

# Game development data analysis visualized with virtual reality

Chih-Wei,Chen,Taoi,Hsu

Department of Digital Multimedia Arts  
#1, Ln.17, Sec.1, Mu-Cha Road,  
Taipei, Taiwan  
+886 910056921, jryuichen@gmail.com

## Abstract

This project aims to improve the information visualization by means of emerging virtual reality technology. A game development and management information is used as the pilot research. Data analysis algorithm is developed in data cluster, data correlation and data regularity verified by the state of art algorithms. To visual the significant analyzed data is employed by the virtual reality allowing game developer and users effectively to mining and support a immersive visualization platform leading a less risk decision making manner. A Unity 3D tools is employed to establish the virtual reality platform.

This research takes K-Means clustering algorithm as the method to present three dimension space and then connect information visualization with virtual reality. This research proposed some advantages in this clustering analysis including (1) highly immersion (2) hidden information mining (3) extra spatial dimension (4) intuition operation mode (5) infinite display space.

**Keywords:** *virtual reality ; information visualization; big data; game data analysis*

## Introduction

21th century is an era that full of information and data. With background of knowledge economy, knowledge is power as well as productivity and competitiveness; not only individual has been kept acquiring knowledge for enhancing self-value, enterprise also constantly long for more information for a better market power. Information is diversity that has been communicated via time and spatial dimension[1].

Visualization is the main method of communication. By the sense of sight, human being should be quite flexible to explore new knowledge; our 80% experiences come from visual information [2]. Images simplified the process of text reading and present the point in an Intuitive way. Information visualization is a process of transforming data in which observation will be extracted, transformed, mapped, highly abstracted and integrated, the information will be then transformed in to graphics and presented by visual interface by means of computer graphics and image processing technique.

Information visualization has become one of important approaches to analyze data , yet complicated data structure and its huge quantity, three-dimensional graphics information visualization could not be well communicated, that will cause leave the audiences confused about the whole message. This

research thus take virtual reality platform as the method to process data mining so as to examine data correlation and data regularity; having embodied multi-dimension information, virtual reality interactive environment and mapping approach are to be employed. From different perspective to observe, the hidden data and its meaning will then be revealed as providing occlusion and mess-solving instantly.

## Information visualization

Information visualization is a process of transforming data into visual form; through information visualization, human visual acuity can observe, browse, discriminate, and understand the interactive graphics as to explore more hidden information as well as for better cognition [3].

### A. information visualization

Information visualization refers to a technique that are applied for big data in non-space, to present data and its structure with multi-dimension and multi-attribute in a straighter way; the audiences could observe from different dimension to analyze message. Here is the flowchart in the following(Fig. 1).

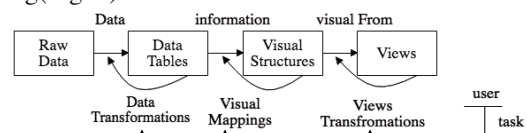


Fig. 1. Information visualization model

- (1) information space : Multi-dimension space that are formed by n, m dimension attribute including time dimension, special dimension, and information dimension.
- (2) information development : To extrapolate through model methods and tools.
- (3) information analysis : Information is extracted, transformed, mapped, highly abstracted and integrated.
- (4) information visualization : technique or tools that transform information into two or three dimension graphics, images, animation.

### B. Visualization Technique

Date visualization has been applied in many fields. For example, weather forecast could be reported by two or three dimension images. And, geographic information visualization could transform the prospecting data into graphics as to present the mineral distribution and oil filed distribution and so on. More, in medical aspect, it also could be seen such as computer tomography, or nuclear magnetic resonance. The visualization technique will be different based on different case.

### C. Two Visualization and Three Visualization

Dimension expanding and its complicity has become the new challenge for understanding data; from scientific perspective, data itself is not being the point that is to be studied but the hidden meaning behind data[4]. Information visualization is thus a crucial point for a better cognition. The process of information visualization is data quantity in which data will be transformed into graphics by means of algorithm. It may initiate audiences' visual acuity, and lead them to recognize the best way to present each different data.

In general, there is no exact difference between two dimension and three dimension as the similar aspect of getting more dimensions, sense of space, readability, occlusion, mess, distortion, scalability, visual angle and so forth. The differences are the data complicity, and the purpose, as well as the software selection for creating the best visualization. Tory and others have been making a lot of experiments, from which they have proposed that two dimension could be employed for the aspect of measure and illustration. As for three dimension, it could get the work done more effectively on the part of navigation and orientation. Having incorporated two dimension and three dimension, the visual performance should be more excellent[5].

