VAST Challenge 2017: Mystery at the Wildlife Preserve

Mark A. Whiting[1], Kris Cook[1], R. Jordan Crouser[2], John Fallon[3], Georges Grinstein[3], Jereme Haack[1], Cindy Henderson[1], Kristen Liggett[4], Diane Staheli[5], Jana Strasburg[1], Jerry Tagestad[1], Carrie Varley[1]

Abstract

The VAST Challenge 2017 offered three mini-challenges and a grand challenge dealing with environmental problems potentially caused by human patterns of life and potentially harmful chemically laden effluent plumes being emitted from factory smokestacks. The data provided included traffic patterns, sensor data though a Preserve, information about the Preserve, multispectral imagery and a map to help an ornithology graduate student particularly concerned with the population decrease of the Rose-Crested Blue Pipit determine who and what might be responsible. Mini-Challenge 1 focused on analysis of vehicles passing through the Preserve over time. Mini-Challenge 2 looked at data collected by air sampling monitors surrounding nearby factories, along with meteorological readings, to understand potential impacts they may be having on the Pipit. Mini-Challenge 3 required investigation into several months of multi-spectral imagery over the area to understand the Preserve's general health. The Grand Challenge asked participants to synthesize across all three mini-challenges to create hypotheses of what is happening and what sensible next steps could be. This year's challenge received 58 submissions and recorded over 1100 unique downloads from 20 countries prior to the submission deadline.

Keywords: Visual analytics, human information interaction, patterns of life, short events analytics, multi-spectral imagery, evaluation, contest

1 Introduction

The Visual Analytics Science and Technology (VAST) Challenge is an annual contest with the goal of advancing the field of visual analytics through competition. The VAST Challenge is designed to help researchers understand how their software would be used in a novel analytic task and determine if their data transformations, visualizations, and interactions would be beneficial for particular analytic tasks. VAST Challenge problems provide researchers with realistic tasks and data sets for evaluating their software, as well as an opportunity to advance the field by solving more complex problems.

The VAST Challenge 2017 focused on a familiar theme, but with mini-challenge problems that have not been presented in the Challenges before. The theme centered on problems within an environmental setting. The Challenge presented variations on this in 2007 and 2014. This year Mini-Challenge 1 looked at environmental problems potentially caused by human patterns of life, Mini-Challenge 2 at effluent plumes being emitted from factory smokestacks containing harmful chemicals, and Mini-Challenge 3 at multispectral analysis of imagery. Contestants could also integrate and synthesize their findings from the mini-challenges in a Grand Challenge to hypothesize and provide evidential support about who and what was the primary contributor to the environmental problems, and the classic why, where and how for the problem.

2 Introducing Multispectral Imaging

In Mini-Challenge 3, participants were provided with a series of multispectral images of the area of concern for environmental problems. Since this represented a new area for the VAST Challenges and perhaps for some participants, we provided a primer to help with analytic approaches. In the mini-challenge, data collected came from a multispectral sensor that measured in six different bands (B1-B6). The sensor was a sensitive scientific instrument and the resulting images were optimized for measurement (that is, were not simply pictures). The B1-B6 bands can be thought of as integrated values associated with six different filters, each of which are useful in assessing land cover. B1, B2, and B3 represent portions of the visible spectrum. B4, B5, and B6 represent longer wavelengths that are beyond human perception. Table 1 shows the colors, wavelength regions and utility of the different bands.

Band	Color	Wavelength (nm)	Useful for Mapping
B1	Blue	450-520	Penetrates water, shows thin clouds and general visible brightness
B2	Green	520-600	Shows different types of plants and general visible brightness
B3	Red	630-690	Vegetation color and certain mineral deposits.
B4	Near Infrared (NIR)	770-900	Partially absorbed by water, sensitive to vegetation structure and chlorophyll
B5	Short-wave Infrared (SWIR) 1	1550-1750	Completely absorbed by liquid water. Sensitive to moisture content of soil and vegetation; penetrates thin clouds
B6	Short-wave Infrared (SWIR) 2	2090-2350	Insensitive to vegetation color or vigor, shows differences in soil mineral content

Table 1 Multispectral bands and their utility

When looking at multi-spectral images, sometimes it is useful to map the values of the different bands to Red-Green-Blue (RGB) image channels. When B3, B2 and B1 are mapped to RGB this generates a "true-color" image. Other band combinations can also be mapped to the RGB image channels creating what are known as "false color" images.

