

Viewpoints: A High-Performance High-Dimensional Exploratory Data Analysis Tool

P. R. GAZIS

SETI Institute, Mountain View, CA 94043; pgazis@sbcglobal.net

C. LEVIT

NASA Ames Research Center, Moffett Field, CA 94035; Creon.Levit@nasa.gov

AND

M. J. WAY^{1,2}

NASA Goddard Institute for Space Studies, 2880 Broadway, New York, NY 10025; Michael.J.Way@nasa.gov

Received 2010 August 12; accepted 2010 October 18; published 2010 November 12

ABSTRACT. Scientific data sets continue to increase in both size and complexity. In the past, dedicated graphics systems at supercomputing centers were required to visualize large data sets, but as the price of commodity graphics hardware has dropped and its capability has increased, it is now possible, in principle, to view large complex data sets on a single workstation. To do this in practice, an investigator will need software that is written to take advantage of the relevant graphics hardware. The Viewpoints visualization package described herein is an example of such software. Viewpoints is an interactive tool for exploratory visual analysis of large high-dimensional (multivariate) data. It leverages the capabilities of modern graphics boards (GPUs) to run on a single workstation or laptop. Viewpoints is minimalist: it attempts to do a small set of useful things very well (or at least very quickly) in comparison with similar packages today. Its basic feature set includes linked scatter plots with brushing, dynamic histograms, normalization, and outlier detection/removal. Viewpoints was originally designed for astrophysicists, but it has since been used in a variety of fields that range from astronomy, quantum chemistry, fluid dynamics, machine learning, bioinformatics, and finance to information technology server log mining. In this article, we describe the Viewpoints package and show examples of its usage.

1. INTRODUCTION

Analysis and visualization of extremely large and complex data sets has become one of the more significant challenges facing the scientific community today. Current instruments and simulations produce enormous volumes of data. These data sets include hyperspectral images from spacecraft, multivariate data of high dimensionality from sky surveys, time-varying three-dimensional flows from supercomputer simulations, and complex interrelated time series from vehicle telemetry, DNA microarrays or physiological monitoring, to name a few.

One representative example from astronomy is the Sloan Digital Sky Survey (SDSS; York et al. 2000). The spectroscopic catalog from the SDSS Data Release 7 (DR7; Abazajian et al. 2009) contains information for over 1.5 million objects. For each object in the joint spectroscopic and photometric catalog

there are over 100 physical measurements. Properties in these catalogs include position on the sky, spectroscopic redshift, photometric magnitude in five bands, energy/width of up to 50 spectral lines, galaxy angular size, bulge-to-disk ratio, etc. In addition, many other physical quantities of interest can be derived from these data, such as absolute magnitudes, photometric redshifts, galaxy type, average distance to nearest neighbors, and other properties of the galaxy's local environment. This is the kind of catalog in which Viewpoints excels on a relatively modern desktop computer.

On the other hand, the DR7 photometric catalog has measured over 350 million unique objects. This is an enormous amount of data by traditional standards, and these kinds of data volumes will continue to grow in size over the coming decade. The actual data volume in the SDSS DR7, including pre- and postprocessing products, is more than 50 TB, while the searchable database for the photometric and redshift catalogs is over 3.5 TB.³ We do not claim that Viewpoints is able to explore a

¹ NASA Ames Research Center, Space Science Division, Moffett Field, CA 94035; Corresponding Author.

² Department of Astronomy and Space Physics, Faculty of Science and Technology, Uppsala, Sweden.

³ See <http://www.sdss.org/dr7>.

data set of this size all at once, as it is beyond the memory capabilities of a normal desktop computer. Still, Viewpoints can quickly explore subsets of a large complex data set like this. When Viewpoints does reach the limits of desktop hardware, one can reach for more capable, but expensive and rare, custom-built hardware systems such as the hyperwall (Sandstrom et al. 2003). The hyperwall can then be utilized efficiently using the knowledge already obtained through exploration via Viewpoints.

Although advances in hardware speed and storage technology have kept up with (and, indeed, have driven) increases in database size, the same is not true of the tools we use to explore and understand these data (e.g., SubbaRao et al. 2008). Modern data sets routinely outstrip the capabilities of contemporary visualization and analysis software, and this problem can only get worse as data volumes from future astronomy-related programs (e.g., Large Synoptic Survey Telescope [LSST],⁴ Supernova Acceleration Probe [SNAP],⁵ Panoramic Survey Telescope and Rapid Response System [Pan-STARRS],⁶ and Dark Energy Survey [DES]⁷) and ever-larger supercomputer simulations increase by orders of magnitude. See Borne (2009) for a review. Data sets of this size and dimensionality defy interactive analysis with standard tools such as MATLAB,⁸ Octave,⁹ Mathematica,¹⁰ R,¹¹ S++,¹² XGobi,¹³ Mirage,¹⁴ gnuplot,¹⁵ etc. To address this problem, we have developed Viewpoints.

