

# Enhancing Parallel Coordinates: Statistical Visualizations for Analyzing Soccer Data

Halldór Janetzko, Manuel Stein, Dominik Sacha (University of Konstanz, Germany)  
Tobias Schreck (Graz University of Technology, Austria)

## Abstract

*Visualizing multi-dimensional data in an easy and interpretable way is one of the key features of Parallel Coordinate Plots. However, limitations as overplotting or missing density informations have resulted in many enhancements proposed for Parallel Coordinates. In this paper, we will include density information along each axis for clustered data. The main idea is to visually represent the density distribution of each cluster along the axes. We will show the applicability of our method by analyzing the activity phases of professional soccer players. A final discussion and conclusion will complement this paper.*

## Introduction

Parallel Coordinates have a long history in information visualization. In 1880, Parallel Coordinate plots (PCP) were used by Henry Gannets visualizing ranks of the states of the United States of America for several categories. Today, the popularity of Parallel Coordinates is originating from the works of Alfred Inselberg systematically investigating the mathematical foundations and showing the usefulness for visualization purposes.

Parallel Coordinates allow to visually explore a high-dimensional data set by connecting for each data point the respective dimension values. With each data point being represented by a polyline, overplotting can cause severe readability problems. Consequently, many techniques enhancing the readability of Parallel Coordinates have been developed and proposed. Sampling and Smart Lenses reduce overplotting by plotting fewer data points. Aggregations and density visualization on each axis have been described to show the frequency distribution along each axis. Coloring of the polylines according to the numerical values of one axis has been discussed to reveal dependencies. We will discuss the proposed enhancements in detail in the subsequent related work section.

In this paper, we focus on the visualization of frequency distributions of categorized data points. Previous enhancements for Parallel Coordinate plots show only the global density distribution and do not regard additional class information. We integrated three state-of-the-art statistical visualizations representing density distributions along each axis regarding class information. Density Distributions will help to detect interesting patterns in highly overplotted Parallel Coordinate plots and interactive filtering and Brushing & Linking techniques help to narrow the analysis down to the interesting portions in the data. The contributions of this paper are as follows:

- Integration of frequency visualizations for multiple classes
- Highly interactive Parallel Coordinates with filtering,

correlation-based ordering, and Brushing & Linking

- Visual Analysis of movement phases in professional soccer matches

Soccer is a representative for geospatial movement with high degrees of overplotting and strong interdependencies between the movements of all actors. In our use cases, we analyze the movement of professional soccer matches recorded and provided by Prozone. We believe that collecting and analyzing soccer movement data will help coaches and professional soccer game analysts in their daily business. We proposed in [10] a comprehensive and versatile framework for the feature-based analysis of soccer matches. Based on this starting point, we clustered the movement into so-called movement phases and implemented Parallel Coordinates to investigate the resulting movement phases. Due to overplotting problems of Parallel Coordinate plots, we decided to enhance them by phase-aware density visualizations along each axis.

In the next section, we will outline previous works and relate them to our approach. In the third technique related section, we describe the technical details of our Parallel Coordinates implementation. Based on the applied clustering approach, use cases will exemplify the benefits of the proposed technique and highlight some interesting findings. In the discussion section, we will discuss our methods and potential perspectives for future work. Finally, we will conclude our paper by a brief summary.

## Related Work

Exploration of high-dimensional data is a fundamental problem studied extensively to date in Information Visualization and Visual Analytics. We briefly discuss related work from high-dimensional data visualization in general, and Parallel Coordinates and its extensions in particular.

### High-Dimensional Data Visualization and Parallel Coordinate Plots

Well-known visualization techniques for high-dimensional data include multivariate glyphs (mapping data variables to glyph shape properties), stacked displays (recursively mapping dimensions to a given schema), and scatter plot matrices (showing pairwise dimension combinations) [15]. The attribute explorer maps multiple data dimensions to parallel, vertical histograms, allowing to the brush and link data records [14]. Parallel Coordinate plots [9] show many data records by polylines intersecting a vertical, parallel coordinate setup at the respective positions. Parallel Coordinate plots support many tasks in visual high-dimensional data analysis, including finding clusters of records, correlating dimensions, and identifying outliers.

