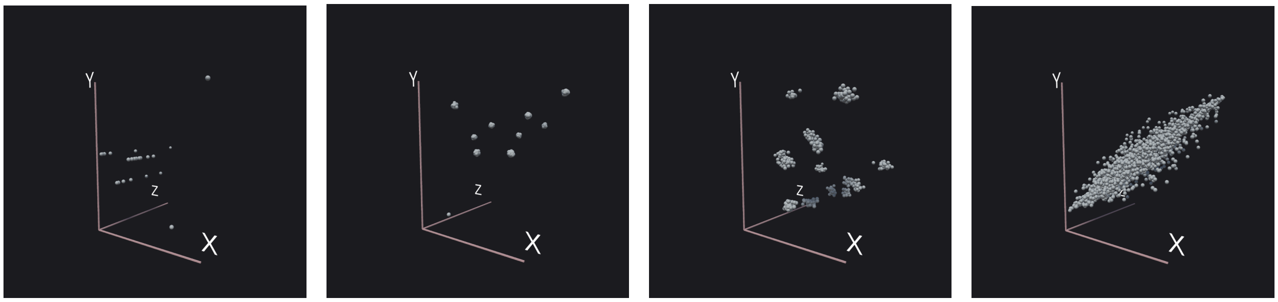
Framework for comparing 3D Scatterplot Analysis in Immersive and Non-Immersive Environments using Smartphones

Kartik Prabhu and Vikram Apilla



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

Fig. 1: Different Datasets: from left to right, the number of data points in the 3D scatterplots are in increasing order.

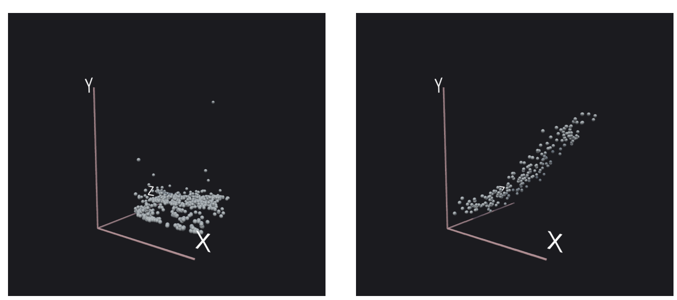
Abstract—The report is a description of the project in which we propose a framework to make a comparison of user performance in data visualization using a smartphone. The data visualization technique and the evaluation method considered here is a 3D scatterplot and a controlled user study respectively. The user study proposed is based on the scatterplot analysis, with tasks consisting of namely *Outlier Detection*, *Cluster Analysis*, and *Correlation Detection*. These tasks would be performed in different environments and on different datasets varying in terms of the size of points as shown in Fig. 1 as a sample. For example, a task, based on *Outlier Detection* would be performed on different datasets, and in different environments(immersive and non-immersive). Further, the framework also includes a few hypotheses and ways to assess them. Based on the results recorded from the user study, one could make a comparison between user performances over different datasets between immersive and non-immersive environments. With the help of this project's framework, conclusions could be drawn about the strengths and limitations of data visualisation using a smartphone.

Index Terms—3D Scatterplots, Augmented Reality, Immersive Analytics, User Study

1 Introduction

We live in an age of data where every aspect of our lives is digitally recorded. With the increasing amount of data, the technologies to visualize it has also been growing at a rapid rate. One could be interested in using the latest technologies to visualize the data to understand the information, patterns and dependencies from it. These technologies include devices such as *VR Glasses* and *HoloLens* which helps in immersive data visualization, but, the caveat here is that they are not always at the reach of a user. One device which has been around for a while and also the one that is frequently accessed on-demand by a user is a mobile phone. Mobile phones have evolved incredibly over the past decade that they are now called *Smartphones*. The idea of this project is to exploit the features provided by the current generation of smartphones in visualizing the data, specifically 3D scatterplots.

Every year, we are witnessing a massive growth in terms of computational power and features that are available through smartphones. Although they are small, these devices are used by almost everyone because of their compactness and functionalities. Every industry, right from entertainment to education, is targeting a smartphone as one of the major platforms. The device offers extensive capabilities such as video streaming, online gaming, and even acting as a virtual assistant. Tech giants, in particular, Google and Apple have invested their time on building augmented reality Software-Development-Kits so that new immersive applications can be developed on mobile platforms with ease. With all these advances and developments in a personal device, one could think, if it is possible to visualize 3D data immersively with the current generation of smartphones with high-resolution displays, and their ability to provide augmented reality features.



