



PolarViz: a discriminating visualization and visual analytics tool for high-dimensional data

Yan Chao Wang¹ · Qian Zhang¹ · Feng Lin² · Chi Keong Goh³ · Hock Soon Seah²

Published online: 14 May 2018
© Springer-Verlag GmbH Germany, part of Springer Nature 2018

Abstract

Visual analytics tools are of paramount importance in handling high-dimensional datasets such as those in our turbine performance assessment. Conventional tools such as RadViz have been used in 2D exploratory data analysis. However, with the increase in dataset size and dimensionality, the clumping of projected data points toward the origin in RadViz causes low space utilization, which largely degenerates the visibility of the feature characteristics. In this study, to better evaluate the hidden patterns in the center region, we propose a new focus+context distortion approach, termed PolarViz, to manipulate the radial distribution of data points. We derive radial equalization to automatically spread out the frequency, and radial specification to shape the distribution based on user's requirement. Computational experiments have been conducted on two datasets including a benchmark dataset and a turbine performance simulation data. The performance of the proposed algorithm as well as other methods for solving the clumping problem in both data space and image space are illustrated and compared, and the pros and cons are analyzed. Moreover, a user study was conducted to assess the performance of the proposed method.

Keywords PolarViz · Customized radial distortion · Focus+context · Discriminating visualization · High-dimensional data analytics · Turbine performance assessment

1 Introduction

In recent years, data visualization has raised great research interest in the field of visual analytics as it provides an intuitive way to display large datasets, especially for high-dimensional datasets. Visualization of the high-dimensional dataset can be considered as a problem of how to map the data to a lower dimension in a useful way. A suitable data projection method enables one to observe and detect underlying data patterns and distributions in exploratory data analysis. A great deal of efforts have been devoted to this topic and various visualization methods have been proposed based on specific application requirements [20,21]. One of the popular

methods in high-dimensional data exploration and analysis is the radial visualization, such as RadViz [11]. However, with the increase in dimensionality, data points tend to clump toward the center of RadViz, known as clumping problem [6]. The patterns and features are thus buried in the clumping data points, so they cannot be detected by visualization functions such as query result display and outlier detection in our turbine performance assessment.

In this paper, we propose a focus+context visualization distortion techniques to ease the clumping problem and increase space utilization in RadViz. The clumping problem was analyzed before and one example is the RadViz with sigmoid function [26]. However, the method changes the values of the original dataset and destroys the cluster information of the dataset.

Our method plots high-dimensional data points into RadViz and analyzes the clumping problem by using a polar coordinate system instead of the Cartesian coordinate system. Based on the new representation, the radial distance shows its potential to alleviate the clumping problem of RadViz. Hence, we propose the idea on the modification of the radial distance distribution. For the distribution of the radial distance, we observe a similar problem in the image processing

✉ Feng Lin
ASFLIN@ntu.edu.sg

Yan Chao Wang
YWANG054@e.ntu.edu.sg

¹ Rolls-Royce@NTU Corporate Lab, Nanyang Technological University (NTU), Singapore, Singapore

² School of Computer Science and Engineering (SCSE), NTU, Singapore, Singapore

³ Rolls-Royce Singapore Pte Ltd, Singapore, Singapore

field. For images with low contrast, it seems that the intensity of pixels is clumping into a small range. Hence, similar to histogram-based operations in image processing, we propose a series of radial operations to manipulate the radius of each point.

This work is an improved and extended version of our previous work [38]. Besides the original contributions, which are the proposed radial equalization operation and radial specification operation, we extend the previous version mainly in three aspects. Firstly, we generalize the radial operations by considering the cumulative distribution functions (CDF) and propose a novel method termed PolarViz. Secondly, to better assess the proposed method, the study on similar methods and direct comparisons with them are conducted in this work. Specifically, we discuss other distortion methods in Sect. 2 and the technical details of these methods in Sect. 3.2. Then PolarViz is compared with existing methods in Sect. 6 from two perspectives: data space and image space, whereas in the previous work, only the comparison among different radial operations was done. The benchmark dataset is used in the experimental comparison for better understanding in this work. Thirdly, a user study in Sect. 7 is added to evaluate the usability of our method.

The main contributions of this paper are:

1. we present the RadViz by using a polar coordinate system. The clumping problem of RadViz can be clearly analyzed and visualized in the polar coordinate system;
2. we define the radial operations including radial equalization and radial specification for the radial distance of RadViz in order to adjust the RadViz plots;
3. we generalize the radial operations by considering the cumulative distribution function (CDF). Any modification on the CDF can be visualized in the new view.

The rest of the paper is organized as follows. Related work is reviewed in Sect. 2. Section 3 discusses the relevant concepts of RadViz and the clumping problem in the Cartesian coordinate system, and then, Sect. 4 discusses our new approach for the polar coordinate system. The detailed methodology is described in Sect. 5 including the radial operations in Sect. 5.1 and the generalization in Sect. 5.2. Applications using the proposed method for data analysis and comparisons with other methods are discussed in Sect. 6. User study to evaluate the usability of the proposed method is described in Sect. 7. Finally, technical discussion and conclusion are presented in Sect. 8.

2 Related work

Our work is a distortion-oriented approach to deal with the clumping problem in RadViz which is used on high-

dimensional data visualization. Hence, the related work is divided into three parts: Dimensionality reduction, RadViz, and distortion-oriented approaches.

2.1 Dimensionality reduction

Visual analysis on the high-dimensional dataset is a big challenge that is known as ‘curse of dimensionality’ [6]. To visualize high-dimensional dataset, one straightforward choice is to implement dimensionality projection techniques and then plot them in a lower dimension, while the other is to display all the dimensionality information at one time.

Dimensionality reduction that maps high-dimensional data point onto low-dimensional space (two-dimension or three-dimension) is crucial in the high-dimensional information visualization. There are two distinct groups in dimensionality reduction depending on the projection methods: linear projection and nonlinear projection. The most popular linear dimensionality reduction methods include Principal Component Analysis [28], and Linear Discriminate Analysis [9]. While for the nonlinear case, Multidimensional Scaling [18], Isomap [37], and Local Linear Embedding [30] are well-known methods adopted to reduce dimensionality. Researchers in the field of machine learning and artificial neural network have also developed methods to present data in low dimensions, such as t-SNE [17,22]. Besides these dimensionality reduction techniques, another choice is to use some visual mapping approaches to draw high-dimensional dataset onto low-dimension plots. Famous methods include scatter plot matrix, parallel coordinates plotting [14], and heatmap [29]. Radial visualizations that map data in a circular fashion are becoming an increasingly popular method in high-dimensional information visualization research. A historical review of radial visualization can be found in [7]. Among various radial visualization methods, RadViz [11] is an emerging leader in recent years.

2.2 RadViz

In RadViz, all the dimensionalities of a data point in data space contribute to the final position of the projected data point in the 2D image space. In the basic RadViz, dimensional anchors (DAs) are evenly spaced around the perimeter of a circle. One end of a spring is attached to each dimensional anchor, while the other end is attached to a certain data point. The number of springs is equal to the data dimensionality. The equilibrium location of this data point in the RadViz spring system is where the sum of each spring force equals to zero. RadViz is well explored in terms of cluster representation, outlier detection and so on. The main challenge for RadViz is the placement of dimen-

sional anchors. Various algorithms have been proposed to place dimensional anchors to obtain desired configurations, followed by quality measurement methods for justification [1,5,32,33,35].

Some researchers focus on designing 3D RadViz with the purpose of getting a better configuration [2,13,25]. In [2], they used the mean value of each dimension as the third coordinate value, while in [25], the Euclidean distance of the projected data point to the origin in the high-dimensional space was used as the third coordinate. The 3D RadViz visualization scheme incorporates a third dimension to visualize the shape and convergence by using the distance to a reference hyper-plane. Their 3D RadViz can effectively visualize the Pareto-optimal fronts with more than three objectives and also can be used to evaluate the performance of an algorithm.

2.3 Distortion-oriented approaches

In the clumping problem, large amount of data points are crowded into a small region. To provide a detailed view of this crowded data region while retaining surrounding context to help keep analysts oriented is exactly a focus + context method [10,27]. Distortion is one of the focus + context techniques that can transform the display region so that focused regions are magnified, while contextual regions are demagnified.

The application of distortion-oriented techniques to data visualization has a relatively long history. The problem arises when using the small display window to view the large information systems [19]. This is quite similar to the clumping problem in RadViz. Trials of the distortion techniques include the polyfocal display [16], bifocal display [36], perspective wall [23], and graphical fish-eye views [34]. Among them, the fish-eye distortion is the most commonly used method which follows the evidence in cameras with wide fields-of-view in computer vision applications. Large amount polynomial and non-polynomial models of fish-eye radial distortion are proposed to simulate the distortion [12]. The detail can be found in the survey papers [12,19] and reference therein.

The main difference between our work and previous distortion-oriented approaches is that in our method, the user can control the transformation function while others fix the transformation function, though they provide other interactions.

In a more general view, the clumping problem can be treated as a visual clutter. Besides distortion-oriented approaches, other clutter reduction methods include sampling [4], filtering, change point size, and so on as mentioned in the survey paper [8]. We did not enclose the comparison between our distortion method and other clutter reduction methods as beyond our purpose.

