Visualization of Large, Time-Dependent, Abstract Data with Integrated Spherical and Parallel Coordinates

J. Walker[1], Z. Geng[1], M. W. Jones[1], and R. S. Laramee[1]

[1]Department of Computer Science, Swansea University, Wales, UK
Email: {csjames, cszg, m.w.jones, r.s.laramee}@swansea.ac.uk

Abstract

Parallel coordinates is one of the most popular and widely used visualization techniques for large, high dimensional data. Often, data attributes are visualized on individual axes with polylines joining them. However, some data attributes are more naturally represented with a spherical coordinate system. We present a novel coupling of parallel coordinates with spherical coordinates, enabling the visualization of vector and multi-dimensional data. The spherical plot is integrated as if it is an axis in the parallel coordinate visualization. This hybrid visualization benefits from enhanced visual perception, representing vector data in a more natural spatial domain and also reducing the number of parallel axis within the parallel coordinates plot. This raises several challenges which we discuss and provide solutions to, such as, visual clutter caused by over plotting and the computational complexity of visualizing large abstract, time-dependent data. We demonstrate the results of our work-in-progress visualization technique using biological animal tracking data of a large, multi-dimensional, time-dependent nature, consisting of tri-axial accelerometry samples as well as several additional attributes. In order to understand marine wildlife behavior, the acceleration vector is reconstructed in spherical coordinates and visualized alongside with the other data attributes to enable exploration, analysis and presentation of marine wildlife behavior.

Categories and Subject Descriptors (according to ACM CCS):   I.3.3 [Computer Graphics]: Picture/Image Generation—Line and curve generation

1. Introduction and Motivation

Parallel coordinates is a widely used visualization technique for large, multi-dimensional data. First introduced by Inselberg [Ins85], each data attribute is represented by an individual axis, the combined set of all axis are aligned in parallel with multiple polylines connecting them. Parallel coordinates is effective in the analysis of large high-dimensional data. However, when visualizing data which contains vector attributes, it is necessary to cognitively integrate each corresponding vector axis to interpret direction, a demanding task for the user.

While parallel coordinates is good for displaying relationships between scalar attributes, it is not as effective for data containing vector (direction) attributes, in our case, tri-axial acceleration data. Instead, as demonstrated by Grundy et al. [GJL*09] tri-axial accelerometer data is better represented in three-dimensional spherical coordinates. This offers a compact representation, enabling visualization, aggregation, exploration and analysis of direction in space.

We join the advantages of spherical coordinates for visualization of tri-axial accelerometry data with those of parallel coordinates for interrogating large time series multi-variate sensor data. The naive joining of the two leads to many problems. Firstly, the computational complexity of parallel coordinates means rendering speeds can be slow, due to the amount of polylines rendered [GPL*11]. Secondly, an occlusion problem occurs from rendering more polylines than available pixels [GPL*11]. It is important to create a visualization which is fast at rendering to enable interaction and exploration of the three-dimensional spherical plot, as well as a visualization which displays a true representation of underlying patterns in the data set.

In this paper, we present a novel, work-in-progress solution, providing a hybrid visualization of integrated parallel and spherical coordinates. Our solution features: a split spherical coordinate sphere to reduce occlusion, curved polylines for enhancing visual perception, and spherical angular histograms for reducing overplotting and the compu-

tational complexity of parallel coordinates. This paper provides the following contributions:

- an integrated spherical and parallel coordinates visualization,
- approaches to address over-plotting and occlusion,
- application of our techniques to real-world time-dependent, high-dimensional, abstract data.

The rest of the paper is organized as follows. In section 2, we provide an overview of related work. In section 3, we present our user-options and associated visualizations, demonstrating the results on marine wildlife behavioral data. Finally, in section 4, we conclude our findings and outline future work.

2. Related Work

The problem of visualizing high-dimensional vector data was recognized by Qu et al. [QCX*07], for weather data from Hong Kong. They proposed drawing polylines from parallel coordinates to a two-dimensional polar axis, a more natural domain for vector data. A simple solution for extending this technique to utilize three-dimensional vector data can be achieved by drawing lines from parallel coordinates to a 3D spherical coordinate plot. However, this can cause occlusion and slower interaction.

We discuss two themes of related work. Firstly, existing methods for visualizing tri-axial accelerometer data, to discover methods of visually representing vector data and secondly, clutter reduction methods to avoid over plotting and slow rendering speeds.