While the game information visualization and virtual reality related topic have not been studied much, by taking D3.js, WebGL, Unity 3D software as the method to convey information, this research states that three dimension visualization could highly abstract data and analyze them more effectively.

### Virtual Reality

Virtual reality application has been published for the first time in her research, "The Ultimate Display", Sutherland proposed virtual reality structure from the perspective of computer monitor display and man-machine interaction. Since then the development of computer images and graphics processing technique has been promoted as well as the research regarding man-machine interaction such as hood display, interactive handle and so forth.

#### A. Virtual reality Definition and its Feature

Feature : Immersion, Interaction, and Imagination. By means of head tracking, spatial positioning, gesture capturing technique, enhance interactive performance. Virtual reality may bring more convenience and change future human life. The development of game market is the best example.

#### B. Virtual Reality Information Visualization Application

Big data visualization analysis has been gradually evolved, super-dimension has become a new challenge to human visual cognition. Most data analysis software could not handle this complicated multi dimension, while data analysis principle is being restrained. Since that the audiences could not explore data from different angle. Thus, many scholars have been proposed some new data visualization tools; under this support, data scientists carry out immersion type virtual reality technique as to embody interactive data space.

Immersion type virtual reality technique has been hugely applied in the field of medicine, forecast, design, industry,

information science[6]. Research have stated that three dimension should create better performance and learning curve than two dimension. In the meanwhile, having connected with virtual reality, better feedback of the result are to be acquired[7]. Virtual reality may provide advantages such as raising consciousness, reducing cognitive load, more real experience, stronger interaction, richer content, multiple people collaboration platform. With positive attitude, research and application that related to virtual reality also could be found in the part of medical surgery and drivers training. From the perspective of virtual reality value as well as its supply and demand, virtual reality related technique potential will be developed more wildly.

### Game Data Analysis

From cassette game to mobile cloud game, we can see the different game playing way in different era; nowadays the players could get resources in the game in an easier and convenient way, that triggering mobile game in to a whole digital era and in the meantime, the development of mobile game and its golden age has begun as well. In Taiwan, game output is about forty billions, in which mobile game is up to sixty percent; and even GooglePlay platform is ranked fourth worldwide. As to cast brick and attract jade, in recent years, the field that related game has become very popular.

Game data analysis may help enhance product quality, extend game life circle and its benefit. Lovell suggested that there are closed and reciprocal relationship between the obtaining new users and keeping old user and game benefit; for obtaining more user, the user's need will have to be recognized so as to improve users' adhered to game. It is clear that game big data is considerable important. Through big data analysis, pro-process decision, in-process optimization, and post-process examination could be done more easily, by which the different desire and need are to be provided for different player. Correct strategies are to be employed for different game life circle.

### Scientific Method

By taking Unity3D, this research created a system that allow the process of collecting information, of analyzing observation, and of visualizing graphics be executed successfully in virtual reality space; and then this research, by means of K-means clustering algorithm, carry out data mining. Each variance and relation will be verified between data; even though having unfixed status, with multi-variance, high dimension, complicated structure, through clustering analysis, high dimension data still could be classified and simplified quickly.

#### A. Normalize

Prior to K-means algorithm, data have to be normalized. Deformation and offset will need to be solved as to keep data unit be consistent with the other.

$$\begin{aligned} \text{avg}(x) &= \sum(x_i) / w \\ \text{std}(x) &= \sqrt{\sum(x_i - \text{avg}(x))^2 / w} \\ x_i' &= (x_i - \text{avg}(x)) / \text{std}(x) \end{aligned}$$

## B. PAA Dimension Reduction

Game data high dimension information including time, user message, user behaviors have to be reduced prior to visualization analysis. Mostly, the procedure are executed based on time series. PAA Dimension Reduction method Eamomm proposed has been employed for reducing time dimension [8], the result of dimension reduction is explained in Fig. 2.

