ARL-TR-8504 ● SEP 2018

US Army Research Laboratory

Reconfigurable Visual Computing Architecture for Extreme-Scale Visual Analytics

by Simon Su, Luis Bravo, and Venkateswara Dasari

NOTICES

Disclaimers

ARL-TR-8504 ● SEP 2018

US Army Research Laboratory

Reconfigurable Visual Computing Architecture for Extreme-Scale Visual Analytics

by Simon Su and Venkateswara Dasari
Computational Information Sciences Directorate, ARL

Luis Bravo
Vehicle Technology Directorate, ARL

Approved for public release; distribution is unlimited.

REPORT DOCUMENTATION PAGE	*Form Approved* *OMB No. 0704-0188*

1. REPORT DATE *(DD-MM-YYYY)*
September 2018

2. REPORT TYPE
Technical Report

3. DATES COVERED (From - To)
October 2017–July 2018

4. TITLE AND SUBTITLE
Reconfigurable Visual Computing Architecture for Extreme-Scale Visual Analytics

5a. CONTRACT NUMBER

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
Simon Su, Luis Bravo and Venkateswara Dasari

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
US Army Research Laboratory
Computational and Information Sciences Directorate (ATTN: RDRL-CIC)
Aberdeen Proving Ground, MD 21005

8. PERFORMING ORGANIZATION REPORT NUMBER
ARL-TR-8504

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution is unlimited.

13. SUPPLEMENTARY NOTES

14. ABSTRACT
Major advancements in computational and sensor hardware have enormously facilitated the generation and collection of research data by scientists; the volume, velocity, and variety of big data research have increased across all disciplines. A visual analytics platform capable of handling extreme-scale data will enable scientists to visualize unwieldy data in an intuitive manner and guide the development of sophisticated and targeted analytics to obtain useable information. The reconfigurable Visual Computing Architecture (VCA) is an attempt to provide scientists with the ability to analyze the extreme-scale data collected. The reconfigurable VCA requires the research and development of new interdisciplinary technological tools that integrate data, real-time predictive analytics, visualization, and acceleration on heterogeneous computing platforms. The reconfigurable VCA will provide scientists with a streamlined visual analytics tool.

15. SUBJECT TERMS
visual analytics, data visualization, immersive analytic, large high-resolution display

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			Simon Su
Unclassified	Unclassified	Unclassified	UU	21	19b. TELEPHONE NUMBER (Include area code) (410) 278-1421

Standard Form 298 (Rev. 8/98)
Prescribed by ANSI Std. Z39.18

Contents

List of Figures

Acknowledgments

This work was supported in part by the DOD High-Performance Computing Modernization Program at the US Army Research Laboratory, Department of Defense Supercomputing Resource Center.

1. Introduction

The reconfigurable Visual Computing Architecture (VCA) is an extreme-scale-enabled visual analytics platform for scientists who work with various components of the system being developed incrementally to fulfill various data analysis and visualization gap. As visual analytics solutions currently used by scientists are each at their own different maturity level, scientists have had to resort to using different data analysis and visualization tools with varying levels of manual intervention to overcome the limitations of the individual tool to accomplish their research goals. These limitations are interfering with their research workflow, resulting in unnecessary delays affecting overall productivity of the research.

The reconfigurable VCA is an attempt to streamline the visual analytics process, providing the scientists with the ultimate visual analytics tool. There are many reasons why the fundamental concept behind the reconfigurable VCA only exists in Hollywood blockbusters like *Avatar* and *Minority Report*. One of the main reasons is because of the inherit nature of visual analytics being a customized process tailored specifically to the type of data and knowledge discovery goal. With limited resources, domain scientists usually choose to address their own research challenges and use existing visual analytics workflows. However, given the scale of their data, scientists may have to get out of their comfort zone in search of a visual analytics tool capable of tackling data at extreme scales. Reconfigurable VCA research areas tend to be studied in isolation in academia, and developing a joint framework will introduce innovative and fundamentally new research problems and solutions.

