A computational model for gestalt proximity principle on dot patterns and beyond

Peng Peng
MOE Key Laboratory for Neuroinformation and University of Electronic Science and Technology of China, Chengdu, China

Kai-Fu Yang
MOE Key Laboratory for Neuroinformation and University of Electronic Science and Technology of China, Chengdu, China

Yong-Jie Li
MOE Key Laboratory for Neuroinformation and University of Electronic Science and Technology of China, Chengdu, China

The human visual system has the ability to group parts of stimuli into larger, inherently structured units. In this article, a computational model inspired by tolerance space theory simulating the human perceptual grouping of dot patterns is proposed. Tolerance space theory introduces a tolerance relation to a discrete set to formulate the continuity of the discrete patterns. The model proposed herein includes one- and two-reach methods based on the assumption that dot patterns can be represented in the proposed extended tolerance space (ETS). Both methods are used to construct a ratio neighborhood graph (RANG), calculate tolerance from the diagram, compute the new RANG, and then rebuild continuous structures from the new RANG with a combinatorial procedure. Experiments are conducted to show the high consistency of the proposed model with human perception for various shapes of dot patterns, its ability to simulate Gestalt proximity and similarity principles, and its potential application in computer vision. In addition, the close relationship of the proposed model with the *Pure Distance Law* is comprehensively revealed, and the hierarchical representation of perceptual grouping is simulated with an adaptation of the proposed model based on the ETS.

Introduction

Perceptual organization, a basic function of the human visual system, is the ability to perceive objects, scenes, and events from visual information. There have been many studies published on the topic of perceptual organization, dealing with a wide range of phenomena (Wagemans et al., 2012), processes (Treisman, 1982), neural mechanisms (Grossberg & Mingolla (1987), models (Compton & Logan, 1993; Kubovy, 1994; Kubovy et al., 1998; Feldman, 2003; Froyen et al., 2015), and theories (Treisman, 1982; Wagemans et al., 2012). Within the wider study of perceptual organization, perceptual grouping is the process by which local elements are organized into connected groups or objects Wagemans (2015). This is important for humans because numerous objects are projected onto the retina in noncontiguous forms because of light conditions, overlapping, and occlusion. However, with the ability of perceptual grouping, humans are able to perceive continuous objects and understand what they are even if they seem discrete.

Simple examples are shown in Figure 1. Perceptual grouping enables us to recognize four groups (vertical lines) consisting of five dots each in Figure 1(a). Similarly, the letter "A" in Figure 1(b) and the hollow disk in Figure 1(c) are also recognized by the perceptual organization of the human visual system.

In essence, visual processing is discrete, starting from the sampling of continuous stimuli of the outside world by the photoreceptor cells of the retina (Chen, 2005). Dot pattern perception is one of the basic components in visual perception research (Kubovy, 1994; Kubovy et al., 1998). Gestalt psychology, which considers a Gestalt as "a whole in itself, not founded on any more elementary objects," emphasizes the importance of both wholes and parts (Wagemans et al., 2012) and provides a list of principles for the empirical, methodological, and conceptual basis of perceptual organization, such as proximity, similarity, closure, good continuation, and common fate (Wertheimer, 1938; Wertheimer & Riezler, 1944). The classical Gestalt laws are often based

Citation: Peng, P., Yang, K.-F., & Li, Y.-J. (2021). A computational model for gestalt proximity principle on dot patterns and beyond. *Journal of Vision*, *21*(5):23, 1–21, https://doi.org/10.1167/jov.21.5.23.

https://doi.org/10.1167/jov.21.5.23 Received December 30, 2020; published May 20, 2021 ISSN 1534-7362 Copyright 2021 The Authors

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License.

Downloaded from abstracts.iovs.org on 04/19/2024

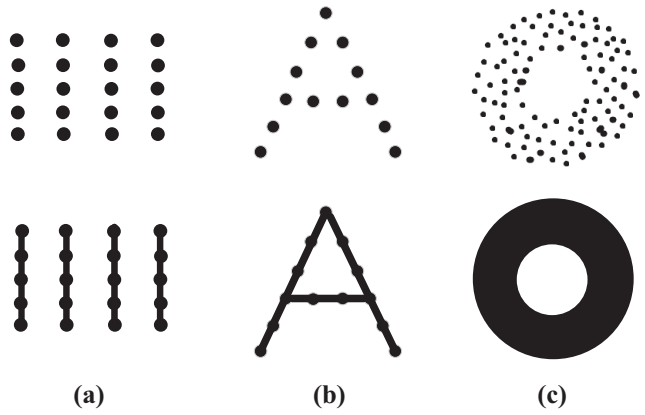


Figure 1. Example of perceptual grouping of dot patterns. The top row shows the original dot patterns, and the bottom row shows the corresponding continuous structures recognized by perceptual grouping. In this article, a computational model to simulate this phenomenon is proposed.

on apparently disconnected dot arrays or lattices, and they aim to explain how these disconnected arrays are organized into perceptually connected wholes. The Pure Distance Law (PDL), summarized by Kubovy et al. (Kubovy, 1994; Kubovy & Wagemans, 1995; Kubovy et al., 1998), quantifies the classical Gestalt laws of grouping by proximity in dot lattices, but does not provide a clear criterion as to which dots should be grouped, which may limit its applicability to discrete patterns, particularly dot patterns of arbitrary shape and density.

Although there are many computational models to simulate the proximity principles (Claessens & Wagemans, 2008; Feldman, 2003; Froyen et al., 2015; Papari & Petkov, 2005; Van Oeffelen & Vos, 1982, 1983), for example, good continuation (Claessens & Wagemans, 2008; Froyen et al., 2015; Lezama et al., 2017), there are no known methods that can simulate the proximity principle and reconstruct perceptual continuous shapes from nonuniform, arbitrarily shaped dot patterns. Therefore, the main purpose of this work is to develop a model to accurately simulate the proximity principle and reconstruct a continuous shape from nonuniform dot patterns.

Gestalt psychologists have reached a consensus on the superiority of the whole over its parts. The topological perception theory proposed by (Chen, 1982, 2005), which also claims "global first," seeks to construct a formal description for perceptual organization on both discrete and continuous patterns. However, the aforementioned computational methods either merely leveraged the local information in dot patterns (Claessens & Wagemans, 2008; Froyen et al., 2015) or simply set the threshold constant (Papari & Petkov, 2005).

Consequently, following the viewpoint of these psychology theories, the contribution of global information over local features of dot patterns is emphasized in the present work. Tolerance space theory (Zeeman, 1962; Sossinsky, 1986) is believed to serve as a proper mathematical approach for the perceptual organization of discrete patterns. It works by introducing a tolerance relation to a discrete set that is reflexive and symmetric to formulate the continuity of the discrete patterns Chen (2005). The details of tolerance space theory are defined in the *Preliminaries* section.

Specifically, the notions of tolerance space theory are borrowed to construct a ratio neighborhood graph (RANG) to explore the global context information of the dot patterns. The notion of tolerance space theory is introduced, and then a new concept of extended tolerance space (ETS) is proposed to promote the generalization capacity of tolerance space theory. Different from the definition of tolerance space, ETS has a special relationship between the data elements and the tolerance. Based on this extended concept, many discrete patterns can be properly represented in ETS; meanwhile, each continuous component satisfies the definition of standard tolerance space on its own. Therefore, finding the proper tolerance relationship for ETS is the core idea of the proposed model.

In this work, a one-reach ratio distance method is proposed to find the proper tolerance. Based on the one-reach strategy, a two-reach method is proposed to handle more complex structures and resist noise and outliers, and to construct a RANG. Subsequently, a combinatorial process is designed based on this RANG to reconstruct continuous structures from discrete dot patterns.

Many researchers claim that the Gestalt *proximity* principle has a close relationship with the *similarity* principle, which states that proximity is one of the conditions of similarity in location (Wagemans, 2015). Therefore, an attempt is made herein to simulate the similarity principle using the proposed model in terms of different features, such as color, shape, size, and velocity. Moreover, the proposed model is also extended to accomplish the task of image segmentation to show its potential in engineering applications.

In summary, in this article, an ETS theory is proposed as the principle underlying our computational model. This model includes a one-reach method and a two-reach method, which simulate perceptual grouping for noisy and nonuniform dot patterns. Physiologically meaningful parameter settings (i.e., the thresholds $th_{ID1} = 0.12$ and $th_{ID2} = 0.035$) are found in our computational model. Then, the consistency between our simulation results and human perception is quantitatively evaluated. Simulations for some Gestalt principles and applications to image segmentation are carried out to show the broad abilities of our model.

Downloaded from abstracts.iovs.org on 04/19/2024

Finally, the close relationship of our model to PDL, the hierarchical representation achieved with our model, simulations of five types of grouping, as well as the psychological significances of our basic parameters are all further discussed.

Related work

There are several subjects related to our work, including computational models simulating dot perception, clustering, surface reconstruction, and image segmentation.

The perception of regular dot patterns, such as dot lattices, was thoroughly studied by Kubovy et al. (Kubovy, 1994; Kubovy & Wagemans, 1995), who identified different types of dot lattices and the notion of the degree of instability of different dot lattices. They later found that the relative strength of grouping dots of a particular orientation approximates a decreasing exponential function of the relative distance to the shortest distance between dots in different orientations, and they referred to this as the PDL(Kubovy et al., 1998). The relationship between the proposed model and PDL is comprehensively analyzed in the Discussion section.

To mimic dot grouping with arbitrary shapes in human perception, Papari and Petkov (2005) adopted the Delaunay triangulation algorithm as an initial graph and then cut the improper edges to build a reduced Delaunay graph (RDG). However, the fixed threshold chosen in the RDG limits its power in modeling perceptual grouping.

Line detection from discrete patterns, such as dot and gabor patterns, was also studied in some previous work. Lezama et al. (2017) proposed an automatic algorithm for detecting good continuation in dot patterns. This algorithm identifies the dots in good continuation and uses them to find the lines in dot patterns. However, because this method calculates dots that are locally quasisymmetric, it only finds lines rather than continuous shapes. Detecting continuous lines in gabor patterns relies on another feature, the contour curvature, which is represented by orientations of gabors and modeled using an association field (Field et al., 1993). Furthermore, the maximum tolerable contour curvature and spatial adjacency of gabors in line were reported in Watt et al. (2008).

Hierarchical (or multiscale) perception is a basic and flexible characteristic of human perception. The normal distribution function (i.e., Gaussian filter) has been widely adopted in human perception simulation and computer vision applications within scale-space theory (Lindeberg, 1994, 2013). To quantify clusters of two-dimensional dot patterns flexibly like human perception, a related pioneering work (Van Oeffelen & Vos, 1982, 1983) proposed the COntour-DEtecting clustering algorithm (CODE), which is based on relative proximity using a binormal distribution function and is invariant to translation, rotation, and size. To thoroughly evaluate the consistency of CODE and its variants with human perception, Compton and Logan (1993, 1999) conducted a series of experiments in which the numerosity, scales, orientations, and shapes of dot stimuli were varied. In contrast with CODE, Rosenholtz et al. (2009) proposed to blur and merge the components in a higher feature space. This approach performs well across diverse patterns with diverse features, such as color and orientation.

For a long time, the lack of a formal way to account for human intuition in computational perceptual grouping models was recognized as a major shortcoming. To address this problem, a logic-based method, the Minimum Model theory, was proposed by Feldman (2003), which chose the most preferred model with least "coincidentalness." Probability views were also popular when building the formal description of human perception. Bayesian methods proposed in (Feldman & Singh, 2006; Froyen et al., 2015), which were based on mixture models, were successfully implemented to yield hierarchical representation for diverse ranges of data including shape skeleton, dot grouping, and image decomposition. Cue combination and prior knowledge integration were also explored in a Bayesian framework proposed by Claessens and Wagemans (2008) to investigate the multistable grouping of zigzag lattices. Although those models can model and interpret many perceptual observations, they are limited in interpreting some other fundamental observations and being applied in computer vision.