Frequently used false-color images include:

B4, B3, B2 mapped to RGB which can be useful in seeing changes in plant health, B5, B4, B2 mapped to RGB which can be useful to show floods or newly burned land, and B1, B5, B6 mapped to RGB which can differentiate between snow, ice and clouds

In other cases it may be helpful to use ratios of different bands to assess vegetation health or detect anomalous signals. A frequently-used ratio is the Normalized Difference Vegetation Index (NDVI):

1. Pacific Northwest National Laboratory, {fn.ln}@pnnl.gov
2. Smith College, jcrouser@smith.edu
3. University of Mass Amherst, {jtfallon, ggrinstein}@umass.edu
4. Air Force Research Laboratory, kristen.liggett@us.af.mil
5. MIT Lincoln Laboratory, diane.staheli@ll.mit.edu

2017 IEEE Conference on Visual Analytics Science and Technology (VAST)
October 1–6, Phoenix, Arizona, USA
978-1-5386-3163-8/17/$31.00 ©2017 IEEE

NDVI = (B4 - B3) / (B4 + B3). When this index is plotted across the image frame, it is easy to distinguish between healthy vegetation and bare ground cover.

Additional potentially useful hints were presented in the primer. While image viewers are fairly simple to construct, a key challenge for Mini-Challenge 3 was to integrate visual analytics into the process of analyzing this kind of data.

3 Scenario and Scope

The VAST Challenge 2017 presented three mini-challenges and a grand challenge for participants to apply visual analytics to help a sad state of affairs in a wildlife preserve:

Mistford is a mid-size city located to the southwest of the Boonsong Lekagul Nature Preserve. The city has a small industrial area with four light-manufacturing endeavors. Mitch Vogel is a post-doc student studying ornithology at Mistford College and has been discovering signs that the number of nesting pairs of the Rose-Crested Blue Pipit, a popular local bird due to its attractive plumage and pleasant songs, is decreasing! The decrease is sufficiently significant that the Pangera Ornithology Conservation Society is sponsoring Mitch to undertake additional studies to identify the possible reasons. Mitch is gaining access to several datasets that may help him in his work, and he has asked you (and your colleagues) as experts in visual analytics to help him analyze these datasets.

In the scenario, in an effort to maintain profit margins, one of the manufacturing facilities began illicitly modifying their night-shift process a few times a week. This modified process improved their manufacturing efficiency, but produced effluents above the regulated limit as well as a significant amount of toxic sludge that could not be disposed of in the facility's usual waste stream without risking detection by local regulators.

As a result, the company made regular trips to a rarely-visited part of the Preserve and dumped the waste. This waste had a negative impact on the local flora and fauna. Specifically the local bird population plummeted and the vegetation in the vicinity of the dumping became stressed.

Mini-Challenge 1 focused on traffic pattern of life behaviors to see if the cars and trucks passing through the Preserve could be impacting the Pipits. Mini-Challenge 2 examined the smokestack effluents from the factories to assess their safety. Mini-Challenge 3 required exploration of multispectral images over the entire Preserve and over several years, to see if long term damage could be observed. The Grand Challenge sought integration over all of the mini-challenges to develop hypotheses and supporting evidence of the root cause of the Pipit's dilemma.

3.1 Mini-Challenge 1: Vehicle Patterns of Life

The Preserve is both a refuge for regional flora and fauna, but also a pleasant place for passing travelers and tourists. Increasing populations in surrounding areas have resulted in increased traffic of many types journeying through the area.

