G-MAP: A High Dimension Data Grand Tour Map

Zhimin Li
University of Utah
Email: zhimin.li@utah.edu

Shusen Liu
University of Utah
Email:shusen.liu.hust@gmail.com

Paul Rosen
University of South Florida
Email:prosen@sci.utah.edu

Abstract—The Grand Tour presents a multidimensional data exploration journey for users that brings them all possible views of their data. However, far too often, steps along the journey are uninteresting and wasteful of time. We instead want to find interesting representative views for the user and static or dynamic exploring interesting views to reveal the feature of the data. To accomplish this, we propose a new high dimensional data analysis pipeline which include four steps. Sampling, we sample 2-dimension projections uniformly distribute along the path of a Grand Tour. Feature detecting, we then use the nine characterizations of Scagnostics to describe the views so that each of them can be differentiated from others. Filtering, from those characterizations, users can select representative views through an combination of feature parallel coordinate and feature SPLOM. Navigation, once the desired projection is selected, we proposed a new technique called feature N-D touch pad to optimize the selected projection view. We demonstrate our approach on a number of commonly used multidimensional data sets.

Keywords—*Grand Tour, statistical graphics, High-dimensional data, visual analytics*

I. INTRODUCTION

High dimensional data set is popular in many different field like science, finance, business and so on. It's common in today that a data set which hundreds or thousands of dimension are generated. However, the ability that human can appreciate this type of data is still limited. Part of the reason comes from the common challenge in data science:curse of dimensionality. Once the number of dimension increase, the volumes of space will extends so fast that available data will become sparse in that dimensional space and that will result into the lost of statistic significant for most data analysis. The other reason may come from natural constrain of human eye[13]. As we grow up in 3d physical world, most object we see will be 3d or lower than which leads to the poor ability to understand the data greater than 3 dimension directly.

Thus, many research choose a solution which transforms the high dimensional data set into low dimensional space(1d, 2d) and explore the data set's feature from it. Even though this operation will result into some loss of information, with the help of some optimizing technique a low dimension view can still reveal some outstanding feature in high dimensional data set. The common static projection technique includes Principle Component Analysis[22] which can present user a low projection view has the maximum variance and Multidimensional Scale[15] that keeps object distance in high dimensional space as well as in low dimensional space. The other famous dynamic projection strategy was proposed by Asimov called grand tour[2], which provided a smooth sequences of 2d projections help user to view all possible direction of the data projection.

This technique generate a number of projections uniformly distribute along the grand tour path and projection interpolation is implemented between each pair of projections to construct a sequence of smooth views. However, most time grand tour costs user too much time to walk around the boring views and user may miss the highlight.

Scagnostics is a technique introduced by John and Paul Tukey. The basic element of it is a scatter plot which reveal the relationship between a pair of variables in the data set. Since the number of plots in the Scagnostic will increase quadratically with dimension, N. Dang and Wilkison develop a tool called ScagExplore[9] . ScagExplore use different metrics[28] to describe each plot in scagnostic and base on those metrics they design a filter mechanism to choose representative plots to explore the data set. However, all the plots in scagnostic will be parallel axis projection and the information in those projection are limit as the dimension increase[1].

In our research, instead of analyzing the parallel axis projection, we will move our step to the linear projection field. Here we introduce a new tool called Grand Map. Instead of viewing a long movie like grand tour or running a time consuming projection pursuit algorithm[11] to get a interesting projection by optimizing the projection pursuit index. Grand Map uses four steps to target an interesting projection and statically or dynamically explore the projection.

1 Sampling: Grand Map sample a large number of 2d projections which uniformly distribute along the path of grand tour.

2 Describing: For each projection, Grand Map use different metrics to describe them and use an intuitive way to show those projections in a map.

3 Filtering: Grand Map has an efficient filtering mechanism to filter less informative projection and target interesting projection.

4 Navigating: Once an interesting projection is selected, user can use data navigation technique or clustering analysis method to optimize the projection or analyze the data set.

The goal of those four steps aims to shorten the time that needs to find the interesting scenery in the grand tour journey. With those projection samples, quality metrics and filtering mechanism, user can easily target relative interesting scenery . Then, the navigating technique will lead user to walk around that place and explore the feature. Compare with projection pursuit, grand map do not aim to find out a projection which has the maximum value reveal feature but it has the ability to

help user to target projection that has the strong feature that fit user's taste. In our research, our contribution are:

- We proposed a new pipeline to analyze the high dimensional data set which include sampling; describing; filtering; analyzing.

