Orientation-Enhanced Parallel Coordinate Plots

Renata Georgia Raidou, Martin Eisemann, Marcel Breeuwer, Elmar Eisemann, Anna Vilanova

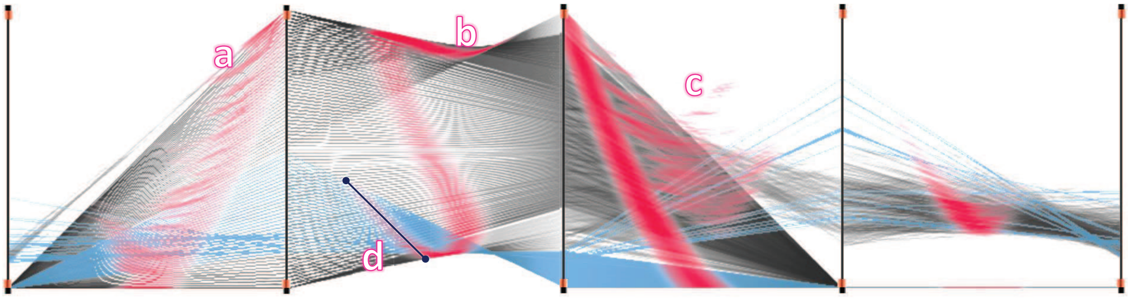

Fig. 1. The proposed OPCPs (red), applied to the `Venus` dataset [2]: (a) Visual enhancement of small patterns between the first two dimensions of the data, i.e., small structures obstructed by a strong pattern. - (b) Facilitated identification of distinct patterns between the second and third data dimension. - (c) Improved readability of outliers, i.e., low density information areas, in the representation. - (d) Efficient and accurate selection (blue) of a specific data structure, using the proposed O-Brushing (dark blue line).

Abstract—Parallel Coordinate Plots (PCPs) is one of the most powerful techniques for the visualization of multivariate data. However, for large datasets, the representation suffers from clutter due to overplotting. In this case, discerning the underlying data information and selecting specific interesting patterns can become difficult. We propose a new and simple technique to improve the display of PCPs by emphasizing the underlying data structure. Our Orientation-enhanced Parallel Coordinate Plots (OPCPs) improve pattern and outlier discernibility by visually enhancing parts of each PCP polyline with respect to its slope. This enhancement also allows us to introduce a novel and efficient selection method, the Orientation-enhanced Brushing (O-Brushing). Our solution is particularly useful when multiple patterns are present or when the view on certain patterns is obstructed by noise. We present the results of our approach with several synthetic and real-world datasets. Finally, we conducted a user evaluation, which verifies the advantages of the OPCPs in terms of discernibility of information in complex data. It also confirms that O-Brushing eases the selection of data patterns in PCPs and reduces the amount of necessary user interactions compared to state-of-the-art brushing techniques.

Index Terms—Parallel Coordinates, Orientation-enhanced Parallel Coordinates, Brushing, Orientation-enhanced Brushing, Data Readability, Data Selection

1 Introduction

Parallel Coordinate Plots (PCPs) [20] are widely used for the visualization of multivariate data. Here, multiple data dimensions are mapped one-by-one to a number of parallel vertical axes. Each multidimensional data object is mapped to a polyline that intersects the axes, connecting the scalar values of every dimension [20]. PCPs efficiently display in a single view all 2D projections of adjacent data dimensions [19, 21, 39], enabling the identification of relations and the detection of data patterns or trends - especially with the help of interaction [17, 34], such as brushing [16] or reordering [4, 30, 36].

A limitation of PCPs is that they might suffer from clutter, due to overplotting [17]. This causes problems in the exploration and interpretation of the data, especially in high density data. Reducing visual clutter in PCPs is an important topic [5, 11, 27, 29, 45]. However, most of the previous solutions are complex and focus mainly on aiding the detection of clusters in the data [5, 45], not in revealing the overall data structure. In other cases, the proposed visualizations may even unintentionally lead to concealing patterns and outliers [5, 29]. Finally, other solutions require interaction to achieve clutter reduction [11, 27].

We propose a simple technique to improve the representation of datasets in PCPs: the *Orientation-enhanced Parallel Coordinates* (OPCPs). Our technique visually enhances specific parts of each PCP line, depending on its slope. Hereby, it enables discerning individual trends and patterns, while it may even reveal patterns that are potentially obscured in traditional PCPs. This enhancement also allows us to introduce a new brushing technique, the *Orientation-enhanced Brushing* (O-Brushing), to facilitate pattern selection in complex data.

Our paper presents the following two major *contributions*:

- The concept of Orientation-enhanced Parallel Coordinates (OPCPs) to improve the view and discernibility of patterns in otherwise cluttered PCPs, without loss of low density data information or outliers.
- A versatile brushing technique based on the OPCPs: the Orientation-enhanced Brushing (O-Brushing). It enables efficient selection of individual data structures, with reduced user interaction.

2 Related Work

Many different techniques have been proposed for the enhancement of data display and clutter reduction in Information Visualization representations [9], including PCPs. Some approaches require the manipulation of the axes of the representation, using *reordering* [4, 30, 36, 42]. These approaches are able to reveal hidden patterns and facilitate data interpretation. However, in data with a large number of points reordering is insufficient. Other approaches involve *visual enhancement* of PCPs by rendering curves instead of lines [3, 14, 35, 45]. Such approaches are especially effective in reducing clutter at the crossings of PCP lines, but they might suppress data patterns, such as outliers.

Another commonly encountered group of techniques requires *clustering*, combined with different kinds of visual enhancements, such

Renata Raidou is with Eindhoven University of Technology and Delft University of Technology. Martin Eisemann is with TH Köln and Delft University of Technology. Marcel Breeuwer is with Eindhoven University of Technology and Philips Healthcare. Elmar Eisemann is with Delft University of Technology. Anna Vilanova is with Delft University of Technology. Correspondance e-mail: r.raidou@tue.nl.

*Manuscript received 31 Mar. 2015; accepted 1 Aug. 2015; date of publication 20 Aug. 2015; date of current version 25 Oct. 2015. For information on obtaining reprints of this article, please send e-mail to: tvcg@computer.org.
Digital Object Identifier no. 10.1109/TVCG.2015.2467872*

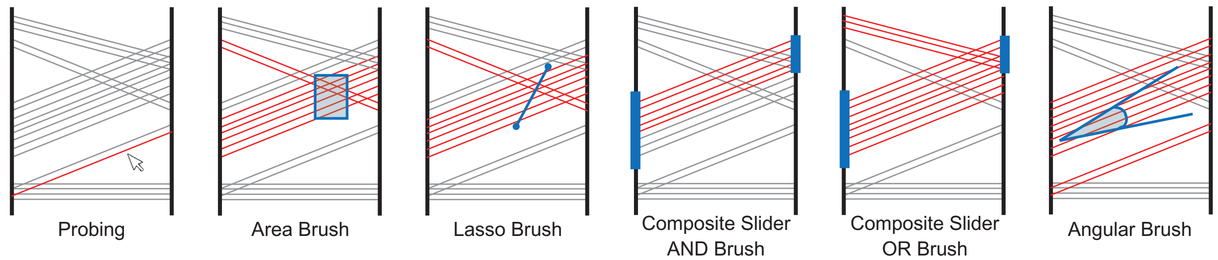

Fig. 2. Overview of different state-of-the-art brushing approaches for PCPs. With red we denote the selections in each case, with blue we denote the brushing operation.

as: manipulating PCPs by averaging polylines and visualizing correlation coefficients between polyline subsets [33]; filtering PCPs based on frequency or density of the data [5]; combining polyline splatting for cluster detection and segment splatting for clutter reduction [44]; using cluster-based hierarchical enhancements and proximity-based coloring schemes to provide a multiresolution view to the data [11]; enabling context visualization at several levels of abstraction, both for the representation of outliers and trends [29]; or using several transfer functions to reveal specific clusters and patterns in the data [22]. All previously mentioned cases involve clustering methods and they focus on detecting and differentiating specific clusters or trends in the data - not data patterns or underlying structures. In certain cases, clustering solutions inevitably lose information in low density areas, when reducing overplotting in high density areas. The approach of Zhou et al. [44] even requires animation, which is not always feasible. Finally, a more artistic approach was proposed by McDonnell et al. [27]. It incorporates a variety of techniques when rendering PCPs, such as edge-bundling, visualization of the distribution and density of the data via opacity and shading, or silhouettes for easy distinction of overlapping clusters. However, not all of these techniques can be used in a single view as some of them do not work well if combined.

