Journal of Engineering Sciences
Vol 11, Issue 6,June/2020
ISSN NO:0377-9254

IDENTIFICATION OF PATTERN CHANGES IN POINT CLOUD DATA FOR COMPONENT FAULT DETECTION

Mrs.Prince Sahaya Brighty[1],Sowmiya R[2] , Saumini B[3], Swathy Renganathan[4]
[1]*Assistant professor ,Ramakrishna Engineering College, Coimbatore, Tamil Nadu, India*
[2,3,4]*Student,Sri Ramakrishna Engineering College, Coimbatore, Tamil Nadu, India*

***Abstract* - In recent decade all the design and manufacturing industries have been shifting from manual design plans to computerized aided designing and manufacturing due to the rise of digitization. The digital world comes in closer to the real-life environment with the advancement of technologies like point cloud technology where we can do 3D modelling, create virtual environment, digital designing and also finding the flaws or defect of any object with more detail and precision. These papers aim to find similarities and identify pattern changes to find any fault in the component with high dimensional point cloud data taken at a constant interval time. This paper utilizes several clustering techniques and multi-data visualization techniques for an efficient comparison across different point cloud data for inferring the changes in the component.**

***Keywords*-Distance measure, Clustering, High dimensionality, Point cloud data**

1. INTRODUCTION

Point cloud data in general are set of points prevailing in space. It stores a huge amount of data that provides intricate detailing of any object or any form of geographical space it has surveyed.

Point cloud data are produced using various 3D scanners with equipment of lasers and remote sensors and LIDAR imaging .In most of the scenarios the object is scanned by scanners emitting the laser beam from a particular dimension while in our the project we have used high dimensional point cloud data, i.e. the object is scanned at varying dimensions in order to get all the minute and precise details of the component. Due to its higher dimensions, it becomes difficult to interpret the object as a physical component. Point cloud data holds huge amounts of data with each point having its own property and attributes. Point clouds are one of the essential form of representation in various fields like in various fields like construction, real-world mapping, architecture, digital design and manufacturing, gaming, augmented reality.

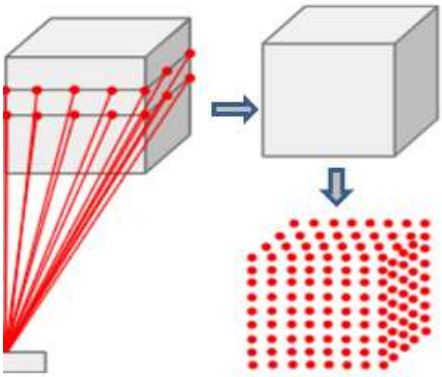

Figure 1 Formation of point cloud data

With the advent of technology, the amount of data collected and processed for any particular application has become immense, and for more unambiguous data we opt for higher-dimensional data. Such high-dimensional data spaces are often encountered in areas such as medicine or biology, where DNA micro array technology can produce a large number of measurements at once, the clustering of text documents, where, if a word-frequency vector is used, the number of dimensions equals the size of the dictionary, and many others, including data integration and management, and social network analysis[1]. For industries like design and machinery it becomes very essential for their product to be clear cut and precise henceforth in this application, we use a point cloud data in higher dimensional space for efficient analysis of an object for better accuracy in finding any defect or flaw if any present.

2. RELATED WORK

[2]Arrays and huge simulation from sensing devices with the huge computing resources resulted in large complex high dimensional data set. Visualization holds important space in exploring data sets. Guidance is made for data practitioner for the navigation over the modular view of recent

inventions inspiring creation of new visualization along with future chances for visualization based research.[3] A large-scale dataset is usually said to be present with high dimensionality. Yet the point cloud possesses a structure that will decrement their intrinsic dimensionality because clusters points are located close with low-dimensional varieties or fine-grained significantly. A test over a recently introduced dimensionality estimator was done, also based on analysing the separability properties of data points, on several constraints and real biological datasets were also used. From this it is illustrated that measure of ID has performance will compete with state of art measures even when having efficiency across a wide range of dimensions also with the performance that will be better in the case of noisy samples. Moreover, permission was allowed to estimate the intrinsic dimension in all situation where the intrinsic manifold the assumption wasn't valid. [4] Computational geometry and topology like fields are in which there is a high potential for analysis of arbitrary high dimensional dataset. To get insight over structure and complexity of dataset, we would preferably plot corresponding point cloud but for orthogonal scatter plot this will work over three dimension like parallel coordinates. There is a the solution to tackle the problem of visualizing point cloud indirectly through having a view over topology of their density distribution. By this approach topology was computed in arbitrary dimensions. For this geometric or topological method analysis it requires generated point cloud data set from original dataset .[5] Cluster analysis divides data into forms of clusters, so that understanding is made more transparent while this clustering as a huge history and the huge amount of clustering methods along with some other fields like data mining pattern recognition. It emphasises on the methods in cluster analysis that focus over challenge of clustering high dimensional data with great knowledge on concept-based clustering.