(a) Outlier Detection (b) Correlation Detection

Fig. 2: Detecting tasks presented to the user

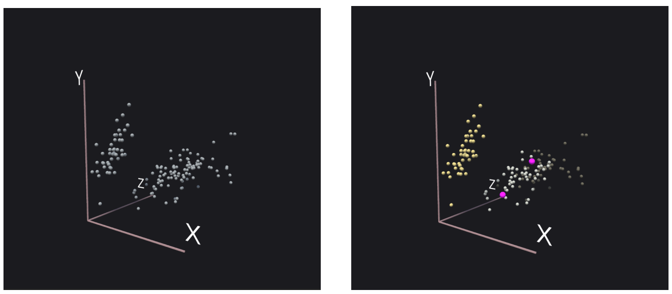
In this project, our focus is on the comparisons made based on user performances in scatterplot analysis using a smartphone supported with augmented reality. With that under consideration, the project can be divided into two parts. The first part helps in addressing, how smartphones could be useful in visualizing different datasets varying in terms of their size. The second part helps in understanding the immersive level with which a user can visualize the data in augmented reality setup using the smartphone as compared to a non-augmented

- *Kartik Prabhu, MSc. in Digital Engineering student[1]. E-mail: kartik.prabhu@st.ovgu.de.*
- *Vikram Apilla, MSc. in Data and Knowledge Engineering student[1]. E-mail: vikram.apilla@st.ovgu.de.*

[1] *Otto-von-Guericke-Universität Magdeburg*

reality one. In this project, we also consider that the user can take advantage of the 3D perception while analyzing a 3D object as presented in the paper [7]. Considering all these aspects, it would be interesting to answer the research question *How effective is 3D scatterplots visualization of different datasets in different environments(immersive and non-immersive) using a smartphone?*

The report is further organized into different sections. Firstly, the *Background* section gives a broad overview of the major components involved in the project. Secondly, a summary of closely related papers is discussed in the *Related Work* to understand prior research. This is followed by the proposed *Hypotheses* of the project. Then, a detailed description of how the controlled user study could be conducted using the framework is given in the *Study Design* section. Furthermore, a short text about the *Evaluation* is made to understand what could be expected of the framework once the user study is completed. Finally, remarks of the project are provided in the *Discussion* and *Conclusions and Future Work* sections.



(a) Clusters Estimation (b) Cluster Assignment

Fig. 3: Estimating tasks presented to the user

2 Background

In this section, we briefly describe the important facets of this project. The purpose is to provide the user with enough knowledge to understand further topics better.

2.1 Augmented Reality

Augmented reality is ”An interactive combination of reality and virtuality” [3]. The rendering is performed in real-time and 3D objects are registered (geometrically). One can view the world around them through their smartphone cameras while projecting an augmented object as overlays, that makes it seem as if the virtual objects were real. The applications in the field of augmented reality are tremendous. One of the most trending games of all time, Pokemon Go, is based on augmented reality, which made the world of Pokemon come to life. Through this, the users got the immersive experience of the gaming world. Similarly, in this project, we make a study of how augmented reality gives an immersive experience with data visualisation to the users.

2.2 Immersive Analytics

Understanding the data and making decisions based on it using the latest technologies such as *Large Displays, Augmented Reality*, and *Virtual Reality* is called *Immersive Analytics* [6][1]. Scenarios including egocentric methods such as data visualization using virtual reality, or multi-user visual analytics using large displays are all good examples of Immersive Analytics. A considerable amount of research has been carried out in this new area of data analytics. Furthermore, the application of Immersive Analytics are a plenty [11], as proposed by Lu et al. In this project, we would like to conduct our part of the research by focusing extensively on the immersive environment Augmented Reality with the help of a smartphone.

2.3 Scatterplots

A scatterplot is a type of plot or mathematical diagram used to represent distributions of 2 or more continuous variables of a dataset. It helps to understand associations between variables and to understand the sparsity of the dataset. Analysis of clusters, correlations and identification of outliers are some of the typical uses of scatterplots [15].

