IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received 22 June 2023, accepted 5 August 2023, date of publication 23 August 2023, date of current version 26 October 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3307718

RESEARCH ARTICLE

Visually Enhanced Parallel Coordinates Plot With Two-Dimensional Kernel Density Scatter Plots

JIA LIU, GANG WAN, SHUAI WANG, PANKE DENG, ZHIJUAN SU, CHU LI, YAO MU, AND YUNXIA YIN
School of Space Information, Space Engineering University, Beijing 101416, China
Corresponding author: Gang Wan (casper_51@163.com)

ABSTRACT Parallel coordinates plots are popular tools for high-dimensional data visualization. To alleviate the difficulties caused by the inherent defects that arise when the dimensions are increased, this study attached two-dimensional kernel density scatter plots to parallel coordinates plot, which integrates the Cartesian and parallel coordinate systems, to combine their benefits and conveniently explore the relationships between paired attributes. The collaborative design combining the two visual types was realized through the corelative axis swap interaction method, which allows users to freely exchange parallel coordinate axes while also updating the associated kernel density scatter plots based on the Cartesian coordinate system. In this study, we obtained six cases of real datasets and performed a field trial study to assess the usefulness of our method in assisting users in evaluating correlations between paired attributes. The results show that our method is more useful for analyzing the relationships between paired attributes.

INDEX TERMS Parallel coordinates plot (PCP), scatter plot, two-dimensional kernel density contour plot, visual analysis, interactive analysis.

I. INTRODUCTION

For the visual interpretation of high-dimensional data, the Parallel coordinates plot (PCP) [1], [2] is a popular method of visual analysis. It has a highly valuable feature in that it can use multiple parallel coordinate axes to visually express multiple attributes and connect attribute values with polylines, rendering it possible to visualize high-dimensional data on a two-dimensional plane. The relationships between paired attributes are perceived through axis-to-axis-parallelism. Since the appearance of PCP, considerable development and positive practical outcomes have been observed in various scientific research fields, including computer science [3], health sciences [4], and ecology [5]. However, as the data records become more comprehensive and detailed, the dimensions continue to increase, leading PCP to plague scientific results by the curse of dimensionality [6], which aggravates polyline overlap and occlusion, increases the uncertainty of identification [7], and conceals the important visual trends in the data on the parallel coordinate axes. The discovery of the relationships between paired attributes is a challenging task. The commonly explored types of relationships are linear, nonlinear, and monotonic. Identifying these relationships can help researchers visualize meaningful trends in the data. Numerous approaches, such as various interaction analysis methods and methods for recording correlation-based parallel coordinate axes, have been proposed to alleviate these limitations, with scientists attempting a range of automated visual analysis methods covering all circumstances, from entirely automatic to entirely interactive [7].

Here, we propose a visually enhanced Parallel coordinates plot using the two-dimensional kernel density scatter plots method (KDEPCP) to explore the correlation between paired attributes in PCP. We present two-dimensional (2D) kernel density scatter plots below the paired axes of PCP. The exploration of correlations between the paired attributes is enhanced by integrating 2D kernel density contour plots and scatter plots using Cartesian coordinate system visualization to study the distribution of the data between the paired parallel coordinate axes in the PCP. Users can reorganize parallel

The associate editor coordinating the review of this manuscript and approving it for publication was Walter Didimo.

coordinate axes and update 2D kernel density scatter plots using the axis swap interaction method, even if the two attributes are far apart. The 2D kernel density scatter plot intuitively reveals the distribution and implicit correlations between and the paired attributes through the distribution of the 2D kernel density contour plots and scatter plots, where the points in the Cartesian coordinate system correspond on a one-to-one basis to the polylines in the parallel coordinate system. To comprehensively and quickly view the relationships between attributes, we also drew a scatterplot matrix (SPLOM) to assist in visual analysis, in which 2D kernel density scatter plots were employed for the lower triangle, scatter plots were used for the upper triangle, and histograms were used for the diagonal line. Based on our work, interaction approaches were designed to help analysts study the details of 2D kernel density scatter plots in a limited screen space.

We demonstrate the practicality of the proposed approach using five datasets from various fields, showing that KDEPCP outperforms in revealing hidden relationships between data attributes. Moreover, we show that KDEPCP can innately extend the histograms through kernel density curves, thereby continuously providing the distribution and development trend of each attribute. Additionally, to evaluate that our method usefulness and more efficient, we conducted a field trial study to assess how KDEPCP assists users in estimating the correlations between paired variables.

The following are the advantages of KDEPCP in particular and the contributions made by this study:

1. Enhance the visual analysis of the relationships between paired attributes in PCP: distribute a two-dimensional kernel density scatter plot along the bottom of the PCP, take advantage of the Cartesian coordinate system's universality and intuitiveness, improve the comprehension of PCP through contrast, and improve the function of exploring the relationship between paired attributes in PCP;

2. Linkage update Interaction technique: the 2D kernel density scatter plots are updated while the axis exchange interaction analysis technique of PCP is employed to combine two non-adjacent attributes for visual analysis;

3, Use of 2D kernel density scatter plots in scatterplot matrix to assist visual analysis: the lower triangle in scatterplot matrix is visualized using 2D kernel density scatter plots without wasting visual layout.

In Section II, we review related studies on PCP, particularly those that utilize scatter plots and histograms. In Section III, we explain the design of the KDEPCP approach and its principles. In Section IV, we explain the process of data exploration using KDEPCP for visual interaction analysis, using five publicly available high-dimensional datasets and detailing user interactions. Then, we conduct a field trial study to evaluate the practical and efficient of our method in Section V. In Section VI, we discuss KDEPCP's shortcomings and implications. Finally, we conclude our work and suggest potential future research directions of KDEPCP.

II. RELATED WORK

PCP provides a comparative view of polyline connections between attributes [8]. However, with the diversification of the recording methods for the datasets and refinement of the recorded content, the data dimensions are continuously increasing, and PCP faces the so-called curse of dimensionality.

The difficulty in exploring the relationships between attributes in PCP is a serious problem and an inherent defect. Two coordinate axes that are far away from each other hide much of their correlations. To solve this problem, many experts have proposed various visual interaction analysis technologies and methods for reordering the axes. For example, highlighting interaction technique highlights a polyline with the mouse hovering over it, surrounded by many other polylines, for a detailed observation. Brushing interaction technique [9] allows the user to filter data by selecting areas on the parallel coordinate axes. In axis swap technique [10], the order of the axes can be tentatively exchanged such that the implicit trends and correlations between different attributes can be obtained by subjective judgment. Even with such excellent visual interaction methods, the intrinsic limitations of PCP continue to present many challenges, often necessitating the sacrifice of the original structure to alleviate them, which significantly diminishes the perceived advantages of this method. A method proposed by Nohno et al. rearranges parallel coordinate axes through spectral analysis of graphs [11]. A parallel coordinate axis reordering approach was presented by Lu et al. to shift the order of the parallel coordinate axes while reducing the computing complexity from the perspective of dimensionality reduction [12]. A method for interactive spacing modification and dimension reordering founded on a parallel coordinate dimension hierarchy was introduced by Yang et al. [13]. Peng et al. introduced the concept of clutter-based dimension reordering. Its aim is to discover views with the lowest degree of visual clutter by altering the dimension order in the visualizations [14]. Koh et al. proposed a cooperative reordering method based on the optimization of the pairwise visual difference in PCP to facilitate their contrastive comparison [15]. The method of rearranging the coordinate axes can only depict the degree of connection between the order of the axes and their separation but not the precise distribution, and different axis-sorting algorithms have different outcomes.