Clustering, another subject closely related to perceptual grouping, aims to group data to several clusters according to their inherent features. There exist many kinds of clustering methods in terms of different perspectives (Saxena et al., 2017), including hierarchical, partitional, grid, density based, and model based.

Image segmentation is another highly relevant and broad research field, encompassing image partition, edge detection, salient object detection, and so on. In order to segment objects from the background, many methods simulating perception and clustering have been adapted into image feature spaces. Shi and Malik (2000) adopted a global view of images within a graph-based technique to find the "better cut" for image partition. Comaniciu and Meer (2002) proposed a general nonparametric technique to analyze and group image pixels in feature spaces. Recently, deep learning has further promoted the performance of image segmentation (Peng et al., 2013). There exist such a huge number of studies using various methods within

Downloaded from abstracts.iovs.org on 04/19/2024

this field that we can only review a few of them; more reviews can be found in these surveys: Peng et al., 2013; Ghosh et al., 2019.

The surface reconstruction process of discrete dot patterns in the first part of the proposed model plays a significant role. Similarly, several computer graphics-based methods have conducted surface reconstruction from scatters using Delaunay triangulation. For instance, the crust (Amenta et al., 1998a; Zhao et al., 2001) adopts the Voronoi diagram and Delaunay triangulation to generate the vertices of triangles from noise-free data; however, these methods are unable to handle noisy and nonuniform data.

Persistent homology is an algebraic tool used to compute topological features for a space at different spatial resolutions by adopting the simplicial complex (Pun et al., 2018). A filtration process with filtration parameters in persistent homology can represent the structure information in the resolution under this parameter. However, persistent homology can only reveal the grouping attributive relationship of dots, whereas the proposed model can provide the continuous and topologically correct structures for the perceptual organization of dot patterns.

Preliminaries

Tolerance space theory

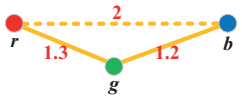
A tolerance is defined in terms of an algebraic relation that is reflective and symmetric but in general not transitive. Two stimuli x_1 and x_2 from a stimulus set X will not be distinguished if they are sufficiently close to each other. More details can be found in (Sossinsky, 1986; Peters & Wasilewski, 2012). One can say that these two stimuli are within a tolerance and that the tolerance ξ is defined as follows:

Definition 1. A tolerance ξ on set X is a binary relation on a set X, that is, $\xi \subseteq X \times X$, and has the following properties. (i) Reflexive: $\forall x \in X, (x, x) \in \xi$; (ii) Symmetric: $\forall x, y \in X$,if $(x, y) \in \xi$, then$(y, x) \in \xi$.

Unlike other binary relations, the tolerance is not in general transitive, which means that even if both (x_1, x_2) and (x_2, x_3) satisfy a tolerance ξ, it does not imply that the pair (x_1, x_3) also satisfies this tolerance. The nontransitive property is the foundation of tolerance theory. It is worth noting that the tolerance ξ is a relation, not a threshold; therefore, the symbol "$\in$" represents "satisfies" rather than "is included." A toy example of tolerance is illustrated in Figure 2.

Definition 2. A tolerance space (X, ξ) is a set X supplied with a tolerance ξ.

The notion of tolerance depicts the abstraction of the least noticeable difference with mathematical terminology. Referring to the analysis in Chen (2005),



$\xi : d_{ij} \leq 1.5$

- $d_{rg} \leq 1.5, then\ (r, g) \in \xi.$
- $d_{gb} \leq 1.5, then\ (g, b) \in \xi.$
- $d_{rb} > 1.5, then\ (r, b) \notin \xi.$

Figure 2. Toy example of tolerance space theory. Tolerance ξ here represents a relation in which the distance between vertices *i* and *j* is less than 1.5. Symbols $\in$ and $\notin$ represent the relationships of "satisfies" and "unsatisfies," respectively. Therefore, according to the edge length marked in red, dots *r* and *g* and dots *g* and *b* satisfy the tolerance ξ, whereas dots *r* and *b* do not.

the special nontransitive property of tolerance emphasizes the global topological properties.

As pointed out in Chen (2005), although Zeeman has proven that the global topological properties are preserved under a tolerance homeomorphism, it is still quite difficult to implement computationally (Zeeman, 1962). Our thought is that recovering the continuous structures of discrete patterns is one of the encouraging solutions that address this problem. Therefore, a new concept called ETS is defined to serve as the basis of the simulation model described in the section that follows.

Delaunay triangulation

Delaunay triangulation is a significant preprocessing technique in computational geometry (Delaunay, 1934; Amenta et al., 1998a, 1998b). A triangulation of a finite point set is a collection of triangulations or graph points, the vertices of which are the given set of points, and the edges of which do not intersect except at the points. Delaunay triangulation has many useful properties, e.g., the smallest angle of any triangle in triangulation is maximized, which ensures that the triangulation result is unique for any point set. In addition, Delaunay triangulation separates the surface where the dots locate with simplexes (triangles) without holes, which facilitates rebuilding a continuous structure from the results.

Methods

The human visual system has a limited capability in perceiving or discriminating cluttered stimuli because of visual acuity. For example, two dots cannot be discriminated when they are sufficiently close. Zeeman (1962) demonstrated this phenomenon as "the least noticeable difference," covering spatial acuity and contrast sensitivity, and advanced a mathematical structure of a "tolerance" to represent the nature of the

Downloaded from abstracts.iovs.org on 04/19/2024

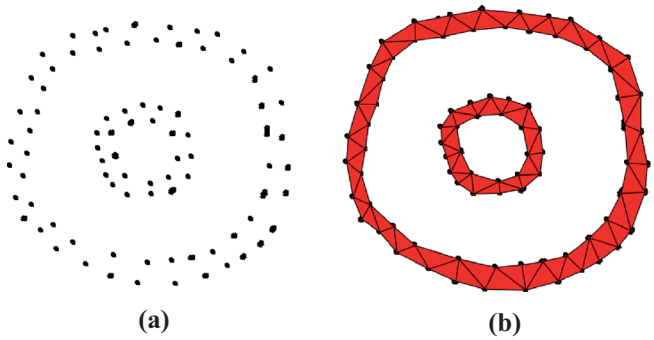


Figure 3. Toy example of ETS theory. (a) Given dot pattern, and (b) continuous map reconstructed from (a) using the proposed model. The form of ETS is used to represent the dot pattern as $(\{X_1; X_2\}, \xi)$, which means that the subsets X_1 and X_2 of dataset X with tolerance ξ meet the definition of the tolerance space; meanwhile, the set X with ξ does not meet this definition.

psychological "least noticeable difference." A specific discussion in Chen (2005) claims that the "tolerance space" may be an appropriate mathematical tool for translating the topological perception theory into the discrete situation of visual perception. More recent comprehensive details about tolerance space can be found in Peters and Wasilewski (2012). In this section, the tolerance space theory is extended to broaden the range of its applications. This extended theory serves as the cornerstone of the proposed model.

Extended tolerance space

Definition 3. Define a set $X = \{x_1, x_2, ..., x_m\}$, where X_n, $(1 \leq n \leq m)$ is the maximum independent subset of X, and each X_n satisfies the definition on tolerance space (X_n, ξ) with the same tolerance relation ξ, while any two subsets do not. Denote $(X, \xi) = (\{X_1; X_2; ...; X_n\}, \xi)$ as the *ETS*.

The following properties can be easily derived from Definition 3:

(a) The union of all of the maximum independent subsets is equal to the whole set, $\{X_1 \cup X_2 \cup ... \cup X_n\} = X$;

(b) There is no intersection between any two maximum independent subsets, $\{X_i \cap X_j; i \neq j, \forall i, j = 1, 2, ..., n\} = \varnothing$;

(c) The standard tolerance space defined in Definition 2 is just a spacial case of ETS, in which there is only one subset.

(d) Changing the tolerance ξ produces various ETSs for a set of data.

The ETS extends the applicability of tolerance space to many scenes. To demonstrate this notion, we present an example in Figure 3. Generally, one will perceive the discrete dot pattern shown in Figure 3(a) as two continuous circles, one nesting within the other; this is the ideal simulation result of the proposed model. It is supposed that all of the locations of dots are represented as a set X, and the dots in the inner and outer circles are denoted as X_1 and X_2, respectively. We can find a tolerance ξ by adopting our model to make the data X satisfy the definition of ETS, while the subsets X_1 and X_2 satisfy the definition of tolerance space. One can determine the tolerance relation ξ with an appropriate threshold to reconstruct a continuous structure that is similar to the perception of the dot pattern in Figure 3(b).

Computational perceptual organization model

Considering a set of discrete data $X_{data} = \{x_1, x_2, ..., x_m\}$, because there may be some other features like color and size, $X = \{v_1, v_2, ..., v_m\}$ is used to denote the dot positions in X_{data}, where m is the number of data points. Similar to the construction of an RDG (Papari & Petkov, 2005), an initial neighborhood graph is constructed on point set X first using Delaunay triangulation (Delaunay, 1934). Then, the unreasonable edges are removed to generate the *RANG* by finding thresholds using the one-reach method proposed in the Section on One-reach Ratio Distance Method or the two-reach method proposed in the Section on Two-reach Ratio Distance Method. This new graph with an appropriate tolerance relation satisfies the definition of ETS. The RANG is the skeleton of the continuous structures of dot patterns. Finally, a continuous map is reconstructed by filling and combining the reasonable triangles in the Section on Reconstruction of Continuous Structures. The output is regarded as the simulation result of the proposed computational model for the dot pattern.

Delaunay triangulation demonstrated in the Section on Delaunay Triangulation can create a neighborhood graph, where a vertex v is connected to its neighbors (Newman, 2010). To use Delaunay triangulation to build the proposed perceptual organization model, a new graph named RANG is defined first, which is a specific subgraph of a neighborhood graph.

It should be noted that two types of algorithms are used in this work to create neighborhood graphs: Delaunay triangulation and k-nearest neighbors (k-NN) (Newman, 2010). For the purpose of continuous structure reconstruction, Delaunay triangulation is adopted because it properly separates the surface using simplexes (triangles) without holes. However, Delaunay triangulation is computationally intensive and unreliable when data are of more than two dimensions. Therefore, k-NN is used when the focus is more on grouping rather than reconstructing continuous structures; for

Downloaded from abstracts.iovs.org on 04/19/2024

example, in the experiments in Sections on Proximity and Similarity Principle Simulation and Image Segmentation.

Definition 4. A RANG is a weighted, bidirected graph, $G = (V, R, W)$, where V is the dot set; $R = (E, L)$ is the ratio distance L of each edge E linking two adjacent dots in V; and W is the weighting of E. W has only two values: either $W = 0$ for removal or $W = 1$ for retention.

To construct the RANG, a *neighborhood graph* is first constructed using the Delaunay triangulation algorithm. Then, the Euclidean distance is calculated for each edge in the neighborhood graph, and the ratio distance of each edge is calculated by normalizing its Euclidean distance with the shortest Euclidean distance of those edges within the scope of its one- or two-reach neighborhood. Here, the one- or two-reach neighborhood of a dot x is defined as the neighboring dots (or nodes in the graph) that can be reached in at most one or two steps, respectively, from node x in the graph. One *step* means a node x can reach its neighborhood y via one edge and y's neighborhood z (not in the neighborhood of x) via two edges.

It should be noted that because the RANG is a bidirected graph, there exists only one Euclidean distance between a pair of neighbors p and q. However, there exist two ratio distances between them ($p \rightarrow q$ and $q \rightarrow p$) because the shortest edge in the neighborhood of a vertex may be different.

The RANG defined in this work is a new graph that consists of edges defined by the ratio distance rather than the absolute distance. This feature endows the proposed model with a good ability to find the intrinsic properties of dot patterns and handle noise and various densities.

