

Image-Based Aspect Ratio Selection

Yunhai Wang, Zeyu Wang, Chi-Wing Fu, Hansjörg Schmauder, Oliver Deussen and Daniel Weiskopf

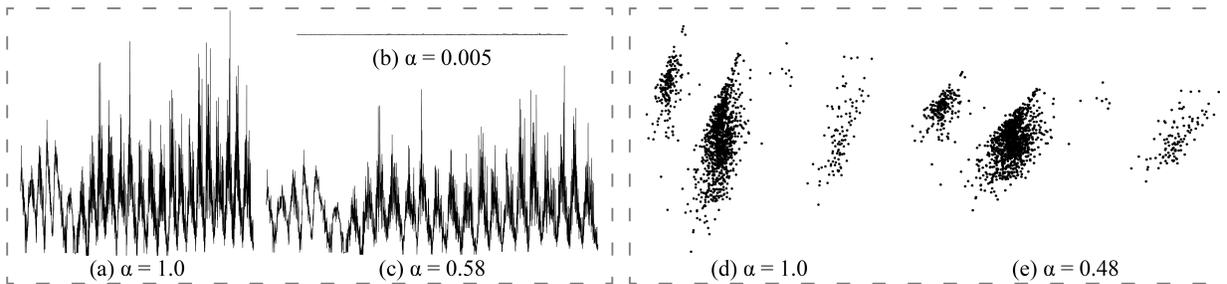


Fig. 1. Influence of aspect ratio α on the perception of trends and cluster separability for the *Sunspot* dataset [21] (a, b, c) and the *Contraceptive Method Choice (CMC)* dataset [10] (d, e). (a) Line chart with the default aspect ratio, and (b) the aspect ratio selected by an existing method (RV) [40], where both methods obscure the trends over the cycles. (c) The aspect ratio selected by our method (imgRV), where detailed cycle oscillations are revealed. (d) Scatter plot of three data clusters with the default aspect ratio, where the visual separation between the two clusters on the left is unclear. (e) The aspect ratio selected by our method shows clearer cluster structures.

Abstract—Selecting a good aspect ratio is crucial for effective 2D diagrams. There are several aspect ratio selection methods for function plots and line charts, but only few can handle general, discrete diagrams such as 2D scatter plots. However, these methods either lack a perceptual foundation or heavily rely on intermediate isoline representations, which depend on choosing the right isovalues and are time-consuming to compute. This paper introduces a general image-based approach for selecting aspect ratios for a wide variety of 2D diagrams, ranging from scatter plots and density function plots to line charts. Our approach is derived from Federer’s co-area formula and a line integral representation that enable us to directly construct image-based versions of existing selection methods using density fields. In contrast to previous methods, our approach bypasses isoline computation, so it is faster to compute, while following the perceptual foundation to select aspect ratios. Furthermore, this approach is complemented by an anisotropic kernel density estimation to construct density fields, allowing us to more faithfully characterize data patterns, such as the subgroups in scatterplots or dense regions in time series. We demonstrate the effectiveness of our approach by quantitatively comparing to previous methods and revisiting a prior user study. Finally, we present extensions for ROI banking, multi-scale banking, and the application to image data.

Index Terms—Aspect ratio, image-based method, Federer’s co-area formula, density field, anisotropic kernel density estimation.

1 INTRODUCTION

Visual attributes like size, shape, and slope greatly influence the expressiveness and effectiveness of a visualization [25]. In particular, the aspect ratio (defined as the fraction, height/width, throughout the paper) has a dramatic effect on the perception of data patterns in a visualization [6]. A well-chosen aspect ratio may reveal trends and relevant clusters that are hidden by a bad aspect ratio.

Figure 1 demonstrates the influence of the aspect ratio. An improper aspect ratio obscures the oscillations over the cycles (Figure 1 (a, b)) and leads to a poor visual cluster separation (Figure 1 (d)). In contrast, an appropriate aspect ratio attempts to reveal major patterns as much

as possible (Figure 1 (c, e)). Therefore, methods for automatically choosing a proper aspect ratio for a given visualization are of high interest for a broad range of visualizations and applications.

Selecting appropriate aspect ratios for 2D line charts has been well studied. Cleveland et al. [8] pioneered the principle of *banking to 45°* and proposed two methods: *average absolute orientation (AO)* and *arc length weighted average absolute orientation (AWO)*. Both methods select aspect ratios by banking the orientation of a line chart’s line segments at around 45 degrees. AWO generally produces reasonable aspect ratios for most data. Guha and Cleveland [29] and Talbot et al. [34] suggested two alternative approaches: the *resultant vector (RV)* and *arc length based (AL)* methods, based on geometric measures such as the resultant vector and the curve’s arc length, respectively. Recent work [40] showed that both methods tend to satisfy the banking to 45° principle and select almost the same aspect ratio as AWO, while RV is usually faster and more robust.

However, all aforementioned methods are specifically designed for line charts and cannot handle other common 2D visualizations such as scatter plots. To facilitate scatter plots, Fink et al. [16] selected aspect ratios based on a Delaunay triangulation [16]. Various geometric criteria were employed, e.g., large minimum angle and total edge length. However, due to the lack of a perceptual foundation, aspect ratios selected by this method might not be favoured by users for certain datasets. Moreover, it is relatively slow; as reported in the paper, it requires several minutes to compute the aspect ratio for a scatter plot with 1,000 points. Talbot et al. [34] proposed an isoline-based approach to selecting aspect ratios by applying existing methods to 2D isolines extracted from the density field derived from a scatter plot. This method

- Y. Wang and Z. Wang are with Shandong University. Email: {cloudseawang,zywangx}@gmail.com. Y. Wang and Z. Wang are joint first authors.
- C.-W. Fu is with the Chinese University of Hong Kong. E-mail: cwfu@cse.cuhk.edu.hk.
- O. Deussen is with Konstanz University, Germany and Shenzhen VisuCA Key Lab, SIAT, China. E-mail: oliver.deussen@uni-konstanz.de.
- H. Schmauder and D. Weiskopf are with VISUS, University of Stuttgart, Germany. E-mail: {hansjoerg.schmauder, Daniel.Weiskopf}@visus.uni-stuttgart.de.

Manuscript received xx xxx. 201x; accepted xx xxx. 201x. Date of Publication xx xxx. 201x; date of current version xx xxx. 201x. For information on obtaining reprints of this article, please send e-mail to: reprints@ieee.org. Digital Object Identifier: xx.xxx/TVCG.201x.xxxxxx

can find proper aspect ratios for most data; however, it depends on (i) the *number of isolines* extracted from the density field and (ii) the *quality of the density field* constructed from the data. Insufficient isolines or an improper kernel for estimating the density field [33] might result in undesirable aspect ratios that obscure patterns in the data.

In this paper, we present a generalized *image-based* approach for finding aspect ratios for 2D diagrams such as scatter plots, density function plots, and line charts. If the input data is not given in the form of a density field but as discrete geometric points, we construct such a field by kernel density estimation (KDE) with Gaussian kernels. In contrast to previous methods, our approach works *directly* with density fields defined over 2D visualizations without the need to generate intermediate data such as the isolines; see Figure 2. We formulate our approach based on *Federer's co-area formula* [14] in geometric measure theory to transform line integrals over sets of isolines of the density field to area integrals over the density field itself. Using our formulation, previous methods that can be rewritten as line integrals can be extended to work directly with density fields. In addition, to faithfully characterize data patterns, we introduce *anisotropic KDE* to construct the density fields, where the kernel associated with each data point adapts to the local structure around the point.

Compared to conventional isoline-based approaches, our method also follows the principle of banking to 45° , but it avoids discrete isolines as intermediate data and works directly with the density fields. Hence, it achieves an accurate and fast computation. Since density fields are often continuous and even smooth, directly computing the aspect ratios from them avoids the limitations of existing methods, e.g., they cannot handle data with spike noise [40] (see Figure 1 (b)). Moreover, our approach is not limited to using a density field of a discrete visualization but can be applied to any non-negative and unnormalized field. Hence, we refer to our approach as an *image-based approach*, and thus, it not only encompasses cases that can be handled by the previous methods, but can also be applied more broadly to continuous scatter plots and even to 2D images in general.

As shown by Wang et al. [40], AWO, AL, and RV can be formulated using line integrals; hence, they can be extended to work directly with density fields. Comparing our image-based versions of these methods with previous isoline-based versions with a large number of isolines, we recommend our method, image-based RV (imgRV), as the default aspect ratio selection method, since it produces results similar to most of the other methods but is the fastest and the most robust. We revisit examples and the user study by Fink et al. [16] and show that choices made by our method are mostly consistent with user preferences.

Last, we show three extensions of our method including region-of-interest (ROI) banking, multi-scale banking, and using images as an input. We summarize the main contributions of our paper below.

- We present a new aspect ratio selection approach that works directly with density fields constructed from a broad range of 2D visualizations. Three image-based aspect ratio selection methods are formulated by extending the line integral forms of existing methods AWO, AL, and RV.
- We introduce anisotropic kernels to better characterize the local structures in the data visualizations.
- We provide a comprehensive evaluation of a variety of methods, showing that the results of our image-based methods and isoline-based methods converge, but our methods are about an order of magnitude faster than existing methods.