By highlighting distinct parts of the data that are difficult to understand through polyline patterns alone, the integration of other visualizations with PCP can assist researchers in visually searching for information. The scatter plot, which depicts the correlations between two attributes, is a popular visualization technique employed for this purpose [16]. Finding data distributions and clusters is rendered easier by visual representations of the data through scatter plot points [17]. A scatter plot is based on the Cartesian coordinate system, which can reliably display datasets with two or three attributes. For visual tasks with higher-dimensional datasets, a SPLOM [18], [19], [19], [20] was used to describe the relationships between paired attributes for all attributes in the entire dataset. Qu et al. placed scatter plots between the paired axes above the PCP to illustrate the distribution of data between the paired axes and observed their possible correlations [21]. To enhance PCP, Janetzko et al. integrated three advanced statistical visual encoding methods, including violin charts and box plots, to represent the density distribution of each axis [22]. To enhance PCP, Bok et al. added color-stacked histograms to parallel coordinate axes [23], enabling users to visually inspect the relationships between attributes even when the attribute axes are separated by a large distance. An accessible approach that enables users to study clustering, linear correlations, and outliers in large datasets without running on the overdraw and clutter issues of the original PCP is a histogram attached to the PCP axis [24] that depicts both the density of the polylines and their slopes. In the geo-coordinated parallel coordinates (GCPC) [25] method, box plots are attached above the parallel axes to illustrate the relationships between the paired coordinate axes. Schmid et al. combined SPLOM, PCP, a permutation matrix, and Andrew's plot [26]. Wong et al. integrated the SPLOM with PCP [27].We discovered that scatter plots, histograms, box plots or violin plots can be attached above the PCP to obtain a coordinated-view visualization. These visual enhancement approaches, based on the frequency of the data and Cartesian coordinate system, can visually express the paired attribute relationships. Histograms, which are useful for providing a summary of the data, have been used in numerous previous studies [28]. This approach is straightforward and effectively displays the clusters and outliers. However, this histogram employs bin mapping. Parameter bins were chosen before the histogram was drawn. The final histogram is substantially different if the bins are different. In addition, the distribution of the histogram is not smooth, meaning that samples in the same bin have the same probability density. Of course, this is frequently inappropriate for the types of analysis in which these methods are applied.

III. DESIGN OF KDEPCP

This section discusses the means by which we implemented the PCP with 2D kernel density scatter plots, with an emphasis on interactive features and visual enhancements.

A. DESIGN RATIONALE

A polyline in a parallel coordinate system represents a point in the Cartesian coordinate system. PCP enables users to visually explore high-dimensional datasets on a 2D plane by connecting the values of each attribute to each data point. However, similar to many other visual encoding methods, PCP is difficult for inexperienced users to interpret, and users must learn its principles. Moreover, to a certain extent, the intrinsic flaws of PCP, such as the order of axis arrangement, overlapping of polylines, and difficulty in pattern recognition between paired attributes, restrict its development.

KDEPCP is designed to estimate the relationships between paired coordinate axes while maintaining the original advantages of PCP. In our design, an SPLOM is also attached to assist in the expression of the overall structure and comprehensively estimate the relationships between various dimensions.

To achieve our design goals and alleviate the curse of dimensionality and difficulties in identifying the attribute relationships between paired coordinate axes in the PCP, as well as the challenges in interpreting the polylines while maintaining advantages of PCP, we attached scatter plots between every pair of coordinates below the PCP and then superimposed 2D kernel density contour plots onto the scatter plots to enhance the density perception. The 2D kernel density contour plot displays the distribution frequency of the data scatter points in a continuous manner, without requiring previous subjective knowledge, by fitting the distribution according to the characteristics of the data. This is better than parameter estimation and can easily detect outliers. To achieve our goal of maintaining the inherent benefits of PCP, our visualization design was efficient and did not require any changes to the original PCP design. The 2D kernel density scatter plots in KDEPCP have superior interpretability and capacity for intuitive comprehension compared to PCP because they employ the most popular and straightforward Cartesian coordinate system on the 2D plane. The initial unintelligible polyline is rendered understandable, since the point and the polyline have a one-to-one correspondence as they are attached below the PCP and updated collectively with the coaxial exchange interaction. Polylines, scatter plots, and kernel density contour plots can be used to visualize the relationships between attributes.

To visualize high-dimensional data, we designed a visual enhancement approach that combines PCP with 2D kernel density scatter plots and follows the visual design principles proposed by Shneiderman [29]. We also designed correlative axis-swap interaction, brushing interaction, and statistical coloring interaction methods for scaling and filtering data. The scatter plots and 2D kernel density contour plots in the 2D Cartesian coordinate system are effective in showing the probability density distributions of the datasets and the connections between paired characteristics, but they cannot effectively help users estimate the relationships between non-adjacent paired attributes in PCP. However, the distribution of each attribute was not reflected. Therefore, we combined the overview of KDEPCP with SPLOM for comprehensive assistance, making them complementary. The diagonal line shows histograms displaying the probability distribution of a single attribute, and the distribution was continually fitted using one-dimensional kernel density curves. The lower triangle shows the 2D kernel density scatter plots and the upper triangle shows scatter plots. Thus, the display area of the screen is not wasted.

B. BUILDING 2D KERNEL DENSITY SCATTER PLOTS

Kernel density estimation [30], [31] is a nonparametric estimation method that differs from commonly used parameter estimation approaches, such as likelihood estimation, in that

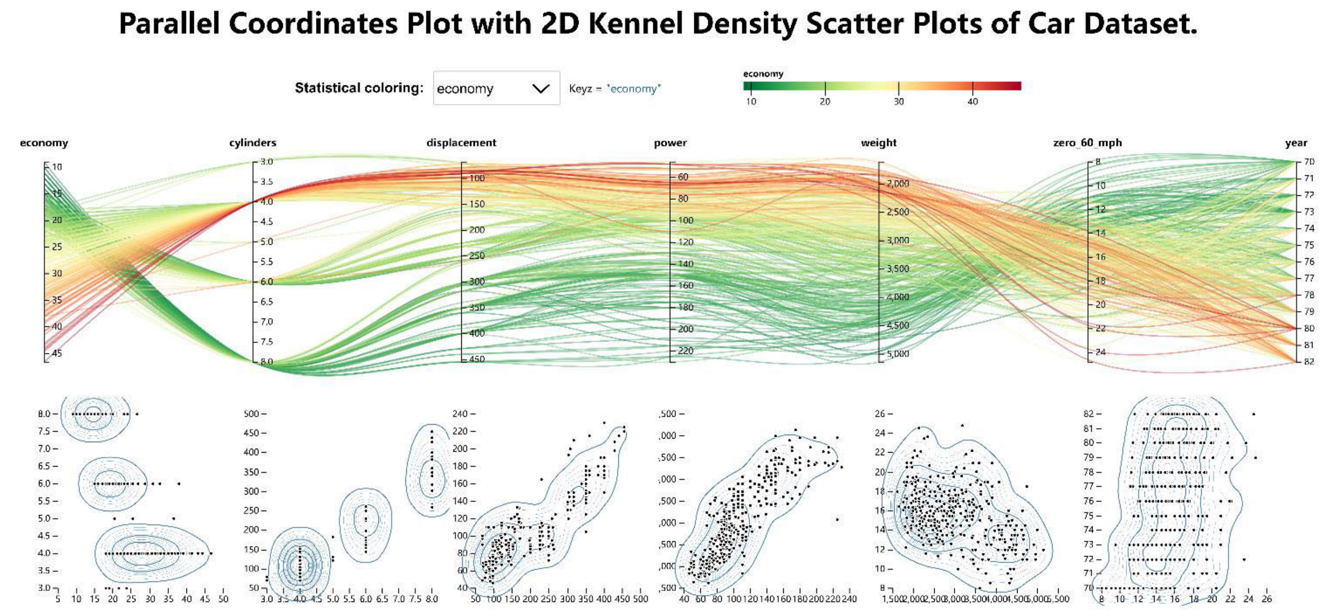

FIGURE 1. Car performance dataset visualization with KDEPCP.

a better model can be obtained by directly using samples to fit the probability distribution of the dataset in accordance with the characteristics and properties of the data, even when one is prior knowledge and the form of the density function of probabilities is unknown. The continuous kernel density curve can be used to represent the density and distribution of the data and alleviate the shortcomings of the uneven distribution of the histograms. Kernel density estimation has made many contributions to visualization [32], [33], [34], particularly as a natural extension of the histogram, has shown promising results in real-world applications [35], [36]. Here, we used 2D kernel density contour plots, which inspired our previous work [37], to enhance the scatter plots. This allowed us to analyze and explore the relationships between paired property associations by perceiving the contour distribution and trend of the curve, as well as the density of the scatter points.

To construct the 2D kernel density scatter plots in KDEPCP, we first used the scatter plots presented in the Cartesian coordinate system, and then superimposed the 2D kernel density contour plots onto the scatter plots. In this case, the visualization graph represents the distribution of the paired attributes and the frequency of the distribution. Owing to the lifting effect of the contour lines, users can investigate the implicit connections between the paired features. Moreover, the visualization based on the Cartesian coordinate system is more precise and effective than that based on the parallel coordinate system, which is also an innate advantage of our KDEPCP design. FIGURE 1 illustrates the method for adding 2D kernel density scatter plots for every two parallel coordinate axes along the PCP.