In the age of big data, the process of data exploration, modeling, and analysis requires a vastly different class of tools compared to the visual analytics process of a few decades ago. It is crucial to be able to quickly analyze large data sets to identify trends, anomalies, and correlations in a rapidly changing and data-driven world. For problems in the information visualization domain, identification of these trends, anomalies, and correlations in data from past events improves our ability to anticipate future events to stay at the forefront of ongoing developments. For problems in the computational sciences domain, the ability to visualize and analyze vast data sets can help improve the accuracy of the computational model. One of the many issues is the need to find more complex relationships among the large numbers of high-dimensional variables. Traditional data exploration using column data formats can be limiting in power, as with spreadsheets, or complex and accessible only to trained data scientists and statisticians, as with programmatic tools like R and Python. Thus, development of useful information visualization tools can meaningfully augment the discovery of relationships in data for a larger pool of users, across more domains.

The US Army test and evaluation (T&E) community tests and evaluates everything the Soldier touches and everything that touches the Soldier, including networks, applications, vehicles, weapons, communication devices, data links, and so on. Therefore, T&E is the single largest producer of data in the Department of Defense research, development, test, and evaluation community. Their big data challenges are diverse, as T&E measures everything conceivable to assess its effectiveness, suitability, survivability, and safety. These requirements produce massive, heterogeneous, distributed data sets requiring new approaches for analysis and exploitation. Furthermore, there are growing number of requirements for real-time or time-critical analysis of the heterogeneous data collected. These are also common requirements throughout the larger Department of Defense T&E community.[1]

Besides the ability to support big data visual analytics, the ever-growing problems faced by data analysts and computational scientists today also require a collaborative team of analysts and scientists to address the challenge. A realistic problem requires dedicated team members with interdisciplinary expertise that can offer a diverse perspective to solve the problem at hand. Therefore, the concept of data visualization has evolved into collaborative large-scale data visualization, with collaboration as one of the grand challenges for visual analytics.[2]

A scalable data visualization framework capable of supporting large, high-resolution, collaborative visual analytics and immersive analytics opens up the possibility for the users to visualize and interact with more of the data present in a big data problem. In scientific visualization, visualization of a data set with hundreds of millions of data points requires a display system with at least that many pixels for an overview visualization of the data. For information visualization, the unstructured nature of the data can unfold in many different ways, driving the large higher-resolution display (LHRD) requirement. Andrews et al. described the potential benefits for using LHRDs for information visualization.[3] A visualization framework capable of supporting multi-user interaction can also help facilitate the collaboration process. All the team members working on a visualization collaboration session should have the ability to interact directly with the content displayed on the visualization system using their own mouse. Team members can then interact with the shared visualization workspace as they would on their own personal laptop, while sharing and discussing their insights resulting from both the visualization and collaboration efforts.

We developed a visualization software suite supporting multiuser collaborative visual analytics using heterogeneous visualization hardware and software. Depending on the visual analytics tasks, the reconfigurable VCA allows the user to use the most suitable combination of visualization hardware and software to support the tasks. Figure 1 shows the design of our overall reconfigurable VCA to support extreme-scale visual analytics.