The Boonsong Lekagul Nature Preserve is used by local residents and tourists for day-trips, overnight camping or sometimes just passing through to access main thoroughfares on the opposite sides of the Preserve. The entrance booths of the Preserve are monitored in order to generate revenue as well as monitor usage. Vehicles entering and exiting the Preserve must pay a fee based on their number of axles (personal auto, recreational trailer, semi-trailer, etc.). This generates a data stream with entry/exit timestamps and vehicle type. There are also other locations in the part that register traffic passing through. While hiking through the various parts of the Preserve, Mitch has noticed some odd behaviors of vehicles that he doesn't think are consistent with the kinds of park visitors he would expect. If there were some way that Mitch could analyze the behaviors of vehicles through the park over time, this may assist him in his investigations.

Contestants were provided with a description of how traffic through the Preserve occurs and how traffic was measured through the sensors. They were given background information about the Preserve and bitmapped files describing the gridded map against which the data was provided. Finally, the data was given in a .csv file. The .csv data contained a timestamp of when the vehicle passed a sensor location, a car-id, a car type (as described in the background information), and a sensor identification.

Mini-Challenge 1 questions were:

1. "Patterns of Life" analyses depend on recognizing repeating patterns of activities by individuals or groups. Describe up to six daily patterns of life by vehicles traveling through and within the park. Characterize the patterns by describing the kinds of vehicles participating, their spatial activities (where do they go?), their temporal activities (when does the pattern happen?), and provide a hypothesis of what the pattern represents (for example, if I drove to a coffee house every morning, but did not stay for long, you might hypothesize I'm getting coffee "to-go"). Please limit your answer to six images and 500 words.
2. Patterns of Life analyses may also depend on understanding what patterns appear over longer periods of time (in this case, over multiple days). Describe up to six patterns of life that occur over multiple days (including across the entire data set) by vehicles traveling through and within the park. Characterize the patterns by describing the kinds of vehicles participating, their spatial activities (where do they go?), their temporal activities (when does the pattern happen?), and provide a hypothesis of what the pattern represents (for example, many vehicles showing up at the same location each Saturday at the same time may suggest some activity occurring there each Saturday). Please limit your answer to six images and 500 words.
3. Unusual patterns may be patterns of activity that changes from an established pattern, or are just difficult to explain from what you know of a situation. Describe up to six unusual patterns (either single day or multiple days) and highlight why you find them unusual. Please limit your answer to six images and 500 words.
4. What are the top 3 patterns you discovered that you suspect could be most impactful to bird life in the nature preserve? (Provide a short text answer.)

Notable solutions identified global patterns among the various vehicle types, single and multiple instances of abnormal behaviors of vehicles traversing the park, changes in long standing patterns, and difficult to explain behaviors for vehicle behavior considering each of the factors of park use, traffic patterns, and expected behaviors for the vehicle types.

3.2 Mini-Challenge 2: Plume Analysis

The Preserve had four factories to the south of the Preserve: Roadrunner Fitness electronics, Kasios Office Furniture, Radiance

ColourTek paints, and Indigo Sol Boards (snow and skate). The four factories in the industrial area were subjected to higher-than-usual environmental assessment, due to their proximity to both the city and the Preserve. Gaseous effluent data from several sampling stations had been collected over several months, along with meteorological data (wind speed and direction), that could help Mitch understand what impact these factories may have had on the Rose-Crested Blue Pipit. These factories were supposed to be compliant with recent years' environmental regulations but Mitch had his doubts that the actual data has been closely reviewed. Substances of concern include:

Appluimonia – In general, most substances that cause odors in the outdoor air are not at levels that can cause serious injury, long-term health effects, or death to humans or animals. However, odors may affect quality of life and sense of well-being. Several odor-producing substances, including Appluimonia, are monitored.

Chlorodinine – Corrosives are materials that can attack and chemically destroy exposed body tissues. They might be hazardous in other ways too, depending on the particular corrosive material. An example is the chemical Chlorodinine. which is harmful if inhaled or swallowed.