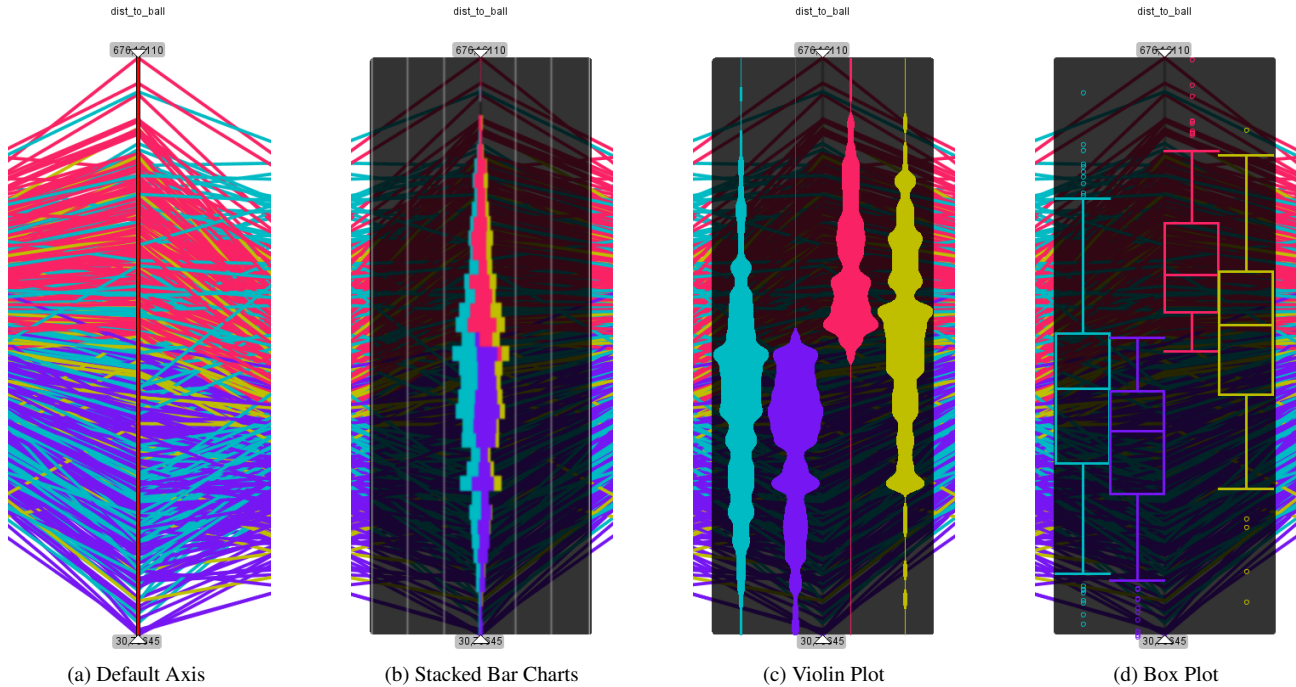


Figure 1: Implemented frequency visualizations for Parallel Coordinates plot to overcome overplotting issues. The cluster frequencies are computed and visualized along the dimension axis.

While typically used for quantitative data, there exist extensions to categorical data [2]. Among others, the idea of parallel coordinate axes has also been applied to glyph layouts as shown in [5]. A good overview of techniques for Parallel Coordinate plots can be found in the state-of-the-art report by Heinrich and Weiskopf in [8].

**Visual Extensions to Parallel Coordinate Plots** Extensions have been proposed to improve the basic PCP, motivated by the need to scale with large numbers of records and dimensions. Sorting of dimensions is useful to reduce clutter and overlaps [4], and help users navigate better. In [16], dimensions are clustered and grouped by similarity, also removing outliers from the display. In [1, 6], the use of opacity bands is proposed to visually group (aggregate) similar polylines, helping to distinguish data clusters and distinguish dimension intervals they reside in. Another visual extensions are the use of so-called edge bundles, which visually aggregate similar polyline segments, again helping with distinguishing groups and reducing clutter. In [12], a density-based clustering is applied per dimension, and used together with an efficient rendering algorithm. In [11], Illustrative Parallel Coordinate plots are proposed. The approach includes several area-based abstractions for clusters, giving rise to a larger abstraction design space to chose from.

**Analytical Extensions to Parallel Coordinate Plots** Also several analytical extensions of Parallel Coordinate plots exist. Depending on the data set and analysis task, there may exist subsets of dimensions which could be aggregated and considered jointly, without dedicating one parallel coordinate to each dimensions. In [17], the authors propose to summarize subsets of dimensions

by means of dimensionality reduction. Specifically, Multidimensional Scaling is applied to form 2D Scatter Plots (SP) for selected dimension subsets: The Scatter Plots are shown inline together with the Parallel Coordinate plots, with connectors relating data between both representations. Furthermore, there are works which define interest measures which can in turn be used to filter dimension subsets or find appropriate dimension orderings. The so-called Pargnostics approach [3] uses measures for pairs of coordinates like number of line crossings to rank and select good views. In [13], among others, the HoughTransform image analysis method was proposed to identify PCP views of potential interest to a user.

**Distinction to our Approach** In our approach, we propose a diagrammatic extension to the Parallel Coordinate plots. We observe that e.g., the Opacity Bundle technique [6] allows well to discern clusters by their overall shape differences (across all axes). However, this and other cluster-based techniques are often less effective in discerning clusters with respect to data intervals spanned on individual axes. To this end, we extend the visual depiction of each PCP axis by different density visualizations, indicating the distribution of clusters across the axis values. Specifically, we propose to use stacked bar charts, violin plots, or box plots visualizing a binning (histogram) over the individual dimensions. Together with a color-coding consistent with the coloring of the polylines, this can provide a more complete representation of cluster properties and differences.

## Technique

In this section, we will describe our implemented Parallel Coordinate plots in detail. We will focus on both the visual en-

hancements and the interaction capabilities. In order to show the implementation in a real application, we provide a video at <http://files.dbvis.de/files/vda2016janetzko.mp4>.

### Visual Enhancements

Overplotting of lines is a common problem occurring in Parallel Coordinates visualizing medium amounts of data. Other methods like sampling or global aggregations hide many of the underlying patterns. As we employ Parallel Coordinates to investigate segments of trajectories being clustered it is important to see local distribution patterns. Therefore, it is crucial to explore the value distributions for the different clusters in each dimension. We consequently support the user in investigating the Parallel Coordinates plot by integrating several state-of-the-art statistical density visualizations. We visualize the frequency distribution of clusters along a dimension axis as exemplified in Figure 1. In Figure 1a), the default axis of a Parallel Coordinates plot is depicted. The painting order and the overplotting are highly impacting the visualization. We implemented stacked bar charts in Figure 1b) being normalized according to all dimensions. Violin plots, shown in Figure 1c), are commonly used in Statistics to show a kernel density estimation of the density distribution. For further statistical measures (e.g., median, quartiles, and outliers), we integrated box plots as depicted in Figure 1d). Although the implemented visualization technique have drawbacks concerning scalability with respect to the number of different clusters, the analyst get a feeling for the feature distribution of clusters and interdependencies of dimensions. We will discuss merits and drawback in detail in a later discussion section.