PCPs have also been used in *combination* with other representations, such as Star Glyphs [10], radviz [6], or scatterplots [43]. As recognized by Holten et al. [18], combining scatterplots with PCPs outperforms many other PCP variants, such as combining with colors, opacity, curved polylines or animations. PCPs have also been combined with histograms [12], to simultaneously show the density and slopes of polylines. This combination enables the exploration of clusters, linear correlations and outliers in large datasets, with more emphasis on data-driven and not pattern-revealing exploration.

Interaction makes local and dynamic data enhancements possible. The use of lenses [8, 40] or brushing are typical examples. As part of the XmdvTool [37], a number of different brushes have been proposed by Martin et al. [26] and Ward [38]. Depending on the information that needs to be shown in the data, different brushes are used for highlighting, linking or masking the underlying data. Additionally, wavelet approximations are used to enhance brushing [41], by showing different parts of the polylines at different resolutions. However, brushing two variables in a non-separable way has only been enabled by the angular brushing proposed by Hauser et al. [16]. The most important state-of-the-art brushing approaches are presented in Fig. 2.

In summary, there are different approaches for data enhancement and readability improvement in PCPs. However, most of the solutions aim at reducing clutter in PCPs, by clustering the data, without giving a better understanding of the overall underlying structure. Data details, such as outliers, are often unintentionally hidden. Additionally, some solutions work better - or only - on a screen, either because they are animated or because they require interaction. Finally, most of the approaches require complex steps, which means that they cannot always be easily reproduced or used. In the following sections, we present our approach to handle these challenges.

3 Orientation-Enhanced Approach for PCPs

Our solution consists of two main components: the Orientation-enhanced Parallel Coordinate Plots (OPCPs) for visual enhancement of PCPs (Sec. 3.2), and the Orientation-enhanced Brushing (O-Brushing) for interactive selection and analysis in OPCPs (Sec. 3.3).

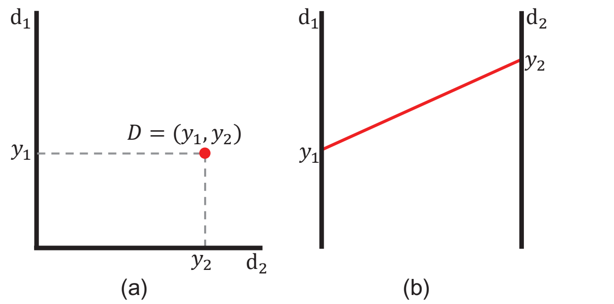

Fig. 3. Schematic representation of the concept of PCPs for the simple case of a two-dimensional point D with dimensions d_1 and d_2 and dimension values (y_1, y_2): (a) In a scatterplot. - (b) In a PCP.

3.1 Background: Parallel Coordinate Plots

In a simple two-dimensional dataset with dimensions d_1 and d_2, a data point $D = (y_1, y_2)$ is plotted as a PCP line, intersecting the two vertical axes d_1 and d_2 at the positions y_1 and y_2, respectively (Fig. 3). When plotting the PCP (poly)lines, it is common to employ opacity, as a simple way of representing the density of the lines [17]. From now on, We refer to this enhancement as *Density PCPs*.

3.2 Orientation-enhanced Parallel Coordinate Plots

Holten et al. [18] conducted an evaluation of PCP variants, where they demonstrated that no other enhancement from the examined alternatives improves PCPs significantly, apart from combining scatterplots with PCPs. Inspired by this paper, we investigated a simple way to combine effectively the two representations to enhance the display of PCPs. In many papers, the combination of PCPs with scatterplots has been limited to having multiple interactive linked views. This might entail memory limitations, as a result of switching the view between the two separate representations. The goal of the proposed visual enhancement of PCPs is to provide a better understanding in the visualized data, by integrating PCPs and their corresponding scatterplots in one view. A similar approach was followed by Yuan et al. [43]. However, this technique is complex and requires bending the polylines to fit to the points of the scatterplots. In contrast, we are looking for a simple approach that keeps the original appearance of PCPs intact. In the proposed OPCPs, the basic principle is to enhance the PCP lines with respect to their slope. This solution links PCPs and the corresponding scatterplot of the neighboring two axes in a natural way.

Mapping. For illustration purposes, we demonstrate our concept using a two-dimensional case. In the following example, we assume for simplicity that the data values for each dimension have been normalized to the range $[0,1]$. A PCP line, as shown in Fig. 4-(a), is defined by its dimension values (y_1, y_2) and a slope α:

$$\alpha = \frac{y_2 - y_1}{d_x}, \quad (1)$$

where d_x is the distance between the two vertical PCP axes.

We map this PCP line to a unique reference point $P = (x_p, y_p)$ in the space between the two PCP axes, with $x_p \in [0, d_x]$ and $y_p \in [0, d_y]$, where d_y is the length of the vertical axis (Fig. 4-(a)). The slope in Eq.(1) is linearly mapped to x_p, while y_p is chosen to make P lie on its corresponding PCP line:

$$x_p = \frac{d_x^2}{2d_y} \cdot \alpha + \frac{d_x}{2} \quad (2)$$

$$y_p = y_1 + x_p \cdot \alpha \quad (3)$$

In Fig. 4-(b), we show an example with multiple PCP lines and their respective reference points. Eqs.(2,3) result into a point-to-point transformation, i.e. a warping, of a 2D scatterplot space to the OPCP space, as shown in Fig. 5. This illustration shows the link between the scatterplot points and the reference point positions on the PCP lines.

Representation. To visually enhance each reference point and to preserve the orientation and context of its PCP line, we create a small line segment: the Orientation-enhanced PCP. It is a small segment that shares the original PCP line orientation and is centered at the reference point P. Assigning a constant intensity and a given length to each segment would result into OPCP segments that would not be visually separated, if they would be very close to each other. Therefore, we vary the intensity of the segments using a kernel smoother: we smooth the edges of the segments and assign higher intensities in the middle, i.e. at the reference point P (Fig. 6). This desired intensity profile can be achieved with peak-shape kernels, such as a Gaussian [13]. The intensity I at a point $S = (x_s, y_s)$ of the OPCP segment, resulting from a reference point $P = (x_p, y_p)$ after applying the Gaussian kernel, is described as:

$$I(x_s, y_s) = k \cdot \exp\left(-\frac{\|x_s - x_p\|^2}{2 \cdot \sigma^2}\right), \quad (4)$$

where σ is the bandwidth of the kernel, which is user-defined and k is a scale factor (height of the peak) given by $\frac{1}{\sigma\sqrt{2\pi}}$. The bandwidth controls the length of the OPCP segments: larger σ values result in smoother and wider-spread segments. Fig. 7 shows an example of OPCPs applied to three simple synthetic cases and the effect of σ on their appearance.

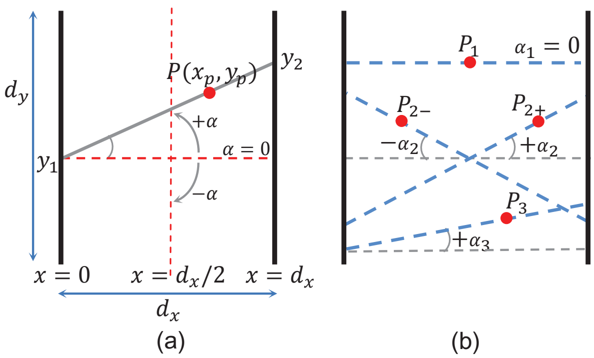

Fig. 4. Schematic representation of the concept behind OPCPs: (a) Mapping of the slope α of the PCP line (y_1, y_2) to the reference point $P = (x_p, y_p)$ between the two PCP axes. - (b) Mapping of the slopes of multiple PCP lines to their corresponding reference points.

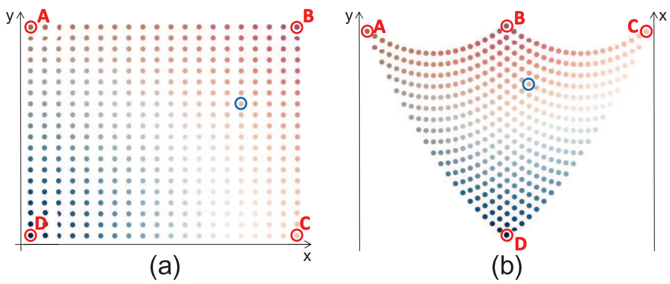

Fig. 5. Schematic representation of the transformation from (a) the scatterplot space to (b) the OPCP space, using a 2D colormap [32] to show the injective point-to-point correspondence.