Visualization of Tri-axial Accelerometer Data Grundy et al. [GJL*09] showed the visualization of tri-axial marine wildlife data using 3D spherical coordinates. Spherical scatter plots are used to show the geometric distribution of data, and spherical histograms show common animal movements. Their method is effective at enabling the exploration of tri-axial data, with two additional attributes forming part of the visualization at any one time.

Blaas et al. [BBG*09] visualize higher-order state transitions between behavior labeled tri-axial data. Their visualization displays state transitions of behaviors within the context of the state they occur in. This is advantageous by providing easy-to-follow transition paths as well as making it possible to perceive common behavior transitions.

Clutter reduction Similar to other information visualization methods [FdOL03, ED07, KK96, UTH06, WB97], the parallel coordinate plot can suffer from overplotting. In this section, we concentrate on previous work on parallel coordinates for large data sets. Generally, the clutter reduction methods for large data sets can be categorized as: alpha-blending, clustering, focus+context and frequency and density plots. We provide a brief overview of the literature.

Alpha Blending: Edward J. Wegman [Weg90, WL96] represented the density of the plots with transparency. This works well with small datasets, however, with large datasets the range of data is much greater and consequently it is more difficult to fully represent the fidelity of complex datasets. It is difficult to obtain a clear understanding of patterns and clusters, and outliers may get lost.

Clustering: Fua et al. define large data sets as containing 10^6–10^9 data elements or more [FWR99]. They adopt Birch's hierarchical clustering algorithm which builds a tree of nested clusters of lines based on proximity information. Proximity-based coloring is introduced to demonstrate clusters, and transparency to show the mean and the extent of each cluster. Johansson et al. [JLJC05] transform each K-means-derived cluster into three high precision textures, namely an animation, outliers and structure texture, and combine them into a polygon.

Focus+Context: Ellis et al. propose a focus+context viewing technique that uses an automatic sampling algorithm and sampling lens for parallel coordinate visualization [ED06]. They investigate three ways to calculate the degree of occlusion from overlapping polylines and describe a raster algorithm as the most efficient metric. Novotny and Hauser develop another focus+context visualization using binned parallel coordinates [NH06]. Binned parallel coordinates provide context views while traditional polyline-based parallel coordinates present focus information. Ruebel et al. [RPW*08] extend Novotny and Hauser's work, and propose adaptive histogram bins which use a higher resolution in areas with high data density.

Frequency and density plots: One of the ways to reduce the clutter in parallel coordinates is based on data frequency. With this approach, data is often aggregated and filtered by the binning process [AdOL04, BBP08, Car91, NH06, JTJ03]. In general, binning is the process of computing the number of values falling in a given interval or bin and storing them in a bin map. The data frequency can then be visually represented by the histogram.

Much previous work adopts bin maps which yields line-based histograms [BBP08, NH06, RPW*08]. They are effective at revealing clusters and outliers while further interaction support is needed to help the user select and brush interesting sub-sets of data and explore useful information. We find that the one-dimensional point-based histogram is effective in revealing an overview of the data [HLD02, Wil96], but such a histogram fails to depict the relations between the data axes. Geng et al. [GPL*11] extend the point-based histogram to a vector-based approach. They use the histograms as the visual aggregation of both the frequency and the direction of polyline-axis intersections. By introducing the angular information from the polyline-axis intersections, angular histograms and attribute curves are able to depict the relationship between data attributes.

3. Hybrid, Multi-Dimensional Visualization

In this section we present our novel visualizations to address occlusion, overplotting and the computational complexity associated with a naive hybrid visualization approach.

3.1. Challenges of Large, Time Dependent, Abstract Data

A naive coupling of parallel and spherical coordinates can be achieved by connecting polylines from the parallel axis to the spherical axis. In the spherical coordinate system the x, y and z components of acceleration are normalized and mapped onto the surface of a 3D sphere. Figure 1 (top) displays the result of this when applied to our time-dependent animal behavior data set consisting of 1,073,053 data samples and 13 attributes. The spherical axis suffers from overplotting and occlusion. In addition, user interaction (rotation and zooming) with the spherical plot is slower due to the large number of polylines. This provides the motivation behind our work, we have incorporated the following user-options and associated visualizations to mitigate these problems:

- sphere splitting to reduce occlusion on the spherical axes,
- curved polylines to enhance visual perception between the parallel and spherical axes.
- spherical, angular histograms to address overplotting and computational complexity on both the parallel and spherical axes,

We now discuss and display the results of these user-options.

