide only a static sky image to view a few objects at a time. Visual overlays of cross-matched data are non-existent and the user interfaces require a steep learning curve.

Attempts to work with gigascale image data have been made in other domains, though none have been applied directly to observational astronomy. Saliency assisted navigation identifies areas of interest in gigapixel images [2]. Through preprocessing and filtering regions of interest, discernible locations in a scene can be presented interactively. Kopf et al. [3] and Machiraju et al. [4] have also developed systems for dealing with gigascale and terascale image data. While these systems have complementary strengths in terms of the storage and the scale of the data being manipulated, each was generally designed for local geospatial data, and not for distributed geospatial sources.

Architectures for multizoom large-scale visualizations have also been explored. Space-Scale Diagrams [5] have been used in many geospatial applications [6], [7], [8]; they serve as a basis for our navigational approach. However, earlier applications were not designed to handle the magnitude of data described in this work. The challenges of indexing astronomy data are discussed in (http://www.star.le.ac.uk/ cgp/ag/skyindex); we use a new indexing approach, based on a *Geohash* [9]. A step further, ZAME [10] has used GPU-accelerated rendering to deliver interactive framerates to multi-scale visualizations. While the ZAME approach is beneficial to client-based applications that are able to provide full graphics support, web-based applications like ours pose more stringent constraints (e.g., limits on how many textures can be passed to a shader at once.) Furthermore, panning and zooming is a common problem among geospatial applications [11], [12], [13]. While many of these works focus on interactive techniques relevant to this project, the focus of this paper is an efficient architecture for viewing and cross-correlating gigabit image data.

Presenting multivariate data visually is also common among geospatial applications. Oriented Slivers provides a method to visualize multivariate information simultaneously on a single 2D plane, but becomes easily cluttered as the dimensionality of the data rises [14]. Heat maps [15] alleviate this problem by assigning each value a temperature and producing a color map based on the resulting heat combinations. While particularly beneficial in giving a general overview of data over large areas, heat maps are less useful in identifying individual data points. The approach we adopt for overlaying information in the sky, data driven spots (DDS), addresses both of these concerns via a pixel-based visualization [16].

The spectra data associated with sky objects are instantiations of ordered single-index tables; in which, however, the index itself is a property. Plain index tables have in general many possible visual mappings, from time-series line charts [17] and bar charts to graph-views, scatterplots [18], colored matrix cells [19] and 3D representations [20]. However, such mappings typically suffer from scalability issues. To address scalability concerns, we follow a pixel-based approach inspired by the compact representations of Keim [21]. Unlike existing pixel-based work which enables comparison through a small-multiple paradigm, however, our approach leverages alignment and resampling of the table data based on the index property. We further employ sorting properties of the object collection in order to generate a single, composite interactive image.

3 Domain Analysis

Our first contribution is a formal analysis of the domain data and tasks. This analysis provides a problem-driven basis on which further visualizations and interactions can be built.

Astronomy surveys cover a wide area of sky by acquiring many smaller images—some of which may overlap—over their targeted region. A given survey usually only covers a small fraction of the whole sky. However, the advent of large telescopes like the Large Synoptic Survey Telescope (LSST)will change dramatically, over the next decade, the scale of these surveys. Different surveys may or may not cover the same area of sky, resulting in possibly completely disparate or overlapping data sets. The Extended Groth Strip [22] for example, is one of the most observed regions of the sky, with upwards of eight different telescopes/surveys collecting data; this region is rich with multi-wavelength observations.

In our experiments we use data from three surveys, the Sloan digital sky survey, the Faint Images of the Radio Sky at Twenty Centimeters (FIRST), and simulated results from the Large Synoptic Survey Telescope. *SDSS* is an optical, wide-field, survey covering a quarter of the sky. Over the past 10 years, it has imaged a half a billion galaxies and taken spectra for a half a million, providing a massive leap in the amount of astronomical data (roughly 15 TB raw image data, stored as $2,048 \times 1,489$ pixel field images). *FIRST* is a radio survey of the sky, following the same path as SDSS. FIRST also covers about a quarter of the sky and contains roughly a million discrete radio sources [23]. *LSST* is a future optical full-sky survey, along the same lines as SDSS but of unsurpassed scale. It will cover ~20,000 sq. degrees of the sky, scanning the entire sky every three nights, in six photometric bands. LSST will image approximately 3 billion galaxies and will archive about 6.8 PB of images a year. As LSST has yet to acquire sky images, the LSST project has generated simulations of images of the sky to mimic and observe the observational prowess of the survey. Seven fields (189 unique image files), each covering ~10 sq. degrees, have been simulated.

3.1 Data Analysis

Astronomers use a variety of data formats to collect, organize, analyze, and share information about the observable Universe. The most common formats used are images and catalogs.

Images are rectangular snapshots (*tiles*) of regions of the sky, typically labeled with the spatial location of the region. Images in astronomy are usually stored as a Flexible Image Transport System (FITS) file. FITS files store image metadata in a human-readable ASCII header, and often include technical telescope details from when the image was taken. FITS files are extremely versatile, capable of storing non-image data such as spectra, 3D data cubes, multi-table databases, and catalog data.

Since the observable Universe is projected onto a sphere, the angle is the most natural unit to use in measuring positions of objects on the sky. Astronomers describe the coordinates of objects in right ascension (RA) and Declination (Dec). Similar to how longitude and latitude describe positions of objects on the Earth from a given reference point, right ascension and declination mark the position, in degrees, of an object with respect to the celestial equator.

Catalogs index all of the objects in a set of images. The catalogs contain spatial location information for every object imaged, along with any non-spatial properties collected or calculated from the observations (e.g., brightness, mass.) Each object in the catalog receives a unique identifier. Catalogs generated from the same survey will use the same unique object identifiers, making cross-matching within a survey straightforward. However, as is often the case, when the same object is observed in different surveys, it is assigned different identifiers for each catalog; this labeling makes cross-survey matching a non-trivial task. While the observed objects in each survey may not overlay exactly due to variation in each telescope's construction and parameters, they may still be physically and visually associated with each other (e.g., radio jets emanating from the center of a galaxy.)

Among the spatial and non-spatial object properties typically stored in catalogs, *spectra* play a particularly important role. Whereas imaging only captures broad features across the entire object, such as color or shape, spectra capture detailed information on the physical processes in and around the object, such as kinematics, temperature, distance from observer (redshift), and elemental content of gas associated with the object. Spectra specify the *wavelength* distribution of electromagnetic radiation emitted by a celestial object, as well as the *flux* (or intensity) of the object at those wavelengths. Large surveys typically acquire one spectrum per object, resulting in an extremely large number of spectra per survey; for example, SDSS has acquired spectra for ~1.6 million objects. Selecting particular classes of objects based on spectra and looking for trends can often lead to valuable insight about that object class.

Table 1 summarizes the data types typical of observational astronomy, as well as the visual mappings proposed in this work. In summary, the observational astronomy domain features large-scale, distributed, overlapping, multivariate data sets consisting of both spatial and non-spatial data: in a nutshell, data characterized by big volume and big variety. *Big volume* characteristics encompass: large image sizes (gigabit), impacting both rendering and interaction

TABLE 1
Data Analysis

Field / Attribute	Data Type (per point)	Visual Mapping
2D Fields		
Optical	RGB Value	Color Overlay
Radio	RGB Value	Color Overlay
Simulated	Intensity	Color Overlay
2D Field Attributes		
Projection Scheme	Formula	Location on Unit Sphere
Stripe Information	Numeric Tuple	Individual Tile Image
Catalogs		
Table of Search Results	Alphanumeric Tuple	Data Driven Spots (DDS)
Object Attributes		
Identifiers	Alphanumeric Value	Detail-on-Demand
Coordinates	Numeric Tuple	Pixel Coordinate
Redshift	Numeric Value	Pixel Intensity
Wavelength & Flux	Numeric Array	Pixel Intensity
Spectra	Numeric Array	Trend Line & 2D Plot
Image	RGB Value	Small Multiple

rates; fragmented images resulting in numerous image tiles; multiple scales; and large numbers of both observations and objects, often indexed in collections. *Big variety* characteristics include: data heterogeneity (e.g., catalogs, spectra and images); multiple data sources (surveys); and complementary domain expertise, e.g., expertise in supernovae as complementary to expertise in transient events. While the data is indexed by object location, uncertainties in the measured position make visual correlation particularly useful.