Visual Enhancement. In the paper of Harrison et al. [15], it is stated that PCPs can emphasize specific correlations more than others. Depending on the data aspects that need to be emphasized, we propose

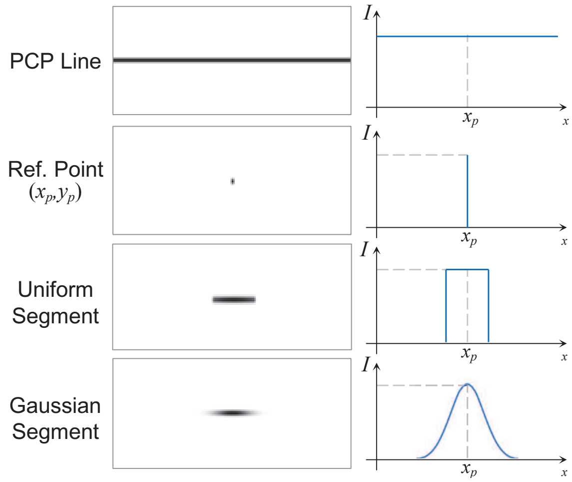

Fig. 6. Alternatives considered for the intensity encoding of the OPCP segments. Next to each case, we show also the intensity profile.

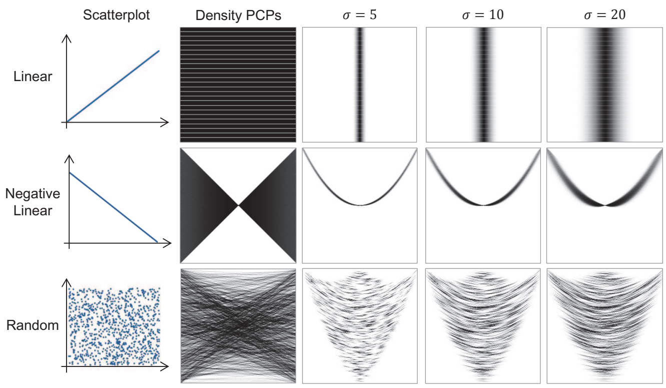

Fig. 7. Effect of the σ value of the Gaussian kernel on OPCPs, for three simple synthetic cases. To increase the visibility of the segments, we have linearly scaled the image intensities to the range [0,1].

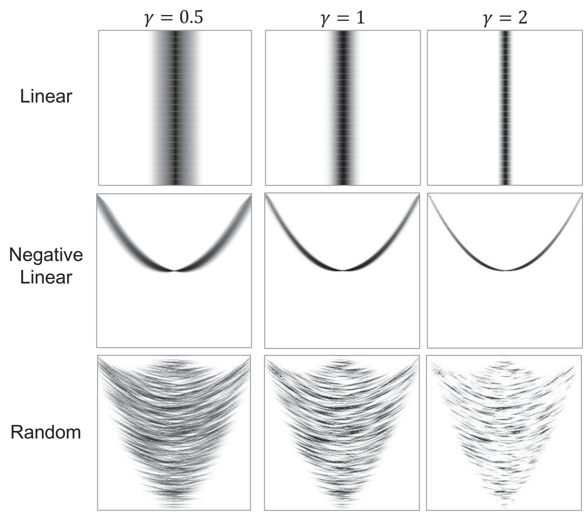

Fig. 8. Effect of the gamma correction on the appearance of the OPCPs. Here, the σ was set to 10 and the image intensities were scaled to the range [0,1].

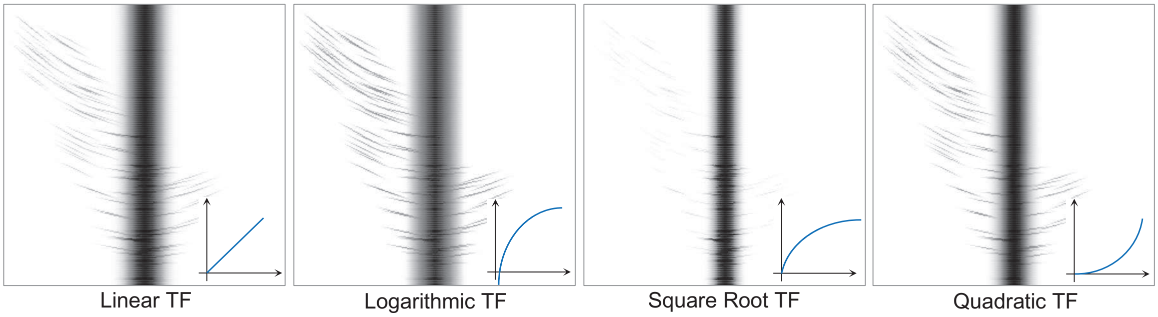

Fig. 9. Enhancement of different data aspects, using transfer functions (TFs). Here, we use a synthetic dataset with a dominant linear relation among the dimensions and a few outliers. The bandwidth σ was set to 10 and the image intensities were scaled to the range [0,1].

to optionally employ three enhancement methods: gamma correction, transfer functions, and histogram equalization.

Gamma correction [13] allows the user to remap the levels of the intensity range in order to discern more details in the darker parts of the OPCP segments. This can be accomplished with low values of gamma, while increasing values of gamma sharpen the OPCPs. Gamma correction is applied per pixel, transforming the intensity I to $I_{gcorr} = I^{\gamma}$. The effect of the parameter γ is depicted in Fig. 8.

The effect of gamma correction can be generalized by applying a transfer function (TF), aiming at controlling the contrast in the representation. As introduced in the work of Johansson et al. [22], different TFs affect the appearence of different data aspects: a linear TF gives an overview on the data, a logarithmic TF enhances low density areas, a square root TF emphasizes outliers in the data and a quadratic TF enhances the high density areas. The effect of the four previously mentioned TFs is shown in Fig. 9.

Optionally, histogram equalization [13] reassigns the intensity values of an image such that the output will exhibit a uniform distribution of intensities. Histogram equalization can create a background-foreground effect and enable better discernibility of different patterns in the data, especially in the presence of noise or of strong patterns. The impact of histogram equalization in OPCPs is depicted in Fig. 10.

Overlay. We enhance PCPs by overlaying the OPCP segments on top of the traditional PCP polylines, e.g., on Density PCPs, using alpha blending [31] (Fig. 11-(a)). Overlaying OPCPs on top of PCPs helps in preserving their main benefit, namely the connectivity across data dimensions. In this way, PCP polyline bundles can still be traced. Additional color encoding of OPCPs can enhance and visually separate them from the underlying PCPs (Fig. 11-(b)). To reduce as much as possible user distraction from overlaying OPCPs on the PCP polylines and interfering with PCP bundle tracking, the appearence of OPCPs can be adjusted. The user can modify the color and opacity of the OPCP segments, but also to fine-tune the σ and γ values to make the OPCPs more or less prominent. For the purpose of this paper, we decided to encode the OPCP intensities as black color values in the explanatory examples and red in the overlay examples.

Parameter values. The parameters involved in the visual enhancement of the OPCPs, such as the bandwidth σ and the gamma correction value γ, should depend on the specific aspects of the data that needs to be brought forward. Therefore, we do not assign a specific set of values, but leave them user-controllable. In our interactive tool, we initially assign a set of values (σ=10 and γ=1), which give already a good result, but can be changed adequately by the user.

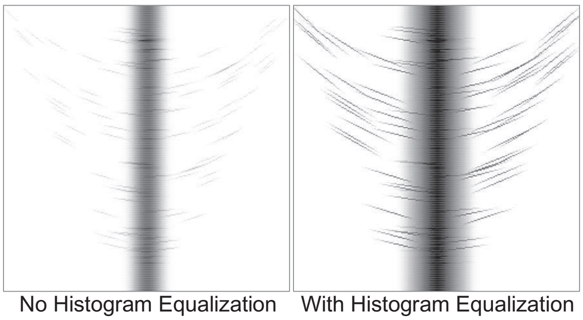

Fig. 10. Effect of histogram equalization on the appearance of OPCPs. Here, the σ was set to 10 and γ to 1.

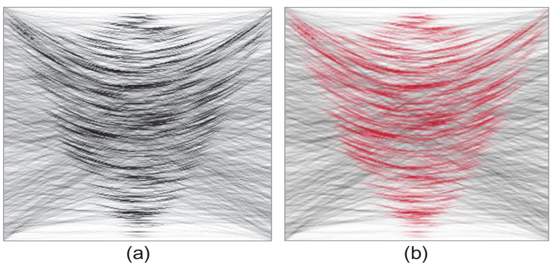

Fig. 11. Example of alpha blending in random noisy data: (a) Alpha blending of histogram equalized OPCPs with Density PCPs - (b) Color encoding of the foreground blended OPCPs (red) and background Density PCPs (black).

3.3 Orientation-enhanced Brush (O-Brushing)

Brushing [17, 34] is a common selection approach in PCPs. However, when the amount of plotted lines increases, selection also becomes difficult. OPCPs have an important property: they establish for each 2D data point a unique position in the space between each pair of the PCP axes. This allows us to introduce a new brushing approach that is applied in the OPCP space, for easier and more efficient selection of specific data patterns: the Orientation-enhanced Brushing (O-Brushing).

