BNL-211843-2019-COPA

An Immersive Visual Analytics Platform for Multidimensional Dataset

B. Sun, W. Xu

Submitted to the 18th International Conference on Computer and Information Science

ICIS 2019 Conference
to be held at Beijing, China
June 17 - 19, 2019

Computational Science Initiative

Brookhaven National Laboratory

U.S. Department of Energy
USDOE Office of Nuclear Energy (NE), Office of Nuclear Reactor Technologies (NE-7)

DISCLAIMER

An Immersive Visual Analytics Platform for Multidimensional Dataset

Bo Sun
Dept. of Computer Science
Rowan University
Glassboro, NJ, USA
sunb@rowan.edu

Aleksandr Fritz
Dept. of Computer and Electrical Enginnering
Rowan University
Glassboro, NJ, USA
fritza7@students.rowan.edu

Wei Xu
Computer Science Initatives
Brookhaven National Lab
Upton, NY, USA
xuw@bnl.gov

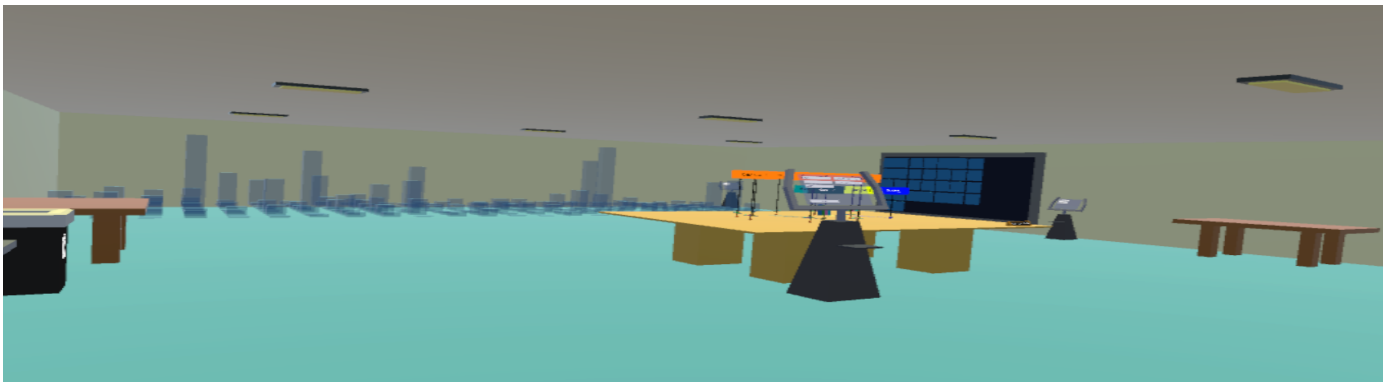


Fig. 1. Overview of VR Laboratory for Visual Analytics

***Abstract*—Virtual Reality is a budding new medium of interaction, but ways in which to use it for data analytics are not immediately obvious. Through our research, we investigated new and useful ways to view multidimensional data in a virtual environment. In this paper, we present a virtual laboratory with collaborative visualizations in different spaces. We focused on geographic view in the center of the room by utilizing spatial freedom in Virtual Reality, which informed both the geographic visualization and time series data. Through this laboratory and the collaborative visualizations housed within, we believe we've found a novel way to interpret data in Virtual Reality.**

Keywords— VR, visual analytics, data analytics and immersive data visualization.

I. Introduction

Exploring and analyzing big data translates information into insight. Visualization has been approved to be a key function in data exploring and analyzing. Visual analytics, as a new field of data analytics, is especially concerned with coupling interactive visual representations with underlying analytical processes such as statistical procedures, data mining techniques so that high-level, complex activities such as sense making, reasoning, decision making can be effectively performed [1]. Visual analytics techniques have been shown to effectively benefit analytical reasoning in many scenarios [2]. Human-information discourse is therefore enabled by applying human knowledge through interaction with the visualization interface. With rapid grow of big data, dataset is not just growing in scales but also in dimensions. A major challenge from traditional 2D visualization is that there are few dimensions to support multi-dimensional information of the dataset. Therefore, it is very common that multiple 2D graphs are produced for side by side comparison, which prevents many discoveries of data correlation.