3.2 Task Analysis

The Universe is a complex structure with many physical processes governing its formation and evolution. While space-based telescopes can observe the full electromagnetic spectrum, cost and technical challenges preclude the design of a single all-purpose telescope. Instead, astronomers rely on many telescopes that observe specific regions of the electromagnetic spectrum and then cross-match the data sets to identify the same objects in each one. Astronomers must also manually seek out data related to a particular object.

Astronomical processes occur on many length scales, from small-scale features such as dust particles to large-scale features such as clusters and superclusters of galaxies. With observations usually pertaining to a specific scale at a time, it can be easy to lose the big picture of how all these processes are connected. Therefore it is advantageous to stitch multiple observations together to create a seamless zoomable image. This would allow astronomers to visually explore how stellar and galactic physical processes relate to the larger picture of galaxy groups and clusters.

Browsing massive astronomy collections of objects provides additional challenges. For example, selecting particular classes of objects based on spectra and looking for trends can often lead to valuable insight gained about that object class. However, when studying trends within astronomy objects of a specific class, typical efforts rely upon inspecting individual objects on an image in the sky or in a database table. The approach takes enormous amounts of time. Furthermore, outliers in the data set can often skew scientific results and must be located and removed before any analysis can be performed. Typical hypotheses relate to identifying trends, common properties, outliers, and discrepancies in collections of objects. Typical operations relate to grouping, selecting and analyzing objects from a collection.

Table 2 summarizes the tasks and challenges typical to observational astronomy. In summary, the observational astronomer workflow involves both queries of the type *what—where—correlated-with-what* and tasks of the type *browse—group—analyze* over multiple surveys at multiple scales. In conjunction with the big volume and big-variety of the data, astronomers seek the ability to interact with and compare multi-field data for a large number of objects and images.

Last but not least, additional requirements gleaned from interviews referred to desired interaction rates, ease of use, learning curve, and cross-platform desirability.

4 Design and Implementation

Based on the domain data and task analysis, we design a pipeline for the interactive exploration of observable astronomy data. Given the multiple, distributed sources of data, and the scale of the data, we follow a client-server model (Fig. 2). The online processes of this architecture are user-demand driven and occur in real-time. While, in a certain sense, our work aims to create a "Scientific Google Sky", we note that due to different requirements Google Sky uses a different—albeit unpublished—infrastructure, organization and implementation than our system.

Our server handles requests for image data and catalogs; it includes an offline module for preprocessing astronomical images. Where applicable, to enhance real-time panning and zooming (and thus help support task types T1 and T3), we assign prefixes to images, then organize and store them in a spatially-indexed, prefix-matching structure (*Geohash*). We store catalogs locally. Our online data driven spots module abstracts catalog results into an image overlay (task type T2). Our additional Trend Image interactive module allows the users to visually construct and browse collections of objects (tasks T4 and T5). Finally, our master Communication Module handles communication with the client, and interfaces with the data and the other server modules.

TABLE 2
Task Analysis

	Task	Visual/Interaction Mapping	Technical Challenge
T1	Pan and zoom in real-time	Panning & Zooming	Real-time infrastructure & interaction
T2	Analyze spatial distribution of objects	Object Overlays	Scalable visual abstraction
T3	Cross-correlate 2D image fields	Filtering on Overlays	Image cross-registration pipeline
T4	Identify trends & outliers in an object-collection	Interactive Trend Images	Visual abstraction
T5	Group objects according to properties	Linked Views	Interaction design
T6	Inspect object properties	Linked Views & Details-on-Demand	Visual design

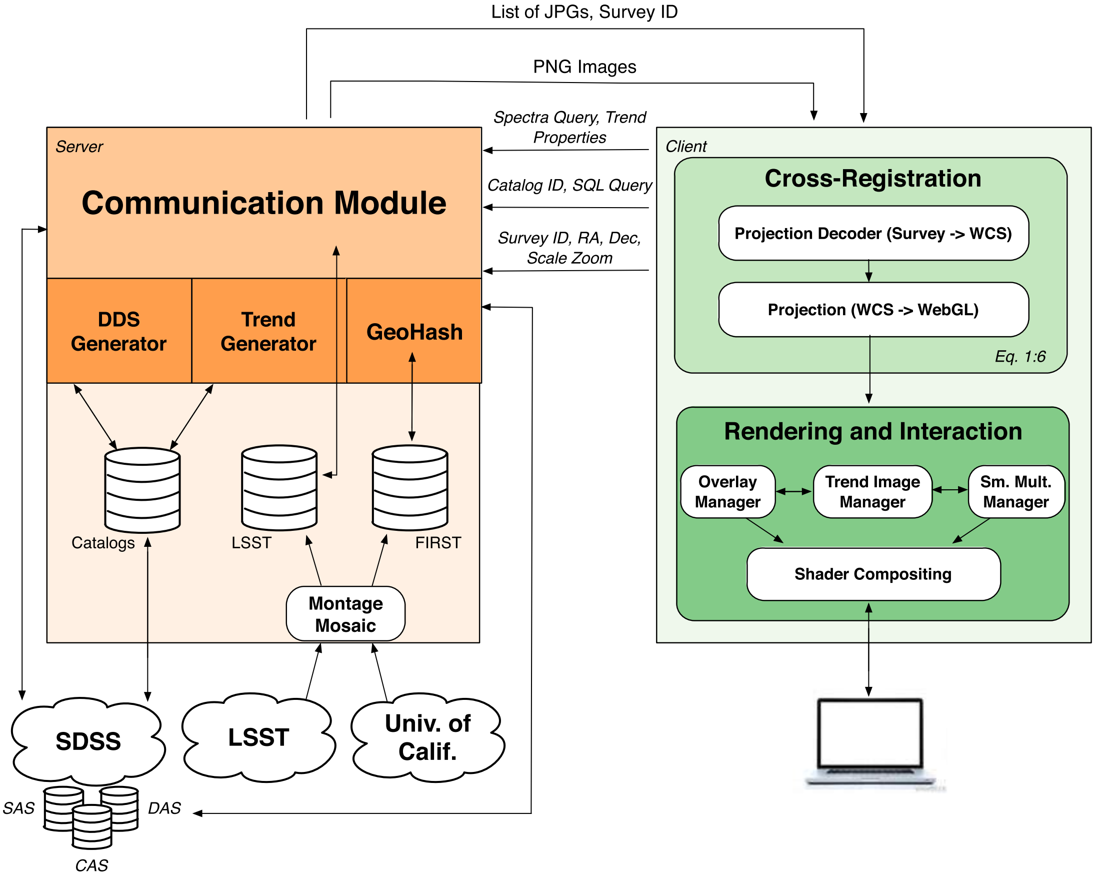

Fig. 2. Client-server architecture for the interactive exploration of observable astronomy data. On the server side, our offline module (light red) preprocesses raw astronomical data sets through *Montage Mosaic*; where applicable, we assign prefixes to offline images, organize and store them in a spatially indexed, prefix-matching structure (*Geohash*). Online server-modules abstract catalog results into visual DDS overlays (DDS Generator) and handle the construction of trend images (Trend Generator). The client handles the tile stitching and cross-registration process into Gigabit, zoomable panoramas, manages the rendering and interaction for the overlay view, the trend and small multiple views; and composites the images using hardware-accelerated shaders. The only third party tools are *Montage Mosaic* and the MongoDB powering the *Geohash*.

The client handles the requests from the user and the view management process. We support through a separate module the tile-stitching and cross-registration pipeline (T1 and T3). A second module controls the rendering and interaction processes through a web-based interface; ultimately, the module presents the stitched images and catalog results to the users in the form of gigabit panoramic overlays, interactive trend images, and small multiples (tasks T1 through T6). As the user navigates the sky, the client queries the server with the current field-of-view or desired catalog information.

Below we describe in detail the server and client modules which, independently and in conjunction, meet the technical challenges identified in Table 2: real-time panning and zooming capabilities, an image cross-registration pipeline, and scalable visual abstractions and interactions. The online modules are implemented using web-based technologies, including WebGL, HTML5 and Javascript.

4.1 Data Retrieval and Preprocessing

We perform image data retrieval and preprocessing on a per-survey basis. To ensure survey and data set compatibility, we extract the RA/Dec coordinates for each image tile so that images from multiple surveys will be properly aligned.

For surveys which benefit from an online programmatic interface, like SDSS, our system implements simple scripts to access the data remotely. Image data for the SDSS survey are stored remotely through various data releases; each release consists of FITS files and frame images. To access the survey data, our server sends SQL queries and fetches online images from the SDSS Data Access Server (DAS) and the SDSS Science Archive Server (SAS).