3 Related Work

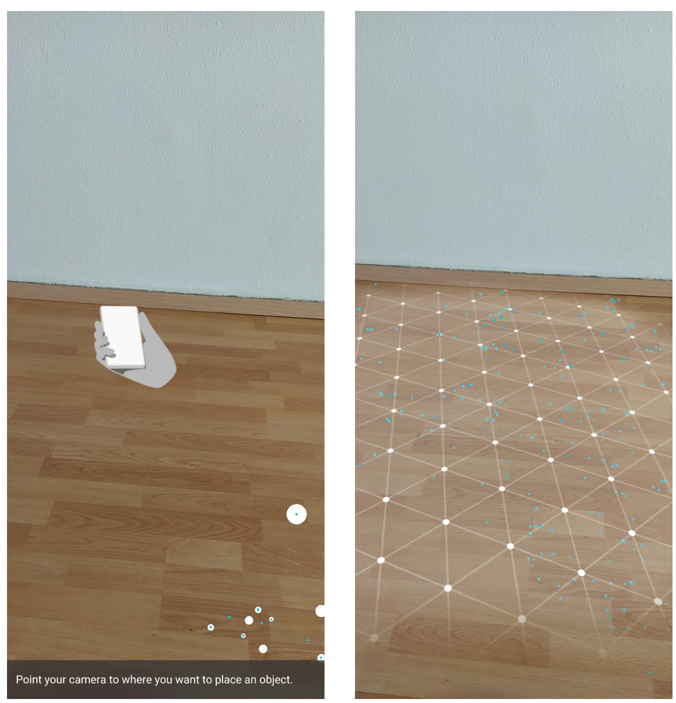
There has been prior research conducted in mobile phone data visualization and augmented reality-based 3D visualizations. In the following, we summarize three closely related research to our work.

In 2006, Buering et al. [4] published their work about data visualization of 7500 data points in a scatterplot using a mobile phone. They made a comparison between two zooming techniques, Geometric-Semantic Zoom and Fisheye Distortion. While the results between these techniques were indistinguishable, it was interesting to see that there was research conducted in mobile data visualization back in the years where smartphones were not as powerful as they are now.

Butcher et al. [5] have proposed a cost-effective solution for Immersive Analytics in 2016. They have successfully made use of the economical Google Cardboard coupled with a smartphone for data visualization using virtual reality. Although they have effectively built an immersive environment, it has to be noted that to take actions such as highlighting and zooming, it proved to be difficult without the hand controllers. In this project, we try to avoid this by switching the environment to augmented reality and thereby providing the user with complete control to take action using their phone through natural gestures.

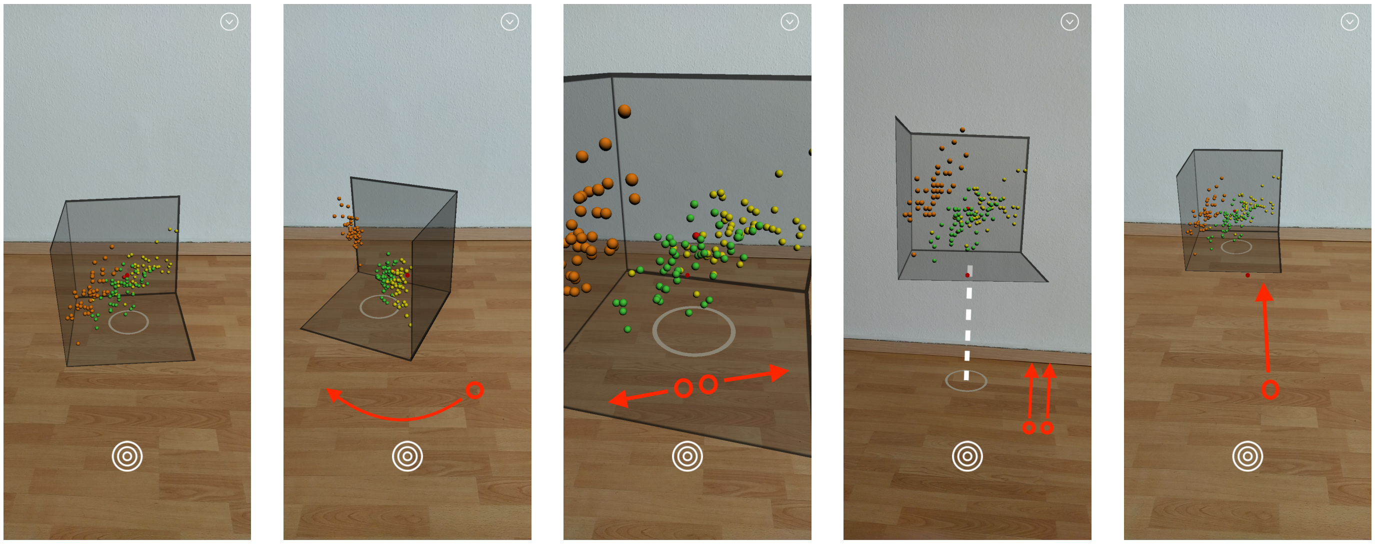
Recently, another research which is closely related to this project was published in the year 2017 by Bach et al. [2]. Their work consisted of a user study based on the comparison of 3D visualizations between three different environments - Desktop, HoloLens and a Tablet. While their results showed that desktop was faster in terms of task completion, HoloLens led the way in immersion.