Viewpoints was inspired to a large extent by the hyperwall (Sandstrom et al. 2003). This facility, developed by the NASA Advanced Supercomputing division at NASA Ames Research Center, is a high-performance visualization cluster that was specifically designed to address the problem of exploring, visualizing, and analyzing large, complex, multidimensional data. See Wong & Bergeron (1997) for a review on the history of multidimensional multivariate visualization. In its original form, the hyperwall was a 50 node Linux cluster with a 7×7 array of tiled displays that could be used to display data in a variety of formats that might range from simple plots to elaborate 3D rendering. It has since been superseded by even larger facilities at NASA Ames Research Center and several other institutions throughout the world.

Hyperwalls have been applied to a wide range of problems in astrophysics, Earth science, aerodynamics, and life sciences

(MacDougall & Henze 2003; Brown et al. 2004; Murman et al. 2004). Unfortunately, tools such as the hyperwall require custom hardware and a team of onsite specialists at a dedicated facility, which places them beyond the reach of most investigators and research programs.

Viewpoints, shown in Figure 1, is intended to provide a small but powerful subset of the hyperwall's functionality on any working scientist's desktop or laptop computer.

Viewpoints provides a high-performance interactive multidimensional visual exploratory data analysis environment that allows users to quickly explore and visually analyze the most common forms of large multidimensional scientific data. It can also be run as a plug-in or external module for scientific programming environments such as MATLAB.

In contrast to programs like Mirage (Ho 2007) or VisIVO (Becciani et al. 2010), which have a large number of supporting features, Viewpoints does a few things very well. Its basic feature set includes linked scatter plots with brushing (e.g., Becker & Cleveland 1987; Davidson & Sardy 2000; Chen 2003; Stump et al. 2004; Co et al. 2005; Jordan et al. 2008), dynamic histograms, normalization, and outlier detection/removal. It also includes a limited capability for Boolean operations with brushes analogous to those of Hurter et al. (2009).

2. OVERVIEW AND CAPABILITIES

Viewpoints is a fast interactive tool for the analysis and visualization of high-dimensional data primarily via linked scatter plots with brushing. It is intended for general-purpose visual data analysis, interactive exploratory data mining, and machine-assisted pattern discovery. The targeted application is the exploration of large, complex, multivariate data sets, with dimensionalities up to 100 or more and sample sizes up to 10^5 – 10^7 . Viewpoints takes advantage of hardware-accelerated graphics (graphics processing units [GPUs]) and other capabilities of modern desktop workstations and laptops (e.g., multi-threading and memory-mapped input/output). As noted previously, it is highly interactive, and almost all of its parameters can be modified in real time to provide immediate visual feedback.

Viewpoints can be invoked either from a desktop icon or from the command line with a set of optional parameters. Once it is running, it is controlled by a conventional graphical user interface (GUI) and/or configuration files saved from previous sessions. Viewpoints reads multidimensional data in the form of ASCII files, binary files, or simple Flexible Image Transport System (FITS) tables and displays these data as an arbitrary number of simultaneously linked 2D and 3D scatter plots (e.g., Comparato et al. 2007) with independently controllable viewing transformations, normalization schemes, overplotting schemes (Zhang et al. 2003), axis labels, and other formatting. These data can then be visualized and explored interactively using a set of graphical, statistical, and semiautomated pattern-recognition techniques. The user can select, flag, and/or

⁴ See <http://www.lsst.org>.

⁵ See <http://snap.lbl.gov>.

⁶ See <http://pan-starrs.ifa.hawaii.edu/public>.

⁷ See <http://www.darkenergysurvey.org>.

⁸ See <http://www.mathworks.com>.

⁹ See <http://www.octave.org>.

¹⁰ See <http://www.wolfram.com>.

¹¹ See <http://www.r-project.org>.

¹² See <http://www.splusplus.com>.

¹³ See <http://www.research.att.com/areas/stat/xgobi>.

¹⁴ See <http://www.bell-labs.com/project/mirage>.

¹⁵ See <http://www.gnuplot.info>.