One-reach ratio distance method

To keep the scale invariant and preserve the global structural information as much as possible, for the dot pairs of x with its j neighbors $N_x = \{n_1, n_2, ..., n_j\}$, the edge ratio distance for each edge $r(x, n_i)$ is calculated by dividing the Euclidean distance of the edge, i.e., $d(x, n_i)$, by the shortest Euclidean distance among the edges of x to N_x. This is written as

$$r(x, n_i) = \frac{d(x, n_i)}{\min\limits_{n_k \in N_x} \{d(x, n_k)\}}. \quad (1)$$

$R(x) = \{r(x, n_1), r(x, n_2), ..., r(x, n_j)\}$ is used to represent the ratio distance sequence of a dot in the neighborhood graph. Therefore, the sequence of all of the dots in the neighborhood graph is represented as $Q_0 = \{R(x_1), R(x_2), ..., R(x_m)\}$. Then Q_0 is sorted in ascending order to obtain a sorted ratio distance sequence Q_1. The results of edge normalization are illustrated in Figure 4.

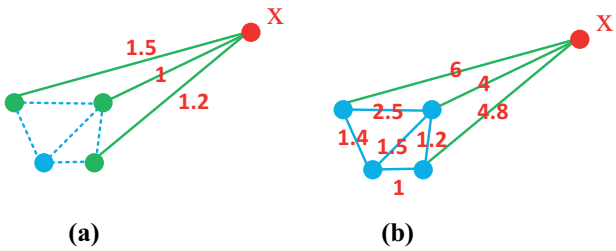


Figure 4. Edge normalization used in the proposed model. (a) One-reach neighborhood and (b) two-reach neighborhood (connected with solid lines) of red dot *x* in neighborhood graph. Ratio distance of the shortest edge (in Euclidean distance) is normalized to 1 in the RANG.

It can be inferred that all of the values calculated by Equation 1 approach 1 if the distances between all pairs of dots are similar. In contrast, if the lengths of edges vary largely, some of the edge ratio distances calculated by Equation 1 will be far greater than 1, which are supposed to be the unreasonable dot relations. Therefore, choosing a proper threshold to eliminate these improper edges is a feasible way of finding the perceptually reasonable dot relations. In the RDG (Papari & Petkov, 2005), the threshold is simply set as a constant value of 1.65, determined empirically, to remove the unreasonable edges defined by the Euclidean distance. Such a criterion neglects the essential distribution properties of different patterns.

Different from RDG (Papari & Petkov, 2005), the aim here is to determine the proper thresholds adaptively for different dot patterns by investigating their intrinsic features from distributions. Fortunately, after plotting the ratio distance Q_1 sorted in ascending order on a diagram, the global intrinsic features can be clearly observed.

An example of the Q_1 diagram is depicted in Figure 5(c). It is clear from the blue curve that there exist some abrupt increases in the ranges of [300, 350] and [400, 450] on the horizontal axis. This fact inspired us to find appropriate thresholds by detecting these abrupt increases and determining the criterion to cut these unreasonably long edges in the RANG, which is shown in Figure 5(b). Considering that the first derivative of the Q_1 curve is noise-sensitive when used to detect the abrupt increases in the Q_1 curve, the first derivative of Q_1 is modulated by multiplying a weighting item to form a one-reach indicator sequence Q_{ID1}, as shown by the red curve in Figure 5(c):

$$Q_{ID1}(k) = W_Q(k) \cdot Q_1'(k), \quad (2)$$

Downloaded from abstracts.iovs.org on 04/19/2024

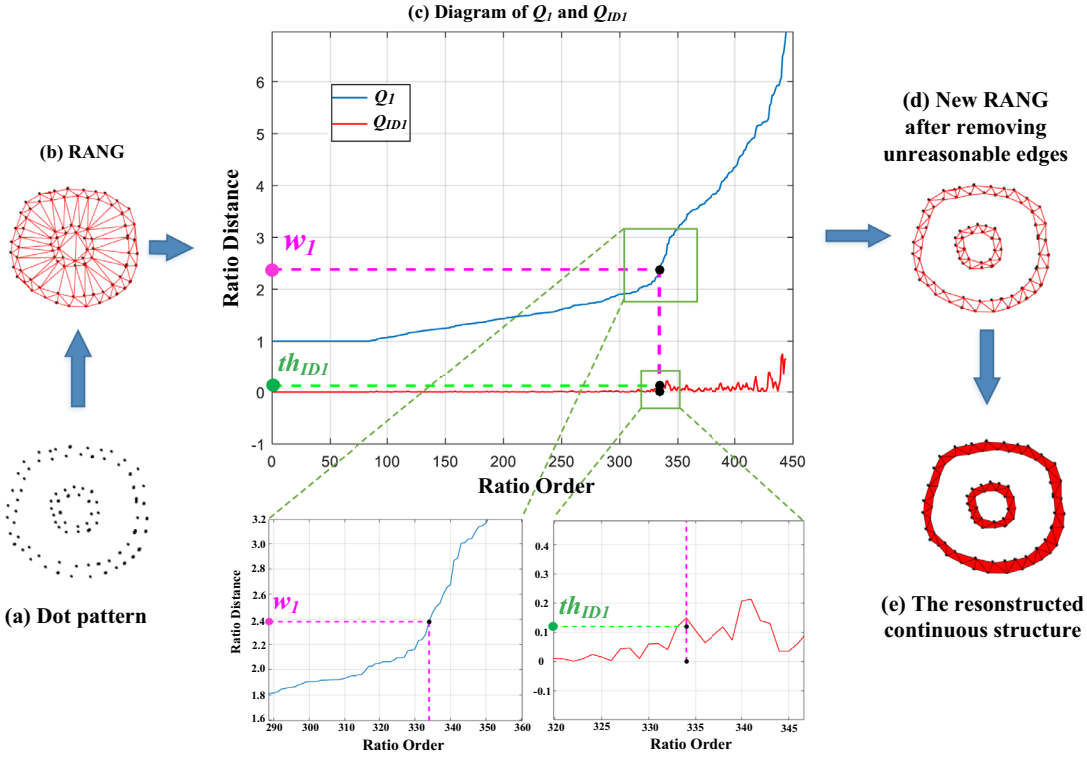


Figure 5. Flowchart of the one-reach method in our computational model. (a) Dot pattern. (b) RANG constructed by using Delaunay triangulation and edge normalization. (c) Diagram of the sorted ratio distance sequence Q_1 and indicator sequence Q_{ID1}. Indicator threshold $th_{ID1} = 0.12$ is empirically set to obtain the minimal ratio threshold w_1 to create the tolerance relation ξ. (d) According to the obtained ξ, edges between dot pairs that do not satisfy ξ will be disconnected to obtain a new RANG. (e) Continuous structure is reconstructed by filling the simplexes (i.e., triangles).

where k is the index of edge in Q_1; and $Q_1^{'}(k)$ is the first derivative of $Q_1(k)$ calculated as

$$Q_1^{'}(k) = \begin{cases} \frac{Q_1(k+1)-Q_1(k)}{2}; & \text{if } 2 \le k \le m-1, \\ 0; & \text{if } k = 1, \\ Q_1^{'}(k-1); & \text{if } k = m. \end{cases} \quad (3)$$

The weighting $W_Q(k)$ is the ratio distance of the current edge value $Q_1(k)$ to the mean cumulative value from 1 to k, which is calculated as

$$W_Q(k) = \frac{Q_1(k)}{\frac{1}{k}\sum_{i=1}^{k} Q_1(i)}. \quad (4)$$

$W_Q(k)$ represents the similarity to the mean of the previous values of Q_1. This means that if $W_Q(k)$ is much greater than 1, this ratio edge is more likely to be an unreasonable edge.

The value of the indicator $Q_{ID1}(k)$ indicates the stability of Q_1 at the k-th location. It was experimentally found that, when setting a threshold $th_{ID1} = 0.12$ for Q_{ID1}, a series of locations with abrupt increases can be stably obtained. Moreover, the Q_1 value at the first location obtained is exactly the one-reach cutting threshold w_1 as shown in Figure 5(b), which is computed as:

$$w_1 = \begin{cases} Q_1(\min \arg_k(Q_{ID1}(k) > th_{ID1})); & \text{if } \exists Q_{ID1}(k) > th_{ID1} \\ max(Q_1); & \text{else.} \end{cases} \quad (5)$$

Therefore, for an edge $L_{(p,q)}$ that links adjacent dots p and q (because each edge is computed twice, both from $p \to q$ and from $q \to p$), their states are denoted as 0 for removal or 1 for preservation.

$$W_{(p,q)} = \begin{cases} 1; & \text{if } Q_{1(p \to q)} \le w_1, \\ 1; & \text{if } Q_{1(q \to p)} \le w_1, \\ 0; & \text{else.} \end{cases} \quad (6)$$

$\{Q_{1(p \leftrightarrow q)} \le w_1\}$ in Equation 6 is the tolerance relationship ξ of an ETS $(\{X_1; X_2\}, \xi)$. Equation 6 demonstrates that if two dots (p and q) in a neighborhood graph accept the ξ that their edge ratio distance is smaller than w_1, the edge linking them

Downloaded from abstracts.iovs.org on 04/19/2024

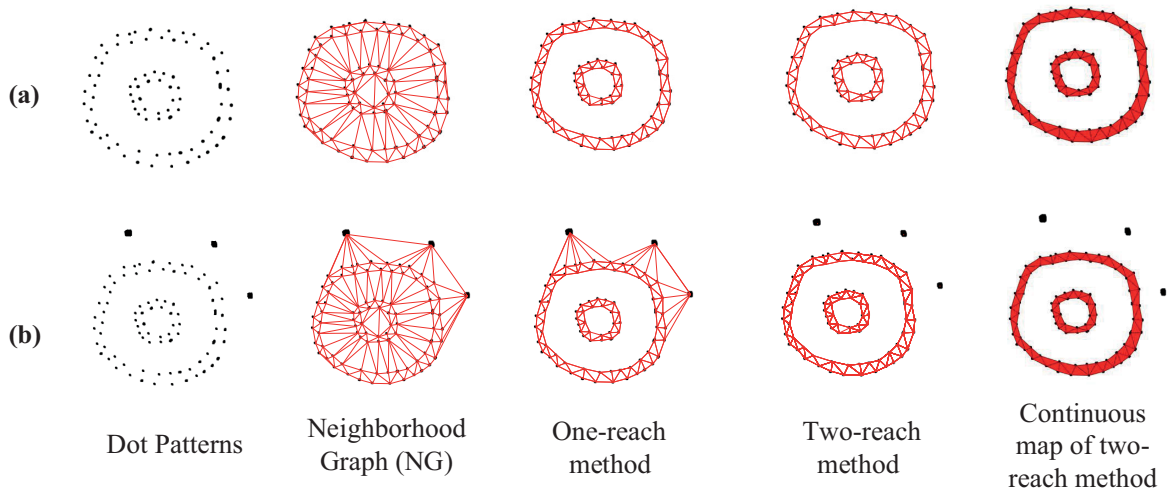


Figure 6. Illustrations of the proposed one- and two-reach methods for continuous structure reconstruction from (a) noise-free and (b) noisy dot patterns.

together is assigned a weighting $W_{(p,q)} = 1$, which means they are connected in the new RANG as shown in Figure 5(d); otherwise, $W_{(p,q)} = 0$, which means they are disconnected. The disconnected edges between adjacent dots are considered *unreasonable* because they are intuitively too long in the neighborhood graph and should be removed as shown in Figure 5. Finally, a new RANG is obtained from the discrete dot pattern shown in Figure 5(d).

It is worth mentioning that when the longest edge in the neighborhood graph approaches zero, which means the dot pattern converges to a continuous pattern, the ratio distance calculated by Equation 1 converges to 1, and all of the indicator values of Q_{ID1} should be 0. As a result, the threshold $w_1 = \max(Q_1) = 0$ is obtained by Equation 5, and eventually no edge will be removed according to Equation 6. This is the ideal result obtained while processing a continuous map. These analyses exhibit the robustness of the proposed model when they are applied to continuous patterns.

