Traffic Data Analysis and Visualization via Iterative Data Image Rotated Bar Graph(iDIRBrG) Method

Jona G. Beceral, Karmina Rosette C. Cabreros,
Richelle Ann B. Juayong, Jasmine A. Malinao, and Henry N. Adorna
Department of Computer Science (Algorithms and Complexity)
University of the Philippines Diliman
Diliman, Quezon City

ABSTRACT

In this paper, we present a visualization method known as the Iterative Data Image Rotated Bar Graph (iDIRBrG) Method that is capable of handling visualizations for large time series traffic data set. This method is especially designed for visualizing data sets having the following characteristics: time series, periodic and multidimensional. It provides a novel way of constructing effective 2D visualizations of such data sets for a more efficient time domain data analysis against previous literature. Even in the absence of complex data transformations, such as Data Signature Construction Methods, Fast Fourier Transforms, Principal Components Analysis, etc., iDIRBrG-based visualizations can highlight peculiarities, irregularities, and patterns that may have been previously unknown to domain users. To test the effectiveness of the new method, we shall use the 2006 North Luzon Expressway (NLEX) traffic volume data set provided by the Manila North Tollways Corporation (MNTC) - Tollways Management Corporation (TMC) through the National Center for Transportation Studies (NCTS).

With the iDIRBrG Method-based investigation done on this data set, we are able to perceive the entirety of the data set in a simple yet comprehensive way. By employing the XMeans clustering algorithm, we extract clusters from the data set. As the iDIRBrG Method possesses an iterative property, we further investigate each of these clusters and perform reclustering within them. By doing so, we extracted very interesting information, detected potential and unexpected outliers from the data set. We discover new patterns and trends, establish relationships across clusters and as well as among the data points. The results are then validated by the experts from NCTS.

Keywords: *iDIRBrG Method, data visualization, traffic data, multidimensional, time series, periodic*

*Funded by an ERDT Project entitled *Information Visualization via Data Signatures*

1. INTRODUCTION

Visualization[1, 3] is a subject matter that can be traced back in our history, from primitive paintings on clays, maps on walls, photographs, and table of numbers. The evolution of visualization was the result of human's increasing needs and advancement in technology.

Visualization is a method that transforms information into pictures. It is the process of transforming objects and concepts into a form visible to the human eyes [2]. There are several information that can be visualized, however this study concentrates on visualizing traffic data.

Needless to say, pictures are more straightforward in projecting interesting behaviors than its numerical equivalent. This is where data visualization takes place. **Data Visualization** is the process of transforming numerical information into visual form enabling users to observe the information through patterns, trends and relationships revealed from the data [3].

Building methods for traffic data analysis is the heart of this study. The complexity of analyzing this type of data comes from its nature of being a) time series, b) multidimensional (with values taken in terms of traffic volume, vehicle speed, weight and length of vehicles, etc.), and c) huge magnitude. It is well known that working with multidimensional data is challenging given that you have to project all the important dimensions and at the same time capture patterns and relationships between these dimensions.

2. DEFINITIONS

2.1 The Data set

This study uses 2006 North Luzon Expressway(NLEX) traffic data obtained from the Manila North Tollways Corporation(MNTC) - Tollways Management Corporation (TMC) through the National Center for Transportation Studies (NCTS), taken at an hourly basis for all 52 weeks of the year for four lanes on both Southbound and Northbound. The data as mentioned is time-series, periodic and multidimensional having the following dimensions (a) total volume of vehicles per hour, (b) spot speed of a vehicle in an hour, (c) volume distribution by length and weight, (d) traffic incidents, and (e) traffic accidents.

2.2 Data Preprocessing

Data preprocessing is done primarily to clean the data and ensure the validity of its values to provide more accurate results and analysis, as well as the partitioning of the data set for visualization purposes. This process is done in each of the dimension of the data set, particularly to those dimensions that show evidence of inaccuracy. Problems handled by this process include having missing values in the data set and having invalid values (values less than the set threshold). To do these task, the knowledge from the domain experts of NCTS was taken.