2 RELATED WORK

Bertin, in his seminal work on the semiotics of graphics [2], already pointed out the importance of a proper aspect ratio for the perception and readability of diagrams. Hereafter, this influence was studied for decades, especially for widely-used visualization techniques such as line charts and scatter plots. A number of approaches for the automatic selection of aspect ratios were developed for these techniques.

Banking methods for line charts. Cleveland et al. [8] were the first to systematically study how the aspect ratio influences the perception

of line charts. Their user studies show that judging the slope ratio between two adjacent line segments is most accurate when the *orientation resolution* (range of orientations of the two involved line segments) is maximized. They also showed that maximizing the orientation resolution is equivalent to centering the absolute value of the mid-angle (average orientation) at 45° . This is known as the banking to 45° principle, which is the foundation of most aspect ratio selection methods. More recent work by Talbot et al. [35] found that the ability to estimate slopes is sub-optimal when the mid-angle is 45° . Nevertheless, most existing methods still focus on optimizing for an aspect ratio that banks lines in the visualization to 45° .

Based on the 45° principle, Cleveland et al. [5–7] and various other researchers developed a number of aspect ratio selection methods, including *median absolute slope* (MS), *average absolute slope* (AS), *average absolute orientation* (AO), and *arc length weighted average absolute orientation* (AWO). All these methods attempt to center the slopes (or orientations) of line segments around a value of one (or 45°). MS and AS bank the median and average absolute slope of line segments to one, whereas AO and AWO bank the average absolute orientations to 45° . AWO weights the average absolute orientation by the lengths of the line segments; doing so it produces more satisfactory results for most cases [7]. Heer and Agrawala [19] developed two methods to directly compute the orientation resolution: *global orientation resolution* (GOR) and *local orientation resolution* (LOR). Since GOR considers all pairs of line segments in a plot, it is usually slow, whereas LOR considers only successive pairs of line segments, so it is much faster, but it tends to select aspect ratios that are excessively large or small. By using an L_1 norm in LOR, L_1 -LOR [40] was found to produce more reasonable aspect ratios for most data.

Rather than following the 45° principle, Guha and Cleveland [29] suggested the *resultant vector* (RV) method, which banks the RV of a line plot to one. RV is the ratio of the total variation of line segments in y and x direction. Though its perceptual foundation is not clear, the method produces good results. Similar to RV, the *arc length based* (AL) [34] method is also not based on the 45° principle; it selects the aspect ratio by minimizing the arc length of a line chart. Talbot et al. [34] showed that RV can be interpreted as AL using the Manhattan distance metric, and thus, both methods share several empirical advantages such as parameterization invariance, robustness, and low computational costs. Recently, Wang et al. [40] showed parameterization invariance of AL, RV, and AWO using line integrals, and proved that RV and AL tend to satisfy the 45° principle. Through a systematic evaluation of most aspect ratio selection methods, they showed that RV and L_1 -LOR have complementary properties for revealing different data patterns of interest, and thus, they proposed a dual-scale banking method that takes advantages of RV and L_1 -LOR.

Banking methods for scatter plots. The aspect ratio selection methods above can only handle 1D functions presented as line charts. However, the 2D data may not be represented as a 1D function in general, e.g., 2D distributions of data samples, 2D scatter plots, as well as many other common forms of 2D visualizations.

For such visualizations, a number of banking methods have been proposed. From a perceptual point of view, banking ellipse-shaped clusters to circles seems to be a preferable solution [38], since ellipses have an orientation, which may interfere with the perception of the clusters [11, 24]. Cleveland et al. [6] fitted continuous LOESS curves (local polynomial regression) to the data of a scatter plot and banked the curves to find the aspect ratio. A drawback of these methods is that if the data does not have a clear correlation (or trend) [30], the LOESS curve will not be able to characterize the data, e.g., by separating clusters appropriately. Talbot et al. [34] converted a scatter plot into a density field by kernel density estimation, computed the isolines, and used the line segments as input to existing aspect ratio selection methods. They showed that AL and MS produce good aspect ratios for most data by banking ellipses to circles. Wang et al. [40] evaluated a variety of aspect ratio selection methods on 2D scatter plots and showed that AWO, RV, and AL usually select similar aspect ratios. Although their approach works well for most data, it may not produce reasonable results if there are insufficient isolines or the density field

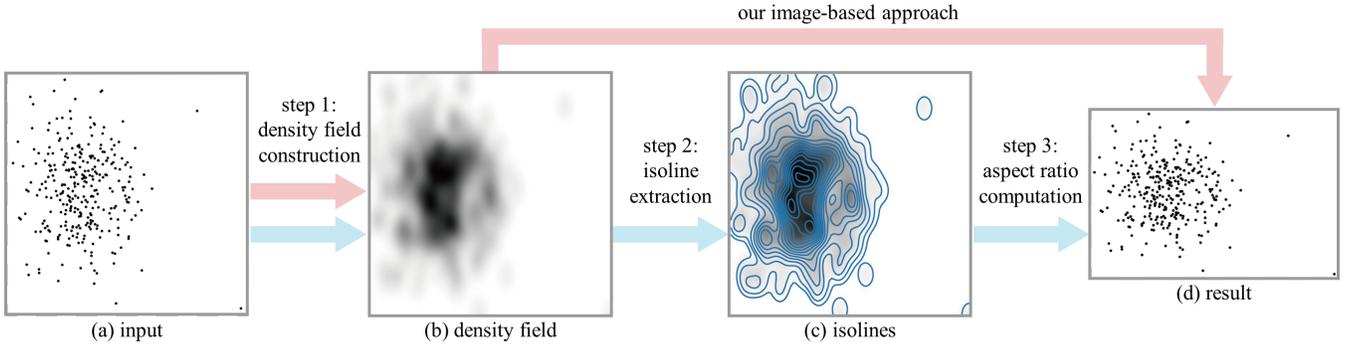


Fig. 2. Comparative overview of the isoline-based (blue arrows) and our image-based approach (red arrows). Our method bypasses the construction of isolines and directly computes the aspect ratio from the density field. Step 1 is obsolete, if the input data is already given as a density field.

does not well characterize the data. In this work, we avoid isolines in the computation, and formulate our method to work directly with the 2D density functions. Thus, we can bypass the isoline extraction, which is typically time-consuming and error-prone.

Fink et al. [16] proposed an alternative approach that works directly with data points in the scatter plot based on a Delaunay triangulation. Although the approach is able to produce reasonable aspect ratios for some datasets, it does not have any perceptual foundation, and thus, its results might not be consistent with the user preference. Moreover, computing and processing a Delaunay triangulation to determine the aspect ratio is an expensive process that takes several minutes for a scatter plot with only a thousand points [16]. As shown by the evaluation, our results are more favored by the users. Furthermore, our extended image-based RV can be solved in a closed form, which is faster than all previous methods, and the result is also close to the isoline-based approach with a large amount of isolines; see Figure 3.

The co-area formula. An important inspiration for our work is the co-area formula [14] from geometric measure theory. This formula expresses the integral over the level sets of a function as the integral of the function itself, where we use the term “level set” as interchangeable with isolines in 2D or isosurfaces in 3D. We are not the first to use this formula for data visualization. It has been successfully employed in other visualization contexts to generalize discrete methods by including density representations over a continuous domain. One example is the computation of histograms (frequency plots) for isovalue statistics of 3D scalar fields visualization [4, 12, 31]. Other examples include the continuous variants of scatter plots [1], parallel coordinates [20], function plots [22], and projected multidimensional attribute spaces [27]. One of our use cases is the aspect ratio optimization for continuous scatter plots. Here, the density representation for continuous scatter plots [1] can be used as the direct input, thus improving the perception of these 2D diagrams. This is complementary to, and can be combined with, other methods, such as feature highlighting in continuous scatter plots [23] and automatic selection between line charts and scatter plots for displaying time series [39].

3 IMAGE-BASED ASPECT RATIO SELECTION MODEL

In this section, we first review isoline-based aspect ratio selection approaches and identify their drawbacks. Then, we derive our image-based approach using the co-area formula [28] to enable us to work directly with the density fields, and finally show how AWO, RV, and AL can be extended through our image-based approach.

