Effects of a time-varying function on multi-dimensional data understanding in VR immersive analytics

Tetsuro Kamiyama[1,*] and Mie Sato[2,*]

[1] *Graduate School of Regional Development and Creativity, Utsunomiya University, 7-1-2, Yoto, Utsunomiya Shi, Tochigi Ken, 321-8585, Japan*

[2] *School of Data Science and Management, Utsunomiya University, 350, Minemachi, Utsunomiya Shi, Tochigi Ken, 321-8505, Japan*

Abstract

In recent years, technology using immersive VR environments has developed rapidly, and significant progress has been made in the field of information visualization, which is referred to as immersive analytics. In this study, we proposed an immersive framework that improves ImAxes, one of the immersive analytics, to visualize multiple data on a single axis that changes in response to user operations. The results of having participants answer time-varying datasets and corresponding questions show that the ImAxes with adding a time-varying function is more useful for better understanding time-varying data in a shorter time and fewer axes than original ImAxes.

Keywords

Immersive Analytics, VR, Data Science

1. Introduction

In recent years, technologies using immersive VR environments have developed rapidly and significant progress has been made in the field of information visualization. The field in which multiple disciplines such as visualization, immersive environments, and human-computer interaction converge to support human data analysis using new technologies is called immersive analytics [9]. This approach goes beyond traditional 2D flat data visualization to enable manipulation and understanding of data in three or more-dimensional space.

An example of a traditional 2D flat data visualization is a tree-map. It is a representation designed for human visualization of complex traditional tree structures, showing arbitrary trees in a 2D space-filling representation. However, this approach suffers from screen space limitations and readability issues when many nodes are displayed simultaneously. This prevents the user from grasping the whole picture [1].

Immersive analysis is a new framework for interactive and intuitive data analysis, and several software toolkits for immersive analysis have recently been released. Kraus et al. [2] found that multidimensional clusters are easier to identify in VR than in 2D desktop displays. Also, Yang et al. [3] found that 3D globes are more effective than 2D projections in conveying the distance and direction of a world map. These studies suggest that immersive displays offer tangible advantages when the data is inherently more than two-dimensional. Another study on immersive map visualization by Yang et al. found that immersive environments allow seamless transitions between 2D views that are optimal for visualizing different aspects of the data [4]. Kwon et al. [5] also found that immersive VR graph layouts enable faster decision making and fewer errors compared to 2D graph layouts.

As part of interaction technologies, immersive analytics emphasizes the possibility of embodied direct manipulation. One of the interaction technologies that can be embodied and directly manipulated is ImAxes [6]. ImAxes employs a method of constructing visualizations through direct manipulation of the 3D axes of a data dimension in 3D space: axes generated from the dataset are placed in the VR space, and their manipulation by the user produces arbitrary visualization results in the VR space. For example, a parallel coordinate plot (PCP) is created by combining any two axes in parallel, and a scatter plot is created by combining them at right angles. With ImAxes, users can try a wide variety of visualization methods with a single tool, such as 2D visualization, 3D visualization, and scatter plots, depending on how you place the axes.

However, ImAxes has some problems. ImAxes can store only one set of data per axis. For example, suppose the user wants to visualize precipitation data for January, February, and March, which changes from month to month. In this case, there are three axes in the VR space named "precipitation in January," "precipitation in February," and "precipitation in March. By combining these axes in parallel, the user can visualize the PCP and visualize changes in precipitation over time. Since the data is "only" precipitation and "only" for three months,

APMAR'24: The 16th Asia-Pacific Workshop on Mixed and Augmented Reality, Nov. 29-30, 2024, Kyoto, Japan
* Corresponding author.
mc236542@s.utsunomiya-u.ac.jp (T. Kamiyama);
mie@is.utsunomiya-u.ac.jp (M . Sato);
0009-0005-7312-1641 (T. Kamiyama);
0000-0002-2190-4629 (M. Sato);

there is no problem with as few as three axes. However, if we add not only precipitation but also temperature data, or data for April, May, and June, the number of axes displayed on the VR space will increase because of an increase in the variety of data. Therefore, it takes time to find the axis of the desired data, which may spoil the immersive experience.