3 Main problems to be solved

3.1 Clumping problem in RadViz

For dataset $D = (d_1, \dots, d_j, \dots, d_m)$ containing m data points, each data point $d_j = (d_{1,j}, \dots, d_{i,j}, \dots, d_{n,j})$ has n -dimensionalities. Let $d'_{i,j}$ be the normalization result for $d_{i,j}$. The normalization equation for $d_{i,j}$ is

$$d'_{i,j} = \frac{d_{i,j} - \min_i}{\max_i - \min_i} \tag{1}$$

where $\min_i = \min \{d_{i,j}\}$ and $\max_i = \max \{d_{i,j}\}$, $\forall j$.

Then the basic RadViz with evenly distributed DAs can be expressed as:

$$R = \begin{cases} x_{1,j} = \frac{\sum_{i=1}^n (d'_{i,j} \cdot \cos(\theta_i))}{\sum_{i=1}^n d'_{i,j}} \\ x_{2,j} = \frac{\sum_{i=1}^n (d'_{i,j} \cdot \sin(\theta_i))}{\sum_{i=1}^n d'_{i,j}} \end{cases} \tag{2}$$

where θ_i refers to the orientation of the i th DA.

The mapping in Eq. 2 is nonlinear. Due to that, an effect that data points tend to clutter in the center of the plotting exists and is often known as the clumping problem as shown in Fig. 1. RadViz's clumping problem was first observed and analyzed in [6], and the effect of diametrically opposed dimensional anchors under the spring-force analogy is considered as the main reason.

To ease the clumping effect by changing the order of DAs is an inefficient way. Firstly, for general cases, the clumping effect exists under every configuration. Secondly, the exploration for the optimal result by exhaustively searching all possible configurations is very time-consuming with a complexity of $\mathcal{O}(n!)$ and practically intractable for not too large n . More importantly, the order of DAs (as well as the orientation of DAs) is often used by researchers to get a better plotting with more information [1]. Hence, it is highly inef-

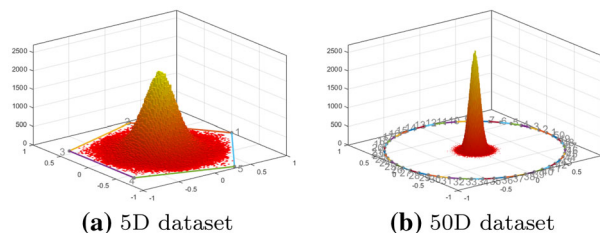


Fig. 1 We demonstrate the clumping effect by plotting the distribution of 100,000 uniformly sampling 5-dimensional and 50-dimensional data points. The 3D view of the distribution shows that most data points are clumped in the center area in 50D case (b) than that in 5D case (a)

ficient to change the order of DAs to tackle the clumping effect.

3.2 Previous approach for clumping problem

The clumping problem in RadViz was analyzed in [26]. They proposed a filtering mechanism to cancel the forces in RadViz and reduce clutter in the center region. Their approach was used for visualization of multi-task and multi-label classification, and applications with validation were given.

The filtering mechanism is a sigmoid weighting method. The filter operates by multiplying each dimension value $d'_{i,j}$ with a zero-one normalized sigmoid function

$$\hat{\sigma}(x, s, t) = \begin{cases} \frac{\sigma(x) - \sigma(0)}{\sigma(1) - \sigma(0)} & \text{if } \sigma(1) \neq \sigma(0), \\ 1 & \text{otherwise,} \end{cases} \quad (3)$$

with

$$\sigma(x) = \frac{1}{1 + \exp(-s(x+t))}.$$

Let d_j^s be the data point after sigmoid function. Then the original data point d_j will be changed to $d_j^s = d_j \cdot \hat{\sigma}(d'_{i,j}, s, t)$. The control parameters s and t are used to build the threshold. Finally, following the expression in Eq. 2, the RadViz with sigmoid function R_s can be expressed as:

$$R_s = \begin{cases} x_{1,j} = \frac{\sum_{i=1}^n (d_{i,j}^s \cdot \cos(\theta_i))}{\sum_{i=1}^n d_{i,j}^s} \\ x_{2,j} = \frac{\sum_{i=1}^n (d_{i,j}^s \cdot \sin(\theta_i))}{\sum_{i=1}^n d_{i,j}^s} \end{cases}. \quad (4)$$

However, as the high-dimensional data points are changed and the projection is nonlinear, the obtained RadViz plotting cannot be estimated. An example is shown in Fig. 2. In Fig. 2a, c, two data points A and B are plotted in RadViz at the same location and their dimension values are labeled. After the sigmoid weighting operation, data point A is moved toward the circle farther than data point B. For data point A, the variance of each dimension is higher than that of data point B. Sigmoid weighting function has less influence on higher values while reduces lower values significantly. Hence, it is obvious that the cluster information of original high-dimensional dataset will be destroyed. The experiment results when handling benchmark dataset and our engine-related datasets in Sect. 6 also show this drawback of sigmoid function.

Distortion methods operated in the image space can be used to ease the clumping problem in RadViz, though they are not designed for this purpose. Fish-eye distortion designed

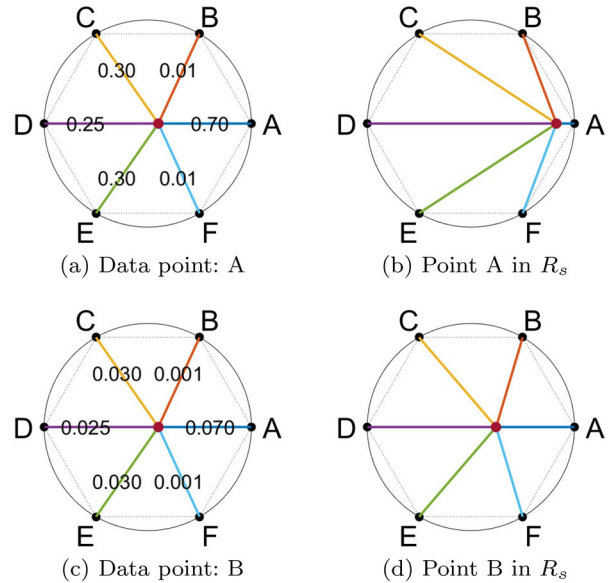


Fig. 2 Data points A and B have the same location in RadViz as shown in (a) and (c). However, after the sigmoid weighting operation, the obtained results are quite different as shown in (b) and (d)

upon the observation from fish-eye camera lens is the most commonly used distortion method, and we take the graphical fish-eye views (GFV) [34] and fish-eye transform (FET) [3] as examples. The transformation functions of these two methods are displayed as follows:

$$T_{GFV} = \frac{(1 + \lambda)x}{\lambda x + 1}, \quad T_{FET} = \frac{\ln(1 + \lambda x)}{\ln(1 + \lambda)} \quad (5)$$

where distortion factor λ controls the amount of the distortion and x is the distance from a point under consideration to the point of focus. x is a normalized distance which can have a value between 0 and 1. To better illustrate their difference, we plot two transformation functions under different distortion factors in Fig. 3.

The distortion model in each distortion method is fixed. Though the user can change the distortion factor and interactively select the focus, the user cannot modify the transformation function on data distribution. Compared with our proposed method in which the user can change the transformation model according to the data distribution, the difference is significant.

4 Proposed approach

4.1 Formulation in polar coordinate

As RadViz is a radial configuration, the projected data points can also be presented in the polar coordinate system. We use radial distance r to present the radial coordinate and orienta-

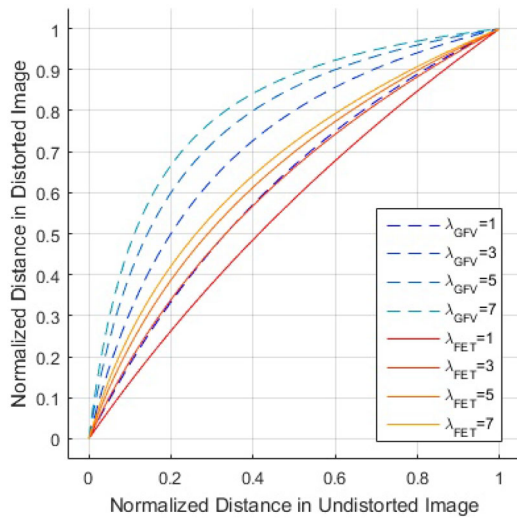


Fig. 3 Transformation functions of GFV and FET under different distortion factors

tion θ to present the angular coordinate. Then the data point $R(x_{1,j}, x_{2,j})$ in Cartesian coordinates can be converted to $R(r_j, \theta_j)$ in the polar coordinates with $0 \leq r_j \leq 1$ and θ_j in the interval $(-\pi, \pi]$ by:

$$R = \begin{cases} r_j = \sqrt{x_{1,j}^2 + x_{2,j}^2} \\ \theta_j = \text{atan2}(x_{2,j}, x_{1,j}) \end{cases}, \tag{6}$$

where $\text{atan2}(y, x)$ is defined as

$$\text{atan2}(y, x) = \begin{cases} \arctan(y/x) & \text{if } x > 0 \\ \arctan(y/x) + \pi & \text{if } x < 0, y \geq 0 \\ \arctan(y/x) - \pi & \text{if } x < 0, y < 0 \\ \pi/2 & \text{if } x = 0, y > 0 \\ -\pi/2 & \text{if } x = 0, y < 0 \\ \text{undefined} & \text{if } x = 0, y = 0 \end{cases}. \tag{7}$$

The orientation θ_j of a data point in RadViz indicates the dimension in which the original high-dimensional data point has larger deviation. Meanwhile, the radial distance r_j shows the relative extent of the deviation. RadViz is designed not for numerical analysis but to gain insights from the plotting. Users can have an overall understanding on the whole dataset in the initial exploration, and then they may know where and how to conduct further numerical analysis.