### Interaction Capabilities

We integrated several interaction and visual boosting techniques into our Parallel Coordinates implementation. In Figure 2, we show a schematic depiction of our Parallel Coordinates implementation for two dimensions with focus on the filtering capabilities. We added to each dimension interactive range selectors providing the following interaction possibilities:

- (I) The upper boundary of the respective selection can be either moved individually or all upper bounds can be moved simultaneously.
- (II) Instead of increasing or shrinking the selected range, the analyst can just drag the selection range on the axis up and down.
- (III) The lower selection limit can be either dragged individually or all lower limits can be moved simultaneously.
- (IV) Mouse hovering will highlight only the data items fulfilling the currently applied filtering criteria.

Allowing the user to simultaneously change all upper and lower limits helps in performing manual nearest-neighbor queries. In this case, the filter intervals would be initialized by the system to fit one single, user-selected Parallel Coordinates line. The analyst is then able to adjust all upper and lower filter boundaries simultaneously. All lines in the Parallel Coordinates plot fulfilling the filter criteria (denoted by the blue hatched area in Figure 2) are drawn unblurred. All other lines are blurred to guide the analyst's pre-attentive focus to date items fulfilling the filtering criteria.

Another very important aspect in Parallel Coordinate plots is the ordering of dimensions. We provide an interactive reorder-

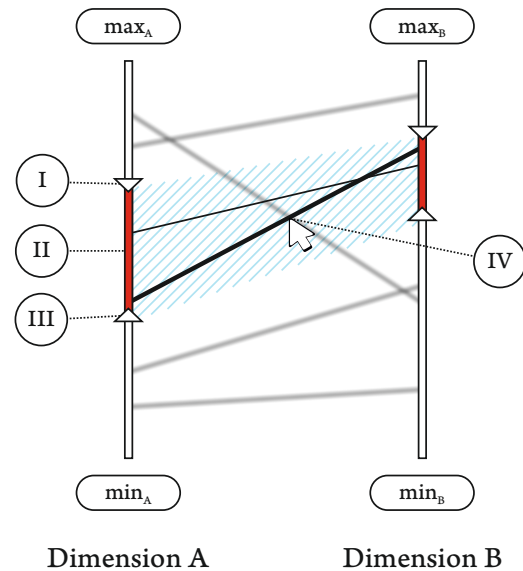


Figure 2: Schematic Parallel Coordinates implementation showing the interaction capabilities and the visual presentation of the filtering results.

ing feature by drag-and-drop. The analyst can freely change the ordering of the dimensions according to his interests. Additionally, we implemented a correlation-based arrangement of the dimensions applying the Pearson product-moment correlation coefficient. By clicking on a dimension labels all other dimensions will be either placed to the left (negative correlated) or placed to the right (in case of a positive correlation). The dimension are ordered by the absolute correlation value with the highest absolute numbers being placed nearest to the chosen dimension.

### Use Case

We will show the applicability of our enhancements for Parallel Coordinates by analyzing recorded soccer movement. The data analyzed in this section was provided within a collaboration with the sport analytics provider Prozone<sup>1</sup>. The data set consists of overall 66 professional soccer matches. For each of the 22 players two-dimensional position data are available with a temporal resolution of 100 milliseconds. Furthermore, the data includes manually annotated events (e.g., fouls, passes, crosses) containing information about position, time and event-specific information as the involved player. These events are less frequent and lack in accuracy as they are manually recorded.

### Clustering Activity Phases for Soccer

In this paragraph, we describe a clustering approach for soccer movements into activity phases previously proposed in [10]. Our use cases and the visual analysis of activity phases is based on the outcome of the clustering process.

In order to detect when a player is actively participating during a match, player features (e.g., speed, acceleration or distance to the ball) will be of use for this kind of analysis. The different activity phases are identified using clustering of the motion and feature patterns. The features being relevant for a single player

<sup>1</sup><http://www.prozonesports.com/>

analysis can be divided into two categories: *Individual Characteristics* (e.g., coordinates and speed) and *Game Context* (e.g., distance to ball). These time series with numerical attributes are analyzed in the subsequently described process.

Similar phases are found by clustering the above mentioned feature dimensions. We exemplify the overall analysis process in Figure 3. In a first step, we first partition all time series into small, fixed-size intervals and aggregate the values into a linear normalized, numerical feature vector describing the respective time interval. Furthermore, a Principal Component Analysis is applied in order to remove noise. We use the library WEKA [7] to perform the PCA and automatically reduce the number of dimensions with a threshold of 95 percent of the variance being still explained. Afterwards, we cluster the intervals with respect to user-chosen parameters. In our analyses, we apply k-Means (allowing us to control the number of resulting clusters) and DBSCAN (being a robust clustering technique with respect to noise and outliers). As a last step, we will merge adjacent intervals to larger phases, if they belong to the same cluster.

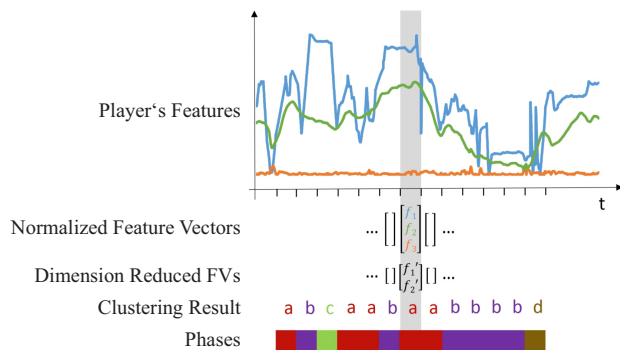


Figure 3: Feature-based approach to detect similar activity phases of a single player. Reprinted from [10].

