A Novel System for Abstraction and Visualization of CAD Images

Shelza
CSE Department
SVIET,
Banur (Punjab), India

Balwinder Singh
CSE Department
Yadavindra College of Engineering,
Talwandi Sabo,India

Abstract:-**Because of the fast rate of technological progress, the volume of CAD data is increasing in gigabytes day by day ,which are hard to make sense and difficult to analysis when time as resource is scarce and engineers need to take immediate inferences from CAD data. Now the challenge is to represent the data in such a way that it provides insights, trends and tendencies at glance and easy visualization by changing the variations options dynamically. In our research, we shall try to make visualization Framework that will incorporate clustering based on features of CAD images. Various data abstraction techniques are used in visualization systems to facilitate analysis from overview to detail. Also, analysts can compare different abstraction methods using the abstraction quality measures- histogram difference measure and nearest neighbor measure to see how well relative data density is maintained and select an abstraction method that meets the requirements of their analytic tasks.**

Keywords:- Clustering, HDM, NNM, CAD

I . INTRODUCTION

Very large multivariate datasets are increasingly common in many applications. This proves true for traditional relational databases and complex 2D and 3D multimedia databases that store images, CAD drawings, geographic information, and molecular biology data. We can view relational databases as high-dimensional databases, since the attributes correspond to the dimensions of the data set. The same also holds true for multimedia data. For efficient retrieval, such data must usually be transformed into high-dimensional feature vectors such as color histograms [11] shape descriptors [12] Fourier vectors [13] and text descriptors [14] Many of the mentioned applications rely on very large databases consisting of millions of data objects with several tens to a few hundreds of dimensions.
However, we can apply a number of different data abstraction algorithms to high-dimensional data

A. Data Abstraction

Data abstraction is the process of reducing a large dataset into one of moderate size dataset, reducing the detail of data while maintaining dominant characteristics of the original dataset Techniques for data abstraction found within information visualization include

1)Sampling: Sampling is the process of selecting and using subsets of observations to estimate some parameters about a population. Various sampling techniques are simple random sampling, stratified sampling and quota sampling.. Sampling techniques have been well studied in statistics and widely applied in social science. In Computer Science, sampling is used for many tasks, including optimizing queries in databases with approximate information from samples . In recent years, faced with increasingly dense visualizations, researchers have begun to explore combining sampling with visualization. Random sampling can make the visualization of large datasets more perceptually effective. Their Astral Telescope Visualizer employs a 2D zooming interface to show data with different sampling levels.A non-uniform sampling algorithm to select less data in dense areas to reduce clutter, and more data in sparse areas to maintain data characteristics.

2) Clustering: Clustering is the process of partitioning a dataset into groups of objects based on similarity between objects or proximity according to some distance measure . Each group, called a cluster, consists of objects that are similar among themselves and dissimilar to objects in other groups. Clustering is an aggregation method, since a cluster is regarded as a higher level object that represents all objects It is widely used because of two reasons:
a) By visualizing cluster attributes rather than the original data, the number of visual elements displayed can be greatly reduced;
b) Clustering itself is a pattern discovering process. Thus visualizing clusters can explicitly reveal hidden patterns to viewers. Many visualization systems have adopted clustering methods to reduce clutter and analyze datasets.

Clutter reduction assigns more screen space to interesting data elements than to less interesting ones. Techniques for clutter reduction include
1) Zooming
2) Distortion

Analysts should be able to compare different abstraction methods, and to select an abstraction level for a specific abstraction method

B. Abstraction Quality Measures

1) HDM (Histogram Difference Measure) and
2) NNM (Nearest Neighbour Measure)

HDM:-The HDM is derived based on the average relative error of aggregation used in approximate query processing of databases as well as image similarity measures used in image retrieval.
NNM:- The NNM is derived based on the nearest neighbour algorithm used in pattern recognition and an image quality measure used in image compression.

We integrated these measures into several multivariate visualizations, including parallel coordinates, scatterplot matrices and glyphs, employing two dynamic bar charts to display the measures for the selected and the unselected regions of data. Several interactive operations have been designed for operating in this quality space, including adjusting the data abstraction level, changing the selected regions, regenerating the abstraction, and setting a desired quality level.