Let matrix A contain the data set for a certain dimension. The column corresponds to an observation taken from one whole year. Each row corresponds to the value at a certain time, note that each observation was taken at an hourly basis. MIN is the specified minimum value. Using expert knowledge, missing/invalid value $A[i][j]$ is derived using the following formula.

$$A[i][j] = \begin{cases} \frac{A[i-1][j]+A[i+1][j]}{2}, if A[i-1][j] \; and \; A[i+1][j] \\ has \; value \; and \; their \; sum \; is > (MIN * 2), \\ \\ otherwise, remove \; the \; entire \; week \; it \; belongs. \end{cases} \tag{1}$$

By applying (1) we may expect to have reduced the number of weeks in that selected dimension, which applies to speed in our case. By doing so, we shall also remove the data points in all other dimensions where the values of speed had been eliminated in the process.

2.3 Xmeans Clustering Algorithm

X-means is a new algorithm that is built from blacklisting that maintains a list of centroids that need to be considered for a given region. It estimates K(the number of clusters) rapidly, making local decisions about which subset of current centroids should split themselves in order to better fit the data [4]. The algorithm is comprised of the following steps:

1. $\mathbf{Improve - Params}$.
   The Improve-Params operation consists of running conventional K-means until the set of chosen centroids do not change in the succeeding iterations.

2. $\mathbf{Improve - Structure}$.
   The Improve-Structure operation finds out if and where the new centroids should appear. This is achieved by letting some centroids split in two.

3. If $K > K_{max}$, the algorithm stops and reports the best scoring model (through the Bayesian Information Criterion) found during the search. Otherwise, proceed to Step 1.

2.4 Data Image

Data image has been suggested as an approach of mapping data attributes to color features for visualization and exploration of higher dimensional data [5]. It is used to easily distinguish elements according to their color for data analysis and identification of clusters as well as detecting outliers.

3. IDIRBRG METHOD

The Iterative Data Image Rotated Bar Graph(iDIRBrG) Method is a new visualization technique introduced in this study for traffic data analysis. It is formulated by studying the strengths and weaknesses of existing methods such as bar graphs, data image, calendar view, 3D sequential graphs to be able to make effective visualization of traffic data.

More specifically, the iDIRBrG method is capable of doing the following tasks:

- Handling and visualizing of traffic data which is time series, periodic and multidimensional.
- Visualizing all the lanes simultaneously for easier analysis.
- Visualizing the clusters into one single graph to better show their similarities and differences.
- Highlighting peculiar patterns or sets of behaviour to detect outliers in the data set.

3.1 Initial Visualization

The iDIRBrG Method starts traffic data analysis by first projecting the entire traffic data set as a data image to provide a birds-eye-view of the characteristics of weeks obtained by preprocessing of the original data set per dimension. The horizontal axis corresponds to the temporal aspect of the data set. With the partitioning done during the preprocessing step, this axis shall correspond to 168 time steps. This is equivalent to the periods of data collection starting from Sunday at 1200H to Saturday 2359H for each dimension. The vertical axis corresponds to remaining weeks arranged in ascending manner from bottom to top of the axis. Colors used in the data image for each week are in correspondence with the overall magnitude of the data taken at each specific time step in each dimension. The overall magnitude may mean the total of a certain dimension taken as the sum of individual values of this dimension from all lanes in the data set. (This can also pertain to the average value of the lanes for the dimension speed). Thus, a row in the data image shall show the trend of the values for all days of the week in a dimension of the data set. A column in the data image shall show the differences of the values for all weeks in the year collected at a specific time of the day. Figure 1 shows an initial visualization example.