4.2 Visual analytics in polar coordinate

The clumping problem can be further explained by considering the radial distance of each projected data points in image

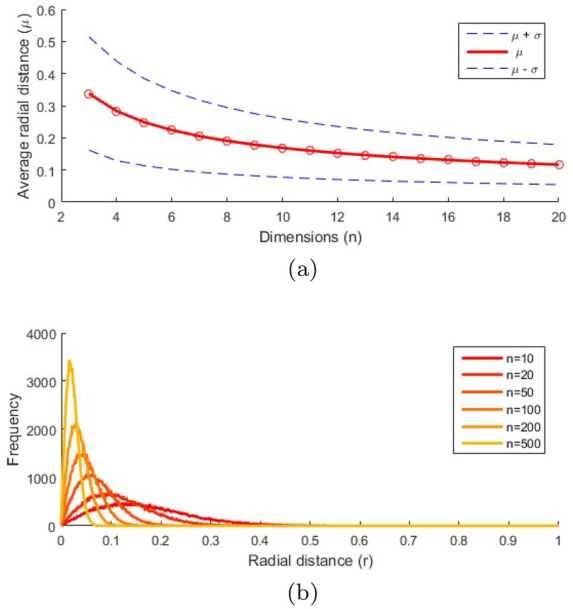


Fig. 4 The average distance of each point to the origin in the display decreases as the dimensionality increases in RadViz, as shown in (a). The plot also includes the average distance $\mu \pm$ one standard deviation σ (dashed lines). The graphics in (b) shows line histograms of the distances plotting 10–500-dimensional uniformly generated random dataset. The clumping effect becomes severe as the dimensionality increases [31]

space. The average radial distance μ and the standard deviation σ can be expressed as:

$$\mu = \frac{\sum_{j=1}^m r_j}{m}, \quad \sigma = \sqrt{\frac{\sum_{j=1}^m (r_j - \mu)^2}{m}}, \tag{8}$$

where m is the amount of data points.

Firstly, how μ and σ change under different dimensionality are explored. The average radial distance μ , $\mu + \sigma$, and $\mu - \sigma$ when plotting datasets with different dimensionality are illustrated in Fig. 4a by considering the maximum range as one unit. The average distances decrease as the dimensionality increases in RadViz. This is one strong illustration to show the clumping effect. Secondly, the distributions of distances to the origin under different dimensionality are also plotted in Fig. 4b. As the dimensionality increases, it can be seen that the peak of the plot is skewing to the left, which means that more data points are clumping toward the center. Besides the analysis in Fig. 4a, b, numerical details including the average distance μ , the standard deviation σ , and $\mu + 3\sigma$ are listed in Table 1. The percentage of data points located in the range of $[0, \mu + 3\sigma)$ is very steady and slightly increasing as the dimensionality increases. However, we should notice the significant decrease in the $\mu + 3\sigma$ value. When plotting the 50D dataset, around 99.4% data points in the uniformly distributed random dataset are plotted in the

Table 1 Numerical analysis on the clumping effect

n	μ	σ	$\mu + 3\sigma$	%
10	0.168	0.091	0.442	99.301
20	0.117	0.062	0.304	99.356
50	0.073	0.039	0.189	99.381
100	0.051	0.027	0.132	99.429
200	0.036	0.019	0.093	99.403
500	0.023	0.012	0.059	99.442

The average distance μ , the standard deviation σ , and value of $\mu + 3\sigma$ for different dimensionalities n are listed. The last column is the percentage of data points that located in the range of $[0, \mu + 3\sigma)$. Though there is a significant decrease in $\mu + 3\sigma$, the percentage is steady

range of $[0, \mu + 3\sigma)$. When it comes to 500D dataset, the range that covers around 99.4% data points is only 0.059.

According to the analysis in Fig. 4 and Table 1, the clumping problem of RadViz $R(r, \theta)$ can be clearly expressed by using the radial distance r , but not the orientation θ . Hence, regarding the clumping problem of RadViz, we can convert it to how to handle the radial distance r in polar coordinate system. Based on this hypothesis, we consider the operations on r to ease the clumping problem in RadViz while keeping the orientation θ unchanged. We come to the ideas on the modification of r .

Change the distribution of radial distance r . This idea is motivated by one observation: the distance distribution in Fig. 4b is similar to the intensity histogram of a low-contrast image. In the image processing field, histogram equalization is widely used to solve this problem [15]. Hence, a straightforward and interesting idea is to use the histogram equalization as well as other histogram operations (histogram specification, histogram local equalization, and so on) to solve the clumping problem by manipulating the projected data points distribution in the basic RadViz. The work in [39] also used probability distribution histogram to enhance visualization. Their method extends a dimension to multiple new dimensions based on the histogram, and finds the optimal placement of dimension anchors for good visual clustering. In comparison, we manipulate the distribution of data points with the histogram. Without creating new dimensions and reordering DAs, our method allows a more intuitive interpretation of the layout. The detailed methodology is described in Sect. 5.1 and then generalized in Sect. 5.2.

5 Detailed methods

5.1 Radial operations

After projecting high-dimensional data points into RadViz in the polar coordinate system, the radial distance r_j of data point d'_j will be modified according to the new distribu-

tion, while the polar angle θ_j remains unchanged. We define our proposed methods as radial operations, including radial equalization (Sect. 5.1.1), radial specification (Sect. 5.1.2), radial movement (Sect. 5.1.3), and radial local equalization (Sect. 5.1.4).

5.1.1 Radial equalization

We propose a radial equalization technique here by combining histogram equalization and basic RadViz. For the RadViz $R(\theta, r)$ in the polar coordinate system, the orientation θ is preserved. The distribution of the radius r of all projected data points is shown in the histogram, and then a histogram equalization method is implemented to manipulate the radial distribution.

The pixel values of an image are discrete in the range of $[0, 255]$. However, in basic RadViz, when using the distance to the origin as the criteria to plot a histogram, the values are continuous. Hence, firstly, we need to specify the bins to discretize the distance values. The polar coordinate system is adopted to represent the basic RadViz in which the distance value can be obtained directly. For a n -dimensional data point d_j in dataset D , its corresponding projected data point in polar coordinate is $R(r_j, \theta_j)$. The whole distance range $([0, 1])$ is digitized into L bins $(\{H_0, H_1, \dots, H_{L-1}\})$. L is set as 1000 in this paper. Let h_k denotes the total number of projected data points whose r_j are located in the range of $[H_k, H_{k+1})$, then the probability density function (PDF) is

$$p(H_k) = \frac{h_k}{m}, \quad 0 \leq k < L, \tag{9}$$

where m is the number of data points in dataset D . The cumulative distribution function (CDF) is defined as

$$c(H_k) = \sum_0^k p(H_k), \quad 0 \leq k < L. \tag{10}$$

Thus the transform function of histogram equalization T can be defined as

$$T_{\text{equal}}(H_k) = H_0 + (H_{L-1} - H_0) \cdot c(H_k). \tag{11}$$

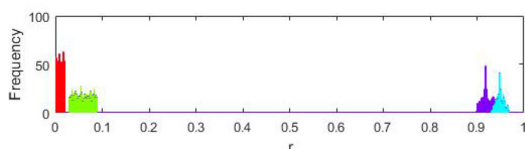
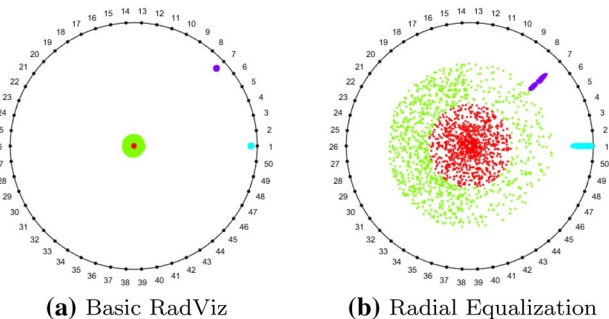
Suppose $R' = \{R'_j(r'_j, \theta_j)\}$ is defined as the RadViz with equalized radial distance r_j , then

$$R' = \begin{cases} r'_j = T_{\text{equal}}(r_j) \\ \theta_j = \text{atan2}(x_{2,j}, x_{1,j}) \end{cases}. \tag{12}$$

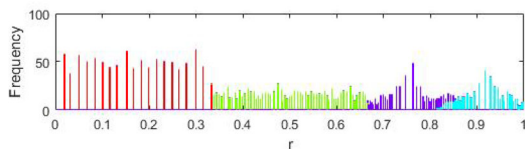
The equalization process is described in Algorithm 1. To better visualize the proposed algorithm, we design a dataset with four clusters ('red,' 'green,' 'blue,' and 'purple') as shown in Fig. 5a. For this synthetic example that

Algorithm 1 Radial Equalization

- 1: Dataset $D = (d_1, \dots, d_m)$;
- 2: Normalize D to get $D' = (d'_1, \dots, d'_m)$;
- 3: Plot D' in RadViz $R(r, \theta)$;
- 4: Set maximum histogram level L ;
- 5: Calculate histogram H by only considering r ;
- 6: Histogram equalization: $H' = T_{equal}(H)$;
- 7: Calculate the corresponding distance r' using H' ;
- 8: Calculate the new RadViz $R'(r', \theta)$;
- 9: Plot $R'(r', \theta)$;



(c) Histogram Before Equalization



(d) Histogram After Equalization

Fig. 5 **a** An example with data points located near the circle edge; **b** RadViz plotting after radial equalization. The histogram of original RadViz is plotted in (c). After histogram equalization, the histogram is plotted in (d)

cannot be handled by stretching, radial equalization method is used to ease the clumping and increase the space utilization. Figure 5b illustrates the result after employing radial equalization method. The distance histograms of Fig. 5a, b are plotted in Fig. 5c, d, respectively.