C. INTERACTION CAPABILITIES

We also designed interactive technologies for KDEPCP to enhance visual analysis. First, we applied the correlative axis swap interactive analysis method, which is a crucial interaction function because both the quantity and quality of the presentation are significantly affected by the order of the axes [38]. Moreover, it is difficult to compare data between nonadjacent dimensions, which may introduce ambiguity [25]. Therefore, KDEPCP includes interactive methods that allow the manual reordering of axes [10]. Users can reorder the axes by simply dragging and dropping operations onto the labels of each attribute's coordinate axis, placing the attributes to be compared together, and reconfiguring the data accordingly. The 2D kernel density scatter plots attached below the PCP were updated with each axis swap operation. It should be noted that the order of the parallel coordinate axes was chosen to highlight important pairwise correlations.

Second, we applied statistical coloring interactive technique to PCP [22], in which the use of color is an important visual channel used to express attribute values. All attributes are included in the drop-down list, and the corresponding color legend is shown above. The color values in the legend are continuous. In this study, we used the color-matching scheme of the D3.js framework. These color schemes are visually appealing and adhere to the accepted range for human color discrimination. Statistical coloring was used to color the polylines according to the attribute value of the data on an axis to reveal the correlations.

Finally, brushing interaction technique is employed in the process of visual information exploration to filter out items that users do not consider important to realize the expansion and contraction of the data items and render the visualization clearer. Brushing interaction technique has been applied to PCP and SPLOM.

IV. CASE STUDY

To demonstrate how useful the KDEPCP is in helping explore relationships between paired attributes, we selected five publicly available datasets from different domains for visual analysis. The first dataset was the car performance dataset, which has been used in many publications and contains 398 data items and eight attributes. It records various performances for different types of cars. The second dataset was the shared bicycle daily dataset, which had 731 records and 10 attributes, including time, weather, and wind speed. The third dataset is the wine quality dataset, which contains 1600 data items and 12 attributes, including wine density, pH value, and citric acid content. The fourth dataset is the Boston housing price dataset, which contains 506 data items and 12 dimensions, recording the relevant factors that affect housing prices. The fifth dataset was a dataset for diabetes. It includes data based on the physiological characteristics of 442 patients and the progression of the illness after 1 year. The dataset has 10 characteristics.

Our software was implemented using the D3.js visualization library [39]. D3.js uses a scalable vector graphics (SVG) file format to generate high-resolution web graphics pages with a high degree of flexibility, which enables developers to design various types of visualization methods and visual interactions.

In KDEPCP, the density is calculated using the d3.contourDensity() function, and the bandwidth() function is employed to determine the appropriate resolution. In this study, it was important to set the bandwidth to 13 and threshold to 30. In the SPLOM, we employed the Epanechinikov function as the kernel function in kernel density estimation, with a bandwidth of 3. And all visualization demos of KDEPCP are available at http://18.223.136.39:8080/kdepcp/index.html.

A. CAR DATASET

We visualized seven attributes from the car performance dataset, which also served as an example dataset for D3.js, as illustrated in FIGURE 1. According to PCP, early cars tend to be less economical with higher fuel consumption. Although this characteristic may have a strong correlation with the number of cylinders, displacement, power, and weight, it leads to rapid acceleration. From this, we again concluded that PCP is an excellent high-dimensional data visualization technique, providing users with a good overview of the data, and the statistical coloring technique is particularly beneficial to users and improves the effectiveness of the analysis.

We discovered an intriguing fact by observing the 2D kernel density scatter plots attached below the PCP: cars with the same number of cylinders had different amounts of fuel consumption and different displacements. The displacement, power, and weight are approximately proportional to each other and inversely proportional to the time required to accelerate from 0 km/h to 60 km/h. In other words, a higher power and faster acceleration are associated with a larger displacement. The labels above the coordinate axes could simply be dragged to render the two axes adjacent if the user wished to examine the relationship between the two nonadjacent coordinate axes. The year attributes also show that the quality of automotive manufacturing is continuously increasing.

Each of the enlarged images in FIGURE 2 is two-dimensional. With the use of statistical coloring in PCP, we observed that (a) the weight and fuel consumption have an inverse correlation, and (b) the number of cylinders and acceleration also have an inverse correlation. By observing the corresponding 2D kernel density scatter plot, consistent conclusions were obtained. With the capacity for interactive exploration and the integrated visualization environment provided by KDEPCP, the task of exploring correlations between attributes has become easier.

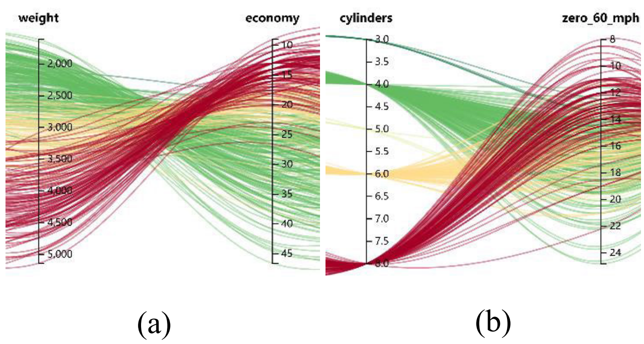

FIGURE 2. Details of car performance dataset visualization with KDEPCP.

B. BIKE SHARING DATASET

There are more attributes in the shared bicycle dataset, which contains data collected daily for 2 years. Nine numerical attributes were visualized, as shown in FIGURE 3. In PCP, brushing interaction technique can be used to view the bicycle usage in each season. We found that more bicycles were used in the summer and autumn than in the winter and spring. The reason for this may be related to temperature. People are more willing to travel bikes during relatively warm weather. We also observed that more people choose to ride bikes when the humidity was appropriate and the wind was not too strong. However, under truly severe weather conditions, almost no one wants to ride bikes. Using 2D kernel density scatter plots, we observed the situation in detail and found several outliers. Occasionally, some people choose to ride under windy and dry weather conditions.

C. RED WINE QUALITY DATASET

In the red wine quality dataset, the attribute dimensions were 12. The entire display could be observed without scrolling through the webpage, because we used a widescreen

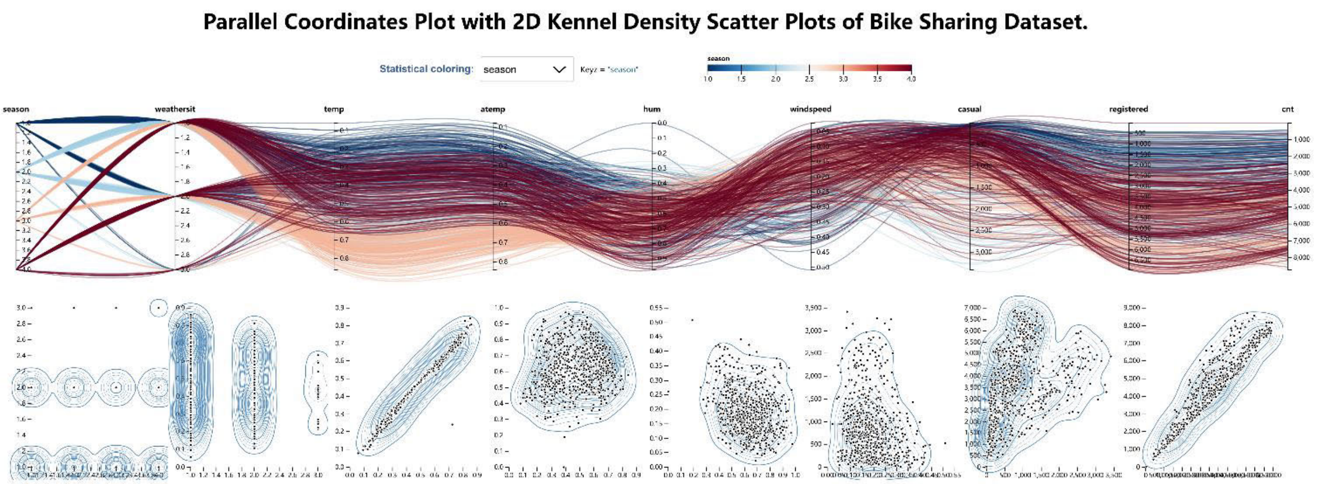

FIGURE 3. Bike sharing dataset visualization with KDEPCP.