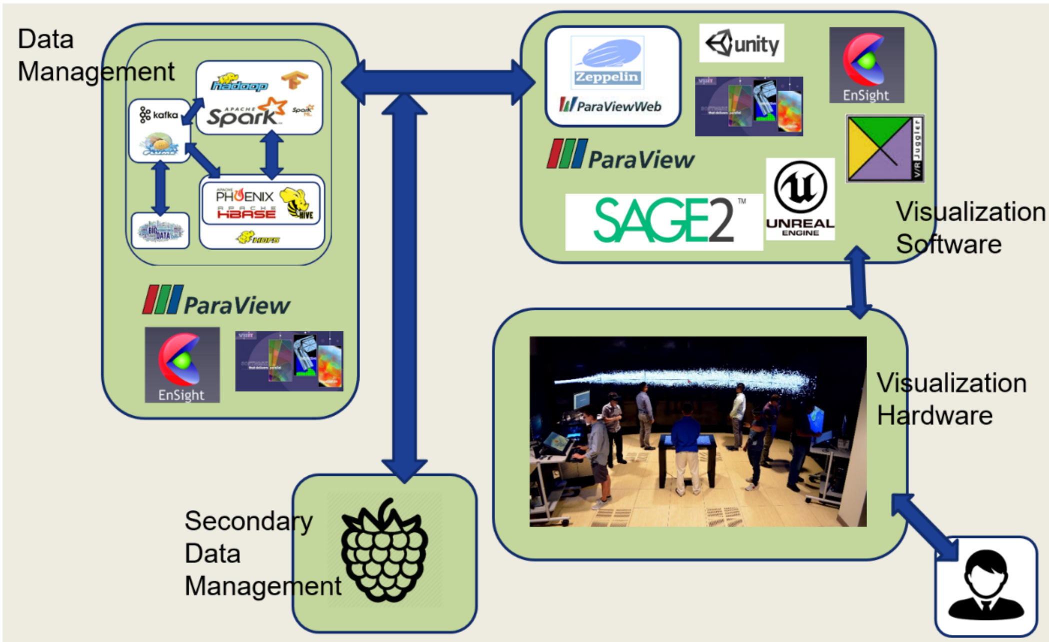

Fig. 1 **Reconfigurable VCA diagram showing the user- and data-centric ecology using heterogeneous visualization to support visual analytics**

2. Visualization Software Suite

This section describes various data visualization software we developed to enable a reconfigurable VCA.

2.1 ParaSAGE: Large High-Resolution Display Scientific Visualization

Figure 2 shows ParaSAGE,[4] a scientific visualization application running on top of the SAGE2 framework that enables high-resolution data visualization. ParaViewWeb was implemented as a SAGE2 application to take advantage of the abstraction provided by SAGE2 that allows use of multiple displays as one large multiuser workspace. The ParaSAGE implementation was tested on a 24-tiled display in a 6 × 4 curved configuration with a total of 49.76 MP available for display. For visualization computing hardware, a visualization cluster with a head node and three visualization nodes with eight outputs each are being used to drive the tiled display system. A simulation data[5] set with 220 million tetrahedrons were tested successfully with ParaSAGE. ParaSAGE application provides a viable high-resolution data visualization alternative to the desktop-based visualization to the scientific community.