In the past decade, visualization has seen profound advances on multi-dimensional data [3][4]. Recently, the emerging of low-cost immersive head-mounted devices (HMD), such as HTC VIVE, Oculus Rift, and Microsoft HoloLens, present another opportunity for data exploration and interaction. These immersive devices adopt stereoscopic techniques to create an engaging and immersive visual environment [5]. They provide by nature the additional freedom to observe and interact with multi-dimensional data compared to conventional 2D or 3D visualization interfaces. There are many works on HMD-based immersive analytics, such as graph visualization [6], urban visualization [7], and geographic visualization [8].

Specifically, geographic visualization is to combine visual interface with geographical representations of the dataset. The map-based interfaces would further support the understanding of information and allows users to get a better sense of spatial datasets by overlaying the data on the map to help guide exploration and analysis [9]. Consequently, users gain an additional sense of analytics reasoning by comparing and connecting with their surroundings [10]. When working in spatial and geographical domains, simulations and virtual reality can lead to better discovery as showed in [10]. Virtual Reality is capable of constructing user experiences that connect time, spatial, interaction and computer graphs all together. This unique combination allows visual analytics process to better relate the spatial, time with other data dimensions, which leads to more analytics reasoning and provides more insight and finding of multidimensional dataset.

To solve the dimension restriction challenge, particularly associated with temporal and spatial dataset, we developed a new immersive data analytics platform by enriching the interaction interface in Virtual Reality. The VR platform allows us to explore the data set through one geographic based view and collaborative visualizations. Users can quickly identify analytic reasoning supported through interactive 3D interface by observing how often peak chemical readings are associated with the factories determined by meteorological data. The new platform allows users to visualize multidimensional data in one scene, speed up the discovery of data correlation, trends and insight, which effectively impacted the analytical reasoning.

The remaining portions of this article are structured as follows. Section II describes the implementation of our application from data preparation to visualization design and the analytical tasks that can be performed by the user. Section III discusses the user interaction the platform provides. Finally, we conclude and mention areas of future work in section IV.

II. Platform Implementation

The data set used in this paper was retrieved from 2017 VAST challenge [11], which provides toxic chemical readings along with meteorological data in a nature preserve that surrounded four factories, Kasios, Radiance, Indigo, and Roadrunner. A local graduate student had found that a rare species of bird had been declining in population. They suspected that the factories in the area were not in compliance with emissions regulations, leading to the population decline. The purpose of the challenge was to identify which, if any, factories had been emitting toxic chemicals above the legal limit using visual analytics approaches. We had previously attempted to visualize and analyze the data through a 3D display using Unity [12], but now we have implemented a more in depth display using Unity in VR.

A. *Mutildementional Temporal and Spatial Dataset*

Specifically, the data consists of sensor readings from a set of air-sampling sensors and meteorological data from a weather station in proximity to the factories and sensors. The Meteorological data represents 3 months of readings in the following format:

TABLE I: METEOROLOGICAL DATA [11]

Date	**Wind Direction**	**Wind Speed (m/s)**
4/1/16 0:00	190.5	4
4/1/16 3:00	203.3	5
4/1/16 6:00	201.1	5.2
4/1/16 9:00	204.9	4.1
4/1/16 12:00	207	3.6

- Date: The date and time that the readings were collected.
- Wind Direction: The compass directions where the wind is originating from, where 360/000 is true north.
- Wind Speed: The speed of the wind.

Each of these reading is taken at the date and time provided. The Sensor data contains 3 months of readings in the following format:

TABLE II: SENSOR DATA [11]

Chemical	**Monitor**	**Date Time**	**Reading**
Methylosmolene	3	4/1/16 0:00	2.68382
Methylosmolene	7	4/1/16 0:00	2.63064
Chlorodinine	3	4/1/16 0:00	1.25917
Chlorodinine	7	4/1/16 0:00	0.943983
AGOC-3A	3	4/1/16 0:00	0.722303

- Chemical: One of the four chemicals detected by the sensors
- Monitor: One of the nine sensors picking up the reading
- Reading: The air sensor detected amount in term of parts per million
- Date Time: The date and time of day of the reading was collected.

In addition, the factories and sensors locations are presented on a 200 x 200 grid using x, y coordinates, with (0,0) at the lower left hand corner (southwest) [11].

B. *Viusalization Laboratory*

We created a room to house the collaborative visualizations. Instead of being an abstract environment, the room view, as showed in figure 1, provided intuitive feeling that make user felt comfortable and familiar enough to navigate. We used the Virtual Reality Toolkit (VRTK) package [13] in Unity [14] to simplify VR functionality implementation. For locomotion in the room, we used teleporting system, as that is the current standard for locomotion to minimize motion sickness as much as possible.