- We suggest a possible way to intuitive show all the projections in a map.

- We proposed our own filtering mechanism to filtering less informative projection.

- We combine heat map and N-D Touchpad navigation technique[20] to create a new data navigation tool which provide more flexible way to view the neighbor of a projection.

Grand Map follows Shneiderman's information seeking strategy: overview first, then detail on demand.In the following, Section 2 will describe the relative work. Section 3 will give an overview of the system. Section 4 will introduce the way to sample projection view and how to pick up representative projection view. section 5 will introduce the data navigation tool and section 6 will illustrate the use of our system to explore two typical data set. Section 7 will give the conclusion and the future work

II. RELATIVE WORK

Dynamic Projection

Most of time, data navigation was used to view the dynamic change of a candidate projection view. In last decade, many technique was proposed relate to this idea. The original tool was Asimov's grand tour show a smooth sequence of 2d frame visit all possible view in high dimensional space, and based on it Dianne cook create Guided tour[7]: instead of walking through all possible view, it brings user go through continuous views which lead to a final view has specific structure; Manual tour[6]:User can control the contribution of one attribute in a projection and how sensitive the projection view relate to this attribute will also be reveal. Manual control can be embed into both grand tour and guided tour. Wegman[25] extend Asimov's grand tour into k dimensional space and user can use scatter plot matrix or parallel coordinate to trace the dynamic change of the projection data. McDonald[18] use the axis-align 2d projection of the data set as the target frame and implement the interpolation between pairs of projection view. Dice Roll[10] proposed a similar framework named scatter plot matrix navigation, but the transition between two scatter plot was implement as an animation. In Hurley's Graphs as navigational infrastructure[14], 2d projection view is regarded as a node in the graph and the transition between each pairs of projection view is represented as a edge. In shusen's view transition graph[17], axis align projection was replace with linear projection and then transition can happen between PCA projection and other interesting projection.

Data Subspace

In the real life data set, the number of dimensional is usually hundreds, thousands or may be more. No all the attributes will contribute to the analyze process, since many attributes are redundant or meaningless. The effective way to
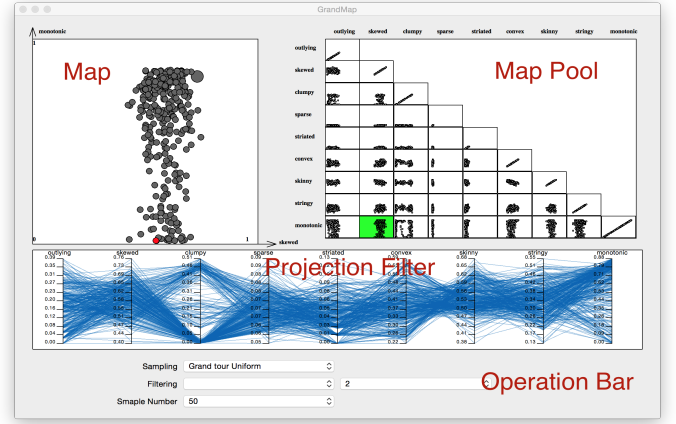

Fig. 1.   Grand Map.

remove the useless and noise attribute is necessary. Jinwook Seo [21] presents an interactive tool called a rank-by-feature framework compare the relation between each dimension. Each pair of attributes in data set can form a Axis Parallel projection, user can select a criterion(eg.Pearson correlation) to measure the projection and ranking the relationship strong between each pair attribute. Entropy-based Subspace talked by Cheng[3] use two entropy based closure property to filter uninteresting subspace and target interesting subspace.

Quality Metrics

In last decades, many metrics was proposed to measure the quality and feature of a scatter plot. In Friedman and Turkey's projection pursuit algorithm, they proposed a projection index measure the inter-point distant and the overall spread of the data. Dianne Cook[8] also proposed projection pursuit index like Holes used to measure the gap between two clusters, Central Mass reveal minor outlier and LDA try to find a projection that have three cluster revealed. There are numerous projection pursuit index[16] and what kind of PP index is chosen depends on what features user are looking for in a projection.class consistency[23]

III. GRAND-MAP

Grand Map(see Fig.1) There are four components on Grand Map.The scatter plot on the left top will be the map and each point in the plot represent a 2d projection in high dimensional space. The X-axis or Y-axis can be one of the nine metrics(outlying, or monotonic) used to describe the plot and the value range from 0 to 1. User can choose different map from the scagnostic on the right which is map poo. The parallel coordinate is the projection filter, users can use it to choose the projection they desire. The Last component is the operation bar, user can choose k-mean, and other filter algorithm to go through the projections or use can re sample all the projections and choose the number of projection they want to have.