3.1 Isoline-Based Aspect Ratio Selection Approaches

The core assumption of existing isoline-based approaches is that the input data plot can be represented by finite isolines extracted from the density field of the plot. Then, we can apply conventional methods to select aspect ratio (denoted as α) based on the extracted isolines. Considering that the input data plot is scaled to a square domain denoted as Ω without loss of generality, and $\vec{X} = \{\vec{x}_1, \dots, \vec{x}_n\}$ is a set of n sample points in the given data plot, the isoline-based approaches have the

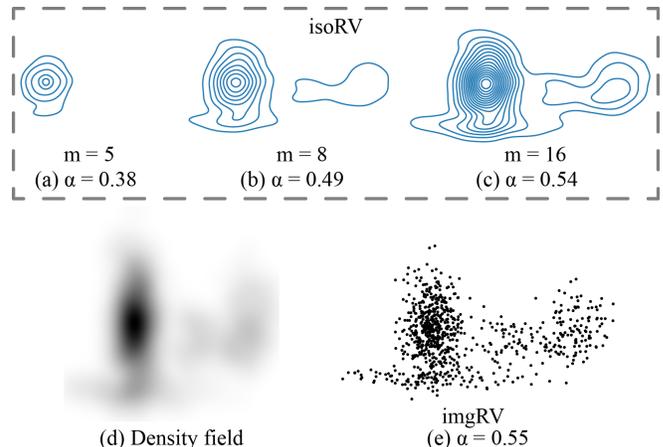


Fig. 3. (a-c) Results of isoRV with increasing m (number of isovalue samples), and equivalently, with increasing number of isolines. (d) Density field input. (e) Result of our imgRV.

following three major steps for computing the aspect ratios (see also the running example shown in Figure 2):

(i) Density field construction. This step is obsolete if the input is already given as a density field (e.g., from continuous scatter plots [1] or from an image). In the case of scatter plots, a density field is constructed from the sample points:

$$\Phi : \vec{x}_i \mapsto \rho(\vec{x}_i), \text{ where } \vec{x}_i \in \Omega \subset \mathbb{R}^2 \text{ and } \rho \in \mathbb{R}. \quad (1)$$

Since the solution of α is unknown at the moment, the density field is simply constructed in Ω with a default aspect ratio $\alpha = 1$. The approach constructs the density field based on the prominence of \vec{x}_i in Ω using Kernel Density Estimation (KDE):

$$\rho(\vec{x}) = \frac{1}{n} \sum_{i=1}^n K_{\vec{H}}(\vec{x} - \vec{x}_i), \text{ for } \vec{x} \in \Omega, \quad (2)$$

where $K_{\vec{H}}(\vec{x})$ is a kernel function with bandwidth matrix \vec{H} . Typical choices for the kernel function are Gaussian and Epanechnikov functions, and we use the Gaussian function here. Typically, the density field is sampled on a uniform grid (image) over the domain of Ω .

(ii) Isoline extraction. Once the density field (ρ) is available, the approach finds the density value range, normalizes ρ to $[0, 1]$, uniformly samples m isovalues (denoted as $t_i, i = 1, \dots, m$) in $[0, 1]$, and extracts a set of isolines (denoted as $L(t_i)$) for each isovalue t_i .

(iii) Aspect ratio computation. In the last step, the approach gathers the line segments of the isolines, and applies an existing aspect ratio selection method (e.g., AL or RV) to determine the aspect ratio.

If AL is used, we denote the overall approach as *isoAL*; if RV is used, we denote it as *isoRV* for the rest of the text. Since some of the methods for selecting the aspect ratio are parametrization-dependent, we need a dense coverage by the isolines to obtain an accurate α .

Figure 3 presents results with a simple data set to illustrate the effect of m (or the number of isolines) on aspect ratio selection. We extract different numbers of isolines (a)–(c) from the same density field (d) using varying m , and then compute the aspect ratios from the extracted isolines using *isoRV*. The results shown in (a)–(c) show that a coarse sampling ($m = 5$) might not sufficiently characterize all structures in the density field, whereas a finer sampling ($m = 16$) would produce many isolines, even for such simple data. Our image-based approach (e) bypasses the isoline extraction and finds the aspect ratio directly from the density field. It is equivalent to considering isolines of all isovalues in the density field domain, thus our aspect ratio result (e) is close to *isoRV*'s result when m is large (c).

Therefore, we can see that the isoline-based approach heavily relies on two factors: (i) a sufficient number of isolines (usually densely sampled) and (ii) a density field that faithfully represents the data. In contrast, our approach bypasses the first aspect and introduces an anisotropic kernel to address the second aspect; by these means, it is able to generalize the aspect ratio selection for general 2D diagrams.

3.2 2D Integral Representation

To characterize all patterns in a 2D visualization, we may increase m to sample more isovalues and extract more isolines (see Figure 3). In theory, when m tends to infinity, we would have considered isolines of all isovalues in the density range. However, this would produce a large amount of isolines, so having a large m will be too costly for the computation. As suggested by Marsden and Tromba [26], *an infinite sum of all isolines is equivalent to the integral over the associated density field*. Their work motivated us to derive a new formulation that works directly with the density field instead of the isolines.

In the following, we denote Ω_α as the domain with aspect ratio α , where Ω_α is obtained by stretching or compressing the square domain Ω (or $\Omega_{1,0}$) in y direction. Moreover, we denote ρ_α and $L_\alpha(t_i)$ as an α -stretched density field and isoline set $L(t_i)$, respectively, in Ω_α .

Now, let us consider a single isoline set $L_\alpha(t_i)$ that corresponds to isovalue t_i . As suggested by Wang et al. [40], several existing aspect ratio selection methods can be formulated as an optimization of a line integral, say f , which depends on the input data \vec{X} and aspect ratio α . Hence, we can integrate over the line segments (or arc length elements ds) along $L_\alpha(t_i)$, and formulate the following objective:

$$\min_{\alpha \in (0, \infty)} \int_{L_\alpha(t_i)} f(\vec{X}, \alpha) ds. \quad (3)$$

Next, since an infinite sum of all isolines is equivalent to integrating the density field, we can consider all possible isolines sampled over all isovalues in the density range through the following integration:

$$\min_{\alpha \in (0, \infty)} \int_0^1 \left(\int_{L_\alpha(t)} f(\vec{X}, \alpha) ds \right) dt. \quad (4)$$

where we use t instead of t_i due to the integration; note that the density range has been normalized to $[0, 1]$, therefore we have $t \in [0, 1]$.

Since a numerical integration over a large amount of curves is computationally very expensive, we apply the co-area formula [15] to the problem and convert the integrals over the range of isovalues to integrals over the domain in which the isolines are defined. In this way, Eq. (4) can be re-formulated as

$$\int_0^1 \left(\int_{L_\alpha(t)} f(\mathbf{X}, \alpha) ds \right) dt = \int_{\Omega_\alpha} f(\mathbf{X}, \alpha) \|\nabla \rho_\alpha(\vec{x})\| d^2\vec{x}, \quad (5)$$

where $f(\mathbf{X}, \alpha)$ is required to be Lipschitz continuous (i.e., f should have bounded gradient) and $\nabla \rho_\alpha(\vec{x})$ is the gradient field of the density field ρ_α . In practice, we work with a sampled version of the density field where the gradient always exists and is bounded. Note also that

the infinitesimal distance element $d\vec{x}$ is in Ω_α , while dx and dy are in $\Omega_{1,0}$; since we scale $\Omega_{1,0}$ in y direction (by α) to obtain Ω_α , we thus have the following relationship between the two domains:

$$\int_{\Omega_\alpha} d^2\vec{x} = \alpha \int_{\Omega_{1,0}} dx dy. \quad (6)$$

Furthermore, the gradient magnitude term $\|\nabla \rho_\alpha(\vec{x})\|$ in Eq. (5) accounts for the uneven sampling of isolines over domain Ω_α . By applying the chain rule, we have

$$\nabla \rho_\alpha(\vec{x}) = \left(\frac{\partial \rho}{\partial x}, \frac{1}{\alpha} \frac{\partial \rho}{\partial y} \right), \quad (7)$$

where $\partial \rho / \partial x$ and $\partial \rho / \partial y$ refer to the gradient of the density field ρ in x and y direction, defined over the domain $\Omega_{1,0}$.

Substituting Eqs. (6) and (7) into Eq. (5), we obtain

$$\int_0^1 \left(\int_{L_\alpha(t)} f(\mathbf{X}, \alpha) ds \right) dt = \int_{\Omega_{1,0}} f(\mathbf{X}, \alpha) \left\| \left(\alpha \frac{\partial \rho}{\partial x}, \frac{\partial \rho}{\partial y} \right) \right\| dx dy, \quad (8)$$

which provides the foundation for extending the existing aspect ratio selection methods to directly process density fields while avoiding isolines. Note that this 2D integral form does not require ρ to be a probability density function; it can be any non-negative and unnormalized field. Hence, we refer our approach as an *image-based approach*, since it can be applied to any non-negative 2D field. It should be noted that this co-area representation can be combined with feature detection approaches [23, 41] so that the selected aspect ratio can highlight features of interest, which is left for future work.

