A Review On Clutter Reduction In Multi-Dimensional Visualization By Using Dimension Reduction

Harpreet kaur, Shelza

Abstract— in this research paper, we have done a exploratory study in which we have found that Big data concepts leads to creation of visualization that are not gain in information gain for doing analysis as it is cluttered, thick and dense. To overcome it there are possibly three techniques to do. The techniques are dimension reduction, dimension reordering and clustering. It has also found that all techniques used to reduce clutter. Clutter may not suit all types of datasets and may also vary with number of dimensions. Typical flow of reducing clutter have also been discussed

Index Terms—multidimensional visualization, dimension reduction, dimension reordering, clustering, visual clutter.

1 INTRODUCTION

Big data is a collection of large and complex datasets which makes difficult for the database management tools and traditional data processing applications to process the data. The challenges of big data include capturing, storing, searching, sharing, transfer, analysis, and visualization of datasets. Big data is used to describe a massive volume of data both structured and unstructured data. Big data contains of billions to trillions of records of millions of some process all from different sources. When dealing with big datasets, organizations face difficulties to create, manipulate, and manage large data. Big data leads to clutter in their visualization. Clutter in data visualization makes it difficult to understand the displayed content.

1.1 Clutter

Clutter is a crowded or disordered collection of graphical entities in information visualization.

Clutter can blur the structure of data. In a small dataset also, clutter makes it hard for the viewer to find relationship and reveal patterns. Clutter is undesirable because it makes viewers difficult to understand the displayed content. When the dimensions or number of data items grow high, it is necessary for users to encounter clutter. Clutter reduces information gain from visualization. The visual clutter diminishes the potential usefulness of visualization. Clutter [13] is a state of confusion that degrades both the accuracy and ease of interpretation of information displays.

There are many techniques which are used to reduce the clutter and make the visualization better. However, many clutter reduction techniques may results in information loss and accuracy of data.

1.2 Why it is important to reduce the clutter

- Increases information gain.
- Increases visibility of hidden datasets.
- Increases insights into datasets.
- Reduces mental overload and stress.

- *Harpreet Kaur is currently pursuing masters degree program in Computer Science engineering at Swami Vivekanand Institute of Engineering and Technology in Punjab Technical University, Punjab, PH-9988250551. E-mail: harpreetmomia@gmail.com*
- *Shelza is currently working as assistant professor in computer science and engineeringdepartment at swami vivekanand institute of engineering and technology, Punjab, PH-9888381430. E-mail: shelzadesires@gmail.com*

- Saves time and improve effectiveness.
- Reorganizing makes it easier to access information and make things more accessible.
- Increases understanding and interpretation analysis of data.

2 VISUALIZATION OF BIG DATA

Visualization is the process of transforming data into graphical representation. The goal of visualization is to facilitate the user to gain a qualitative understanding of the information. Data visualization is the set of techniques used to turn a set of data into visual insight. A good visualization clearly reveals structure of the data. An ideal visualization needs to maximize the visibility of patterns and structure and minimize the clutter present. Earlier visualization was done by constructing a visual image in mind but nowadays visualization is like a graphical representation that supports in decision making which extracts a lot of information in one vision without reading a lot of data files. Data abstraction techniques are used in visualization to facilitate analysis from overview to detail. Data abstraction maintains dominant characteristics of the original dataset. Data abstraction reduces the details of data or it is the process of reducing a large dataset into moderate size dataset. The abstraction quality measure measures how well relative data density is maintained. The techniques for data abstraction in visualization include sampling and clustering.

3 A REVIEW OF METHODOLOGY

The methodology to reduce the clutter incorporates the following steps:

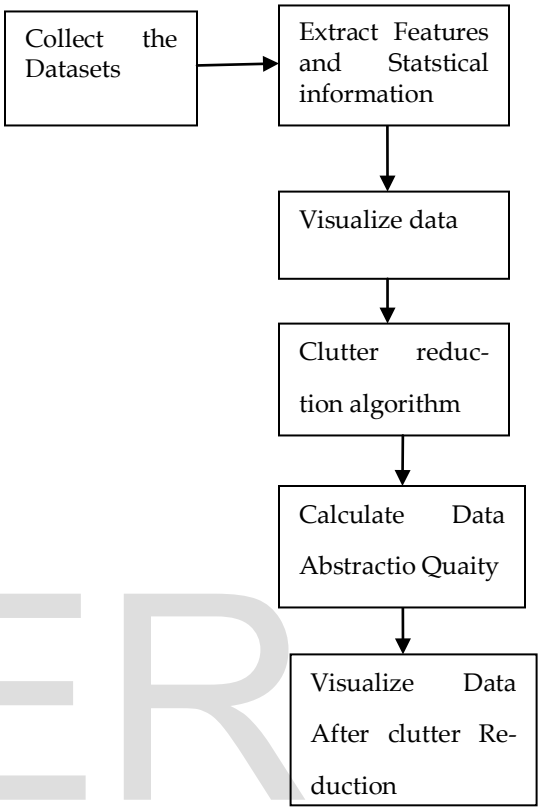

FIG. 1 PROCEDURE FOR CLUTTER REDUCTION IN MULTI-DIMENSIONAL VISUALIZATION

3.1 Collect the datasets

The first step for the clutter reduction is to collect the dataset. The dataset can be a collection of images, numerical data etc.