However, when the absolute distances between dots are large enough, there will be no continuous map perceived, but discrete scatters will be. Fortunately, if required, it is possible to add criteria in Equation 6 to specify the maximal absolute length of the longest edge of dot patterns.

Sometimes outliers exist in the discrete data, such as the dot pattern shown in Figure 6(b). Unfortunately, the proposed one-reach method can hardly handle this situation since the outliers have similar *reasonable* ratio distances to other dots. However, searching for the shortest edges in a two-reach neighborhood and calculating the ratio distance can avoid this dilemma.

Two-reach ratio distance method

Different from the one-reach method searching for the shortest edge within the one-step reachable neighborhood in the neighborhood graph, the shortest edge in our two-reach method is obtained by searching within a two-step reachable neighborhood from the RANG obtained by the one-reach method. This is illustrated in Figure 4. If it is assumed that the red dot is one of the outliers, the green dots are in its one-reach neighborhood [Figure 4(a)], and the blue dots are all in its two-reach neighborhood [Figure 4(b)]. The smallest ratio edges within these two kinds of neighborhoods are different. Consequently, if one sets a cutting threshold (for instance, 1.1), there still exist edges in the one-reach neighborhood connecting the outlier, whereas no edge will exist in the two-reach neighborhood. This property of the two-reach neighborhood largely improves the robustness to outliers.

It should be mentioned that directly searching for the shortest edge in the two-step neighborhood produces a heavy computational burden because the run time complexity of the computation procedure is $O(n^2)$. Therefore, applying the two-reach method on the RANG obtained by the one-reach method is natural and feasible. The sorted two-reach ratio sequence Q_2 is obtained in a similar way as the one-reach sequence Q_1. Additionally, the indicator Q_{ID2} detecting the abrupt increases of Q_2 is computed as

$$Q_{ID2}(k) = \frac{Q_2^2(k)}{\frac{1}{k} \cdot \sum_{i=1}^{k} Q_2(i)} \cdot Q_2''(k), \quad (7)$$

Downloaded from abstracts.iovs.org on 04/19/2024

where $Q_2''(k)$ is the second derivative of $Q_2(k)$ calculated by Equation 8.

$$Q_2''(k) = \begin{cases} \frac{Q_2(k+1)+Q_2(k-1)-2Q_2(k))}{4}; & \text{if } 2 \le k \le m-1, \\ 0; & \text{if } k = 1, \\ Q_2''(k-1); & \text{if } k = m. \end{cases} \quad (8)$$

The two-reach cutting threshold w_2 is obtained by Equation 9 when setting $th_{ID2} = 0.035$ empirically for Q_{ID2} in this work. The processing of the new RANG obtained by the two-reach method with w_2 is similar to that of the one-reach method, as demonstrated in Equation 6:

$$w_2 = \begin{cases} Q_2(\min \arg_k(Q_{ID2}(k) > th_{ID2})); & \text{if } \exists Q_{ID2}(k) > th_{ID2}, \\ max(Q_2); & \text{else.} \end{cases} \quad (9)$$

The examples shown in Figure 6 exhibit the different capabilities of the two proposed methods for dealing with outliers. The proposed one- and two-reach methods share the same core idea to disconnect the unreasonable links, but they differ in calculation and application scenarios. The two-reach method can deal with various data, even data with outliers, but it is computationally intensive, whereas the one-reach method is weaker in processing outliers but less computationally intensive. The more computationally intensive versions, such as three (or more)-reach methods might not be practical since the outliers are already resisted by two-reach neighborhoods.

Reconstruction of continuous structures

Inspired by the combinatorial process (Zomorodian 2010) that combines the graph to the Vietoris–Rips complex, a similar process is designed to reconstruct a continuous map from the RANG. The process is to simply fill all of the triangles with a certain color. The difference between Zomorodian (2010) and the proposed model is that we only fill the triangles instead of finding the highest-dimensional simplexes. For the examples shown in Figure 6, the proposed two-reach method can properly reconstruct the continuous structures.

Experiments

Continuous structure reconstruction

Visual results of the proposed perceptual organization model and a comparison with those of the RDG (Papari & Petkov, 2005) are shown in Figure 7. The figure shows that all of the methods perform well on simple dot patterns. The one-reach method cannot resist the noise, whereas the two-reach method and the RDG can. The RDG has poor performance in complex structures such as those shown in Figures 7(e)-(j). On the contrary, our two-reach method obtains better results on all of the complex dot patterns tested here.

To quantitatively evaluate the performance of continuous structure reconstruction, the results of the proposed methods and RDG (Papari & Petkov, 2005) were compared with the human depicted graphs built in this work. However, the commonly used metrics for surface reconstruction are suitable for three-dimensional tasks, but not for flat surface reconstruction. The *F-score* is commonly used as a measure of accuracy in statistics and engineering (Beitzel, 2006), and it takes into consideration both precision and recall, which seems suitable to evaluate the connectedness of adjacent points. Therefore, we adopted F-score to evaluate the performance of the perceptual organization algorithms in this paper.

Observers: Ten observers with ages ranging from 22 to 28 with normal or corrected vision. All of the observers were naive to the purposes of the experiment.

Stimuli: Ten dot patterns (shown in Figure 7) and the corresponding neighborhood graphs pre-obtained using Delaunay triangulation.

Apparatus: Subjects were seated in a chair in front of a computer screen with a resolution of 1024 × 768 at a distance of 80 cm. All of the programs were written and run in MATLAB.

Procedure: We first ran the program, which simultaneously presented the dot patterns and corresponding neighborhood graphs in parallel. The instruction to the observers was to cut the redundant lines in the neighborhood graphs to form a continuous graph according to the original dot patterns until all of the stimuli were presented.

Results: The results of the human observers are regarded as the ground truths, and the mean of all of the observer data are denoted as G. The results obtained by specific algorithms are denoted as S. Some significant parameters such as the numbers of true positives, false positives, and false negatives are calculated as $TP = card(S \cap G)$, $FP = card(S) - card(S \cap G)$, and $FN = card(G) - card(S \cap G)$, where the function $card(\cdot)$ counts the element number of a finite set. *Precision* (P) and *recall* (R) are calculated as follows Beitzel (2006):

$$P = \frac{TP}{TP + FP} = \frac{card(S \cap G)}{card(S)}, \quad (10)$$

$$R = \frac{TP}{TP + FN} = \frac{card(S \cap G)}{card(G)}. \quad (11)$$

Downloaded from abstracts.iovs.org on 04/19/2024

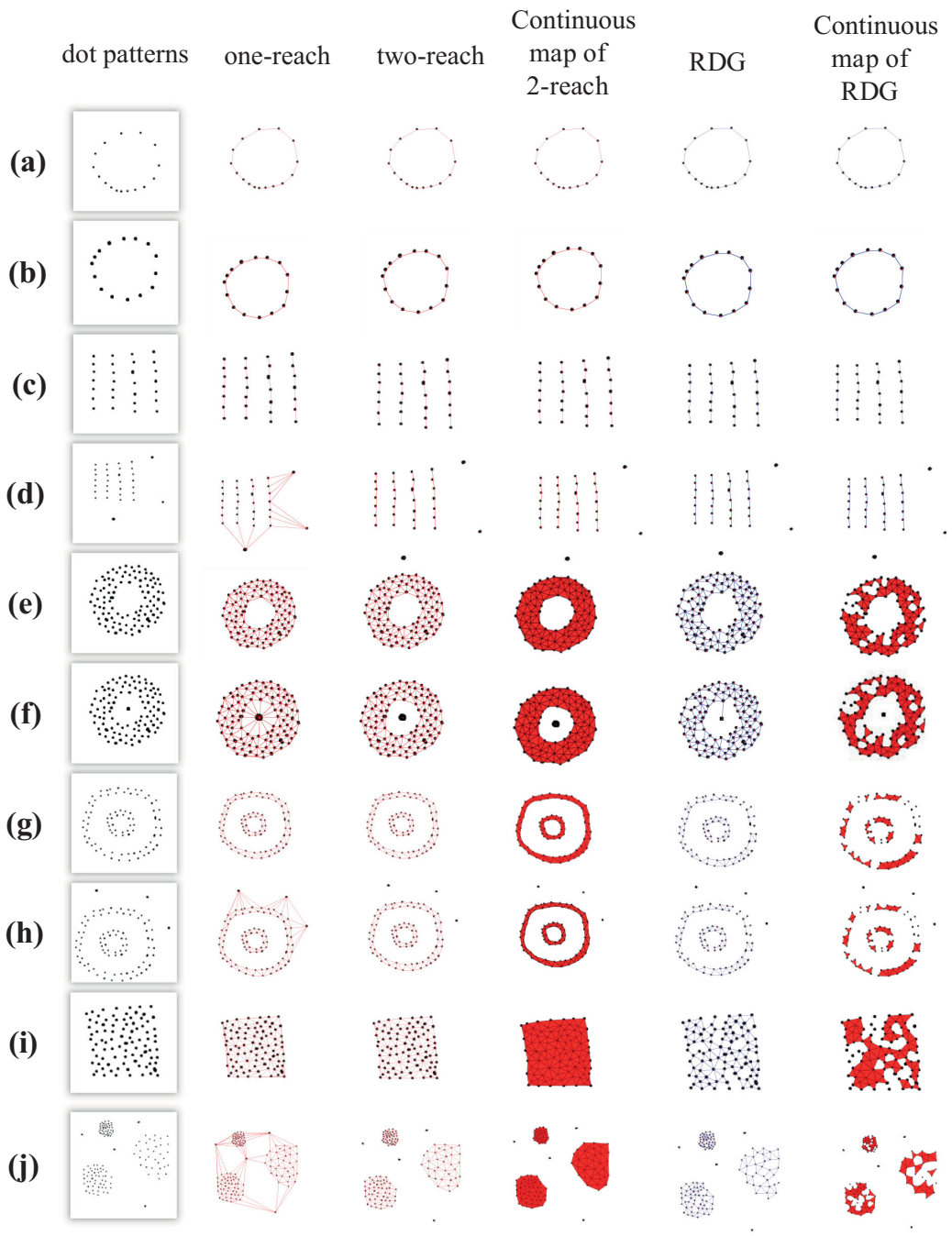


Figure 7. Quantitative evaluation of our model and RDG. Columns from left to right are dot patterns, simulation results of the one-and two-reach methods, continuous maps generated by the two-reach method, and the simulation results and continuous maps of RDG.

Then, the F-score (F) is calculated as

$$F = \frac{2 \cdot P \cdot R}{P + R}. \quad (12)$$

Precision and recall measure accuracy from two different perspectives, while the F-score can measure comprehensive performance. Intuitively, a higher F-score in the range (0, 1) means better results. It is worth noting that, because not all of the observers reported the same results, the F-scores of simulated results and even individual observers for some patterns never got the result of 1 in this experiment, such as (c) and (d) in Figure 8.