3.3 Image-Based Extensions

Using the formulation we derived in Eq. (8) and the line integral forms from Wang et al. [40], we are now ready to create our image-based versions of several aspect ratio selection methods:

ImgAL. The arc length based (AL) method [34] selects α by minimizing the total arc length of line segments in the plot, while preserving the area under the plot. Since the arc length is inherently a variable of the line integral, the objective of AL can be reformulated (from [40]) as the following line integral:

$$\min_{\alpha \in (0, \infty)} \sum_{i=1}^n \frac{1}{\sqrt{\alpha}} \|\Delta x_i, \alpha \Delta y_i\| = \min_{\alpha \in (0, \infty)} \int_C \frac{1}{\sqrt{\alpha}} ds, \quad (9)$$

where C is the set of line segments in the plot. Comparing Eq. (9) with Eq. (3), we can see that $f(\mathbf{X}, \alpha)$ of AL is simply $1/\sqrt{\alpha}$. Hence, by substituting $f = 1/\sqrt{\alpha}$ into Eq. (8), we obtain the objective of the image-based version of AL (denoted as *imgAL*):

$$\min_{\alpha \in (0, \infty)} \int_{\Omega_{1,0}} \left\| \left(\sqrt{\alpha} \frac{\partial \rho}{\partial x}, \frac{1}{\sqrt{\alpha}} \frac{\partial \rho}{\partial y} \right) \right\| dx dy, \quad (10)$$

where the gradient ($\nabla \rho$) at each point in $\Omega_{1,0}$ is multiplied with the following area-preserving squeeze mapping:

$$\mathbf{S}_\alpha := \begin{pmatrix} \sqrt{\alpha} & 0 \\ 0 & 1/\sqrt{\alpha} \end{pmatrix}.$$

ImgAWO. AWO [5] selects an aspect ratio that banks the weighted average absolute orientations to 45° :

$$\min_{\alpha \in (0, \infty)} \left| \frac{\sum_i |\theta_i(\alpha)| l_i(\alpha)}{\sum_i l_i(\alpha)} - \frac{\pi}{4} \right|,$$

where $\theta_i(\alpha)$ is the absolute orientation and $l_i(\alpha)$ the length of each line segment. Hence, AWO aims at finding a proper α , such that the

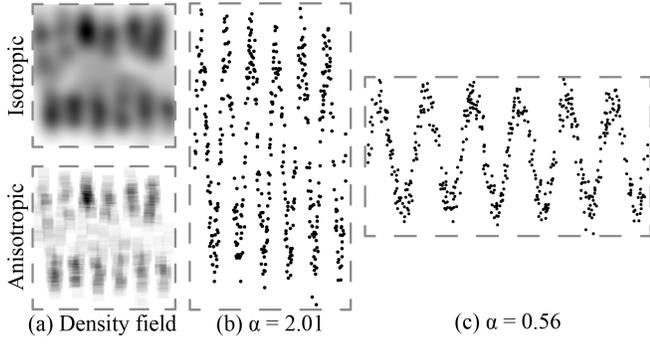


Fig. 4. (a) Isotropic (top) vs. anisotropic (bottom) density fields. (b, c) imgRV results produced from isotropic and anisotropic density fields, resp.

mean density of the plotted curve is close to $\pi/4$. Like AL, AWO can be formulated using line integral terms:

$$\min_{\alpha \in (0, \infty)} \left| \frac{\int_C |\theta(\alpha)| ds}{\int_C ds} - \frac{\pi}{4} \right|. \quad (11)$$

However, since Eq. (11) is not a single line integral as AL, we cannot directly apply Eq. (8) to Eq. (11). Hence, we consider C in each line integral term as sets of isolines over all isovalues and apply Eq. (8) to transform each line integral term in Eq. (11). By further expressing θ in the formulation using gradients $\frac{\partial \rho}{\partial x}$ and $\frac{\partial \rho}{\partial y}$ in $\Omega_{1,0}$, we can obtain the following objective of imgAWO:

$$\min_{\alpha \in (0, \infty)} \left| \frac{\int_{\Omega_{1,0}} \tan^{-1} \left(\alpha \frac{\partial \rho}{\partial x} / \frac{\partial \rho}{\partial y} \right) \left\| \left(\alpha \frac{\partial \rho}{\partial x}, \frac{\partial \rho}{\partial y} \right) \right\| dx dy}{\int_{\Omega_{1,0}} \left\| \left(\alpha \frac{\partial \rho}{\partial x}, \frac{\partial \rho}{\partial y} \right) \right\| dx dy} - \frac{\pi}{4} \right|, \quad (12)$$

which is subject to the same optimization process as AL. For the derivation details, please refer to the supplemental materials.

imgRV. The *resultant vector* (RV) method [29] takes the ratio of the total variation of line segments in x and y direction as the aspect ratio. Hence, we can write it in the following line integral form:

$$\alpha = \frac{\sum_i |\Delta x_i|}{\sum_i |\Delta y_i|} = \frac{\int_C |\cos \theta| ds}{\int_C |\sin \theta| ds}. \quad (13)$$

Similar to the derivation for imgAWO, we consider C as sets of isolines over all isovalues and apply Eq. (8) to each line integral term. Then, we replace $\cos \theta$ and $\sin \theta$ by using the gradients $\frac{\partial \rho}{\partial x}$ and $\frac{\partial \rho}{\partial y}$ of the density field in $\Omega_{1,0}$ and obtain the following objective of imgRV:

$$\alpha = \frac{\int_{\Omega_{1,0}} \left\| \frac{\partial \rho}{\partial y} \right\| dx dy}{\int_{\Omega_{1,0}} \left\| \frac{\partial \rho}{\partial x} \right\| dx dy}, \quad (14)$$

which can be regarded as the ratio between the total variations in x and y direction over the whole density field. Again, please refer to Appendix A for the derivation details. Since imgRV has a closed form, it is faster to compute than imgAWO and imgAL, for which we have to solve a minimization problem; see Section 5 for more details.

4 ANISOTROPIC KERNEL DENSITY ESTIMATION

A particularly interesting aspect of our method is that the kernels for constructing the density field are *selected in relation to the data*. The KDE associated with such kernels is referred to as anisotropic KDE.

Isotropic KDE. Before introducing anisotropic KDE, we first briefly review isotropic KDE. Given a 2D domain, a Gaussian kernel function models the density in a local region defined by

$$K(\vec{x}) = \frac{1}{2\pi\sqrt{|\vec{H}|}} \exp\left(-\frac{1}{2}\vec{x}^T \vec{H}^{-1} \vec{x}\right), \quad (15)$$

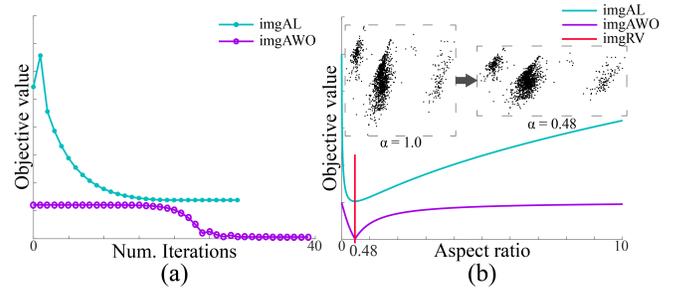


Fig. 5. (a) Convergence of the conjugate gradient optimization of imgAL and imgAWO. Shown are the objective values against the number of iterations, where imgAL reaches a good solution in less than 30 iterations, whereas AWO requires 39 iterations. (b) Objective values plotted over the aspect ratio, while the red line marks the aspect ratio selected by imgRV. In this example (CMC dataset [10]), all three methods find a similar aspect ratio of 0.48, showing three clusters in circular shapes.

where position \vec{x} is expressed as a column vector. The bandwidth matrix \vec{H} scales and rotates the region. Thus, the choice of \vec{H} affects the precision of the density field in characterizing the data patterns. Typically, KDE adopts Silverman's rule of thumb [33] to determine a diagonal matrix for the bandwidth matrix \vec{H} with two elements

$$\vec{H} = \begin{pmatrix} h_x & 0 \\ 0 & h_y \end{pmatrix}.$$

For example, h_x can be computed by

$$h_x = 1.06 \times \min\{s_x, \frac{R}{1.34}\} n^{-1/5}, \quad (16)$$

where s_x is the standard deviation of the input data points \mathbf{X} in x direction and $R = x_{[0.75n]} - x_{[0.25n]}$ is the difference between the 25% and 75% quantiles of $\{x_i\}$. The value of h_y is computed similarly.

Such Gaussian kernels result in isotropic, axis-aligned elliptical density regions of the same size around each point, so the resulting density field may not represent the data patterns well (see Figure 4 (a)).

Anisotropic KDE. To better characterize the data patterns by the density fields, we introduce anisotropic Gaussian kernels. These kernels are selected in a data-aware manner, such that each kernel creates an elliptical density region aligned to the main orientation of the data around each data point. For each data point \vec{x}_i , we perform a principal component analysis (PCA) on its k nearest neighbors and obtain the corresponding eigenvectors $\{v_i^1, v_i^2\}$ and the associated principal directions. Since the principal directions maximize data variance, we use them to adjust the shape of the Gaussian function:

$$\vec{H}_i = \vec{H} [v_i^1; v_i^2], \quad (17)$$

where \vec{H}_i is the bandwidth matrix of the i -th point. Thus, each point has its own \vec{H}_i and its associated elliptical region with an orientation that adapts to the local data pattern. Figure 4(a) shows the density fields generated by isotropic KDE and anisotropic KDE, where we can see that the local orientation is clearly characterized. Accordingly, the plot shown in Figure 4(d) with the aspect ratio selected by using the anisotropic density field clearly reveal the sin function, while the one in Figure 4(c) is not able to reveal the pattern.