3.2. Sphere splittng

We add a user-option to enable cutting the spherical coordinate system in half, displacing each sphere half by a user-defined distance. Our approach allows the user to change interactively the displacement size. This enables the user to specify a split which reduces occlusion. Figure 1 shows the results of a displaced coordinate system on our large data set. As well as reducing occlusion, this method can also reduce visual clutter. In addition, we provide the user with the ability to change the alpha value used to blend the polylines. This helps to fine-tune the visualization.

3.3. Curved Polylines

To enhance visual perception we present a user-option to utilize spline curves instead of straight polylines. We replace straight edges connected to the sphere with smooth curves that are orthogonal to the sphere's surface. Figure 1 demonstrates this user-option. Polylines now display a curved structure connected to the acceleration vector value it represents in the spherical coordinate system. This makes it easier to perceive a polyline's corresponding data value and assists in discovering a polyline's intersection point on the reference sphere.

We utilize hermite spline curves [Fol93] in the following way. The start point is on the parallel axis, its tangent vector is orthogonal to the parallel axis, the endpoint is the coordinate of the normalized acceleration vector, and its tangent vector is orthogonal to the sphere at the endpoint. The user may control the vector coefficients so they can modify the curvature of the polylines to enhance their perception.

3.4. Spherical Angular Histograms

Angular histograms, introduced by Geng et al [GPL*11] provide a method of displaying an overview of the trends in a data set within a parallel coordinates visualization. We adopt and extend angular histograms in order to address both overplotting and computational complexity. This technique is used as an overview to locate regions of interest with in the data set, prior to using a polyline based visualization to show a detailed representation of the data. We modify the original visualization by splitting angular histograms on parallel coordinate axes neighboring the spherical coordinate plot when the spherical coordinate system is split, this encodes both upward and downward polyline information.

To incorporate both frequency and angular information into the spherical coordinate system, we extend the spherical histogram technique presented by Grundy et al [GJL*09] to encode angular information. Frequency is represented by bar height, and angular information through the bars curvature together with neighboring axes histograms. As seen in Figure 1 (bottom), the curve is effectively pulled towards the average angle. A playback function is used to represent the temporal component in our visualization. This is used to show the evolution of histograms over time. Because each curved bin represents polyline intersection frequency, the visualization is naturally suited for large data sets.

4. Conclusion and Future Work

In this paper we present work-in-progress solutions to integrated parallel coordinates with a spherical coordinate system. A naive approach suffers from multiple limitations including overplotting, computational complexity and occlusion. We presented the following user options to mitigate these: split sphere, curved polylines and spherical angular histograms. In future work we hope to investigate applying multiple spherical axis for data sets with multiple vector attributes, in addition to researching intuitive methods of integrating magnitude as well as direction into our visualization technique.
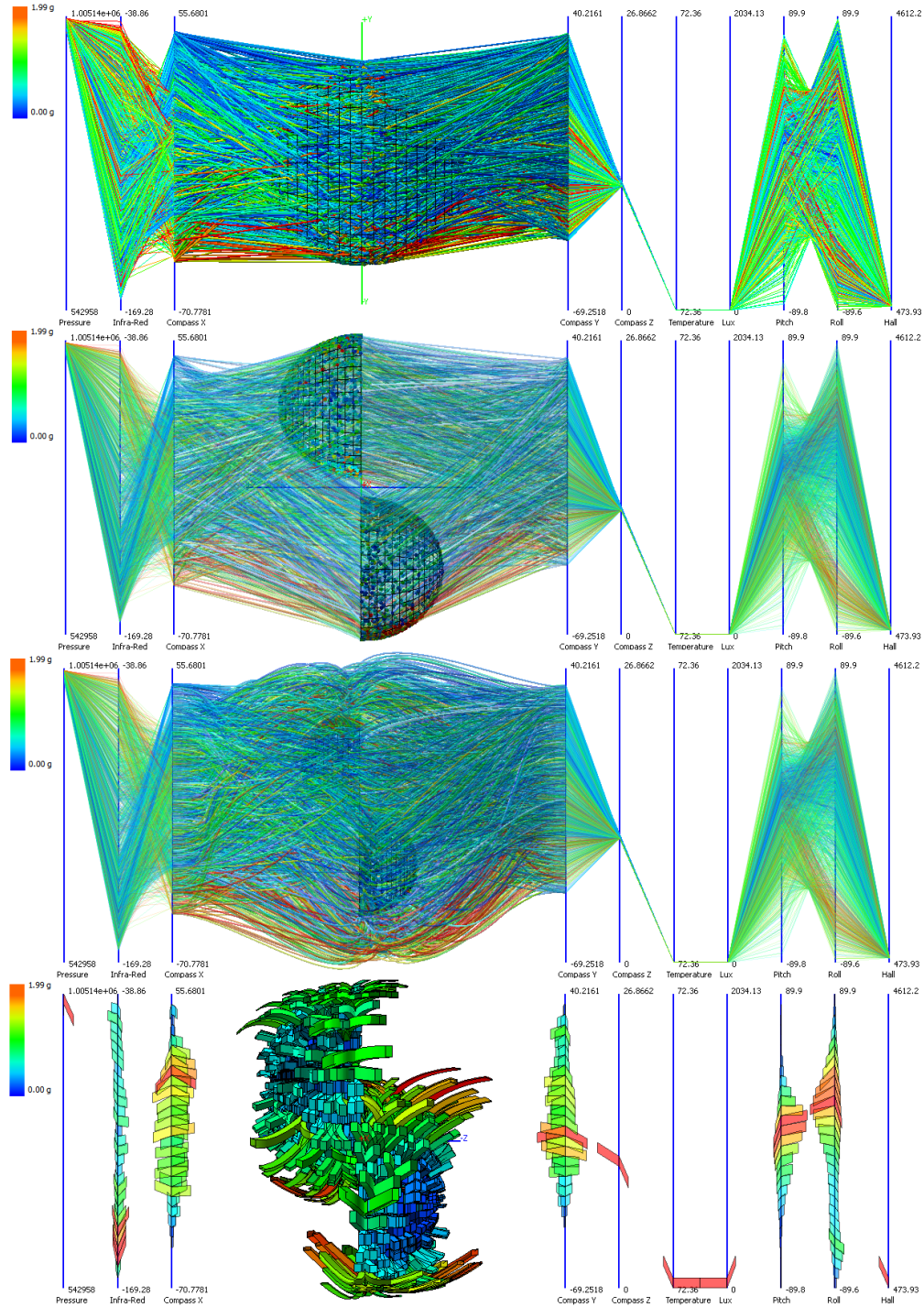
5. Acknowledgments