### Single Player Analysis

The first step of our analysis is to investigate and explore movement patterns of single players. Analyzing single players will already result in overplotting issues making enhanced and interactive Parallel Coordinate plots necessary. We will first focus on the analysis of a defender and later on investigate the movement of a forward.

### Feature Analysis for Defender Movement

We introduced in a previous section our Parallel Coordinates implementation allowing interactive filtering and additionally visualizing the cluster distribution on each axis. In this section, we will investigate the clustering and segmentation results for a defender. We clustered the movement data using the following four dimensions: speed, acceleration, distance to ball, and distance to the nearest opponent. We applied k-Means clustering with a desired cluster number of four. Four clusters relate to the four phases in a soccer game described by ball possession and whether the teams are organized or not during ball possession switches. The resulting phases are depicted in Figure 4 with color representing the four clusters.

Without any further visualizations, the analyst is unfortunately not able to interpret the clusters completely. Nevertheless, there are some patterns visible by coloring the trajectory accord-

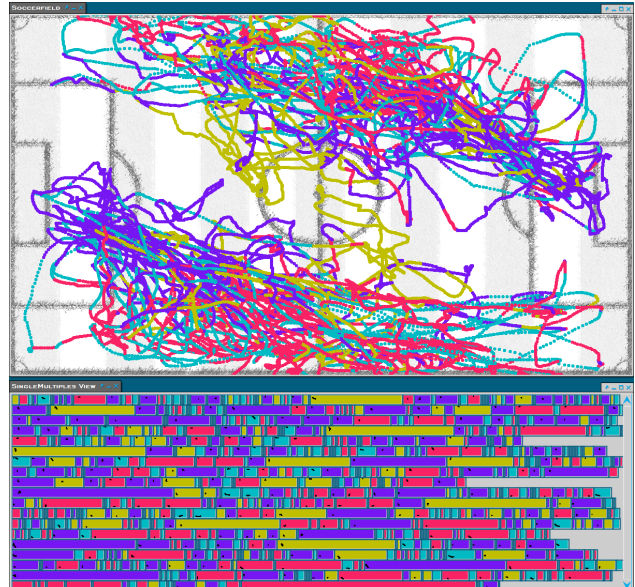


Figure 4: Clustering and segmentation results of a defender. The movement is colored according to cluster (top) and the temporal changes is depicted by colored bars with the width representing the length of the respective phase (bottom).

ing to cluster membership as shown in Figure 4. From a spatial perspective, the defender stays always on his assigned right side. More interesting and insightful is that the purple phases seem to be the only ones occurring around the own goal. All other clusters are mostly located outside the penalty area. We will further discuss this finding when analyzing the corresponding Parallel Coordinates visualization.

There seems to be no clear spatial explanation for the other three clusters (red, yellow, and turquoise). For this purpose, we integrated Parallel Coordinates visualization and enhanced them by a distribution visualization introduced previously. We visualize all phases of the defender's movement in a Parallel Coordinates plot and represent each single phase as one data item (one line in the Parallel Coordinates plot). We compute average values of each phase and use them in the Parallel Coordinates plot. The corresponding visualization are depicted in Figure 5.

We show in Figure 5 two different filtering steps during the analysis process. In the upper figure, the analyst selected one single phase to investigate the corresponding Parallel Coordinate line (highlighted by black borders). The filtering intervals will be automatically adjusted to fit the selected phase. As the analyst wants to understand the properties of yellow phases, he moves all range sliders simultaneously starting from the single selected yellow phase (lower figure). The analyst hovers over the previously selected line on the axis labeled distance to the nearest opponent, in order to similar phases with the same distance to the nearest opponent. All Parallel Coordinate lines at the mouse position fulfilling the filter criteria will be highlighted by black borders and will be rendered unblurred (lower figure). Analyzing the phases visualized in Figure 5 we were able to derive the following findings:

- Yellow phases correspond to movement with high distances to the nearest opponent, low speed, and low to medium dis-

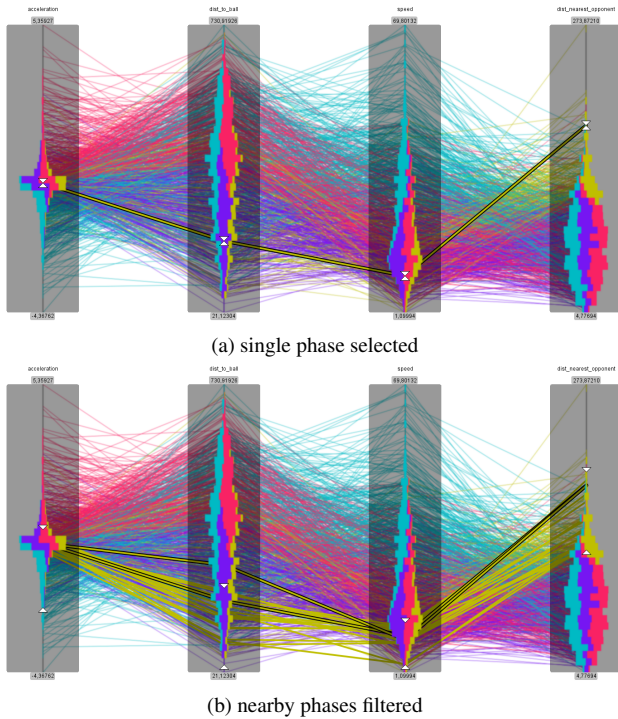


Figure 5: Parallel coordinate plots for segmentation results with interactive filtering. A phase of interest is selected (top) and interactively the filtering range is increased resulting in similar phases being selected (bottom). The data items emphasized by black borders are highlighted either by phase selection (top) or mouse hovering (bottom).

tances to the ball.