5 NUMERICAL IMPLEMENTATION

Compared to isoline-based approaches, our image-based approach consists of only two steps: (i) density field construction and (ii) aspect ratio computation. To accelerate the k nearest neighbor search for constructing the anisotropic kernel function, we build a k -d tree over the input data points. In addition, since imgAL, imgAWO, and imgRV all need the gradient information, we sample a 2D uniform grid at a resolution

of $w_g \times h_g$ over the density field ρ and perform the Sobel operator to compute the gradient field $\nabla\rho$, where w_g and h_g are empirically set as 1000. Doing so, we are able to accelerate the computation of the integrals in Eqs. (10), (12), and (14), which are implemented as a Riemann sum over all grid points in the domain.

To calculate the optimal aspect ratio for imgAL and imgAWO, we numerically find the value that minimizes the objective functions from Eqs. (10) and (12) by using the conjugate gradient method [17]. Like Talbot et al. [34], we parametrize the 1D search problem with $\log(\alpha)$, this lets the optimization converge in less than 30 iterations. The optimal aspect ratio for imgRV can also be easily obtained by directly using Eq. (14). Figure 5 presents an exemplary result in which all three methods find very similar optimal aspect ratios; a more comprehensive evaluation will be presented in Section 6.1.

We implemented our approach in R (see code in supplemental materials) and ran it on a machine with an Intel[®] Core[™]i5-4200H with 2.8 GHz dual-core CPU in double precision. Our experiments on 100 datasets showed that imgRV mostly took around 50 to 300 ms, thus allowing for interactive manipulation; see Section 6.1 for details. The performance could be further boosted up by exploiting parallel computation on the GPU.

6 EVALUATION

We performed four experiments to evaluate our image-based methods. First, we quantitatively compared our methods with isoline-based methods (Section 6.1) for two purposes: (i) learn to see if our methods produce similar aspect ratios as isoline-based methods but in less time, and (ii) find out, which one of our three methods is fastest and most robust. Second, we studied the convergence of our image-based methods with isoline-based methods and the corresponding methods designed for line charts (Section 6.2). Next, we presented a parameter analysis to find proper parameters for the best method recommended by the first experiment (Section 6.3). Lastly, we configured the best method with proper parameters, and revisited the user study by Fink et al. [16] to show that choices made by our method are mostly consistent with the user preferences (Section 6.4).

6.1 Quantitative Comparison with Isoline-based Methods

Comparing extensively with various methods, Wang et al. [40] showed that the RV method not only produces similar results to classical methods such as AWO and AL, but also is the fastest and most robust. On the other side, Talbot et al. [34] extended various aspect ratio selection methods designed for line charts (including RV) into isoline-based methods to work for scatter plots. Therefore, we selected the isoline-based RV method (isoRV) as a representative of isoline-based methods, but extracted a large number of isolines for isoRV to capture most of the structures in the data. To be specific, we computed a 1000×1000 density field for each dataset. In addition, we took out the effect of anisotropic KDE, and used isotropic KDE for both isoRV and our image-based methods. We only summarize the comparison results with isoRV in the following, detailed comparison results with most methods are presented in the supplementary material.

Datasets. For a comprehensive comparison, we gathered 100 scatter plots with substantial variations in terms of the number of data points, ranging from 200 to 5000. Among them, 71 datasets show the dimensionality reduction results provided by Sedlmair et al. [32] and the rest are real bivariate datasets collected from the UCI repository [10].

Measure. To find out if our methods produce similar aspect ratios as isoRV, we measured the relative deviation of the aspect ratios selected by our methods from the ones selected by isoRV:

$$\text{deviation} = \frac{\alpha_{\text{img}X} - \alpha_{\text{isoRV}}}{\alpha_{\text{isoRV}}} * 100\%, \quad (18)$$

where $\alpha_{\text{img}X}$ refers to α selected by imgRV, imgAL, or imgAWO.

Results. To summarize the deviation results for the 100 datasets, we created the boxplots shown in Figure 6(a). These boxplots show that the deviations of aspect ratios selected by our methods are very small,

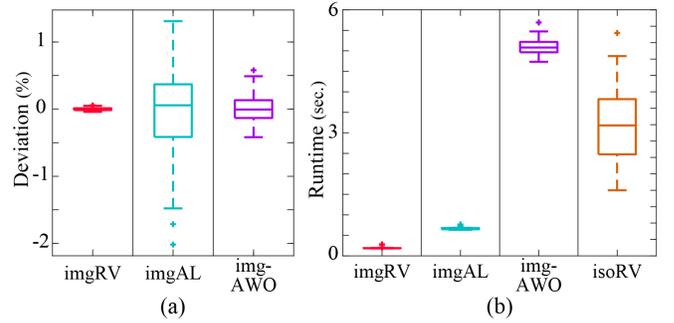


Fig. 6. (a) Deviation of aspect ratios selected by our image-based methods (imgRV, imgAL, and imgAWO) from those selected by isoRV; (b) runtime of our methods and isoRV for aspect ratio selection.

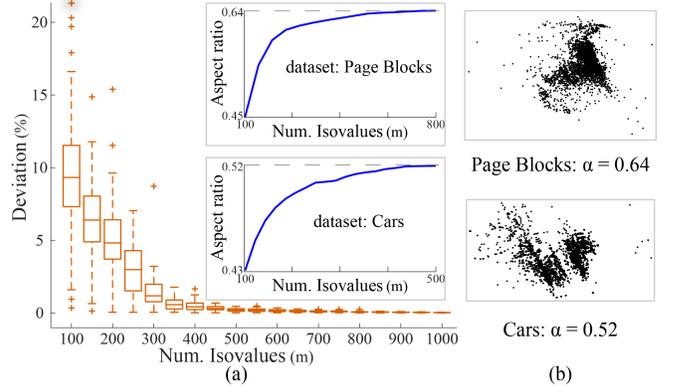


Fig. 7. (a) Boxplots showing the convergence of imgRV and isoRV over 100 datasets with increasing number of isovalues (m). On top left, we show the convergence plots for two typical datasets, *Page Blocks* [10] and *Cars* [36], indicating that when m reaches over 500, imgRV and isoRV select very similar aspect ratios, i.e., very low deviations. (b) Resulting dataset plots for the aspect ratios chosen by imgRV and isoRV. In (a) the result for 100 isovalues is out of the plot range, we indicated it by a dark transparent shadow.

ranging only from -2% to 1.5% ; particularly, the range for imgRV is typically very small, revealing that the method typically selects the same aspect ratio as isoRV. Note also that we aim for precision in this part of the experiment, so we considered a large amount of isolines for isoRV with $m = 1000$.

Figure 6(b) summarizes and compares the time performance of all four methods. In this part of the experiment, we used $m = 500$ to extract isolines for isoRV, since this is sufficient for isoRV to handle most datasets (see Section 6.2), and isoRV can perform faster. As seen from Figure 6(b), imgRV is the fastest, it took less than 0.37 seconds consistently for all datasets; imgAL is the second, but it still took around 0.7 seconds, which is roughly two times slower than imgRV. In contrast, the other two methods, isoRV and imgAWO, required around 3.2 and 5.2 seconds. Therefore, we consider imgRV to be able to support interactive aspect ratio selection for typical datasets, which is about an order of magnitude faster than isoline-based methods.

Recommendation. After carefully inspecting the comparison results, we come up with the following two observations: (i) all three image-based methods select similar aspect ratios as isoRV, where the maximal absolute deviations of the three methods are less than 2% ; and (ii) imgRV produces most similar results to isoRV, while it is the fastest among the four methods being compared. Therefore, we recommend imgRV as the representative method for our image-based approach.

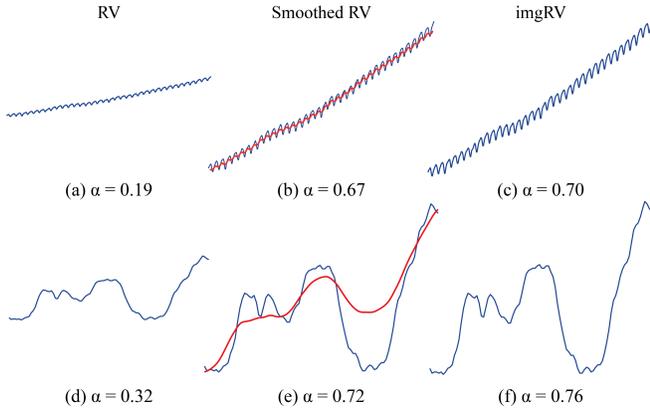


Fig. 8. Illustrating the relationship between RV and imgRV using two time-series datasets: CO_2 [6] (top) and $Computer$ [9] (bottom). (a,d) Line charts and aspect ratios selected by directly applying RV to the curves; (b,e) Smoothed line charts and aspect ratios selected by applying RV to the smoothed curves shown in red; (c,f) Line charts and aspect ratios selected by directly applying imgRV to the curves.