Methylosmolene – This is a trade name for a family of volatile organic solvents. This chemical was strictly regulated in the manufacturing sector. Liquid forms of Methylosmolene are required by law to be chemically neutralized before disposal.

AGOC-3A – New environmental regulations and consumer demand have led to the development of low-VOC (Volatile Organic Compounds) and zero-VOC solvents. Most manufacturers now use one or more low-VOC substances and Mistford's plants have wholeheartedly signed on. These new solvents, including AGOC-3A, are less harmful to human and environmental health.

Mitch discovered that the state government has been monitoring the gaseous effluents from the factories through a set of sensors, distributed around the factories, and set between the smokestacks, the city of Mistford and the nature preserve. The state gave Mitch access to their air sampling data, meteorological data, and locations map.

For this mini-challenge the data included meteorological data that provided wind speed and direction for certain periods of time and sensor data for three months' worth of readings with the chemical detected, the sensor id, the reading in parts per million, and the date/time.

The Mini-Challenge 2 questions were:

1. Characterize the sensors' performance and operation. Are they all working properly at all times? Can you detect any unexpected behaviors of the sensors through analyzing the readings they capture? Limit your response to no more than 9 images and 1000 words.
2. Now turn your attention to the chemicals themselves. Which chemicals are being detected by the sensor group? What patterns of chemical releases do you see, as being reported in the data? Limit your response to no more than 6 images and 500 words.
3. Which factories are responsible for which chemical releases? Carefully describe how you determined this using all the data you have available. For the factories you identified, describe any observed patterns of operation revealed in the data. Limit your response to no more than 8 images and 1000 words.

Notable submissions in this mini-challenge specified their assumptions about the data, focused on visual analytics solutions, and made reasonable hypotheses about the factories that appeared to be responsible for hazardous emissions.

3.3 Mini-Challenge 3: Multispectral Imagery

The Boonsong Lekagul Preserve covered many miles of diverse terrain and to best understand what was happening to flora and fauna over time, Mitch realized that image analysis could shed insight into long term processes and impacts. Images were provided to him via a professor from his college:

Perhaps, the professor mused, there have been changes in the flora that are related to issues with fauna. Mitch thought this kind of associated study may be informative, so the professor provided him with some images of the Preserve collected by the National Space Service. However, they are multi-spectral image files, which Mitch had never dealt with before. The image analysis packages he found online were very complicated to work with, so he has asked you, a visual analytics expert, to help him view and understand this data.

A challenging component of this dataset was that it was multispectral, that is, there were three other wavelength channels included beyond red, green, and blue. As mentioned, we provided a multispectral analysis primer to assist contestants with their analysis. The questions for this mini-challenge started from the basic question of how large (in a linear measurement) was a pixel on the images supplied, to what normal phenomenon can you match to what is looking anomalous. Specifically, the questions were:

1. Boonsong Lake resides within the Preserve and has a length of about 3000 feet (see the Boonsong Lake image file). The image of Boonsong Lake is oriented north-south and is an RGB image (not six channels as in the supplied satellite data). Using the Boonsong Lake image as your guide, analyze and report on the scale and orientation of the supplied satellite images. How much area is covered by a pixel in these images? Please limit your answer to 3 images and 500 words.
2. Identify features you can discern in the Preserve area as captured in the imagery. Focus on image features that you are reasonably confident that you can identify (e.g., a town full of houses may be identified with a high confidence level). Please limit your answer to 6 images and 500 words.
3. There are most likely many features in the images that you cannot identify without additional information about the geography, human activity, and so on. Mitch is interested in changes that are occurring that may provide him with clues to the problems with the Pipit bird. Identify features that change over time in these images, using all channels of the images. Changes may be obvious or subtle, but try not to be distracted by easily explained phenomena like cloud cover. Please limit your answer to 6 images and 750 words.

Reviewers were looking for submissions that went beyond simple displays of the image data and employed interesting visual analytics to support investigation into the changes in the area over time.