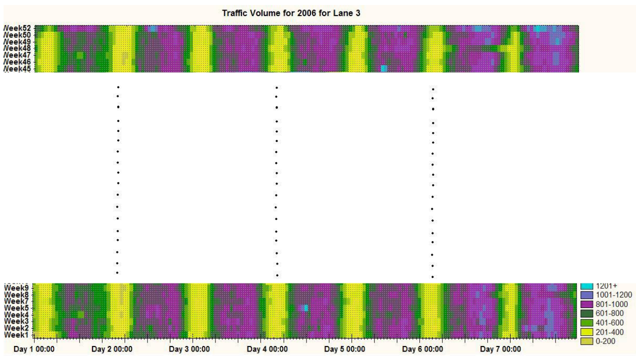

Figure 1: Data Image Initial Visualization Example

3.2 Refining Initial Visualization via X-Means Clustering

The initial visualization provides the user a very rough view of the entire data set for a given dimension. This visualization automatically highlights individual weeks which tend to follow their own unique pattern compared to others providing the users an insight of behavior of the outliers in the data set. However, this visualization still needs to be refined to effectively show trends for a number of weeks in the data set to derive a full characterization of the traffic flow behavior in a given year.

To address this concern, we use the clusters obtained from X-means algorithm to refine and restructure the data image in the initial visualization. The weeks belonging to one cluster are structured contiguously in the data image, thus simplifying the task for cluster analysis. At this stage, domain experts can easily discriminate the overall behavior of one cluster against the other and pinpoint range of values that each cluster takes through time. Furthermore, this visualization can already explain why certain weeks were assigned or not assigned to a specific cluster in terms of trends observed in them.

3.3 Linking Visualization to Traffic Accidents and Business Decisions

By the new data image obtained by using the results of X-Means, special instances such as an unusually prolonged set of large (or small) and unexpected values for a given dimension can now be scrutinized. By looking at the time frames wherein these instances occur, we find rationale to them by zooming in to that part of the data set to produce a visualization of the specific subset of data we are interested in. For this case, we project each individual lane as a rotated bar graph showing this subset of the data as shown in Figure 2. The horizontal axis corresponds to the overall magnitude of the data taken at each specific time step in each dimension. Each bar represents each time step, thus the vertical axis corresponds to 168 time steps, the periods of data collection starting from Sunday at 1200H to Saturday 2359H. The color of the bars tell us overall magnitude of another dimension.

By looking at simultaneous events found in all lanes as seen in Figure 2, domain experts would be able to ascertain probable reasons to justify what has been previously observed as an area of interest. Prolonged large values of traffic volume may point to possible key areas in an expressway where accidents frequently occur or may have problematic sections inducing constant heavy traffic. Thus, business decisions pertaining to doing alterations to a current traffic system can be drawn out to address these concerns.

3.4 Visualization as an Iterative Process for Traffic Data Analysis

The iDIRBrG method incorporates the iterative method of visualization by drawing out subclusters for each known cluster in the data set. In a case wherein users would opt to probe more of a cluster's characteristics, X-means shall again be performed with an input equal to the cluster itself. After doing such, users shall perform again the subsequent tasks indicated above to perform effective and efficient traffic data analysis.

4. EMPIRICAL TESTING

The method was tested using the 2006 NLEX Balintawak Northbound (BLK-NB) traffic volume data set provided. The data set used for the following results is 2006 NLEX traffic volume data composed of 52 weeks. The values are taken on an hourly basis through an automated detector. Initially, we focused on visualizing volume and speed of the data. Learning from initial investigations that the dimension speed is erroneous because of inaccurate measurements, the speed data undergo the preprocessing step so as to ensure that every value is valid. As a result, 42 out of 52 weeks retained for further analysis. And as stated earlier, the weeks eliminated from speed data should also be removed in all other dimensions such as volume. Table 1 shows the summary of the data set after preprocessing. Those in red represents the weeks deleted and will be disregarded during analysis.

In order to reveal the patterns and normal trends of the data, we visualize its total volume that is partitioned into weeks. Its value corresponds to the sum of all lanes in a particular time. We then applied X-means to the data so that weeks with similar behavior