Quality measures are recomputed whenever the above operations are performed. The measures and interactions together form an environment in which analysts can explore multi resolution visualization with abstraction quality information available. Data abstraction is the process of reducing a large dataset into one of moderate size, reducing the detail of data while maintaining dominant characteristics of the original dataset. Some data abstraction methods select a subset of the original dataset as the abstraction, such as sampling and filtering, while other data abstraction methods construct a new, more abstract representation, such as clustering and summarizing. Measurement generally refers to the process of estimating the magnitude of a quantitative property [9]. Measurement is essential for scientific research; with measurement, researchers can compare different objects and evaluate the effectiveness of programs or processes

In this section, we will describe two abstraction quality measures in detail. To facilitate explanation of these measures, we define the DAL & DAQ

DAL:-Data Abstraction Level (DAL) as the ratio between the size of the abstracted dataset and the original dataset, and **DQL**:- Data Abstraction Quality (DAQ) as the degree to which the abstracted dataset represents the original dataset.

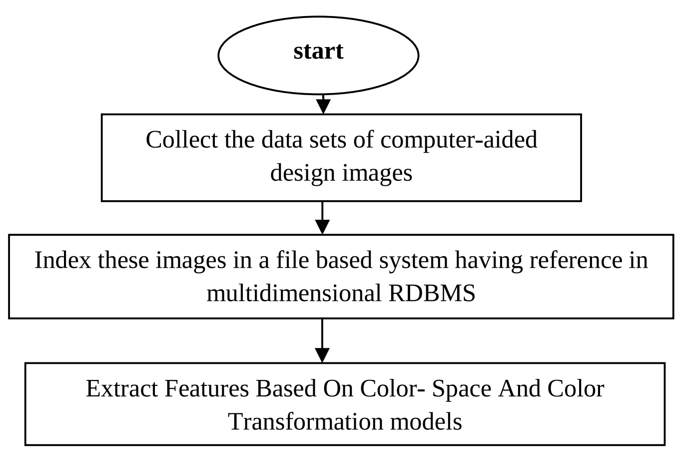
At a given DAL, the DAQ will vary based on the different abstraction methods used or even on different invocations of a given abstraction operation. A good abstraction method should maximize the data abstraction quality and minimize the variance of data abstraction quality in different invocations. The DAL can be considered as a very coarse data abstraction quality measure. Other data abstraction quality measures will, in general, be better descriptors than the DAL.

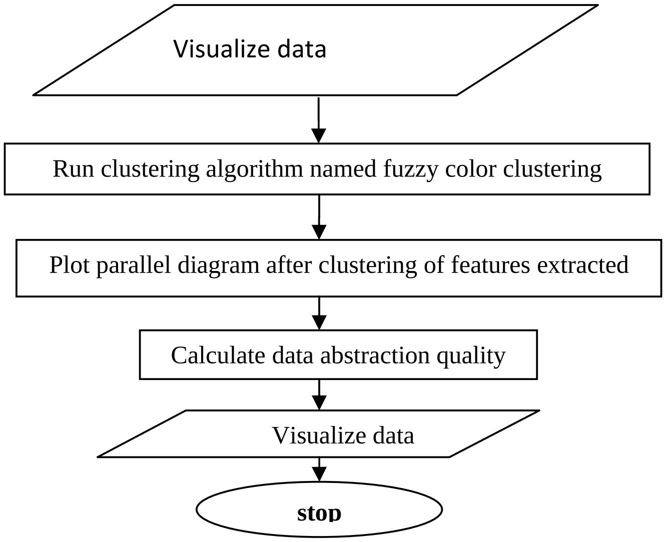
II. RELATED WORK

Several researchers have proposed measures for visualization and data abstraction.Kathrin Anne Meier [2] explored data abstractions based on density estimation. The method is to estimate the density of scientific data sets is based on the directory of a multidimensional data access structure. This data density estimator is called directory estimator. It is based on multidimensional adaptive histograms and is therefore computationally efficient, even for large data sets and many dimensions. They describe the methodology in general and focuses on the estimator's accuracy in particular. The accuracy of the directory estimator depends on the parameters of the access structures used, such as the bucket capacity. They evaluate the choice of bucket capacity theoretically as well as empirically with the ISE (Integrated Squared Error) being the measure of error and using a gridfile as the data access structure..Qingguang Cui,[1] define two data abstraction quality measures for computing the degree to which the abstraction conveys the original dataset: the Histogram Difference Measure and the Nearest Neighbor Measure. They have been integrated within XmdvTool, a public-domain multiresolution visualization system for multivariate data analysis that supports sampling as well as clustering to simplify data. Several interactive operations are provided, including adjusting the data abstraction level, changing selected regions, and setting the acceptable data abstraction quality level. Conducting these operations, analysts can select an optimal data abstraction level. PAULA FREDERICK [9] use appropriate data structures for storing sets of graphical objects ,it can lead to great performance improvements in the visualization of large figures, such as maps and CAD drawings. They present here a study on the performance of persistent data structures composed by an R-tree and V-trees, for storing and efficiently retrieving 2D graphical objects. Results are shown to demonstrate the efficiency of the proposed solutions when applied to large maps. Alexander Hinneburg, [3] describes an advanced clustering algorithm combined with new visualization methods interactively clusters data more effectively. Experiments show these techniques improve the data mining process.Huy [8] describes new emerging abstractions for parallel data processing, in particular computing clouds, can be leveraged to support large-scale data exploration through visualization. They take a first step using MapReduce framework to implement large-scale visualization techniques.