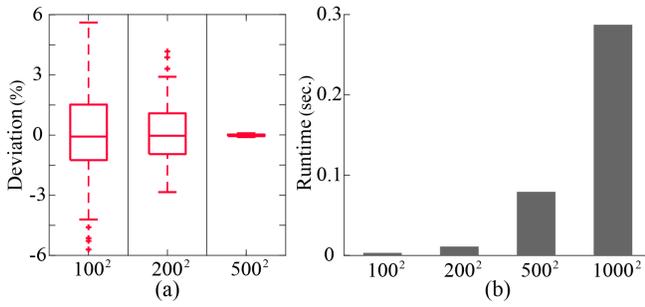


Fig. 9. (a) Deviation (in %) and computing time (b) for creating aspect ratios for 100 datasets in domains of various resolutions: 100×100 , 200×200 , 500×500 , and 1000×1000 .

6.2 Convergence Study

We refer the isoline-based methods and the original methods designed for line charts as geometric methods, since both work with line segments. In the second experiment, we compare their convergence with our image-based methods. Since the previous experiment showed that imgRV produces almost the same results as isoRV, we take imgRV as the representative of image-based methods. Furthermore, we take isoRV as the representative of isoline-based methods, and RV as the representative of the original methods designed for line charts.

imgRV vs. isoRV. Unlike our imgRV, the quality of aspect ratio selected by isoRV strongly depends on the amount of isolines. Hence, we tried different m over the 100 datasets in the previous experiment when using isoRV to select aspect ratios, and computed the deviations in aspect ratios using Eq. (18). Figure 7 (a) summarizes the results. From the boxplots, we can see that when isoRV has a sufficiently large amount of isolines (i.e., large m) as inputs, the aspect ratios selected by imgRV and isoRV converge. However, isoRV typically requires at least $m=500$ for most datasets. Thanks to the image-based formulation, our approach (imgRV) can bypass the isoline construction and directly compute the aspect ratios from the density fields.

imgRV vs. RV. By converting line charts to density fields, our methods can also be used to find aspect ratios for line charts. Since density fields are smooth representations of the original line charts, our imgRV may not produce similar aspect ratios as RV, which works on the unsmoothed polygonal data. Hence, if we properly smooth the data, imgRV and RV might produce similar aspect ratios. Figures 8 (a, b, c) show results produced with RV, smoothed RV, and imgRV using the CO_2 dataset [6], where a similar upward trend is clearly observable in Figures 8 (b, c).

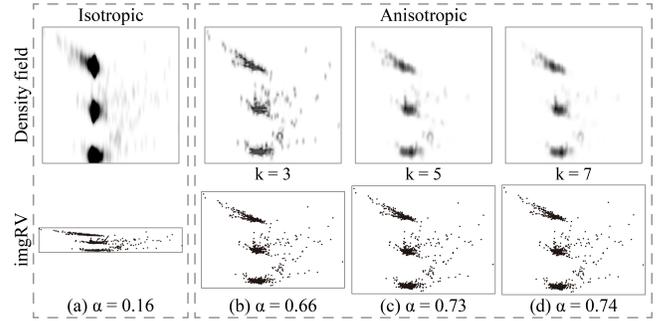


Fig. 10. Influence of the number of nearest neighbors (k) on the constructed anisotropic density field and the aspect ratios selected by imgRV on the $eFashion$ dataset [13]. (a) Density field constructed by isotropic KDE and its resulting plot; (b, c, d) Density fields constructed by anisotropic KDE with different k and the resulting plots.

Figures 8 (d, e, f) show the $Computer$ dataset [9], which further confirms our observation that applying RV to smoothed data produces aspect ratios similar to the ones selected by imgRV.

6.3 Parameter Analysis

In the third experiment, we studied how to select the resolution of density fields and number of nearest neighbors for our image-based methods (typically, imgRV):

(i) Resolution of density fields. Patterns in data are generally better characterized by density fields of higher resolution, but this increases the required computing time. Hence, we performed an experiment to find the proper resolution for revealing all necessary aspects of the data while consuming less computing time. We considered four different resolutions: 100×100 , 200×200 , 500×500 , and 1000×1000 , and applied imgRV to compute the aspect ratio for each of the 100 datasets described in Section 6.1. To study the influence of the resolution, we again used the deviation measure defined in Section 6.1 and took the aspect ratios computed with the highest resolution as the references.

Figure 9 (a) summarizes the results for the 100 datasets. It shows that the deviations for a resolution of 500×500 are in a small range of $[-0.2\%, 0.15\%]$, thus the aspect ratios selected with domain resolutions 500×500 are similar to those selected with 1000×1000 . Figure 9 (b) summarizes the time performance of imgRV for different resolutions; it shows that the computational time increases non-linearly with the domain resolution. Considering these two results, we recommend 500×500 as the default resolution for our density fields.

(ii) Number of nearest neighbors (k). Next, we attempted to empirically find a proper value of k for constructing the anisotropic density fields; see Figures 4 and 10 for results on two typical datasets. From the analysis, we found that similar k tends to produce similar aspect ratios, but if k is too large, it might lead to obscuration of interesting patterns; see Figure 10 (d). Therefore, we set the default k as five and use this value to produce all the results shown in the paper.

Initially, we attempted to perform an analysis with the deviation measure to explore imgRV and isoRV for different k , similar to the previous experiments. However, k affects the density fields and so are imgRV and isoRV, so both methods produce similar aspect ratios for varying k . However, since imgRV allows interactive exploration, users may adjust k to explore the plot like multi-scale banking. Indeed, k is data-dependent; in the future, we plan to investigate ways to automatically decide proper values of k in terms of the data characteristics.

6.4 Revisiting the Study of Fink's Method

In the last experiment, we revisited the study in [16] to show that our method yields results that are quite close to the ones produced by the best two criteria in [16], but also consistent with the user preferences. Fink et al. [16] studied various geometric criteria to select aspect ratios, and found that the criteria of minimizing the total length and minimizing the mean uncompactness yielded the best results, so we compared

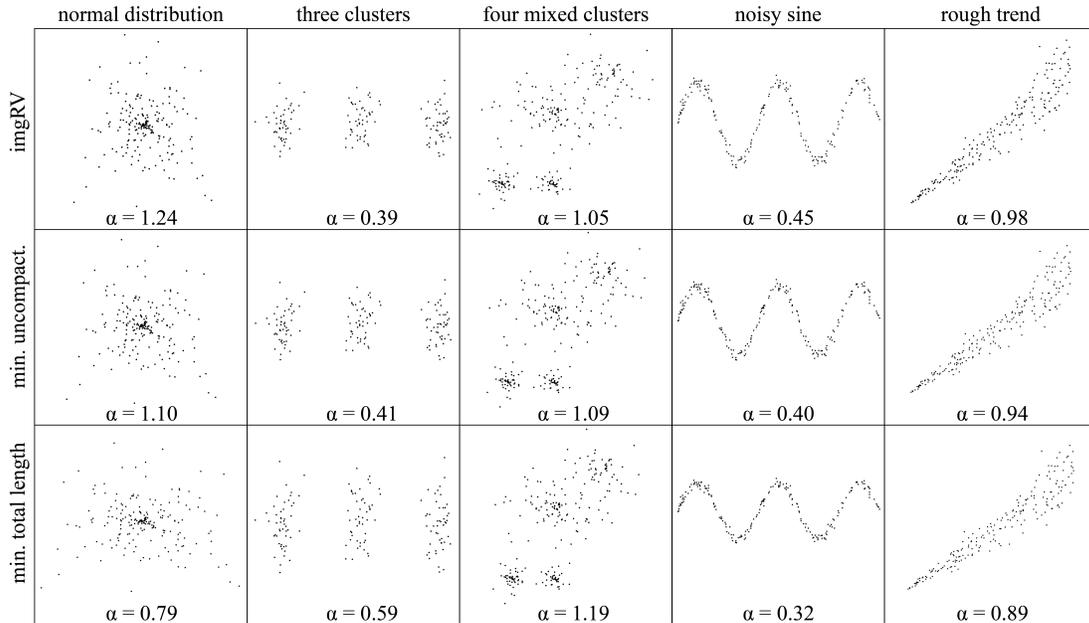


Fig. 11. Comparison between imgRV and two methods of Fink et al. [16]: (i) minimizing uncompactness, and (ii) minimizing total length, over the five synthesized datasets. Comparing the result, we can see that our results are close to the ones produced by minimizing uncompactness.

our method mainly with these two criteria. Since neither the implementation nor the user study result of [16] is available, we applied our method only to the available data shown in the paper [16].

Five test datasets from [16]. The effectiveness of [16] was demonstrated using five synthesized datasets, among which three had varying number of clusters and two had different data trends. Note that we obtained the datasets from the authors of [16]. By applying imgRV to the datasets, we obtained the results shown in Figure 11, showing that our results are quite close to the ones generated by the “minimizing the mean uncompactness” criteria, where all clusters and trends are clearly visible. In this sense, the results from our method and [16] are comparable. We compared the performance by running our method on the similar machine. For the set with four mixed clusters, their method required around 5 sec. while our method only took 0.2 sec.