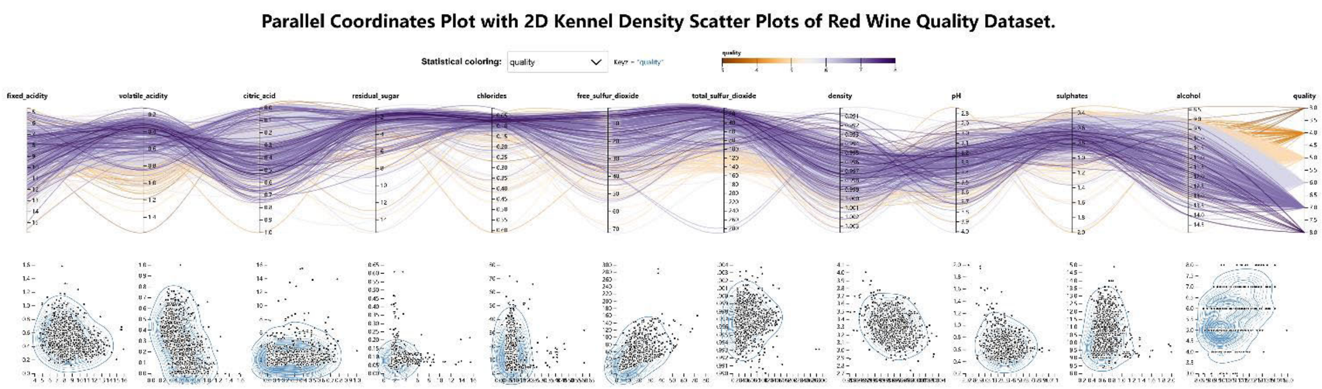

FIGURE 4. Red wine quality dataset visualization with KDEPC.

monitor. First, statistical coloring was adjusted to sort and colorize the data according to the attributes of quality, as shown in FIGURE 4. Although there were no discernible differences in quality between the classes, we found that lower-quality wines had greater levels of volatile acidity, chloride, and total sulfur dioxide.

By switching the quality attribute axes individually, we can compare and analyze the correlations between the quality attribute and other attributes, which is our area of interest in this study. However, in our approach, the auxiliary SPLOM, as shown in FIGURE 5, can be used to make general observations. The last row shows 2D kernel density scatter plots of red wine quality and other attributes, whereas the last column shows scatter plots of red wine quality and other attributes. Here, SPLOM helps visualize the desired features in the most efficient manner.

D. HOUSE PRICE IN BOSTON DATASET

Boston housing price dataset was used as the object of analysis to identify the main factors affecting housing prices and has the largest dimensionality of all datasets utilized in this study, with a total of 14 dimensions. As long as the screen is sufficiently large, we believe that KDEPCP can still be used to perform effective visual analysis even with greater dimensions. Because the factors influencing the prices of homes are those that we were most interested in studying, we chose the parallel coordinate axis, which represents the price attribute as the statistical coloring object, as shown in FIGURE 6. Housing prices were evidently high in places with low per capita crime rates and high concentrations of lower economic classes. These two elements can be seen to have significant impacts on housing prices. The 2D kernel density scatter plots, which demonstrate that these two parameters are inversely related to house value, provide more support this conclusion.

E. DIABETES DATASET

The diabetes dataset included 11 dimensions and was subjected to mean centralization to facilitate quantification, where the target was the progression of diabetes. We adjusted

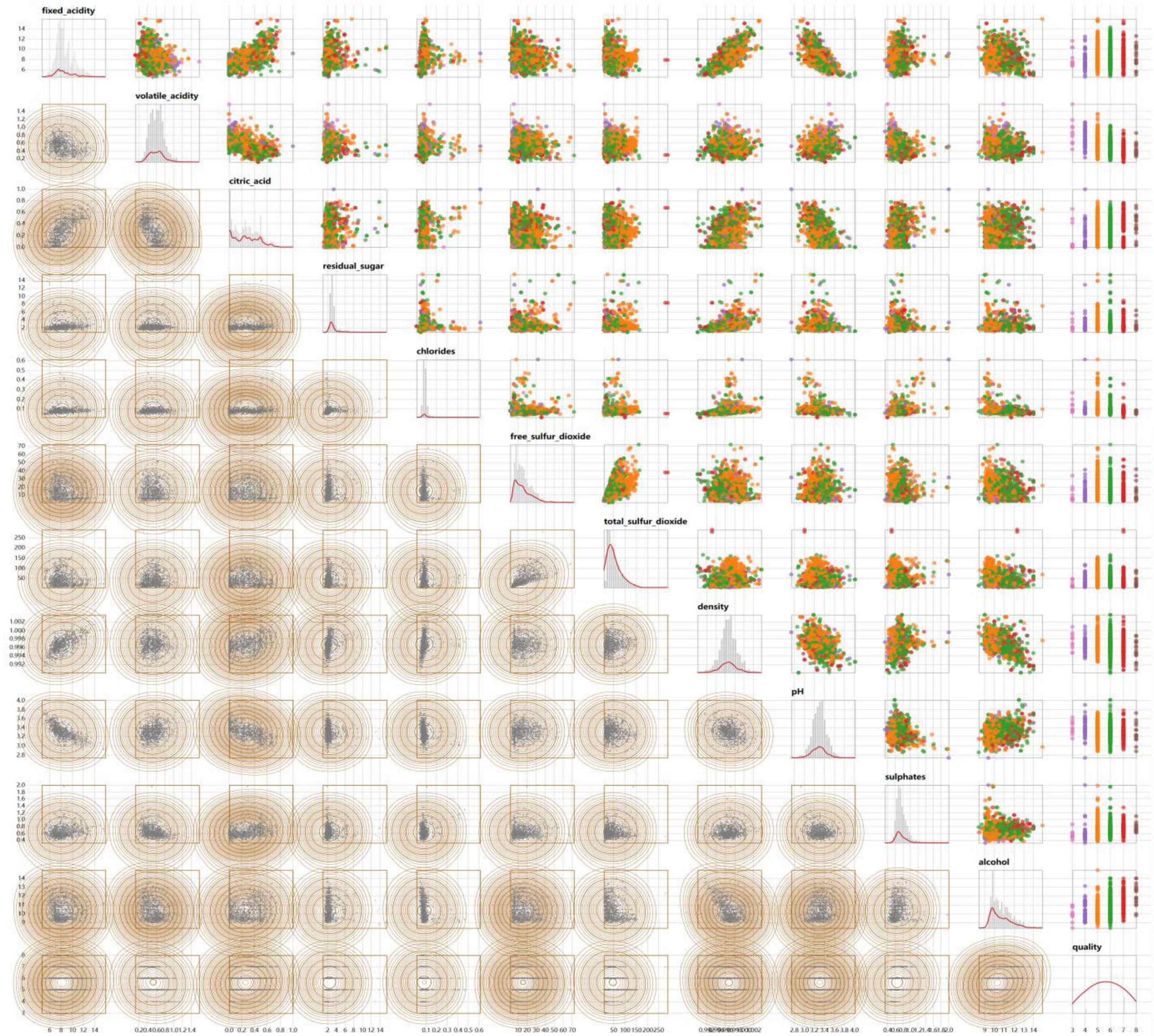

FIGURE 5. Red wine quality dataset visualization with scatterplot matrix.

the statistical coloring applied to this axis, as shown in FIGURE 7. The 2D kernel density contour plots outline of the data distribution, allowing users to quickly determine the distribution profile and correlation by observing the shape. KDEPCP users can quickly detect factors, such as those that are more closely associated with the course of diabetes, allowing for the possibility of performing early diabetes intervention. This would take longer if conventional PCP identification were to be used alone.

V. EVALUATION

To evaluate the overall performance of KDEPXP in the correlation analysis between paired attributes, we conducted a field trial study. The field trial study is open-ended analysis that give analysts the freedom to investigate and analyze as part of a typical data analysis activity.

We recruited 92 participants from the faculty members of a university, including 51 males and 41 females aged 17–55 years. Of these, 40 were undergraduates, 18 were master's students, 11 were PHD students, 3 were professors, 4 were associate professors, 5 were lecturers, and 11 were civilian staff. These participants were unfamiliar with the visualization methods but were proficient in using computers and web browsing. To show the visualizations throughout the experiment, we used 34-inch Lenovo widescreen monitors with a resolution of 2560*1440.