Data Navigation(see Fig.2) The scatter plot on the left is the actual projection that you think may be interesting. The polygon on the right is ND Touch Pad and user can use it
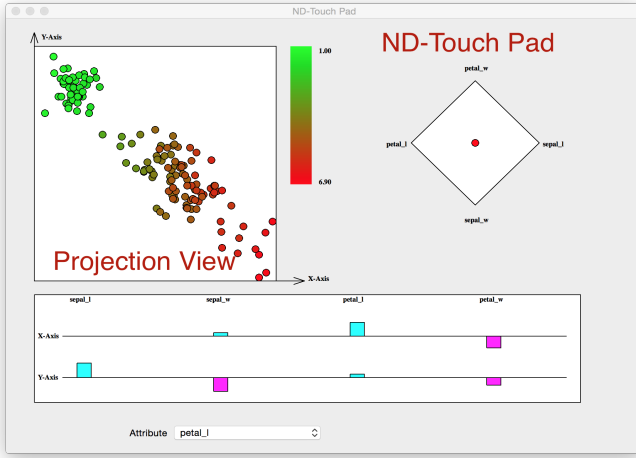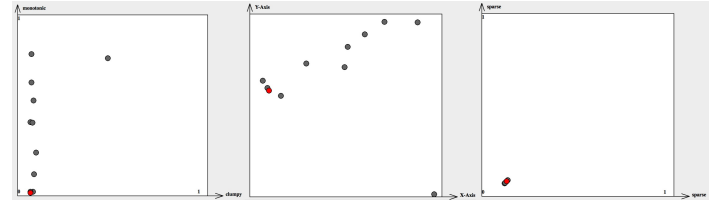
Fig. 2. Data Navigation Tool.

Fig. 3. There are three different scatter plots in this image. The x-axis of the plot on the left is clumpy and y-axis is monotonic. From left to right projection's clupmy feature will increase and from bottom to top monotonic feature will become stronger. The axis of plot in middle doesn't have real meaning, it's generated by multi-dimensional scale and properties of each describing feature(see section V). The projection close to each other will have similar feature value. The plot on the right has the same mean x-axis and y-axis. All the point cluster near the original point and it means spare feature is weak in all the projection.

to navigate the projection view to see the data in different direction. The Bar chart on the bottom shows the weight that each dimension contribute to the current projection view. The list box on the bottom allow user to choose different attribute value to be shown as color gradient in the projection view.

IV. SAMPLING

How to choose the subset projections from all projection of the high dimensional data set will great affect the result of the analysis. The process of sampling can reference to how the path of tour was generated. When generating a smooth tour, a target projection need to choose and interpolation need to be implemented in the a geodesic path between current projection and target projection. How a new target projection generated is different in various tour. In grand tour, Daniel Asimov propose a few methods:Torus method generate a sequence of projection with a step-size in the path of the grand tour. At-random method generate a list of independently projections which are uniformly distribution in the path of grand tour. In Dianne cook's projection pursuit grand tour, instead of walk through all the possible projection view, the tour only go through the path lead to the high feature revealing view. These is the other way which was proposed by J.H. Conway[5], a fixed number of projection was chosen and they are approximate equi-spaced.Wegman proposed a faster algorithm generate an Pseudo-Grand[26] however, the space cover of the algorithm is limit.

Grand Map choose two different ways to sample the projection view in high dimensional space. One of them are the at-random method in grand tour, which almost guaranteed that the projection will uniformly distribute along the path of grand tour. The other method is called metrics driven sampling.

At-random

1) **Step1**: Generate a vector $\mathbf{U}$ with n components(n is the dimension of the data set), each component is a pseudo-random number in unit interval and $0 < U \cdot U < 1.0$

2) **Step2**: The same as step1, generate another new vector $\mathbf{V}$.
3) **Step3**: Apply Gram-Schmidt procedure to $\mathbf{U}$ and $\mathbf{V}$ to obtain a pair of orthonormal vectors $\mathbf{W}$ and $\mathbf{X}$

$$W = \frac{U}{||U||}$$

$$X = \frac{V - (V \cdot W)W}{||V - (V \cdot W)W||}$$

There are two reasons that this method is chosen. Firstly, this method can almost guarantee that the sequence of projection will be uniformly distributed in the path of grand tour. Secondly, this method is very easy to be implemented.