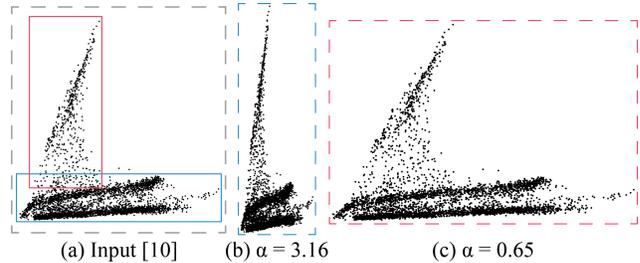


Fig. 13. ROI banking on the *Satimage* dataset [10]. Our method can take user-specified ROIs (red and blue boxes in (a)) as inputs and select aspect ratios (b, c) that focus on revealing patterns inside the ROIs.

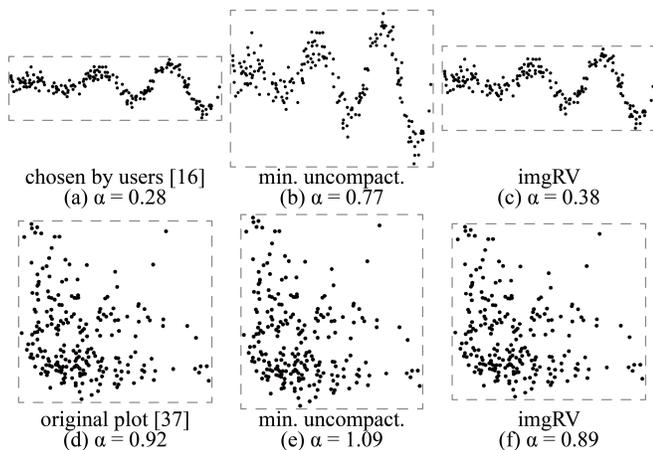


Fig. 12. Comparing the aspect ratios selected by our method (c, f) and the method of Fink et al. [16] (b, e) to user preferences (a, d): function $y = x \sin x$ used in [16] (top) and a data plot extracted from [37] (bottom).

User selection. Next, we further revisited the user study in [16] to learn how well our selected aspect ratios match the user preferences. However, the user study result in [16] is unavailable, so we took only the user study result in the paper [16] for comparison; see the top row

in Figure 12, which has clearly shown that our result is closer to the user preferred one.

In [16], a data plot shown in a *Nature* article [37] was taken as an input to see whether the method [16] produced a similar aspect ratio. Assuming that the authors of [37] have carefully adjusted the aspect ratio for producing the plot, we took the same plot as input and applied our method to select an aspect ratio, which was also found to be closer to the original aspect ratio; see the bottom row in Figure 12. In the future, we plan to perform a large-scale user study to confirm the findings.

7 EXTENSIONS

In this section, we present three extensions that utilize the capability of our image-based methods for a proper aspect ratio selection.

Region-of-interest (ROI) banking. Our image-based approach operates directly on the density field, so it inherently allows the user to select an ROI and then banks the diagram for the data patterns in the ROI. Figure 13 shows an example dataset that has two major patterns: an upward trend with sparse data points on top and a dense cluster at the bottom. By specifying an ROI (red and blue boxes shown in Figure 13 (a)), a user is able to highlight patterns in the ROI, since our method adapts the aspect ratio accordingly (see Figure 13 (b, c)).

Multi-scale banking. For line charts, a single aspect ratio might not show all patterns of interest. Therefore, Heer and Agrawala [19] proposed the idea of multi-scale banking, which combines spectral analysis with the banking to 45° principle to automatically select multiple aspect

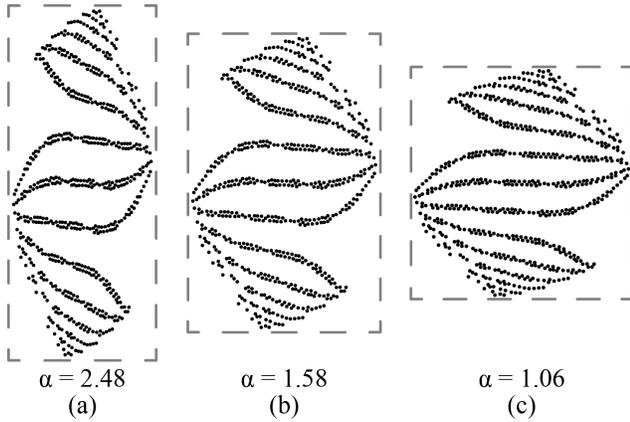


Fig. 14. Multi-scale banking on the *Balance Scale* dataset [10]. Three aspect ratios are picked to show patterns of interest at different scales.

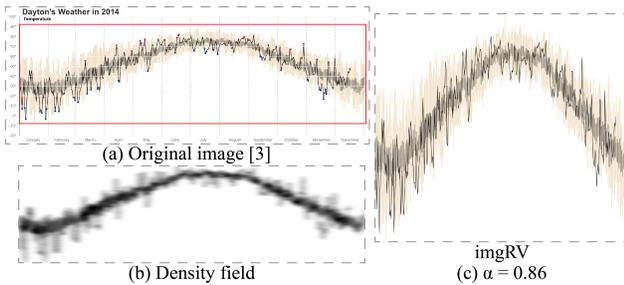


Fig. 15. Selecting a good aspect ratio for an image from the Internet [3]: (a) crop a relevant image part; (b) density field converted from the cropped image; and (c) result plotted with our selected aspect ratio.

ratios and to show patterns of interest at different scales.

Our image-based approach can also be extended to support multi-scale banking. For this, we build a Gaussian pyramid on the input density field and apply our imgRV method to compute the optimal aspect ratio for each image in the Gaussian pyramid. Once we obtain the aspect ratios, we follow [19] to ignore similar aspect ratios, i.e., aspect ratios that are less than 1.25 times of the previous (larger scale) ratio. Figure 14 shows a result with three selected aspect ratios for highlighting patterns of interests at different scales.

General image data. Our image-based approach can also handle general image data, simply by converting input images into scalar density fields, e.g., by using color2gray [18] (for color images) to obtain grey scale images, and then applying a Gaussian filter to smooth them.

Figure 15 shows an example image collected from the Internet [3]. After cropping the relevant part of the image in Figure 15 (a), we converted it to a density field (b) and then applied our approach to finding a better aspect ratio. The image scaled by applying our selected aspect ratio in Figure 15 clearly shows the Gaussian envelope of the weather variation, which is obscured in the original input. Figure 16 shows another example for a density function plot (or continuous scatterplot), which is from a CT *facial deformity* dataset [41]. Our image-based method enables us to directly compute a good aspect ratio from the density field for showing the arch structure more prominently in the density-based image data.

8 CONCLUSION AND FUTURE WORK

In this paper, we introduced a generic image-based approach for selecting a good aspect ratio for 2D diagrams, which allows to find similarly good aspect ratios as existing methods in less time. Derived from Federer’s co-area formula, we are able to generalize any previous aspect ratio selection method based on a line integral representation to an image-based method. To more faithfully characterize the data patterns, we adopted anisotropic kernel density estimation to construct the

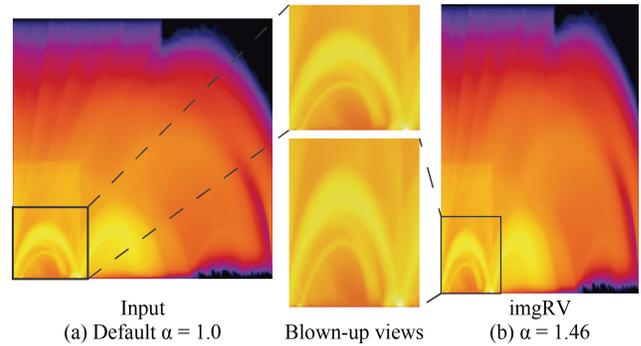


Fig. 16. Our approach can compute an aspect ratio directly from a density function plot. (a) The input dataset is a continuous scatterplot of the CT *facial deformity* data [41]. (b) Rescaled using α selected by imgRV.

necessary density fields. A comprehensive quantitative comparison to previous isoline-based methods shows that our image-based resultant vector (imgRV) method is faster and more robust than existing methods. In addition, we revisited the evaluation of Fink et al. [16], and showed that our approach is able to perform similarly to the best method of Fink et al. for the tested data, while our selected aspect ratios better match the user preferences. Finally, we presented three extensions for our approach: region-of-interest banking, multi-scale banking, and selecting aspect ratios for general image data.

For future work, we plan to conduct a large-scale user study to find out to what extent the aspect ratio selected by our approach is able to reveal different patterns in the visualizations. Furthermore, we want to spotlight some fundamental topics that have so far been neglected. For example, we see a need to better examine the perceptual foundations of aspect ratio optimization, i.e., there is a need for more refined models that might go beyond banking to 45° . As recognized by Talbot et al. [35], the assumption that a line with a 45° angle is optimal for perception is disputable, since in our opinion it cannot differentiate between $\pm 45^\circ$. It would also be interesting to study the perceptual implications of dealing with data clusters in scatter plots, i.e., whether circles are in fact optimal in this scenario. Lastly, it would also be interesting to explore the selection of k for anisotropic KDE by using the scagnostic approach from [42] in order to characterize the data more accurately.