- Red phases describe movement with a high distance to the ball. Red phases have a positive acceleration and by trend lower speed compared to turquoise phases.
- Turquoise phases are independent of the distance to the ball and describe movement with negative acceleration. Negative acceleration values will only occur if the speed is sufficiently high.
- The purple phases being very visual salient in the geospatial representation are described by below-average values of distance to the ball, speed, and distance to the nearest opponent. Furthermore, the acceleration values are around zero.
- The difference between purple and yellow phases is only dependent on the distance to the nearest opponent. This is reflected in the spatial visualization as opponents are mostly near to defenders when opponents attack and defenders should be near their own goals during opposite attacks.

From these observations, we see that we need several views to the data. For instance, the difference between yellow and purple phases could be only fully understood when combining the spatial and the multi-dimensional feature visualization. We believe that combining several views and connecting them interactively by Brushing & Linking is an effective way to support the analyst. Visualizing frequencies along the dimension axes enabled us to reveal these interesting patterns.

### Analysis of a Forward

In the second part of our use case we focus on detecting offensive phases of a forward. A forward is a player that is mainly responsible for attacks of a team. Therefore, he tries to get in positions where he is able to score a goal. As the opposite team will try to hinder him he needs to be fast and move in a way that he has no player of the opposite team nearby. To analyze his phases accordingly we decided to focus on features that might describe dangerous phases. We initially decided for speed, distance to the next opponent, distance to the goal of the opposite team, and distance to the ball.

To compare the distribution of our different clusters we choose violin plots, see Figure 6. The turquoise cluster stands out for phases with high speed, which is very interesting for us because high speed indicates situations where a player is participating in, for example, a fast counter attack. As a result we filter for the turquoise cluster and phases with a higher speed. As we can see in the violin plots the clusters of the other features are more equally distributed, hence we switch to the box plot distribution visualization to earn additional details about the data distribution.

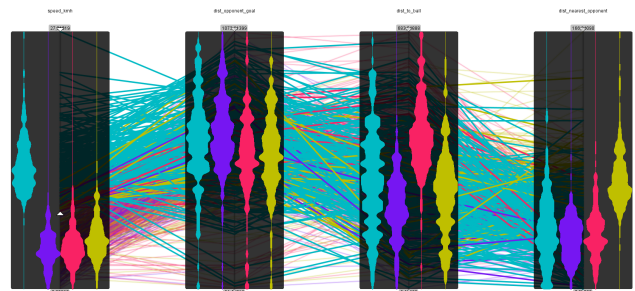


Figure 6: Violin plots denoting the density distributions of phases for a forward. The turquoise phases are dependent on the speed dimension.

We are interested in the distributions and the differences of the clusters, therefore we switch to the box plot visualization. As expected, the turquoise clusters differs significantly from the others in the speed dimension. Interestingly, there is no significant difference regarding the distance to the opponent goal meaning that the clustering did not use this dimension for partitioning. The whiskers in the third dimension (distance to the ball) describe that the lowest value in this dimension are common for the turquoise cluster.

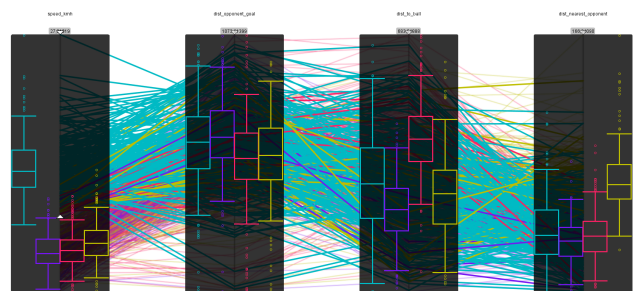


Figure 7: Box plots denoting density and statistical information for each cluster in each dimension. Statistically significant differences can be easily detected by comparing the boxes of the box plots.

## Multi-Player Analysis

In this section we used our Parallel Coordinates plot in combination with the soccer pitch trajectory visualization in order to investigate the behavior of multiple players at the same time. This enables us to compare feature characteristics (Parallel Coordinates) as well as the spatial movement (trajectories) in combination. Therefore we applied the clustering (as described in the previous section) to all players of interest.

**Back Runs** This use case investigates the behavior of the defense players within the back-four-formation of a team. First, we are interested in investigating typical dangerous defense situations which are fast attacks by the opposing team (also known as counter attacks). In this case, all the defenders have to run back towards their own goal in order protect it. Therefore we applied a clustering with many features that may be of interest. Note that it is not necessary to chose the optimal feature set because our experiences and evaluations revealed that an analyst is able to make sense of phase clusters by interactively exploring the Parallel Coordinates plot. As a first result the system shows visualizations that suffer from overplotting due to the massive amount of visualized data items. To investigate the data we can effectively make use of the enhanced axis-visualizations and filtering capabilities of the Parallel Coordinates visualization which is linked to the soccer pitch visualization. First, we filter the data for all back movements (towards the right hand side of the soccer pitch) by setting the filter for the x-direction axis (Figure 8-A). The resulting visualizations reveal some interesting patterns (Figure 8). From the Parallel Coordinates plot we can clearly see that the yellow phases are of interest. By investigating the axis plots (in this case the violin plots) we can see that the yellow phases cover higher speed (Figure 8-B) and we therefore do not need to filter this dimension further in order to detect runs (speed axis). In addition, these phases are much more straight than the others (Figure 8-C). From the x-axis plots and the selected (and yellow) data items we clearly see that the defenders run back most of the time (positive distribution of yellow phases in the x-pos axis) instead of towards the opposing goal (Figure 8-D). Also the soccer pitch visualization reveals a typical back run pattern with all the sprints (yellow phases) starting in the middle of the soccer pitch an ending in the own penalty area (Figure 8-E). In the soccer pitch all the selected situations for all defenders are drawn. Interestingly, all the phases ending within the penalty area are followed by a purple phase (Figure 8-F). By investigating the purple axis violin plots, we find that these phases cover less speed, low distance to the ball, and are near the own goal which can be interpreted as counter-attack-ending/finishing phases.