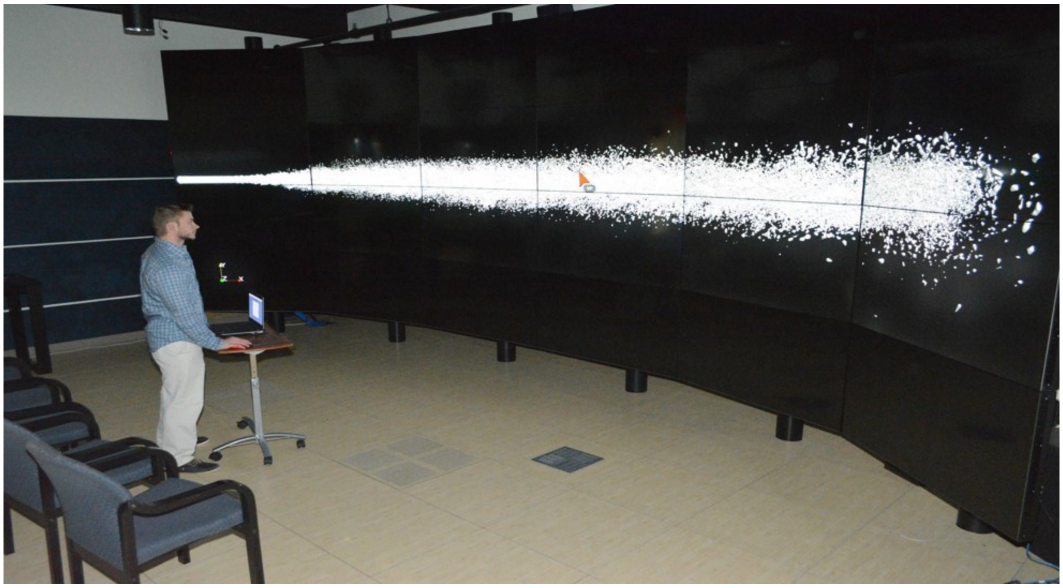

Fig. 2 ParaSAGE: scientific visualization on an LHRD

2.2 ParaView on zSpace: Hybrid 2- and 3-D Scientific Visualization

Figure 3 shows a hybrid 2- and 3-D scientific visualization workflow, which allows users to add 3-D immersive visualization capability to their day-to-day desktop scientific visualization workflows.[6] A 3-D immersive and interactive visualization capability was added to the widely used visualization tool ParaView running on the zSpace semi-immersive virtual reality display system using ray tracing for rendering at an interactive rate. The zSpace display system supports head-tracked stereoscopic display and stylus-based 3-D interactive interaction. Further, the zSpace virtual reality system requires very little calibration or maintenance after its initial system driver setup and configuration. For software, we extended an existing ParaView plugin, the pvOSPRay renderer plugin, to work with the Immersive ParaView plugin to support 3-D immersive and interactive visualization. We tested the workflow presented using a data set with 220 million tetrahedrons[5] at an interactive rate. We also evaluated this visualization workflow in a pilot user study comparing 2-D visualization and 3-D immersive and interactive visualization using a scientific simulation data set common to our users. We collected performance data and user feedback via a usability questionnaire. The 3-D immersive and interactive visualization workflow was found to be preferred over the 2-D visualization workflow, together with some very constructive user feedback on the 3-D immersive and interactive visualization system.

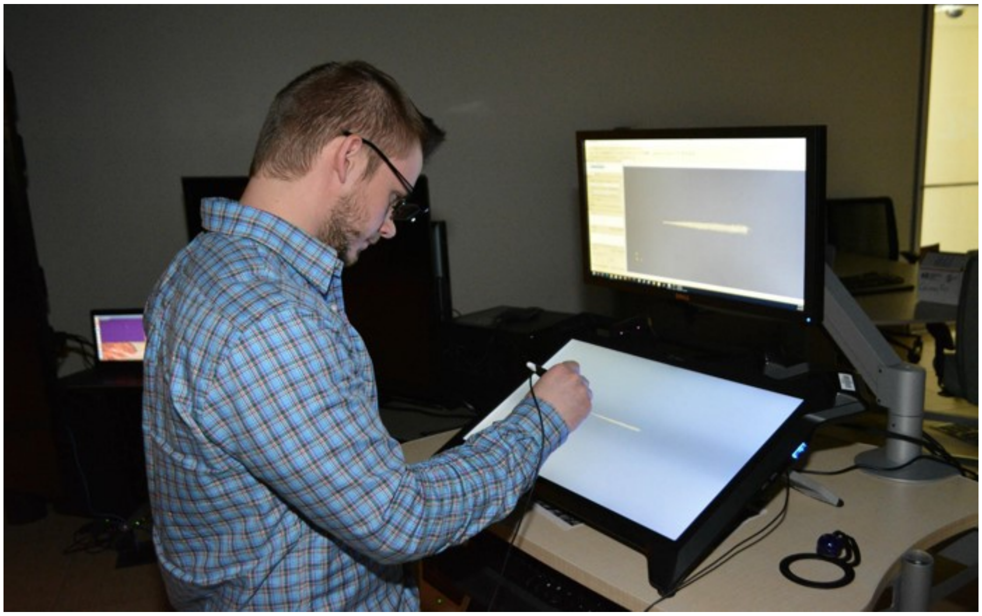

Fig. 3 ParaView running on zSpace showing hybrid 2- and 3-D data visualization on a semi-immersive virtual reality system

2.3 ParaView on Oculus Rift: Complete Immersive Scientific Visualization

Figure 4 shows a complete immersive scientific visualization workflow, which provides the users with the ability to be completely immersed in their scientific data. Similar to using the zSpace, HTC Vive and ParaView's OpenVR plugin[7] added immersive visualization to the widely used visualization tool ParaView. The HTC Vive head-mounted virtual reality system supports head-tracked stereoscopic display and two-handed 3-D interactive interaction. Further, the setup requires very little calibration or maintenance after the initial system driver setup and configuration. For software, we used ParaView's OpenVR plugin developed by Kitware to extend ParaView to support 3-D immersive and interactive visualization. We tested the workflow presented using a data set with 220 million tetrahedrons at an interactive rate.