The F-score was also adopted to compare the results of the proposed one- and two-reach methods with those of the RDG in Figure 8. The tested dot patterns were

Downloaded from abstracts.iovs.org on 04/19/2024

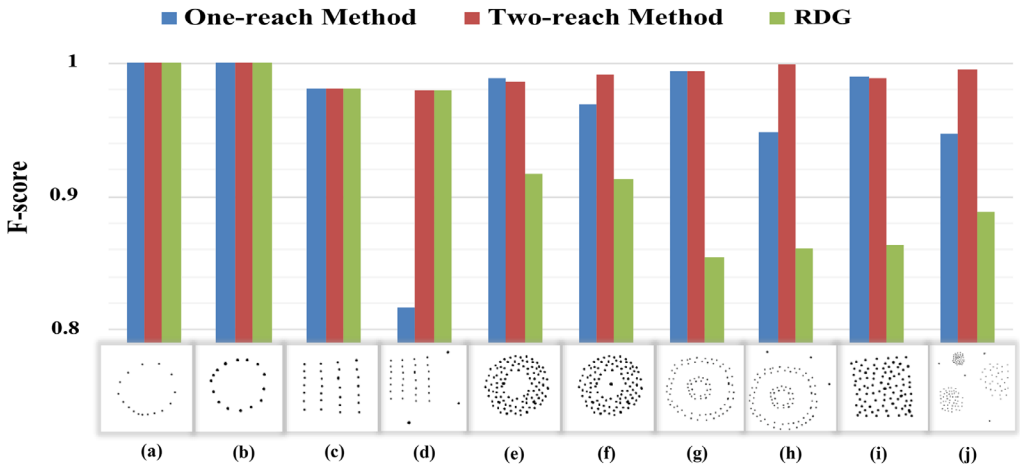


Figure 8. Quantitative evaluation and comparison of proposed methods and RDG.

drawn by hand, and each listed F-score is an average of the F-scores from human observers. The results shown in Figure 8 indicate that the performance of our one-reach method is good but sensitive to noise, as illustrated in the patterns in Figures 8(d), (f), (h), and (j). The proposed two-reach method has more stable performance on all of the data, whereas the RDG has worse performance on complex data like the patterns in Figures 8(e)–(j).

Constructing a surface from a graph obtained by Delaunay triangulation has been widely adopted. Consequently, it is naturally and reasonable to evaluate our methods and similar methods by leveraging the neighborhood graphs. However, a potential drawback is that the connected neighborhood graph may have some influence on the perceptual grouping of the original dot patterns. This potential distractor will be avoided in the experiment conducted in the next subsection, Perceptual Grouping.

To summarize, the main purpose of this experiment is to obtain the human labeled data of surface reconstruction from the neighborhood graph created by Delaunay triangulation and then evaluate the performance of our model and that of RDG. The results show that the proposed methods (especially the two-reach method) both have better performance than RDG, and the simulation results are highly consistent with human labeled data. This observation demonstrates that the proposed model can simulate the perceptual organization (mainly on continuous structure reconstruction) of dot patterns well.

Perceptual grouping

Perceptual grouping is also one of the functions of our methods. Experiments in the previous sections qualitatively exhibited the grouping abilities of our methods, as Figure 7 shows. In this section, we first introduce how we obtained the ground-truth through a psychophysical experiment, and then how we used two widely used metrics for evaluating the simulation results of our methods and RDG in terms of different transformations.

Observers: Eight observers with ages ranging from 22 to 28 with normal or corrected vision. All of the observers were naive to the purposes of the experiment, and they are different observers from the experiment in previous subsection.

Stimuli: Five dot patterns ((a),(b),(c),(g), and (h) from Figure 7) with different scales (i.e., 0.5 and 2 times the original image size) and rotations (i.e., 45° and 90°) were pre-obtained. The corresponding label-images with labeled dots (initially all of the dots were colored in red, i.e., all of the dots were of the same class) were provided. The reason why we chose the patterns in Figure 7 rather than randomly distributed dot patterns as used by Compton and Logan (1993) is that our method extracts the global intrinsic features of dot patterns. Therefore, it is inappropriate to evaluate the performance of our method on randomly distributed patterns. To evaluate the robustness of our model, following the work of Compton and Logan (1999), we also chose to rescale and rotate the original stimuli to certain scales and orientations.

Apparatus: Subjects were seated in a chair in front of a computer screen with a resolution of 1024 × 768 at a distance of 80 cm. All of the programs were written and run in MATLAB.

Procedure: We first ran the program, which simultaneously presented the dot patterns and corresponding label-images in parallel. The instructions to the observers were to observe the randomly presented original dot images and then label the label-images in different colors according to their intuitive sense using the mouse. When the observer finished and saved the

Downloaded from abstracts.iovs.org on 04/19/2024

	Rescale				Rotation							
	0.5		2		$\pi/4$		$\pi/2$		Original		Average	
Dot patterns	NMI	RI	NMI	RI	NMI	RI	NMI	RI	NMI	RI	NMI	RI
(a)	1.000	1.000	1.000	1.000	0.875	0.903	0.875	0.903	1.000	1.000	0.950	0.961
(b)	/	1.000	/	0.933	/	0.948	/	1.000	/	0.957	/	0.968
(c)	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
(g)	0.982	1.000	0.953	0.994	0.959	0.994	0.959	0.994	0.982	1.000	0.967	0.996
(h)	/	0.760	/	0.770	/	0.850	/	0.713	/	0.714	/	0.761
Average	0.994	0.952	0.984	0.939	0.945	0.939	0.945	0.922	0.994	0.934	0.972	0.937

Table 1. The NMI and RI of the results of our two-reach method. "/" means this value does not exist.

	Rescale				Rotation							
	0.5		2		$\pi/4$		$\pi/2$		Original		Average	
Dot patterns	NMI	RI	NMI	RI	NMI	RI	NMI	RI	NMI	RI	NMI	RI
(a)	1.000	1.000	1.000	1.000	0.875	0.903	0.875	0.903	1.000	1.000	0.950	0.961
(b)	/	1.000	/	0.933	/	0.948	/	1.000	/	0.957	/	0.968
(c)	0.647	0.880	0.647	0.880	0.647	0.880	0.647	0.880	0.647	0.880	0.647	0.880
(g)	0.694	0.887	0.674	0.881	0.679	0.882	0.679	0.882	0.694	0.887	0.684	0.884
(h)	/	0.760	/	0.770	/	0.850	/	0.713	/	0.714	/	0.761
Average	0.780	0.905	0.774	0.893	0.734	0.892	0.734	0.875	0.780	0.888	0.760	0.891

Table 2. The NMI and RI of the results of RDG.

data of one image, another one was presented, until all of the images were presented.

Evaluating the performance of clustering is too complicated to use the F-score adopted in the previous section. Therefore, we adopted two of the widely used metrics of clustering, that is the Normalized Mutual Information (NMI) Vinh et al. (2010) and Rand index (RI) Rand (1971). RI is complementary to NMI because the latter fails in some cases, particularly when both the predicted data and ground truth have only one class.

Results: The results are shown in Tables 1 and 2. Table 1 shows the results of our two-reach method. For most of the dot patterns (i.e., (a)-(g)), NMI and RI were higher than 0.9. For the complex pattern (h) they were larger than 0.7. These results show that our method is insensitive to scale and rotation. The averaged results in terms of stimuli and in terms of metrics were both higher than 0.9, indicating the effectiveness of our method. Table 2 shows various results of RDG (Papari & Petkov, 2005). RDG was also stable across different transforms, but the performance of RDG was worse than the performance of our method, especially for the complex dot patterns (c), (g), and (h).

The method of CODE, which was proposed in Van Oeffelen and Vos (1982, 1983), also groups dots using relative proximity. However, there are two main differences between CODE and our methods. First, our methods find the adjustive thresholds by an indicator function when searching the global properties of dot patterns, whereas CODE simply sets them by observing local properties. Second, because the relative proximity is represented by Euclidean distance, our methods are more flexible and hence easier integrate with other features compared to CODE, which uses Gaussian functions to represent the relative proximity.

Proximity and similarity principle simulation

The proximity principle is commonly considered a special case of similarity in location Wagemans (2015). and Kubovy and Van Den Berg (2008) reported that the grouping of proximity and similarity is additive in the logarithmic form of the PDL. Since the proposed model is designed to simulate the proximity principle, it may also be possible to simulate the similarity principle according to these observations.

Figure 9(a) shows a list of examples of discrete patterns of similarity in terms of various features f, including color (C), shape (D), size (E), orientation (F), and common fate (G), which can also be treated as a

Downloaded from abstracts.iovs.org on 04/19/2024

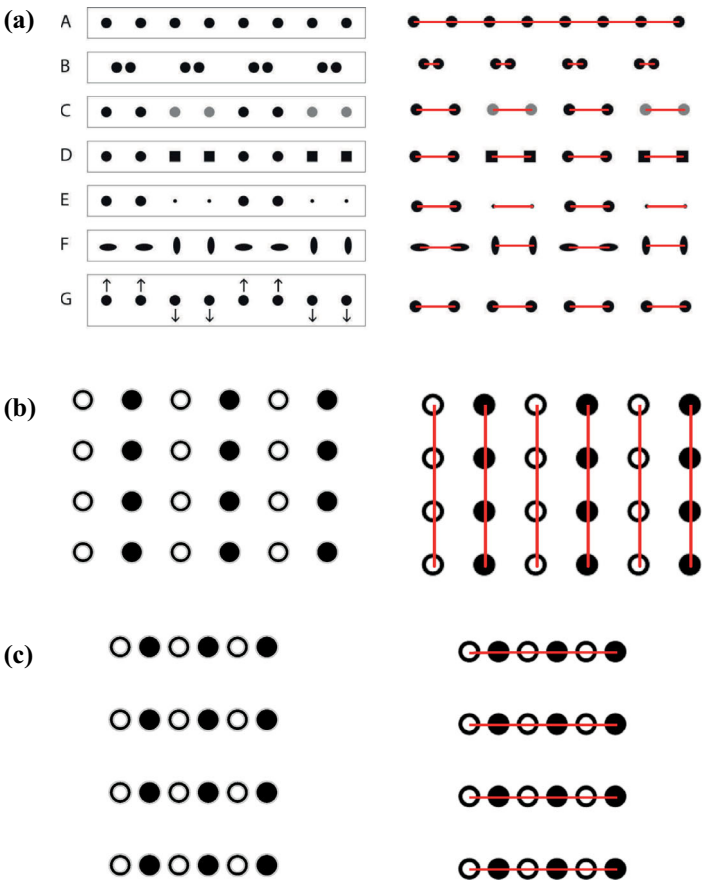


Figure 9. Simulations of gestalt proximity and similarity law by using the proposed model. (a) Left: dot arrays grouped by gestalt principles of proximity (B) and similarity in terms of color (C), shape (D), size (E), orientation (F), and velocity (G). Right: corresponding simulations using the proposed one-reach method. It should be noted that no dot-wise grouping observed in A is equivalent to the observation that all dots are grouped to one line. (b-c) When proximity and similarity principles simultaneously occur, a change in the weighting w_f in Equation 13 can appropriately simulate vertical grouping (b) and horizontal grouping (c).

similarity of velocity. The edge distance linking dot pair (p, q) is computed as

$$d_{(p,q)} = \sqrt{\|\Delta x\|_2^2 + \|\Delta y\|_2^2 + \|w_f \cdot \Delta f\|_2^2}, \quad (13)$$

where Δx, Δy, and Δf represent the distance in their respective dimensions, and w_f is the weighting of feature f. Then, the RANG of this pattern is constructed based on this edge distance, and, with proper w_f in terms of features, the grouping of these patterns is appropriately simulated by the proposed model (one-reach method) as shown on the right-hand side of Figure 9(a). Dot lattice examples with two kinds of features of dots are exhibited in Figures 9(b and c).

To summarize, these results demonstrate that the proposed model can simulate proximity and similarity appropriately. Interestingly, the ability of simulating proximity and similarity make the proposed model able to cluster data. Therefore, in the Image Segmentation subsection, the proposed model will adopt the distance between super-pixels calculated by Equation 13 to segment the digital images.

Image segmentation

One of the natural applications of grouping in proximity (or similarity) in the field of computer vision is image segmentation, which partitions a digital image into segments. Numerous methods have been proposed for image segmentation (Shi & Malik, 2000; Chan & Vese, 2001; Felzenszwalb & Huttenlocher, 2004; Pablo et al., 2011; Ronneberger et al., 2015).

Referring to the Proximity and Similarity Principle Simulation section, the proposed model can simulate the grouping of dot patterns in gestalt laws of similarity and proximity, and it should also be possible for it to accomplish the task of image segmentation. In this section, the two-reach method proposed in the section (Method) is adopted to segment the images.

