

VR-Explorer: A Demonstration of a Virtual Reality Interactive Data Exploration System

LI HAOYUN, University of Science and Technology of China, China

CAO YUKUN, University of Science and Technology of China, China

XIKE XIE, University of Science and Technology of China, China

Interactive Data Exploration (IDE) is a common way to comprehend large volume datasets, which can help user gradually discover his/her data interest. The core idea of IDE is discovering user's data interest through multiple rounds of user interactions, such as labelling data tuples. However, there is still a gap between IDE and users, especially the lack of user interactivity research and data representation design. In this work, we designed and implemented VR-Explorer, an virtual reality interactive system on data exploration that connects IDE with VR technology. We designed virtual interactive interfaces for each exploration step according to the workflow of IDE and demonstrated the good usability of VR-Explorer through a study.

CCS Concepts: • Human-centered computing → Interactive systems and tools.

Additional Key Words and Phrases: virtual reality; interactive data exploration; human-computer interaction.

ACM Reference Format:

Li Haoyun, Cao Yukun, and Xike Xie. 2023. VR-Explorer: A Demonstration of a Virtual Reality Interactive Data Exploration System. In *The 2023 ACM International Conference on Supporting Group Work (GROUP '23) Companion (GROUP '23), January 8–11, 2023, Hilton Head, SC, USA*. ACM, New York, NY, USA, 4 pages. <https://doi.org/10.1145/3565967.3570976>

1 INTRODUCTION

Interactive Data Exploration (IDE) [3] relies on system assistance and interactive guidance to extract data ranges of interest to users from the database without the clear specification of queries. The IDE's problem is challenging due to the two-fold nature of the user's data interests: 1) User interest is intangible and thus requires incremental refinement/exploration [1]; 2) User interest is indescribable because it is often too complex to be expressed by the user through traditional query languages (e.g., SQL) [4]. A typical Explore-by-Example paradigm [2] of IDE views the exploration process as an iterative training process for the classifiers. Users only need to iteratively determine whether the database tuples are “interesting” so that the classifiers can infer the interesting data region after being trained on the labelled tuples.

IDE system has a wide range of application prospects. Starting from an IDE scenario in a comic form, as shown in Figure 1, a reader wants to find a book of interest in the library. However, his interest is so uncertain that it is hard for him to express it accurately. Meanwhile, he cannot use complex information retrieval tools like SQL or filters. Alternatively, it is easy for him to communicate with an IDE assistant, which probes and explores his query interests by offering a few samples for him to label, and takes the feedback for recommending quality data samples and refining his query interests iteratively. For example, an IDE assistant is on the library's navigation website. The assistant repeatedly

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).

© 2023 Copyright held by the owner/author(s).

Manuscript submitted to ACM

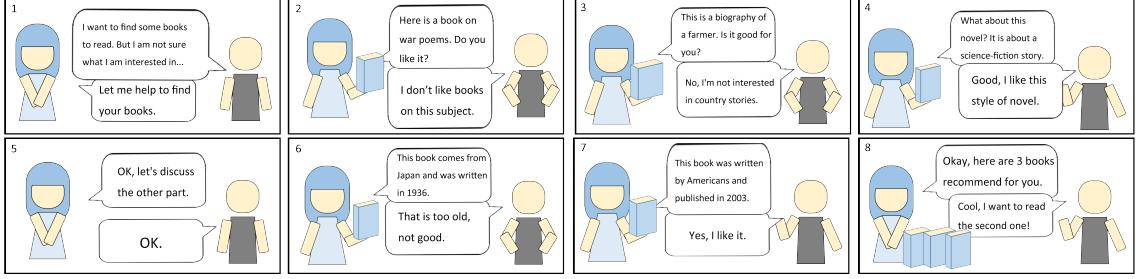


Fig. 1. A reader (the right one) who doesn't know how to use the retrieval tools interacts with the IDE assistant (the left one) to find his favorite books from the library database.

recommended some books to the reader and constantly located his interest according to his feedback. During this process, readers can choose to stop or continue the interactions with the assistant.

As a technique for human-computer interaction, IDE needs to provide users with an interactive interface. So far, existing interactive system prototypes [5] work as a backend platforms [1] [2], which are mostly on the server side. These systems, however, lack of user interactivity research and data representation design, which makes it difficult for users to understand the data.