Before the task started, all participants were trained to use KDEPCP for visual analysis. We provided participants

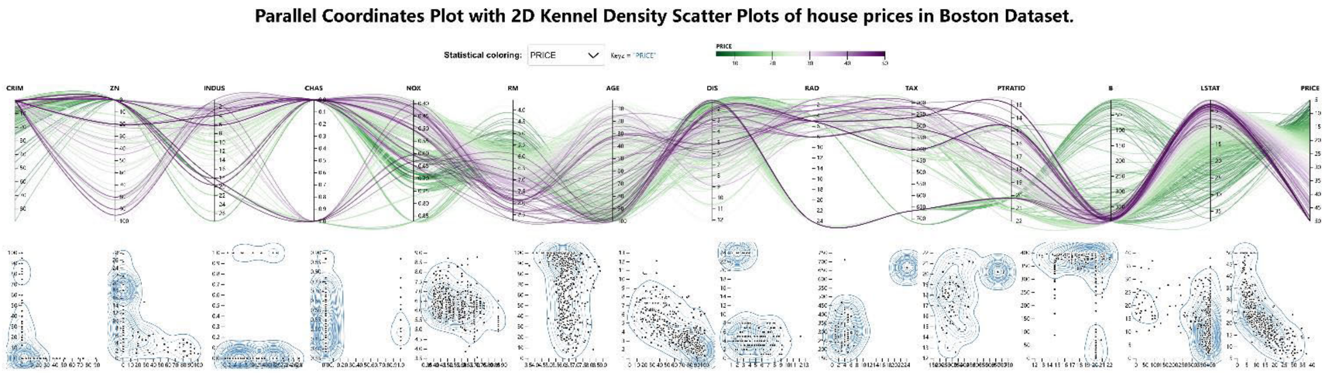

FIGURE 6. Housing prices in Boston dataset visualization with KDEPCP.

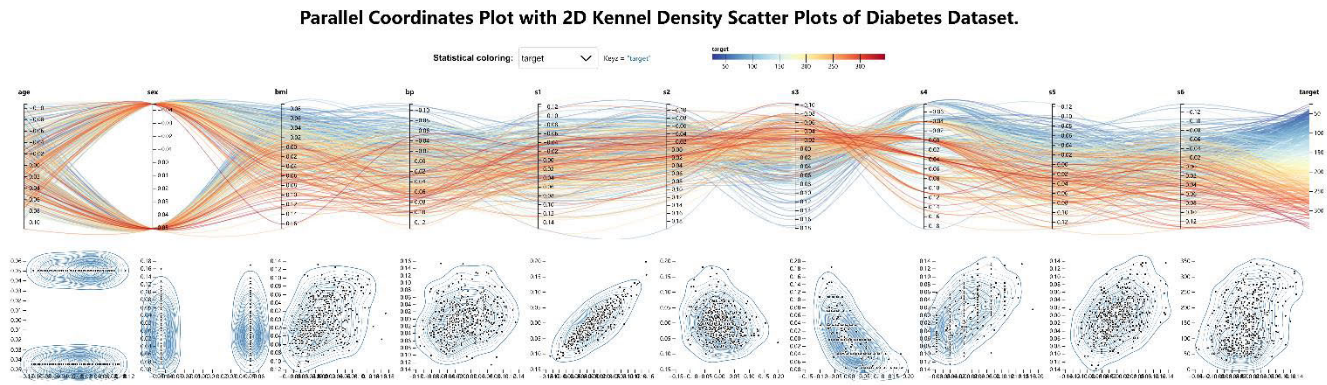

FIGURE 7. Diabetes dataset visualization with KDEPCP.

with instructions on how to visually analyze the relationships between paired attributes and evaluate the results. The training goal was to familiarize users with visual analysis and enable them to understand the subsequent analysis tasks.

A. STUDY DESIGN

The goal of the field trial study is to offer empirical proof of the performance and usefulness of KDEPCP method. We chose another two visual analysis methods for comparison with KDEPCP: PCP and SPLOM. PCP was chosen as the control condition to demonstrate the degree to which our approach improved. SPLOM was chosen because it is commonly used and is considered the best way to visually analyze the relationships between two attributes and is more accurate and efficient than analyzing the relationships between paired attributes through PCP [40]. We did not use histograms for comparison because they did not show any relationships between attributes. Moreover, the kernel density plot is an improvement over the histogram.

We conducted the content of the comparison open-ended analysis, including four instructions: (1) without any interactions, analyze the factors influencing iris species; (2) exchange the order of the parallel axes to analyze the factors affecting iris species again, where the SPLOM showed no such interactions; (3) use the brushing interaction to analyze the factors affecting iris species (the SPLOM also included brushing interaction and certain points could be selected for analysis); and (4) perform free analysis, using the original conditions and interactions of each visualization method in order to analyze the factors that determine iris species. With these guidance, we can more thoroughly evaluate the effects of interactive methods. Then the participants' reaction times and errors were recorded.

After the comparison open-ended analysis, the participants were invited to participate in our interviews, aiming to further investigate the usability of KDEPCP and obtain their responses to the following questions: (1) Which of the three visualization methods do you think is the most useful? (2) Are the interactions created by KDEPCP helpful for increasing the effectiveness of visual analysis? (3) What problems did you face when using the KDEPCP? (4) What functions or interactive technologies do KDEPCP currently not offer? Do you want to be available to implement?

B. SETTINGS

The field trial study was conducted with the publicly accessible iris dataset, which has 150 entries and five variables, including the categorization and the lengths and widths of the iris calyx and petals. This dataset was chosen because it is often used for data mining, has a small number of dimensions, and allows one to make clear comparisons.

VI. RESULTS

A. RESPONSE TIME

FIGURE 8 illustrates the total response time required for the participants, using the three visual analysis methods, to complete the analysis assignment, and it reveals that KDEPCP required the least amount of response time. This indicates that the participants spent less time analyzing the relationships between attributes and drawing conclusions, confirming the high efficiency of KDEPCP. The effect of the SPLOM is also evident. We speculate that the variations in the response time were due to participants' greater familiarity with the SPLOM constructed in the Cartesian coordinate system. As the dimensions increase, the SPLOM become less effective because they are visually displayed as matrix, which causes the webpage to become congested and chaotic, thereby reducing the effectiveness of the visual analysis.

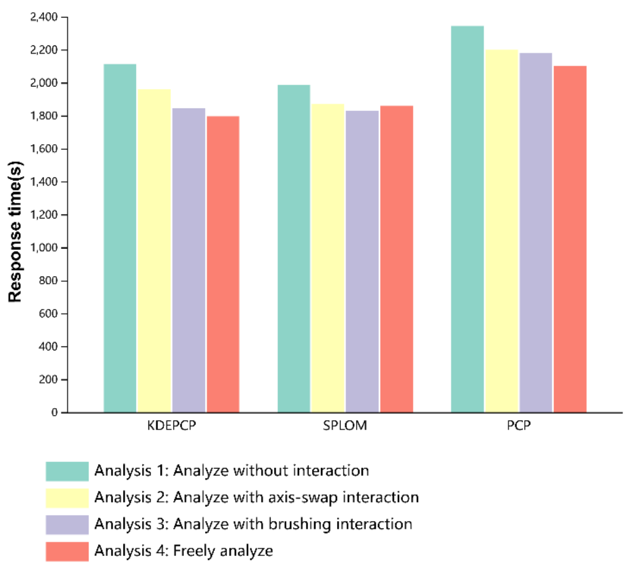

FIGURE 8. Results regarding the response time of the field trial study.

B. ERROR RATE

FIGURE 9 demonstrates that the error rate of KDEPCP is substantially lower than that of the other methods. Most of the time, there was no trade-off between the response time and error rate; hence, better performance did not lead to increased error rates. The error rate also decreased as further analysis was performed, proving that an effective interactive analysis design can enhance visual encoding performance and improve visual analysis accuracy.