**Forward Runs** We analyze in this use case the defensive back-four formation in attack situations. In order to distinguish the different players we changed the color coding for all the phases. Instead of assigning a color to a specific cluster, all phases of the same player are drawn with the same color. All the phases of the players are shown in Figure 9. From the y-pos axis in the Parallel Coordinates plot we can clearly see that the purple and red players are the wide-defenders which are differentiated from the central defenders (yellow and blue, Figure 9-A). In order to reveal the forward runs of all the defenders we make use of the

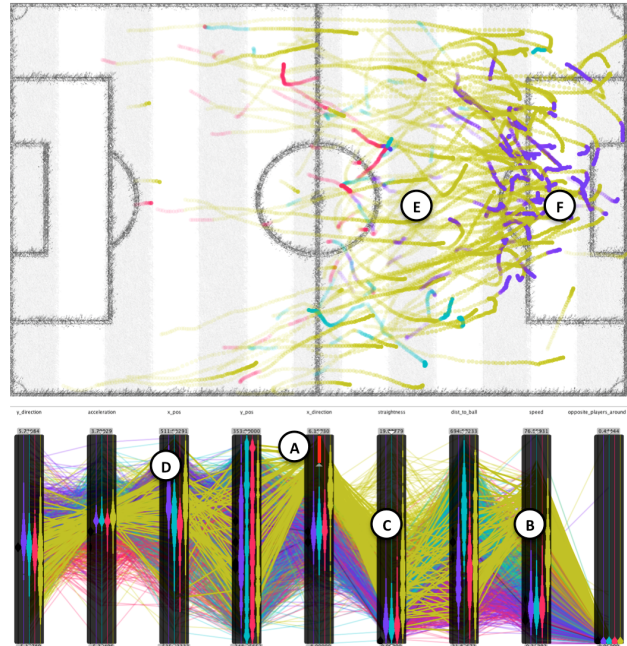


Figure 8: Back runs of all back four formation defenders. Top: Spatial trajectory visualization, bottom: Parallel coordinates enhanced with violin plots. Color represents the different activity phases of the players. Yellow back run phases are followed by purple finishing/ending phases near the penalty area of the own goal.

filtering capabilities. This time we set the filters for negative x-positions (Figure 9-B), and negative x-direction (Figure 9-C) in order to identify the movements on the left hand side of the soccer pitch and towards the opposing goal. In order to filter further, we set the speed filter for high speed phases (Figure 9-D). As a result the soccer pitch visualization shows all the situations where one of the defenders had a fast forward run. From the spatial visualization we are able to interpret that only the wide defenders participated in fast attacks (purple and red trajectories near the opponent's goal, Figure 9-E).

## Intensity Analysis

This section illustrates the benefits of the axis visualizations which enable to investigate vast amounts of data items within a single Parallel Coordinates plot. Therefore, we include all phases of all players of a whole soccer match in our analysis. In this case the colors represent again the different players. This last example analyzes stop moments of the different players which are especially interesting for fitness coaches because stop moments are high intense actions that have to be performed and are a valuable fitness indicator for soccer players. In order to identify the stop moments we created a Parallel Coordinates plot with the features *acceleration*, *straightness*, and *speed*. The box plots allow to compare the different player features directly within each axis. For example, we are able to compare the players by their speed distributions in order to find fast and slow players. We set the filters for high speed (Figure 10-A) and high-negative acceleration (Figure 10-B). The result is shown in Figure 10 where all the moments are also drawn onto the soccer pitch. A fitness coach is

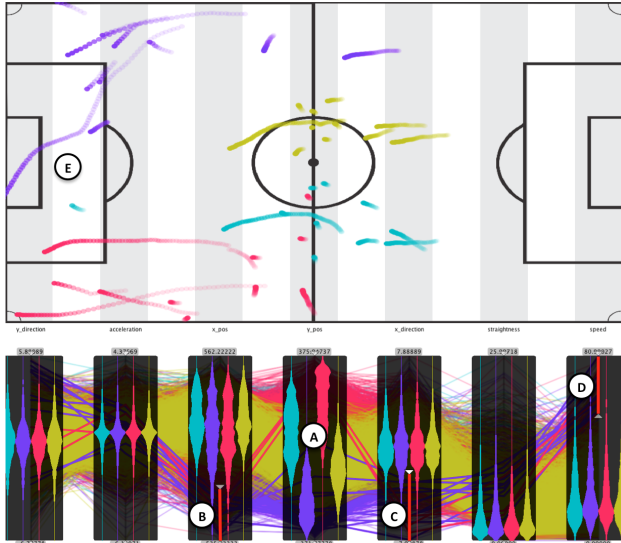


Figure 9: Forward runs of all back four formation defenders. Top: Spatial trajectory visualization, bottom: Parallel coordinates enhanced with violin plots. Color represents the different players. Wide defenders (purple and red) are involved in side attacks.

now able to hover over a trajectory in order to highlight all stop moments of this player which can be used to compare these moments also spatially. As another interesting aspect, all high intense phases (which are the stop moments) are mainly taking place on the left hand side of the soccer pitch (Figure 10-C) which is an indication that the team which is playing from right to left dominates the other team (tactically or physically). In order to analyse this further it would be interesting to observe the two teams separately (or to assign another color coding).