Figure 1: *Our user-options and their associated visualizations applied to a sample of 2764 time steps from our animal behaviour data.* **Top:** *A spherical plot naively added to a parallel coordinates plot. Polylines are drawn from the neighboring parallel coordinates axis to the spherical axis. The spherical plot is overplotted causing occlusion.* **Second from top:** *A split spherical coordinate system displaced by a user defined amount. The problems of occlusion and visual clutter are reduced.* **Third from top:** *Hermite spline curves are used instead of straight polylines to enhance visual perception. Data values are encoded into their corresponding polylines. It is easier to perceive a lines corresponding data value and its intersection point on the sphere.* **Bottom:** *Spherical angular histograms on the parallel and spherical axes display an overview of frequency and angles between neighboring axes. Histograms are split when the spherical axis is displaced to show bidirectional polyline information.*

References

[AdOL04] ARTERO A. O., DE OLIVEIRA M. C. F., LEVKOWITZ H.: Uncovering clusters in crowded parallel coordinates visualizations. In *Proceedings of the IEEE Symposium on Information Visualization* (Washington, DC, USA, 2004), IEEE Computer Society, pp. 81–88. 2

[BBG*09] BLAAS J., BOTHA C. P., GRUNDY E., JONES M. W., LARAMEE R. S., POST F. H.: Smooth graphs for visual exploration of higher-order state transitions. *IEEE Transactions on Visualization and Computer Graphics 15*, 6 (2009), 969–976. 2

[BBP08] BLAAS J., BOTHA C., POST F.: Extensions of parallel coordinates for interactive exploration of large multi-timepoint data sets. *IEEE Transactions on Visualization and Computer Graphics 14* (2008), 1436–1451. 2

[Car91] CARR D. B.: *Looking at large data sets using binned data plots*. Springer-Verlag New York, Inc., New York, NY, USA, 1991, pp. 7–39. 2

[ED06] ELLIS G., DIX A.: Enabling automatic clutter reduction in parallel coordinate plots. *IEEE Transactions on Visualization and Computer Graphics 12*, 5 (2006), 717 –724. 2

[ED07] ELLIS G., DIX A.: A taxonomy of clutter reduction for information visualisation. *IEEE Transactions on Visualization and Computer Graphics 13*, 6 (2007), 1216 –1223. 2