C. *Geographic Display*

We placed a geographic view in the center of the laboratory. The new spatial view used the advantage of virtual reality to render spatial relations of the dataset.

1) *Geographic View*

The factories and sensors were plotted in the new geographic display based on the given coordinates. As we were only given latitudinal and longitudinal data, so heights were assumed to be uniform.

In the geographic view, users could view the chemical readings as a 7-days time series. The chemical reading associated with time series data could be played back dynamically as showed in figure 2 (a), or users could manually

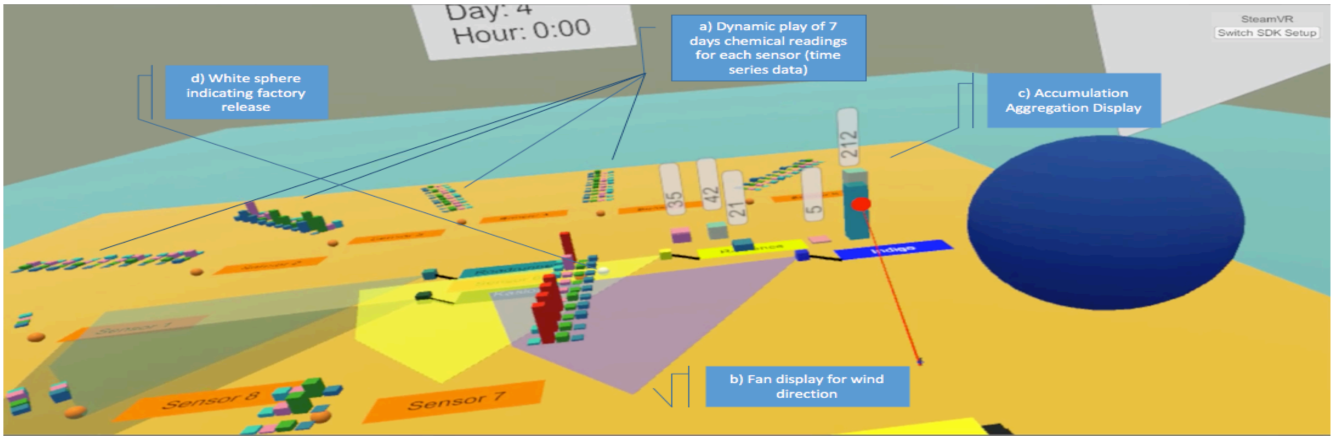


Fig. 2. Geographic View

move to a particular date and hour using dial control to see the readings of the sensors at that time. Along with showing chemical readings, the meteorological wind data was displayed. In figure 2 (b), the wind was displayed as four fans for the wind direction, each originating from one of the four factories. This allowed us to intuitively see if a factory's output was contributing to the chemical reading on one or more sensors. In addition to that, an accumulation was shown above the factories as showed in figure 2 (c) along with a moving white sphere as seen in figure 2 (d) in the center of the fan to illustrate the confirmed accumulation. The accumulations shown were the amount of chemical readings that were determined to be caused by the specific factory. We determined that a reading was contributed by a factory through a vector system [12]. If the dot product of the normalized vectors of the wind direction and the direction from a factory to a sensor was above a set threshold, it was determined that the reading of that particular sensor was due to the factory. To filter out natural environmental readings, only peak readings were considered when determining factory accumulation.

2) Time Control

Dials at the right side of the table, as showed in figure 3, allow users to choose the time period to observe the dynamic data visualization on the geographic table.

Playback is controlled through a blue sphere. This sphere has a display attached showing the current date and time of the playback. To use the sphere, users must grab it and press the trigger of the Vive controller to play or pause the playback. For feedback, the sphere turns purple when the trigger is held, and turns back to blue when trigger is released. There are two spheres in the area so that users may also place a display showing the current date and time wherever they choose, alongside the display of the date and time above the geographic view.

3) Accumulation Aggregation

We added a function to the geographic view to make it easier to compare accumulation levels of different factories. Now, users can press a button to create a snapshot of all of the current accumulations. This snapshot is created on top of a grabble sphere so that the display can be placed wherever the user desires. There is also an accumulation scaling dial on the right side of the table that uniformly scales all accumulations on the table so that either larger disparities can be easily seen by scaling down, or smaller differences can be investigated by scaling up as seen in figure 4.

D. Collaborative Visualizations

Additionally, we implemented different stations located throughout the room to support visual analytics by observing the dataset from different aspects.