All of these prior research has laid a good foundation for further research to be conducted in the area of Immersive Analytics, mainly in this case, using a smartphone.



(a) Hand Gesture Animation (b) Detected Plane

Fig. 4: Plane detection system in ARCore



(a) Initial Position (b) Rotation (c) Zooming (d) Elevation (e) Translation

Fig. 5: Multi-touch gesture support using ARcore manipulation system. A tap and drag outside the object rotates, pinch for zoom, 2 finger tap and vertical drag for elevation, tap on object and drag for translation (left to right).

4 Implementation

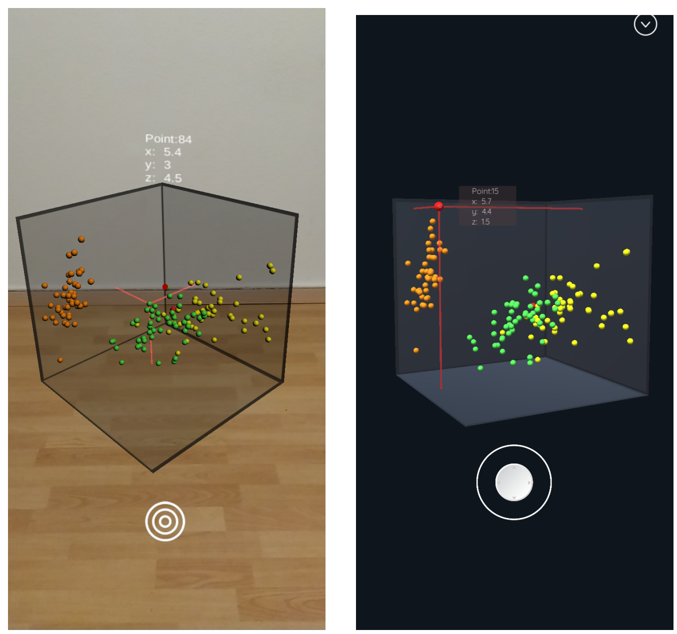
The application for scatterplot analysis was built on the Android platform with the support of ARCore library on Unity3D. ARCore enables a supported smartphone to sense and interact with the environment around it. It is based on 3 key features - motion tracking, environmental understanding and light estimation. The library makes use of sensors (accelerometer, gyroscope, light sensor) present in the smartphones to enable these features.

ARCore also supports a manipulation system as illustrated in Fig. 5. The user can *Scale* the entire plot uniformly in all directions with a pinch gesture. *Rotation* and *Translation* interactions are similar to Kruger et al.[8]. A central circular area (≈25% of the object size) is defined where the object is translated. With a contact point outside, a movement with a pointing device or touch rotates the object. Using 2 fingers perpendicular motion, the user can elevate the plot. All these manipulations have only one degree of freedom.

The augmented plot object can be placed on a detected plane. Plot placement follows the general guideline that the object should move along a supporting surface or an artificial interaction plane, in this case, the detected plane surface. The plot placement supports two interaction tasks as mentioned in [14]: a new object is inserted in the virtual environment and an existing object is selected, picked up and placed elsewhere, but the 3rd task of translating incrementally is not supported, as mobile devices don't have explicit keys. An animation to guide the user on how to detect a plane is also implemented as shown in Fig. 4.

For the selection of glyph points in the scatterplots, raycasting is implemented. Considering the large number of data points that would be used in the user study, spatial cues such as shadows, grid lines, and ghost cuts would not be helpful as shown by Luboschik et al. [12]. Hence, they are not included in the framework application. Rather, to assist the user, the co-ordinates and point number is displayed in a tooltip which appears above the plot while always looking at the camera as shown in Fig. 6a.

In the case of non-augmented reality environment, for selection of glyphs, a cross-hair selector controlled by a joystick on screen is implemented as illustrated in Fig. 6b. The manipulation system consists of scaling and rotation with y-axis as the pivot. Tooltip support, just like in the augmented reality environment is given.