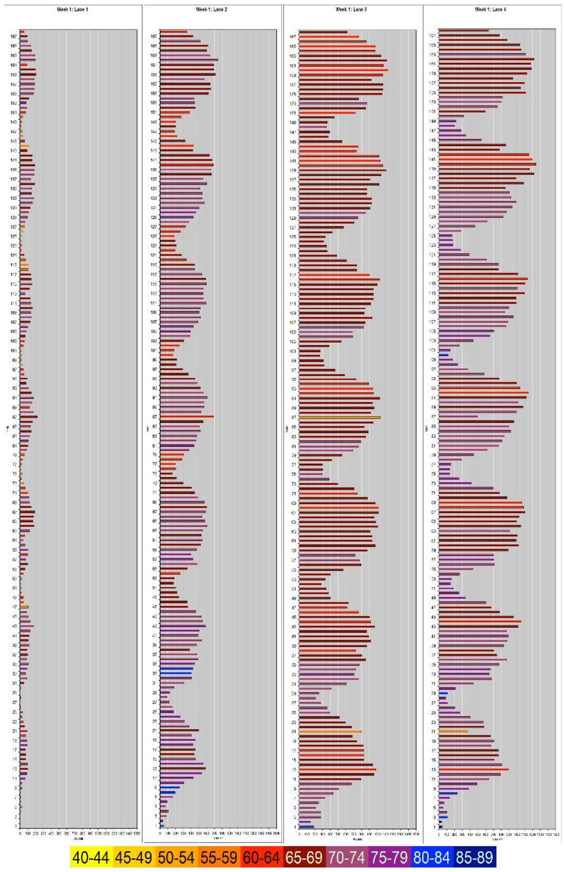

Figure 2: Sample Visualization of Rotated Bar Graph showing Lane 1 to Lane 4 simultaneously from Left to Right

Mo(s).	Wk	Date	Mo(s).	Wk	Date
Jan	1	01-07	(Jun-Jul)	27	02-08
	2	08-14		28	09-15
	3	15-21		29	16-22
	4	22-28		30	23-29
Jan-Feb	5	29-04	Jul-Aug	31	30-05
	6	05-11		32	06-12
	7	12-18		33	13-19
	8	19-25		34	20-26
Feb-Mar	9	26-04	Aug-Sep	35	27-02
	10	05-11		36	03-09
	11	12-18		37	10-16
	12	19-25		38	17-23
Mar-Apr	13	26-01		39	24-30
	14	02-08	Oct	40	01-07
	15	09-15		41	08-14
	16	16-22		42	15-21
	17	23-29		43	22-28
Apr-May	18	30-06	Oct-Nov	44	29-04
	19	07-13		45	05-11
	20	14-20		46	12-18
	21	21-27		47	19-25
May-Jun	22	28-03	Nov-Dec	48	26-02
	23	04-10		49	03-09
	24	11-17		50	10-16
	25	18-24		51	17-23
Jun-Jul	26	25-01		52	24-30

Table 1: 52 Weeks of 2006. The weeks colored in red are the weeks eliminated during preprocessing.

group together. The algorithm then partitioned the data set into 3 disjoint clusters as seen in Table 2.

Cluster	Weeks
1	wk15
2	wk1,wk2,wk3,wk4,wk5,wk7,wk8,wk9,wk10,
	wk11,wk12,wk25,wk27,wk28,wk29,wk30,
	wk31,wk32,wk33,wk35,wk37,wk38,wk39,
	wk41,wk42,wk44,wk45,wk46,wk47,wk48
3	wk13,wk14,wk16,wk18,wk19,wk20,wk23,
	wk43,wk49,wk50,wk52

Table 2: Clusters obtained using Xmeans

Next, to simultaneously visualize volume and speed of all four lanes, another technique - Rotated Bar Graph is used. Through this, the unusual behavior seen from the Data Image can be verified thus giving the domain users more additional information regarding the behavior.

From the original cluster it was noticed that there were some weeks that showed behavior different from the other weeks in the cluster where it belonged. Hence re-clustering was done to provide more detailed information for better analysis. Table 3 shows reclustering results done on Cluster 2 and 3. Results from these steps were visualized, analyzed and validated by the domain experts.