Virtual reality(VR) technology can simulate user's physical presence in a virtual environment, allow user to immersively interact with virtual objects. By turning abstract data into virtual entities in VR scenes, user can understand data intuitively. In this work, we combined IDE with VR technology and proposed a virtual reality interactive system on data exploration, called VR-Explorer. We implemented the specific exploration procedure of user interaction according to the workflow of IDE, and designed user-friendly virtual interfaces incorporating human-computer interactions. Finally, we conducted a study on the car database and demonstrated that the VR-Explorer is of good functionality and usability for public users.

2 SYSTEM DESIGN

2.1 Procedure

According to the workflow of IDE [1], we divided the interaction procedure into five stages, *Attribute Selection*, *Attribute Group*, *Initial Labelling*, *Iterative Labelling*, and *Result Display*. We take the car exploration as the case and make a storyboard to describe the interaction process and the design of VR-Explorer (see Figure 2). Note that VR-Explorer is a general-purpose system that supports data exploration in any field. Just upload the dataset and replace the 3D models in the virtual scene to make the system adapt to the corresponding field.

After selecting the dataset (Figure 2(a)), the user will enter a sci-fi style virtual room where he/she can walk freely. The first task is *Attribute Selection* (Figure 2(b)). In this stage, the attributes of the dataset are displayed on the virtual scene. User should pick the attributes that they care about. The system explores the data based on the selected attributes. To avoid considering all the selected attributes at once, in *Attribute Group* stage (Figure 2(c)), user can divide the selected attributes into groups according to his/her knowledge of the relations of the attributes. Alternatively, user can also directly use the groups recommended by the system.

In the *Initial Labelling* stage (Figure 2(d)), there is a batch of insightful samples represented as virtual entities in the scene, which are the tuples extracted from the dataset. The user can interact with virtual entities through gesture-based

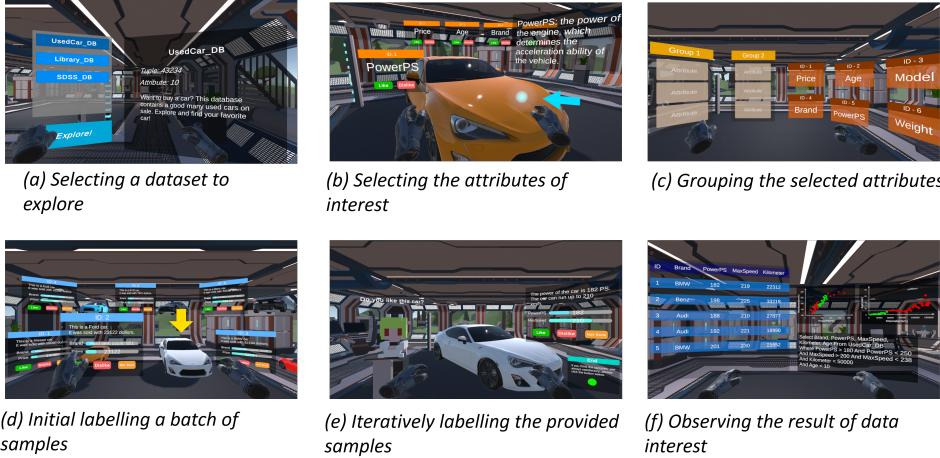


Fig. 2. The storyboard of VR-Explorer, defining the major procedure of data exploration

manipulations. By manipulating the VR controllers, user can move, rotate, or scale the entities and observe the features of the samples from various directions and levels. And the user can click the “Like” button with green color to indicate that this sample is of interest to him/her, or click the “Dislike” button with red color. User’s like on these samples enables the system to roughly determine the user’s interest. Then, in the *Iterative Labelling* stage (Figure 2(e)), the system iteratively provides updated samples to the user, and asks for his/her labelling feedback to gradually grasp the user’s interest. User can click the “End” button to terminate the labelling if he/she thinks the values of samples meet his/her interest.

At last, in the *Result Display* stage (Figure 2(f)), user can browse the data recommended by the system as the result of the data exploration. The system also provides data plots for summarizing and demonstrating the statistical information of the dataset and showing the relation between interest region and the entire dataset. The SQL statements are utilized to describe the user interest at an abstract level for user with professional skills further process the data.

2.2 Implementation

The architecture of the VR-Explorer follows the C/S architecture. We choose the Unity engine to implement the client of VR-Explorer. The ways of interacting with the interface were coded using C#. The virtual scenes and 3D models are built based on the plugins from the Unity Asset Store. The server of VR-Explorer is running on Centos7.9. The communication between the server and the client is completed through the Django framework. We choose to use HTC Vive Pro 1.0 for immersive interaction, which includes a head-mounted display device, two wireless handheld controllers, and two 3D spatial laser-tracking stations.