Metrics Driven Sampling

Before doing any analysis, if user already know what kind feature they are looking for in a projection, then there is not need to show all the projects in front of user. Grand Map use metrics driven sampling, which will only generate projection that has a few or other specific feature. Before the generation of projection view, user will set a threshold value for one or a few properties. Once a random projection is generated, it will be checked and see whether it has the feature that user prefer to. It the projection doesn't satisfy the requirement. It will be discarded and won't be shown up in the further analysis. Here we use the method that Anushka Anand in her paper mention about random projection[1].

Once there are large amounts of projections, how to represent those projects and how to organize them on computer screen will be a problem. In TripAdvisor [**?**], the author proposed a solution which contain two maps: one is global sight map and the other is sight glyph. Sight glyph is a projection scatter plot with two bar chars show on x-axis and y-axis to reveal each projection vector value. Global sight map include many sight glyphs and position of each sight glyph is decided by multi-dimensional scale algorithm and 2n vector which is a combination of two orthonormal n component vector. In ScagExplore, TN, Dang use force-layout map and nine metrics of scagnostic each scatter plot has to decide the final map of projections. Both of those solution will be helpful to reveal the relationship between each projection when the number of projections are small. Once the size of subset projections is

large, their representation will be cluttered. Grand Map simply use scatter plot to organize projections and each projection is a point in the plot. However, what x-axis and y-axis represents will affect how user comprehend the plot. In figure 3, there are three different scatter plots and the position of each projection on the plot will have different meaning. Grand Map can use different plot to choose desire projection or check the value distribution of projection's specific feature.

Even though, we use scatter plot and point to organize those projection. The possible projection show on the map may be large. It will be shown as a big cluster of points and user will be hard to choose their desire projection. Here we use the zooming[4] technique to solve this problem.

V.   DESCRIBING

In graph-theoretic scagnostics[28], the author use convex hull, alpha hull and minimum spanning tree three different geometric graphs to construct 9 different metrics which used to measure the trend, density, outlier, shape, coherence of scattered points. Grand Map use those metrics to measure the quality of each projection. **Outlier**

- outlying: This metrics is calculated based on the edges length of the MST.

$$\omega = q75 + 1.5(q75 - q25)$$

  $q75$ is the 75 percentile length, $q25$ is the 25 percentile length. The edge length longer than ω will be considered as a outlier. The sum of outliers' length compare to the sum of the all edges' length in the MST will be the metrics value.

$$Outlying = \frac{length(outlier)}{length(MST)}$$

Shape

- Convex: is the ratio of the area of convex hull with the alpha hull.

$$Convex = \frac{area(alphashape)}{area(convexhull)}$$

- Skinny:This method used compare the area of the alpha hull and the perimeter of the alpha hull.

$$Skinny = 1 - \frac{\sqrt{4\pi area(alphahull)}}{perimeter(alphahull)}$$

- Stringy: the ratio of the length of longest shortest path go thought the MST and the sum of the length of all edge in the MST.

$$Stringy = \frac{diameter(T)}{length(T)}$$

- Straight: compare the euclidean distance between the two end points of longest shortest path with the longest shortest path length of MST.

$$Straight = \frac{dis(i,j)}{diameter(T)}$$

Trend

- Monotonic: is calculated by using the Pearson correlation. Since the domain of Pearson correlation is in $[-1, 1]$, the value of monotonic will be the square of the Pearson correlation, which make sure the value range will be in $[0, 1]$

$$Monotonic = PCorrelation^2$$

Density

- Skewed: the edge length distribution of MST tree can be used to measure the relative density of the scatter points.

$$Skew = \frac{(q_{90} - q_{50})}{(q_{90} - q_{10})}$$

- Clumpy: indicate the possibility of clustering. In the following equation, j is the index of the edge of MST, and k is the index of the runt set[12] related to the j indexed edge in MST.

$$Clumpy = \max_{j}[1 - \frac{\max_{k}[length(e_k)]}{length(e_j)}]$$

Coherence

- Straited: measure the relative smooth path of the MST.$V^{(}2)$ is a set of vertex which is degree of 2 in MST.