When a programmatic interface does not exist (e.g., FIRST or LSST) we fetch the sky images a-priori and store them locally. There are 30,500 FIRST image files, requiring 300 GB storage. To ensure compatibility between surveys, the raw data is processed using the third-party tool Montage Mosaic [24], to extract the image data from the raw FITS format; the resulting images are named according to the RA/Dec center of the image.

In the case of the FIRST survey, we perform further optimization to reduce the rendering load when a large area of the sky is being viewed. To this end, we use custom Matlab code to generate a pyramid of image tiles, with four levels (number of levels empirically determined for demonstration purposes) of decreasing resolution. We obtain tiles through repeated Gaussian filtering followed by subsampling. We perform this entire preprocess once for the data set, averaging a 30 second generation time per tile. Once the

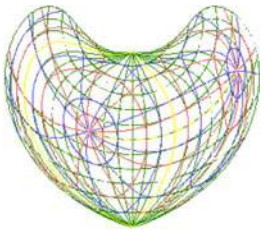

Fig. 3. Depending on the survey, astronomical data can appear in different map projections (shown in different colors above) [26]. While about 25 different projections are common to astronomy, there is no limit to the number of possible projections available.

local images are preprocessed, we spatially index them for quick access into the geospatial index powered by MongoDB [9], an open source document-oriented NoSQL database system. We hash the coordinates of the image tiles as string-based prefixes through MongoDB's *Geohash* table.

We do not map locally, however, LSST images to multiple levels of detail, since the domain experts anticipate a future programmatic online interface. Because this small LSST test data set is privately owned and accessed, we perform the entire procedure a-priori and store locally all images. *Montage Mosaic* and *MongoDB* are the only third-party tools in our system.

Catalog Data. Aside from sky survey panoramas, we create more specific overlays from user-performed searches over catalogs. We retrieve catalog data from the SDSS server and store the data locally into a mySQL database.

4.2 Cross-Registration and Online Overlays

To support task types T1 (pan and zoom), T2 (analyze spatial distributions) and T3 (cross-correlate images), we follow a cross-registration and online overlaying approach.

Sky Panoramas. We create sky panoramas (T1) and cross-correlate images (T3) by stitching together multiple astronomy images into a seamless, zoomable, pixel-based abstraction. Depending on the survey, astronomy images can appear in different map projections. While about 25 different projections are common, the number of possible projections is not limited (Fig. 3). In our approach we use the World Coordinate System (WCS) specification [25]. Our custom code converts to WCS a variety of image coordinates given in different projection schemes. SDSS's projection scheme is Gnomic (TAN), an azimuthal projection, given by [26], equations (54) and (55)]. The FIRST radio survey uses a Slant Orthographic projection (SIN), also an azimuthal projection, and is given by equations (59) and (60) in [26]. The LSST simulated data set uses the TAN projection scheme, similar to SDSS.

For the actual cross-registration and stitching we project the images on a viewing sphere. Our sphere is an abstraction of the sky as viewed from Earth, with the camera located at the center of the sphere.

To overlay sky images for viewing, our next step is to convert the WCS coordinates into the native WebGL graphics coordinates. The standard WCS Cartesian coordinate system is a right-handed coordinate system with the positive x, y, and z axes pointing outward, to the right, and up, respectively. In the WCS spherical coordinate system, the angle θ increases clockwise starting from the positive z-axis, and the angle ϕ increases counter-clockwise starting from the positive x-axis. In contrast, in the right-handed WebGL graphics coordinate system the positive x, y, z, axes point to the right, up, and outwards, respectively. Furthermore, the angle θ increases clockwise starting from the negative x-axis, and the angle ϕ increases counter clockwise starting from the negative y-axis. Due to these differences between the standard and WebGL coordinate systems, a transformation has to be applied to convert from the world RA/Dec coordinates to the WebGL spherical and Cartesian graphics coordinates. We derive this transformation as

$$\phi = (90^\circ - Dec) \tag{1}$$

$$\theta = (270^\circ - RA) + 360^\circ \quad ; when \quad RA > 270^\circ \tag{2}$$

$$\theta = (270^\circ - RA) \quad ; when \quad RA \leq 270^\circ \tag{3}$$

$$x = \sin(\phi) * \cos(\theta) \tag{4}$$

$$y = \cos(\phi) \tag{5}$$

$$z = \sin(\phi) * \sin(\theta), \tag{6}$$

where Equation (1) spells out, for clarity, the $Declination$ rotation into the WebGL angle ϕ, and Equations (2) and (3) spell out the $Right\ Ascension$ rotation into the WebGL angle θ.

Following the above transformation, our approach maps sky images to the unit viewing sphere. Aside from projecting each image tile onto the sky, our cross-registration and overlay module also facilitates zooming in and out to account for telescope parameters, and changing transparency.

DDS Overlays. To support task type T2 (analyze spatial distribution of object), we generate custom data-driven-spots overlays from catalog data. In response to the client requirements—desired resolution of the output image, the minimum and maximum RA/Dec values, attribute thresholds, desired color-mapping, and any other optional filters on the other parameters present in the catalog database—our server creates one or more new PNG images from the catalog database. The RA/Dec columns in each tuple are used to position the drawing within the image. The closer the value of the key attribute is to the maximum threshold, the brighter the pixel will be at that location. All data tuples are added to the images, which we then compress and return over the network to the client application.

Online Compositing. To create a visual abstraction of multiple data sources (task T3), pixels are further composited online into transparent overlays using the WebGL GLSL fragment shader. WebGL has the advantage of performing computations exclusively on the client machine GPU, leaving the CPU available for user interaction. To optimize Javascript memory use and texture loading we implement local garbage collection; this optimization helps prevent HTML5 from bottlenecking interaction while rendering texture objects.

The client we implement receives the images, cross-registers them through the approach described above, computes the pixel-based overlay and displays it. The client finally renders the scene, where the visualization scenegraph consists of the viewing sphere with the camera at the center looking out.

4.3 Interactive Trend Images

As outlined in Table 2, grouping and regrouping spatial objects according to their properties is a common astronomy task (task types T4 and T5). The custom overlays we enable—based on catalog queries (Section 4.2)—facilitate tasks which analyze scalar properties of the objects. Opacity-control further enables establishing correlations amongst multiple object properties.

However, some object properties are non-scalar, but ordered tuples or indexed-arrays for which the index itself is an object property. For example, object spectra are ordered tables of (key, value) pairs in which each key is a specific wavelength and the value is the flux at that wavelength. Analyzing such table properties across collections of objects, in order to establish trends or identify outliers, can be enormously time-consuming. We note that disparate features such as data artifacts that exist in a small set of spectra (amongst the larger pool) tend to be difficult to identify algorithmically or analytically, due to their ill-defined nature and randomness. To facilitate this process, we propose a second compact, pixel-based visual abstraction called an interactive trend image.

Trend images are a novel visual abstraction which relies on aligning and resampling property-indexed table data into a pixel-based representation. The abstraction further requires and leverages sorting properties of the data across an object collection. The trend abstraction builds on the astronomy concept of a 2D composite image [27].

The input to the trend image abstraction is a collection of objects. Because of their nature, the objects are guaranteed to have at least one scalar property suitable for sorting, p_{sort}—the distance from Earth to each object. Each object also has an indexed-table property in which the keys (a.k.a. index) are themselves a property of the object. We map the values in each object's table to a row of pixels, and then combine all the rows corresponding to the object collection into a trend image as described below.