3. METHODOGY

In general, point cloud data usually contains a set of (x ,y, z) coordinates which is restricted to a three-dimensional model in space. In our problem, we adapt to a higher level of dimensionality of the data. Due to the dimensions there occurs a phenomenon called the curse of dimensionality which is caused to the exponential increase of data in a particular space where there can also be noisy and irrelevant data included. Hence, it is necessary to clean and organize the obtained point cloud data into a particular structure. When we scan the component at the varying point of time the data points generated may differ each time due to environmental factors while the dimensions remain intact. The below figure explains the workflow of this project:

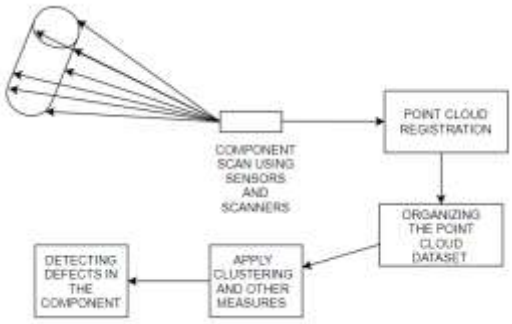

Figure 2work flow

The first phase involves scanning the component using essential techniques like the LIDAR technology, Photogrammetry and 3D scanners . The point cloud acquisition time and processing depends on the level of accuracy we intent for the particular task . Multiple scans might be required to achieve a wholesome coverage of the object. These scans are amalgamated together in the process of *point cloud registration.*

The next stage of the project is organizing the point cloud data set in which we structure the raw point cloud data into an understandable form of a high dimensional data matrix where each column represents a dimension with each row consider as one point in space with several dimensions. The component is scanned at multiple intervals of time henceforth we will get several datasets with the same dimensions yet varying number of points.

Next phase is a crucial part where we apply techniques to order to find similarities and finding pattern changes in the varying datasets. Below are two methods implicated in this paper:

A. DISTANCE METRICS

The distance metric is used to find similarity amongst data points using mathematical distance functions. Therefore, closing the distance measure to zero the more similarity is inferred from the set of data points. The correctness of the metric varies for each application. The distance measure forms as a proximity matrix which contains pairwise similarities

between the component being considered [5]. Some commonly used distance metrics are:

- Manhattan Distance
- Euclidean Distance
- Cosine Similarity
- Chebyshev Distance
- Minkowski Distance
- Correlation

For a higher level of dimensions the L1 and L2 distance metric that is the Manhattan and Euclidean distance is much preferable [6]. Each dimension is taken into account separately to achieve a clarity of the object is taken.

B. HIGH DIMENSIONAL CLUSTERING

The purpose of clustering is to divide the data into groups based on the norms on which you want to group them. In our scenario, we cluster data points based on their similarity trait. Clustering analysis is used in various fields like pattern recognition, machine learning, statistics, big data analysis whereas high dimensional clustering can also be applied to the field of bioscience to understand the genes and proteins of humans and also used for efficient data-based retrieval on large scale. In our the application we apply clustering on point cloud data of a component to get pattern changes across varying datasets are taken at different points of time. Some of the clustering techniques implied in this paper are :-

K-means clustering: It creates K clusters based on centroid and associating data points closer to it.

Grid based clustering: Divides the whole space into grid and merges data to generate clusters.

Density Based clustering: locates regions and separate themselves as lower and higher dense regions with the use of epsilon value.

Hierarchical clustering: Step by step clustering process creating nested clusters from all data points.