## Discussion

Instead of applying sampling or aggregation of polylines, we add further density visualizations to overcome overplotting issues in Parallel Coordinate plots. Cluster-aware density visualizations help to reveal data inherent structures otherwise hidden in overplotting polylines. The interactive filtering capabilities help furthermore to narrow down analyses to the interesting data portions.

While our density visualization is a useful extension for Parallel Coordinate plots analysis, there remain points to improve or understand better. First, as we use a stacked bar chart representation, we require an ordering of the classes per dimension histogram, which allows good perception of classes across dimensions. While such sorting can be done straightforward for individual axes, for a Parallel Coordinate plot with several axes this may be a difficult problem. On the one hand, a globally stable ordering of classes across all axes is desirable to retain the users' mental maps for comparing classes. This gives rise to a global sorting problem, and we would need to define appropriate tradeoffs for input to a sorting algorithm. The latter may depend also on user preferences and class sizes, among others.

Then, a natural limitation in using stacked bar charts is the limited perceptual distinction of colors. It is assumed that usually one cannot distinct more than about ten different classes by color. While other visual attributes like texture or shape could be

used to define other density histograms, how to do this and where the scalability limits are, remains to be researched in future work. Further, regarding human perception, there may be interrelations between the stacked bar charts and the Parallel Coordinate poly-lines. Specifically, for many records and classes, there may be clutter effects and/or unwanted pre-attentive perception effects. A solution may be to define a local quality function for the display that rates the perceptual effects. For perceptual difficult areas, based on rules the system could resort to techniques for lowering the clutter and perception effects. Sampling of polylines, or switching to edge bundles and routing the connectors appropriately around density histogram areas, could be possible as a response.

Recent works in Parallel Coordinate plots extension have included dimension reduction and adapted axis displays (i.e., Scatter Plots in [17]) to scale Parallel Coordinate plots with large numbers of dimensions. The idea of dimensionality reduction is also relevant to our approach. We may, for example, combine the idea of class-density visualization with in-line SP representations proposed in [17]. A first idea to do so could be to run a kernel density estimation for the resultant SPs, apply a contour-based density visualization, and link this to the colors of the remaining stacked bar chart representations. Finally, interest measures for the density plots could be defined and added into existing Parallel Coordinate plots quality measure systems like Pargnostics [3], eventually supporting automatic selection of best views or Parallel Coordinate plots segments.

## Conclusion

We presented in this paper an extension for Parallel Coordinate Plots tackling the overplotting problem. In comparison to other approaches, we use cluster information to show local density distribution by state-of-the-art density visualization techniques. We applied our implemented prototype to analyze real-world soccer matches. The investigation of clustered movement patterns with the help of our enhanced Parallel Coordinate Plots revealed interesting findings. We believe Parallel Coordinate Plots to be really powerful and beneficial for the in-depth analysis of clustered or annotated data.

## References

- [1] Gennady Andrienko and Natalia Andrienko. Parallel coordinates for exploring properties of subsets. *Coordinated and Multiple Views in Exploratory Visualization, International Conference on*, 0:93–104, 2004.
- [2] F. Bendix, R. Kosara, and H. Hauser. Parallel sets: visual analysis of categorical data. *IEEE Symposium on Information Visualization, 2005. INFOVIS 2005.*, (1):133–140.
- [3] Aritra Dasgupta and Robert Kosara. Pargnostics: Screen-space metrics for parallel coordinates. *IEEE Trans. Vis. Comput. Graph.*, 16(6):1017–1026, 2010.
- [4] Geoffrey P. Ellis and Alan J. Dix. A taxonomy of clutter reduction for information visualisation. *IEEE Trans. Vis. Comput. Graph.*, 13(6):1216–1223, 2007.
- [5] Elena Fanea, Sheelagh Carpendale, and Tobias Isenberg. An interactive 3d integration of parallel coordinates and star glyphs. *Information Visualization, IEEE Symposium on*, 0:20, 2005.
- [6] Y.-H. Fua, M. Ward, and E. Rundensteiner. Hierarchical parallel