5. RESULTS AND DISCUSSION

Initially, we obtain the data image of the total volume from the traffic data that has been preprocessed. From the first image acquired, Figure 3, we would be able to see the patterns and behaviors of the

Cluster	Weeks
1	wk15
2	
(2.1)	wk1,wk27,wk29,wk30,wk31,wk32,wk33
	wk35
(2.2)	wk2,wk3,wk4,wk5,wk7,wk8,wk9,wk10,
	wk11,wk12,wk25,wk37,wk38
(2.3)	wk28,wk39
(2.4)	wk41,wk42,wk45,wk46,wk47
(2.5)	wk44
(2.6)	wk48
3	
(3.1)	wk13,wk14,wk16,wk18,wk19,wk20,wk23,
	wk43,wk49,wk50
(3.2)	wk52

Table 3: Subclusters obtained after reclustering

remaining 42 weeks, but this result only presents limited information, thus we use the X-means clustering algorithm to our data and obtain 3 disjoint clusters.

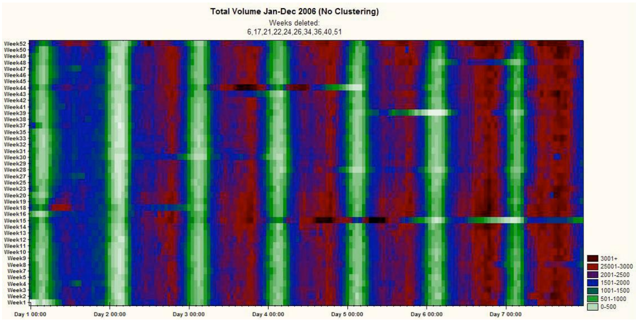

Figure 3: Total Volume of 2006 NLEX BLK-NB data set

We restructure the initial data image visualization such that members of a specific cluster are contiguously shown in the new visualization. Figure 4 shows this visualization, and by examining more closely the second Data Image, we can deduce the following analysis:

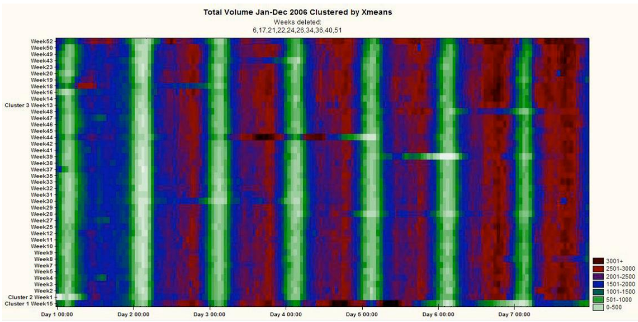

Figure 4: 2006 NLEX BLK-NB Total Volume clustered using X-means

Cluster 1 shows a decrease in volume from the mean values (1501-2000). For the majority of the days of this week, there is a significant increase in volume except for Day 6 compared to most weeks in the data set. Largest values occurred on days 4 and 5 compared to Cluster 2 and Cluster 1. From late Day 5 to Day 7 there is a prolonged occurrence of low values compared to Cluster

2 and 3 and prolonged high values from mid Day 4 to mid Day 5. Week 15 as shown in Table 1 is the Holy Week (in Catholic beliefs) for the year 2006. This week had been known as an outlier due to the unique trend it possesses due to this religious event.

Cluster 2 shows the regular traffic flow for volume. The values scale up per day until Saturday. It also follows the normal increasing trends from Day 1 to Day 7. There is also prolonged occurrence of high values during mid Day 3 and start of Day 4. While prolonged mean values (1501-2000) occurred during mid Day 4 compared with other weeks. One notable week is Week 44 because even though it has been found to be an outlier from the previous clustering algorithm (K-means) it still grouped with Cluster 2 because most of the days in week 44 followed the regular trend for clustering. From the visualization it is easier to recognize that Week 44 has a unique behaviour. Another significant week is Week 30, because from previous clustering we know that Week 30 has a set of low values from Day 2 to Day 3, thus it is considered an outlier. But for X-means it was grouped to Cluster 2 because majority of its behaviour again follows the trend in Cluster 2.

Cluster 3 presents largest values that are consistent from the start until the end of the week. Again, from our previous clustering algorithm we have concluded that the largest set of values comes from Week 51(which is deleted) and Week 52, and X-means validated these conclusions. One notable week in Cluster 3 is Week 18 where there is an increase of volume during Day 1 and low values for Day 2. It seems that the increase was because of the incoming Labor Day in Day 2. It can also be predicted that Day 2 has low values because of Labor Day vacation. Another interesting week to consider is Week 43. It has a set of low values for all lanes that you can verify via data image of individual lanes. Thus, as seen from the Data Image of total volume, it showed a significant difference compared to the weeks in its cluster. However, majority of its trend still follows the behavior of its comembers in Cluster 3.