Let N be the total number of objects in the collection. Let M be the total number of samples in each object's table-property—for example, flux, $\rho^i_{1:M}$, indexed by ordered wavelength, $\lambda^i_{1:M}$; where i is the object index $(1:N)$. To generate the trend image, we resample and align the data across the object collection. We generate first a uniformly-spaced basis for the collection of objects returned by the client query. The uniform basis b ranges from the minimum and maximum keys over all the N queried objects, and it is incremented by the desired horizontal resolution r of the image (specified by the client)

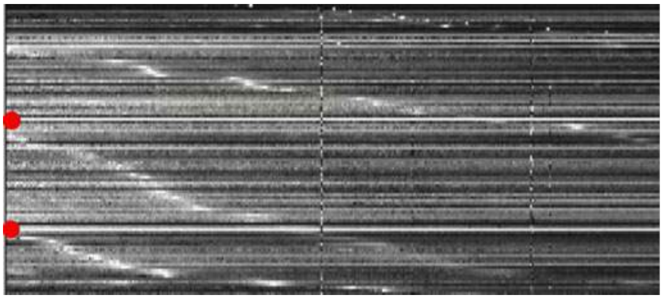

Fig. 4. Trend image for a test data set containing 100 quasi-stellar (quasar) objects. Each pixel-row corresponds to the spectrum of a quasar object; quasars are sorted vertically according to their redshift (a distance-related measurement). Note how outliers—quasars with unusual spectra (marked in red)— immediately stand out.

$$b = \left[\min_{\substack{i=1:N\\ j=1:M}} \lambda_j^{(i)} : r : \max_{\substack{i=1:N\\ j=1:M}} \lambda_j^{(i)} \right]. \tag{7}$$

We next generate for each object a normalized table of values, ρ_{norm}, through resampling and interpolation over the uniform basis

$$\rho^{(i)}_{norm(k)} = \rho^{(i)}_{k-1} + \left(\rho^{(i)}_{k+1} - \rho^{(i)}_{k-1}\right) * \frac{\left(b_k - \lambda^{(i)}_{k-1}\right)}{\left(\lambda^{(i)}_{k+1} - \lambda^{(i)}_{k-1}\right)}, \tag{8}$$

where $i = 1:N$, $k = 1:|b|$

Next, a pixel row is generated for each object. Pixels are organized from left to right in the table order and are individually mapped to a color in the HSV space based on the property values in the normalized table.

For color encoding, we calculate the hue of each pixel, H, its saturation S and value V as follows:

$$H^i_k = (\gamma_{high} - \gamma_{low}) * \left(1 - (b^{(i)}_k - b_{min})/b_{max}\right) \tag{9}$$

$$S^i_k = 1 \tag{10}$$

$$V^i_k = \left(\rho^{(i)}_{norm(k)} - \rho_{low}\right)/(\rho_{high} - \rho_{low}), \tag{11}$$

where $i = 1:N$, $k = 1:|b|$, and $(\gamma_{low}, \gamma_{high})$ and $(\rho_{low},\ \rho_{high})$ are the user-desired color range, respectively the desired property-mapping range.

In the example above, H encodes the wavelength and V encodes the flux value. Other mappings are also possible: for example, encoding flux alone in a grayscale image (Fig. 4); encoding rest-frame wavelength ($b^i_k/(1+z^i)$, where z^i is the object redshift) as hue; or encoding magnitudes for objects which do not have spectra, but do have broadband photometric colors which span specific wavelength regions.

To collect data for an interactive trend image, our client sends to the server queries for sets of spatial objects, and their desired properties. The server Trend Generator module (Fig. 2) fetches and processes the data for each object matching the query, and caches the object properties of interest into a local mySQL database. The user may also specify a color mapping range (γ_{high} and γ_{low}), the desired property-mapping range, and the desired horizontal resolution r of the trend image.

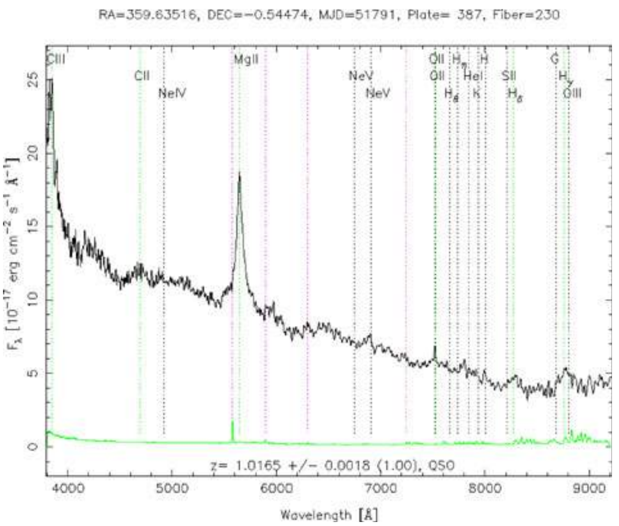

Fig. 5. On-demand spectrum for a Fig. 4 quasar object. The spectrum plot contains dotted lines to indicate the wavelengths at which atoms are absorbed or emitted.

The server returns to the client the collection of pixel-based rows, as highly compressed PNG images. In turn, the client assembles the composite image from the individual rows using the sorting property p_{sort}—for example, the object distance from the observation point. The client also facilitates alternative sortings, where available, and further interaction with the data. Sorting the objects reveals trends in the data, and helps identify outliers. Further interaction through a fish-eye lens and details-on-demand enables the individual analysis of object properties.

Fig. 4 shows an example grayscale trend image generated for a test data set containing 100 quasi-stellar (quasar) objects—extremely remote and massive celestial objects. Since quasars are visually similar in appearance to dim stars, they are difficult to identify from examination of image overlays alone. Instead, astronomers differentiate quasars from other stars by analyzing their spectrum (Fig. 5). The trend image shows the quasar spectra as horizontal rows, while the vertical sorting property is the quasar redshift. In this representation, key identifying features become apparent, such as trends in emission lines present in quasars (log-style curves in the figure). Outliers such as quasars with incorrect redshifts or with unusual spectra immediately stand out, as well.

The trend image abstractions allow the user to quickly, and intuitively, identify which spectra amongst the large data set are unlike the others, or those that belong together in groups. Furthermore, as we show in Section 5, the identification of outliers from the trends is key in identifying both (1) new objects of interest that will bring insight; (2) problems with the previous steps in the data analysis and processing.

4.4 Rendering and Interaction

Panning and Zooming Large-Scale Overlays. To enable interactive panning and zooming, an Overlay Manager maintains the current viewing location and parameters, as well as a list of the image tiles currently in the view. The manager sends to the server requests for new images, when needed. Panning the view maps mouse motion to updates in the view range. Zooming also computes and maps the new scale to updates in the viewing range. If the updated range covers images that have not been fetched yet, the manager requests for those image tiles to be sent out to all overlays that are listening to the current view. Each of those requests is handled asynchronously.

Interactive trend image rendering and the on-demand spectra are handled as a separate process, in a side panel from the sky-view. Similar to the custom-overlay creation process, users can specify the parameters to be used in the construction of the trend image and then interact with the resulting set of spectra.

Trend Image Interaction. On the client side, we provide interaction techniques to further help support workflows related to object-group analysis. Hovering over the trend image highlights individual rows for inspection; a GLSL shader fish-eye lens can further be activated to magnify a selected row.

Details-on-demand (task T6) are also provided: such as the spectrum for that object, the object ID, and its coordinates. Selecting the plot opens a new tab in the browser that links to the object's SDSS reference page. This allows the user to drill-down for additional properties.

Small Multiples. The objects in the current collection may not be located in the same area of the sky, and thus may not be visible or distinguishable in a single gigabit overlay view. To facilitate the analysis of such collections (task T6), we provide an additional small multiples view of the collection of objects represented by the trend image. In order to retrieve a postage-stamp image of each object, the client sends the list of object coordinates to the server. The server then connects to the SDSS Science Archive Server and requests cutout field images centered around the coordinates it received. Finally, the images are returned to the client and rendered in the small multiple view.

Linked Views. Dynamic queries and view linking further allow the users to refine the collection of objects represented by the trend image (task T5). Through query interaction in a linked panel the client can both filter out incorrect results as well as add new entries to the visual abstraction. Right-clicking on an object line in the trend view enables the user to jump to the object's image in the gigabit overlay panel, and thus make further inferences based on the object's celestial neighborhood.

Overall, the use of the trend pixel-based abstraction allows viewing and analyzing data for large collections of objects, while efficiently using the available screen real estate. Along with the zooming, filtering, details-on-demand, dynamic queries and linked-views interaction techniques described above, the small multiples representation aides in both identification and filtering of the results. Furthermore, condensing the data from multiple FITS files into PNG images enables us to avoid the transfer of large FITS files to the client. The approach reduces thus the bandwidth usage between the client and server.