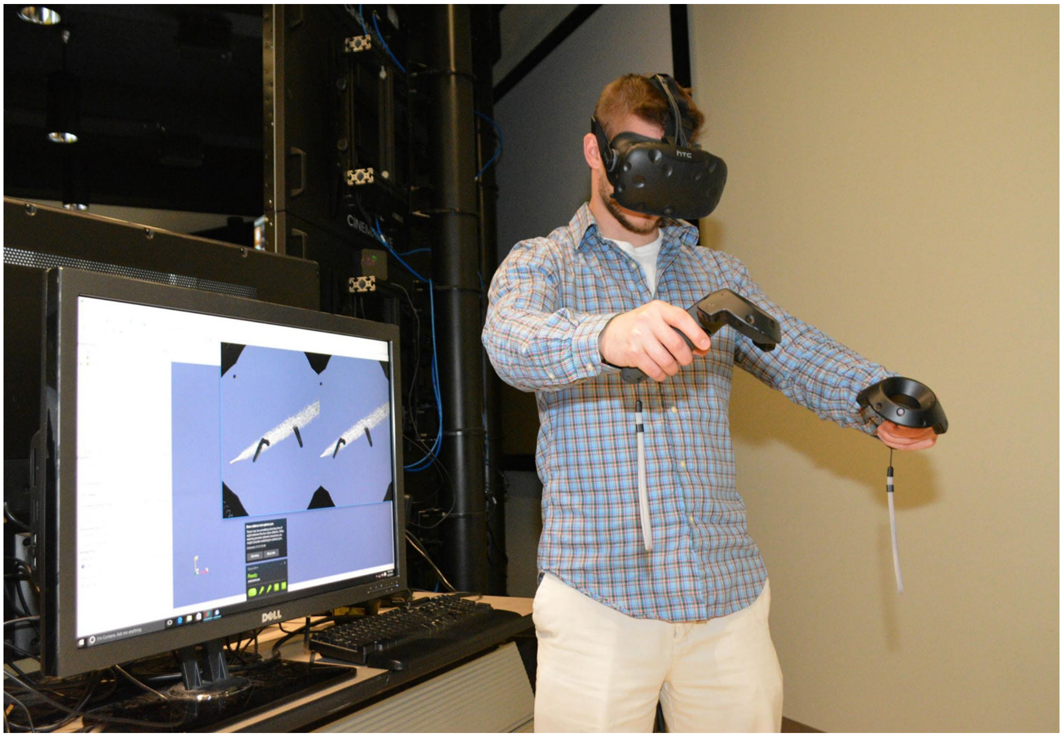

Fig. 4 ParaView running on HTC Vive showing complete immersive data visualization on a head-mounted virtual reality system

2.4 Hybrid 2- and 3-D Visual Analytics on Large High-Resolution Display and Immersive Environment

Figure 5 shows a data-flow-oriented scalable and extensible visualization system for supporting hybrid 2- and 3-D visual analytics. Our application[8] allows users to visually analyze the results of a complex, multivariate Monte Carlo simulation. The simulation outputs variables describing various properties of 3-D objects interacting in a dynamic 3-D environment. Our system uses 2-D charting tools to visualize the statistical relationships among simulation variables. We developed a Unity3D application to animate the 3-D simulation in a virtual environment showing the time-varying results of the dynamics in a 3-D environment. The Unity3D application runs on both a 2-D high-resolution display system and a completely immersive, head-mounted display device. The 2-D visualization framework running on our LHRD system supports multiple coordinated views across all the 2- and 3-D visualization components. Preliminary result shows our data-centric design has resulted in a user-centric visualization tool that can greatly enhance the analytical process and speed up the derivation of insights from data.