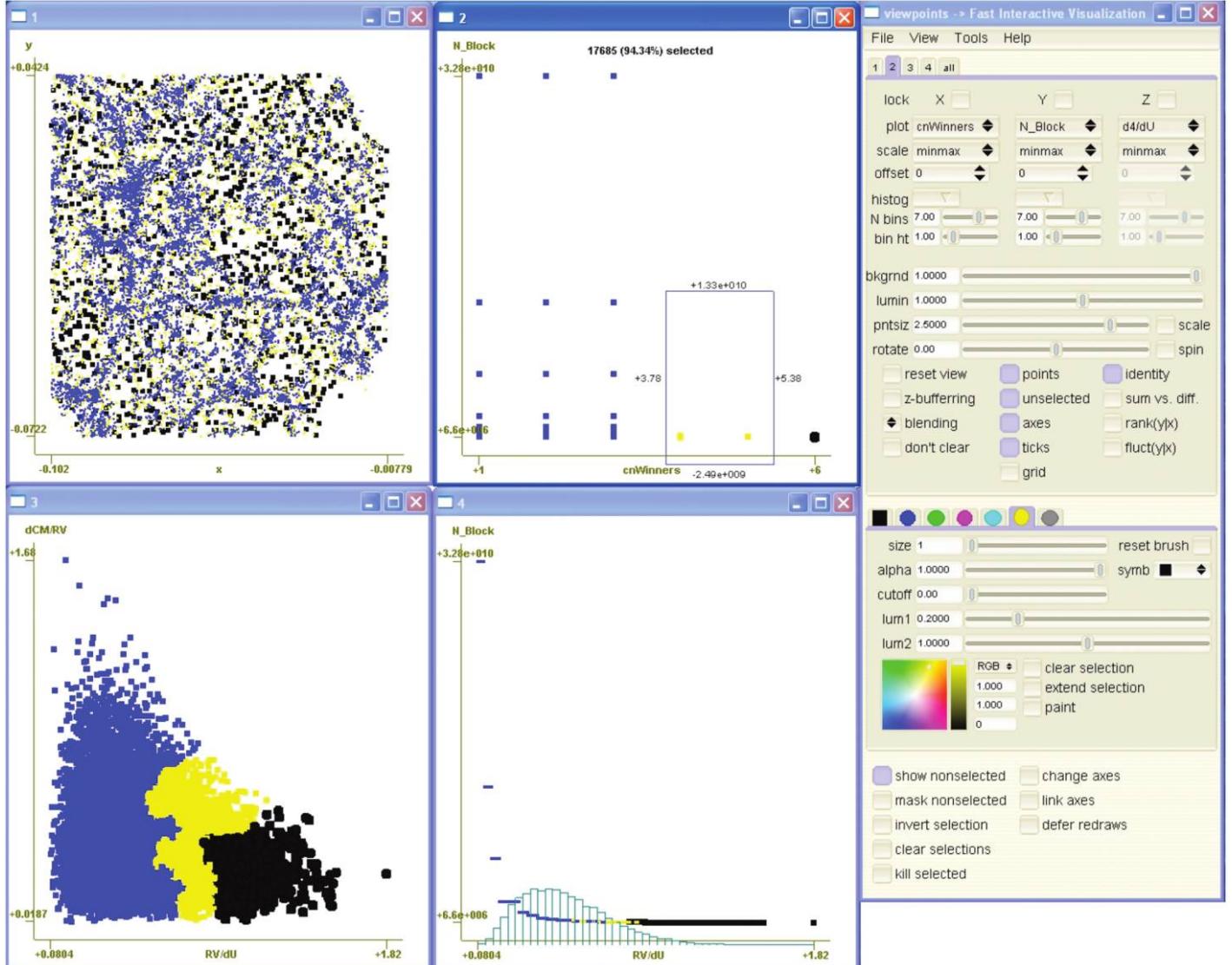


FIG. 1.—Viewpoints displaying a 2×2 array of scatter plots of SDSS data. *Upper left:* xy positions of $\sim 20,000$ galaxies from the SDSS. Colors indicate different classes identified by a clustering algorithm. *Upper right:* Density classes identified by a Bayesian partitioning scheme vs. class identified by a self-organizing map. *Lower left:* Local density vs. local gradient. *Lower right:* Local density vs. density classes identified by a Bayesian partitioning scheme. *Far right:* The Viewpoints control panel.

delete subsets of the data using brushes, and the modified data set can be written back to disk in ASCII, binary, or FITS format.

The program's capabilities were carefully selected using a minimalist philosophy. We have chosen to provide a small core of essential features that are universally useful and keep the program size small, execution speeds high, and the learning curves shallow. Figure 1 shows an example of Viewpoints displaying a small 2×2 array of linked scatter plots. This figure is described in detail in § 5. In practice, one can simultaneously display an arbitrary number of linked scatter plots, and we routinely use up to 30 windows to examine data sets of high dimensionality. Viewpoints has a conventional GUI, shown in the right panel of Figure 1, with an assortment of menus, buttons, and sliders

that provide access to its features. The most important of these capabilities can also be invoked by mouse gestures and/or keyboard shortcuts. It has a full range of basic help facilities, such as usage messages, help files, and tool tips. It also produces a diagnostic output stream when invoked via the command line.

Viewpoints has a tightly integrated set of features. A partial list of these features is provided next. Some were in the original release, others were added in response to users requests, and a few were contributed to the open source repository¹⁶ by a small community of developers.

¹⁶ See <http://www.assembla.com/wiki/show/viewpoints>.