(a) AR tooltip with raycast selector (b) Non-AR tooltip with joystick

Fig. 6: Selectors and tooltips in different environments

5 Study Design

This section covers the details about the controlled user study to evaluate the user performance [9].

5.1 Device

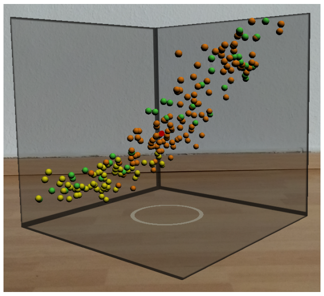
The device used for the study is Xiaomi Mi A3, supported on Android-one program, running on Android 9 OS. It meets the minimum SDK version of 27 to support ARCore library(1.15.0). The device consists of Qualcomm® Snapdragon™ 665 processor with a Super AMOLED Display of 15.46cm (6.08) and HD+ display (1560×720). The advantage of having a smartphone in the experiment is that the user can use regular gestures for manipulating the object on the screen.

5.2 Environments

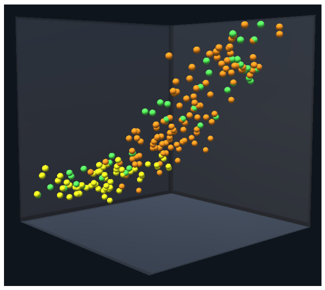
As the tasks would be based on the comparison between augmented reality-based and non-augmented reality-based scatterplots, the application used for the study would contain both the environments.

5.2.1 *With Augmented Reality*

With augmented reality, the user would be able to register the plot in their desired location and use the phone's ability of the movement to analyse the plot.



(a) Augmented Reality



(b) Non-Augmented Reality

Fig. 7: Different environments for user study

5.2.2 *Without Augmented Reality*

In the case of the non-augmented environment, the plot would be displayed with a fixed background. While the provision of using simple finger gestures for manipulating the plots is provided, other features of the smartphone, such as the camera, gyroscope and accelerometer, are disabled. Thus, the user can only analyse the plot with limited movement like any other non-augmented reality application.

5.3 Procedure

The application for the survey is built in a quiz format, where the users are able to record their responses either through a numerical or a multiple-choice answer. For the survey itself, the participants(users) would be split into two groups. One group allowed to perform the tasks with augmented reality, and the other without augmented reality. The reason for different groups is to introduce counterbalancing and avoid any learning effects resulting from the sequence of different experiments. Also, before the user study, every participant is provided with a ten minutes tutorial on how to complete the user study.

The participants' tasks and responses are both presented and recorded respectively using an android application. The application would contain two different branches. One without augmented reality, and the other with augmented reality as shown in Fig. 7.

Along with the response to a task from the participants, the task-completion time is also recorded automatically. It would be the basis for the qualitative and summative evaluation[10]. During the task, the number of interactions using different manipulation techniques is also observed. To understand immersion levels however, the responses are recorded based on users personal experience with the two given environments using the Likert scale.

5.4 Datasets

The datasets used for the user study are listed in the table 1 with its name, and the task it is used for and the number of points it contains.

All of them except the *Randomly Generated* were sourced from the internet which was free to download and publicly available. The assignment of the dataset to a task was done appropriately. For example, a dataset which could only exhibit correlation and had no clusters was not chosen for a cluster assignment task and vice-versa. Further, datasets such as *Iris* and *Seeds* exactly had three quantitative variables whereas others such as *Ebola* and *Cancer* had more than three numerical features. In such cases, we have decided to display three randomly chosen quantitative variables.

Dataset Name	Task	Data Points
Ebola	Outlier Detection	32
Haberman's Survival	Outlier Detection	307
Cancer	Outlier Detection	700
Iris	Cluster Assignment	151
Graph	Cluster Assignment	1025
Dimensions	Cluster Assignment	2027
Big City	Correlation Estimation	50
Dome-Shape	Correlation Estimation	99
Seeds	Correlation Estimation	211
Binary	Correlation Estimation	4097
Randomly Generated	Performance Test	20000

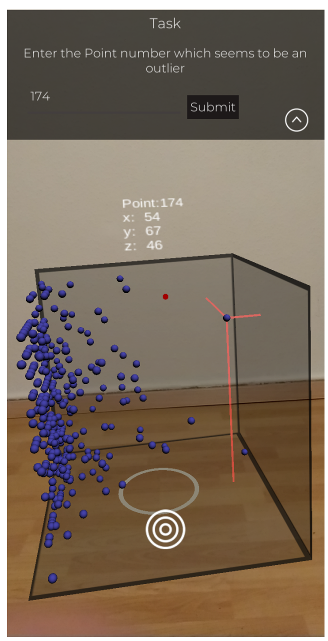
Table 1: Datasets with their task assignment and size