O-Brushing is performed on OPCP segments, in two ways: either with a traditional brush metaphor (O-Brush), or with a prober (O-Prober) (Fig. 12). The O-Brush acts as a lasso brush [17], applied only in the OPCP space and requires two user interactions, i.e., two clicks (Fig. 12-(a)). The O-Prober is an interactive rectangle that can be resized and moved around the representation (Fig. 12-(b)). It works similarly to an area brush, applied only in the OPCP space and re-

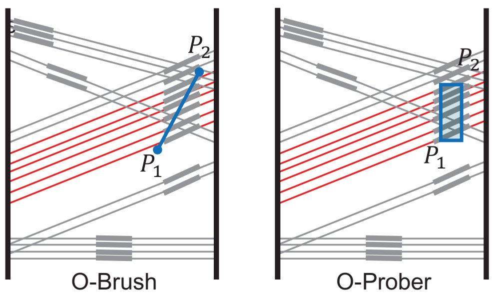

Fig. 12. Schematic representation of the concept behind O-Brushing. The thick gray segments represent OPCPs for each underlying PCP line. With red we denote the selections in each case, while with blue the brushing operation.

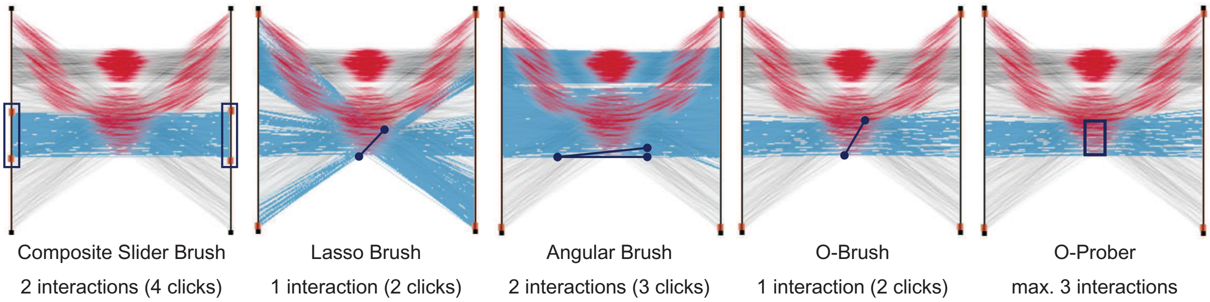

Fig. 13. Example for the comparison of O-Brush and O-Prober against traditional brushing methods, when attempting to select the same part, i.e., data points with middle values of both dimensions. All brushes have been applied individually to the data. In this example, the lasso and area brush do not succeed in selecting the specific data region. We show also the number of user interactions, i.e., the number of clicks, required for each of the brushes. The composite slider brush requires maximally four clicks, the O-Brush requires two clicks and the O-Prober requires maximally 3 user interactions (resize in both dimensions and translate).

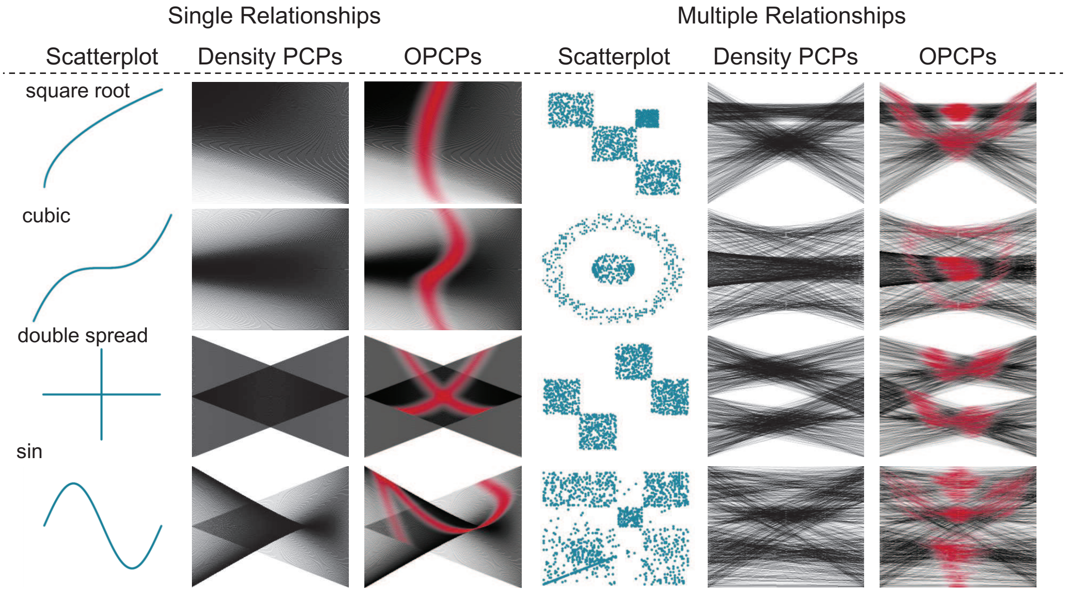

Fig. 14. Examples showing that the OPCPs allow the discernibility of (multiple) patterns or clusters in the synthetic data.

quires maximally three interactions, i.e., 2D resizing and translation. Compared to the traditional brushing methods, the O-Brush and O-Prober act only in the OPCP space, hereby allowing for a more precise and local selection and resulting in a reduced amount of required user interaction. The O-Prober and the O-Brush can produce the same result, but their main difference is that the former can be used to probe through the dataset for multiple similar patterns, i.e., lines with a given slope or a range of slopes, with minimal user interaction. In our current implementation, the O-Prober is a simple movable rectangle of user-defined size, but it could be easily extended to any arbitrary shape.

Fig. 13 shows a comparison of the O-Brushing methods with alternative PCP brushing methods. It depicts for each brushing technique the best achieved result and the user interactions required to select one specific pattern of interest. In this example, the specific selection is only possible with the composite slider brush and the two proposed O-Brushing methods. However, the O-Brushing methods require fewer user interactions than the composite sliders. In our interactive tool, we included also the state-of-the-art brushes, to enable users to perform selections both in the traditional PCP space and the OPCP space.

4 Results

In this section, we present the results obtained by the application of OPCPs and the O-Brushing to different datasets, intending to provide a deeper understanding into the OPCPs space and its characteristics. To this end, the visualization of the OPCPs was implemented in Python on the GPU, using OpenCL. The interaction for the brushing was realized using the Visualization ToolKit (VTK).

We tested the OPCPs on two different types of data. First, inspired by previous work [22, 43], we tested the behavior of OPCPs on a number of synthetic cases with two-dimensional data with 1000 data points, containing predefined patterns and structures. PCPs and OPCPs are meant for multidimensional data, but we employ these two-dimensional examples for illustration purposes. Secondly, to demonstrate a real usage scenario of OPCPs, we used multivariate data, obtained from various sources [1, 2, 23, 25].

4.1 Results with two-dimensional synthetic stimuli

In Fig. 14, we show our approach as applied to the synthetic stimuli, together with their corresponding scatterplots and Density PCPs. The transformation, i.e., the warping of the scatterplot space on the PCPs that was described in Section 3.2, becomes apparent. We confirmed that the OPCPs facilitate the discernibility of (multiple) data patterns, data outliers and also data structures obstructed by noise compared to PCPs, which we illustrate with examples, in the following paragraphs.

Discernibility of (multiple) data patterns. We include two main subcategories of relationships in the data: either there is a single relationship in the data, but it is not immediately recognizable, or there are multiple and more complex relationships (Fig. 14). In the first category, we included four different stimuli. For the cubic and square root stimuli, the OPCPs facilitate the identification of the different patterns compared to PCPs, because of the visible correspondance to the scatterplot space. Additionally, the double spread and sinusoidal stimuli, have a similar appearance when shown in the Density PCPs. However, the OPCPs allow to see that these are different patterns. Finally, for the second category, it is also easier to identify the multiple relations between the two data dimensions - or data clusters - when employing the OPCPs, as depicted in Fig. 14.

Discernibility of data outliers. We use two synthetic stimuli for which the dimensions are linearly correlated (Fig. 15). However, the second stimulus contains some outliers. By overlaying the OPCPs on the Density PCPs, we enhance the main pattern in the data, i.e., the linear relationship, without obscuring the outliers.

Discernibility of noise-obstructed data structures. We created two stimuli with initially no correlation between the two dimensions (Fig. 16). An additional pattern, i.e., a structure with a linear relationship between the dimensions, was then added to the second stimulus. In Density PCPs, this structure is hidden. By overlaying the OPCPs on the Density PCPs, we can visually enhance the obstructed structure in the data and recover the linear relationship.