This increases the number of times the axis is moved, causing fatigue and possibly inducing a "gorilla arm" effect [7]. In Immersive analytics, there is a need to reduce fatigue within the immersive environment. Some research has been conducted to develop devices to reduce fatigue problems caused by 3D interaction in the air using a VR controller [8].

This study proposes an immersive framework that improves ImAxes and visualizes multiple data for a single axis, which changes in response to user manipulation. This will improve the efficiency and user comfort of data visualization in an immersive environment and contribute to reducing fatigue. We asked participants to view time-varying data visualizations and explored the advantages and disadvantages of time-varying data.

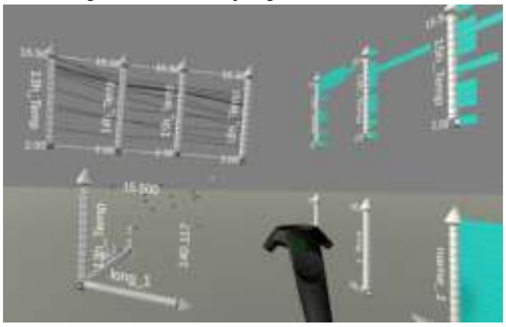

Figure 1: Image of the participant during the task

2. Design and implementation of time-varying axes

The design of this system is intended to complement the manipulation system that ImAxes lacks; one of the advantages of ImAxes is that users can freely move and position axes within the virtual environment, thus forming creative visualizations. Suppose, however, that the user wants to visualize data from multiple time periods. However, since one axis can only store data from one time period, when the user visualizes data from one time period and then wants to visualize data from the next time period, user needs to have an axis that contains that data. This would require more manipulation and could potentially compromise the immersive experience.

Therefore, the purpose of this research is to add a function that stores multiple data on a single axis and switches the type of data output based on user input. This function allows visualization of data that continuously changes over time and is expected to deepen the user's understanding of the data.

In addition, the color of the data visualization was also changed. In the original ImAxes, the data can be freely colored according to the size of the data. This allows the user to confirm the size of the data by looking at the color. In this study, we wanted to investigate the effect of a time-varying function on data comprehension, so all data in scatter plots and parallel coordinate plots were colored black. This allows participants to distinguish between large and small data only by the position of the data, and not by the color. The color of all values to fix the maximum and minimum values of the data was white. When a 2D or 3D scatter plot is created with the ImAxes with adding a time-varying function, a white point is placed on the opposite side of the origin. This point represents the maximum value of the time-varying data, which helps us to understand the data by looking at the scatterplot.

3. Evaluation through visualization of time-varying data

In this study, a participant experiment was conducted to explore the advantages and disadvantages of the original ImAxes and the ImAxes with adding a time-varying function: one question was prepared for each dataset, and the time to answer the question and the number of axes used to answer the question were counted.

3.1. Experiment

This system was developed using the game engine Unity and the VR device HTC VIVE, which has a head-mounted display and two controllers that allow flexible viewpoint control and interactive label axis operation simultaneously.

The left-hand-operated device for controlling time changes has a knob; time moves forward one step when turned to the left and backward one step when turned to the right. Figure 2 shoes an image of participant during the task. Figure 3 shows the left-hand-operated device and a VIVE controller used in the experiment. Since a touchpad on the VIVE controller already had this function, we prepared a left-hand control device. Turning the knob clockwise would advance time, and vice versa, backward.

Figure 2: Image of the participant during the task

Figure 3: The left-hand-operated device and a VIVE controller used in the experiment

3.2. Data to be used in the experiment

In the first dataset, the axes consist of “temp”, “lat”, and “long”. “temp” means template and stores the temperature at each location. “lat” means latitude and “long” means longitude. The interval is set to August 24~30, 2024. Latitude and longitude indicate the latitude and longitude of AMeDAS stations [10] that measure precipitation in Tochigi Prefecture, respectively. Precipitation indicates the amount of precipitation observed at a particular AMeDAS station on that day.

In the second dataset, the axes consist of “wind”, “lat”, and “long”. “wind” means wind speed and stores the temperature at each location. The interval is August 14~20, 2023. Latitude and longitude indicate the latitude and longitude of AMeDAS stations that measure precipitation in Tochigi Prefecture, respectively. Precipitation indicates the amount of precipitation observed at a particular AMeDAS station on that day.