3.4 Grand Challenge

The Grand Challenge involved integration and synthesis over all of the mini-challenges to understand the situation, where blending of information and discoveries led to enhanced insight. This year's grand challenge asked for a coherent and comprehensive story about the impacts to the Pipit. Some additional data was provided to the contestants in the form of newsletters describing the Mistford factories' activities over several years. The newsletters contained information about charity fundraisers and employee accomplishments, but also some interesting financial news and other data that could be woven into the story of the Pipit. The following were the questions for the Grand Challenge:

1. Provide your best hypothesis with supporting evidence of what is happening in the Lekagul Preserve that is affecting the Rose-crested Blue Pipit. Your answer should include identification of who is responsible for the impactful activities, what they are doing, where impactful events are occurring, when and how often these occur, how these events are taking place, and why they are happening.
2. Provide a timeline that comprehensively describes the relevant activities in Mistford, the Industrial Park, and the Preserve that helps concisely describe the events identified in Question 1.
3. How confident are you in your hypothesis? What factors impact your confidence in your hypothesis? What additional information would help strengthen your hypothesis?
4. What are your proposed next steps? Do you have a course of action to correct the problems in the Lekagul Preserve and help the Rose-crested Blue Pipit?

A desirable Grand Challenge entry would include information and analysis from each mini-challenge and the Grand Challenge with hypotheses, evidence and confidence levels in these estimations as supporting evidence.

4 Review Process

The VAST Challenge relies on an anonymous peer review process to provide feedback to the participants and to recommend submissions for award consideration. All submissions are reviewed by researchers in the visual analytics community. In addition, all submissions are reviewed by subject matter experts who do not typically participate in visual analytics research but understand the application areas being addressed. This year, we included experts who work in plume analysis and multispectral image analysis, and who are not regular participants in the visual analytics community. A total of 72 reviewers provided comments for up to ten submissions. Each submission received at least three reviews.

Reviewers were provided with guides containing background on the mini-challenges and descriptions of the types of responses the committee anticipated. The evaluation process relied on the reviewers' expert judgment, supported by these review guides. Each reviewer was asked to provide an overall rating for the submission on a 1-5 scale, along with a written rationale for the rating. The remaining review questions were tailored to correspond to a specific Mini-Challenge.

All reviewers were given an opportunity to nominate deserving submissions for awards. Reviewers were given freedom to identify any aspect of the submission that they deemed worthy of recognition. When nominating a submission for recognition, the reviewer was asked to suggest an award title and to provide a rationale for their nomination. Upon receiving the reviewer feedback, the VAST Challenge committee held a two-day meeting to consider the submitted award nominations and finalize the selection of submissions for recognition. We provided two categories of awards: awards and honorable mentions.

5 Award Winners and Honorable Mentions

The submissions recognized for awards and honorable mentions in 2017 are listed in Table 2. Additional information about all completed entries can be found in the VAST Challenge papers included in the electronic proceedings of the 2017 IEEE Conference on Visual Analytics Science and Technology (VAST)

Mini-Challenge 1
Eindhoven University of Technology (#103) Award: Elegant support for hypothesis generation and testing
Singapore Management University (#107) Award: Actionable and detailed analysis
SAS (#115) Award: Outstanding clarity of presentation
University of Konstanz (#145) Award: Excellence in spatio-temporal graph analytics
University of Korea (#165) Honorable Mention: Novel use of experimental coordinated visualizations
Mini-Challenge 2
giCentre, City University London (#147) Award: Comprehensive MC2 solution
University of Illinois – Chicago (#140) Honorable Mention: Clarity in visual communication
Mini-Challenge 3
University of South Florida (#106) Honorable Mention: Good facilitation of single image analysis
University of Illinois – Chicago (#141) Honorable Mention: Good interactive image explorer for temporal analysis
Multi-Challenge
Purdue University (#119 & #108) Award: Aesthetic design
HKUST (#124 & #126) Award: Compelling synthesis of information
Singapore Management University (#152 & #155) Award: Clear presentation of hypotheses and supporting evidence
Peking University (#142, #172, & #160) Award: Combining automated and visual analytics