4.2 Results with multivariate synthetic/real data

In a real-world analysis, PCPs are used to visualize multivariate data. To additionally assess our approach using more complex data with more realistic data patterns across their dimensions, we employ four well-known datasets from various databases. The employed datasets are the `apartments` dataset from [1] with 2290 data points, the `Venus` dataset from [2] with 8784 data points, the `Out5d` dataset from [2] with 16384 data points and the `household` dataset from [25] with 2075259 data points. Fig. 17 shows the results of using OPCPs to represent the above mentioned datasets. The OPCP advantages discussed in 4.1 are again apparent in these cases.

In the `apartments` dataset, especially between the first two data dimensions, multiple data patterns are emphasized and more discernible when using OPCPs (Fig. 17-a). This also occurs between the second and third dimension of the same dataset (Fig. 17-b). In this dataset, OPCPs are also able to bring forward outliers, e.g., between the second/third and third/fourth data dimensions, which were not easily discernible in the Density PCPs (Fig. 17-c). In the `Venus` dataset, the OPCPs facilitate the identification of distinct patterns, e.g., three patterns between the second and third data dimension (Fig. 17-e). Additionally, OPCPs allow to visually enhance the multiple small clusters between the first two dimensions (Fig. 17-d), as well as outliers between the third and fourth data dimension that are not visible in the respective Density PCPs (Fig. 17-f). In the `Out5d` dataset, pattern identification becomes easier throughout all dimensions, especially in parts of the representation, where the patterns are obstructed by noise, e.g., in the last three dimensions of the data (Fig. 17-g,h,i). Finally, in the `household` dataset, between the first two dimensions, but also between the third and fourth dimensions of the data, patterns that were not visible in the traditional Density PCPs are brought forward with the use of OPCPs (Fig. 17-j,k). Especially, in Fig. 17-k, there are two main patterns in the data, which appear as a single pattern in the Density PCPs. Also, between the last two data dimensions there are several outliers, which are significantly enhanced with the OPCPs (Fig. 17-l).

Based on the added benefits of OPCPs, it is expected that O-Brushing will facilitate the selection of the respective data structures in the OPCP space, in comparison to state-of-the-art brushing methods, which act in the PCP space, such as the lasso brush, the angular brush and the composite slider brush, as shown in Fig. 13 and 18.

4.3 Performance

The performance of our approach was tested on several datasets from various databases [1, 2, 23, 25]. The datasets vary between 1000 and 2 million data points. The test was conducted on an Alienware Aurora R4 with an Intel Core i7-4820K @ CPU 3.70GHz Processor, 16GB RAM and NVIDIA GeForce GTX 780. The performance results are depicted in Fig. 19. The system is implemented on the GPU and enables interactive brushing. The O-Brush and O-Prober can be employed for almost real-time data-driven selection.

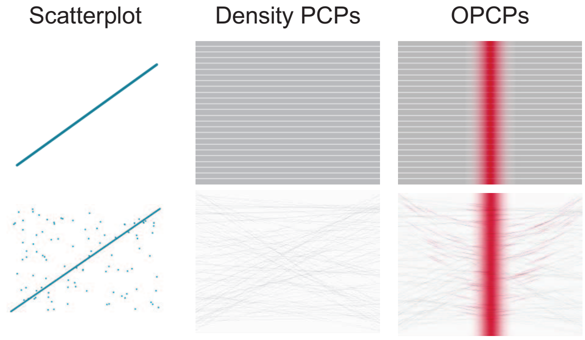

Fig. 15. Examples showing that the OPCPs enable the discernibility of outliers in the synthetic data.

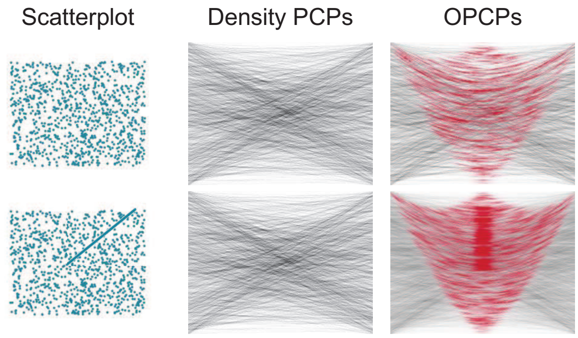

Fig. 16. Examples showing that the OPCPs enable the discernibility of noise-obstructed structures in the synthetic data.

5 EVALUATION

To test our approach with respect to the state-of-the-art, we conducted a user evaluation. We used the implemented interactive prototype, which enables visualizing data with Density PCPs and OPCPs, as well as data selection with all five brushing techniques: composite slider brushes, classical lasso brush [17], angular brushing [16], as well as our O-Brush and O-Prober.

The evaluation was designed based on the paper of Lam et al. [24] and consisted of two main parts. The first part was a controlled user study to measure User Performance [24] with the OPCPs and O-Brushing, against Density PCPs and traditional brushing. For this part, we performed three experiments, which are described in detail in the following subsections. The second part consisted of answering a questionnaire to measure User Experience [24], using Likert scales, ranking, and open questions.

We employed 16 participants, with various backgrounds: Computer Science (11, out of which 5 from Computer Graphics and 4 from Visualization), Electrical Engineering (3), Physics (1) and Biomedical Engineering (1). Most of them (9) had preliminary knowledge of PCPs, although only one participant had worked with PCPs before. Before the evaluation, we gave a short introduction, we demonstrated the functionality of the prototype, e.g., how to perform data selections with each method, and we allowed participants to use it, until they felt confident with it. In average, people spent around 5 minutes on the prototype before the experiment.

5.1 First Part: User Performance

We performed three experiments. The *first experiment* aimed at measuring the performance of users in discerning (1) patterns, (2) outliers, and (3) data structures obstructed by noise using Density PCPs or OPCPs. We created two comparable, two-dimensional synthetic datasets per case and we visualized them with both representations. Then, we showed static images of the representations to the users in a randomized order, and we asked them to perform tasks such as identifying and pointing out patterns in the data, outliers and noise-obstructed data structures. For each one of the static images, we mea-

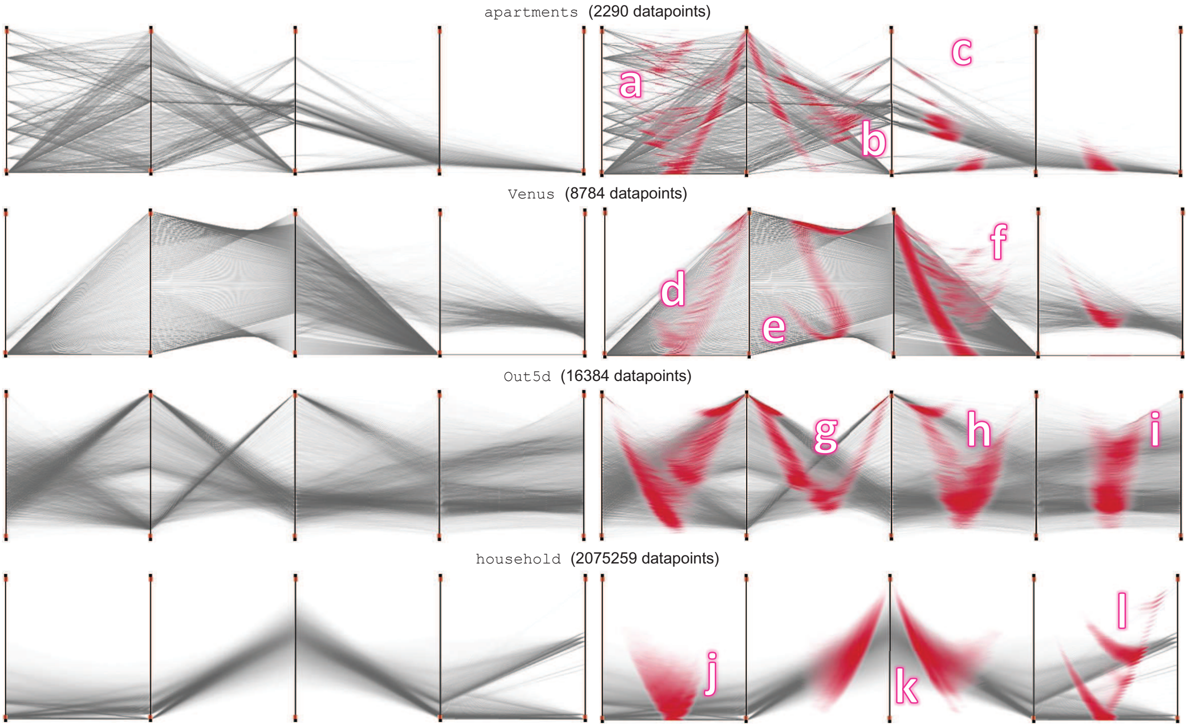

Fig. 17. Application of OPCPs to multivariate synthetic or real data obtained from various databases [1, 2, 25].