The third dataset has axes named “temp”, “lat”, and “long”, and varies with time. The interval is from 11:00 to 17:00 on July 19, 2014. Latitude and longitude indicate the latitude and longitude of the station measuring precipitation in Tochigi Prefecture, respectively. The precipitation amounts represent the amount of precipitation observed at a particular AMeDAS station in one hour.

In the fourth dataset, the axis names are “wind,” “lat,” and “long” and vary with time. The interval is from 5:00 to 11:00 on August 9, 2013. Latitude and longitude indicate the latitude and longitude of the precipitation measuring stations in Tochigi Prefecture, respectively. Precipitation represents the average wind speed observed at a particular AMeDAS station in one hour.

For each dataset, a question was asked. For the first dataset, the question was “On which day is the station with the highest precipitation observed? For the first dataset, we asked the question, “On which days are the stations with the highest precipitation? For the second dataset, we asked the question, “On which day is the station with the highest wind speed observed? For the third dataset, we asked the question, “From which direction did the rain clouds move? For the fourth dataset, “How did the wind speed change over time for Tochigi as a whole?

The reasons for choosing this dataset/question are as follows. The first and second datasets/questions needed to compare data from distant dates. In this case, we assumed that the ImAxes with adding a time-varying function, which cannot be compared side-by-side, are not suitable for this problem. The third and fourth datasets/questions were chosen because of the need to understand the flow of time-varying data. In this case, we assumed that the ImAxes with adding a time-varying function is suited for this problem because it can visualize time-varying data continuously.

The daily precipitation (before and after improvement) and hourly precipitation (before and after improvement) were randomly displayed, and questions were asked. The following questions were then answered on a 7-point Likert scale.

2. Was the axis operation intuitive?
3. Was there discomfort during use?
4. (Only original ImAxes only) Would you like the time-varying function?
5. (Only ImAxes with adding a time-varying function) Did the added functionality help me to better understand the data?

After four trials, participants were asked to visualize ImAxes with adding a time-varying function that included data such as latitude and longitude, precipitation, temperature, and wind speed that changed over successive time intervals of 15 minutes, and to see if they could find relationships between the data.

As a post-experiment questionnaire, participants were also asked about their impressions and understanding of the data before and after the improvement.

4. Results

Nineteen male participants and one female participant tool part in the experiment. The mean age was 22.95 years, with a standard deviation of 1.07. All 20 participants had previous experience with VR.

One question was set for each of the four datasets, and each participant was asked to answer either Original ImAxes or ImAxes with adding a time-varying function in a random combination. For the first, second, and fourth datasets, the percentage of correct answers was 100% for both the original ImAxes and the ImAxes with adding a time-varying function. For the third dataset, however, correct answers were 60 % for the original ImAxes and 90 % for the ImAxes with adding a time-varying function.

Data obtained from the experiment were analyzed by considering condition as a within- participants factor. Friedman tests were performed on each data set, and Wilcoxon signed rank sum tests were performed when significant differences were found.

A box-and-whisker diagram of the time to answer a question is shown in Figure 4. Across the entire dataset, the time to answer a question was significantly less with the ImAxes with adding a time-varying function than with the original ImAxes ($p < 0.05$).

A box-and-whisker diagram of the time to answer a question is shown in Figure 5. The number of axes used to answer the question was significantly less in the ImAxes with adding a time-varying function than in the original ImAxes ($p < 0.05$).

A box-and-whisker diagram of answers of question 2 is shown in Figure 6. The ImAxes with adding a time-varying function was significantly more intuitive to use than the original ImAxes ($p < 0.05$).

A box-and-whisker diagram of answers of question 3 is shown in Figure 7. The ImAxes with adding a time-varying function caused significantly less discomfort during use than the original ImAxes ($p< 0.05$).