Table 2. VAST Challenge 2017 Awards and Honorable Mentions

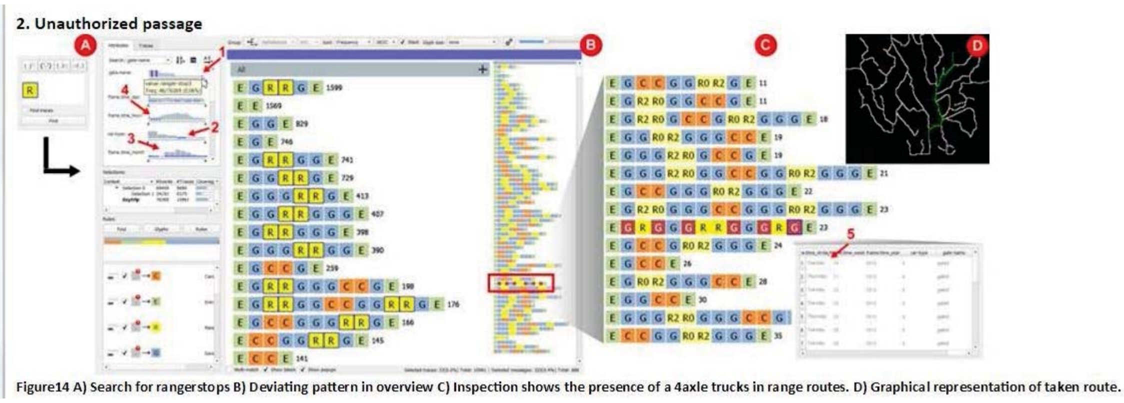

Figure 1. Using EventPad in examination of Mini-Challenge 1 (Eindhoven, Cappers)

and in the Visual Analytics Benchmark Repository. We present a sample of the award winning submissions below to highlight excellent work.

Eindhoven University (Bram C.M. Cappers) submitted an entry to Mini-Challenge 1 highlighting the use of EventPad, a custom tool to visualize events through glyph sequences (Figure 1). Through the application of command-line regular expressions and predicate logic, users can search and color glyphs based on event attributes of interest. This entry assigned colors to camping events, gate traversals, and ranger stops. From there, they were able to visually stack similar sequences and perform sorting and searching. Figure 1 illustrates the use of this tool to discover unauthorized trucks passing through a ranger stop, not an allowable activity in the Preserve. This discovery was a key element of the ground truth for Mini-Challenge 1, as these trucks were involved in illegal dumping activities.

GiCentre - City University London (Jo Wood) contributed a submission featuring a highly detailed analysis of the sensor data to obtain an understanding of sensor function over time and to examine the supplied data for anomalies and other characteristics of interest. This submission showed a zoomable map for the sensors and the factories enabling visual examination of chemical emission events, and allowing users to clearly see and understand activities at the factories (Figure 2). The analysis was sharp and detailed, and the figures included in the submission provided full-sized versions when clicked on—a feature greatly appreciated by the reviewers.

Peking University and Qihoo 360 provided entries to the three mini-challenges and was given a “multi-challenge award” for its entries in mini-challenges 1, 2 and 3. These entries, submitted by different teams, had notable attributes and received praise from the reviewers for their contributions across the various components of the challenges. Figure 3 provides an example of their work on Mini-Challenge 3. While the question asked participants to estimate the size of a pixel for the images provided, the Peking MC3 team took extra steps to ensure appropriate alignment of the two provided images to obtain a correct measurement. That extra effort encouraged the award committee to recognize the teams and the University for the skill demonstrated across the three mini-challenges.

6 Observations from the VAST Challenge Committee

In this section, we discuss characteristics and trends observed in this year’s challenge submissions. We also discuss the implications of the observation on future Challenges and related activities.

6.1 Exploration and Creativity in Visual Analytics

The VAST Challenge provides a forum for researchers to create and express new forms of visualization, interaction, and exploration against the offered datasets. We seek and recognize innovation.