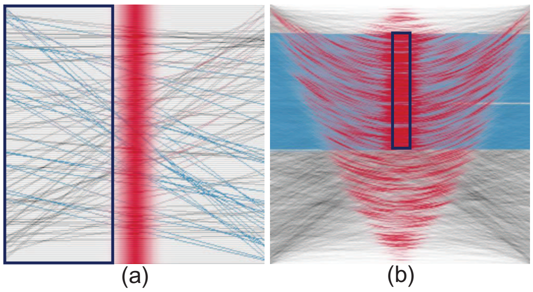

Fig. 18. Examples showing that the OPCPs allow the selection of (a) outliers and (b) noise-obstructed structures in the data.

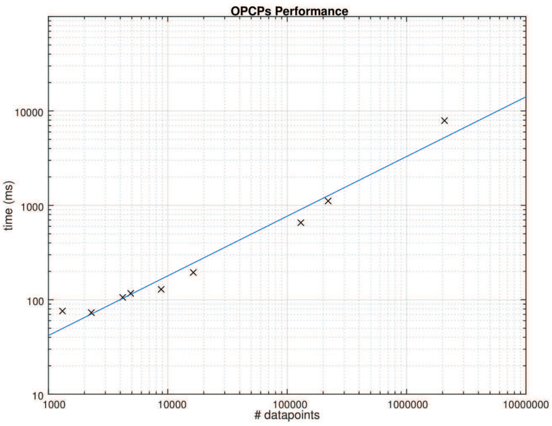

Fig. 19. Performance times of OPCPs for multivariate synthetic or real data from various databases [1, 2, 23, 25].

sured the time that users needed to give a conclusive answer and accuracy of their answers. Since the data were synthetic, we already knew the exact number of, e.g., patterns in the data. Thus, every wrong or unidentified pattern was penalized in the accuracy measurement.

The *second experiment* aimed at measuring user performance in selecting (1) specific patterns, (2) outliers, and (3) noise-obstructed data structures, with state-of-the-art brushing or O-Brushing. We created a two-dimensional synthetic dataset per case. We showed static images to the users, explaining which part of the data needed to be selected. Then, we asked them to perform a selection of the previously specified data part, using only one of the brushing techniques at a time, in a random order. All tasks were possible with all methods. Again, we measured time, accuracy, and number of interactions, i.e., number of clicks, required for task completion.

The *third experiment* aimed at measuring performance with multivariate, complex data and tasks. We created two comparable five-dimensional synthetic datasets, using the PCDC tool [7]. Then, we designed a set of questions, which were related to identifying and/or selecting data patterns, outliers and noise-obstructed structures. The users were asked to apply traditional brushing to one of the datasets with PCPs, and O-Brushing to the other dataset with OPCPs to perform the given tasks. The order of the dataset and approach, as well as their combination, was alternated randomly to reduce bias from learning. We measured completion time, accuracy and number of interactions, i.e., clicks required from the user, for the task completion.

The outcome of the statistical analysis of the experiments is summarized in Fig. 20. The first and third experiment were analyzed with Paired t-tests, while the second was analyzed with ANOVA and Tukey's HSD test, for statistical significance. The results of the *first experiment* indicate that identification of patterns, outliers and noise-obstructed structures is more accurate with OPCPs than with PCPs ($\rho<0.01$). Especially, in case where a structure is obstructed by noise in the data, the OPCPs were much more accurate ($\mu = 1, \sigma = 0$) than PCPs ($\mu = 0.06, \sigma = 0.25$). The distinction of noise-obstructed structures is also faster ($\rho<0.05$) in OPCPs: users required half the time to recognize these kinds of structures in the data with OPCPs than with PCPs. For pattern and outlier detection, there is no conclusive result for the time performance, but the accuracy is significantly improved with OPCPs. The outcome of the *second experiment* shows that O-Brushing is faster and more accurate ($\rho<0.01$) in all cases. For pattern and outlier selection, O-Brushing also requires significantly less interactions ($\rho<0.05$). From Tukey's HSD test, it results that there is no statistically significant difference between the performance of users when using the O-Brush or the O-Prober. Based on this test, the overall

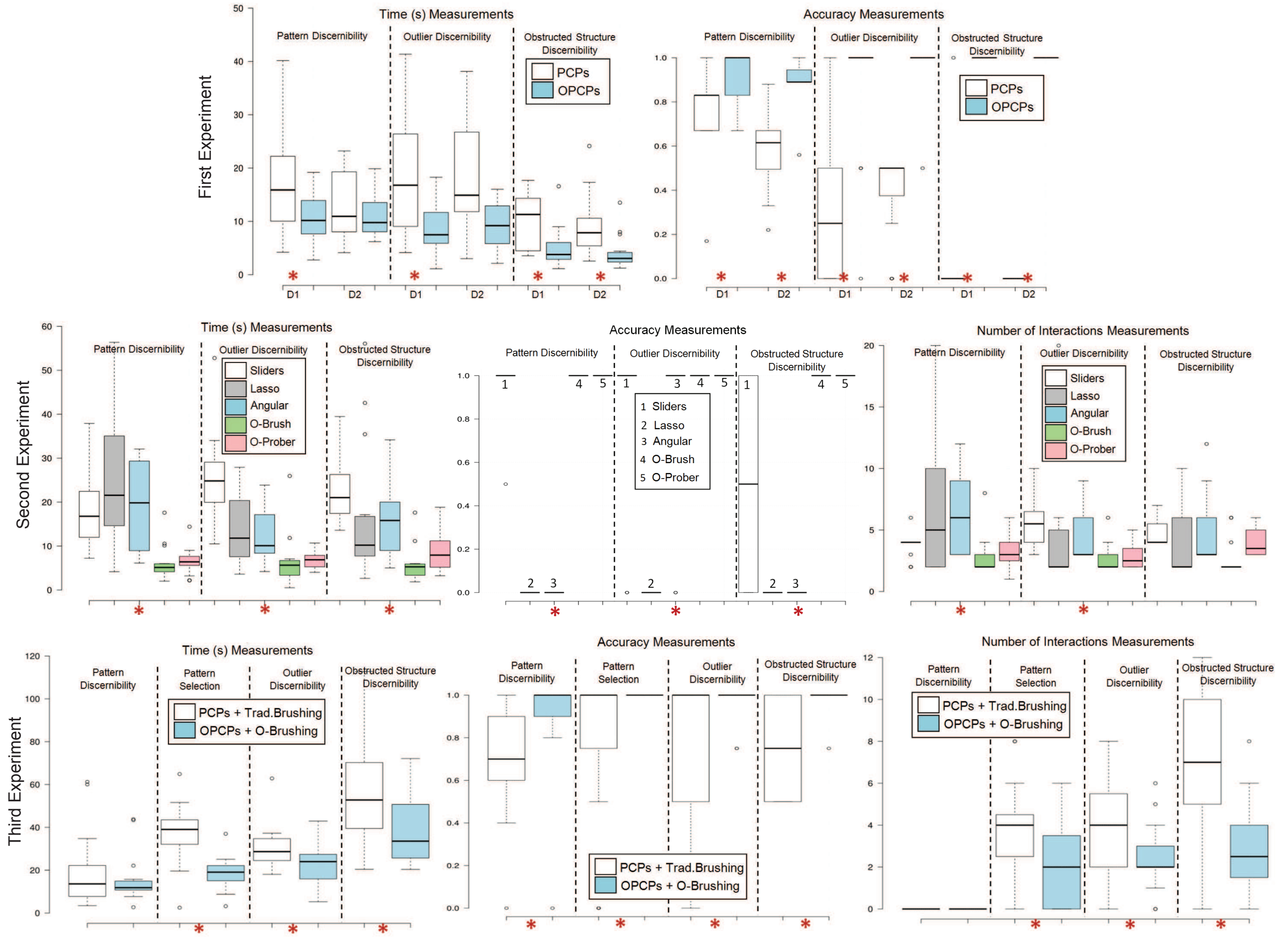

Fig. 20. Results for the experiments conducted as part of the evaluation, for the User Performance part. The small white circles denote outliers in the measurements. The asterisks denote a statistically significant difference (ρ <0.05) between the measurements, as it resulted from our statistical analysis using ANOVA and Tukey's HSD test.

ranking of the different brushing techniques for the three investigated tasks results as: the two variants of O-Brushing, angular brushing, composite brushing using sliders and lasso brushing. The results of the *third experiment* demonstrate that our approach is more accurate than the state-of-the-art approach for the four given tasks. The combined use of OPCs with O-Brushing had an average accuracy of 0.96 for all tasks, while traditional PCPs with standard brushing only 0.75. In this experiment, there were no indications that pattern discernibility requires less time with OPCPs. However, for the other three tasks the use of OPCPs and the proposed O-Brushing makes a big difference in performance times. For example, for pattern selection our approach requires half the time of the traditional approach. Overall, there is an indication that when selection is involved, our approach is also significantly faster and requires less interaction (ρ<0.05).