3.2 Extract features and statstical information

The features and statistical information are extracted according to import features and dataset is created.

3.3 Visualize data

After creating the dataset, the data is visualized.

The visualization of data is done before the reduction of clutter from data.

3.4 Implementing Clutter reduction algorithm

Different clutter reduction algorithms have been implemented. Sampling is a widely used technique that is used for clutter reduction. Sampling [2][12] affects the appearance of the data item. The problem with sampling is that it cannot avoid overlapping of data and it does not satisfy the criterion of discriminate points or lines. Clustering is a widely used technique used for clutter reduction.in K-means clustering, can deal with large number of variables, K-means is computationally fast. The disadvantage with K-means is that it does not work well with non-globular clusters. The Yang et al. [9] proposed a visual hierarchical technique that creates meaningful lower dimension spaces with representative dimensions from original data. This technique generates a low dimensional subspace to reduce clutter but the problem with this technique is that some information in original data space is lost. All these approaches cause information loss.

The nearest neighbor and greedy algorithms have been implemented. These algorithms are good for parallel coordinates and scatterplot matrices displays but the problem is to direct measure the dimension relationship. With heuristic algorithm, the work with higher dimension data is easy and with relatively good results. Yang et al. [8] imposed a hierarchical structure over the dimensions. In this the complexity of ordering problem is reduced. However, dimensions are reordered according to one particular measure.

The dimension reduction technique is used for high-dimensional datasets. The dimensional visualization has different techniques to reduce the clutter. This technique works with higher dimensional datasets and makes the visualization better.

3.5 Calculate data abstraction quality

Data abstraction is the process of reducing a large dataset into moderate dataset. The data abstraction quality should be close to zero or one. Zero shows the best abstraction quality.

3.6 Visualize data

After calculating the data abstraction quality, the data is visualized. The data is visualized after reducing the clutter. The goal of visualization is to facilitate the user to gain a qualitative understanding of the information.

TABLE 1. COMPARISON OF DIFFERENT CLUTTER REDUCTION TECHNIQUES.

Sr. no.	Technique	Advantage	Disadvantage
1.	Sampling	Preserves the spatial information and maintains the density much better.	It cannot discriminate points/lines
2.	K-means Clustering	Easy to implement and gives best result in some cases. It	It cannot handle non-globular data of different sizes and den-

		is fast for low dimensional data.	sities and it will not identify outliers. it loses the relative density.
3.	Dimension reordering	It can deal with high-dimensional data with relatively good results	This technique is difficult to implement. It cannot facilitate the interpretation of visualization much better.

4 Conclusion

In this paper, we have proposed the concept of visual clutter reduction using different techniques in multi-dimensional visualization. After implementing these techniques, there is some clutter present in data. One single technique is not useful to reduce the clutter. Employing more techniques can be useful to reduce the clutter. The combination of clutter reduction approaches with dimension reduction or hierarchical data visualization can be used to measure the effectiveness of these techniques. There are many visualization techniques that have not experimented yet. The dimension reduction technique in clutter reduction can be helpful to reduce clutter and it is useful for high-dimensional datasets.

References

[1] Andre Maximo, Ricardo Marroquim, ClaudioEsperanc, Rodrigo Weberdos Santos, Cristiana Bentes and RicardoFarias "High Performance Volume Rendering for 3DHeart Visulization ",. 2009.

[2] A Taxonomy of Clutter Reduction for Information Visualization by Geoffrey Ellis and Alan Dix

[3] I.A. Essa, "Ubiquitous Sensing for Smart and Aware Environments ", *IEEE Personal Communications*, 2000, IEEE

[4] M. Derthick, M.G. Christel, A.G. Hauptmann and H.D. Wactlar, "Constant Density Displays Using Diversity Sampling", *Proc.InfoVis'03*, Seattle, pp. 137-144, 2003, IEEE

[5] A. Dix and G.P. Ellis, "by chance: enhancing interaction with large data sets through statistical sampling", *Proc. AVI'02*, L'Aquila, Italy, pp. 167-176, 2002, ACM Press

[6] G.P. Ellis and A. Dix, "Enabling Automatic Clutter Reduction in Parallel Coordinate Plots", *IEEE Trans. Visualization and Computer Graphics (Proc. InfoVis'06), 12(5)*, pp. 717-723, Sept 2006

[7] M.O. Ward, "A taxonomy of glyph placement strategies for multidimensional data visualization", *Information Visualization, 1*, pp.
194–210, 2002

[8] J. Yang. W. Peng, M.O. Ward, E.A Rundensteiner. Interactive hierarchical dimension ordering, spacing and filtering for exploration of high dimensional datasets. Proc. IEEE Symposium on information visualization, pages 105-112, 2003.

[9] J. Yang. W. Peng, M.O. Ward, E.A Rundensteiner, and S. Huang. Visual hierarchical dimension reduction for exploration of high dimensional datasets. Eurographics/IEEE TCVG Symposium on visualization, pages 19-28, 2003.

[10] Clutter Measurement and Reduction for
Enhanced Information Visualization by Natasha Lloyd, Computer Science December 2005

[11] Measuring Data Abstraction Quality in Multiresolution Visualizations by Qingguang Cui. May 2007.

[12] S. Thompson. Sampling John Wiley and Sons, Inc., New York, 2th edition, 1992.