5.5 Tasks

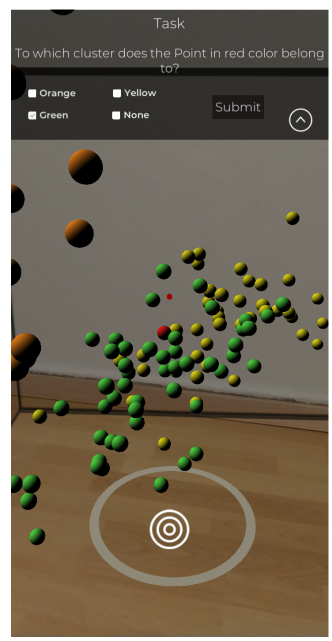
Three types of tasks are presented to the participants in the user study. The tasks are based on scatterplot analysis namely *Outlier Detection*, *Cluster Analysis*, and *Correlation Detection*. Each task is presented thrice to the participant, every time the number of data points in the plot would vary. For example, consider *Outlier Detection*, a participant would be asked to detect outliers in three different scatterplots, each plot with an increasing number of data points

5.5.1 *Outlier Detection*

A 3D scatterplot is presented to the user as shown in Fig. 8a. Based on this scatterplot, the user would be asked to find the data point that is maximally away from the rest of the points. Although sounds simple, it would be hard to choose between two data points that have almost the same distance away from others in the scatterplot.



(a) Outlier detection



(b) Cluster assignment

Fig. 8: Tasks in the immersive environment

5.5.2 *Cluster Analysis*

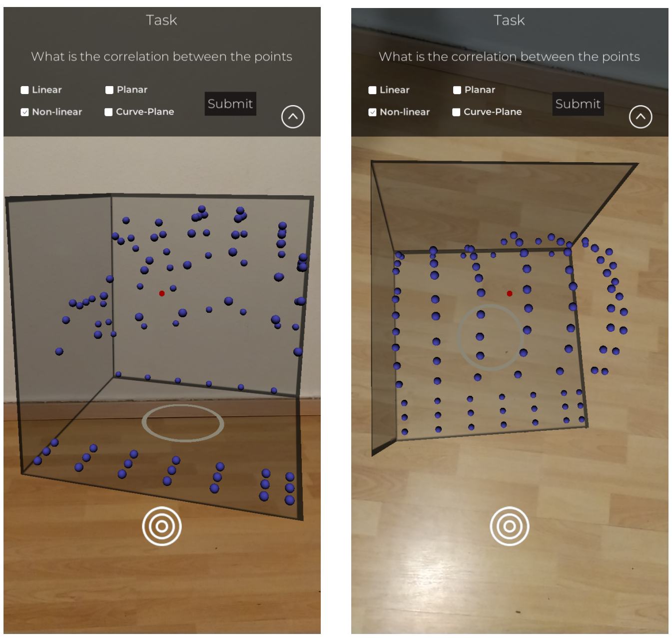
The cluster analysis task is composed of two different sub-tasks as shown in Fig. 3.

i *Clusters Estimation*: a point cloud without colour is presented to the user. Based on this point cloud, the user has to estimate the total number of clusters present in the dataset.

ii *Cluster Assignment*: the user has to assign a highlighted point in the point cloud to an appropriate cluster. For example, in Fig. 8b, the points that are enlarged and highlighted with red colour has to be assigned to one of the three clusters in orange, yellow, or green.

5.5.3 *Correlation Detection*

For the correlation detection, the user has to respond whether the scatterplot that is being presented has a strong correlation, weak correlation, positive correlation or negative correlation, linear, non-linear or planar. A sample of the correlation detection task can be found in Fig. 1 (d) and Fig. 2 (b). In both of these figures, the datasets contain a strong positive correlation. The same task in immersive environment is shown in Fig. 9.



(a) Front view (b) Top view

Fig. 9: Correlation detection task shows disordered shape from front view, but shows planar correlation from the top view

5.6 Participants

The user study could be conducted within the range of 10 to 20 in each participants' group. Preferably, participants with a background in Data Mining, Machine Learning, Visualization or Visual Analytics must be chosen.