III. METHODOLOGY

First, data set of around 1200 images is created,then features are extracted from images according to color and shape and stored in database. After feature extraction two clustering algorithms are implemented for clustering of features extracted from CAD images and visualization is done for the data set which is abstracted by best clustering algorithm.





A. The Density Based Clustering

Some multidimensional data access structures provide a directory that divides the data space into mutually exclusive regions while organizing the data. The gridfile is an example of such a data structure. In the following we only consider gridfiles, but for all the data structures providing mutually exclusive and exhaustive partitions of the data space a similar theory could be built up.

When inserting the data into a gridfile, the data space will be divided into mutually exclusive regions, called bucket regions, according to a certain split strategy and a fixed bucket capacity. The information about the resulting data space partition and the number of points included in each region is captured in the gridfile directory, which actually maps the bucket regions to a list of bucket addresses. The data space partition can be considered as a collection of bins (multidimensional intervals) and the number of points per region as the counts

Given a sample of *n* k-dimensional observations **X** = **{x1,**. . . , **xn,}** whose underlying density is to be estimated, the *directory estimator* is formed by inserting the points of **X** in a gridfile and counting the number of points **xi,** in each bucket region of the gridfile directory.

Given a split strategy and a fixed bucket capacity, the bucket regions, represented by the k-dimensional intervals $B_i = [n_{i1}, \mathbf{v_{i1}})$x $[n_{i2}, v_{i2})x. . .$ x $[n_{ik}, v_{ik})$ $(\boldsymbol{i} = 1,.. . ,\boldsymbol{b})$(where **ni,** and **vi,** denote the near and far vertices of the bucket region *B*, and *b* is the total number of bucket regions), define the partition of the data space. Note that **ni** and **vi,** are both expressed in coordinates of the data space, i.e. the gridfile directory information together with the gridfile scale information is used to compute **ni,** and **vi,.** The *directory estimator* normalized to 1 is then defined as:

f(x)=1 /n (number of xi in same bucket as x/volume of the bucket region containing xi)

The directory estimator is, in principle, a *multivariate adaptive histogram estimator.* f is a density in **X,** i.e. it is nonnegative and integrates to one. The directory estimator is determined by the sample **X** = **{xi…….xn} fr**om f(x) and the gridfile directory.

B. K-Means

K-Means is a well-known partitioning algorithm .Objects are classified as belonging to one of the k groups k chosen is priori cluster membership is determined by calculating the centroid for each group .and assigning each object to the group with the closet centroid. This approach minimizes the overall with in- cluster dispersion by iterative reallocation5 of clusters members

Data abstraction quality is calculated for both clustering methods & best clustering algorithm is chosen based on higher abstraction quality and visualization technique is applied on data set

Parallel Co-Ordinates:-

It is a visualization technique in which different classes identified by clustering algorithm based on distance classifier are visualized with different colors. A parallel co-ordinate plot basically gives holistic view of how data is lying semantically in the complete data set. It works on the principle of vector space model in which a multi class is plotted against no. of data points.

Visual analysts can benefit from data abstraction quality measures in several ways.

1), these measures give analysts a confidence level in the given abstraction they work with and thus also for any observation made based on the abstracted dataset. This enables them to make more accurate decisions.

2) These measures make analysts aware of the abstraction quality of the dataset. Better yet, interactive mechanisms are available for the analysts to control the abstraction quality. Thus they can find an appropriate abstraction level by trading off the accuracy of representing the data subset and the degree of visual clutter.