Each image is first resized to 500 pixels in its larger dimension, and the superpixel map $S = \{s_1, s_2, ..., s_n\}$ is obtained using SLIC (Achanta et al., 2012) with size parameter $regionSize = 50$ and regular parameter $regularizer = 500$. Then, five features $f = \{f_p, f_c\}$, including the average position $f_p = \{x, y\}$ and the color feature $f_c = \{r, g, b\}$ in color space, are extracted for each superpixel and normalized to [0, 1].

It is held that the position features represent the proximity of superpixels and that the color features demonstrate the appearance similarity. Because there are five feature dimensions (two for location and three for color) of an RGB image, k-NN with $k = 5$ is adopted to replace Delaunay triangulation when constructing the neighborhood graph, because Delaunay triangulation is slow and unstable in processing high-dimensional data. Usually, image segmentation results are not unique, but they may vary from coarse to fine (Pablo et al., 2011). The simulation experiment of proximity and similarity described in the previous section revealed that it is feasible to balance the weightings between position and color features to obtain results in various scales.

The feature distance between super-pixels s_i and s_j is computed as

$$\begin{aligned} dist(s_i, s_j) &= \|\Delta f\|_2 \\ &= \sqrt{\|w_f \cdot \Delta f_p\|_2^2 + \|\Delta f_c\|_2^2}, \end{aligned} \quad (14)$$

Downloaded from abstracts.iovs.org on 04/19/2024

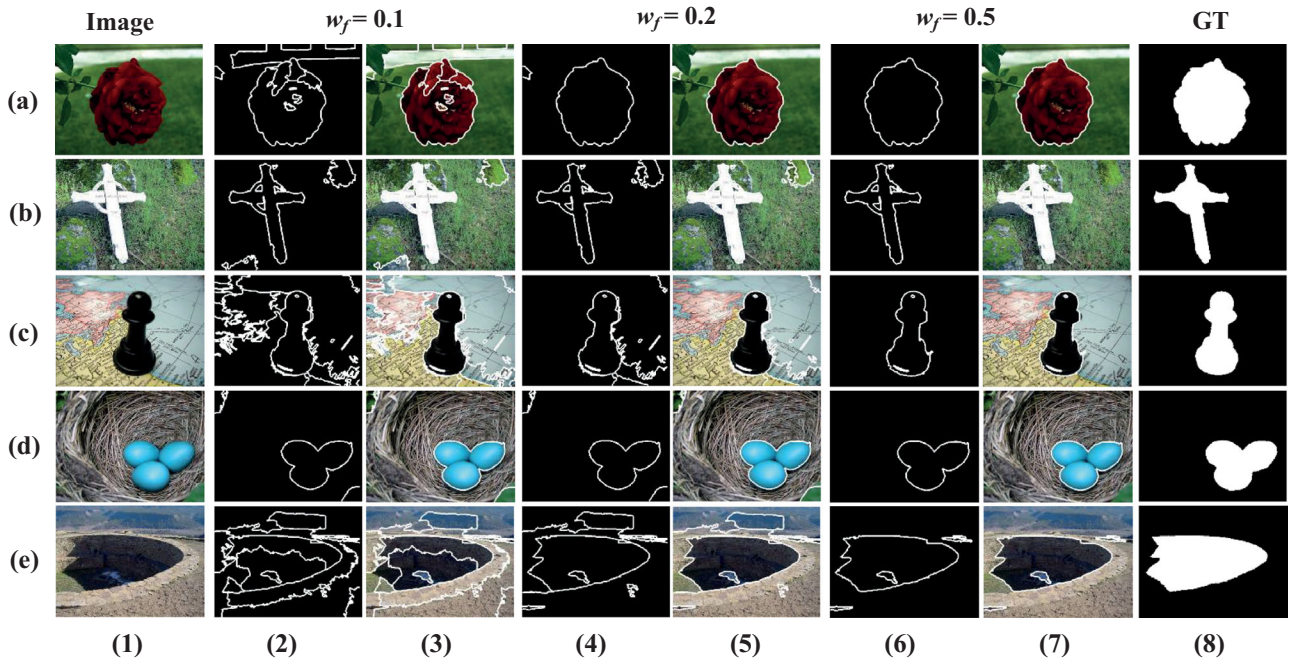


Figure 10. Visual results of image segmentation using the proposed model by slightly adapting weighting w_f. Column (1): Source images. Columns (2)–(7): Segmentations with $w_f = 0.1$, $w_f = 0.2$, and $w_f = 0.5$. Column (8): Human labeled salient objects of each images. The results show that the objects (or foregrounds) are correctly segmented from their backgrounds by using our model.

where w_f is a balance parameter between regional proximity and appearance similarity, and Δf is the difference of average feature values between two superpixels, including the location and color feature. Various neighborhood graphs are constructed with different balanced weightings w_f that can be used in the proposed methods. Finally, the proposed two-reach method combines two super-pixels if they are physically close to one another and have a similar appearance.

The visual results of the proposed model with different w_f are shown in Figure 10. As shown in this figure, one can obtain various plausible image segmentation results with different balance weightings, where $w_f = \{0.1, 0.2\}$ are relatively finer.

Compared with human labeled salient objects from Achanta et al. (2009), segmentation with $w_f = 0.5$ can correctly segregate the object from the background in each image, which shows the practical potential of the proposed model in computer vision.

Our model can also accomplish edge detection. Figure 11 shows some visual examples in which the edge maps can produce relatively good segmentations for each image, and the average edge maps are quite consistent with the human labeled edge ground truth from Arbelaez et al. (2010). However, row (d) in the figure exhibits a difficult case, where the leaves in the background are not well segmented owing to the limited ability of our model.

Because our model is a prototype of a segmentation method transferred from perceptual grouping, there are still several deficiencies to be further improved. One is that only location and color features (i.e., low-level or mid-level features) were used in our experiment, but more sophisticated features (e.g., texture, orientation, luminance, and even high-level cues) may benefit the segmentation ability of our method. Additionally, the parameters, such as regional size, number of superpixels, thresholds th_{ID1}, and th_{ID2}, weighting w_f, and even the methods for distance calculation can be more flexibly adjusted for better segmentation performance.

Discussion

Relationship to PDL

In this section, the close relationship of our model with PDL is analyzed, and it will be shown that the proposed computational model (including one- and two-reach methods) is a possible extension of the computational formulation of PDL proposed by Kubovy et al. (Kubovy & Wagemans, 1995; Kubovy et al., 1998).

Kubovy et al. (Kubovy and Wagemans, 1995; Kubovy et al., 1998) drew a conclusion on the grouping strength along two different orientations via the proximity rule, which is termed the Pure Distance Law (PDL). This law demonstrates that the relative strength of grouping into

Downloaded from abstracts.iovs.org on 04/19/2024

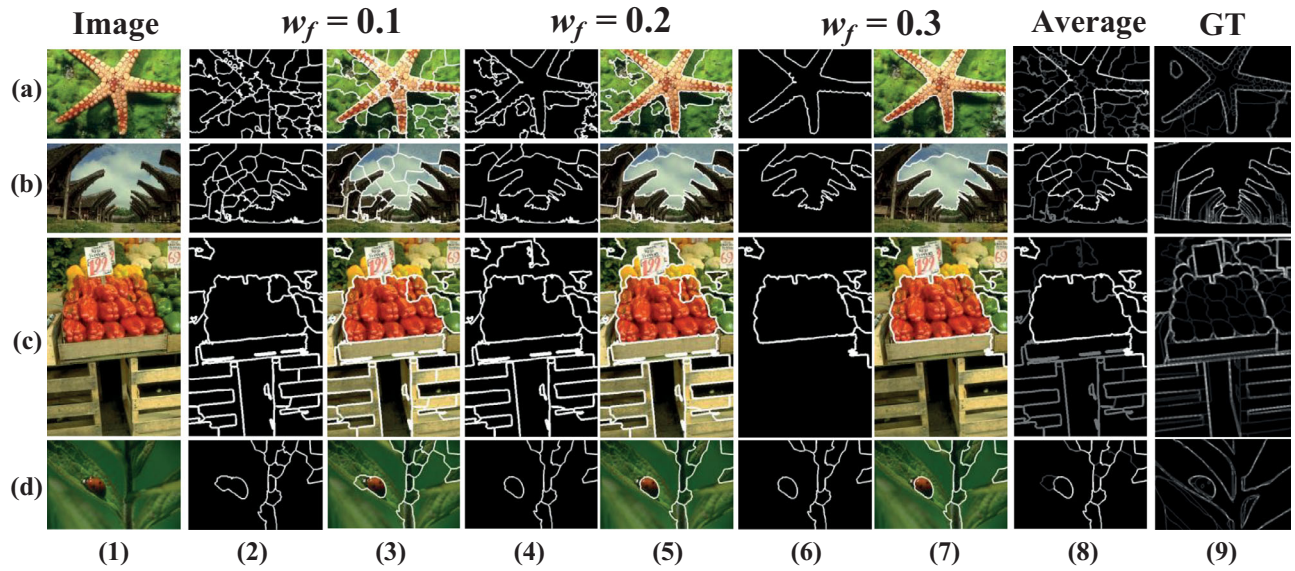


Figure 11. Visual results of edge detection using the proposed model by slightly adapting weighting w_f. Column (1): Source images. Columns (2)–(7): Segmentations with $w_f = 0.1$, $w_f = 0.2$, and $w_f = 0.3$. Column (8): Average edge maps obtained by our model. Column (9): Human labeled edge ground truths. (d) The source image presented a difficult edge detection case for our model.

one orientation of dot lattices approximates a decaying exponential function of the relative distance between dots in that orientation (Kubovy & Wagemans, 1995; Kubovy et al., 1998).

Specifically, considering a dot in the dot lattices shown in Figure 12(b), $|v|$ represents the distance of an edge v, which denotes a direction of any edge from $\{a, b, c, d\}$ to link the adjacent dots. a is the shortest edge in the neighborhood of a dot. $p(v)$ is the probability of each edge of the perceptual organization. $\frac{p(v)}{p(a)}$ represents the responses (i.e., grouping strengths) along the direction of v.

Therefore, the PDL is formulated as an *attraction function*:

$$ln\frac{p(v)}{p(a)} = \alpha\left(\frac{|v|}{|a|} - 1\right), \quad (15)$$

or

$$\frac{p(v)}{p(a)} = e^{\alpha\left(\frac{|v|}{|a|}-1\right)}, \quad (16)$$

where $\alpha < 0$ is a proximity attraction to show the grouping tendency. An illustration of this law is provided in Figure 12(a). The first subgraph shows this function, where the dots around the dashed line represent the ratios of various dot lattices. Taking a lattice as an example, the second sub-graph shows a lattice with $\gamma = \pi/2$. There are many edges, but only four kinds of ratios, that is, ($\frac{|a|}{|a|}$, $\frac{|b|}{|a|}$, $\frac{|c|}{|a|}$, and $\frac{|d|}{|a|}$) in this lattice, and all are plotted in the linear attraction function graph as the third sub-graph shows. It is clear that the grouping strength in the logarithm along a is greater than in others, so this lattice tends to be grouped in the vertical direction.

However, when the dot pattern is irregular, there exist many more kinds of ratios than lattices. The top part of Figure 13(a) shows such an example. The PDL only provides a description of grouping strength and tendency, but it does not suggest which ratios a dot should accept to form a cluster in such a dot pattern.

The proposed model (especially the one-reach method) gives a possible answer to address this problem. The intrinsic feature is explored from the all-ratio sequence and the threshold is found through an indicator function as shown in Figure 5. The simulation result of the dot lattice is given in Figure 12(b), where the ratio of $\frac{|a|}{|a|}$ is separated from others as shown in Figure 12(d), which means the dot lattice is grouped in the vertical direction.