Bearing in mind the weeks that demonstrate unusual behaviors, we would want to further examine it and scrutinize what had happened in each lane of NLEX on those weeks. Thus we construct another set of data images that will show the activities of vehicles for each lane. We sample Lane 2 and Lane 4 traffic volume and visualize their traffic volume for the entire 2006 NLEX Balintawak Northbound data set as seen in Figures 5 and 6.

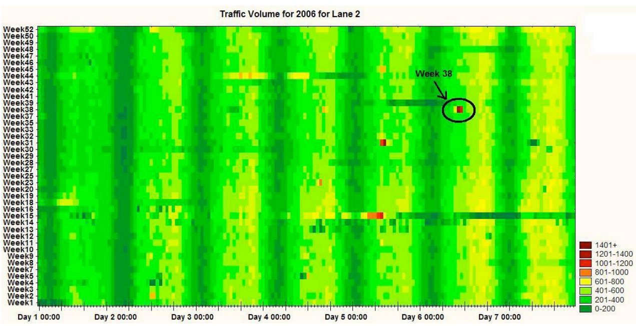

Figure 5: 2006 NLEX BLK-NB Traffic Volume for Lane 2

Notice that the data image produces a sensible amount of information that is beneficial in doing analysis, but since traffic data is multidimensional in nature thus we introduce Rotated Bar Graphs. These are capable of visualizing two dimensions at the same time. In our case, volume and speed were the chosen dimensions. Along with its capability to handle two dimensional data, Rotated Bar Graphs also shows all four lanes in just one visualization. This

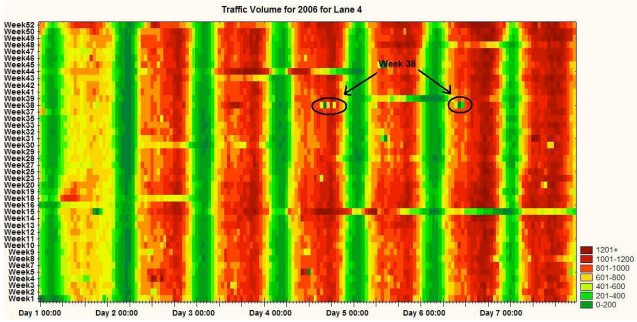

Figure 6: 2006 NLEX BLK-NB Traffic Volume for Lane 4

feature allows easier analysis of all four lanes enabling the domain experts to see the relationship of speed with volume, as well as how each lane behave and affect other lanes at a certain time.

By comparing the data images obtained from the four lanes we could possibly have a an idea of what time and lane accidents usually happen over the year, which is an interesting site for another investigation. Take for example the mid afternoon during day 6 of Week 38, the Data Image for Lane 2, Figure 5, shows a tremendous increase in number of volume for three straight hours while the Data Image for Lane 4, Figure 6 shows an opposite behaviour for the same time frame. Rotated Bar Graphs give the viewer a better perspective and understanding of this unusual pattern.