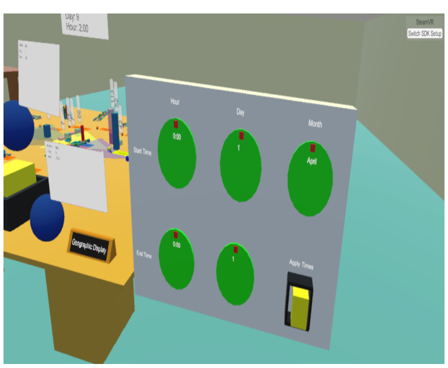


Fig 3: The dial controls and the application button

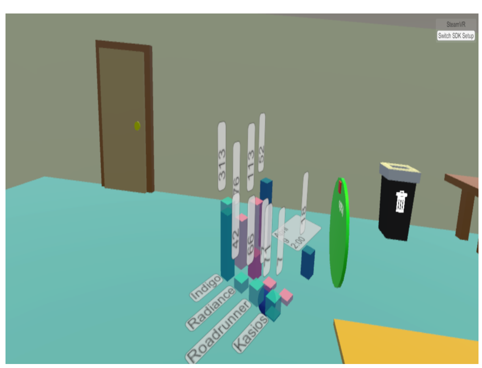
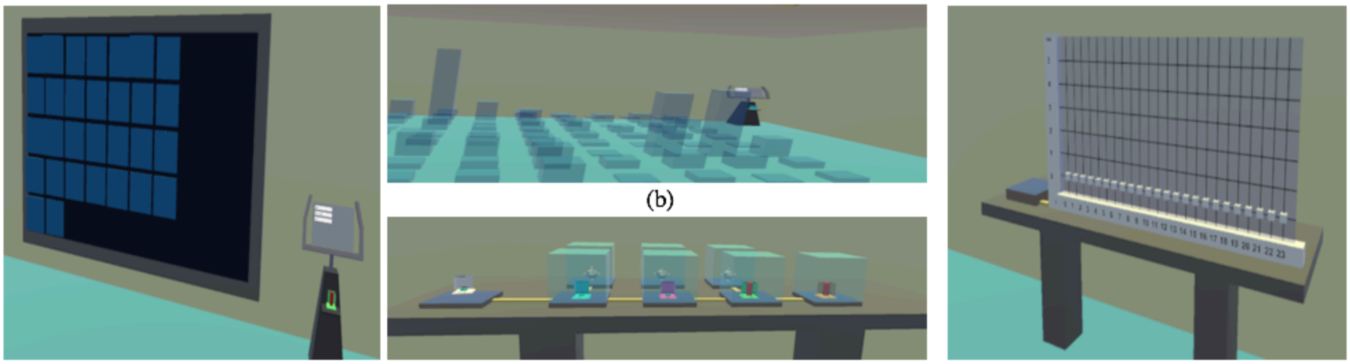


Fig. 4. The accumulation snapshot and accumulation scaling dial

Fig. 5. Analysis Table



(a) (b) (c) (d)

Fig. 6. (a) Calendar Display; (b) Day Bar Graph; (c) Aggregation Table (d) Inactive Line Plot

1) Analysis Table

This station as showed in figure 5 reads the information of a reading object with the aggregation level of one hour. This displays all of the base information of that data point to the user. It also works as a debugging tool during development, since it checks if data is being passed and read properly when creating reading objects.

2) Calendar View

This station as seen in figure 6 (a) shows the average reading of a certain chemical at a certain day in a month. It lets users see the average readings in a month at a glance. Since each bar shown is a day, it also works as a way to extract day readings from month readings.

3) Day Bar Graph

This station as seen in figure 6 (b) shows the readings for a day of a certain chemical. The short axis is the sensor number, one through nine, and the long axis is the hour, 0:00 through 23:00. The vertical axis is the reading of that particular sensor and hour. The bars are transparent, not opaque, to prevent readings from being occluded by larger readings surrounding them. Since each bar shown is a set of hour, it also works as a way to extract hour readings from day readings.

4) Day Line Plot

This station shows a line plot of readings over a day of a certain chemical at a certain sensor. It is similar in function to the Day Bar Graph, but it is on a table instead of encompassing a large part of the room. It also only shows a single sensor as opposed to all sensors. Like the Day Bar Graph, it works as a way to extract hour readings from day readings as showed in figure 6 (d).

