A Modular Approach to State-of-the-Art Big Data Visualization

Keith A. Maki

Abstract

In our technical demo, we will present a novel approach to visibly large data visualization using a modular segmented display design. We demonstrate that this approach is both large and displays data, and we compare to more established methods of big data display, finding that the demonstrated approach can display very big data indeed.

Introduction

Recently, there has been a stir in the computational research community over the need for research conducted on big data. This is an incredibly promising area, and much work can surely be done provided the data is big enough. However, most computer scientists work at desks with a small number of monitors that are only so big, which severely limits the size of the data they can work with. We believe that to empower the field of big data science new technologies must be developed to meet the demand for ever-increasing sizes of data. Our technical demonstration will provide a novel application of common indicator technologies in a modular configuration highly amenable to scalable visualization of really big data. We compare the presented design to several alternative approaches of varying degrees of conventionality and we demonstrate its ability to visualize both really big numbers and really big text in a variety of colors.

Background and Related Work

The needs of the big data research community have rarely been the focus of computer engineers developing computing technologies, which have long been optimized for computational capabilities, power efficiency, portability, and cost. Ironically, as electronics capabilities have improved, data storage has become more and more limiting from a big data standpoint, such that in order to store data on a standard 3.5in hard drive, data is physically constrained to have a surface area no larger than the largest planar cross section of the case volume, which is not very large. Gone are the days when magnetic tape reels the size of hubcaps stored data, although to be fair in order to store data larger than the tape width a method such as striping had to be used, and putting the data back together for any practical purposes had in our experience always resulted in a tangled mess.

To overcome these limitations, data is generally analyzed with the help of a visualizer, such as a computer monitor, that attempts to compensate for the smallness of the data by biggifying it so it can be worked with as though it were big data. However, these technologies are limited in their capabilities and can only compensate so much, leading to problematic results such as that shown in Figure 3.

Nevertheless, a number of technologies have attempted to provide solutions to these challenges. For example, the electronic scoreboard is perhaps the most familiar and well-known example of a big data visualizer. These displays are typically implemented using incandescent or LED technology in either segmented, eggcrate, or dot matrix format, although flip-disc and vane display technologies have also been used. These displays can display very big data but are often bulky and impractical for transporting, and are capable of visualizing only a fixed number, format, and color of numeric digits, which limits their usefulness in many big data applications, particularly text.

Similarly, projection systems have also found their way into standard use, and these overcome the limitations of scoreboards on data format, color, and even size, being able to project colored numeric and text data at sizes proportional to the distance between the projector and the visualization surface. However, the data visualizations produced by these systems suffer from a noticeable dimming in brightness inversely proportional to their area. Additionally, while very large, crisp projection systems, such as IMAX[1], exist which are capable of visualizing big data with thousands of digits, the scalability in terms of the number of digits is limited by the physical dimensions of the projection surface and the optical resolution properties of the projector, and there will be a point at which the data is either too small or too indistinct to reliably work with. Additionally, such research-grade big data projection systems are generally not portable, which further detracts from their utility.

The presented system will attempt to address all of these shortcomings, providing a portable, scalable design using modular display panels. Our work is most similar to that of Overpass Light Brigade[2], who make use of LED signboards to display the individual characters that make up a piece

[1] http://www.imax.com/

[2] http://overpasslightbrigade.org/

of data. However, where OLB uses monochrome LEDs in hardwired letter configurations, our modular display panels are both programmable and wicked cool, allowing for dynamic visualization of both numeric and text data in variable color.

Proposed Design

We propose the use of LED indicators positioned in a segmented display pattern of the desired size. The LED layout for a single panel is shown in Figure 1. This configuration was chosen to allow for compatibility with standard 16-segment display layout alphanumeric characters, but the design is flexible enough to allow for similar but modified layouts. For our prototype system, we make use of WS2812B RGB LEDs, which have an integrated shift register to manage PWM control for each pixel separately. We wire the LEDs such that each panel has the same configuration, with a data in port on one side and a data out port on the other. In this way, color programming information can be shifted in down a chain of several panels. For the LED server node, we use a BeagleBoneBlack running the LEDscape[3] framework, to handle the buffering of several chains of LEDs simultaneously; and we send frames to the server over an ethernet connection using the Open Pixel Control[4] communications protocol.