$$Striate = \frac{1}{V^{(}2)} \sum_{v \in V^{(}2)} |cos\theta_{e(v,a)e(v,b)}|$$

VI.   FILTERING

One of the challenge in exploratory data analysis is to find an high informative view in multi-dimensional space. In Grand Map, the number of projections will be large. Thus a powerful filter technique are necessary to help users target their desired projection. Many previous research had proposed efficient filter mechanism to help user identify interesting plot from scagnostics. In ranking by feature framework, author rank each plot by quality metrics value. The scatter plot has high informative content can be targeted by user quickly. Grand Map combine feature parallel coordinate[9] and feature SPLOM[27] together to help user identify their desire projection.Figure 4 is a feature parallel coordinate and Figure 5 is a feature SPLOM. Both of those method have been used to help user choose desire plot in scagnostics. However, the combination of those two technique has more powerful functionality. For feature scatter plot in figure 5, each point is a projection and user can check projections' convex and monotonic value distribution. User can tell which projection has more feature reveal than others from their position. However, the possible feature of projection is more than 2, thus feature parallel coordinate can filter some less useful projection before user to do further analysis.

VII.   NAVIGATING

Once an interesting projection(eg. high clumpy feature) is picked from the large amount of projects, it worth to check its neighbor projection and see if there is more features can be figure out.
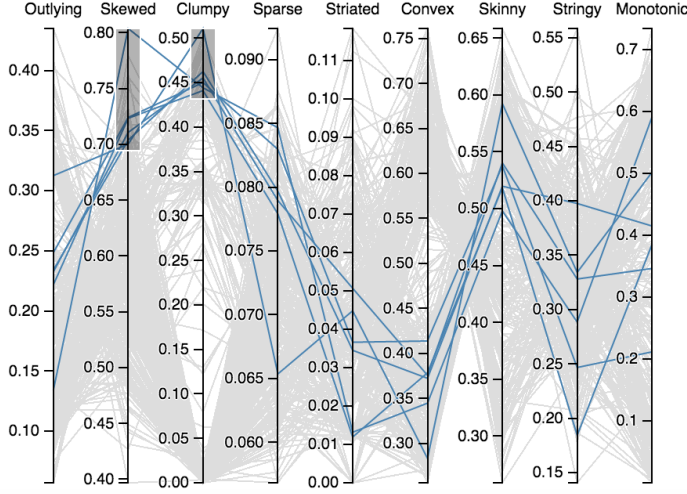
Fig. 4. In the figure, each polygon line represent a projection with feature value in each feature axis. For each axis, the ticket near it indicate the feature value range of current sample projection subset. User can use the parallel coordinate to choose user's desire subset projections. In this figure, user choose projection with high Skewed and Clumpy feature.
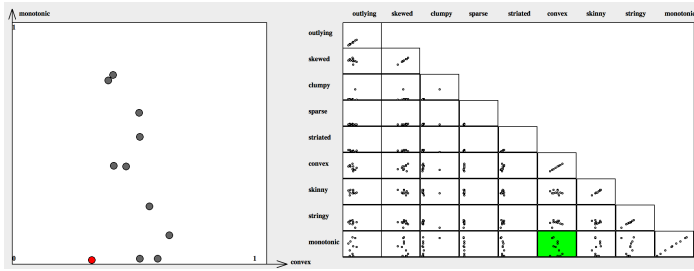

Fig. 5. The scagnostic on the right is the feature SPLOM, and the scatter plot on the left is the feature plot show the projections' monotonic and convex value distribution.
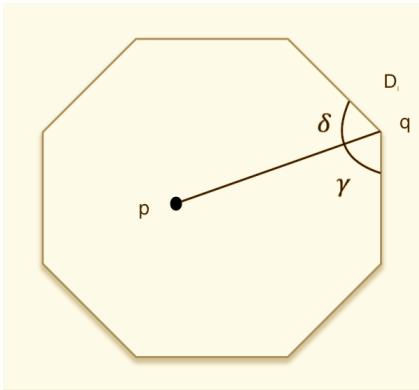

Fig. 6. This is a 8-sided polygon and it can be used to navigate eight dimension data. p represent a position inside the polygon and q represent a vertex of dimension D_i. δ, γ is the angle line pq intersect with polygon. The interpolate weight for D_i is decided by the position of p and the angle δ, γ.

Here we combine the technique of ND-touch pad and target projection pursuit and create a more flexible data navigate tool to help user explore the neighbor of a projection.