5) Aggregation Table

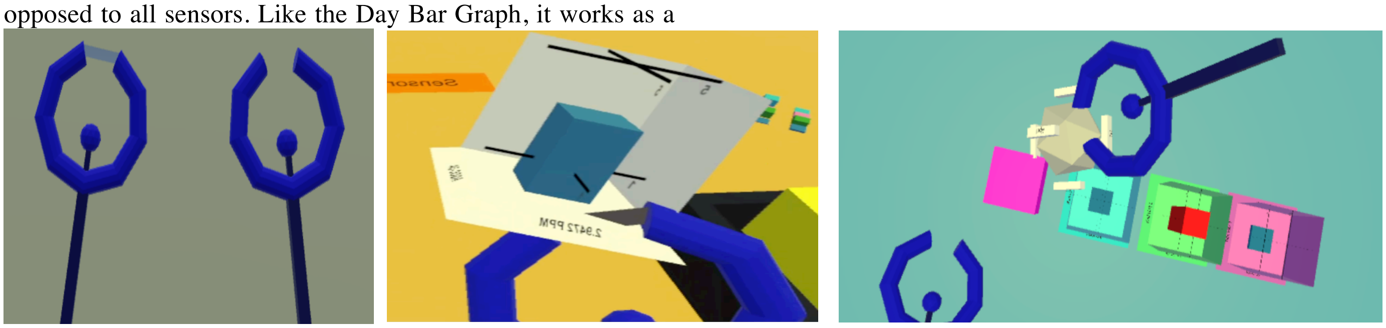
This station takes in readings and aggregates according to various time segments such as per day, per week and per month. It can take any aggregation level of reading and creates objects of every aggregation level as seen in figure 6 (c).

III. Design Principles

To standardize interactivity and best teach users how to use the VR based platform, we came up with a set of principles. These principles guide our decisions when designing new displays and interactions.

A. User Interaction

The user can interact with the world mainly through teleportation and grabbing as provided through the VRTK package. When grabbing with the base controller models, it was found that the collision between the controllers and the environment was not clear. To fix this, a pair of claw controllers was created (see figure 7.a). These models were structured in lower polygon count to help with performance. The claws have a sphere in the middle that is the origin of the pointers used for teleportation and basic GUI interaction. The grab collision is between the two end points of the claw's pincers, and a gray bar is shown when grabbing to visually communicate the location of the collision box. The physical reading objects can be held and dropped by the claw controllers by pressing the grip buttons on the side of the Vive controllers.



(a) (b) (c)

Fig. 7. (a) User controllers, left controller is grabbing (b) A single physical reading (c) Toolbelt holding readings

Readings are the basis of data interpretation. This is why we chose to highlight the data with fantastical functionality. Whenever a reading is shown in any display, the user can grab it to extract an object as seen in figure 7 (b). This object shows the relevant information pertaining to the specific reading, including value, dimension, and aggregation level. Time reading objects, also called chronospheres, also exist and are interactable. The presence of these objects both give easy insight into specific readings and also extend the functionality of stations, as explained in section II. D. 5.

User may want to hold multiple readings at once. To remedy this, a toolbelt *VR* was added. This tool belt is made up of multiple object drop locations. The drop location functionality was provided by the VRTK package. When a valid object is held over a drop location, the location is highlighted to show its position and the object's validity. If a valid object is dropped inside a drop location, it anchors to the point. These objects can then be picked back up from the drop location for future use. With multiple drop locations situated around the user's waist, a toolbelt is created that lets the user carry multiple reading objects with them as seen in figure 7 (c).

B. Be Fantastical Sparingly

Grounded interactions and displays are more familiar and intuitive to user. This means that the more workspace is grounded, the less the user needs to be taught. Fantastical interactions are also useful such as grabbing readings to highlight a specific part of data, or teleporting for locomotion as they are the best way to implement certain functionality.

C. Standardization of interaction and Form

Standardization creates knowledge crossover. We attempted to create the same interaction for all visualization stages presented in the room so that user can quickly adopt a new display as most of the functionality is shared between different displays.

D. Comfort Awareness

When creating a realistic environment, user comfort is key, especially if they are to inhabit the environment for extended periods of time, particularly for visual analytics process in this case. Standard practice for keeping users comfortable is limiting harsh sensory input, no loud sounds or overly bright colors. Specifically, ceiling lights were added to the room to give a mental source for the lighting in the room. After minor testing in the beginning, it was found that users could feel trapped and closed in, so a door was added. Even though the door is not functional, it makes the room feel more real and provides an escape route on a subconscious level.