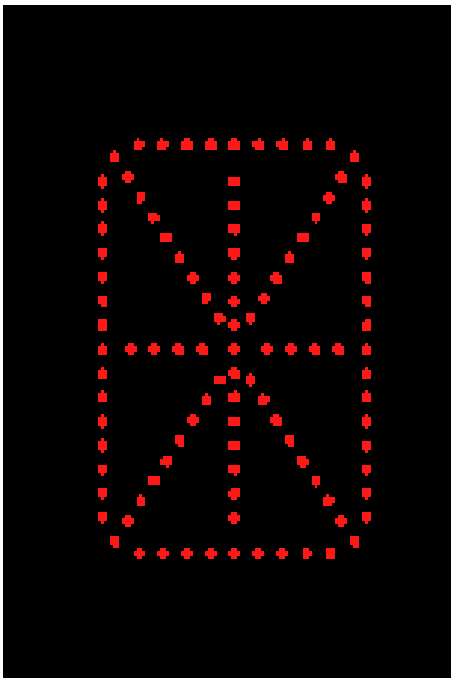

Figure 1: The layout of LEDs on each panel

[3] https://github.com/Yona-Appletree/LEDscape/

[4] http://openpixelcontrol.org/

Figure 2: Big data scientist conducting a visualization test

Experiments

We compared the proposed approach with a number of alternative technical solutions for big data visualization. Each approach was coded by two raters for four binary attributes:

1. **Big data** Whether the technology constitutes big data visualization.
2. **Portable** Whether it's feasible to transport the technology on a semi-regular basis.
3. **Scalable** Whether the technology scales for big data with an arbitrary number of characters.
4. **Nerdy factor** Whether people will know you're a real nerd if you say you visualize things with the technology.

The inter-rater agreement was strong (Cohen's $\kappa = 0.85$). The results of our comparisons are shown in Table 1.

We have manufactured eight prototype panels, controlled using four panels on each of two channels on one node to demonstrate the potential of our system, but the prototype setup is scalable to five panels on each of 48 channels on each of at least 16381 nodes, for a total of altogether too many panels[5]. With each digit measuring at nearly two feet tall, our prototype panels definitely constitute "big", and we've had a lot of fun using them to visualize data in several colors and combinations. An image of a big data scientist conducting a data visualization test using the prototype system is shown in Figure 2.

Conclusion and Future Work

In this report, we have described the capabilities of our novel big data visualization technology. We have explored the technical advantages of our system when compared to other data visualization technologies with respect to state-of-the-art big data research desiderata. Future work will explore more freely configurable pixel layouts such as modular

[5] We did the math.

Technology	Big data	Portable	Scalable	Nerdy factor	Notes
Beach	Yes	No	Yes	No	Not programmable
Calculator	No	Yes	No	Yes	Not a big data solution
IMAX	Yes	No	No	No	I doubt your lab has space for one
Projector	Yes	Yes	No	Yes	Brightness varies with size of data
Scoreboard	Yes	No	No	Yes	Limited data format
Ulexite	No	Yes	No	Yes	Only a rocky solution
Our system	**Yes**	**Yes**	**Yes**	**Yes**	Fails insignificance test ($p = i$)

Table 1: A comparison of the proposed approach with various data visualization technologies

dot matrix designs, allowing for flexible data size stretching multiple panels in any direction. Additionally, it would be interesting to explore the use of drone swarm technology to visualize big data in higher dimensions, but we definitely would need grant money for that.

Acknowledgements

Special thanks to Bob Rudenborg and Sam Selders for all their help with the soldering.

Figure 3: A piece of data which likely exceeds the physical limits of your display technology