3 STUDY AND VALIDATION

We took car exploration as the case study to evaluate the usability of our system. We loaded a Car [5] dataset to the system database, with 43K tuples and 10 attributes of second-hand car information on eBay. And we invited 10 participants (6 males and 4 females, all aged between 24-53, they are labelled as P1-P10) with the desire to buy a car to undergo the study. The assistant instructed participants to wear HTC VR equipment and enter the virtual scene. And

participants use VR-Explorer to discover their favorite cars. After the study, participants are asked to participate in a 15-minute semi-structured interview, to share their experience on VR-Explorer.

3.1 Result Analysis

Most participants found the ideal cars from different perspectives through VR-Explorer. As P3 said: “I need a car to commute, so I mainly consider its cost performance. I particularly like the Nissan car that it finally recommended. It’s cheap and low in fuel consumption.” P7 said, “I care more about the horsepower of the car. I hope I can drive fast on the highway, and it showed me some good coupes.”

In addition, exploring data in a virtual environment gives participants a rich interactive experience. As P2 said: “I feel like I have walked into the car shop, and these cars are right beside me, and I could even touch them.” And P8 said: “I can take a close look at the car and zoom in on some of its details.”

Moreover, we found that the interaction mode based on labelling enables user to quickly understand the use of the system. As P1 said: “All I have to do is express my like or dislike, so that it can find a fancy car for me.” And P4 said: “It’s easier to answer the judgment questions than the essay questions. This data query method is much simpler than other methods.”

4 CONCLUSIONS

In this work, we designed and implemented the VR-Explorer, a virtual reality interactive system on data exploration that connects IDE with VR technology. We made virtual interactive interfaces to cover the workflow of the IDE for enhancing user interactivity and data representation. The final study demonstrated that the VR-Explorer is of good functionality and usability for public users.

ACKNOWLEDGMENTS

This work is supported by NSFC (No.61772492, 62072428), the CAS Pioneer Hundred Talents Program. Xike Xie is the corresponding author.

REFERENCES

- [1] Kyriaki Dimitriadou, Olga Papaemmanouil, and Yanlei Diao. 2014. Explore-by-example: an automatic query steering framework for interactive data exploration. In *International Conference on Management of Data, SIGMOD 2014, Snowbird, UT, USA, June 22–27, 2014*, Curtis E. Dyreson, Feifei Li, and M. Tamer Özsu (Eds.). ACM, 517–528. <https://doi.org/10.1145/2588555.2610523>
- [2] Enhui Huang, Liping Peng, Luciano Di Palma, Ahmed Abdelkafi, Anna Liu, and Yanlei Diao. 2018. Optimization for Active Learning-based Interactive Database Exploration. *Proc. VLDB Endow.* 12, 1 (2018), 71–84. <https://doi.org/10.14778/3275536.3275542>
- [3] Stratos Idreos, Olga Papaemmanouil, and Surajit Chaudhuri. 2015. Overview of Data Exploration Techniques. In *Proceedings of the 2015 ACM SIGMOD International Conference on Management of Data, Melbourne, Victoria, Australia, May 31 – June 4, 2015*, Timos K. Sellis, Susan B. Davidson, and Zachary G. Ives (Eds.). ACM, 277–281. <https://doi.org/10.1145/2723372.2731084>
- [4] Tova Milo and Amit Shomech. 2020. Automating Exploratory Data Analysis via Machine Learning: An Overview. In *Proceedings of the 2020 International Conference on Management of Data, SIGMOD Conference 2020, online conference [Portland, OR, USA], June 14–19, 2020*. ACM, 2617–2622.
- [5] Xuedi Qin, Chengliang Chai, Yuyu Luo, Nan Tang, and Guoliang Li. 2020. Interactively Discovering and Ranking Desired Tuples without Writing SQL Queries. In *Proceedings of the 2020 International Conference on Management of Data, SIGMOD Conference 2020, online conference [Portland, OR, USA], June 14–19, 2020*, David Maier, Rachel Pottinger, AnHai Doan, Wang-Chiew Tan, Abdussalam Alawini, and Hung Q. Ngo (Eds.). ACM, 2745–2748. <https://doi.org/10.1145/3318464.3384695>

Received 30 September 2022; revised 1 October 2022; accepted 3 October 2022