The length of time series  $x$  is marked as  $n$  in  $N$  space, that will be functioned as the vector  $X = x_1, \dots, x_n$ ; the  $i^{\text{th}}$  element of  $X$  are to be functioned in the following :

$$\bar{X}_i = \frac{w}{n} \sum_{j=\frac{n}{w}(i-1)+1}^n x_j$$

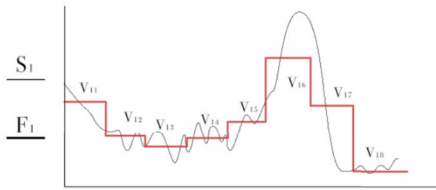


Fig. 2. PAA Dimension Reduction Diagram

## C. K-means Clustering

k-means clustering is a method of vector quantization in which the data that has been collected will be partitioned into  $k$  clusters; in the process of Euclidean Distance, the average value of each element in cluster  $w$  to obtain the minimum value of group square and WCSS therefore each observation is to be classified into different cluster that with the nearest mean. And in the next following steps of the process, the new observation data will be further classified in to existing clusters(Fig. 3.) Here is the clustering process in the following :

- (1). Initiate Random Partition and coordinate them :  $C_1(0), C_2(0), \dots, C_k(0)$ .
- (2). Partition  $X_j$  in to cluster that with the nearest mean, a certain cluster ( $S^i$ ) for each individual observation( $X_p$ ).
- (3). Based on the second step the new data is to be further classified :

$$C_i^{(t+1)} = \frac{1}{|S_i^{(t)}|} \sum_{x_j \in S_i^{(t)}} x_j$$

- (4). Partition new cluster from existing clusters.
- (5). Keep classifying each data into the cluster that with the nearest mean, just back to the second step and keep going the same process if possible until the data cannot be classified anymore.

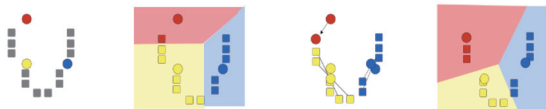


Fig. 3. Clustering process

## D. Similarity measure

Euclidean Distance is a very popular method for testifying similarity. The distances between two points in Euclidean space is a straight line that show the similarity value; the longer the distance, the higher similarity. Here is the formula for the following :

$$d(x, y) = \left( \sum_{i=1}^p (x_i - y_i)^2 \right)^{1/2}$$

## E. UNITY3D

Unity 3D is a game engine that is popular worldwide for the development of three dimension video game, with the best graphic technique DirectX and OpenGL, it could initiate physical engine, particle system, virtual reality, and so forth. Even VR product supplier have promoted their product SDK only for Unity3D users.

## F. data in three dimension

After accessing in to system, the data has been selected will be classified in to the value from one to ten, the value world coordinate will then be converted in to local object coordinate.(Fig. 4) Finally, the object are to be instantiated.

$$Y = (x - x_{\text{MinValue}}) / (x_{\text{MaxValue}} - x_{\text{MinValue}})$$

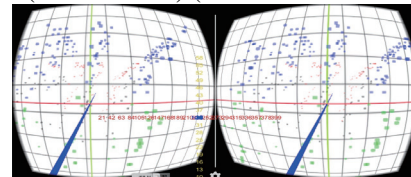


Fig. 4. data Concretization

## G. color data in cluster

After K group data is obtained by means of K-means, the color of object will need to be changed so as to help the audience recognize the correctness and difference of the information; through Euclidean Distance the distance value will be calculated, and then take up HSV method to color the component. Hue (H) is to be classified into K cluster. Color saturation (S) is the distance of data and group center point; the longer the distance, the lower the saturation. The lowest saturation is set as 30% forehand for better color cognition.

## H. Operation method

With VR glasses, the user should immerse in to virtual reality and analyze observation data quickly and easily without using hands. Different from the traditional tools as Gaze, attach Leap Motion controller to process the complicated part such as the move of push, of pull, of press, and toggle is be to an easier and convenient method with great 3D value [9].

## I. Game Data source

The players are crucial point for game industry yet the process of obtaining the related information is a hard work. This research take the information that Yeng-Ting Lee has collected from World of Warcraft as the object for the research. Through searching the player list the information is to be collected from game server every ten minutes. After three years collection, this research has obtained the observation data including time, serial number, character name, Gulid Class, level, race, Zone.



Fig. 5. data Concretization in virtual reality

### Research Finding

After comparing this research software and SPSS, this research states that this research software have some advantages such as the aspect of rotation, zoom, selection, the sense of space, the sense of distance, extra dimension, better cognition mode and so on.

In terms of game data, the data with higher dimension includes time, space, strategies for each need all of which could not be found in the traditional way. By taking K-means clustering analysis as the method, process data mining to examine each relation between each .

With virtual reality platform, mapped data into three dimension space as to make 360° analysis and examine every single change between data in order to help the audiences have a better understand of abnormal value and outlier. The data status could not be embodied in traditional way only images could be used for presenting data difference and distance.