ND-touch pad is an navigator interface first proposed by S.Garg. Later, J Talbot also discuss a similar tool named Linear Combination Widget[24] which combine with visual ensemble matrix to construct a ensemble classifier for machine learning. This technique base on the generalized Barycentric coordinates formula talked by M. Meyer[19] which does an excellent job in scale interpolation over irregular N-sided polygons. For navigation, ND-touch pad was described by a regular N-sided polygons and each vertex represent one dimension(see figure 1). The interpolate value w_i for D_i is:

$$w_i = \frac{cot(\gamma) + cot(\delta)}{||p - q||^2}$$

Assume we have N vertex, then the interpolated value at p with data vector v will be

$$pv = \sum_{i=1}^{N} v_i a_i \quad a_i = \frac{w_i}{\sum_{k=1}^{N} w_k} \quad \sum_{i=1}^{N} a_i = 1$$

In TripAdvisor Julia E. Nam talk about how the order of each dimension in the polygon vertex will affect the result and the numerical unstable at the edge of the polygon. For numerical unstable, Talbot suggest the linear interpolate method when the point p close to the edge. As the author mention the possible space that ND-touch pad can cover is limited.

Based on previous technique, Grand Map add new functionality to the touch pad.

Color Guider

User can use ND-touch pad to explore the selected projection by checking it's neighbor projection. However, like grand tour, research without professional training can't catch interesting pattern from the smooth changing projection. Grand Map add heat map to the nd-toupad component create a feature touch pad to help user identify informative projection easier. see figure 7. The color gradient change from blue(weak feature) to red(strong feature). The place have deep red color means move the touch pad controller to that place will show up a projection has strong feature value. Figure 7 is a clumpy feature touch pad. The feature touch pad will lead user to view the projection that has the strong feature. The way to check the neighbor of a projection will help user to reveal the a few variable that result into this feature. What it does is that it will add more weight to those variable which result into the desire projection view. At the end, user will get a projection with strong specific feature.

Re-Center

In order to increase the possible space that user can explore. Grand Map modify the touch pad which allow it re-centering. ND-toupad can use to explore a projection's neighbor but once the neighbor is more interesting than the selected projection. Feature ND-toucpad can recenter and make the more interesting projection as the selected projection and explore it again.
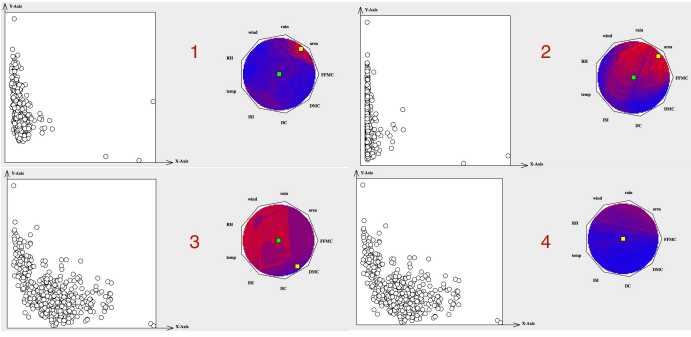
Fig. 7. There are 4 image labels from 1 4 in this figure. Image 1 is the initial projection selected from grand map with the feature touch pad. Image 2 take the projection in image 1 as the center of navigation. Image 3,4 take the projection in image 2,3 as the center of navigation.
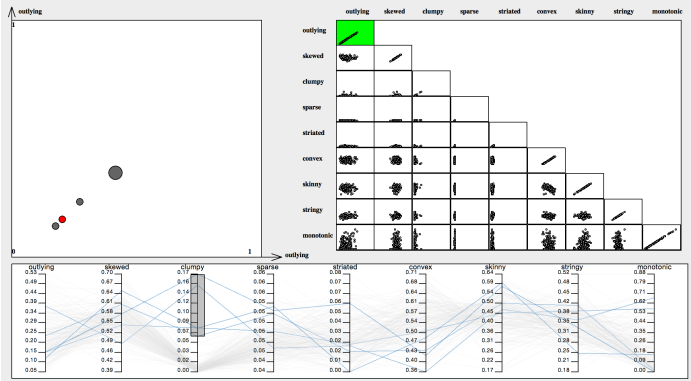

Fig. 8. In this figure, user choose the outlier as the x-axis and y-axis. In parallel coordinate, the clumpy feature was used to filter the less useful information that user believe. User click the projection which has the highest outlier value with relative high clumpy feature.

Once user selected a projection whose specific feature has high value revealing, user can use the feature touch-pad to explore it and find the local maximum.

VIII. EVALUATION

As we claim before, grand map helps user to choose a projection which has their desire properties to analyze. And the independent can be used the feature touch pad to optimize the feature.