Radial equalization can automatically calculate the transformation function to reshape the distribution. This eases the burden of users. In some cases, however, the configuration obtained by the radial equalization may not be as desired by the users and the original patterns may be destroyed. The radial specification can address this issue, which allows the user to specify a histogram distribution. By employing these techniques, users can see a specific part of the dataset and explore the patterns or features.

5.1.2 Radial specification

The radial specification technique accepts a user-specified histogram distribution as an input to reshape the RadViz. When the clustering of projected data points is not clear or users have special requirements on the distribution, the radial specification can generate a required configuration. The dataset in Fig. 5a is again used here as an example to show the radial specification results. In Fig. 6c, a specified histogram is given and the corresponding plotting is illustrated in Fig. 6b. Compared with the radial equalization result in Fig. 5b, which plots clusters together, the radial specification results in Fig. 6b show the pattern in a visually clearer way.

The advantage of the radial specification is that it displays the data points in RadViz exactly as the user’s desire, and the radial specification operation will not destroy the relative distance of points to the origin. Let H' denotes the user-specified histogram and $c'(H)$ be the corresponding CDF. Then the specification process can be described in Algorithm 2. Compared with the radial equalization operation in Algorithm 1, the radial equalization is a special case of radial specification by using a uniform specified histogram as user input.

5.1.3 Radial movement

Besides bringing in a new histogram, the user also can edit the current histogram. Radial movement operation allows the user to select a range of bars in the histogram, and move the selected bars to another location. No collision with other bars is allowed during this movement in order to preserve the relative relationship. Based on the histogram in Fig. 5c, the middle cluster is selected and moved to the position of around the 400th bin, and the result is shown in Fig. 7c. With this modified histogram, the original RadViz in Fig. 5a is changed into Fig. 7a. As shown, cluster ‘green’ is now separated from cluster ‘red’ via the radial movement.

Algorithm 2 Radial Specification

- 1: Dataset $D = (d_1, \dots, d_m)$;
- 2: Normalize D to get $D' = (d'_1, \dots, d'_m)$;
- 3: Plot D' in RadViz $R(r, \theta)$;
- 4: Set maximum histogram level L ;
- 5: Calculate histogram H and the CDF $c(H)$ by only considering r for the original RadViz;
- 6: Calculate the CDF $c'(H')$ for the user specified histogram;
- 7: Find the histogram level H'_j for which $c(H_i) = c'(H'_j)$;
- 8: Then the histogram specification transformation function: $H'_j = T_{specify}(H_i)$;
- 9: Calculate the corresponding distance r' using H' ;
- 10: Calculate the new RadViz $R'(r', \theta)$;
- 11: Plot $R'(r', \theta)$;

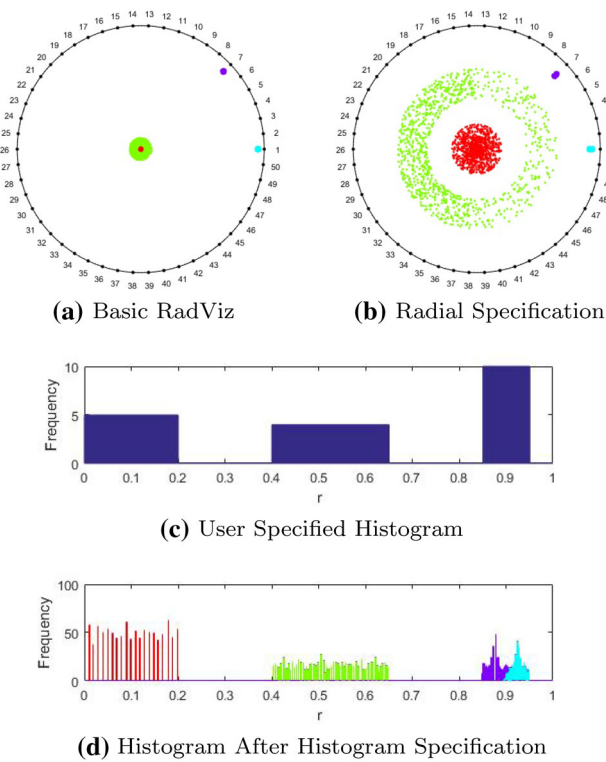


Fig. 6 **a** An example with data points located near the circle edge; **b** after radial specification, the RadViz is illustrated. The histogram in **(c)** is user specified. After radial specification, the histogram is plotted in **(d)**

5.1.4 Radial local equalization

Radial local equalization is conducted in the histogram by selecting a partial range and then implementing radial equalization in this range only. This operation is used to increase local contrast. As the example shown in Fig. 5a, after radial movement, we implement radial local equalization on cluster ‘green’ to increase the local contrast. The range [100, 650] is selected for local equalization. The resulting histogram and RadViz plotting are shown in Fig. 7b, d, respectively.

5.2 PolarViz

The proposed histogram-based radial operations in the previous four subsections can flexibly control the radial distribution by importing specified distribution or editing current distribution or both. Although these operations are defined in the histogram, the CDF is used for calculation. All the operations can be treated as the matching between the CDF before transformation and the CDF after transformation. Hence, we can generalize the radial operation by considering the matching of different CDFs. Let $c(H)$ be the CDF of the current RadViz while $c'(H)$ being the new CDF, then the histogram level H'_j in $c'(H)$ for which $c(H_i) = c'(H'_j)$ is found. The

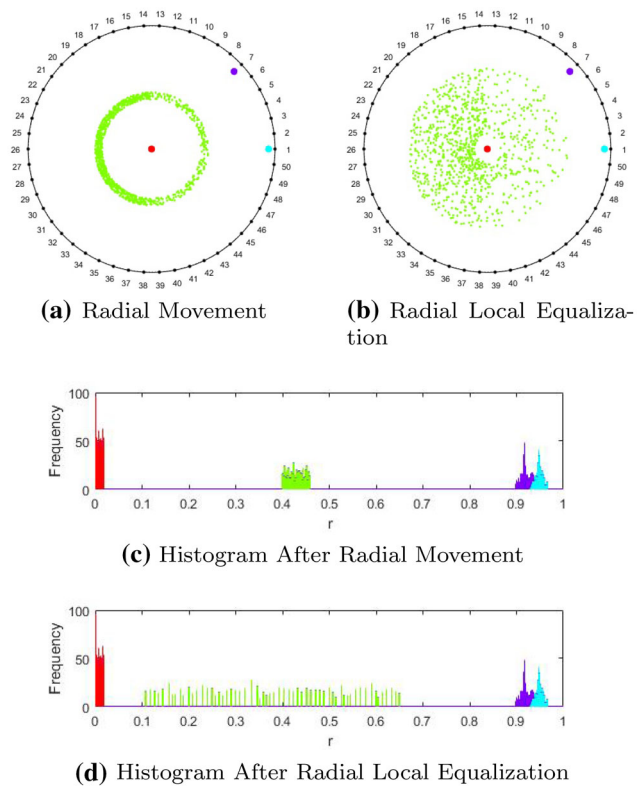


Fig. 7 **a** The RadViz plotting after radial movement; **b** the RadViz plotting after radial local equalization. After radial movement, the obtained histogram is plotted in **(c)**. After radial local equalization, the obtained histogram is plotted in **(d)**

result of this matching is the generalized radial transformation function.

Based on this, we propose the PolarViz plotting which can be treated as a series of radial operations to modify the radial distribution of RadViz. Let $T(\cdot)$ be the operator of the generalized radial distribution transformation, then the PolarViz R_{polar} can be expressed as

$$R_{polar} = \begin{cases} r_j^p = T(r_j) \\ \theta_j^p = \theta_j = atan2(x_{2,j}, x_{1,j}). \end{cases} \tag{13}$$

6 Comparative studies

The comparison with other methods that can be used to solve the clumping problem is conducted in this section. As the clumping problem is formed during the projection from data space to image space, the comparative studies are divided into two parts: comparison with methods operated in the data space (Sect. 6.1) and comparison with methods operated in the image space (Sect. 6.2).