3) These measures can be used to compare the effectiveness of different abstraction methods. Analysts can thus select the abstraction method via a compromise between the relative data density, the degree to which outliers are preserved, and response time

IV. CONCLUSION

The methodology presented in this paper open a novel set of tools and possibilities for data abstraction and visualization. Volume of CAD data is very large, the ability of researchers in the scientific and engineering community to generate or acquire data far outstrips their ability to analyze it. This problem is even more pronounced when the data is of high dimensionality. Visualization has been identified **as** a critical technique for exploring data sets and for this best abstraction technique is chosen based upon data abstraction quality from the number of available data abstraction techniques.

REFERENCES

[1] .Qingguang cui .Mathew O. ward ,Elke A . rundensteiner & Jing jang . measuring data abstraction quality in multiresolution visualizations. IEEE transactions on visualization and computer graphics,vol. 12 ,NO. 5 ,pages 709-716 ,2006
[2] Kathrin Anne Meier . data abstraction through density estimation by storage management . proc. IEEE,pages 39-47. 1997
[3] Alexander Hinneburg , Daniel A. Keim and Markus Wawryniuk . hd-eye : visual mining of high – dimensional data. IEEE computer graphics and applications , pages 22-31 , 1999
[4] Jianbing Huang,and Michael B. Carter, Interactive Transparency Rendering for Large CAD Models ,IEEE transactions on visualization and computer graphics, VOL. 11, NO. 5, pages 584-595, 2005
[5] Pierre Georgel, Pierre Schroeder and Nassir Navab , Navigation Tools for Augmented CAD Viewing,IEEE Computer graphics and applications,2009
[6] Kai-Mo Hu, Bin Wang, Bin Yuan and Jun-Hai Yong.,Automatic Generation of Canonical Views for CAD Models,12th International Conference on Computer-Aided Design and Computer Graphics,pages 17-24 ,2011
[7] Jeffrey LeBlanc, Matthew 0. Ward, norman wittels, Exploring N-Dimensional databases,IEEE, pages 230-237 ,1990
[8] Huy T. Vo', Jonathan Bronson ,Brian Summa ,Joao L.D. Comba, Juliana Freire, Parallel Visualization on Large Clusters using Map Reduce,IEEE Symposium on Large Data Analysis and Visualization ,pages 81-88 ,2011
[9] .**P**aula frederick, marcelo gattass, mauricio riguette mediano, Efficient Visualization of Graphical Objects,IEEE,2002
[10] A. Inselberg and B. Dimsdale, “Parallel Coordinates: A Tool for Visualizing Multidimensional Geometry,” *Proc. Visualization 1990*, IEEE CS Press, Los Alamitos, Calif., 1990, pp. 361-370.
[11] H. Shawney and J. Hafner, “Efficient Color Histogram Indexing,” *Proc. Int’l Conf. on Image Processing*, IEEE Press, Piscataway, N.J., 1994, pp. 66-70.
[12] R. Methrotra and J.E. Gray, “Feature-Index-Based Similar Shape Retrieval,” *Proc. 3rd Working Conf. on Visual Database Systems*, Chapman and Hall, London, 1995, pp. 46-65.
[13] T. Wallace and P. Wintz, “An Efficient Three-Dimensional Aircraft Recognition Algorithm Using Normalized Fourier Descriptors,” *Computer Graphics and Image Processing*, Academic Press, Vol. 13, 1980, pp. 99-126.
[14] K. Kukich, “Techniques for Automatically Correcting Words in Text,” *ACM Computing Surveys*, Vol. 24, No. 4, 1992, pp. 377-440.

AUTHORS PROFILE

Er.Shelza is currently working as Lecturer in Computer Science and Engineering Department at Swami Vivekanand Institute of Engineering and Technology, Banur. She is pursuing M.Tech in Computer Engg. from Yadwindra College of Engineering and Technology, Talwandi Sabo affilated to Punjabi University Patiala .she holds the degree of B.Tech in Computer Science and Technology from Sant longowal institute of Engineering & Technology, Longowal Distt Sangrur,Punjab.She has more than 7 years of teaching experience. Her research interest included programming and data mining.She has authored papers in National Conference and has attended short term course at NITTTR,Chandigarh

Er. Balwinder Singh is working as Assistant Professor (Computer Science) at Yadavindra College of Engineering, Punjabi University Guru Kashi Campus at Talwandi Sabo (India). He received B.E. degree from Guru Nanak Dev Engg. College, Ludhiana (India) and M.Tech. degree from Department of Computer Science & Engineering, Punjabi University, Patiala (India). He has more than 10 years of teaching and research experience. His research interests include Image Processing, Programming and Natural Language Processing.