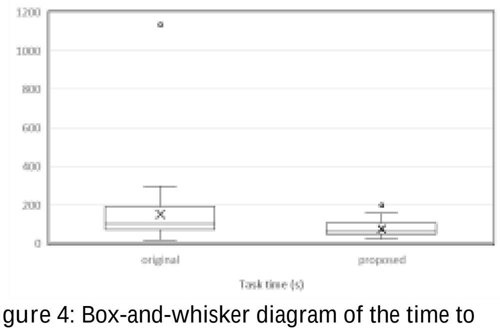

Figure 4: Box-and-whisker diagram of the time to answer a question

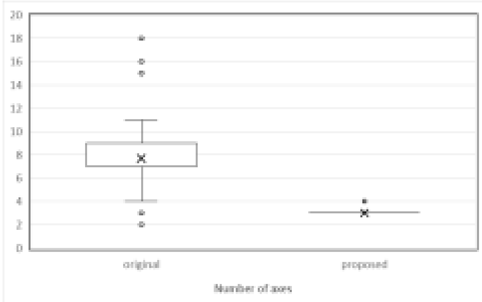

Figure 5: Box-and-whisker diagram of the number of axes to answer a question

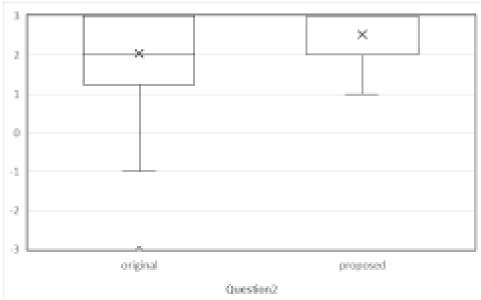

Figure 6: Box-and-whisker diagram of the answer to Question 2

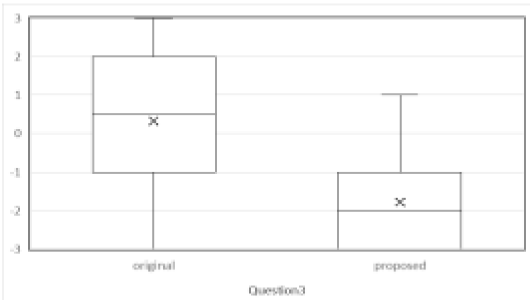

Figure 7: Box-and-whisker diagram of the answer to Question 3

Next, we investigate the difference in results between the time-varying dataset per day and the time-varying dataset per hour.

A box-and-whisker diagram of the time to answer the question with different datasets is shown in Figure 8. In Both the time-varying datasets per day and the time-varying datasets per hour, the time to answer questions was significantly shorter for the ImAxes with adding a time-varying function than for the original ImAxes and ($p < 0.05$).

A box-and-whisker diagram of the number of axes to answer with different datasets is shown in Figure 9. In both the time-varying datasets per day and the time-varying datasets per hour, the number of axes used before answering a question was significantly lower for the ImAxes with adding a time-varying function than for the original ImAxes ($p < 0.05$).

A box-and-whisker diagram of the answer question 2 with different datasets is shown in Figure 10. In both the time-varying dataset per day and the time-varying dataset per hour, then, there was no significant difference in intuitive use between the original ImAxes and ImAxes with adding a time-varying function ($p = 0.0545$, $p = 0.0975$).

A box-and-whisker diagram of the answer question 3 with different datasets is shown in Figure 11. In Both the time-varying dataset per day and the time-varying dataset per hour, the ImAxes with adding a time-varying function were significantly less uncomfortable to use than the original ImAxes ($p < 0.05$).

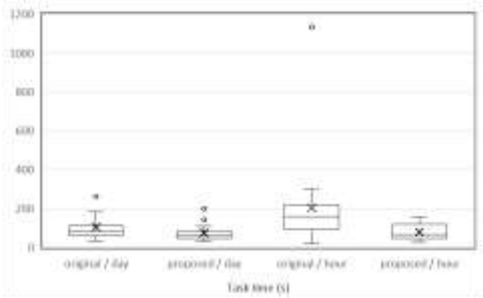

Figure 7: Box-and-whisker diagram of the time to answer the question with different datasets