The same simulation results are also obtained for the irregular dot patterns, such as those shown in the top part of Figure 13(a). Figure 13(b) shows our plausible results, including the grouping simulation and separated ratios in the logarithm and power-function graph. To demonstrate that the proposed method is better, it was compared with k-means (Likas et al., 2003) (Figure 13(c)) and OTSU Otsu (1979) (Figure 13(d)). The visual comparison results show that the proposed method obtains better grouping results than others.

Downloaded from abstracts.iovs.org on 04/19/2024

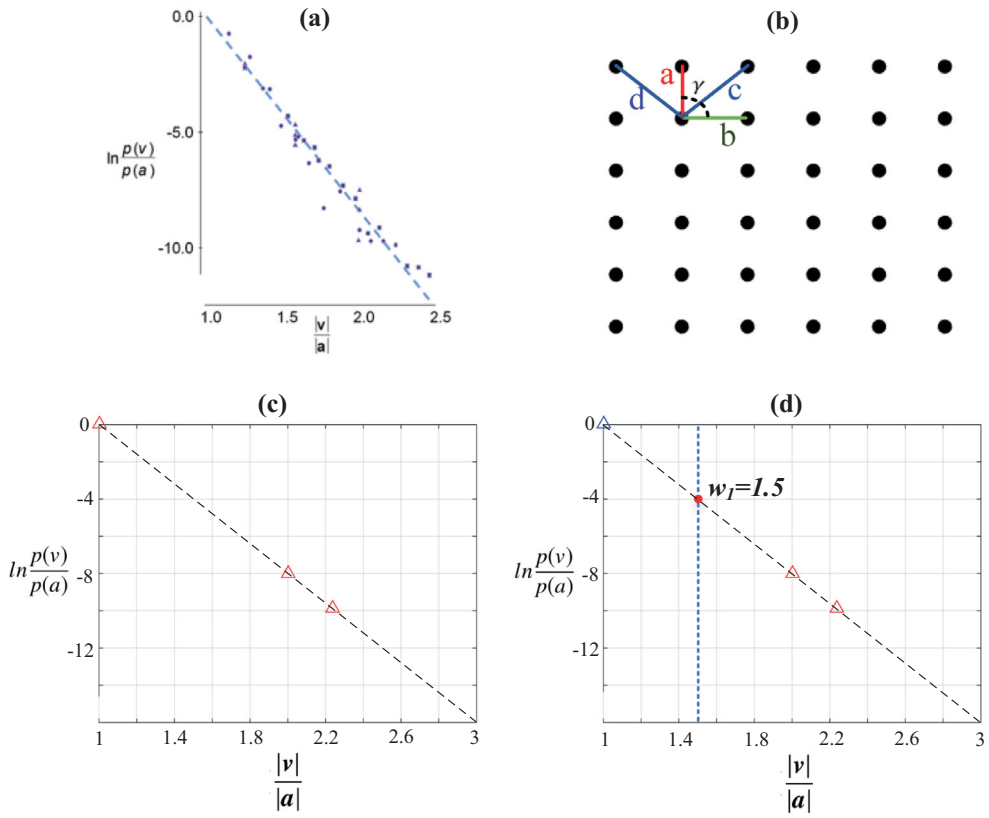


Figure 12. Illustration of the PDL on a dot lattice and how the proposed model finds the threshold for the PDL to determine the grouping direction. (a) Diagram of PDL. Ratios of various dot lattices are plotted in it. (b) Example of dot lattice, where *a* represents the shortest edge. γ is the angle of the horizontal and vertical edges. (c) Ratios of this dot lattice in the logarithm form of PDL. (d) Proposed one-reach method computes a threshold $w_1 = 1.5$ for this dot lattice to retain only the edge connected by *a*, which means that the dot lattice is grouped in the vertical direction.

In summary, the proposed model can conceivably provide a criterion for the PDL when deciding (or simulating) which direction the dot lattices are grouped along. Beyond this, the PDL is extended herein to group random, nonuniform dot patterns and reconstruct their continuous structures.

Hierarchical perceptual grouping

The viewpoint of perceptual grouping tends to be hierarchical has been reported by many studies (Palmer, 1977; Marr & Nishihara, 1978; Lee & Mumford, 2003). Referring to Def. 3, the ETS is a set of maximum independent subsets with a tolerance denoted as $(X, \xi) = (\{X_1; X_2; ...; X_n\}, \xi)$. However, when the tolerance ξ is altered, the maximum independent subsets and the grouping results can be different, which enables the proposed model to accomplish hierarchical grouping. There are many factors related to the task of hierarchical perceptual grouping, such as the distance between the eyes and objects. In this section, we discuss the possibility of our model simulating hierarchical perceptual grouping with a distance-modulated example.

The computational models described in the previous sections only simulate the perceptual organization of the human visual system at a specific distance. However, the perceptual organization of structures is also modulated by the distance between objects and the eyes. The human visual system can perceive the finer local geometric properties of an object at shorter distances, whereas more coarse and global properties are perceived at longer distances. This phenomenon is common in daily life. For example, the advertisements on LED billboards seem smooth and continuous at a distance, but many tiny, colored lamps will be seen up close. The essence of this phenomenon is that the ability to distinguish the retina image is influenced by distance and is due to the limited acuity and sensitivity of the retina. An illustration of this effect is shown in Figure 14(a).

This phenomenon was simulated by extending the proposed model to a distance-modulated model. To a certain extent, this distance-modulated model can be treated as a generalization of the proposed two-reach

Downloaded from abstracts.iovs.org on 04/19/2024

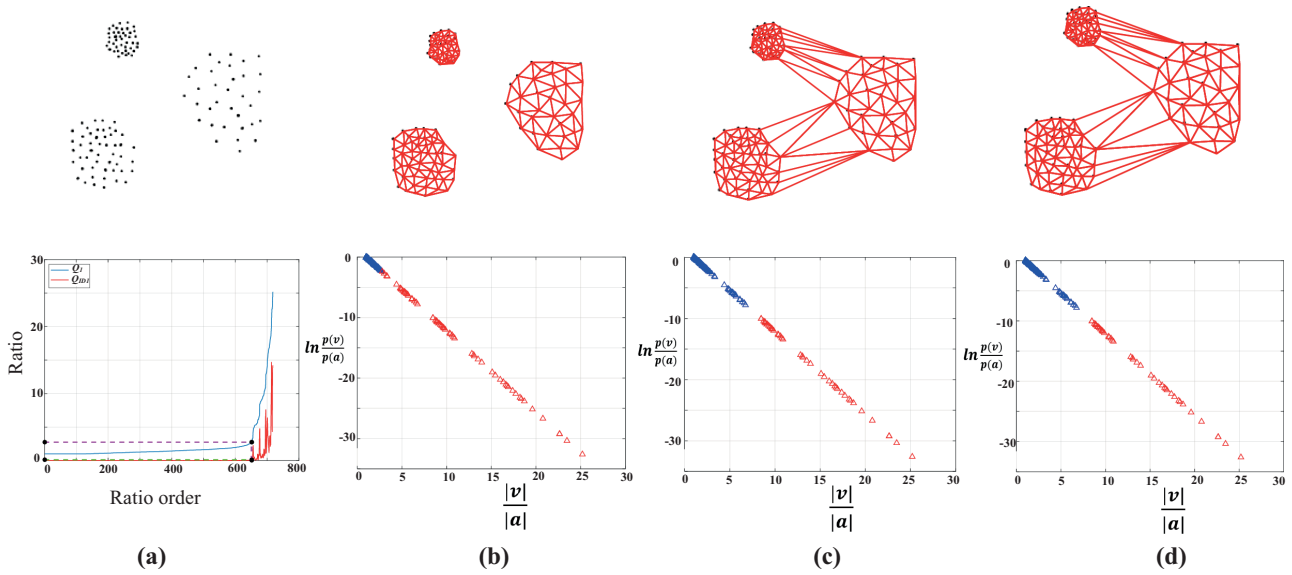


Figure 13. Illustration of finding a threshold for the PDL diagram for nonuniform data by using the proposed model, k-means, and OTSU. (a) Top: An example of a dot pattern. Bottom: Diagram of the proposed one-reach method when simulating the perceptual grouping of this dot pattern. (b) Top: Simulation result of proposed method. Bottom: Ratios plotted in the PDL diagram are split to blue and red using the threshold computed by proposed method. The ratio values marked in blue are accepted. (c) Top: Simulation of using k-means to group the ratios with a group number of 2. Bottom: Segregated ratios using k-means. (d) Top: Simulation of using OTSU to group ratios. Bottom: Segregated ratios using OTSU.

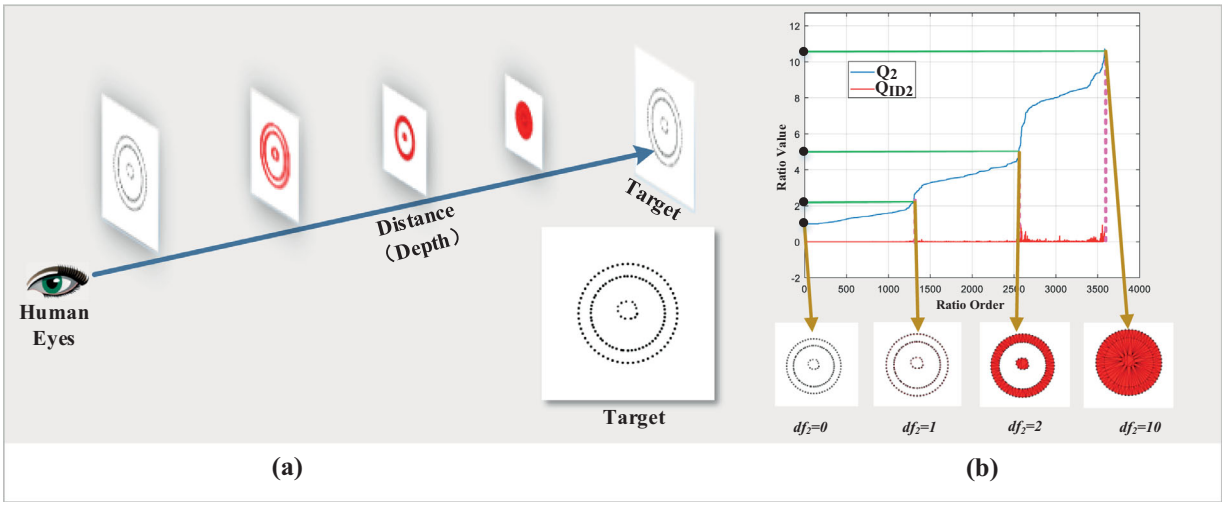


Figure 14. Simulation of distance-modulated perceptual grouping by modifying the proposed model. (a) Illustration of how humans see and perceive a dot pattern at different distances; (b) simulation of this process using our model. With various distance factors, different cutting thresholds [represented by the black dots along the *y*-axis in (b)] are obtained to reconstruct various continuous maps from the target.

method, because the proposed two-reach method finds the empirically best tolerance for removing the "most unreasonable" edges, and in the distance-modulated model, various levels of tolerance are found by altering the indicators th_{ID1} and th_{ID2}.

The diagram of Q_2 in the two-reach method depicts the structural properties from global to local, and the indicator sequence Q_{ID2} computed by Equation 7 depicts the stability state of Q_2. By adjusting the indicator thresholds th_{ID1} and th_{ID2} based on the *distance factors* df_1 and df_2 as follows, various structures are obtained.

$$th_{ID1} = 0.12 \cdot df_1, \quad (17)$$

$$th_{ID2} = 0.035 \cdot df_2. \quad (18)$$

Downloaded from abstracts.iovs.org on 04/19/2024

Here, df_1 and df_2 act as distance modulation factors. The core operation to simulate the distance effect in the proposed model is to change the distance factors to df_1 and df_2, respectively. Setting a larger value for df_2 simulates a longer distance between the human eyes and the objects, whereas a smaller value simulates a shorter distance. Therefore, one must preserve all of the structural information in the first step by setting df_1 as a large value. $df_1 = 1,000$ is usually enough for most patterns.