For the methods operated in the data space, we execute the Ono’s method [26] in which the data point in the high-dimensional space is filtered by a sigmoid function. In the

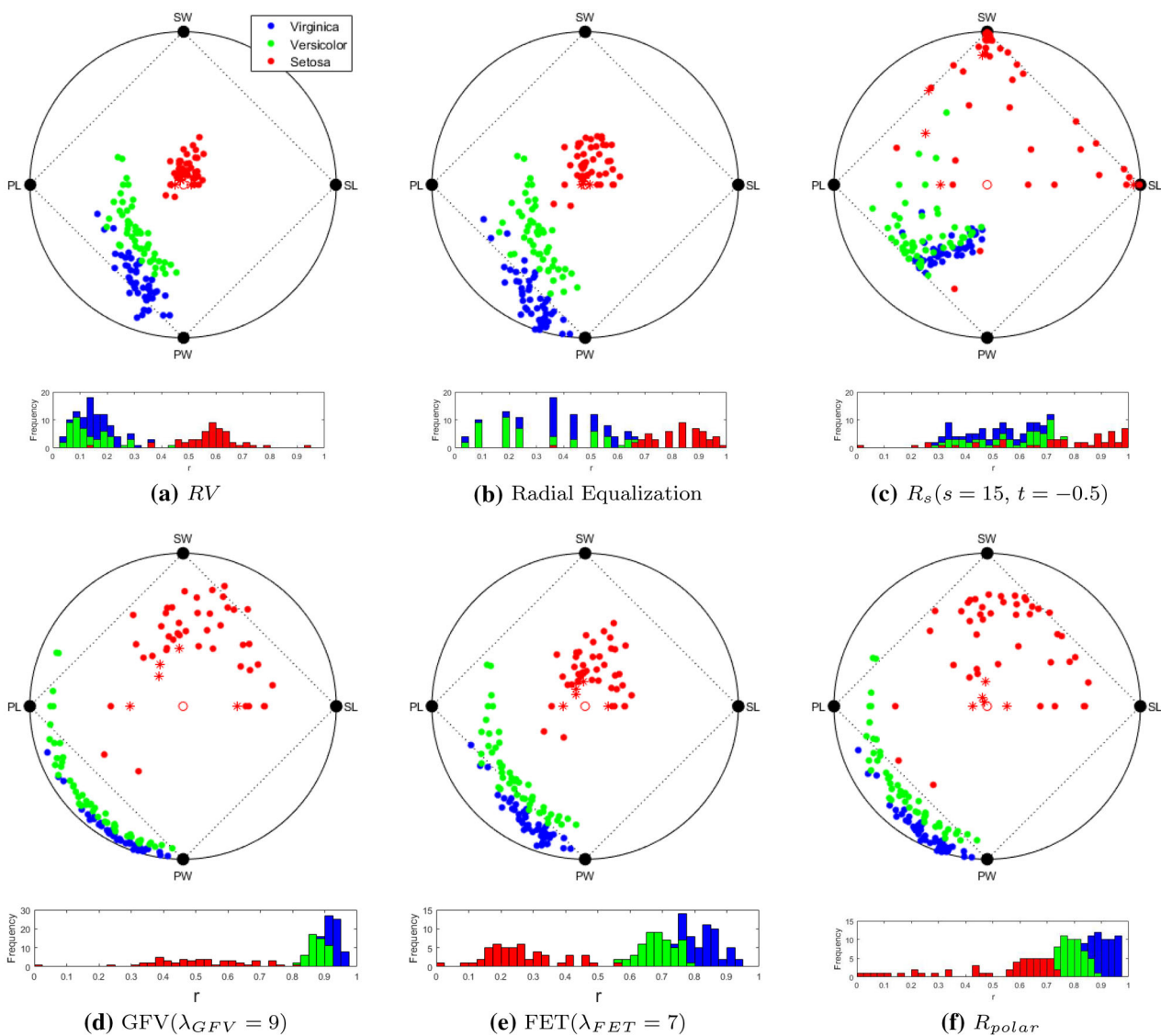


Fig. 8 We color clusters and use different symbols for different query-related data points. The original result using basic RadViz (RV) is shown in (a), while radial equalization result is shown in (b). The result

of the RadViz with sigmoid function (R_s) is illustrated in (c). The results of two distortion methods, GFV and FET, are illustrated in (d) and (e), respectively. In the end, the proposed PolarViz (R_{polar}) is shown in (f)

experiment, we set the parameter of the RadViz with sigmoid function (R_s) with $s = 15$ and $t = -0.5$.

For the methods operated in the image space, distortion methods for radial layout (i.e., GFV [34] and FET [3]) are implemented for comparison. For GFV, we set λ_{GFV} to 9 while setting λ_{FET} for FET to 7 in our experiment. The values of λ_{GFV} and λ_{FET} can be changed as well as the s and t for R_s .

Two datasets are used in the comparison experiment. The first one is the four-dimensional dataset ‘IRIS’ in which three clusters are involved, while the second dataset is a simulation dataset generated by the EngineSim. The EngineSim developed by NASA [24] is a turbine engine simulator used in our project to assess the performance of engine. We designed

3444 input data points that have 16 parameters for the Turbo Fun engine simulator and then obtained 3444 output data points. The 3444 output data points with 38 dimensionalities are used here.

With these two datasets, we aim to visualize the center region of the views after projection into image space. The center point is padded with zero value for each dimension and few nearest data points are highlighted with different symbols. Due to the clumping problem, these points are hard to view in the basic RadViz plot. Different methods are compared at this stage to provide a detailed view of the center region while retaining surrounding context to help keep analysts oriented [10].

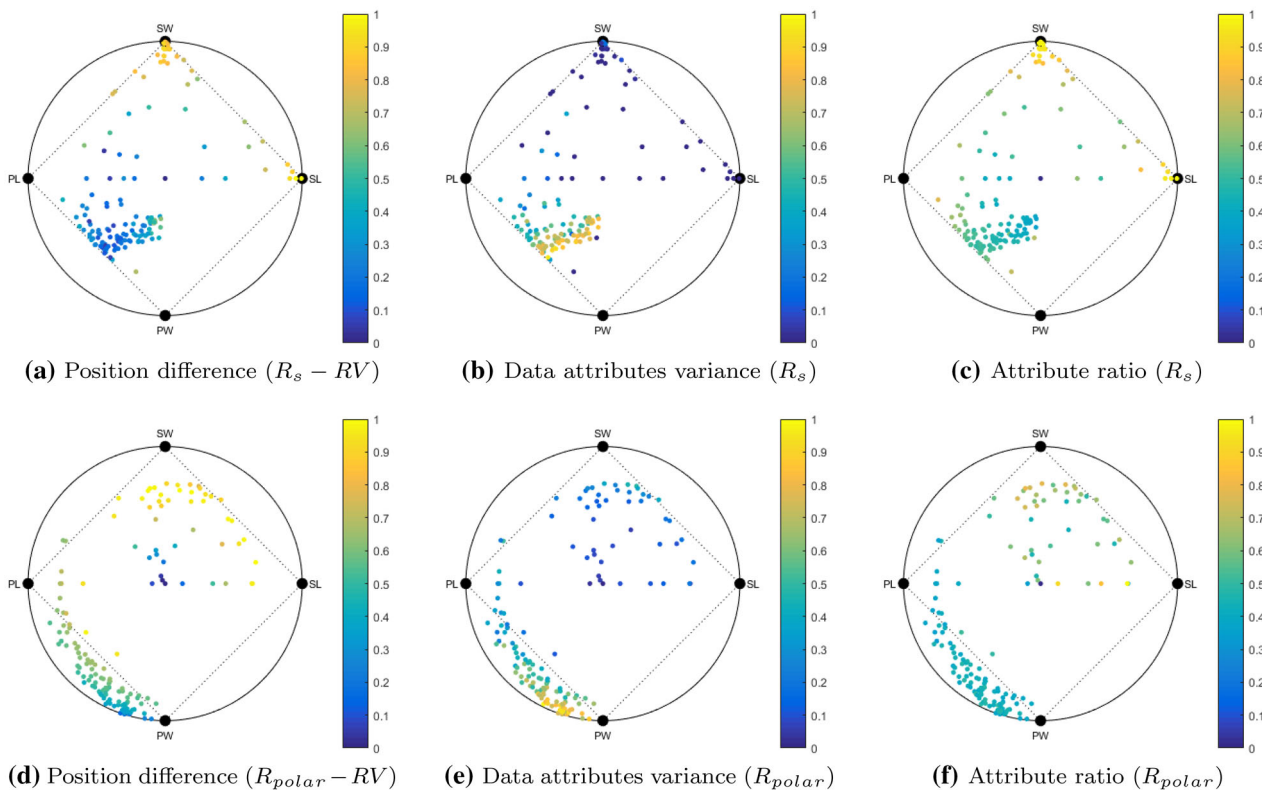


Fig. 9 **a** The position difference of same data point is plotted in R_s (Fig. 8c) and basic RadViz RV (Fig. 8a). **b** The attribute values variance of a data point is plotted in R_s . **c** The ratio of the largest attribute value to the mean attribute value is plotted in R_s . **d** The position differ-

ence of same data point is plotted in R_{polar} (Fig. 8f) and basic RadViz RV (Fig. 8a). **e** The attribute values variance of a data point is plotted in the PolarViz (R_{polar}). **f** The ratio of the largest attribute value to the mean attribute value is plotted in the PolarViz (R_{polar})

6.1 RadViz with sigmoid function

6.1.1 Iris

As the filter in R_s is nonlinear and the projection method in RadViz is also nonlinear, the result view of R_s will have quite different distribution when compared with basic RadViz. By using the benchmark dataset ‘IRIS,’ we plot the views obtained from the basic RadViz (RV) and R_s and the PolarViz (R_{polar}) in Fig. 8a, c, f, respectively.