1. Multiple linked scatter plots with dynamic brushing rendered via OpenGL with update rates of greater than 10^7 points s^{-1} for most modern graphics cards.
2. The ability to brush separate selections using separate colors and/or symbols.
3. The ability to map densely overplotted regions smoothly to different hues and/or intensities.
4. The ability to save the program's complete state to an XML-formatted configuration file at any time.
5. The ability to initialize and/or restore the program's state from a configuration file.
6. The ability to append and merge additional data and create or destroy additional plot windows on the fly.
7. Per-axis one-dimensional equiwidth marginal histograms with dynamically adjustable bin widths and dynamic update of the selected fraction in each bin. Histograms are individually selectable and adjustable for both the x and y axes of each plot separately.
8. Per-axis automatic normalizations based data limits, quantiles, moments, rank, or simple Gaussianization.
9. Elementary outlier detection and removal.
10. The ability to change most parameters globally across all plots simultaneously, or locally on a single plot.
11. The ability to lock (and unlock) individual plots so that they are immune from (susceptible to) global operations.
12. Rudimentary semi-automated browsing through high-dimensional data by permuting all (unlocked) axes on all plots.
13. Real-time display of quantitative data for individual axes and for the current selection.
14. Basic online help.
15. A comprehensive set of error tests along with graceful error recovery if a user attempts to read corrupt and ill-formatted data or configuration files.
16. The ability to replace missing values with user-specified default values.
17. Basic ability to plot time series, analogous to that of Akiba & Ma (2007).
18. Viewpoints also achieves smooth interactive performance on large data as long as:
 - a) The (binary) array corresponding to the data set fits into real system (CPU) memory.
 - b) The coordinates of all vertices being displayed fit into available graphics (GPU) memory. For example, on a 512 Mbyte graphics card, Viewpoints can simultaneously update and display $(4 \text{ plots}) \times (8 \times 10^6 \text{ points plot}^{-1}) \times (3 \text{ vertices point}^{-1}) \times (4 \text{ bytes vertex}^{-1}) = \sim 400 \text{ Mbytes}$.

3. SOFTWARE DESIGN

Viewpoints uses hardware-accelerated OpenGL 1.5 on all platforms and performs rendering via OpenGL¹⁷ vertex buffer

¹⁷ See <http://www.opengl.org>.

objects to provide extremely high graphics performance. The package has cross-platform availability and can be easily ported to other architectures. Table 1 has a list of operating systems and architectures known to work to date. Viewpoints has a preliminary interface to MATLAB and Octave. It has the potential to work with scripting languages such as Python¹⁸ or Perl.¹⁹ Most importantly, Viewpoints is specifically engineered for use with large data sets. As noted previously, the targeted application is the interactive analysis of large, complex, multivariate data sets, with dimensionalities that may surpass 100 and sample sizes that may exceed 10^5 – 10^7 .

In the interests of portability, extensibility, and efficiency, Viewpoints is written in C++. This allows us to program at a high level of abstraction and still achieve excellent runtime performance on all target platforms. Most importantly, C++ allows us to leverage vast amounts of code written by other people. Identical code compiles and runs on Linux, Microsoft Windows, and Apple OS X.

Viewpoints is a relatively small program. The compiled Linux binary is only 5.5 Mbytes.²⁰ Since it makes extensive use of libraries, including some that leverage C++ template metaprogramming (e.g., Blitz++²¹ and Boost²²), the Viewpoints source tree itself is also rather small (approximately 40,000 lines). The libraries it depends on are listed in Table 2.

4. GPU USAGE AND OPENGL IMPLEMENTATION

Graphics boards that were available at the time development began, such as NVIDIA Geforce 7950, can transform and display over a billion floating-point vertices per second and can drive two or more 4 Mpixel displays simultaneously. This capability continues to increase. In its default mode, Viewpoints uses GPUs explicitly to take advantage of this tremendous processing and rendering power. This allows the package to provide smooth high-frame-rate interactivity while manipulating multiple linked views, each containing millions of data points. If for some reason a GPU is not available, Viewpoints can also be instructed to use a system's ordinary CPU, at some cost in performance.

The current release uses the following advanced OpenGL features:

1. GPU-resident vertex buffer objects (VBOs) for point coordinates. This conserves main memory (by offloading relevant data to GPU memory), increases graphics performance (by caching frequently rendered vertex data locally on the GPU), and frees up essentially all of the CPU bandwidth and bus bandwidth for other uses.

¹⁸ See <http://www.python.org>.

¹⁹ See <http://www.perl.org>.

²⁰ The compiled file contains a large number of statically linked libraries.

²¹ See <http://www.oonumerics.org/blitz>.

²² See <http://www.boost.org>.

TABLE 1
PLATFORMS KNOWN TO RUN VIEWPOINTS

Operating System	OS revision	Architecture
Apple	Mac OS 10.3–10.6	Intel
Apple	Mac OS 10.3–10.6	PPC
Linux	Most distributions with kernels 2.4.x–2.6.x	Intel
Windows	2000	Intel
Windows	XP	Intel
Windows	Vista	Intel
Windows	7	Intel

2. GPU-resident VBOs for point indices. A single copy of the vertex index arrays (which encode the selection status) is shared by all plots.

3. Squared alpha blending with separate blending control per plot. For data with wide variations in projected density this provides fast, adjustable mapping of overplot density to varying hue and/or intensity.