C. ANALYSIS OF VARIANCE

Analysis of variance (ANOVA) is a useful technique for determining whether and to what extent each element has an impact on the given outcomes [41], [42], [43], [44], [45]. Using the social science statistical online tool SPSSPRO (https://www.spsspro.com/), we first conducted a two-way analysis of variance (two-way ANOVA) of the response time of each participant in the four phases of the visual analysis assignment using three visual analysis methods. We put the following two hypotheses into the test, assuming that there are three levels of factors affecting the visual analysis method and four levels of factors affecting the task stage:

$$\begin{cases} H_{01}{:}\alpha_1 = \alpha_2 = \alpha_3 = 0, \\ H_{11}{:}\alpha_1, \alpha_2, \alpha_3 \textit{ are not all zero.} \end{cases} \tag{1}$$

$$\begin{cases} H_{02}{:}\beta_1 = \beta_2 = \beta_3 = \beta_4 = 0, \\ H_{12}{:}\beta_1, \beta_2, \beta_3, \beta_4 \textit{ are not all zero.} \end{cases} \tag{2}$$

where α represents the visual analysis method factor, and β represents the visual analysis task stage factor. The results are shown in TABLE 1.

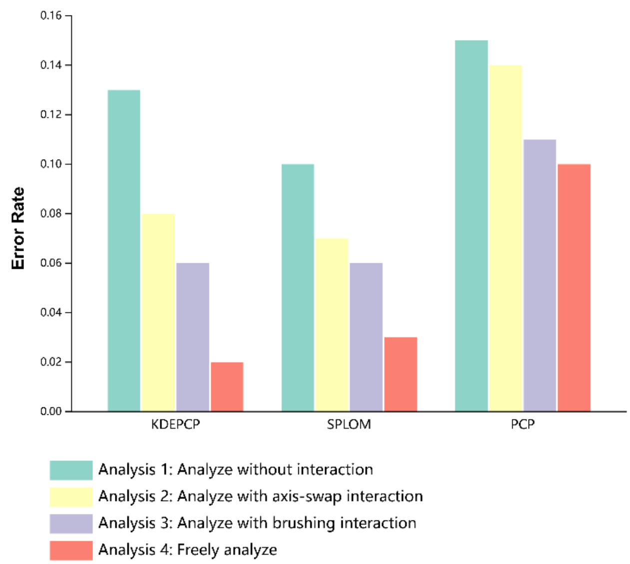

FIGURE 9. Results regarding the error rate of the field trial study.

TABLE 1. Results of two-way ANOVA for response time of field trial study.

Item	Sum Of Squares	Degree Of Freedom	Mean Square	F	P
Intercept	526981.363	1	526981.363	22184.67	0.000***
Method	2616.739	2	1308.37	55.079	0.000***
stage	1012.494	3	337.498	14.208	0.000***
Error	26082.224	1098	23.754		NaN

Hypotheses H_{01} and H_{02} are rejected because the time required to analyze high-dimensional data using the different visual analysis methods differs considerably, as shown by the values of $F_{0.05}\,(2, 1098) = 19.5 < 55.079, F_{0.05}\,(3, 1098) = 8.53 < 14.208$. Additionally, the amount of time required for the visual analysis of high-dimensional data using various visual interaction analysis methods varies considerably. In other words, both interactive and visual analysis methods significantly affected the duration of the visual analysis process.

For the error rate, we also performed a two-way ANOVA, and the results are reported in TABLE 2. It is considered that there are considerable variances in the error rate produced by employing different visual analysis methods to examine high-dimensional data, as well as large differences in the error rate generated by different visual interaction analysis methods. In other words, the visual analysis error rate is significantly influenced by both the interactive technique and the visual analysis approach.

TABLE 2. Results of two-way ANOVA for the error rate of visual analysis task.

Item	Sum Of Squares	Degree Of Freedom	Mean Square	F	P
Intercept	8.452	1	8.452	3279.212	0.000***
Method	0.787	2	0.393	152.584	0.000***
stage	0.867	3	0.289	112.133	0.000***
Error	2.83	1098	0.003		NaN

D. RESULTS OF INTERVIEW

In the interviews, in addition to providing their opinions on, and details about their comprehension of, the KDEPCP, the participants also provided results for the comparison open-ended analysis, focusing primarily on the following two points: (1) KDEPCP performs better in analyzing the relationships between paired attributes than other models. Because users can select comparison factors, the interactive axis swap approach is particularly useful. (2) SPLOM is also quite practical, allowing one to fully understand the relationships between paired attributes even in the absence of interactions. However, the 14-dimensional Boston housing price dataset formed a very large matrix to be presented in full, because widescreen monitors were used. If an ordinary monitor is to be used, the need to constantly scroll through the webpage would weaken the advantages of the SPLOM.

We also collected many useful recommendations, mainly focusing on the following two points. (1) Despite being comparatively unrestricted, the attribute comparison lacks a scientific foundation. If some reliable quantification tools, such as correlation coefficients, could be combined to render the axis swap interaction more purposeful, instead of taking the form of a random comparison, the efficiency would be greatly enhanced, especially when the dimensions increase, because the increase in dimensions would render the paired attribute exploration process more time-consuming. Future improvements in KDEPCP will primarily focus on this direction. (2) Because a polyline in the parallel coordinate system is equivalent to a spot in the Cartesian coordinate system, if the filtered data are also highlighted in the 2D kernel density scatter plot while the data are filtered on the parallel coordinate axes, the visual analysis effect can be enhanced by sensing the implicit connections generated by all highlighted matches and analyzing the correlations between attributes.

In summary, it can be concluded that KDEPCP has a significant effect on the correlation analysis between attributes, is more widely applicable, and, in particular, outperforms other visualization methods in terms of visual enhancement and interactive features.

E. DISCUSSION

The intrinsic drawbacks of PCP include difficulty in comprehension, overlapping occlusion of lines, curse of dimensionality, and ambiguous relationships between non-adjacent variables, as can be observed in the plots in the previous section. Most non-experts would also require some training for the parallel coordinate system because they are more accustomed to the Cartesian coordinate system. As the dataset's number of records increase, it becomes impossible to avoid polyline overlap, even if the polylines are rendered translucent or partially reduced by brushing interactions. The relationships between paired attributes are unclear because the axes are far apart from each other. KDEPCP integrates 2D kernel density scatter plots in the Cartesian coordinate system and PCP in the parallel coordinate system, thus combining the advantages of these two types of visualizations and allowing for comparisons to be made between these two coordinate systems.

Similar to many other PCP-based visualizations, the pattern-recognition ability of the original PCP can be useful. Users can seamlessly use it with KDEPCP, which also supports interactive analysis methods, including brushing, axis swap, and statistical coloring interactions. As shown in FIGURE 10, for the iris dataset, without statistical coloring interaction, the original PCP cannot effectively observe the boundaries of the classification, and it is difficult to explore the relationships between paired attributes without axis swapping. Furthermore, the attached 2D kernel density scatter plots can improve the exploration efficiency of the original PCP and facilitate an understanding of the relationships between the paired attributes. As shown in FIGURE 10, the iris can be classified according to the length and width of the petals. In the corresponding figure of the Cartesian coordinate system, we can also see that the lengths and widths of the petals of different varieties of irises are diverse.

The histogram of each attribute is displayed on the diagonal line in the SPLOM, which provides visualization assistance in KDEPCP approach. To alleviate the drawbacks of the histogram's lack of smoothness, we combined the histogram with a one-dimensional kernel density curve and fitted the data by smoothing the peak function. In KDEPCP, 2D kernel density contour plots are combined with scatter plots, and continuous density curves are used to describe the distribution forms of the variables. Thus, the visualization has strong robustness and weak model dependence.

The scatter plot is a particularly effective visualization tool for representing the distribution of the two attributes.

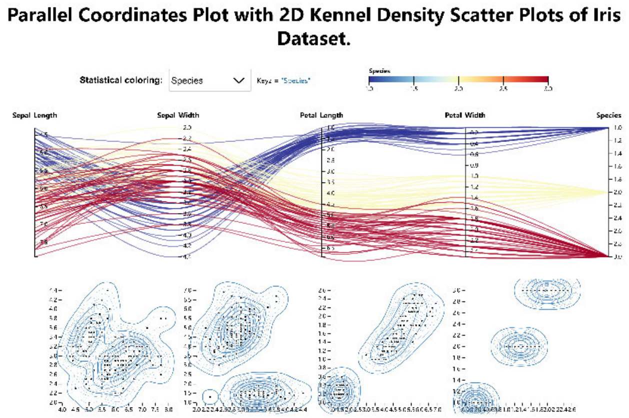

FIGURE 10. Iris dataset with KDEPCP.