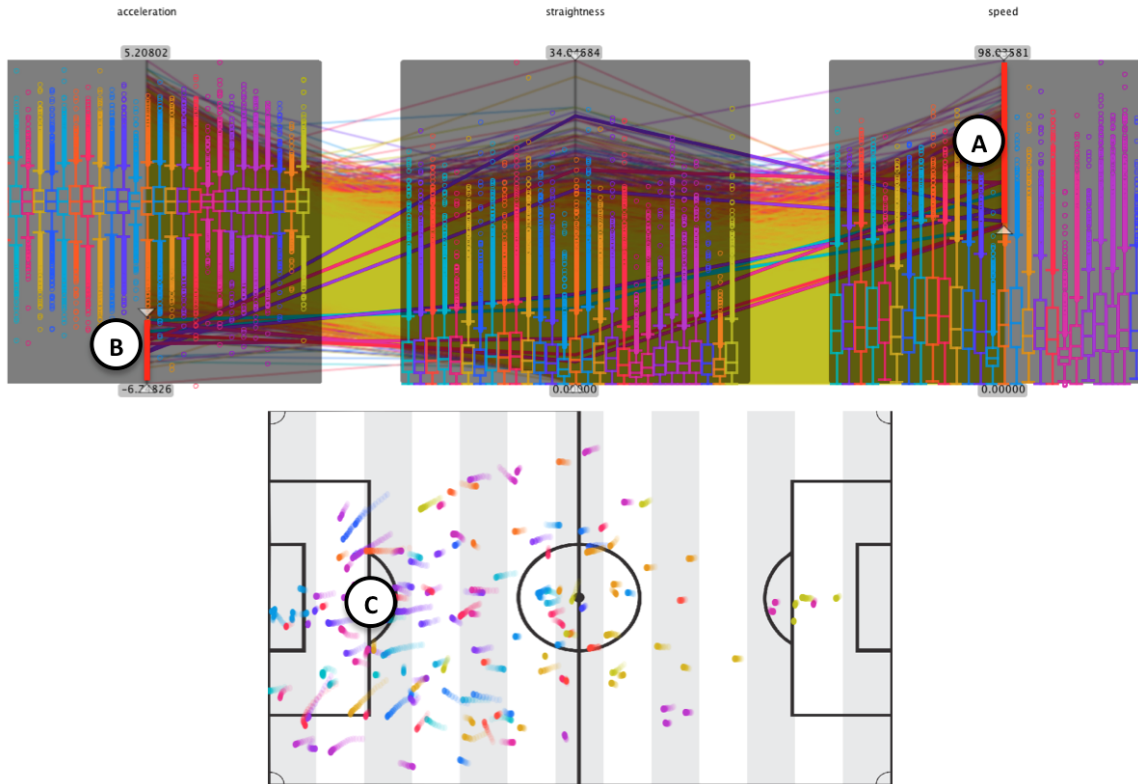


Figure 10: Stop moments of all players. Top: Parallel coordinates enhanced with axis box plots, bottom: Trajectory visualization. Color represents the different players. Stop moments mainly occur on the left hand side of the soccer pitch.

coordinates for exploration of large datasets. In *Proc. IEEE Conference on Visualization*, pages 43–50, 1999.

- [7] Mark Hall, Eibe Frank, Geoffrey Holmes, Bernhard Pfahringer, Peter Reutemann, and Ian H. Witten. The weka data mining software: An update. *SIGKDD Explor. Newsl.*, 11(1):10–18, November 2009.
- [8] Julian Heinrich and Daniel Weiskopf. State of the Art of Parallel Coordinates. In M. Sbert and L. Szirmay-Kalos, editors, *Eurographics 2013 - State of the Art Reports*. The Eurographics Association, 2013.
- [9] Alfred Inselberg. *Parallel Coordinates: Visual Multidimensional Geometry and Its Applications*. Springer, 2009.
- [10] Halldór Janetzko, Dominik Sacha, Manuel Stein, Tobias Schreck, Daniel A. Keim, and Oliver Deussen. Feature-Driven Visual Analytics of Soccer Data. *Proceedings of the 2014 IEEE Symposium on Visual Analytics Science and Technology (VAST '14)*, 20(12), December 2014.
- [11] Kevin T. McDonnell and Klaus Mueller. Illustrative parallel coordinates. *Comput. Graph. Forum*, 27(3):1031–1038, 2008.
- [12] G. Palmas, M. Bachynskyi, A. Oulasvirta, H.-P. Seidel, and T. Weinkauff. An edge-bundling layout for interactive parallel coordinates. In *Proc. IEEE PacificVis*, Yokohama, Japan, March 2014.
- [13] Andrada Tatu, Georgia Albuquerque, Martin Eisemann, Peter Bak, Holger Theisel, Marcus A. Magnor, and Daniel A. Keim. Automated analytical methods to support visual exploration of high-dimensional data. *IEEE Trans. Vis. Comput. Graph.*, 17(5):584–597, 2011.
- [14] L. Tweedie, B. Spence, D. Williams, and R. Bhogal. The attribute explorer. In *CHI 1994 Conference Companion*, pages 435–436, 1994.
- [15] M. Ward, G. Grinstein, and D. Keim. *Interactive Data Visualization: Foundations, Techniques, and Applications*. CRC Press, 2015.

- [16] Jing Yang, Wei Peng, Matthew O. Ward, and Elke A. Rundensteiner. Interactive Hierarchical Dimension Ordering, Spacing and Filtering for Exploration of High Dimensional Datasets. In *Proc. IEEE Symposium on Information Visualization*, pages 105–112, 2003.
- [17] Xiaoru Yuan, Peihong Guo, He Xiao, Hong Zhou, and Huamin Qu. Scattering points in parallel coordinates. *IEEE transactions on visualization and computer graphics*, 15(6):1001–8, 2009.

## Author Biography

*Halldór Janetzko received his M.Sc. (2010) and PhD (2015) in computer science from the University of Konstanz. Since 2010, he worked as a research associate at the Data Analysis and Visualization group at the University of Konstanz. Halldór Janetzko works currently as a post doctorate researcher at the Data Analysis and Visualization group at the University of Konstanz.*

*Manuel Stein received his M.Sc. in computer science from the University of Konstanz (2014). Since then he is a research associate at the Data Analysis and Visualization group at the University of Konstanz.*

*Dominik Sacha received his M.Sc. in computer science from the University of Applied Sciences Konstanz (2012). Since 2013, he works as a research associate at the Data Analysis and Visualization group at the University of Konstanz.*

*Tobias Schreck received his M.Sc. (2002) and PhD (2006) in computer science from the University of Konstanz. He held positions as a Postdoc Researcher and research group leader with TU Darmstadt (2007–2011) and as an Assistant Professor with University of Konstanz (2011–2015). Since then, he is a Professor with Graz University of Technology, Austria.*