6 Hypotheses

In this project, we propose few hypotheses in the following. It would be interesting to see which of these would be proven right and which of them would be proven wrong by conducting a user study.

1. *Participants to perform better in the immersive environment than a non-immersive environment* because of the degree of freedom provided by augmented reality. In the augmented reality environment, the user can freely walk and look around the object as opposed to being stationary in a non-immersive environment.

2. *Accuracy of detecting the outliers would not differ among the different datasets* because outliers are generally clearly evident since they are far away from the average position of other data points.

3. *Cluster assignment of data points would be harder as the number of clusters and data points increase* because when there are many clusters and data points, it would be difficult to understand the cluster membership of a data point.

4. *User interactions to be less in an immersive environment than a non-immersive environment* because of freedom. Consider an example, if the user wants to closely examine a data point, the user might use zooming in a non-immersive scene, whereas, the user could walk towards the point in an immersive scene.

7 Evaluation

The user study results could help in making interesting findings as mentioned below.

1. Comparison of participants performance over different datasets in terms of the number of data points.

2. Comparison of participants performance between the environment with augmented reality and without augmented reality.

3. Immersion levels of participants over the two environments.

4. Number of interactions of users in augmented and non-augmented reality environments.

8 Discussion

In this section, we would have ideally discussed the data collected from the user study. Analyse the graphs from the error rates and time taken to complete each task and summarise the questionnaire filled by the user to check overall preference towards Augmented reality. Unfortunately, at the time of writing this report, the world was hit by an epidemic in the form of COVID-19[16], and the survey was cancelled. Therefore, we tweaked the project to be more of a framework. Nonetheless, we mention some of the observations below which we found while working on the project.

8.1 Device Limitations

A striking limitation is caused by holding the phone for a long time. Even though we are used to, it could lead to fatigue to the user. Further, the screen size of the phone must also be considered, although tiny, it is good enough to analyze a medium-sized dataset but not for a large one.

Another limitation is the performance of the mobile device when the points in the dataset are more than approximately 5000. The rendering starts lagging and manipulation becomes difficult while at the same time, thousands of points are occluded as shown in Fig. 10.

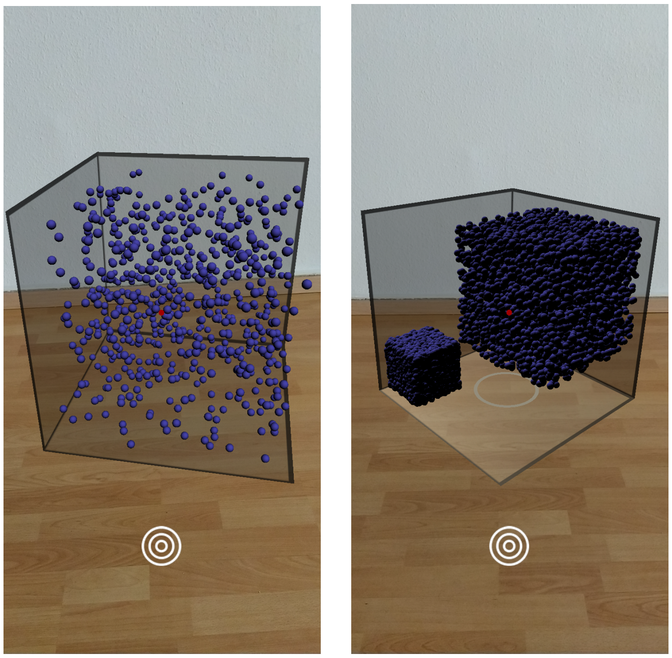
8.2 Environment Differences

It has to be noted that both the environments were given the options of zooming and rotating around the only Y-axis. The materials used for the rendering were also the same.

The augmented reality environment showed problems due to differences in light, whereas the non-augmented reality scene showed no effects. In the non-augmented reality, however, the problems were more related to the perception of the data points. This was due to its constant dark background. It is hard to perceive shapes such as the dome-shaped scatterplot in Fig. 9 without additional rotations along X and Z axes. However, in the augmented reality environment, the user can perceive such shapes without the need for rotation across different axes because of the degree of freedom to walk around the object.

9 Conclusions and Future Work

In this project, we have illustrated a framework for data visualization using the current generation of smartphones. Besides, a controlled user study is proposed on how to measure the effectiveness of the user performance in data analysis using a smartphone. This idea was to show that the current mobile phones have the capabilities to extend



(a) Sparse dataset of 700 points (b) Dense dataset of 20,000 points

Fig. 10: Datasets with varying densities

their use beyond the typical activities of capturing photos and playing games. We made use of the latest technology and research conducted in the Immersive Analytics as a baseline for the current work. Even though the smartphones are tiny devices, their current capabilities make them useful for interesting research in Immersive Analytics.