5 Results

In this section we report on the performance of our approach. We first measure the precomputation of the

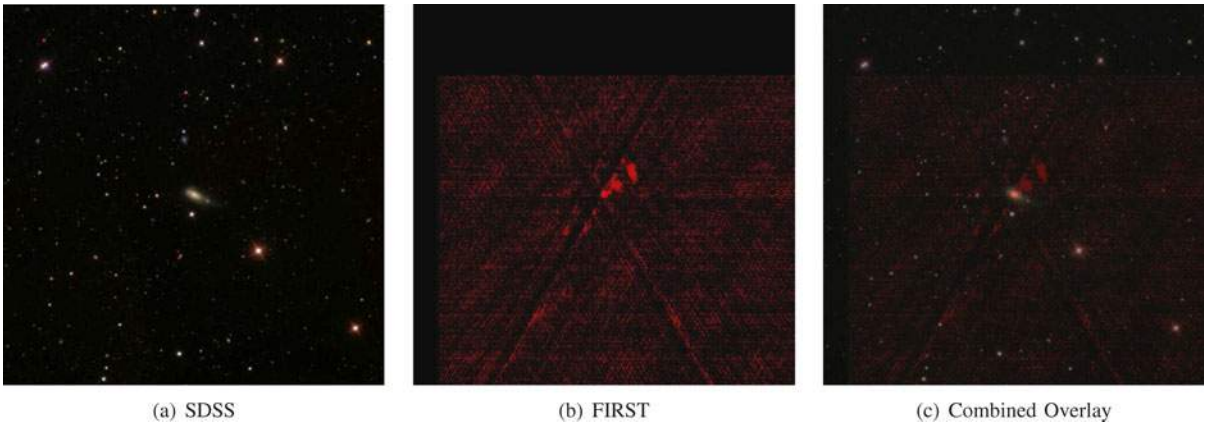

Fig. 6. Images of UGC 08782 from two surveys. Fig. 6(a) shows an optical image of the galaxy from the SDSS while Fig. 6(b) shows a radio image of the same galaxy from FIRST. When overlaid in Fig. 6(c), the connection between the two as radio emission emanating as jets from the central black hole of the galaxy becomes immediately clear.

FIRST images stored on the back end of the pipeline; conversion to raw images, mosaicking, and reprojection. Next we report rendering speeds with varying amounts of image data presented to the user. We then present a case study where domain experts perform an overlay-based analysis with our tool and report their findings. A second case-study examines the benefits of interactive trend images to browsing and grouping tasks. The third and most complex case study follows an integrated workflow through our system. Finally, we report feedback from repeated evaluation with a group of five astronomy researchers, as well as from three astronomy workshops where the tool was demonstrated and made available to astronomers for testing. The workshops correspond to separate interest-based groups of astronomers associated with particular sky surveys; each workshop featured more than 30 participants. The implementation of our approach is in its beta release.

Preprocessing. Offline precomputation of the FIRST images is the most time consuming part of the pipeline; however, this stage only has to be performed once when the data is first acquired for a survey. Each image takes between 30 and 40 seconds to generate, with 20 seconds of the process dedicated to reprojecting the image into the WCS map projection. Depending on the sky coverage of the survey, this preprocessing can take anywhere from a week to a month. In the case of FIRST, it took fifteen days to compute all of the images needed for tiles using a server running CentOS 6, Dual 6 Core processor at 24 GHz, and 32 GB RAM.

Performance. The initial data retrieval and loading stage varies depending on the source the images arrive from. To retrieve FIRST images from our server, a loading time of 50-200 ms is incurred for sizes varying between 400-700 KB. Retrieving LSST images from our server incurs a loading time between 200-400 ms with sizes varying between 4-5 MB. Finally, SDSS loading times are slightly higher, typically incurring 750-1250 ms with sizes varying between 60-70 KB. These speeds can vary greatly depending on the bandwidth and load of the SDSS servers at the time of use.

Trend image generation takes between 25 seconds (for a 50 object collection) to 130 seconds (for a 200 object collection), on a Macbook Pro, 2.3 GHz Quad Core i7 with 8G of RAM. Returning the spectrum associated with an object takes between 180 and 220 ms, dependent on the speed of the network.

Once the images are fetched, the rendering speed hovers at 45 frames per second on a Windows 7 Machine, Quad Core i5, 16 GB RAM. This allows interactive panning and zooming to regions of interest. Our web-based implementation has been tested on multiple browsers such as Safari, Chrome and Firefox.

5.1 Case Study: UGC 08782 - A Dusty Elliptical

Fig. 6 shows how the cross-correlation and interactive visual navigation of SDSS and FIRST can be used in tandem for immediate gains in astronomy. Two of our co-authors are senior astronomy researchers and provide the following case study (completed in minutes) and feedback.

Fig. 6a shows an optical image from UGC 08782, a bright elliptical galaxy at a redshift of 0.045. The morphology of this galaxy was originally ambiguous between a spiral and a dusty elliptical, exhibiting dust lanes and disturbed morphological features. Dusty ellipticals are often seen to show signatures of an active galactic nucleus (AGN) [28], [29]. Some of these AGN exhibit jets, which tend to be perpendicular to the dust lanes. One way to test if UGC 08782 fits these trends is by checking its SDSS spectrum, viewing the optical image, and searching for radio counterparts [30]. Fig. 6b shows radio observations from the FIRST survey of the same region, which detected several interesting features. The image in the radio looks quite different. There is a single bright point where the optical galaxy ought to be and two bright patches extending to the upper right. Due to the differing resolutions and sensitivities of the surveys, it is unclear looking at the individual images whether the FIRST emission is from a unique object or associated with UGC 08782. Without our system, associating the FIRST emission with an optical counterpart would require manually searching optical catalogs for nearby objects and match on position, ranking by closest proximity; and then carefully overlaying the two locations using photo-editing software. The astronomers estimate the process would require 30 minutes to one hour, end to end.

In contrast, when the images are viewed together (Fig. 6c), in under one minute, using our online overlays,

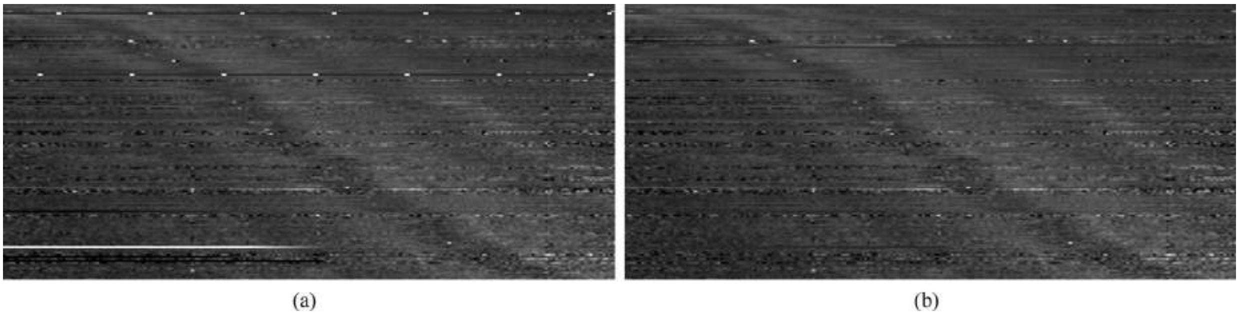

Fig. 7. Two trend images generated for the Type Ia Supernovae case study. This interactive visualization allows users to see the spectrum of interest in order to identify unusual outliers. In the first image, 200 Type Ia Supernovae objects are ordered in increasing redshift top ($z = 0.15$) to bottom ($z = 1.0$) and increasing wavelength left to right. The general broad features of the Type Ia can be seen in the dark and light bands that represent characteristic features of explosions. These bands smoothly trace-back over redshift, indicating that these objects form a consistent class. The first trend image (a) shows clear evidence of potential outliers: objects with unusual spectra that do not match the group. Upon further inspection of these objects, the astronomer removed the confirmed outliers and generated the reduced data set shown in (b).

the association between these two sources from different surveys is immediate. The bright radio point source lines up on the center of the optical galaxy, as it would if it were the nucleus of the galaxy. The two patches of radio emission in the upper right appear to emanate from the central point source, as a radio jet might. Not only does the overlay allow for a more efficient cross-matching, it also provides a nice framework for understanding the physical processes observed in each survey and how those processes are connected to one another.

5.2 Case Study: Trends in Type Ia Supernovae

The second case-study, completed by two different senior astronomy researchers, showcases the benefits of trend images in the analysis of large collections of Type Ia (“one-A”) Supernovae. Supernovae are stars that are undergoing catastrophic explosions, which can be classified according to their spectra. In particular, the spectrum of a Type Ia Supernova is characterized by a lack of hydrogen lines, a strong absorption line at 6,550 angstroms near maximum, and late-time spectrum iron-group emission lines. Identifying these spectrum-based features via direct catalog querying is, however, a laborious and intensive process. The process is further prone to errors such as inclusion of problematic data, or exclusion of a good object. To alleviate these shortcomings, the researchers were looking for ways to group, analyze trends, and regroup potential collections of Type 1a Supernovae.