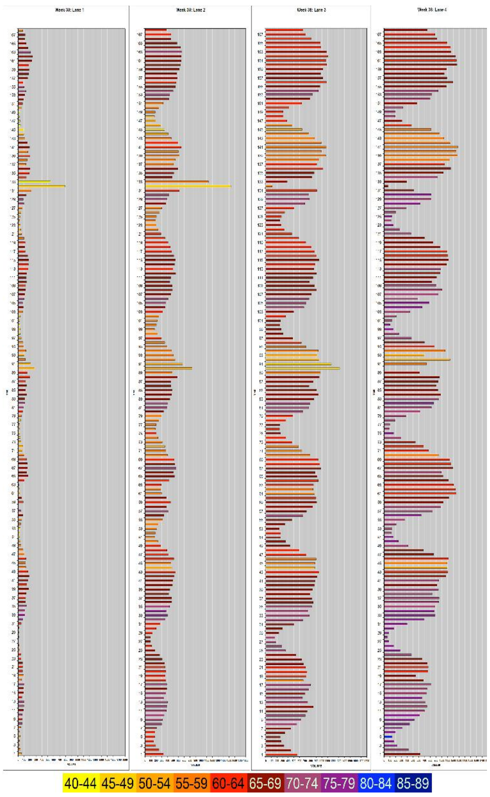

Figure 7: Lanes 1 to 4 Volume and Speed of 2006 NLEX BLK-NB for Week38

Information gathered from the data images were then validated by the Rotated Bar Graphs. If we take a look on mid afternoon of Day 6 we can observe the same behavior as in the previous data images but this time we see it in a four lane version. By just simply looking at the image, we can already presume that an incident might have occurred in Lane 4 (which is the fast lane), due to occurrence of remarkably low values and a decrease in the speed of the vehicles. We then consider the behavior of the other three lanes and assume that because of the incident in Lane 4 vehicles have transferred and compressed in the other lanes explaining the great increase in its volume and a drop off of its speed as seen in figure 7.

Data images compares each week for a general view of what the data exhibits, while Rotated bar graphs validates information from these data images and gives additional understanding by studying and comparing lane per lane of a given week. Working with these two visualization technique can assure a better analysis of the traffic data.

After analysis was done from the previous results it was found out that there were still some weeks that show discrepancy within its cluster. As an example, from Figure 4 which shows the Data Image of the original cluster obtained through X-means, it is remarkably noticeable that Week 44 exhibits a behavior different from the other members of cluster 2 where it belongs to. Week 44 have prolonged high values during Day 3 to mid Day 4 and a volume drop on mid Day 4 to Day 5 as compared to the other weeks. Aware of this unusual behavior we then conducted another X-means on each of the existing clusters. The result was subclusters for each known clusters.

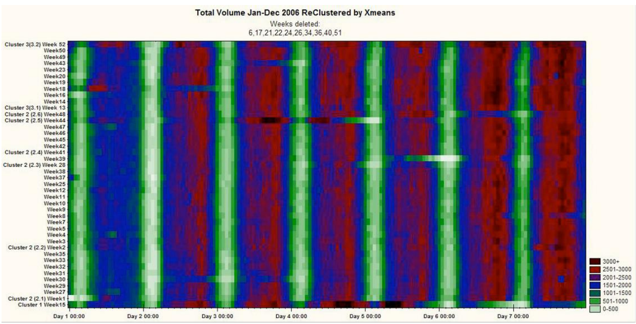

Figure 8: 2006 NLEX BLK-NB Total Volume re-clustered using X-means

Figure 8 shows the new data image for the reclustered data where the following analysis have been deduced:

From Cluster 2, six new subclusters were generated.

Subcluster 2.1 shows normal traffic behaviour but when evaluated with the other subclusters this contains lowest set of values for traffic volume. As mentioned in the previous analysis of Cluster 2, Week 30 was considered as an outlier, but after reclustering it remained to be grouped with other weeks which exhibited very low volume values.

Subcluster 2.2 contains 13 weeks and shows normal traffic behaviour similar to Subcluster 2.1. These weeks possess much higher values compared to the previous subcluster specifically for Days 2, Day 3 and Day 4. The Weeks 4 and 37 may evidently have peculiar trend during Day1, the number of days with trends that are similar to their comembers is high enough to overcome these peculiarities. Thus, they did not form their own cluster.

Subcluster 2.3 contains only 2 weeks, i.e. Weeks 28 and 39. It is very evident from the Data Image that these two weeks projected have same behavior thus giving us an idea that it will be aggregated together. As compared to the other subclusters these two weeks contains values that are extremely low during mid Day 4 to early Day 6. Week 39, apparently a potential outlier, shows a more drastic change as it appears green (0-500,501-1000) during mid Day 5 rather than violet to red (2001-2500, 2501-3000) as the normal color of other weeks in Cluster 2. A super typhoon named "Milenyo" had hit the National Capital Region from September 25 - 30[6] of that year and may explain the lesser degree of utilization for the expressway.