Different objects are located in the room to provide extra functionality to supplement the stations. Trash Cans are located in multiple areas and are explicitly labeled in the front with trash can icons. When reading objects are no longer useful, users can deposit them in the trash cans to delete them. This functionality is grounded and intuitive, to the point that in every test, users figured out the use for them without direction. A broom is also located in the room to sweep up readings that may have accumulated over time. In many test cases, enough readings piled up that individually throwing away each one was time consuming, so the broom was added. Desks are also located in the room for users. These provide users a space away from the stations to collect or compare readings.

IV. Conclusions

The VR laboratory presents a collaborative visualization platform to conduct visual analytic in immersive environment. We used the spatial freedom of VR to render multidimensional dataset, specifically associate with time and geographical attributes. The geographical visualization in the center of the lab integrated meteorological data with sensor reading data, it took privilege of VR and presents a unique user experience for analytics reasoning. The center station combines dynamic visualization of the time series data with the spatial location of sensor and manufactory, users can directly observe interrelationship between peak reading and wind direction; and the accumulation bar further verify the observation, which results to an intuitive view to users. The amount of data dimensions presented in the VR platform, 2D interface is not capable to hand effectively.

Additionally, the collaborative visualization presents several stations to support other analysis aspects including aggregations, calendar view, day bar graph and line plots. Our future work includes further developing line plots and investigating the best user friendly navigation in immersive environment.

Acknowledgment

We would like to acknowledge the funding support of Brookhaven National Lab. This research was supported by two Lab Directed Research and Development projects 16-041 and 17-029 of Brookhaven National Laboratory.

References

[1] J. J. Thomas, "Illuminating the path: the research and development agenda for visual analytic", IEEE Computer Society, 2005.

[2] D. A. Keim, "Information visualization and visual data mining", IEEE Transactions on Visualization and Computer Graphics, vol. 8, no. 1, pp. 1–8, 2002.

[3] A. Kerren, J.T. Stasko, J.D. Fekete, C. North. "Information Visualization." ACM Sigmod Record, vol. 4950, no. 25, pp.14-15, 2008. doi:10.1007/978-3-540-70956-5.

[4] S. Liu, D. Maljovec, B. Wang, P. T. Bremer, V. Pascucci, "Visualizing high- dimensional data: Advances in the past decade", IEEE Transactions on Visualization and Computer Graphics vol. 23, no. 3, pp. 1249–1268, 2017. doi:10. 1109/TVCG.2016.2640960.

[5] B. Bach, R. Dachselt, S. Carpendale, T. Dwyer, C. Collins, and B. Lee. "Immersive analytics: Exploring future interaction and visualization technologies for data analytics", In ACM Proceedings on Interactive Surfaces and Spaces, 2016.

[6] O. Kwon, C. Muelder, K. Lee, and K. Ma. "A study of layout, rendering, and interaction methods for immersive graph visualization", IEEE Transactions on Visualization and Computer Graphics, 2016. doi: 10.1109/TVCG.2016.2520921

[7] Z. Chen, H. Qu, and Y. Wu. “Immersive Urban Analytics through Exploded Views”, Workshop on Immersive Analytics: Exploring Future Visualization and Interaction Technologies for Data Analytics. Phoenix, AZ. 2017.

[8] Terrain Texture. Last accessed at https://assetstore.unity.com/packages/2d/textures-materials/floors/terrain-textures-snow-free-samples-54630 in Jan. 2018.

[9] Q. Ho and M. Jern, “Exploratory 3D Geovisual Analytics”, IEEE International Conference on Research, pp. 276–283, 2008.

[10] A. Moran, V. Gadepally, M. Hubbell, J. Kepner. “Improving Big Data Visual Analytics with Interactive Virtual Reality”. IEEE High Performance Extreme Computing Conference (HPEC '15), 2015.

[11] VAST Challenge. Last accessed at http://www.vacommunity.org/VAST+Challenge+2017 in Jan. 2018.

[12] A. Fritz, B. Sun, W. Xu. “Immersive Visual Analysis To Explore Mystery At Wildlife Preserve”, IEEE Virtual Reality and 3DUI, Reutlingen, Germany, Mar. 2018. doi: 10.1109/VR.2018.8446324

[13] VRTK - Virtual Reality Toolkit. Last accessed at https://vrtoolkit.readme.io/ in May 2018.

[14] Unity3d. Last accessed at http://unity3d.com/ in March 2019.