Forest Fires Data Set

We use the forest fires data set which we download from UCI machine learning repository. In this data set, there are 13 different attributes. From figure 8, we use the tool to filter projections which has less clumpy feature. Then, we choose the projection has the highest outlier feature from the subset. As we can see the initial projection we choose has good cluster feature. There are two different clusters show on the projection plot. The outlier feature is relative strong, there is a point on the top of the plot which can be identify as outlier. However, as we can see from the feature touch pad, the outlier feature is stronger near the rain, area, FFMC attribute. Once we use the feature touch pad navigate the current projection to the area which has stronger outlier feature. The projection plot on the left give us a better view about outlier.
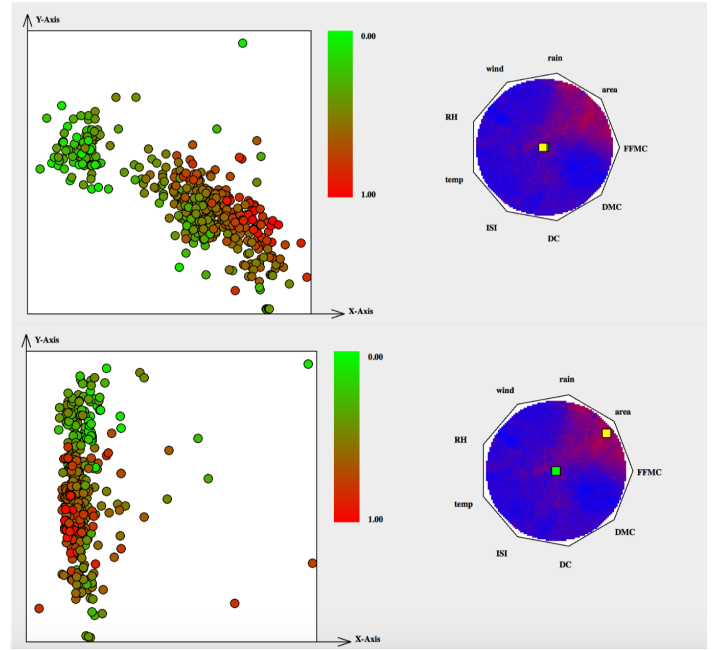

Fig. 9. In this figure, there are two picture. The picture on the top represents the initial projection we choose from the grand map which has highest outlier feature with relative high clumpy feature. The heat map show on the feature touch indicate the outlier feature value distribution on the touch pad. The picture on the bottom show that under the guider of feature nd-touch pad, we navigate the current project to the area which has higher outlier feauture.

Blood Transfusion Service Center Data Set

We will use the other data set which get from UCI machine learning repository is blood transusion data set. In this data set, there are 5 different attributes,

IX. CONCLUSION

In our research, we present a new pipeline to visualize the high dimensional data set which include four key step: sampling, describing, filtering and navigating. Base on this mechanism, we use Qt platform to implement a high dimensional data analysis tool called Grand Map. Instead of analyze individual projection directly, grand map show all the projection which is described by a few quality metric on a plot. This ideas are presented by a few other researchers but most of those work focus the projection on the axis parallel domain. Our research aim to analyze the projection in linear projection domain. We combine the common feature SPLOM, feature parallel coordinate to create a map which has more flexible ability to reveal the relationship between each projection base on the quality metrics. At the same time, we modify the previous research about nd-toucpad and create a new data navigation tool called feature touch pad which will lead user to navigate the current feature projection to the place which some feature has local maximum. Unlike grand tour and projection pursuit which helps user to view all possible projection or present the optimizing projection. Grand map use the interactive lead user to the projection which has relative feature. It's easier to implement and more intuitive for user to use it.

X. NEXT STEP

In our tool we selected subset projection from infinite projection in high dimensional space to analyze the feature of high dimensional data. However, the number of projections that are enough to reflect all the feature of high dimensional data set is still challenge. There are many method proposed to generate projection in high dimensional space but which method is more robust and which method is more useful is still unknown. Howe to measure the quality of our sampling will be the future research that we want to explore more.

XI. ACKNOWLEDGMENT

REFERENCES

[1] A. Anand, L. Wilkinson, and T. N. Dang. Visual pattern discovery using random projections. In *Visual Analytics Science and Technology (VAST), 2012 IEEE Conference on*, pages 43–52. IEEE, 2012.

[2] D. Asimov. The grand tour: a tool for viewing multidimensional data. *SIAM journal on scientific and statistical computing*, 6(1):128–143, 1985.