Subcluster 2.4 shows a set of values that are significantly larger than all other weeks belonging to other subclusters of Cluster 2.

Subcluster 2.5 having Week 44, a known outlier, as its only member was previously noted to exhibit a unique behavior. During Day 3 and start of Day 4 it contains high values and abrupt change to low values during mid Day 4 to early Day 5. As shown in Table 1, this week points to the All Saints and All Souls Day, thus, these events may seem to justify such unique behavior. Unlike all other weeks in the data set, this week possesses the highest peaks for traffic volume during the middle part of the week.

Subcluster 2.6 also contain only a single week, i.e. Week48. It can be observed that there exists a set of prolonged very large values in Days 2, 3, 4, and 5. This is followed by an abrupt prolonged low values in Day6 - i.e. Dec. 1, 2006. This loss of the usual volume of traffic flow in NLEX may have been triggered by the onslaught of the Super Typhoon "Reming" in the country (esp. Luzon) on that day[6]. The dominant behavior of this week resumed in Day7.

Cluster 3 was divided into 2 subclusters. Subcluster 3.1 contains all the same weeks except for Week 52. Week 52(an expected outlier) was clustered alone as subcluster 3.2 as it shows a very different behavior as compared to the other weeks in cluster 3. Week 52 contains the highest value rather that any other week in the whole year. This can be expected as Week 52 is the last week of the year where most people went home for Christmas and New Year vacations. The high values seen in this week are consistent in that it occurs as in every day of the week.

6. CONCLUSION

We have presented a new visualization technique that is mainly suitable for traffic data analysis. Aside from handling large traffic data set this method is purposely designed to deal with the data's following characteristics: time series, periodic and multidimensional. The Data Images and Rotated Bar Graphs generated from this method aid in a more efficient and effective data analysis for its domain users by highlighting and establishing peculiar patterns in the image. We had successfully incorporated the X-means clustering algorithm in detecting such patterns. It can also be concluded that this method is also capable of identifying and detecting normal trends, interesting and unexpected behaviors as well as outliers through iDIRBrG Method's iterative nature. Therefore, the results of this study can be a great tool for domain experts to implement better traffic planning, identifying critical regions of the North Luzon Expressway and in formulating more specific business decisions.

7. ACKNOWLEDGMENTS

The authors would like to thank Dan Pelleg and the Auton Lab of Carnegie Mellon University's School of Computer Science for the implementation of X-means used in this research.

J. Malinao is supported by DOST-SEI through the Accelerated Human Resource Development Program. R.A. Juayong is supported by supported by the Engineering Research and Development for Technology(ERDT) Scholarship. H. Adorna is supported partially by DOST-PCASTRD through an ERDT project entitled "Information Visualization via Data Signatures".

The authors would like to thank Ma. Sheilah Napalang, Dr. Engr. of NCTS and Regin Regidor, Dr. Engr. current Director of NCTS for use of data and for the help in verifying the results.

8. REFERENCES

[1] C. Johnson and C. Hansen. *The Visualization Handbook.* Academic Press, Inc., Orlando, FL, USA, 2004.

[2] W. Shroeder, K. Martin, and B. Lorensen. *The Visualization Toolkit, An Object Oriented Approach to 3D Graphics.*, 3rd Ed., Kitware Inc., Clifton Park, N.Y., 2002.

[3] C. H. Yu and S. Stockford. *Evaluating spatial- and temporal-oriented multi-dimensional techniques, Practical Assessment, Research and Evaluation*, 8:17, 2003. Retrieved August 19, 2009 from http://PAREonline.net/getv.asp?v=8n=17

[4] D. Pelleg and A. Moore. *X-Means: Extending k-means with efficient estimation of the number of clusters.* In Proceedings of the Seventeenth International Conference on Machine Learning, pp. 727-734, 2000.

[5] M. C. Minnotte and W. West. *The data image: a tool for exploring high dimensional data sets.* In Proceedings of the ASA Section on Statistical Graphics, pp. 25-33, 1998.

[6] www.typhoon2000.ph/stormstats/WorstPhilippineTyphoons.com