The natural extension to the current work would be to conduct a user study that was discussed in the earlier sections. Based on the results gathered from the user study, further remarks could be made on the effectiveness of data visualization through a mobile phone. To broaden the scope of the current work, another potential comparison between the mobile phone and the HoloLens could be made within an immersive environment. Through this comparison, effects of fatigue, display resolution and semi-transparency could be studied over these two devices. Further, a qualitative evaluation[13] could be conducted based on collaborative task solving in this application. This could be analyzed by recording the device screen and the conversation between the two subjects.

10 Acknowledgements

We would like to thank the developers of *Unity* and *ARCore* library for providing *easy to work* features. In addition, we would like to also extend our gratitude towards *Plotly* for their datasets which are public and freely available to download.

11 Attachments

We have added a short video (link) for the reader to get a glimpse of the application discussed in this project.

References

[1] 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 *Proceedings of the 2016 ACM International Conference on Interactive Surfaces and Spaces*, pages 529–533, 2016.

[2] B. Bach, R. Sicat, J. Beyer, M. Cordeil, and H. Pfister. The hologram in my hand: How effective is interactive exploration of 3d visualizations in immersive tangible augmented reality? *IEEE transactions on visualization and computer graphics*, 24(1):457–467, 2017.

[3] M. Billinghurst, A. Clark, and G. Lee. A survey of augmented reality, 2014.

[4] T. Buering, J. Gerken, and H. Reiterer. User interaction with scatterplots on small screens-a comparative evaluation of geometric-semantic zoom and fisheye distortion. *IEEE Transactions on Visualization and Computer Graphics*, 12(5):829–836, 2006.

[5] P. W. Butcher, J. C. Roberts, and P. D. Ritsos. Immersive analytics with webvr and google cardboard. *Posters of IEEE VIS*, pages 30–32, 2016.

[6] T. Chandler, M. Cordeil, T. Czauderna, T. Dwyer, J. Glowacki, C. Goncu, M. Klapperstueck, K. Klein, K. Marriott, F. Schreiber, and E. Wilson. Immersive analytics. Sept. 2015.

[7] G. Cliquet, M. Perreira Da Silva, F. Picarougne, Y. Prié, and T. Vigier. Towards HMD-based Immersive Analytics. In *Immersive Analytics Workshop in IEEE VIS 2017*, 2017.

[8] R. Kruger, S. Carpendale, S. D. Scott, and A. Tang. Fluid integration of rotation and translation. In *CHI 2005: Technology, Safety, Community: Conference Proceedings - Conference on Human Factors in Computing Systems*, pages 601–610, 2005.

[9] H. Lam, E. Bertini, P. Isenberg, C. Plaisant, and S. Carpendale. Empirical studies in information visualization: Seven scenarios. *IEEE Transactions on Visualization and Computer Graphics*, 2012.

[10] J. R. Lewis. Usability Testing. In *Handbook of Human Factors and Ergonomics: Fourth Edition*, pages 1267–1312. John Wiley and Sons, mar 2012.

[11] A. Lu, J. Huang, S. Zhang, C. Wang, and W. Wang. Towards mobile immersive analysis: A study of applications. In *2016 Workshop on Immersive Analytics (IA)*, pages 25–30. IEEE, 2016.

[12] M. Luboschik, P. Berger, and O. Staadt. On spatial perception issues in augmented reality based immersive analytics. In *Proceedings of the 2016 ACM Companion on Interactive Surfaces and Spaces*, pages 47–53. 2016.

[13] B. Preim, T. Ropinski, and P. Isenberg. A critical analysis of the evaluation practice in medical visualization. *Eurographics 8th EG Workshop on Visual Computing for Biology and Medicine,Sep 2018, Granada, Spain.*

[14] R. J. Teather and W. Stuerzlinger. Guidelines for 3D positioning techniques. In *Proceedings of the 2007 Conference on Future Play, Future Play '07*, 2007.

[15] E. R. Tufte. *The visual display of quantitative information.* (Graphics Press, Box 430, Cheshire, CT 06410, USA), 1983.

[16] T. P. Velavan and C. G. Meyer. The COVID-19 epidemic. *Tropical Medicine and International Health*, 2020.