The trend image mechanism enabled the researchers to quickly query the SDSS DR7 survey, then analyze and group 200 potential Type Ia Supernova objects (Fig. 7). In these trend images, the spectrum of objects is plotted as intensity along the *X* axis while the objects themselves are sorted along Y according to their redshift. Redshift is the factor by which the wavelengths of light have been stretched as it travels by the expansion of the Universe and provides a sorting in terms of how far back in time we are looking for each object, or equivalently a sorting in distance. Sorting along the redshift property let the researchers immediately see how spectral features (such as emission lines from particular elements) move in observed wavelength with redshift, and immediately observe whether the strength of those features is changing over time. Objects with incorrect redshifts or with unusual spectra immediately stand out because of the great mismatch with their neighbors.

Producing this first trend image from the raw observations (Fig. 7a) brought immediate attention to two distinct outlier classes: 1) the periodic bright spots in the otherwise grey lines (one line a few rows down from the top and another 1/3 down from the top), as well as three half-length continuous lines near the bottom (one white, two black); and 2) the individual bright pixels in spectra scattered among the data. These classes are visibly noticeable as outliers to the rest of the spectra in the image that—according to the researchers—would have taken many man-hours to discover by combing through each spectrum in the data set individually. Upon further examination of the individual spectra, the first class turned out to be a data reduction issue: unreasonable values in the released telescope data. The second class turned out to be an astrophysically interesting issue: emission lines from the region of the host galaxy underlying the supernovae.

After investigation and correction of the first issue, the researchers arrived at the new visualization in Fig. 7b. This second trend image has the powerful property that features which are vertical lines (constant observed wavelength) represent information about detector problems or atmospheric lines, whereas the properties of the supernovae themselves follow the curves of constant rest wavelength seen in the dark and light bands. The smooth trends seen in this visual representation of supernova spectra confirm that they exhibit similar features with one another and can be grouped together as one class of objects. The researchers took about 10 minutes with our system to complete this case study, with most of the time spent examining the outlier spectra. They estimate that, without our system, the study would require on the order of weeks of work.

5.3 Case Study: Spectroscopic Analysis of Galaxies

The third case study follows an integrated workflow through our system. In this study, a senior astronomer is interested in the spectroscopic analysis of galaxies, and in particular, in understanding how object colors, and absorption and emission features in spectra from different classes relate to each other. The astronomer began his analysis by examining first the properties of galaxy spectra in contrast

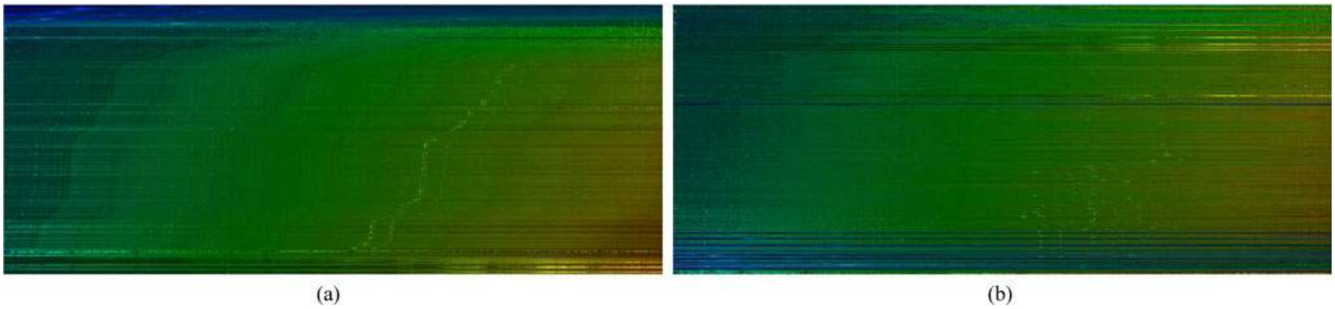

Fig. 8. Trend image for 200 galaxies, stars, and quasars in the direction towards the Galactic North Pole, from the Sloan digital sky survey. In the left image (a), each row is an object's spectrum, plotted along x in observed wavelength, and sorted along y in redshift, with smaller values at the bottom. Galaxies occupy the majority of the parameter space, in the central part of the image; stars are grouped in the noisier, bottom part; while quasars (top band) appear entirely blue. In the right image (b), the same object rows have been sorted by g-r color, with smaller (bluer) values at the bottom and larger (redder) values at the top. In this image, star-forming galaxies with strong emission lines features are blue in color, reflecting their young age. In contrast, elliptical galaxies tend to be redder, reflecting their older ages. At the very top, the researcher noticed a fair number of extremely red stars in the Milky Way North Pole.

to those of stars and quasars. To this end, he used our system to do a broad query over the Sloan digital sky survey for 200 astronomical objects within 1 degree of RA, Dec: 191.0, 26.0, a region pointing towards the Galactic North Pole. Specific search parameters included the RA, Dec, redshift, and g-r color (a star with a high g-r color is redder than a star with a low g-r color.)

Continuing to use the system, the researcher then generated the trend image corresponding to the spectra in this resulting data set (Fig. 8). To explore the trends in the data set, he sorted the trend image first by redshift. Sorting by redshift shows the stars, galaxies, and quasars existing in three distinct locations in this parameter space. Galaxies occupy the majority of the parameter space, in the central part of the image, with many clear absorption and emission features visible as dark, and bright lines, respectively, trending through the visual representation. For example, at the right of the image, the researcher could clearly see the strong H-alpha and SII emission lines in star-forming galaxies trending upwards and to the right. These features were also apparent in the one-dimensional spectra detail on demand. The sudden disappearance of the feature at higher redshift revealed a fair number of elliptical galaxies lacking these features. He noticed similar trends in the dark-banded absorption lines.

Stars and quasars occupy less of the parameter space in this particular region of the sky but were still instantly visible to the researcher in the trend image. The stars occupy the "noisier", lower part of the image and make their presence known by "breaking up" the nice trends seen in the galaxy spectra. Zooming in on the stellar spectra revealed to the researcher the presence of vertical absorption lines bands, that exhibit no trend with redshift. The researcher recalled that the stars had a redshift close to 0, due to the stars location within the Milky Way Galaxy. The quasars are visible in the upper part of the image. As quasars are galaxies that predominantly exist at high redshifts, they looked entirely blue.

Sorting on g-r color (Fig. 8 right) further revealed that the star-forming galaxies exhibiting strong emission lines features are also blue in color, reflecting their young age. In contrast, elliptical galaxies tended to be redder, reflecting their older ages. At the very top, the researcher noticed a fair number of extremely red stars in the Milky Way North Pole. He concluded these stars are thus likely a part of the Milky Way halo, a spheroid surrounding the disk containing clusters of old, red stars that have existed since the earliest formations of the Milky Way.

Having explored the trends in the various objects found towards the Galactic North Pole, the astronomer decided to inspect the objects themselves along with their local environment by bringing up the thumbnail images of the data set (see accompanying video, which can be found on the Computer Society Digital Library at http://doi.ieeecomputersociety.org/10.1109/TVCG.2014.2312008.) He found the galaxies in this direction to be an average mix of galaxies exhibiting a wide variety of properties.

Curious about the positional relationship between the North Pole stars, the galaxies, and quasars, the astronomer then decided to display an overlay of all the objects in the data set, color-coded by redshift, with marker differentiating between the three classes of objects. He found that all three classes appeared evenly distributed within the region specified in his query.

Finally, examining the full sky image of this region (see video, available in the online supplementary material), the astronomer visually noticed a large numbers of red stars in the halo of our Galaxy, supporting his deductions from the trend image. Wondering if any of the galaxies or quasars in the data set are producing radio emissions, he overlaid the FIRST radio survey and explored the overlapping regions. Following these observations, with no noticeable overlapping features between the radio and optical images, the astronomer decided to explore, in a future study, the question of whether or not stars and galaxies in the directions toward the Galactic center and anti-center exhibit similar colors as seen in the direction of the Galactic North Pole.