Fig. 5 Interactive hybrid 2- and 3-D visual analytics on an LHRD and immersive environment

2.5 SyncVis: ParaViewWeb-based Big Data Visual Analytics

As illustrated in the Data Management bubble in Fig. 1, we used various open-source project to ingest, store, and manipulate our data. Figure 6 shows our work on an integrated data querying, processing, and visualization system. Zeppelin is a web application that allows flexible interactive data analysis through its notebook-based user interface. Analysts can separate their code into distinct paragraphs, and within each, use various languages such as Scala, R, Python, and SQL, along with their respective ecosystem of existing libraries, to accomplish a given processing and analysis step. From the data, while there are visualization capabilities built into Zeppelin, there are some limitations in scalability in this particular arena. SyncVis is a data-visualization web application, whose strength lies in visualizing multiple coordinated views over one particular data set, potentially a very large one. It allows us to do specific detailed analysis using visual filters as part of a closed-loop, visual analytic solution. However, it cannot match Zeppelin's programmatic flexibility in tasks during a detailed analysis process of aggregating or further data mining. Further, the data to be visualized must be prepared and formatted correctly and then manually loaded into SyncVis. Analysts often want to iterate among data munging, data mining, and visualization as part of the overall visual analytics process. Therefore, there is a need to integrate data processing and visualization tools. We developed a solution that can fully exploit the respective strengths of Zeppelin and SyncVis, which allows data that have been mined within Zeppelin to be visualized seamlessly within SyncVis.

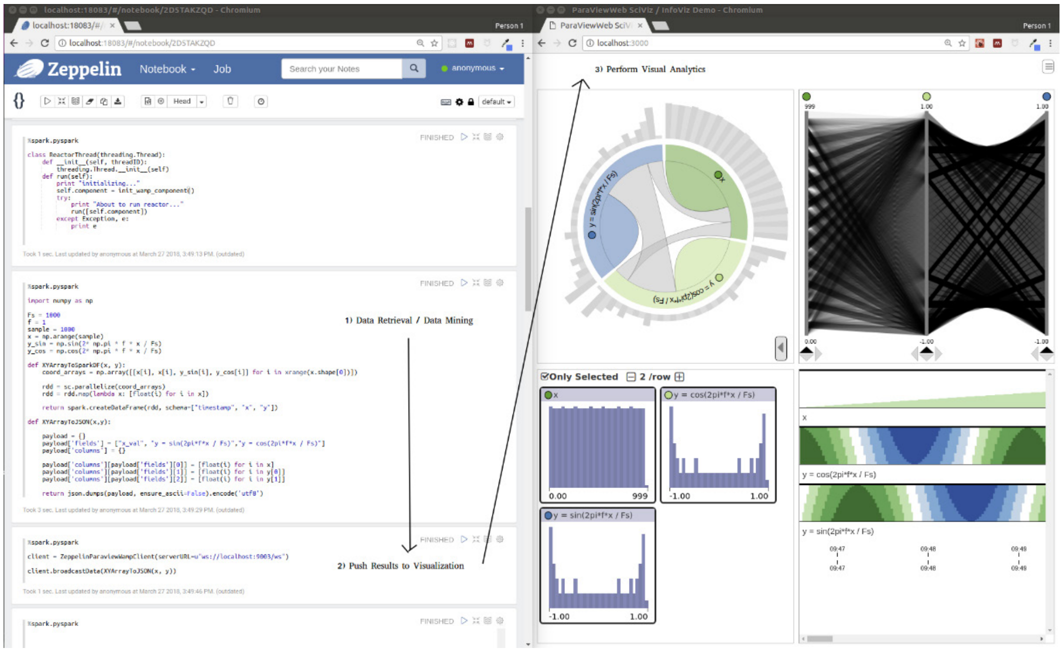

Fig. 6 SyncVis visualization of Zeppelin

On ParaView server side, the ParaView server utilizes the Autobahn Python library and the twisted Web Application Messaging Protocol (WAMP) server implementation. This allows the server to communicate to other connected devices via the publish/subscribe model, as well as the ability for any connected device to remotely call procedures (remote procedure call [RPC]) on another remote device. These are both features we need so that SyncVis and Zeppelin and interoperate in a reactive manner.