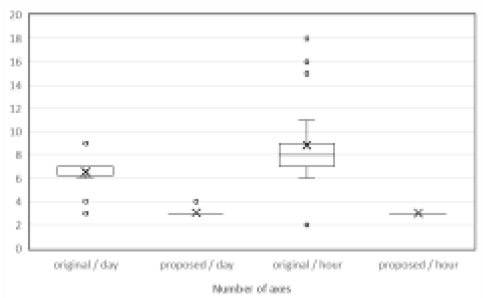

Figure 9: Box-and-whisker diagram of the number of axes to answer a question with different dataset

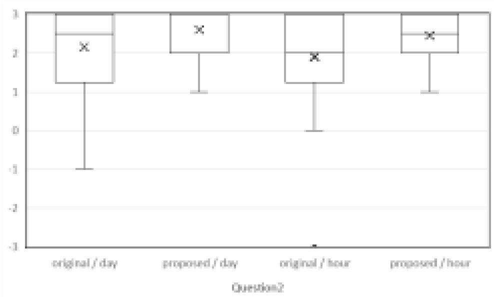

Figure 10: Box-and-whisker diagram of the answer to Question 2 with different dataset

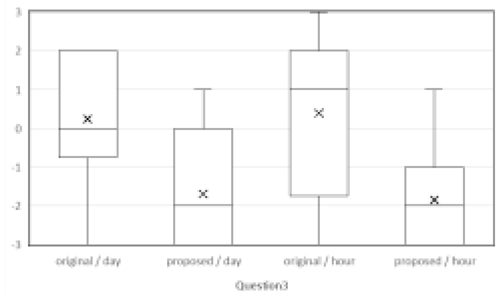

Figure 11: Box-and-whisker diagram of the answer to Question 3 with different dataset

A box-and-whisker diagram of the answer to question 4 with different datasets is shown in Figure 12. There was no significant difference between the time-varying dataset per day and the time-varying dataset per hour for participants who thought the time-varying function was necessary ($p = 0.097$).

A box-and-whisker diagram of the answer to question 5 with different datasets is shown in Figure 13. There was no significant difference between the time-varying dataset per day and the time-varying dataset per hour for participants who thought the time-varying function helped them better understand the data ($p = 0.517$).

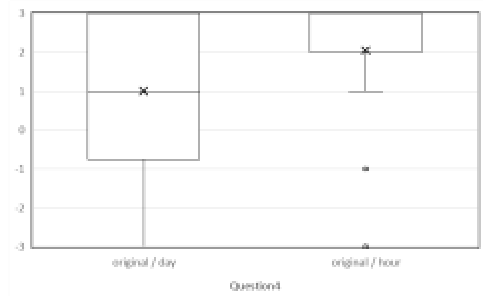

Figure 12: Box-and-whisker diagram of the answer to Question 4 with different dataset

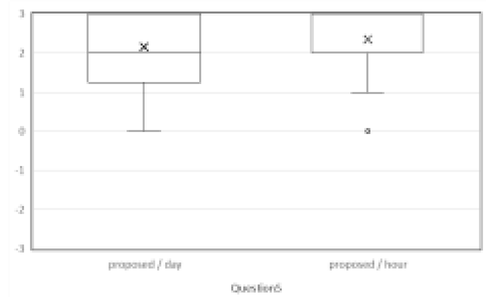

Figure 13: Box-and-whisker diagram of the answer to Question 5 with different dataset

5. Discussion

With respect to the percentage of correct answers, only the third datasets had incorrect answers. This is because the third dataset and question were more difficult than the others, since the correct answer could only be given if the participants correctly understood the locations of AMeDAS stations and movement in areas with high rainfall. In the case of the original ImAxes, participants had to remove the axis they were currently looking at and prepare a new axis if they wanted to look at the next time-varying axis, which we consider having lowered the percentage of correct answers. The fact that the percentage of correct answer for the questions changed before and after the addition of the time-varying function is evidence that the added function had an impact on the understanding of the data, and we would like to use this as a reference when creating future questions.

In the entire dataset and in the time-varying dataset per hour, the time to answer was shorter for the ImAxes with adding a time-varying function than for the original ImAxes. In the time-varying dataset per hour, there was difference in time to answer the questions, but the *p*-value was larger than for time-varying datasets per day. This indicates that the time-varying function is useful for solving all problems, and the ImAxes with adding a time-varying function is especially useful when one wants to see a continuous flow of time-varying data.