Compared with the radial equalization operation on the basic RadViz result (Fig. 8b), R_s confuse the meaning of RadViz plot in the image space. Although R_s can significantly solve the clumping problem in this case, the position relationship among data points in the image space is destroyed. In Fig. 8c, the neighbors of the center point are not the nearest data points to the center anymore. To better understand the position change in these plots, we compare the data point position difference between R_s and RV in Fig. 9a while plotting the position difference between R_{polar} and RV in Fig. 9d. The position difference is calculated by considering the Euclidean distance of same data point in different views.

For each comparison, the difference is normalized into the interval of [0, 1] and then colorized accordingly. The color of data point in Fig. 9a indicates that the position change in R_s is irregular. Meanwhile, in Fig. 9d, a gradient ramp in the radial direction can be observed which indicates a more regular radial modification.

It does not mean that R_s is useless in data visualization, as the design of R_s is to highlight the attributes with large values [26]. In this case, we consider R_s from two perspectives. Firstly, we calculate the attribute value variance for each data point in R_s and R_{polar} (Fig. 9b, e). Secondly, we pay attention to the ratio of the largest attribute value to the attribute mean for each data point (Fig. 9c, f). The attribute variance and attribute ratio for each data point d_j is calculated as follows:

$$\text{AttributeVariance} = \sqrt{\frac{1}{n} \sum_{i=0}^n (d_{i,j} - \bar{d}_j)^2} \quad (14)$$

$$\text{AttributeRatio} = \frac{\max \{d_{i,j}\}, \forall i}{\sum_{i=0}^n d_{i,j}} \quad (15)$$

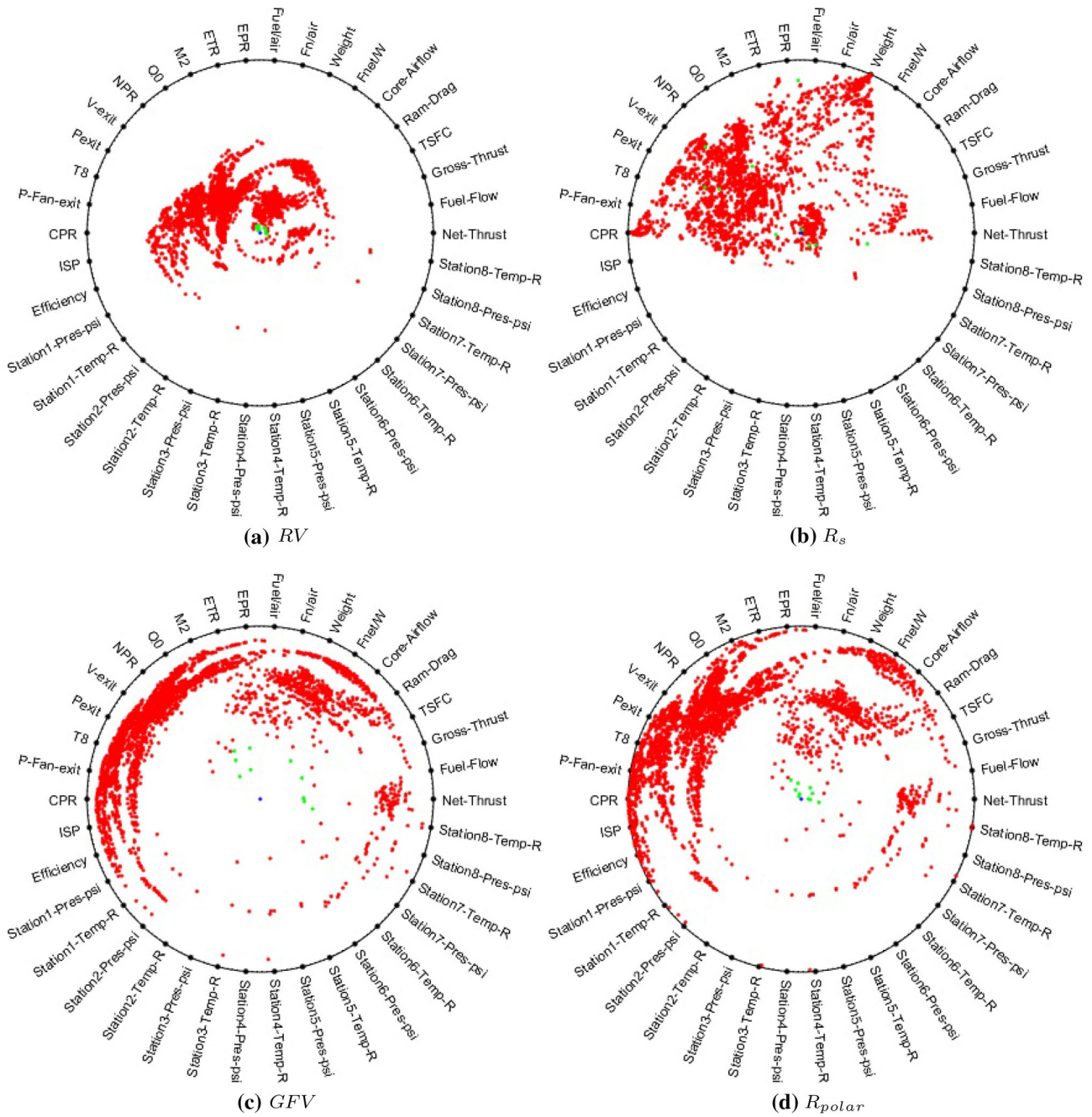


Fig. 10 The basic RadViz (a), R_s (b), GFV with $\lambda_{GFV} = 9$ (c), and R_{polar} (d) are used to plot the EngineSim dataset

where $\bar{d}_j = \frac{1}{n} \sum_{i=0}^n d_{i,j}$. The basic RadViz plots the data point which has small attribute variance near the center and plots the data point with large attribute variance away from the center. This character is preserved in R_{polar} as shown in Fig. 9e. It seems that the attribute variance in R_s is not associated with that in R_{polar} according to Fig. 9b. The main property of R_s is that after using the sigmoid function to filter the original dataset, the attribute with large values previously will have larger results while small values will have smaller

results. Hence, we further calculate the ratio of the attribute with largest value to the mean value in each data point. R_s shows clear insight in Fig. 9c. R_s plots data points which has largest values in attribute ‘SL’ and ‘SW’ near these two DAs, respectively. The dots located between ‘PL’ and ‘PW’ indicate that these data points have two attributes with similar values. Meanwhile, R_{polar} in Fig. 9f cannot show this information.

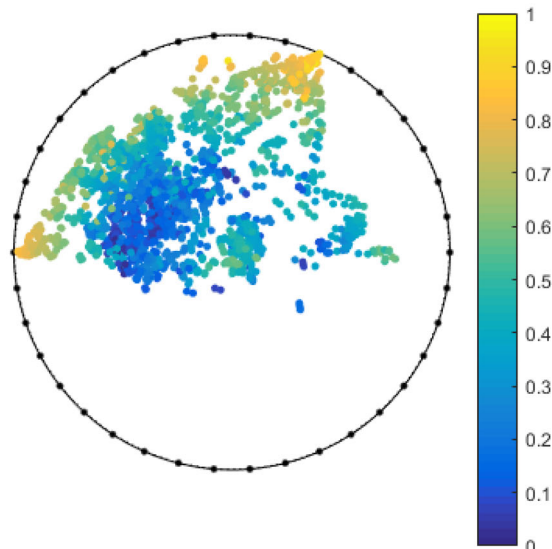
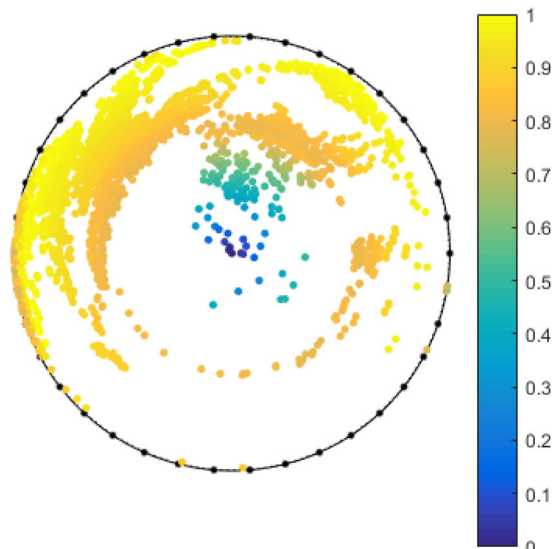
(a) Position difference ($R_s - RV$)(b) Position difference ($R_{polar} - RV$)

Fig. 11 In (a), we plot the position difference of data points in R_s (Fig. 10b) and basic RadViz RV (Fig. 10a). Then in (b) we compare the position difference between R_{polar} (Fig. 10d) and basic RadViz RV (Fig. 10a)

6.1.2 NASA EngineSim

Another example is using the EngineSim simulation dataset. Compared with 'IRIS,' there is no cluster label in the EngineSim dataset. We plot the visualization result of RV , R_s , and R_{polar} in Fig. 10a, b, d, respectively. As more data points are involved in the EngineSim dataset, the result of R_s still has many points crowded in the center region after the sigmoid filtering. In this case, the advantage of R_{polar} over R_s is more obvious.