The ParaView server registers a RPC referenced by the string "pvserver.datatable.set" that other connected devices can call to update the data set being visualized. As mentioned previously, the ParaView server registers an RPC that will, given deserialized data coming from any connected device, update the visualization data in the server with the new data, reinitialize any metadata and parameters, and recompute any postprocessed data necessary that the data visualizations need. A WAMP topic referenced by the string "pvserver.data.ready" is published after setting these variables. It is each connected client's responsibility to subscribe to this topic so that it can be notified when the visualization data have changed.

On Zeppelin Python Module, we design a Python module that can be dynamically loaded up within a PySpark Zeppelin paragraph. This module is designed for the purpose of sending new data to a WAMP server to which SyncVis or other visualization clients can also connect. In particular, the module implements a ZeppelinParaviewServerWAMPConnection singleton class. It is responsible for connecting to the ParaView server, given a websocket URL. The run component method will take in a spark data frame or SQL query string, run the query and/or collect the results from spark memory, convert the resulting data frame to JSON format, and call the RPC on the ParaView server to pass the data set over. The Autobahn library takes care of the data serialization/deserialization on the ParaView server.

2.6 Data Visualization on Microsoft HoloLens: Augmented Reality Scientific Visualization

Figure 7 shows an augmented reality scientific visualization workflow, which provides the users with the ability to visualize their polygonal data using Microsoft HoloLens. Similar to using other virtual reality display system, this augmented reality setup also allows the user to use hand gestures and voice commands to interact with their 3-D simulation data in 3-D. The Microsoft HoloLens mixed reality system is the first fully self-contained holographic computer running Windows 10 that allows the user to place holograms in their physical environment. For software, we used Unity to develop the visualization applications and deploy them on the Microsoft HoloLens. For the 3-D model shown in Fig. 7, we had to reduce the polygon count of the 3-D model to run the application on Microsoft HoloLens at an interactive rate.

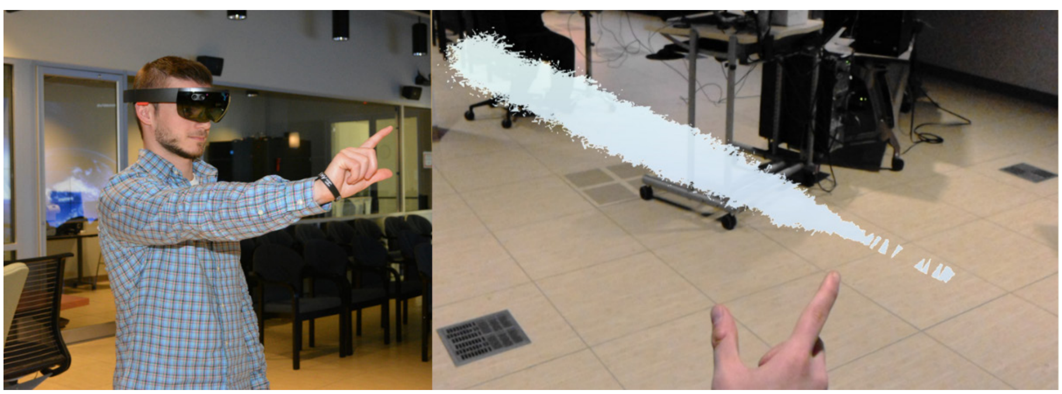

Fig. 7 Unity-based augmented reality application running on Microsoft HoloLens showing augmented reality data visualization

3. Visualization Systems

As shown in Fig. 8, our reconfigurable VCA supports various immersive and non-immersive visualization systems. It ranges from the tiled LHRD, as shown in Fig. 6, to multitouch displays, to personal workstations, to tablets, to head-mounted, completely immersive virtual reality systems (Oculus Rift and HTC Vive), to semi-immersive virtual reality systems (zSpace), to augmented reality display systems (Microsoft HoloLens), to the users themselves.