[FdOL03] FERREIRA DE OLIVEIRA M., LEVKOWITZ H.: From visual data exploration to visual data mining: a survey. *IEEE Transactions on Visualization and Computer Graphics 9*, 3 (2003), 378 – 394. 2

[Fol93] *Introduction to Computer Graphics*. Addison-Wesley Professional, 1993. 3

[FWR99] FUA Y.-H., WARD M. O., RUNDENSTEINER E. A.: Hierarchical parallel coordinates for exploration of large datasets. In *Proceedings of the conference on Visualization 1999* (Los Alamitos, CA, USA, 1999), VIS '99, IEEE Computer Society Press, pp. 43–50. 2

[GJL*09] GRUNDY E., JONES M. W., LARAMEE R. S., WILSON R. P., SHEPARD E. L. C.: Visualisation of sensor data from animal movement. *Comput. Graph. Forum 28*, 3 (2009), 815–822. 1, 2, 3

[GPL*11] GENG Z., PENG Z., LARAMEE R., WALKER R., ROBERTS J.: Angular histograms: Frequency-based visualizations for large, high dimensional data. *IEEE Transactions on Visualization and Computer Graphics 17*, 12 (2011), 2572 –2580. 1, 2, 3

[HLD02] HAUSER H., LEDERMANN F., DOLEISCH H.: Angular brushing of extended parallel coordinates. In *Proceedings of the IEEE Symposium on Information Visualization (InfoVis 2002)* (Washington, DC, USA, 2002), INFOVIS 2002, IEEE Computer Society, pp. 127–130. 2

[Ins85] INSELBERG A.: The plane with parallel coordinates. *The Visual Computer 1*, 2 (1985), 69–91. 1

[JLJC05] JOHANSSON J., LJUNG P., JERN M., COOPER M.: Revealing structure within clustered parallel coordinates displays. In *Information Visualization, 2005. InfoVis 2005. IEEE Symposium on* (2005), pp. 125 – 132. 2

[JTJ03] JR. J. F. R., TRAINA A. J. M., JR. C. T.: Frequency plot and relevance plot to enhance visual data exploration. *Graphics, Patterns and Images, SIBGRAPI Conference on 0* (2003), 117. 2

[KK96] KEIM D., KRIEGEL H.-P.: Visualization techniques for mining large databases: a comparison. *IEEE Transactions on Knowledge and Data Engineering 8*, 6 (1996), 923 –938. 2

[NH06] NOVOTNY M., HAUSER H.: Outlier-preserving focus+context visualization in parallel coordinates. *IEEE Transactions on Visualization and Computer Graphics 12*, 5 (2006), 893 –900. 2

[QCX*07] QU H., CHAN W.-Y., XU A., CHUNG K.-L., LAU K.-H., GUO P.: Visual analysis of the air pollution problem in hong kong. *IEEE Transactions on Visualization and Computer Graphics 13*, 6 (2007), 1408 –1415. 2

[RPW*08] RÜBEL O., PRABHAT, WU K., CHILDS H., MEREDITH J., GEDDES C. G. R., CORMIER-MICHEL E., AHERN S., WEBER G. H., MESSMER P., HAGEN H., HAMANN B., BETHEL E. W.: High performance multivariate visual data exploration for extremely large data. In *Proceedings of the 2008 ACM/IEEE conference on Supercomputing* (Piscataway, NJ, USA, 2008), SC '08, IEEE Press, pp. 51:1–51:12. 2

[UTH06] UNWIN A., THEUS M., HOFMANN H.: *Graphics of Large Datasets: Visualizing a Million (Statistics and Computing)*, 1 ed. Springer, July 2006. 2

[WB97] WONG P. C., BERGERON R. D.: 30 years of multidimensional multivariate visualization. In *Scientific Visualization, Overviews, Methodologies, and Techniques* (1997), IEEE Computer Society, pp. 3–33. 2

[Weg90] WEGMAN E. J.: Hyperdimensional data analysis using parallel coordinates. *Journal of the American Statistical Association 85*, 411 (1990), 664–675. 2

[Wil96] WILLS G. J.: Selection: 524,288 ways to say "this is interesting". In *Proceedings of the 1996 IEEE Symposium on Information Visualization (InfoVis 1996)* (Washington, DC, USA, 1996), INFOVIS 1996, IEEE Computer Society, pp. 54–. 2

[WL96] WEGMAN E. J., LUO Q.: High dimensional clustering using parallel coordinates and the grand tour. *Computing Science And Statistics 28* (1996), 361–368. 2