4. Point sprites for displaying symbols and/or smooth kernels instead of just simple points (splatting).

5. Anti-aliased points and/or point sprites of variable color, brightness, and size per brush and/or per plot.

6. Floating-point color and alpha values, allowing both more sensitivity and more dynamic range for overplotting.

These design features provide dramatic graphics performance. When one of our early beta releases was run on a 1997 MacBook Pro laptop, we measured throughput of over 70×10^6 vertices s $^{-1}$. This consisted of six linked plots, each with 2 million data points simultaneously updating at 5 frames s $^{-1}$.

Viewpoints has proved quite useful on small-sized (e.g., netbooks) and medium-sized computer platforms with as little as 1 Gbyte of RAM. We do most of our demos on a laptop displaying up to 18–27 variables in 6–9 windows on a single monitor. To take full advantage of this package we recommend using a high-end workstation with high-end graphics card, two high-resolution displays, and several gigabytes of RAM. Experience suggests that this is a good impedance match to the visual perception of a scientist or a team of up to three collaborators.

The best way to think about the limitations of Viewpoints with respect to common desktop computer hardware is that

the data should be able to literally fit in RAM memory. Specifically, Viewpoints can achieve smooth interactive performance on data as long as:

1.The (binary) array corresponding to the data set fits into real system (CPU) memory

2.The coordinates of all vertices being displayed fit into available graphics (GPU) memory.

For those looking to the future, please note that the relationship between the amount of memory on a graphics card and the allowable size of the data set appears to be linear. So the latest graphics cards with 1.5 Gbyte GPU memory should display four plots with $\sim 25 \times 10^6$ points per plot, with an update rate of a few frames per second. The update rate is harder to estimate, since “vertices per second” is no longer a common figure of merit in GPU specs. However, this appears to scale roughly along with clock speed and GPU memory.

When visible plots exhaust GPU memory, one can run Viewpoints with the “–no_vbo” flag, and check the “defer redraws” box in the lower right-hand control panel. We have run 100 million point data sets on a laptop this way and it is quite usable.

5. EXAMPLES

Figure 1 shows a simple example of Viewpoints usage with a small 2×2 array of linked scatter plots of data from the SDSS (York et al. 2000), several methods of calculating local density, and a cluster detection scheme. The upper left panel shows xy coordinates of $\sim 20,000$ galaxies from the SDSS. When used this way, a scatter plot becomes a map of a region of space that is approximately 30 Mpc across. The blue, yellow, and black colors indicate cluster galaxies, halo galaxies, and field galaxies identified by a partitioning scheme based on self-organizing maps (SOMs). It can be seen that cluster galaxies (blue) are concentrated into filaments, while the field galaxies (black) are distributed more or less uniformly. The upper right panel shows density classes identified by a Bayesian partitioning scheme versus classes identified by the SOM-based algorithm. This panel is where the selection capability of Viewpoints was used. The rectangle in the lower right side of the panel is the selection box that was used to flag the halo galaxies (yellow points). The lower left panel shows local density versus local gradient to

TABLE 2
C++ LIBRARIES USED IN VIEWPOINTS

Library	Description
libc++	Standard runtime library for basic math, input/output, and memory management
STL	Standard template library for data structures and algorithms
OpenGL	Standard cross-platform library for hardware-accelerated graphics
FLTK	Lightweight platform-independent open source GUI library
Blitz++	Library of extensions that add FORTRAN 90 style array expressions (and FORTRAN 90 style numerical performance) to C++
GSL	GNU scientific library, includes functionality such as statistics and linear algebra
Boost serialization	Library of extensions for saving and restoring the state of c++ objects

display the distribution of the three galaxy classes in parameter space. The lower right panel shows local density versus density class identified by a Bayesian partitioning scheme. This panel shows a simple example of the histogram capability of Viewpoints. Figure 1 provides a real-world example of the abilities of Viewpoints to quickly deduce scientifically useful features for research purposes. The example provided here is directly related to a recent publication by two of the authors of this article (Way et al. 2010). All of the figures comparing the three different clustering technologies described in Way et al. (2010) were originally devised using Viewpoints. Publication-quality plots were later produced using MATLAB.

Figure 2 shows a more elaborate usage involving a 2×2 array of linked scatter plots of astrometric data from the *Tycho*

mission (Perryman et al. 1997; Høg et al. 2000). For clarity, the control panel has been eliminated from this figure.