Fig. 8 Visualization ecology showing heterogeneous visualization systems supporting the reconfigurable VCA

4. Discussion

We are applying this reconfigurable VCA to Army test data for automotive systems, communications systems, and logistics. Automotive and communications system tests produce 10s of terabytes of data for a single system. ParaViewWeb allows building an automated workflow, from data ingestion to visualization and interrogation of the data. Our reconfigurable VCA is still a work in progress with the eventual goal to build a workflow for customer applications and hand off the tool to the customer while we continue development of improved capabilities and expand the tools to encompass additional customer requirements. Having T&E as part of the requirements and development team provides a natural two-way communication for US Army Research Laboratory to understand user requirements, for the customer to understand the visualization capability, and to naturally transition the capability to the customer already trained in its use. The ongoing relationship facilitates sustaining the capability with bug fixes, updates, new features, and ports to new platforms.

5. Conclusion

We do not believe one size fits all in visualization. Our reconfigurable VCA gives the user the flexibility to display the data using the software and hardware best suited for the specific analysis and exploration task. Depending on the computational requirements, the user can tap into HPC resources to shorten the processing time required. As interactive visualization requires near-real-time processing of the data, we further improve the overall performance of the ecology using middleware like Cloudberry.[9] Moving forward, we plan to streamline the process by making available additional visualization techniques, automating the data ingest step, expanding the use on additional visualization devices, and building automated workflows for our customer applications.

6. References

1. Barton JM, Namburu R. Data-intensive computing for test and evaluation. ITEA J. 2017 June;38(2):144–152.

2. Thomas JJ, Cook KA. Illuminating the path: the research and development agenda for visual analytics. Richland (WA): National Visualization and Analytics Center; 2005.

3. Andrews C, Endert A, Yost B, North C. Information visualization on large, high-resolution displays: Issues, challenges, and opportunities. Info Vis. 2011;10(4):341–355.

4. Kobayashi D, Su S, Bravo L, Leigh J, Shires D. ParaSAGE: Scalable web-based scientific visualization for ultra resolution display environment. IEEE VisWeek. 2016.

5. Bravo L, Kim D, Ham F, Powell C, Kastengren A. Effects of fuel viscosity on the primary breakup dynamics of a high-speed liquid jet with comparison to X-ray radiography. Proc Combustion Institute. 2018. issn: 1540-7489. doi: https://doi.org/10.1016/j.proci.2018.05.064.

6. Su S, Bravo L. Hybrid 2-D and 3-D immersive and interactive user interface for scientific data visualization. Aberdeen Proving Ground (MD): Army Research Laboratory (US); 2018 Jan. Report No.: ARL-TR-8113.

7. Martin K, DeMarie D, Jhaveri S, Ayachit U. Taking ParaView into virtual reality. In: KitWare blog. Clifton Park (NY): KitWare; 2016 Sep 22 [accessed 2018 Aug 1]. https://blog.kitware.com/taking-paraview-into-virtual-reality/.

8. Su S, Perry V, An M, Chen M. Real-time interactive hybrid 2D and 3D visual analytics on large high resolution display and immersive virtual environment. SEARIS 2018 11th Workshop on Software Engineering and Architectures for Realtime Interactive Systems; IEEE Virtual Reality; 2018, accepted for publication.

9. Jia J, Li C, Zhang X, Li C, Carey M J, Su S. Towards interactive analytics and visualization on one billion tweets. In: Proceedings of the 24th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems, GIS '16; 2016; New York, NY. ACM; c2016. 85:1–85:4.

List of Symbols, Abbreviations, and Acronyms

2-D	2-dimensional
3-D	3-dimensional
LHRD	large higher-resolution display
RPC	remote procedure call
SQL	Structured Query Language
T&E	test and evaluation
URL	Uniform Resource Locator
VCA	Visual Computing Architecture
WAMP	Web Application Messaging Protocol

1	DEFENSE TECHNICAL
(PDF)	INFORMATION CTR
	DTIC OCA
2	DIR ARL
(PDF)	IMAL HRA
	RECORDS MGMT
	RDRL DCL
	TECH LIB
1	GOVT PRINTG OFC
(PDF)	A MALHOTRA
4	ARL
(PDF)	RDRL CIC
	S SU
	RDRL CIH S
	B PANNETON
	V DASARI
	RDRL VTP
	L BRAVO