However, when a dataset with n attributes is visualized, it needs to be presented in $n\times n$ scatter plots, and the space efficiency of this visual analysis method can also present problems, as shown in FIGURE 5. Each scatter plot shrinks and the SPLOM becomes more congested as the number of attributes increases, making it more difficult to determine how the attributes are related. Moreover, the matrix is diagonally symmetric, and because half of the space is wasted, we use 2D kernel density scatter plots to represent this portion of the matrix. However, when visualizing multiple attributes, the spatial efficiency of the visual analysis methods is critical because it directly affects the number of attributes that can be presented within the constrained screen area. KDEPCP can depict the correlations between numerous paired characteristics in a more space-efficient manner than SPLOM, which is advantageous for users who want to discover information on the relationships between several attributes.

The 2D kernel density scatter plots used in KDEPCP are based on the Cartesian coordinate system, which is in line with human comprehension tendencies. The learning process required to understand these plots is short or may even be unnecessary. This method can also be used to enhance the information search processes for other visuals. KDEPCP allows users to analyze data more efficiently by sorting the attributes according to their relevance or similarity.

VII. CONCLUSION

The KDEPCP visual analysis approach was presented in this work with the intent of overcoming the drawbacks of PCP in exploring nonadjacent attribute associations, improving the experience of investigating correlations between paired attributes, and resolving the dimensionality problem. While maintaining the perceptual benefits and features of PCP, kernel density contour plots and scatter plots, KDEPCP uses the 2D kernel density scatter plot visualization method based on a Cartesian coordinate system, integrates with a PCP, visually expresses the relationships between paired attributes on a joint basis, and visually mines implicit relationships and trends between attributes. With KDEPCP, users can employ axis swap interaction technique with their own subjective intuition to identify potential trends and implicit relationships in the data. In addition, the 2D kernel density contour plot has an effect on the detection of outliers, and we believe that a more comprehensive analysis and application of this problem could be a fascinating subject for the future. KDEPCP also provides several interactions, including axis swapping, brushing, and statistical coloring, to help users quickly and efficiently study the details of kernel density scatter plots within a limited screen space. In our case studies, we demonstrated the use of KDEPCP on real datasets. We evaluated KDEPCP's performance on tasks involving visual data analysis. The results indicate that, even with high dimensions, the KDEPCP method has excellent correlation estimates and strong applicability, enabling it to meet the basic data analysis requirements of various fields.

When visualizing datasets with higher dimensions, the KDEPCP process for analyzing the relationships between paired attributes is limited. If a high-dimensional dataset contains n dimensions, there will be $(n-1)!$ possible combinations of paired attributes. Consequently, the user's execution will be limited, and the study of the paired attributes will result in blindness. Therefore, the user will no longer be able to explore the relationships between paired attributes by continuously and actively switching the axes. effectiveness of the analysis, we must utilize axis-reordering methods or correlation coefficient calculations. Future research efforts will follow this direction.

REFERENCES

[1] A. Inselberg, "The plane with parallel coordinates," *Vis. Comput.*, vol. 1, no. 2, pp. 69–91, Aug. 1985.

[2] E. J. Wegman, "Hyperdimensional data analysis using parallel coordinates," *J. Amer. Stat. Assoc.*, vol. 85, no. 411, pp. 664–675, Sep. 1990.

[3] H. Choi, H. Lee, and H. Kim, "Fast detection and visualization of network attacks on parallel coordinates," *Comput. Secur.*, vol. 28, no. 5, pp. 276–288, Jul. 2009.

[4] A. A. Mitku, T. Zewotir, D. North, and R. N. Naidoo, "Exploratory data analysis of adverse birth outcomes and exposure to oxides of nitrogen using interactive parallel coordinates plot technique," *Sci. Rep.*, vol. 10, no. 1, pp. 1–9, Apr. 2020.

[5] O. Alminagorta, C. J. G. Loewen, D. T. de Kerckhove, D. A. Jackson, and C. Chu, "Exploratory analysis of multivariate data: Applications of parallel coordinates in ecology," *Ecological Informat.*, vol. 64, Sep. 2021, Art. no. 101361.

[6] D. L. Donoho, "High-dimensional data analysis: The curses and blessings of dimensionality," *AMS Math. Challenges Lect.*, vol. 1, p. 32, Aug. 2000.

[7] A. Dasgupta, M. Chen, and R. Kosara, "Conceptualizing visual uncertainty in parallel coordinates," *Comput. Graph. Forum*, vol. 31, no. 3pt2, pp. 1015–1024, Jun. 2012.

[8] H. Nguyen and P. Rosen, "DSPCP: A data scalable approach for identifying relationships in parallel coordinates," *IEEE Trans. Vis. Comput. Graphics*, vol. 24, no. 3, pp. 1301–1315, Mar. 2018.

[9] J. Johansson and C. Forsell, "Evaluation of parallel coordinates: Overview, categorization and guidelines for future research," *IEEE Trans. Vis. Comput. Graphics*, vol. 22, no. 1, pp. 579–588, Jan. 2016.

[10] H. Makwana, S. Tanwani, and S. Jain, "Axes re-ordering in parallel coordinate for pattern optimization," *Int. J. Comput. Appl.*, vol. 40, no. 13, pp. 43–48, 2012.

[11] K. Nohno, H.-Y. Wu, K. Watanabe, S. Takahashi, and I. Fujishiro, "Spectral-based contractible parallel coordinates," in *Proc. 18th Int. Conf. Inf. Visualisation*, Jul. 2014, pp. 7–12.

[12] L. F. Lu, M. L. Huang, and J. Zhang, "Two axes re-ordering methods in parallel coordinates plots," *J. Vis. Lang. Comput.*, vol. 33, pp. 3–12, Apr. 2016.

[13] 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 *Proc. IEEE Symp. Inf. Visualizat.*, Mar. 2003, pp. 105–112.

[14] W. Peng, M. O. Ward, and E. A. Rundensteiner, "Clutter reduction in multi-dimensional data visualization using dimension reordering," in *Proc. IEEE Symp. Inf. Visualizat.*, 2004, pp. 89–96.

[15] E. Koh, M. Blumenschein, L. Shao, and T. Schreck, "Reordering sets of parallel coordinates plots to highlight differences in clusters," in *Proc. EuroVA Int. Workshop Vis. Analytics*, 2022.

[16] R. G. Raidou, M. Eisemann, M. Breeuwer, E. Eisemann, and A. Vilanova, "Orientation-enhanced parallel coordinate plots," *IEEE Trans. Vis. Comput. Graphics*, vol. 22, no. 1, pp. 589–598, Jan. 2016.

[17] X. Yuan, P. Guo, H. Xiao, H. Zhou, and H. Qu, "Scattering points in parallel coordinates," *IEEE Trans. Vis. Comput. Graphics*, vol. 15, no. 6, pp. 1001–1008, Nov. 2009.

[18] R. A. Becker, W. S. Cleveland, and A. R. Wilks, "Dynamic graphics for data analysis," *Stat. Sci.*, vol. 2, no. 4, pp. 355–383, Nov. 1987.

[19] R. A. Becker and W. S. Cleveland, "Brushing scatterplots," *Technometrics*, vol. 29, no. 2, pp. 127–142, May 1987.

[20] R. Becker and W. Cleveland, "Brushing a scatterplot matrix: High-interaction graphical methods for analyzing multidimensional data," AT&T Bell Laboratories Tech. Memorandum, Murray Hill, NJ, USA, Tech. Rep., 1984.

[21] H. Qu, W.-Y. Chan, A. Xu, K.-L. Chung, K.-H. Lau, and P. Guo, "Visual analysis of the air pollution problem in Hong Kong," *IEEE Trans. Vis. Comput. Graphics*, vol. 13, no. 6, pp. 1408–1415, Nov. 2007.

[22] H. Janetzko, M. Stein, D. Sacha, and T. Schreck, "Enhancing parallel coordinates: Statistical visualizations for analyzing soccer data," *Electron. Imag.*, vol. 28, no. 1, pp. 1–8, Feb. 2016.

[23] J. Bok, B. Kim, and J. Seo, "Augmenting parallel coordinates plots with color-coded stacked histograms," *IEEE Trans. Vis. Comput. Graphics*, vol. 28, no. 7, pp. 2563–2576, Jul. 2022.

[24] Z. Geng, Z. Peng, R. S. Laramee, J. C. Roberts, and R. Walker, "Angular histograms: Frequency-based visualizations for large, high dimensional data," *IEEE Trans. Vis. Comput. Graphics*, vol. 17, no. 12, pp. 2572–2580, Dec. 2011.