The upper left panel shows a plot of right ascension versus declination for 1 million stars. This panel illustrates the effect of overplotting with alpha blending. The galactic plane, with its greater concentration of stars, is plainly visible as a curving white band that meanders through the image. The upper right panel shows a plot of the number of transits versus parallax error. The selection box (outlined in blue) has been used to select points for which more than 250 transits were available. The legend at the top of the panel reports that these involved 32,503 points or 3.14% of the total. This selection is repeated in the other panels. In the upper left panel described previously, it appears as two parallel and roughly sinusoidal bands that mark the

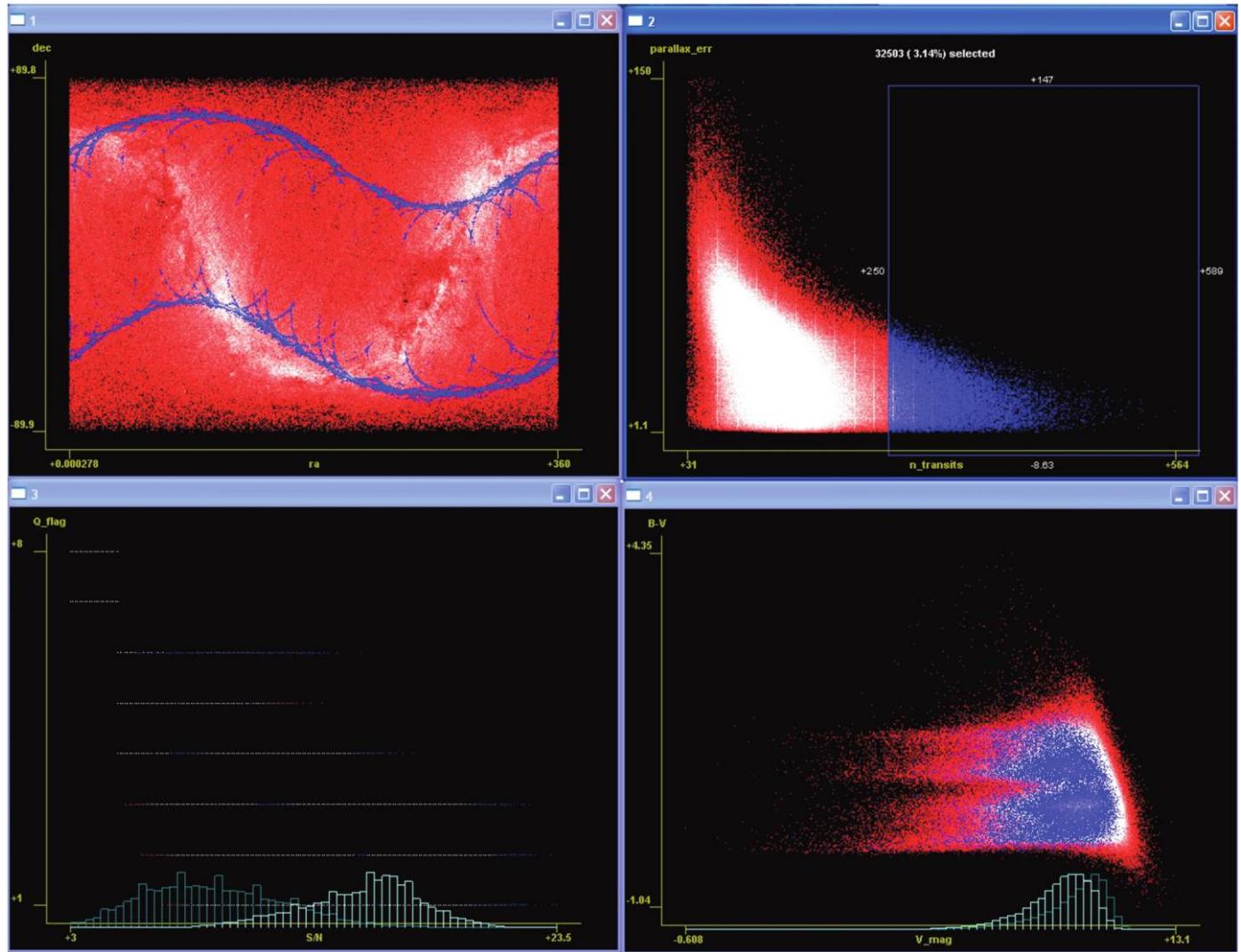


FIG. 2.—Viewpoints displaying a 2×2 array of linked scatter plots of *Tycho* data for 1 million stars. *Upper left:* Right ascension vs. declination. *Upper right:* Number of transits vs. parallax error, with a selection box outlined in white. *Lower left:* S/N vs. a quality flag. *Lower right:* *V*-band magnitude plotted vs. the $B - V$ magnitude difference.

part of the spacecraft's orbit from which it was possible to obtain this number of transits. The lower left panel shows a plot of signal-to-noise ratio (S/N) versus a quality flag. This is included to illustrate use of a marginal (conventional) and a conditional histogram. The latter peaks at a significantly higher S/N than the former, as would be expected, since measurements that involve more transits should have a higher S/N . In the lower right panel the V -band magnitude has been plotted against the $B - V$ magnitude difference to produce a typical color-magnitude diagram.

Figure 3 shows an example of Viewpoints used to visualize the set of solar wind data that is provided with the Viewpoints package for demonstration purposes.