Extra dimension space can be provided by virtual reality. As seen the diagram above(Fig.5), with Vive glasses accessing in to virtual reality space then analyze data by Leap Motion; the data users has been selected will be transformed in to graphics and mapped in to three dimension space. Only with VR glasses, the users and data could be connected immediately in the same space; and only with gesture capturing technique, the graphics could be embodied, dimension content could be switched, and, two data could be analyzed then. As just turning around the head, the related information could be acquired. Therefore, the details of data are to be shown in the sight of the users. By means of data clustering analysis, the mode of move, rotation, zoom and so on, should be easily processed, and the problem of occlusion , mess (distortion), visual angle could be solved as well.

### Conclusion

This research aims to take information visualization to analyze game data; having constructed immersive visualization system, the audiences can directly interact with data in an infinite three dimension space, by which the hidden data will be manifested by means of clustering data analysis.

Virtual reality is a technique that constantly make progress and may make change on future human life. What has been know that virtual reality refer to an application technique with high interaction and high immersion; a better cognition could be verified after the process of encoding data. This research connects information visualization and virtual reality to collect game data such as time, users, game context; diversity data will then be provided for making decision instantly, by which the audiences could discover data details from different perspective, and the statement that three dimension could not work well in market will be broken. Therefore, information could be conveyed in different way.

been a hot topic for discussion to many; related seminars have been held in order to create new strategies for analyzing data effectively. In term of game market in Taiwan, consumption capacity has been ranked ahead of most countries, yet Taiwan are not major country export game .Most game Taiwanese people are playing is the game imported from other countries. However, there are many

Taiwanese enthusiastic game producers who with great creativity, thus how to meet users' need as to make great success in game market will be the core value.

Mechanism could be the method for clustering data analysis, for example, computer software. However, the usability of the result is depended on observers' selection and discrimination; in the process of operating virtual reality mode, this research find out some advantages including (1) highly immersion, (2) hidden information mining, (3) extra spatial dimension, (4) intuition operation mode, and (5) infinite display space, from which the difference between data could be further recognized deeply.

### ACKNOWLEDGMENTS

This study support from the Shih Hsin University grants P10605.

### REFERENCES

- [1] S. Haag and M. Cummings , Management Information Systems for the Information Age, New York, NY: McGraw-Hill, Inc, 2009.
- [2] A. A. Berger, Seeing is Believing: An Introduction to Visual Communication, Mayfield Publishing Company, 1998.
- [3] A. Sears and J. A. Jacko, Human-Computer Interaction: Design Issues, Solutions, and Applications, CRC Press, 2009.
- [4] C. Donalek, S. Djorgovski, S. Davidoff, A. Cioc, A. Wang, G. Longo, J. S. Norris, J. Zhang, E. Lawler, S. Yeh, A. Mahabal, M. Graham and A. Drake, "Immersive and Collaborative Data Visualization Using Virtual Reality Platforms," *2014 IEEE International Conference on Big Data*, Washington, DC, USA, 2014.
- [5] M. Tory , A. Kirkpatrick , M. Atkins and T. Moll, "Visualization task performance with 2D, 3D, and combination displays," *IEEE Transactions on Visualization and Computer Graphics*, 12 (1), pp. 2-13, 1 2006.
- [6] L. P. Berg and J. M. Vance, "Industry use of virtual reality in product design and manufacturing: a survey," *Virtual Reality*, 21(1), pp. 1-17, 3 2017.
- [5] J. Gray, L. Chambers and L. Bounegru, The Data Journalism Handbook How Journalists Can Use Data to Improve the News, 1 editor , O'Reilly Media, 2012.
- [6] J. Steele and N. Iliinsky, Beautiful Visualization: Looking at Data through the Eyes of Experts, O'Reilly Media, Inc., 2010.
- [7] S. A. W. Andersen, P. T. Mikkelsen, L. Konge, P. C.-. Thomasen and M. S. Sørensen, "The effect of implementing cognitive load theory-based design principles in virtual reality simulation training of surgical skills: a randomized controlled trial," *Advances in Simulation*, 17(6)2016.
- [7] S. Dübel , M. Röhligh, H. Schumann and M. Trapp, "2D and 3D presentation of spatial data: A systematic review," *2014 IEEE VIS International Workshop on 3DVis (3DVis)*, Paris, France, 2014.
- [8] E. Keogh, K. Chakrabarti, M. Pazzani and S. Mehrotra, "Dimensionality Reduction for Fast Similarity Search in Large Time Series Databases," *Knowledge and Information Systems*, 3(3), pp. 263-286, 8 2001.
- [9] J. C. Coelho and F. J. Verbeek, "Pointing Task Evaluation of Leap Motion Controller in 3D Virtual Environment.," *Creating the Difference*, 2014.