The forum is especially useful as participants provide posters and 2-page papers to describe their work, and great feedback can be obtained from the VIS community during the workshop day and throughout the conference.

6.2 Visual Analytics for Multispectral Analysis

The multispectral images of Mini-Challenge 3 may have been new to this community. Domain experts in the area are excited to see the innovations visual analytics can provide. This year provided an introduction to this area and the submissions provided an excellent foundation for future research. Our next VAST Challenges will provide additional opportunities for participants to explore more complex problems.

6.3 Blending Different Computing Technologies with Visual Analytics

A few submissions explored using machine learning with visual analytics. This kind of blending technologies is very exciting to the VAST Challenge committee and we would like to encourage such future explorations. We also welcome design features feedback we could include in future challenges to better enable the synergy.

6.4 Opportunities for Integrated Analysis

Problems posed by this year’s VAST Challenge offer several interrelated datasets for analysis. While analysis of each dataset individually can provide insight, the truly outstanding teams examine their data iteratively, using data from one dataset to help test their hypotheses developed in another dataset or mini-challenge. Looking for agreement or contradiction among data sources is an important element of the analysis process for any problem including the VAST Challenges, but a technique that was employed only by the most sophisticated submissions.

7 Conclusion

Now in its twelfth year, the VAST Challenge continues as a resource for the visual analytics research community. Through the generous support of the University of Maryland and the University of Massachusetts in maintaining the Visual Analytics Benchmark Repository [1,2], archived datasets from the past decade of VAST Challenge competitions remain freely available for use in student research projects and Visual Analytics courses worldwide. In addition, the research community is able to use these datasets along with the ground truth provided in the solution for evaluation and testing of new analytic approaches.

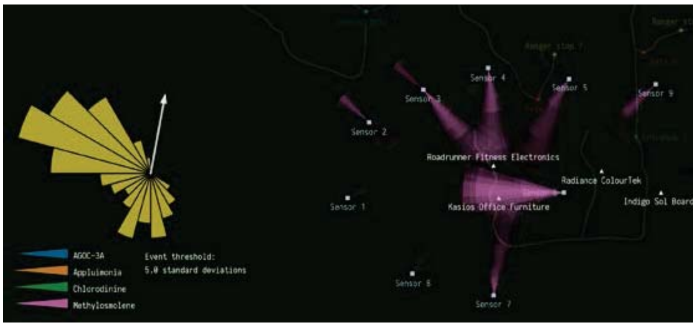

Figure 2. Emission Plumes and Wind Indicators (giCentre, Wood)

8 Acknowledgments

The committee acknowledges the US Department of Energy, Pacific Northwest National Laboratory, MIT Lincoln Laboratory, the University of Massachusetts Amherst, Smith College, and the US Department of Defense for supporting the VAST Challenge. The committee also wishes to thank Catherine Plaisant at the University of Maryland and Brad Fritz at Pacific Northwest National Laboratory.

References

[1] University of Maryland, "Visual Analytics Benchmark Repository". [Online]. http://hcil2.cs.umd.edu/newvarepository/.

[2] University of Massachusetts Amherst, "Visual Analytics Benchmark Repository". [Online]. http://vast.cs.umass.edu/vast/repo/.

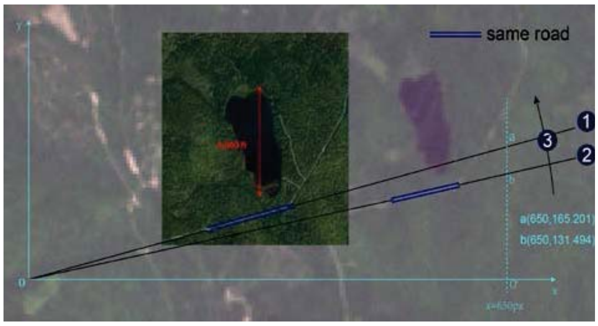

Figure 3. Alignment of imagery for accurate measurement (Peking and Qihoo 360)