4. IMPLEMENTATION

The implementation of this application is done using MATLAB. Various clustering method has experimented with the varying point cloud datasets. The histogram is plotted using a distance metric of two adjacent component datasets to find the similarity and the ratio of dissimilarity the below figure is taken with Euclidean Distance metric. Clustering algorithm like k-means, mean shift clustering, hierarchical, DBSCAN, grid-based clustering have been implemented. All the dimensions cannot be plotted individually as the models are restricted up to 3d model. There are multi-data visualization techniques to view higher dimensional data like the heat map which indicates the data point with the variation of colour intensity.

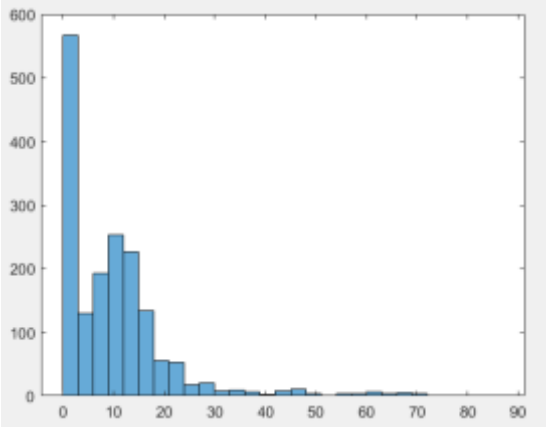

Figure 3Histogram Analysis

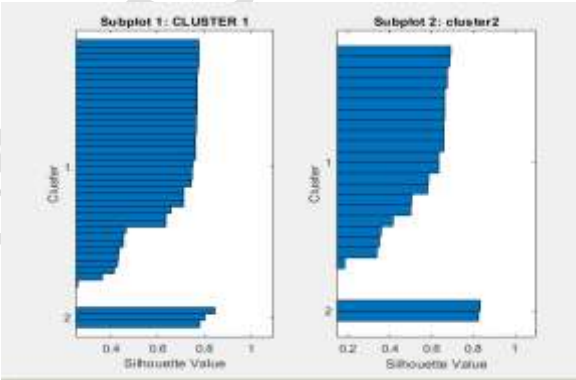

Figure 4 k-means silhouette plot

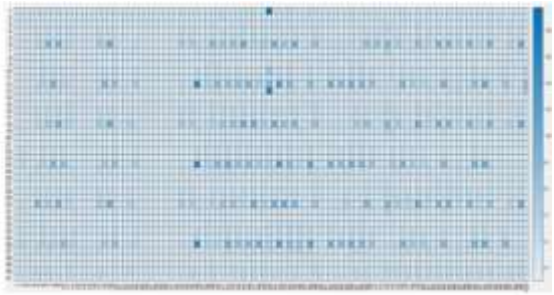

Figure 5HeatMap

5. CONCLUSION

The Paper aims to find pattern changes in order to find the defect and flaws of a component. High enduring dimensional point cloud data has been captured and used for better accuracy. Various techniques are deployed to find the defects. K-means and mean-shift clustering algorithms as well as Euclidean and Manhattan distance metrics have

found to be more accurate for this particular application. For Design and manufacturing industries precision is very primal for the component they produce, hence this methodology can be implicated for better accurate outcomes.

REFERENCES

[1] International Workshop on
Clustering High-Dimensional Data
Palazzo Serra di Cassano, Via Monte di Dio 14, Naples (Italy), May 15th, 2012

[2] Visualizing High-Dimensional Data:
Advances in the Past Decade
Shusen Liu*, Dan Maljovec*, Bei Wang, Member, IEEE,Peer-Timo Bremer, Member, IEEE, Valerio Pascucci, Member, IEEE

[3] Estimating the effective dimension of large biological datasets using Fisher separability analysis
Luca Albergante, Jonathan Bac, Andrei Zinovyev

[4] "On the Generation of Point Cloud Data Sets: Step One in the Knowledge Discovery Process", Andreas Holzinger, Bernd Malle, Marcus Bloice, Marco Wiltgen, Massimo Ferri, Ignazio Stanganelli, and Rainer Hofmann-Wellenhof

[5] "Visual Analysis of High Dimensional Point Cloud using Topological Landscapes" , Patrick Oesterling, Christian Heine , Heike Janicke Gerik Scheuermann

[6] The Challenges of Clustering High Dimensional Data , Michael Steinbach, Levent Ertöz, and Vipin Kumar

[7] On the Surprising Behavior of Distance Metrics in High Dimensional Space
Charu C. Aggarwal, Alexander Hinneburg, and Daniel A. Keim