The astronomer estimates that completing the present study in the absence of our system would have required weeks of collecting, examining, and grouping the data. Typically, researchers would have approached the problem by first identifying the type of data they are interested in through papers, catalogs, or a perhaps a specialized interface. These data are specific to the question they are attempting to answer, and are seldom restricted to one particular survey. Our researcher and his group would then manually download locally all the data and/or catalogs

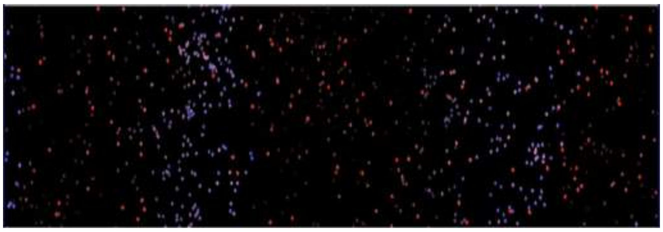

Fig. 9. 831 points resulting from searches over the Sloan digital sky survey catalog database are visualized efficiently using pixel-based overlays. Two query results based on two different attributes are overlaid (red for redshift, blue for the focal ratio of the telescope; brighter intensities correspond to greater values), revealing spatial patterns in conjunction to attribute overlaps.

satisfying their initial criteria. He would then display the imaging or spectroscopic data in a local software package, manually inspecting individual objects in this data set one at a time, all the while marking outliers, objects of interest, or interesting patterns. Classification of spectroscopic objects is often done by the identification of specific emission—or absorption—line features visible in the spectrum. This approach is rather time-consuming, involving the initial data location and acquisition, combined with manual inspection of spectra for spectral-line identification. Manual inspection, on an individual basis, of an object's spectral class could take anywhere from 1-30 seconds, depending on the astronomer's goals. Cross-survey correlations, and marking interesting or outlying objects, for further analysis may often take much longer. Performing this set of tasks for hundreds of objects at a time was an unreasonable time commitment for the astronomer.

In contrast, collecting, viewing and visually analyzing the results in our system highlighted practically instantly many aspects of the spectroscopic data set that would have been difficult to glean by examining the spectra one at a time. Completing this case study (highlighted in the accompanying video, available in the online supplementary material) with our system has only required 5 to 10 minutes. The astronomer has adopted our system as a research tool.

5.4 Domain-Expert Feedback

Feedback from repeat evaluation meetings showed enthusiasm for the tool. The domain experts considered the approach "an exciting and effective tool for visualizing all-sky surveys. Many of the tools required have been implemented effectively." The ability to compare images of the sky taken at different wavelengths simultaneously and to visually query catalogs was particularly appreciated, while the interactive navigation was considered on par with the much appreciated Google Sky interface. The interactive trend images attracted enthusiastic feedback. Astronomers stated that the interactive trend images allow them to more easily and quickly identify patterns and outliers in the data. The researchers are eager to use the tool in their research and in classrooms.

The workshop expert-users particularly appreciated the ability to combine separate sources of information without having to resort to cumbersome, external tools for image processing. As shown in the example in Fig. 9, overlaying catalog search results visually further enables queries of the *what-where-correlated-with-what* type. In this example, more than 800 points resulting from searches over the Sloan digital sky survey catalog are visualized efficiently using pixel-based overlays: two query results based on two different attributes are overlaid (red for redshift, blue for the focal ratio of the telescope; brighter intensities correspond to greater values), revealing vertical spatial patterns in conjunction to attribute overlaps. Fig. 1 further shows three cross-correlated overlays (partial coverage shown in the figure solely for static illustration purposes) of optical observations, radio-emission observations, and simulation results from the SDSS sky survey, the FIRST sky survey, and the LSST data set. Transparency can be interactively controlled for each overlay, enabling cross-spectrum analysis. The workshop researchers are interested in applying this prototype to specific problems such as browsing large sets of objects and galaxy identification. In toy demonstrations, the interactive trend images have already been used to *browse–group–analyze* several collections of objects, from galaxies to quasars; to great feedback and requests for immediate release to the astronomy community. Several astronomy research groups have expressed keen interest in integrating their data with our tool.

6 Discussion and Conclusion

Our approach enables the visual cross-correlation of sky surveys taken at different wavelengths, as well as the visual querying of catalogs. Furthermore, the combination of prefix-matching indexing, a client-server backbone, and of pixel-based overlays makes possible the interactive exploration of large scale, complementary astronomy observations.

New surveys can be flexibly added to the system, provided they specify the raw image data and the projection information of the telescope in standard FITS files. For surveys which benefit from a programmatic interface, our system would implement a simple script to access the data from the online interface. If a programmatic interface does not exist, the images would first need to be downloaded, organized in indices, and stored on local servers.

Our results show that pixel-based overlays and geohashing have the potential to generate scalable, interactive, graphical representations of astronomy data. This approach may allow us to overcome bandwidth and screen-space current limitations in astronomy visualization. The advantages of this approach are its versatility, flexible control on the client side, and visual scalability (to the pixel level), enabling the visual analysis of large data sets. Accessing graphics hardware through WebGL further provides the users with a rich, graphics-accelerated web experience.

The trend image visual abstractions naturally highlight the trends within objects of a given class. They also support the rapid identification of outlier objects in an object collection—be they outlier objects characterized by poor data/identifications, or outlier objects which have unusual physical properties. Interactive trend images similar to those depicted in Fig. 7 may be constructed with almost any property of the object of interest—such as distance, color, or time since a transient event began—for the Y sorting. These interactive representations provide a rapid way to look for correlations between properties of objects, but also take advantage of the human eye's ability to recognize patterns and detect outliers.

Finally, evaluation on three case studies, as well as overwhelmingly positive feedback from astronomers emphasize the benefits of this visual approach to the observational astronomy field. In terms of limitations, relying on streaming the data from remote sources is a concern as certain surveys do not provide programmatic access to their images.

In conclusion, we have introduced a novel approach to assist the interactive exploration and analysis of large-scale observational astronomy data sets. Our approach successfully integrates large-scale, distributed, multi-layer geospatial data while attaining interactive visual mining, panning and zooming framerates. From a technical perspective, we contribute a novel computing infrastructure to cross-register, cache, index, and present large-scale geospatial data at interactive rates. Large local image data sets are partitioned into a spatial index structure that allows prefix-matching of spatial objects and regions. In conjunction with pixel-based overlays and trend images, this web-based approach allows fetching, displaying, panning and zooming of gigabit panoramas of the sky in real time. In our implementation, images from three surveys (SDSS, FIRST, and LSST), and catalog search results were visually cross-registered and integrated as overlays, allowing cross-spectrum analysis of astronomy observations.

From the application end, we contribute an analysis and model of the observational astronomy domain, as well as three case studies and an evaluation from domain experts. Astronomer feedback and testing indicates that our approach matches the interactivity of state-of-the-art, corporate educational tools, while having the power and flexibility needed to serve the observational astronomy research community. Being able to quickly aggregate and overlay data from multiple surveys brings immediate clarity to inherently complex phenomena, reducing time spent managing the data while allocating more time for science.

Acknowledgments

This research was funded through US National Science Foundation (NSF)-OIA-1028162, NSF-IIS-0952720, and NSF GRFP-2012144824. They gratefully acknowledge the enthusiastic help and unwavering support of our collaborators Jeffrey Newman and Arthur Kosowski: no one could wish for better collaborators. They thank the AEGIS, SKA and LSST astronomy communities, the anonymous reviewers and the Pitt VisLab for feedback and interesting discussions; Rebecca Hachey, Ruhsary Rexit, Boyu Sun, and the cs1699 students for contributions to prototype codes; and the ADMT lab for occasional testing and feedback. Further acknowledgments to Noel Gorelick and Jeremy Brewer, for generously sharing their Google Sky development experience.

References

[1] T. Luciani, B. Cherinka, S. Myers, B. Sun, W. Wood-Vassey, A. Labrinidis, and G. Marai, "Panning and zooming the observable universe with prex-matching indices and pixel-based overlays," in *Proc. IEEE., Large-Scale Data Anal. Vis. Symp.*, Oct. 2012, pp. 1–8.

[2] C. Y. Ip and A. Varshney, "Saliency-assisted navigation of very large landscape images," *IEEE Trans. Vis. Comput. Graph.*, vol. 17, no. 12, pp. 1737–1746, Dec. 2011.

[3] J. Kopf, M. Uyttendaele, O. Deussen, and M. Cohen, "Capturing and viewing gigapixel images," in *Proc. SIGGRAPH Conf. Comput. Graph.*, vol. 26, no. 3, July 2007.

[4] R. Machiraju, J. E. Fowler, D. Thompson, W. Schroeder, and B. Soni, "EVITA: A prototype system for efficient visualization and interrogation of terascale datasets," Eng. Res. Center, Mississippi State Univ., Tech. Rep. MSSU-COE-ERC-01-02, 2000.

[5] G. W. Furnas and B. B. Bederson, "Space-scale diagrams: understanding multiscale interfaces," in *Proc. SIGCHI Conf. Human Factors Comput. Syst.*, 1995, pp. 234–241.