ACKNOWLEDGMENTS

This work is supported by the grants of the National Key Research & Development Plan of China (2016YFB1001404), NSFC (61772315), NSFC-Guangdong Joint Fund (U1501255), Leading Talents of Guangdong Program (00201509), Shandong Provincial Natural Science Foundation (ZR2016FM12), the Open Research Fund of Beijing Key Laboratory of Big Data Technology for Food Safety, Beijing Technology and Business University, the Fundamental Research Funds of Shandong University, and the German Research Foundation (DFG) within projects A01/B01 of SFB/Transregio 161. We thank Marcel Hlawatsch for fruitful discussions.

REFERENCES

- [1] S. Bachthaler and D. Weiskopf. Continuous scatterplots. *IEEE Trans. Vis. & Comp. Graphics*, 14(6):1428–1435, 2008. doi: 10.1109/tvcg.2008.119
- [2] J. Bertin. *Semiology of Graphics*. University of Wisconsin Press, 1983.
- [3] B. Boehmke. Dayton’s weather in 2014. https://rpubs.com/bradleyboehmke/weather_graphic. Accessed January 2, 2015.
- [4] H. Carr, B. Duffy, and B. Denby. On histograms and isosurface statistics. *IEEE Trans. Vis. & Comp. Graphics*, 12(5):1259–1266, 2006. doi: 10.1109/tvcg.2006.168
- [5] W. S. Cleveland. A model for studying display methods of statistical graphics. *Journal of Computational and Graphical Statistics*, 2(4):323–343, 1993. doi: 10.2307/1390686
- [6] W. S. Cleveland. *Visualizing Data*. Hobart Press, 1993.

- [7] W. S. Cleveland. *The Elements of Graphing Data*. Wadsworth Advanced Books and Software, 1994.
- [8] W. S. Cleveland, M. E. McGill, and R. McGill. The shape parameter of a two-variable graph. *Journal of the American Statistical Association*, 83(402):289–300, 1988. doi: 10.1080/01621459.1988.10478598
- [9] Y. Croissant. *Ecdat: Data Sets for Econometrics*, 2016. R package version 0.3-1.
- [10] D. Dheeru and E. Karra Taniskidou. UCI machine learning repository, 2017. <http://archive.ics.uci.edu/ml>.
- [11] T. M. Dijkstra, B. Liu, and S. Oomes. Perception of ellipse orientation: data and Bayesian model. *Journal of Vision*, 3(9):11–11, 2003. doi: 10.1167/3.9.11
- [12] B. Duffy, H. Carr, and T. Möller. Integrating isosurface statistics and histograms. *IEEE Trans. Vis. & Comp. Graphics*, 19(2):263–277, 2013. doi: 10.1109/tvcg.2012.118
- [13] F. Färber, S. K. Cha, J. Primsch, C. Bornhövd, S. Sigg, and W. Lehner. Sap hana database: data management for modern business applications. *ACM SIGMOD Record*, 40(4):45–51, 2012. doi: 10.1145/2094114.2094126
- [14] H. Federer. Curvature measures. *Transactions of the American Mathematical Society*, 93(3):418–491, 1959. doi: 10.2307/1993504
- [15] H. Federer. *Geometric Measure Theory*. Springer, 1996.
- [16] M. Fink, J.-H. Hauernt, J. Spoerhase, and A. Wolff. Selecting the aspect ratio of a scatter plot based on its Delaunay triangulation. *IEEE Trans. Vis. & Comp. Graphics*, 19(12):2326–2335, 2013. doi: 10.1109/tvcg.2013.187
- [17] J. C. Gilbert and J. Nocedal. Global convergence properties of conjugate gradient methods for optimization. *SIAM Journal on Optimization*, 2(1):21–42, 1992. doi: 10.1137/0802003
- [18] A. A. Gooch, S. C. Olsen, J. Tumblin, and B. Gooch. Color2gray: saliency-preserving color removal. *ACM Trans. on Graph.*, 24(3):634–639, 2005. doi: 10.1145/1073204.1073241
- [19] J. Heer and M. Agrawala. Multi-scale banking to 45°. *IEEE Trans. Vis. & Comp. Graphics*, 12(5):701–708, 2006. doi: 10.1109/tvcg.2006.163
- [20] J. Heinrich and D. Weiskopf. Continuous parallel coordinates. *IEEE Trans. Vis. & Comp. Graphics*, 15(6):1531–1538, 2009. doi: 10.1109/tvcg.2009.131
- [21] R. Hyndman. fma: Data sets from forecasting: methods and applications by Makridakis, Wheelwright & Hyndman (1998). *R package version*, 2, 2009.
- [22] O. D. Lampe and H. Hauser. Curve density estimates. *Computer Graphics Forum*, 30(3):633–642, 2011. doi: 10.1111/j.1467-8659.2011.01912.x
- [23] D. J. Lehmann and H. Theisel. Discontinuities in continuous scatter plots. *IEEE Trans. Vis. & Comp. Graphics*, 16(6):1291–1300, 2010. doi: 10.1109/tvcg.2010.146
- [24] G. Loffler. Perception of contours and shapes: Low and intermediate stage mechanisms. *Vision Research*, 48(20):2106–2127, 2008. doi: 10.1016/j.visres.2008.03.006
- [25] J. Mackinlay. Automating the design of graphical presentations of relational information. *ACM Trans. on Graph.*, 5(2):110–141, 1986. doi: 10.1145/22949.22950
- [26] J. E. Marsden and A. Tromba. *Vector Calculus*. Macmillan, 2003.
- [27] V. Molchanov, A. Fofonov, and L. Linsen. Continuous representation of projected attribute spaces of multifields over any spatial sampling. *Computer Graphics Forum*, 32(3):301–310, 2013. doi: 10.1111/cgf.12117
- [28] F. Morgan. *Geometric Measure Theory: A Beginner's Guide*. Academic Press, 2016.
- [29] G. Saptarshi and W. S. Cleveland. Perceptual, mathematical, and statistical properties of judging functional dependence on visual displays. Technical report, Purdue University Department of Statistics, 2011.
- [30] A. Sarikaya and M. Gleicher. Scatterplots: Tasks, data, and designs. *IEEE Trans. Vis. & Comp. Graphics*, 24(1):402–412, 2018. doi: 10.1109/tvcg.2017.2744184
- [31] C. E. Scheidegger, J. Schreiner, B. Duffy, H. Carr, and C. T. Silva. Revisiting histograms and isosurface statistics. *IEEE Trans. Vis. & Comp. Graphics*, 14(6):1659–1666, 2008. doi: 10.1109/tvcg.2008.160
- [32] M. Sedlmair, A. Tatu, T. Munzner, and M. Tory. A taxonomy of visual cluster separation factors. *Computer Graphics Forum*, 31(3):1335–1344, 2012. doi: 10.1111/j.1467-8659.2012.03125.x
- [33] B. W. Silverman. *Density Estimation for Statistics and Data Analysis*. CRC Press, 1986.
- [34] J. Talbot, J. Gerth, and P. Hanrahan. Arc length-based aspect ratio selection. *IEEE Trans. Vis. & Comp. Graphics*, 17(12):2276–2282, 2011. doi: 10.1109/tvcg.2011.167
- [35] J. Talbot, J. Gerth, and P. Hanrahan. An empirical model of slope ratio comparisons. *IEEE Trans. Vis. & Comp. Graphics*, 18(12):2613–2620, 2012. doi: 10.1109/tvcg.2012.196
- [36] A. Tatu, G. Albuquerque, M. Eisemann, J. Schneidewind, H. Theisel, M. Magnork, and D. Keim. Combining automated analysis and visualization techniques for effective exploration of high-dimensional data. In *Proc. IEEE Symposium on Visual Analytics Science and Technology*, pp. 59–66, 2009. doi: 10.1109/vast.2009.5332628
- [37] W. Volkmuth and R. Austin. DNA electrophoresis in microlithographic arrays. *Nature*, 358(6387):600, 1992. doi: 10.1038/358600a0
- [38] J. Wagemans. Detection of visual symmetries. *Spatial Vision*, 9(1):9–32, 1995. doi: 10.1163/156856895x00098
- [39] Y. Wang, F. Han, L. Zhu, O. Deussen, and B. Chen. Line graph or scatter plot? Automatic selection of methods for visualizing trends in time series. *IEEE Trans. Vis. & Comp. Graphics*, 24(2):1141–1154, 2018. doi: 10.1109/tvcg.2017.2653106
- [40] Y. Wang, Z. Wang, L. Zhu, J. Zhang, C.-W. Fu, Z. Cheng, C. Tu, and B. Chen. Is there a robust technique for selecting aspect ratios in line charts? *IEEE Trans. Vis. & Comp. Graphics*, to appear, 2018. doi: 10.1109/tvcg.2017.2787113
- [41] Y. Wang, J. Zhang, D. J. Lehmann, H. Theisel, and X. Chi. Automating transfer function design with valley cell-based clustering of 2D density plots. *Computer Graphics Forum*, 31(3):1295–1304, 2012. doi: 10.1111/j.1467-8659.2012.03122.x
- [42] L. Wilkinson, A. Anand, and R. L. Grossman. Graph-theoretic scagnostics. In *Proceedings of the IEEE Symposium on Information Visualization*, pp. 157–164, 2005. doi: 10.1109/infvis.2005.1532142