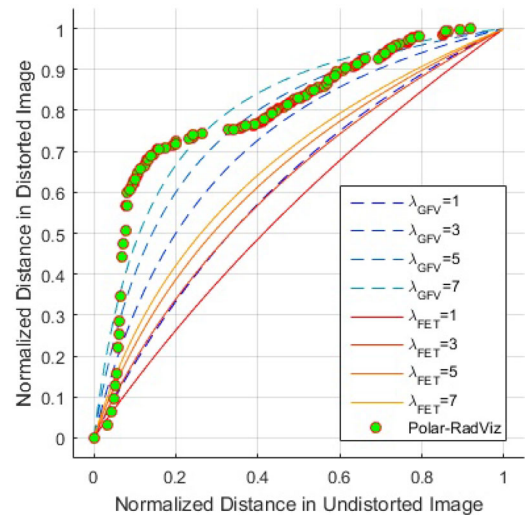


Fig. 12 We compare the transformation functions of the view in Fig. 8f with Fig. 3

We also plot the position difference of these three plots and the results are shown in Fig. 11. Similarly, like the result in Fig. 11, the position change in R_s confuses the meaning of RadViz plot.

6.2 Distortion methods

Techniques that can be used to solve the clumping problem in the image space, such as distortion methods (GFV and FET), are also compared with the proposed PolarViz upon the two datasets, the 'IRIS' and the EngineSim dataset. The results of distortion methods on two datasets are plotted in Figs. 8d, e and 10c, respectively.

Distortion methods are also operated along the radial direction. Hence, they can produce similar results with the PolarViz when dealing with the clumping problem. However, the major difference is that in the distortion methods the transformation functions are fixed. Once the distortion parameter is determined, the user cannot focus only on one region while keeping other parts unchanged. The distortion generated is hard to control in these distortion methods, making them difficult to meet the various requirements. Take the result shown in Fig. 10c as an example. The center region is too distorted, while most of the data points are pushed toward the bounding circle. In this case, as the transformation function is fixed, the user cannot modify the view. In the contrast, R_{polar} has more space for the rest of data points while clearly viewing the center region. The user can specify the distribution to satisfy the requirement. The transformation function of R_{polar} as well as those of distortion methods with different parameters are plotted in Fig. 12.

The attribute variance, as well as the attribute ratio, are not calculated in the comparison between distortion methods and

Table 2 The user study process and the corresponding questionnaire are described in this table

User study	Questionnaire
Introduce the RadViz plot by using the spring-force model; then using the equations in the Cartesian coordinate system; then using the equation in the Polar coordinate system	Q1: Sort the description models (the spring-force model, the equations in the Cartesian coordinate, and the equations in the polar coordinate) of RadViz from 'easy-to-understand' to 'difficult-to-understand'
Introduce the clumping problem in RadViz by using the Cartesian coordinate system; then a polar coordinate system	Q2: Sort the description models (the Cartesian coordinate system and the polar coordinate system) of the clumping problem in RadViz from 'easy-to-understand' to 'difficult-to-understand'
Introduce the parameters of each method (RadViz, R_s , GFV, FET, and R_{polar})	
Participants were asked to modify the parameters of each method to achieve a focus + context purpose	Q3: Sort the methods for the clumping problem regarding performance and operation difficulty from easy to difficult

the PolarViz. The PolarViz and the distortion methods are operated in the image space. Hence, they will not change the original data values. The attribute variance and the attribute ratio for each data point are persevered during plotting.

The amount of bins used in the radial operations is of great importance. The more bins used, the more accurate the result is. In Fig. 8b, f, when plotting the radial equalization result and R_{polar} , we use 1000 bins. Meanwhile, to better illustrate the overlapping of different clusters, we use 40 bins in the corresponding histograms.

7 User study

We performed a user study to assess the usability of the proposed PolarViz. The user study was divided into two parts. In the first part, participants in the user study were taught some background knowledge about the RadViz, the clumping problem, and the methods for solving the clumping problem. In the second part, participants were asked to explore different methods by themselves with a certain target and then sort these methods based on the performance.

In the background knowledge introduction part, we first introduced the RadViz plot by using the spring-force model and then using equations in the Cartesian coordinate system and then followed by equations in a polar coordinate system. Secondly, we introduced the clumping problem in RadViz by using the Cartesian coordinate system and a polar coordinate system, respectively. Finally, we briefly introduced the parameter of each method while keeping the theory of all methods uninvolved. In the second part, three datasets

including the 'IRIS' dataset, the EngineSim dataset, and a random dataset were used. We plotted the three datasets into the basic RadViz and colorized the center point and its nearest neighbors. The RadViz plot is used as a reference. We then plotted the three datasets by using R_s , GFV, FET, and R_{polar} , respectively. Participants were asked to modify the parameters of each method to achieve a focus + context purpose. It requires a focus on the center region as well as a clear view of the highlighted data points. Meanwhile, the pattern of the rest should be preserved as much as possible. The user study content is summarized in the left column of Table 2.

Fourteen graduate students and researcher staff (11 male, 3 female) participated in our user study. None of them reported to be familiar with information visualization before, and 11 of them reported to be familiar with the general spring-force model as well as the equation expression, and 10 of them knew about the fish-eye distortion. The average length of the study was about 30 minutes as around 15 minutes were taken for the background knowledge introduction. Three questions were asked in the user study. In the first question, participants were asked to sort the description models of RadViz from 'easy-to-understand' to 'difficult-to-understand.' In the second question, participants were asked to sort the description model of the clumping problem in RadViz from 'easy-to-understand' to 'difficult-to-understand.' In the third question, participants were asked to sort the methods for the clumping problem regarding performance and operation difficulty. The questionnaire is summarized in the right column of Table 2. The overall opinion indicated that the proposed R_{polar} overcomes other methods. Specifically, the opinions for each question are as follows:

- Q1:** All the participants favored the spring-force model. Thirteen of them mentioned that the equations of the RadViz in the Cartesian coordinate are not difficult to understand while only 1 participant supported the equation expression in the polar coordinate rather than the Cartesian coordinate. One reason is that using the spring-force model to explain the RadViz plot makes sense. Meanwhile, the evidence that most of the participants were familiar with the spring-force model as well as the equation expression in the Cartesian coordinate may be another reason.
- Q2:** Ten participants voted the clumping problem analysis in the polar coordinate system while 4 participants thought the analysis in the Cartesian coordinate system and in the polar coordinate system are the same.
- Q3:** Eleven participants voted R_{polar} as the easiest method while 3 participants favored one of the distortion methods (GFV or FET). The reason why we put GFV and FET together is that during the study, comments provided by some participants indicate that they failed to find the difference between GFV and FET. Hence, we treated the GFV and FET as one method. In this case, R_{polar} was 11 times in the first position and 3 times in the second position. Meanwhile, the GFV or FET methods were 3 times in the first position and 11 times in the second position. The R_5 was 14 times in the third position which means that all the participants found it difficult to meet the objective.

The user study results support the observation that the proposed R_{polar} , which uses the histogram to modify the radial distribution, is easy to understand and provide more flexibility.

8 Discussion and conclusion

Regarding operation accuracy and computational performance, the proposed PolarViz method not only has a strong relationship with the dimensionality n and the total amount of data points m , but also largely depends on the number of bins L when doing digitization. To plot PolarViz, firstly the high-dimensional dataset needs to be projected into a low dimension space. The computational complexity of this mapping is $O(mn)$. Secondly, once we have the mapping result, the complexity of digitization is $O(m)$. Then no matter how many data points we have, after digitization there are L bins. Thirdly, for the radial operations, the complexity to handle all these bins is $O(L)$ and the complexity of re-plotting data points is $O(m)$. Hence, the total complexity of PolarViz is $O(mn + m + L + m)$. The value of n , m , and L will influence the final complexity of PolarViz.

Nevertheless, the proposed techniques have a limitation that the current distortion is only focusing on the center region of the plot. If the PolarViz is implemented in another region, confusion on the plot understanding and radial operations may arise.

In conclusion, this paper focuses on the clumping problem of RadViz by providing a series of radial-related operations to uncover hidden patterns and increases the space utilization. We can manipulate the distribution of data points with histogram by radial equalization and other radial specified operations. The latter shows the flexibility of our technique for different user requirements. Experimental results indicate the advantage of the PolarViz. Compared with other methods, the PolarViz not only preserves the pros of RadViz but also provides a flexible radial modification on the views.

Acknowledgements This work was conducted within the Rolls-Royce@NTU Corporate Lab with support from the National Research Foundation (NRF) Singapore under the Corp Lab@University Scheme. The work is largely extended from our CGI 2017 paper [38].