5.2 Second Part: User Experience

The second part of the evaluation consisted of conducting a survey. First, we asked users to grade PCPs and OPCPs and, also, the five previously used brushing methods, using *Likert scales*. The outcome of the statistical analysis of the experiments is summarized in Fig. 21. From the statistical analysis, it resulted that the PCPs were easier to understand, but the OPCPs were considered significantly easier to use (4.13), more useful (4.44) and also more suitable for the identification of patterns (4.44), outliers (4.38) and noise-obstructed data structures (4.31), compared to traditional PCPs. Moreover, the composite sliders and O-Brushing were considered easier to use and useful, while the easiest to understand were composite sliders and the O-Prober. The sliders and O-Brushing were considered most suitable for pattern selection, while for outlier and obstructed structure selection only O-Brushing was preferred. The next part of the questionnaire consisted on ranking the two representations using the same scale and the five brushing methods. The OPCPs were ranked significantly higher (8.31) than the PCPs (5.81) (ρ<0.05), while the O-Brush and the O-Prober were ranked significantly higher (8.25 and 8.29, respectively) than the sliders (6.50), the lasso (4.81) and the angular brushing (5.25) (ρ<<0.01). The questionnaire was concluded with open questions. The evaluation participants replied that the OPCPs "*can be very strong in structure detection in the data*", especially "*when there is a lot of overlap in the data*". However, "*the OPCPs take more time to get used to*" and "*might require some training for naïve users*". Also, finding "*simple correlations across dimensions can be easier sometimes with PCPs only*". O-Brushing makes it "*easier to select patterns locally*", but "*O-Prober could be improved by using also different shapes, other than the rectangle*". Most users commented that our approach supported them more in the identification and selection of patterns and outliers, in particular. For simple cases, due to the fact that OPCPs require prior familiarization and training, they might be less suitable. However, for cluttered data, the advantages are straightforward.

6 Discussion

The results of the application of OPCPs on synthetic and real datasets presented in Sec. 4, as well as the evaluation results of Sec. 5 raised

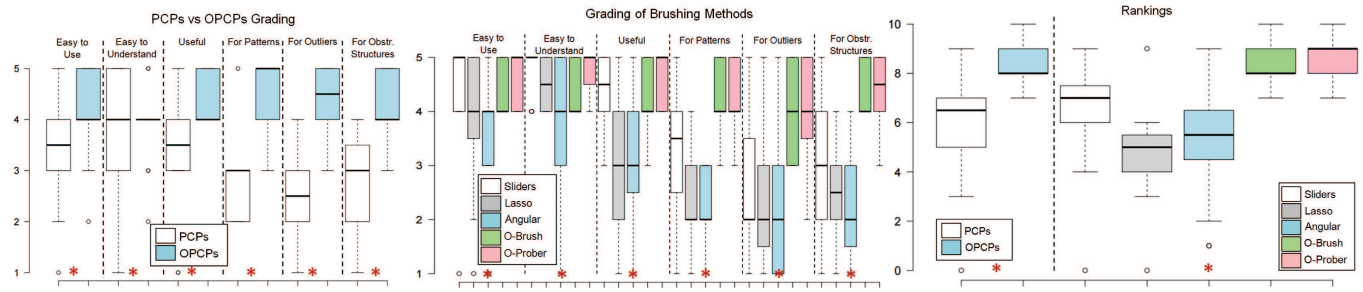

Fig. 21. Results for the experiments conducted as part of the evaluation, for the User Experience part. The asterisks denote a statistically significant difference ($\rho < 0.05$) between the measurements, as it resulted from our statistical analysis using ANOVA and Tukey's HSD test.

several points for discussion and limitations.

Firstly, the proposed OPCPs are a visual enhancement of PCPs for more efficient discernibility of patterns, outliers and noise-obstructed structures in the data. In the paper of Holten et al. [18], it is stated that combining scatterplots with PCPs can result in significant performance gains for the users. In many papers, the combination of PCPs with scatterplots is limited to having multiple linked views, where interactive linking and brushing can reflect selections from one representation to the other. However, in this case, users need to switch between windows and use their mental memory for data exploration and analysis, e.g., when the user performs an operation and sees the result in another window. Our OPCPs, instead, are not aiming at substituting scatterplots or at using linked scatterplot views. They focus on giving a better understanding of the data represented by PCPs, by integrating in a seamless way the two representations in one, combining their benefits and reducing the memory limitations that result from switching between the two separate representations. We consider that a comparison between scatterplots linked to PCPs against solely OPCPs is out of the scope of this paper, as the latter is a visual enhancement of PCPs and not a new representation on its own. In fact, as noticed by Holten et al. [18], PCPs might still be better than scatterplots in showing the actual shape of clusters, which is evident in our visualization. Still, OPCPs could be combined with linked scatterplots and it would also be interesting to investigate a comparison between scatterplots linked to PCPs and scatterplots linked to OPCPs.

Additionally, from the evaluation, it resulted that the interpretability of the patterns might not be straightforward and requires a certain level of familiarization with the enhancement. However, during the evaluation, the users were able to identify patterns more accurately than with traditional PCPs. Also, the cognitive load of OPCPs is not so significant to slow down the analysis of the data. As it can be seen in the evaluation results, in the vast majority of the tasks, the time needed to perform an operation using OPCPs and the related O-Brushing is significantly less than the time needed to perform the same operation with the state-of-the-art techniques. This is a first indication that the interpretability of patterns in OPCPs is not compromised. In a future additional evaluation, it would be interesting to research this further.

Moreover, there was no evidence so far in the user evaluation that the use of OPCPs might be distractive for the user or interfering with bundle tracking, which is the main advantage of the use of PCPs. OPCPs are indeed a new visual enhancement that requires some training, as pointed out by users. However, in our interactive tool the appearance of OPCPs can be adjusted by fine-tuning the σ and γ values to make the enhancement as prominent as the user would like. Also, there is always the option to adjust the color and opacity of the OPCP segments, to interfere less with the underlying PCPs and the polyline bundles. In the user evaluation, we included tasks where bundle tracking was necessary. In these cases, the users could perform the tasks without problems. However, for a more conclusive answer to this point, a more extensive study would be needed.

For the brushing functionality, in the interactive version of our tool, the user can select in which of the two spaces, i.e., PCP space or OPCP space, he/she would like to brush. In the OPCP space, the two proposed O-Brushing methods can be employed, while in the PCP space, state-of-the-art brushing, such as angular or lasso or composite brushing, can be used. As some users stated during the evaluation, having this possibility to choose the space to perform selections on the data is useful in different occasions: for example, if the user needs to perform selections based on the range values of some dimensions, the state-of-art brushing methods are more appropriate and more straightforward to use. However, if specific patterns, or outliers or structures in the data need to be selected, then O-Brushing is more efficient.

Limitations. We foresee some limitations of our approach. First, OPCPs require some familiarization, as they are not immediately intuitive. Additionally, they require a wider spacing between the dimension axes as compared to traditional PCPs in order to be effective. Moreover, the OPCPs should be accompanied by PCPs, to preserve context and connectivity across dimensions. Finally, the O-Prober could improve by using free-hand shapes or a scribbling interface instead of the predefined rectangle. This would enable easier, faster and more accurate selection of specific patterns with OPCPs, similarly to the shape-based method by Muigg et al. [28] for traditional brushing.

7 Conclusions and Future Work

Parallel Coordinate Plots exhibit overplotting, which results in a cluttered view on the data. Therefore, discerning the underlying data information and selecting interesting patterns can become difficult. We proposed a new technique, the Orientation-enhanced Parallel Coordinate Plots, to improve the view and discernibility of patterns in otherwise cluttered PCPs. We achieved our goal by visually enhancing parts of each PCP line with respect to its slope, hereby incorporating information from 2D scatterplots in the representation [18]. Compared to the state-of-the-art, our approach is simple and provides better discernibility of data patterns, especially when there are multiple overlapping patterns or when there are outliers and structures, obstructed by noise. We evaluated our approach with several synthetic and real-world datasets. One of the main advantages of OPCs is that thy allow a new and versatile selection method, the Orientation-enhanced Brushing. Brushing in the OPCPs space enables an efficient selection of individual data structures involving a reduced user interaction when compared to the state-of-the-art selection tools in PCPs. On the other hand, OPCPs require more training, compared to PCPs.