[3] C.-H. Cheng, A. W. Fu, and Y. Zhang. Entropy-based subspace clustering for mining numerical data. In *Proceedings of the fifth ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 84–93. ACM, 1999.

[4] A. Cockburn, A. Karlson, and B. B. Bederson. A review of overview+ detail, zooming, and focus+ context interfaces. *ACM Computing Surveys (CSUR)*, 41(1):2, 2008.

[5] J. H. Conway, R. H. Hardin, and N. J. Sloane. Packing lines, planes, etc.: Packings in grassmannian spaces. *Experimental mathematics*, 5(2):139–159, 1996.

[6] D. Cook and A. Buja. Manual controls for high-dimensional data projections. *Journal of computational and Graphical Statistics*, 6(4):464–480, 1997.

[7] D. Cook, A. Buja, J. Cabrera, and C. Hurley. Grand tour and projection pursuit. *Journal of Computational and Graphical Statistics*, 4(3):155–172, 1995.

[8] D. Cook, A. Buja, E.-K. Lee, and H. Wickham. Grand tours, projection pursuit guided tours, and manual controls. In *Handbook of data visualization*, pages 295–314. Springer, 2008.

[9] T. N. Dang and L. Wilkinson. Scagexplorer: Exploring scatterplots by their scagnostics. In *Pacific Visualization Symposium (PacificVis), 2014 IEEE*, pages 73–80. IEEE, 2014.

[10] N. Elmqvist, P. Dragicevic, and J.-D. Fekete. Rolling the dice: Multidimensional visual exploration using scatterplot matrix navigation. *Visualization and Computer Graphics, IEEE Transactions on*, 14(6):1539–1148, 2008.

[11] J. H. Friedman and J. W. Tukey. A projection pursuit algorithm for exploratory data analysis.

[12] J. Hartigan and S. Mohanty. The runt test for multimodality. *Journal of Classification*, 9(1):63–70, 1992.

[13] G. Hatfield. *Perception as unconscious inference*. Wiley Online Library, 2002.

[14] C. Hurley and R. Oldford. Graphs as navigational infrastructure for high dimensional data spaces. *Computational Statistics*, 26(4):585–612, 2011.

[15] J. Kruskal and M. Wish. *Multidimensional Scaling*. Number no. 11 in 07. SAGE Publications, 1978.

[16] E.-k. Lee. Projection pursuit methods for exploratory supervised classification. 2003.

[17] S. Liu, B. Wang, J. Thiagarajan, P. Bremer, and V. Pascucci. Visual exploration of high-dimensional data through subspace analysis and dynamic projections.

[18] J. A. McDonald. Orion i: interactive graphics for data analysis. *Dynamic Graphics for Statistics*, pages 179–200, 1988.

[19] M. Meyer, H. Lee, A. Barr, and M. Desbrun. Generalized barycentric coordinates on irregular polygons.

[20] J. E. Nam and K. Mueller. Tripadvisor^{ND}: A tourism-inspired high-dimensional space exploration framework with overview and detail. *Visualization and Computer Graphics, IEEE Transactions on*, 19(2):291–305, 2013.

[21] J. Seo and B. Shneiderman. A rank-by-feature framework for interactive exploration of multidimensional data. *Information Visualization*, 4(2):96–113, 2005.

[22] J. Shlens. A tutorial on principal component analysis. *arXiv preprint arXiv:1404.1100*, 2014.

[23] M. Sips, B. Neubert, J. P. Lewis, and P. Hanrahan. Selecting good views of high-dimensional data using class consistency. In *Computer Graphics Forum*, volume 28, pages 831–838. Wiley Online Library, 2009.

[24] J. Talbot, B. Lee, A. Kapoor, and D. S. Tan. Ensemblematrix: interactive visualization to support machine learning with multiple classifiers. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 1283–1292. ACM, 2009.

[25] E. J. Wegman. The grand tour in k-dimensions. In *Computing Science and Statistics*, pages 127–136. Springer, 1992.

[26] E. J. Wegman and J. L. Solka. On some mathematics for visualizing high dimensional data. *Sankhyā: The Indian Journal of Statistics, Series A*, pages 429–452, 2002.

[27] L. Wilkinson, A. Anand, and R. Grossman. High-dimensional visual analytics: Interactive exploration guided by pairwise views of point distributions. *Visualization and Computer Graphics, IEEE Transactions on*, 12(6):1363–1372, 2006.

[28] L. Wilkinson, A. Anand, and R. L. Grossman. Graph-theoretic scagnostics.