In the top three panels, solar wind speed, density, and temperature have been plotted versus day since 1950 to produce a conventional time series. These show a typical structure in the solar

wind, in which high-speed/low-density plasma (visible on the left of each panel) overtakes low-speed/high-density material (visible on the right) to produce an interaction region (visible in the middle) that is characterized by intermediate speeds, high density, and high temperature. These interaction regions are often subdivided into a leading portion and trailing portion with different ranges of densities and temperatures separated by a well-defined boundary. These are plainly visible in this time-series plot. The selection capability of Viewpoints has been used to select these regions as blue and yellow. The bottom three panels show scatter plots of velocity versus density, velocity versus temperature, and density versus temperature. It can be seen that the leading portion of this interaction region (blue) was associated with a large range of densities and moderate range of temperatures, while the trailing portion was associated with a moderate range of densities and large range of temperatures.

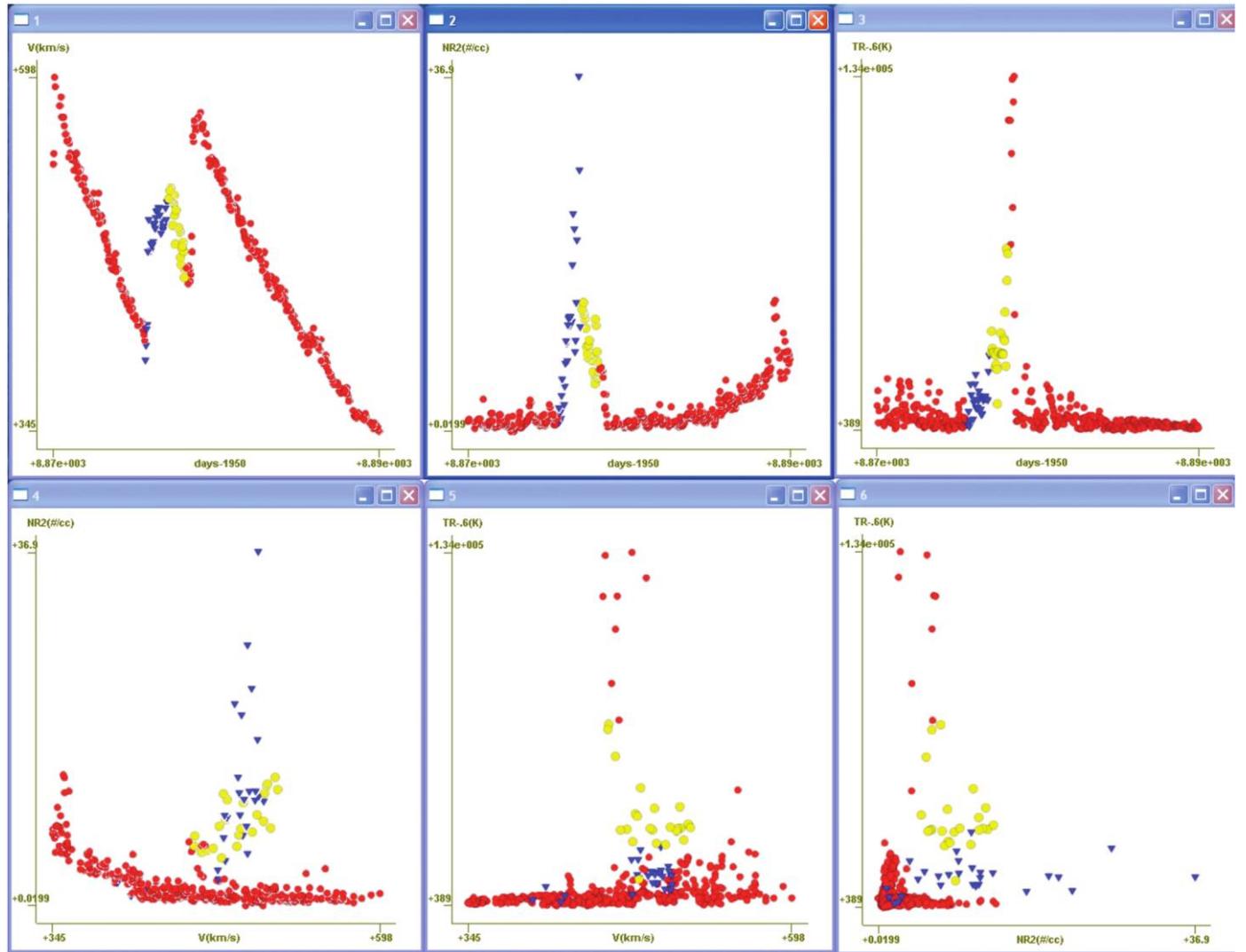


FIG. 3.—Viewpoints displaying a 2×3 array of linked scatter plots of solar wind data. *Upper left:* Solar wind speed vs. time. *Upper middle:* Solar wind density vs. time. *Upper right:* Solar wind temperature vs. time. *Lower left:* Density vs. speed. *Lower middle:* Temperature vs. speed. *Lower right:* Temperature vs. density.

As mentioned in the Abstract, Viewpoints has been used extensively in a variety of fields to advance specific scientific goals. Further examples include:

Geophysics.— Relationships between earthquake occurrence frequency, location, season, and time of day.