A direction for future work includes employing color transfer functions in the OPCPs for better discrimination of the different data patterns, or even clustering. Moreover, it would be interesting to extend the evaluation of our proposed visual enhancement, but also of the related brushing method, to cover the points discussed in Sec. 6. Finally, the extension of the O-Prober to other shapes should allow easier, faster, and more interactive selections of data patterns.

8 Acknowledgements

This work was partially supported by the FP7 European Project DR THERAPAT, the FP7 European Project Harvest4D, and the Intel VCI at Saarland University. The authors would like to thank the reviewers for their helpful feedback.

References

[1] Institute for Information Systems at the TU Braunschweig (Germany).

[2] XmdvTool Datasets: http://davis.wpi.edu/xmdv/datasets.html.

[3] D. F. Andrews. Plots of high-dimensional data. *Biometrics*, pages 125–136, 1972.

[4] M. Ankerst, S. Berchtold, and D. A. Keim. Similarity clustering of dimensions for an enhanced visualization of multidimensional data. In *Information Visualization, 1998. Proceedings. IEEE Symposium on*, pages 52–60. IEEE, 1998.

[5] A. O. Artero. Uncovering clusters in crowded parallel coordinates visualizations. In *In Proceedings of the IEEE Symposium on Information Visualization 2004*, pages 81–88. IEEE Computer Society, 2004.

[6] E. Bertini, L. Dell'Aquila, and G. Santucci. Springview: Cooperation of radviz and parallel coordinates for view optimization and clutter reduction. In *Coordinated and Multiple Views in Exploratory Visualization, 2005.(CMV 2005). Proceedings. Third International Conference on*, pages 22–29. IEEE, 2005.

[7] S. Bremm, T. Von Landesberger, M. Heß, and D. Fellner. PCDC-On the highway to data-a tool for the fast generation of large synthetic data sets. In *EuroVis Workshop on Visual Analytics*, pages 7–11, 2012.

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

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

[10] E. Fanea, S. Carpendale, and T. Isenberg. An interactive 3D integration of parallel coordinates and star glyphs. In *Information Visualization, 2005. INFOVIS 2005. IEEE Symposium on*, pages 149–156. IEEE, 2005.

[11] Y.-H. Fua, M. O. Ward, and E. A. Rundensteiner. Hierarchical parallel coordinates for exploration of large datasets. In *Proceedings of the conference on Visualization'99: celebrating ten years*, pages 43–50. IEEE Computer Society Press, 1999.

[12] Z. Geng, Z. Peng, R. S. Laramee, J. C. Roberts, and R. Walker. Angular histograms: Frequency-based visualizations for large, high dimensional data. *Visualization and Computer Graphics, IEEE Transactions on*, 17(12):2572–2580, 2011.

[13] R. C. Gonzalez and R. E. Woods. *Digital Image Processing (2nd Edition)*. Prentice-Hall, Inc., Upper Saddle River, NJ, USA, 2002.

[14] M. Graham and J. Kennedy. Using curves to enhance parallel coordinate visualisations. In *Information Visualization, 2003. IV 2003. Proceedings. Seventh International Conference on*, pages 10–16. IEEE, 2003.

[15] L. Harrison, F. Yang, S. Franconeri, and R. Chang. Ranking visualizations of correlation using weber's law. *Visualization and Computer Graphics, IEEE Transactions on*, 20(12):1943–1952, 2014.

[16] H. Hauser, F. Ledermann, and H. Doleisch. Angular brushing of extended parallel coordinates. In *Information Visualization, 2002. INFOVIS 2002. IEEE Symposium on*, pages 127–130. IEEE, 2002.

[17] J. Heinrich and D. Weiskopf. State of the Art of Parallel Coordinates. 2012.

[18] D. Holten and J. J. Van Wijk. Evaluation of cluster identification performance for different pcp variants. In *Computer Graphics Forum*, volume 29, pages 793–802. Wiley Online Library, 2010.

[19] C. Hurley and R. Oldford. Pairwise display of high-dimensional information via eulerian tours and hamiltonian decompositions. *Journal of Computational and Graphical Statistics*, 19(4), 2010.

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

[21] A. Inselberg. *Parallel Coordinates: Visual Multidimensional Geometry and Its Applications*. Springer-Verlag New York, 2009.

[22] J. Johansson, P. Ljung, M. Jern, and M. Cooper. Revealing structure within clustered parallel coordinates displays. In *Information Visualization, 2005. INFOVIS 2005. IEEE Symposium on*, pages 125–132. IEEE, 2005.

[23] S. Johansson and J. Johansson. Interactive dimensionality reduction through user-defined combinations of quality metrics. *Visualization and Computer Graphics, IEEE Transactions on*, 15(6):993–1000, 2009.

[24] H. Lam, E. Bertini, P. Isenberg, C. Plaisant, and S. Carpendale. Empirical studies in information visualization: Seven scenarios. *Visualization and Computer Graphics, IEEE Transactions on*, 18(9):1520–1536, 2012.

[25] M. Lichman. UCI machine learning repository, 2013.

[26] A. R. Martin and M. O. Ward. High dimensional brushing for interactive exploration of multivariate data. In *Proceedings of the 6th Conference on Visualization'95*, page 271. IEEE Computer Society, 1995.

[27] K. T. McDonnell and K. Mueller. Illustrative parallel coordinates. In *Computer Graphics Forum*, volume 27, pages 1031–1038. Wiley Online Library, 2008.

[28] P. Muigg, J. Kehrer, S. Oeltze, H. Piringer, H. Doleisch, B. Preim, and H. Hauser. A Four-level Focus+ Context Approach to Interactive Visual Analysis of Temporal Features in Large Scientific Data. In *Computer Graphics Forum*, volume 27, pages 775–782. Wiley Online Library, 2008.

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

[30] W. Peng, M. O. Ward, and E. A. Rundensteiner. Clutter reduction in multi-dimensional data visualization using dimension reordering. In *Information Visualization, 2004. INFOVIS 2004. IEEE Symposium on*, pages 89–96. IEEE, 2004.

[31] T. Porter and T. Duff. Compositing digital images. In *ACM Siggraph Computer Graphics*, volume 18, pages 253–259. ACM, 1984.

[32] P. L. Rheingans. Task-based color scale design. In *28th AIPR Workshop: 3D Visualization for Data Exploration and Decision Making*, pages 35–43. International Society for Optics and Photonics, 2000.

[33] H. Siirtola. Direct manipulation of parallel coordinates. In *Information Visualization, 2000. Proceedings. IEEE International Conference on*, pages 373–378. IEEE, 2000.

[34] H. Siirtola and K.-J. Räihä. Interacting with parallel coordinates. *Interacting with Computers*, 18(6):1278–1309, 2006.

[35] H. Theisel. Higher Order Parallel Coordinates. In *VMV*, pages 415–420, 2000.

[36] A. Unwin, C. Volinsky, and S. Winkler. Parallel coordinates for exploratory modelling analysis. *Computational Statistics & Data Analysis*, 43(4):553–564, 2003.

[37] M. O. Ward. XmdvTool: Integrating Multiple Methods for Visualizing Multivariate Data. In *Proceedings of the Conference on Visualization '94*, VIS '94, pages 326–333, Los Alamitos, CA, USA, 1994. IEEE Computer Society Press.

[38] M. O. Ward. Creating and manipulating n-dimensional brushes. In *Proceedings of Joint Statistical Meeting*, pages 6–14. Citeseer, 1997.

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

[40] N. Wong, S. Carpendale, and S. Greenberg. Edgelens: An interactive method for managing edge congestion in graphs. In *Information Visualization, 2003. INFOVIS 2003. IEEE Symposium on*, pages 51–58. IEEE, 2003.

[41] P. C. Wong and R. D. Bergeron. Multiresolution multidimensional wavelet brushing. In *Visualization'96. Proceedings.*, pages 141–148. IEEE, 1996.

[42] J. Yang, W. Peng, M. O. Ward, and E. A. Rundensteiner. Interactive hierarchical dimension ordering, spacing and filtering for exploration of high dimensional datasets. In *Information Visualization, 2003. INFOVIS 2003. IEEE Symposium on*, pages 105–112. IEEE, 2003.

[43] X. Yuan, P. Guo, H. Xiao, H. Zhou, and H. Qu. Scattering points in parallel coordinates. *Visualization and Computer Graphics, IEEE Transactions on*, 15(6):1001–1008, 2009.

[44] H. Zhou, W. Cui, H. Qu, Y. Wu, X. Yuan, and W. Zhuo. Splatting the lines in parallel coordinates. In *Computer Graphics Forum*, volume 28, pages 759–766. Wiley Online Library, 2009.

[45] H. Zhou, X. Yuan, H. Qu, W. Cui, and B. Chen. Visual clustering in parallel coordinates. In *Computer Graphics Forum*, volume 27, pages 1047–1054. Wiley Online Library, 2008.