[25] M. El Meseery and O. Hoeber, "Geo-coordinated parallel coordinates (GCPC): Field trial studies of environmental data analysis," *Vis. Informat.*, vol. 2, no. 2, pp. 111–124, Jun. 2018.

[26] C. Schmid and H. Hinterberger, "Comparative multivariate visualization across conceptually different graphic displays," in *Proc. 7th Int. Work. Conf. Scientific Stat. Database Manage.*, 1994, pp. 42–51.

[27] P. C. Wong and R. D. Bergeron, "Multivariate visualization using metric scaling," in *Proc. Vis.*, Oct. 1997, pp. 111–118.

[28] H. Hauser, F. Ledermann, and H. Doleisch, "Angular brushing of extended parallel coordinates," in *Proc. IEEE Symp. Inf. Visualizat.*, Jan. 2002, pp. 127–130.

[29] B. Shneiderman, "The eyes have it: A task by data type taxonomy for information visualizations," in *The Craft of Information Visualization*. Amsterdam, The Netherlands: Elsevier, 2003, pp. 364–371.

[30] E. Parzen, "On estimation of a probability density function and mode," *Ann. Math. Statist.*, vol. 33, no. 3, pp. 1065–1076, Sep. 1962.

[31] M. Rosenblatt, "Remarks on some nonparametric estimates of a density function," *Ann. Math. Statist.*, vol. 27, no. 3, pp. 832–837, 1956.

[32] G. Zhang, "Detecting and visualizing observation hot-spots in massive volunteer-contributed geographic data across spatial scales using GPU-accelerated kernel density estimation," *ISPRS Int. J. Geo-Inf.*, vol. 11, no. 1, p. 55, Jan. 2022.

[33] T. N. Chan, P. L. Ip, B. Choi, and J. Xu, "SWS: A complexity-optimized solution for spatial–temporal kernel density visualization," *Proc. VLDB Endowment*, vol. 15, no. 4, pp. 814–827, Dec. 2021.

[34] T. N. Chan, P. L. Ip, W. H. Tong, S. Mittal, Y. Li, and R. Cheng, "KDV-explorer: A near real-time kernel density visualization system for spatial analysis," *Proc. VLDB Endowment*, vol. 14, no. 12, pp. 2655–2658, Jul. 2021.

[35] Y. Li, M. Abdel-Aty, J. Yuan, Z. Cheng, and J. Lu, "Analyzing traffic violation behavior at urban intersections: A spatio-temporal kernel density estimation approach using automated enforcement system data," *Accident Anal. Prevention*, vol. 141, Jun. 2020, Art. no. 105509.

[36] Y. Zheng, Y. Ou, A. Lex, and J. M. Phillips, "Visualization of big spatial data using coresets for kernel density estimates," *IEEE Trans. Big Data*, vol. 7, no. 3, pp. 524–534, Jul. 2021.

[37] M. C. Minnotte, D. J. Marchette, and E. J. Wegman, "The bumpy road to the mode forest," *J. Comput. Graph. Statist.*, vol. 7, no. 2, pp. 239–251, Jun. 1998.

[38] S. Johansson and J. Johansson, "Interactive dimensionality reduction through user-defined combinations of quality metrics," *IEEE Trans. Vis. Comput. Graphics*, vol. 15, no. 6, pp. 993–1000, Nov. 2009.

[39] M. Bostock, V. Ogievetsky, and J. Heer, "D^3 data-driven documents," *IEEE Trans. Vis. Comput. Graphics*, vol. 17, no. 12, pp. 2301–2309, Dec. 2011.

[40] J. Li, J.-B. Martens, and J. J. van Wijk, "Judging correlation from scatterplots and parallel coordinate plots," *Inf. Visualizat.*, vol. 9, no. 1, pp. 13–30, Jan. 2010.

[41] Y. Orooji, E. Ghasali, N. Emami, F. Noorisafa, and A. Razmjou, "ANOVA design for the optimization of TiO_2 coating on polyether sulfone membranes," *Molecules*, vol. 24, no. 16, p. 2924, Aug. 2019.

[42] N. Atia, A. Benzaoui, S. Jacques, M. Hamiane, K. E. Kourd, A. Bouakaz, and A. Ouahabi, "Particle swarm optimization and two-way fixed-effects analysis of variance for efficient brain tumor segmentation," *Cancers*, vol. 14, no. 18, p. 4399, Sep. 2022.

[43] H. M. Lee, Y. Hua, J. Xie, and H. P. Lee, "Parametric optimization of local resonant sonic crystals window on noise attenuation by using Taguchi method and ANOVA analysis," *Crystals*, vol. 12, no. 2, p. 160, Jan. 2022.

[44] M. J. Blanca, J. Arnau, F. J. García-Castro, R. Alarcón, and R. Bono, "Non-normal data in repeated measures ANOVA: Impact on type i error and power," *Psicothema*, vol. 35, no. 1, pp. 21–29, 2023.

[45] D. Naganaidu and Z. M. Khalid, "ANOVA assisted variable selection in high-dimensional multicategory response data," *Math. Statist.*, vol. 11, no. 1, pp. 92–100, Jan. 2023.

JIA LIU was born in Tianjin, China, in 1984. She received the B.S. degree in communication engineering from the Tianjin University of Commerce, in 2007, and the M.S. degree in software engineering from the Beijing University of Technology, in 2010. She is currently pursuing the Ph.D. degree in information and communication engineering with Space Engineering University.

From 2010 to 2013, she was a Lecturer with the Information University of China. Since 2014, she has been a Software Engineer and has developed several software applications. Her current research interests include high-dimensional spatiotemporal data visualization and visual analysis, digital earth, and big data analysis.

GANG WAN was born in Hebei, China, in 1972. He received the Ph.D. degree in cartography and geographic information engineering, in 2006.

He is currently a Professor with the School of Space Information, Space Engineering University, the Director of the Department of Surveying and Mapping and Space Environment, and a Ph.D. Advisor. He is the author of more than ten books, more than 100 articles, and more than 20 inventions. His current research interests include virtual geographic environment, spatiotemporal big data, and spatial environment exploration.

Dr. Wan was a member of the System Simulation Society's Simulator Professional Committee and the Society of Surveying and Mapping's Professional Committee for cartography and geographic information engineering, in 2011. He has received two second-place awards: the National Teaching Achievements Award and the National Science and Technology Progress Award. He was elected for the Ministry of Education's New Century Outstanding Talents Support Program, in 2008. He received the Henan Province Youth Science and Technology Award, in 2010. He was selected as the Special Expert of the Ministry of Civil Affairs' Key Laboratory for Disaster Assessment and Risk Prevention, in 2012. He received the 2013 Xia Jianbai Surveying and Mapping Entrepreneurship and Technological Innovation Award. He received the National University GIS Innovation Talent Award, in 2014.

SHUAI WANG received the Ph.D. degree in communication and information systems from Space Engineering University, in 2018. He is currently a Lecturer with Space Engineering University. His research interest includes three-dimensional imaging.

PANKE DENG received the M.S. degree in mechanical engineering from Beijing Jiaotong University, Beijing, China, in 2020. He is currently pursuing the Ph.D. degree with Space Engineering University. His main research interests include information communication, remote sensing, deep learning, and visual analysis.

ZHIJUAN SU received the M.S. degree in cartography and geographic information system from Hebei Normal University, Hebei, China, in 2015. She is currently pursuing the Ph.D. degree with the School of Space Information, Space Engineering University. Her main research interests include deep learning, and satellite video target detection and tracking.

CHU LI received the B.S. degree in navigation engineering from Space Engineering University, Beijing, China, in 2022, where he is currently pursuing the master's degree in information and communication engineering. His main research interests include visualization, mars remote sensing, and deep learning.

YAO MU received the Ph.D. degree in environmental science and engineering from the China University of Geosciences.

Since 2019, she has been a Lecturer with Space Engineering University. She is the author of more than ten articles. Her current research interests include space weather forecast and spacecraft microbiome.

Dr. Mu received the Outstanding Graduate Student Award, in 2014 and 2015, and received the Top Academic Scholarship for five years running. She also received the National Scholarship for Graduate Student, in 2017.

YUNXIA YIN was born in Shaanxi, China, in 1974. She received the M.S. degree in software and network engineering from the Beijing University of Technology, in 2012. She is currently an Associate Professor with Space Engineering University. Her research interests include information visualization, big data visual analysis, artificial intelligence, and computer vision.