Information Technology.— Relationships between query response time, query source location, time of day, type of query, etc. These were causing slow response and high system loads on commercial servers which Viewpoints helped resolve quickly.

Finance.— Currency arbitrage relationships and portfolio clustering in statistical arbitrage.

6. OPENING UP TO THE OPEN SOURCE COMMUNITY

Two significant concerns with any software tools are long-term support and response to the user community. Viewpoints has been released as an open source and is on a collaborative-oriented public wiki site called Assembla.²³ We now have a significant and growing user community that is successfully and enthusiastically applying Viewpoints to a huge variety of problem spread across multiple disciplines.

7. CONCLUSIONS AND THE FUTURE OF VIEWPOINTS

Viewpoints was adapted from its alpha prototype to general use with seed funding from the NASA Applied Information

Systems Research Program.²⁴ It is now available on all major computing platforms (see Table 1) and is in use by scientists and engineers across all NASA mission directorates, as well as by some early adopters in industry and academia. Its operation, interface, source code, and impressive performance are the same on all platforms. Viewpoints is now directly callable from MATLAB. There is a Viewpoints wiki and Web site²⁵ that allows the open source community to take part in development and affords users access to the latest information regarding bug fixes and program enhancements. In the near future we expect more development to come from the open source community and plan to modify the base code to take advantages of new advances in video chip technology as it continues to evolve.

Many thanks go to Chris Henze and Jeff Scargle for their input and ideas. The authors would also like to thank Joe Bredekamp and his NASA Applied Information Systems Research Program for pilot funding of Viewpoints. M. J. W. would like to thank the Astronomy Department of Uppsala University for their kind hospitality. This research has made use of NASA's Astrophysics Data System. Viewpoints has been released under the NASA Open Source Agreement and is considered an open source license. The source code and precompiled binaries can be freely downloaded from the Viewpoints Web site. If you use Viewpoints in your research, please consider referencing this work.

²³ See <http://www.assembla.com>.

²⁴ See <http://aisrp.nasa.gov>.

²⁵ See <http://www.assembla.com/wiki/show/viewpoints>.

REFERENCES

- Abazajian, K., et al. 2009, ApJS, 182, 543
 Akiba, H., & Ma, K. 2007, Eurographics/IEEE-VGTC Symp. Vis. (Washington: IEEE)
 Becciani, U., et al. 2010, PASP, 122, 119
 Becker, R. A., & Cleveland, W. S. 1987, Technometrics, 29, 127
 Borne, K. 2009, Next Generation in Data Mining (Boca Raton: Taylor & Francis), 91
 Brown, J., Tobak, M., Prabhu, D., & Sandstrom, T. 2004, 37th AIAA Thermophysics Conference (2004-2285) (Reston: AIAA)
 Chen, H. 2003, IEEE Symp. Inf. Vis. (Washington: IEEE), 181
 Co, C. S., et al. 2005, Eurographics/IEEE-VGTC Symp. Vis. (Washington: IEEE), 279
 Comparato, M., et al. 2007, PASP, 119, 898
 Davidson, A. C., & Sardy, S. 2000, J. Comput. Graph. Stat., 9, 750
 Ho, T. K. 2007, in ASP Conf. Ser, 371, Statistical Challenges in Modern Astronomy IV, ed. G. J. Babu, & E. D. Feigelson (San Francisco: ASP), 391
 Høg, E., et al. 2000, A&A, 355, L27
 Hurter, C., Tissoires, B., & Conversy, S. 2009, IEEE Trans. Vis. Comput. Graph. (Washington: IEEE), 151017
 Jordan, D. D., et al. 2008, AAS/AIAA Space Flight Mechanics Meeting (AAS 08-245; Springfield: AAS), 499
 MacDougall, P., & Henze, C. 2003, 226th ACS National Meeting, (New York: Wiley), 499
 Murman, S. M., Aftosimis, M. J., & Nemec, M. 2004, 22nd AIAA Applied Aerodynamics Conf., (2004-5076; Reston: AIAA)
 Perryman, M. A. C., et al., 1997, The Hipparcos and Tycho Catalogues, (SP-1200; Noordwijk: ESA), 117
 Sandstrom, T. A., Henze, C., & Levitt, C. 2003, Coordinated & Multiple Views in Exploratory Visualization, (Piscataway: IEEE), 124
 Stump, G. M., et al. 2004, Aerospace Conference (Piscataway: IEEE), 6, 3885
 SubbaRao, M. U., Aragón-Calvo, M. A., Chen, H. W., Quashnock, J. M., Szalay, A. S., & York, D. G. 2008, NJP, 10, 125015
 Way, M. J., Gazis, P. R., & Scargle, J. S. 2010, ApJ, submitted (arXiv:1009.0387v1)
 Wong, P. C., & Bergeron, R. D. 1997, Scientific Visualization Overviews Methodologies and Techniques (Washington: IEEE), 3
 York, D. G., et al. 2000, AJ, 120, 1579
 Zhang, J., et al. 2003, IEEE International Conference on Data Mining (Washington: IEEE)