[6] B. B. Bederson and J. D. Hollan, "Pad++: a zooming graphical interface for exploring alternate interface physics," in *Proc. 7th Annu. ACM Symp. User Interface Softw. Technol.*, 1994, pp. 17–26.

[7] G. Furnas, "Generalized fisheye views," *Proc. SIGCHI Conf. Human Factors Comput. Syst.*, vol. 17, no. 4, pp. 16–23, Apr. 1986.

[8] H. Lieberman, "Powers of ten thousand: Navigating in large information spaces," in *Proc. ACM Symp. User Interface Softw. Technol.*, 1994, pp. 15–16.

[9] K. Chodorow and M. Dirolf, *MongoDB: The Definitive Guide.*1st ed. Sebastopol, CA, USA: O'Reilly Media, Sept. 2010.

[10] N. Elmqvist, T.-N. Do, H. Goodell, N. Henry, and J. Fekete, "Zame: Interactive large-scale graph visualization," in *Proc. IEEE Pacific Vis. Symp.*, 2008, pp. 215–222.

[11] C. Appert, O. Chapuis, and E. Pietriga, "High-precision magnification lenses," in *Proc. SIGCHI Conf. Human Factors Comput. Syst.*, 2010, pp. 273–282.

[12] W. Javed, S. Ghani, and N. Elmqvist, "GravNav: Using a gravity model for multi-scale navigation," in *Proc. Int. Working Conf. Advanced Vis. Interfaces*, 2012, pp. 217–224.

[13] W. Javed, S. Ghani, and N. Elmqvist, "Polyzoom: Multiscale and multifocus exploration in 2d visual spaces," in *Proc. SIGCHI Conf. Human Factors Comput. Syst.*, 2012, pp. 287–296.

[14] C. Weigle, W. Emigh, G. Liu, R. Ii, J. Enns, and C. Healey, "Oriented sliver textures: A technique for local value estimation of multiple scalar fields," in *Proc. Graph. Interface.*, 2000.

[15] D. Fisher, "Hotmap: Looking at geographical attention," in *Proc. IEEE Symp. Inform. Vis.*, 2007.

[16] A. Bokinsky, "Multivariate data visualization with data-driven spots," Ph.D. dissertation, Dept. Comput. Sci., The University of North Carolina at Chapel Hill, NC, 2003.

[17] P. McLachlan, T. Munzner, E. Koutsofios, and S. North, "Liverac: Interactive visual exploration of system management time-series data," in *Proc. SIGCHI Conf. Human Factors Comput. Syst.*, 2008, pp. 1483–1492.

[18] P. Craig and J. B. Kennedy, "Coordinated graph and scatter-plot views for the visual exploration of microarray time-series data," in *Proc. IEEE Inf. Vis*, 2003, pp. 173–180.

[19] D. Albers, C. N. Dewey, and M. Gleicher, "Sequence surveyor: Leveraging overview for scalable genomic alignment visualization," *IEEE Trans. Vis. Comput. Graph.*, vol. 17, no. 12, pp. 2392–2401, Dec. 2011.

[20] J. J. Van Wijk and E. R. Van Selow, "Cluster and calendar based visualization of time series data," in *Proc. IEEE Symp. Inform. Vis.*, 1999, p. 4.

[21] D. Keim, "Designing pixel-oriented visualization techniques: Theory and applications," *IEEE Trans. Vis. Comput. Graph*, vol. 6, no. 1, pp. 59–78, Jan. 2000.

[22] M. Davis, P. Guhathakurta, N. P. Konidaris, J. A. Newman, M. L. N. Ashby, A. D. Biggs, P. Barmby, K. Bundy, S. C. Chapman, A. L. Coil, C. J. Conselice, M. C. Cooper, D. J. Croton, P. Eisenhardt, R. Ellis, S. Faber, T. Fang, G. G. Fazio, A. Georgakakis, B. Gerke, W. M. Goss, S. Gwyn, J. Harker, A. Hopkins, J.-S. Huang, R. J. Ivison, S. A. Kassin, E. Kirby, A. Koekemoer, D. C. Koo, E. Laird, E. Le Floc'h, L. Lin, J. Lotz, P. J. Marshall, D. C. Martin, A. Metevier, L. A. Moustakas, K. Nandra, K. Noeske, C. Papovich, A. C. Phillips, R. M. Rich, G. H. Rieke, D. Rigopoulou, S. Salim, D. Schiminovich, L. Simard, I. Smail, T. A. Small, B. Weiner, C. N. A. Willmer, S. P. Willner, G. Wilson, E. Wright, and R. Yan, "The All-Wavelength Extended Groth Strip International Survey (AEGIS) Data Sets," *The Astrophys. J. Lett.*, vol. 660, pp. L1–L6, May 2007.

[23] R. H. Becker, R. L. White, and D. J. Helfand, "The VLA's FIRST Survey in," *Astron. Data Anal. Softw. Syst. III*, vol. 61, p. 165, 1994.

[24] J. C. Jacob, D. S. Katz, G. B. Berriman, J. C. Good, A. C. Laity, E. Deelman, C. Kesselman, G. Singh, M. Su, T. A. Prince, and R. Williams, "Montage; a grid portal and software toolkit for science; grade astronomical image mosaicking," *Int. J. Comp. Sci. Eng.*, vol. 4, no. 2, pp. 73–87, July 2009.

[25] E. W. Greisen and M. R. Calabretta, “Representations of world coordinates in fits,” *Astron. Astrophys.*, vol. 395, pp. 1061–1075, Dec. 2002.

[26] M. R. Calabretta and E. W. Greisen, “Representations of celestial coordinates in FITS,” *Astron. Astrophys.*, vol. 395, pp. 1077–1122, Dec. 2002.

[27] X. Fan, “Evolution of high-redshift quasars,” *New Astron. Rev.*, vol. 50, pp. 665–671, Nov. 2006.

[28] C. G. Kotanyi and R. D. Ekers, “Radio galaxies with dust lanes,” *Astron. Astrophys.*, vol. 73, Mar. 1979.

[29] S. S. Shabala, Y.-S. Ting, S. Kaviraj, C. Lintott, R. M. Crockett, J. Silk, M. Sarzi, K. Schawinski, S. P. Bamford, and E. Edmondson, “Galaxy Zoo: Dust lane early-type galaxies are tracers of recent, gas-rich minor mergers,” *ArXiv e-prints*, July 2011.

[30] C. Moellenhoff, E. Hummel, and R. Bender, “Optical and radio morphology of elliptical dust-lane galaxies-Comparison between CCD images and VLA maps,” *Astron. Astrophys.*, vol. 255, pp. 35–48, Feb. 1992.

Timothy Basil Luciani received the BSc degree in computer science from the University of Pittsburgh in 2011. He is working toward the PhD degree in the Department of Computer Science, University of Pittsburgh, and a fellow under the National Science Foundation (NSF) GRF Program. His research focuses on large-scale data visualization and realtime rendering. He is a student member of the IEEE.

Brian Cherinka received the PhD degree in physics and astronomy from the University of Pittsburgh in 2012 and is currently a post-doctoral fellow at the Dunlap Institute for Astronomy and Astrophysics in Toronto, Ontario. His research focuses on galaxy-gas interactions, as well the development of software for data analysis and astronomy visualization.

Daniel Oliphant received the BSc degree in computer science from the University of Pittsburgh in 2010. He is currently a software engineer at Google, Pittsburgh. He has a background in realtime rendering, data visualization, and big data analytics.

Sean Myers received the BSc degreein computer science from the University of Pittsburgh, in 2013 and is working toward the MSSc degree in the Department of Computer Science, University of Pittsburgh. His research interests focus on data visualization and distributed systems.

W. Michael Wood-Vasey received the PhD degree in physics from the University of California, Berkeley, in 2004. He is currently an associate professor, Department of Physics and Astronomy, University of Pittsburgh. His research focuses on dark energy.

Alexandros Labrinidis received the PhD degree in computer science from the University of Maryland, College Park in 2002. He is currently an associate professor, Department of Computer Science, University of Pittsburgh. He received the National Science Foundation (NSF) CAREER Award in 2008 and researches data management systems; he contributes to this work the Geohash description and implementation.

G. Elisabeta Marai received the PhD degree in computer science from Brown University in 2007. She is currently an assistant professor in the Department of Computer Science, University of Pittsburgh. She has received the National Science Foundation (NSF) CAREER Award in 2010 and multiple Best Paper Awards and Teaching Awards for her work in computational representations for scientific modeling and data visualization. She is a member of the IEEE.