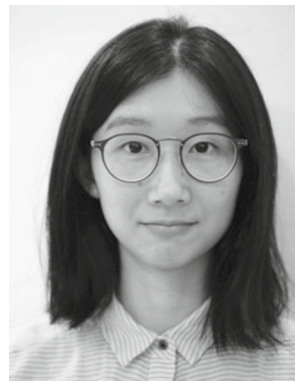
References

1. Albuquerque, G., Eisemann, M., Lehmann, D.J., Theisel, H., Magnor, M.: Improving the visual analysis of high-dimensional datasets using quality measures. In: IEEE Symposium on Visual Analytics Science and Technology (VAST), 2010, pp. 19–26. IEEE (2010)
2. Artero, A.O., de Oliveira, M.C.F.: Viz3d: effective exploratory visualization of large multidimensional data sets. In: Proceedings of 17th Brazilian Symposium on Computer Graphics and Image Processing, 2004, pp. 340–347. IEEE (2004)
3. Basu, A., Licardie, S.: Alternative models for fish-eye lenses. Pattern Recognit. Lett. **16**(4), 433–441 (1995)
4. Bertini, E., Santucci, G.: By chance is not enough: preserving relative density through nonuniform sampling. In: Proceedings of 8th International Conference on Information Visualisation, 2004. IV 2004, pp. 622–629. IEEE (2004)
5. Chen, K., Liu, L.: iVIBRATE: interactive visualization-based framework for clustering large datasets. ACM Trans. Inf. Syst. (TOIS) **24**(2), 245–294 (2006)
6. Daniels, K., Grinstein, G., Russell, A., Glidden, M.: Properties of normalized radial visualizations. Inf. Vis. **11**(4), 273–300 (2012)
7. Draper, G.M., Livnat, Y., Riesenfeld, R.F.: A survey of radial methods for information visualization. IEEE Trans. Vis. Comput. Graph. **15**(5), 759–776 (2009)
8. Ellis, G., Dix, A.: A taxonomy of clutter reduction for information visualisation. IEEE Trans. Vis. Comput. Graph. **13**(6), 1216–1223 (2007)
9. Fisher, R.A.: The use of multiple measurements in taxonomic problems. Ann. Eugen. **7**(2), 179–188 (1936)
10. Heer, J., Shneiderman, B.: Interactive dynamics for visual analysis. Queue **10**(2), 30 (2012)
11. Hoffman, P., Grinstein, G., Marx, K., Grosse, I., Stanley, E.: DNA visual and analytic data mining. In: Proceedings of Visualization'97, pp. 437–441. IEEE (1997)
12. Hughes, C., Glavin, M., Jones, E., Denny, P.: Review of geometric distortion compensation in fish-eye cameras. In: IET Conference Proceedings, pp. 162–167. Institution of Engineering and Technology (2008)

13. Ibrahim, A., Rahnamayan, S., Martin, M.V., Deb, K.: 3D-RadViz: visualization of Pareto front in many-objective optimization. In: IEEE Congress on Evolutionary Computation (CEC), 2016, pp. 736–745. IEEE (2016)
14. Inselberg, A.: *Parallel Coordinates*. Springer, Berlin (2009)
15. Jain, A.K.: *Fundamentals of Digital Image Processing*. Prentice-Hall, Inc., Upper Saddle River (1989)
16. Kadmon, N., Shlomi, E.: A polyfocal projection for statistical surfaces. *Cartogr. J.* **15**(1), 36–41 (1978)
17. Kohonen, T.: The self-organizing map. *Proc. IEEE* **78**(9), 1464–1480 (1990)
18. Kruskal, J.B.: Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. *Psychometrika* **29**(1), 1–27 (1964)
19. Leung, Y.K., Apperley, M.D.: A review and taxonomy of distortion-oriented presentation techniques. *ACM Trans. Comput. Hum. Interact. (TOCHI)* **1**(2), 126–160 (1994)
20. Liu, S., Cui, W., Wu, Y., Liu, M.: A survey on information visualization: recent advances and challenges. *Vis. Comput.* **30**(12), 1373–1393 (2014)
21. Liu, S., Maljovec, D., Wang, B., Bremer, P.T., Pascucci, V.: Visualizing high-dimensional data: advances in the past decade. *IEEE Trans. Vis. Comput. Graph.* **23**(3), 1249–1268 (2017)
22. Maaten, Lvd, Hinton, G.: Visualizing data using t-sne. *J. Mach. Learn. Res.* **9**, 2579–2605 (2008)
23. Mackinlay, J.D., Robertson, G.G., Card, S.K.: The perspective wall: detail and context smoothly integrated. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 173–176. ACM (1991)
24. NASA: Enginesim version 1.8a (2014). <https://www.grc.nasa.gov/www/k-12/airplane/ngnsim.html>
25. Nováková, L., Štepanková, O.: Radviz and identification of clusters in multidimensional data. In: 13th International Conference on Information Visualisation, 2009, pp. 104–109. IEEE (2009)
26. Ono, J.H.P., Sikansi, F., Corrêa, D.C., Paulovich, F.V., Paiva, A., Nonato, L.G.: Concentric RadViz: visual exploration of multi-task classification. In: 28th SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI), 2015, pp. 165–172. IEEE (2015)
27. Packer, J.F., Hasan, M., Samavati, F.F.: Illustrative multilevel focus+ context visualization along snaking paths. *Vis. Comput.* **33**(10), 1291–1306 (2017)
28. Pearson, K.: LIII. On lines and planes of closest fit to systems of points in space. *Lond. Edinb. Dublin Philos. Mag. J. Sci.* **2**(11), 559–572 (1901)
29. Pryke, A., Mostaghim, S., Nazemi, A.: Heatmap visualization of population based multi objective algorithms. In: International Conference on Evolutionary Multi-Criterion Optimization, pp. 361–375. Springer (2007)
30. Roweis, S.T., Saul, L.K.: Nonlinear dimensionality reduction by locally linear embedding. *Science* **290**(5500), 2323–2326 (2000)
31. Rubio-Sánchez, M., Raya, L., Diaz, F., Sanchez, A.: A comparative study between radviz and star coordinates. *IEEE Trans. Vis. Comput. Graph.* **22**(1), 619–628 (2016)
32. Russell, A., Daniels, K., Grinstein, G.: Voronoi diagram based dimensional anchor assessment for radial visualizations. In: 16th International Conference on Information Visualisation, 2012, pp. 229–233. IEEE (2012)
33. Russell, A., Marceau, R., Kamayou, F., Daniels, K., Grinstein, G.: Clustered data separation via barycentric radial visualization. In: Proceedings of the International Conference on Modeling, Simulation and Visualization Methods (MSV), p. 1. The Steering Committee of the World Congress in Computer Science, Computer Engineering and Applied Computing (WorldComp) (2014)
34. Sarkar, M., Brown, M.H.: Graphical fisheye views of graphs. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 83–91. ACM (1992)
35. Sharko, J., Grinstein, G., Marx, K.A.: Vectorized radviz and its application to multiple cluster datasets. *IEEE Trans. Vis. Comput. Graph.* **14**(6), 1427–1444 (2008)
36. Spence, R., Apperley, M.: Data base navigation: an office environment for the professional. *Behav. Inf. Technol.* **1**(1), 43–54 (1982)
37. Tenenbaum, J.B., De Silva, V., Langford, J.C.: A global geometric framework for nonlinear dimensionality reduction. *Science* **290**(5500), 2319–2323 (2000)
38. Wang, Y.C., Zhang, Q., Lin, F., Goh, C.K., Wang, X., Seah, H.S.: Histogram equalization and specification for high-dimensional data visualization using RadViz. In: Proceedings of the Computer Graphics International Conference, CGI '17, pp. 15:1–15:6. ACM (2017)
39. Zhou, F., Huang, W., Li, J., Huang, Y., Shi, Y., Zhao, Y.: Extending dimensions in RadViz based on mean shift. In: IEEE Pacific Visualization Symposium (PacificVis), 2015, pp. 111–115. IEEE (2015)



Yan Chao Wang is a Ph.D. student in the Rolls-Royce@NTU Corporate Lab, School of Computer Science and Engineering, Nanyang Technological University, Singapore from 2015. His current research interests include Visual Analytics, Information Visualization, Optical Metrology, and Optimization. He received his Bachelor and Master degree from University of Electronic Science and Technology of China (UESTC) and National University of Singapore (NUS), respectively.



Qian Zhang is a research engineer at Nanyang Technological University Singapore. Her research interests include visualization, computer graphics, and computer vision. Zhang has a B.S. in electronics and information engineering from Huazhong University of Science and Technology.



Feng Lin is an Associate Professor in the School of Computer Engineering, Nanyang Technological University, Singapore. And he is the Director of Bioinformatics Research Centre. His research interest includes biomedical informatics, bioimaging, computer graphics and visualization, and high performance computing. He is a Senior Member of IEEE.



Hock Soon Seah is a Professor and Director of the Multiplatform Game Innovation Centre (MAGIC) at the School of Computer Engineering, Nanyang Technological University. His current research areas are in geometric modelling, image sequence analysis with applications to digital film effects, automatic in-between frame generation from hand-drawn sketches, virtual and augmented reality, computer animation and game engine.



Chi Keong Goh currently holds the position of R&T Team Lead, Computational Engineering, at the Rolls-Royce Applied Technology Group (ATG) in Singapore. His research interest includes ensemble and transfer learning, extraction and incorporation of expert knowledge into data analytics, and multi-objective optimization. Chi-Keong has served as a reviewer for various international journals and on conference program committees in the fields of Computational Intelligence and Data Mining.