Participants used fewer axes with the ImAxes with adding a time-varying function than with the original ImAxes. This indicates that when participants wanted to see a 3D scatter plot of time-varying data, the ImAxes with adding a time-varying function allows them to combine three axes to create a 3D scatter plot and then turn a knob to see the time variation, thus maximizing their understanding of the data with a minimum number of axes.

Participants felt that the operation of the ImAxes with adding a time-varying function axes was more intuitive than that of the original Imaxes axes. This was an unexpected result. We expected that the operation feel of ImAxes would be constant regardless of whether the time-varying function was included, and question Q3 was asked to confirm whether the addition of the time-varying function had a negative impact on the operation feel of the ImAxes axes. However, in reality, many participants answered that the presence of the time-varying function made the operation of the axes intuitive. This is thought to be because the time-varying function caused participants to use only the minimum number of axes necessary, which improved the intuitiveness of axis operation.

Participants felt less discomfort with the ImAxes with adding a time-varying function than with the original ImAxes. This may be due to the fact that the knob facilitates visualization of time-varying data, which may have reduced discomfort.

Participants often desired the time-varying function in solving the original ImAxes problem. They wanted the time-varying function in answering the question of time-varying dataset per hour more. This indicates that the time-varying function is a feature that was desired by the participants.

6. Conclusion

This study proposed an immersive framework that improves ImAxes and visualizes multiple data for a single axis, which changes in response to user manipulation. To demonstrate the usefulness of the time-varying function, we showed participants the original ImAxes and the ImAxes with adding a time-varying function and asked them to visualize the data. The time-varying function was shown to be useful for understanding of time-varying data for short periods of time or for short intervals with a small number of axes.

However, the current mainstream in data visualization is software on 2D displays. In the future, we would like to compare the differences between the time-varying function and 2D displays and verify the usefulness of each.

References

[1] B. Shneiderman, University of Maryland, Tree Visualization with Tree-Maps: 2-d Space-Filling Approach, ACM Transactions on graphics (TOG), 1992.

[2] M. Kraus, N. Weiler, D. Oelke, J. Kehrer, D. A. Keim, and J. Fuchs, The Impact of Immersion on Cluster Identification Tasks. IEEE Transactions on Visualization and Computer Graphics 26, 1, 2020, 525–535.

[3] Y. Yang, B. Jenny, T. Dwyer, K. Marriott, H. Chen, and M. Cordeil, Maps and Globes in Virtual Reality, Computer Graphics Forum 37, 3, 2018, 427–438. J. Cohen (Ed.), Special issue: Digital Libraries, volume 39, 1996.

[4] Y. Yang, T. Dwyer, K. Marriott, B, Jenny, and S. Goodwin, Tilt map: Interactive transitions between Choropleth map, Prism map and Bar chart in immersive environments, IEEE Transactions on Visualization and Computer Graphics, 2020.

[5] O- H. Kwon, C. Muelder, K. Lee, and K.-L. Ma, A Study of Layout, Rendering, and Interaction Methods for Immersive Graph Visualization, IEEE Transactions on Visualization and Computer Graphics 22, 7, 2016, 1802– 1815.

[6] M. Cordeil, A. Cunningham, T. Dwyer, B. H. Thomas, and K. Marriott, ImAxes: Immersive axes as embodied affordances for interactive multivariate data visualization, UIST 2017 - Proceedings of the 30th Annual ACM Symposium on User Interface Software and Technology, 2017, 71–83.

[7] J. David Hincapié-Ramos, X. Guo, P. Moghadasian, and P. Irani, Consumed endurance: A metric to

quantify arm fatigue of mid-air interactions, In Conference on Human Factors in Computing Systems - Proceedings (CHI '14). Association for Computing Machinery, New York, NY, USA, 1063–1072.
[8] M. Cordeil, B. Bach, Y. Li, E. Wilson, and T. Dwyer, Design space for spatio-data coordination: Tangible interaction devices for immersive information visualization, IEEE Pacific Visualization Symposium.
[9] K. Marriott et al. Immersive analytics, Springer, 2018.
[10] List of Regional Weather Observation Stations, URL: https://www.data.jma.go.jp/utsunomiya/kishou/amedas_titen4.pdf