The two-reach method can be regarded as a special case of this model with specific th_{ID1} and th_{ID2}. However, since the final result of the proposed model is the combination of filled triangles, missing any triangle will make the results inaccurate. Each value of th_{ID2} will produce a series of cutting thresholds (i.e., w_2) based on the sequence Q_{ID2} [e.g., the red curve in Figure 14(b)]. Therefore, it is necessary to find several representative thresholds. As shown by the Q_{ID2} curve in Figure 14(b), most of the possible thresholds (corresponding to the locations with large fluctuations in the Q_{ID2} curve) can be clustered into several groups. DBSCAN (Ester et al., 1996) is adopted to find these groups, and the first value from each group is selected as a representative cutting threshold.

In the example presented in Figure 14, two parameters of DBSCAN are set as $epsilon = 45$ and $minPts = 15$. When the dot pattern is placed close enough to the eyes, only local features are perceived, that is, the isolated dots, rather than lines or rings. However, when the structure is placed farther away, more global properties are perceived. Therefore, from near to far, one first perceives the dot patterns, then three circles, then a disc nested in a ring, and finally a disc.

In this section, we exhibited the flexible ability of our methods to model the hierarchical representation of perception with a distance modulated example. It is also possible to replace the distance with other factors in this model to form other factor-modulated models.

For different types of grouping

Perceptual grouping is not a single process as it seems. There are various conceptual differences when referring to perceptual grouping in different research fields. One of the clear classifications was reported in Wagemans (2018), where the grouping was distinguished as *clustering*, *segregating*, *linking*, *layering*, and *configuring*. In this section, we discuss the ability of our model when applied to simulating those different types of grouping using by Equation 14.

According to the definitions in Wagemans (2018), *clustering* is "the process of treating individual items as members of a larger ensemble, basically extracting their common feature and ignoring others." *Segregating* is defined as "the process of treating one subset of items as members of a larger set, while at the same time also distinguishing this set from another subset of items." *Layering* is "the process of segregating two sets of items, with an additional indication of which subset is figure and which is ground." All can be regarded as different groups of elements owing to different features, which can be easily simulated by our model when adding visual features (e.g., orientations here) using Equation 14. Although our model can separate the different areas of discrete patterns, it lacks the necessary information to discriminate which is foreground and background, as shown by the segmentation results in Figure 10. Fortunately, leveraging some assumptions is one of the solutions and has been widely used in the field of salient object detection, such as feature contrasts (Itti et al., 1998; Cheng et al., 2014), boundary priors (Zhang & Sclaroff, 2015; Huang & Zhang, 2017), and even contour priors (Yang et al., 2016; Liu et al., 2017).

Linking is "the process of connecting individual items in specific ways, often as a sequential spreading of pair-wise couplings" (Wagemans, 2018). It has been regarded as a very general process in human perception from rather simple to higher-order cases. Finally, *configuring* is " the process of organizing individual items in larger, structured wholes or Gestalts with configural properties." Linking discrete elements into certain shapes such as a horse, circle, etc., will lead to configurations with those shapes. However, because our model is mainly designed for continuous structure reconstruction and grouping based on the global properties of relative proximity, one of the challenges is how to create such proper relative proximities between their neighbors. Consequently, our model would be able to simulate linking and configuring after extending the proposed model with proper feature representation and high-level priors (e.g., the shape coding of a horse).

Further discussion

We have proposed a computational model for dot pattern perception and shown the relevant applications in the above sections. Our model is flexible since its parameters, such as th_{ID1}, th_{ID2} in Equation 17 and Equation 18 and w_f in Equation 14, can be adjusted to produce hierarchical representations of discrete patterns and image segmentation. The parameters recommended by the authors, $th_{ID1} = 0.12$ and $th_{ID2} = 0.035$, enable the model to perform well across many tasks.

The main idea of our model is to simulate human perception to extract global information from discrete patterns, which was validated by our experiments to be consistent with human perception. In addition, PDL has refined the theory that relative proximity is closely related with the human perception of dot lattices. We have discussed that our model can be a possible extension of PDL when applied to broader

Downloaded from abstracts.iovs.org on 04/19/2024

situations. Therefore, it seems reasonable to speculate that maybe the human visual system accomplishes such tasks in a similar way, capturing global properties from relative proximity and making decision based on some thresholds. Thresholds like th_{ID1} and th_{ID2} may suggest to find the similar principles in human visual system, which provides new perspectives to this research field and is worth further study in our future work.

We have also exhibited the potential of our model for applications in data mining, computer graphics, and computer vision. Although too naive to accomplish these tasks perfectly, it still shows desirable performance in the tasks of clustering, surface reconstruction, image segmentation, edge detection, and salient object detection. In addition, to make our model more powerful across different tasks, it is also very flexible in terms of adjusting parameters and incorporating more features, including low-level and high-level information.

Conclusions

In this article, a computational model for the proximity principle for dot-pattern grouping inspired by tolerance space theory is proposed. This model includes a one-reach method, a two-reach method, and a combination process to reconstruct the continuous structure of the dot patterns. The edge ratio sequence in the so-called diagram is investigated to find a tolerance for the ETS, which serves as a threshold for removing unreasonable edges.

Experiments and quantitative evaluations on continuous structure reconstruction, proximity, similarity principle simulation, and image segmentation reveal the effectiveness of the proposed model in perceptual grouping and computer vision. The close relationship between the proposed model and the PDL is comprehensively analyzed, and the ability to adapt the proposed model to simulate distance-modulated perceptual grouping of dot patterns is discussed. Finally, the possible physiological significance and potential applications of our model have also been discussed, offering new perspectives for these research fields.

Keywords: dot patterns, perceptual grouping, computational model, one-reach method, hierarchical representation

Acknowledgments

The authors thank Accdon for its linguistic assistance during the preparation of this manuscript.

Supported by Guangdong Key R&D Project (#2018B030338001) and Natural Science Foundations of China (#61806041, #62076055).

Commercial relationships: none.
Corresponding author: Yong-Jie Li.
Email: liyj@uestc.edu.cn.
Address: MOE Key Lab for NeuroInformation, University of Electronic Science and Technology of China, Chengdu, China.

References

Achanta, R., Hemami, S., Estrada, F., & Susstrunk, S. (2009). Frequency tuned salient region detection. In *IEEE Conference on Computer Vision and Pattern Recognition.* Orlando, FL. USA (pp. 1597–1604), IEEE.

Achanta, R., Shaji, A., Smith, K., Lucchi, A., Fua, P., & Süsstrunk, S. (2012). SLIC superpixels compared to state-of-the-art superpixel methods. *IEEE Transactions on Pattern Analysis & Machine Intelligence, 34*(11), 2274–2282.

Amenta, N., Bern, M., & Eppstein, D. (1998a). The crust and the beta-skeleton: Combinatorial curve reconstruction. *Graphic Models and Image Processing, 60*(2), 125–135.

Amenta, N., Bern, M., & Kamvysselis, M. (1998b). A new voronoi-based surface reconstruction algorithm. In *Proceedings of the 25th Annual Conference on Computer Graphics and Interactive Techniques.* Orlando, FL. USA (pp. 415–421), IEEE.

Arbelaez, P., Maire, M., Fowlkes, C., & Malik, J. (2010). Contour detection and hierarchical image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 33*(5), 898–916.

Beitzel, S. (2006). On understanding and classifying web queries (ph. d. thesis). *IIT. CiteSeerX, 10*(1.127), 634.

Chan, T. F., & Vese, L. A. (2001). Active contours without edges. *IEEE Transactions on Image Processing, 10*(2), 266–277.

Chen, L. (1982). Topological structure in visual perception. *Science, 218*(4573), 699–700.

Chen, L. (2005). The topological approach to perceptual organization. *Visual Cognition, 12*(4), 553–637.

Cheng, M.-M., Mitra, N. J., Huang, X., Torr, P. H., & Hu, S.-M. (2014). Global contrast based salient region detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 37*(3), 569–582.

Claessens, P. M., & Wagemans, J. (2008). A bayesian framework for cue integration in multistable grouping: Proximity, collinearity, and orientation priors in zigzag lattices. *Journal of Vision, 8*(7), 1–23.

Downloaded from abstracts.iovs.org on 04/19/2024

Comaniciu, D., & Meer, P. (2002). Mean shift: A robust approach toward feature space analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 24*(5), 603–619.

Compton, B. J., & Logan, G. D. (1993). Evaluating a computational model of perceptual grouping by proximity. *Perception & Psychophysics, 53*(4), 403–421.

Compton, B. J., & Logan, G. D. (1999). Judgments of perceptual groups: Reliability and sensitivity to stimulus transformation. *Perception & Psychophysics, 61*(7), 1320–1335.

Delaunay, B. (1934). Sur la sphere vide. *Izv. Akad. Nauk SSSR, Otdelenie Matematicheskii i Estestvennyka Nauk, 7*(793-800), 1–2.

Ester, M., Kriegel, H. P., Sander, J., & Xu, X. (1996). A density-based algorithm for discovering clusters in large spatial databases with noise. In Proceedings of the 2nd International Conference on Knowledge Discovery and Data Mining, E. Simoudis, J. Han, and U. Fayyad (Eds.), AAAI Press, *96*(34): 226–231.

Feldman, J. (2003). Perceptual grouping by selection of a logically minimal model. *International Journal of Computer Vision, 55*(1), 5–25.

Feldman, J., & Singh, M. (2006). Bayesian estimation of the shape skeleton. *Proceedings of the National Academy of Sciences, USA*, *103*(47), 18014–18019.

Felzenszwalb, P. F., & Huttenlocher, D. P. (2004). Efficient graph-based image segmentation. *International Journal of Computer Vision, 59*(2), 167–181.

Field, D. J., Hayes, A., & Hess, R. F. (1993). Contour integration by the human visual system: Evidence for a local association field. *Vision Research, 33*(2), 173–193.

Froyen, V., Feldman, J., & Singh, M. (2015). Bayesian hierarchical grouping: Perceptual grouping as mixture estimation. *Psychological Review, 122*(4), 575.

Ghosh, S., Das, N., Das, I., & Maulik, U. (2019). Understanding deep learning techniques for image segmentation. *ACM Computing Surveys (CSUR), 52*(4), 1–35.

Grossberg, S., & Mingolla, E. (1985). Neural dynamics of perceptual grouping: Textures, boundaries, and emergent segmentations. *Perception & Psychophysics, 38*(2), 141–171.

Huang, X., & Zhang, Y.-J. (2017). 300-fps salient object detection via minimum directional contrast. *IEEE Transactions on Image Processing, 26*(9), 4243–4254.

Itti, L., Koch, C., & Niebur, E. (1998). A model of saliency-based visual attention for rapid scene analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 20*(11), 1254–1259.

Kubovy, M. (1994). The perceptual organization of dot lattices. *Psychonomic Bulletin & Review, 1*(2), 182–190.

Kubovy, M., Holcombe, A. O., & Wagemans, J. (1998). On the lawfulness of grouping by proximity. *Cognitive Psychology, 35*(1), 71–98.

Kubovy, M., & Van Den Berg, M. (2008). The whole is equal to the sum of its parts: A probabilistic model of grouping by proximity and similarity in regular patterns. *Psychological Review, 115*(1), 131.

Kubovy, M., & Wagemans, J. (1995). Grouping by proximity and multistability in dot lattices: A quantitative gestalt theory. *Psychological Science, 6*(4), 225–234.

Lee, T. S., & Mumford, D. (2003). Hierarchical bayesian inference in the visual cortex. *JOSA A, 20*(7), 1434–1448.

Lezama, J., Randall, G., Morel, J.-M., & Gioi, R. Grompone von. (2017). An unsupervised algorithm for detecting good continuation in dot patterns. *IPOL. Journal Image Processing on Line*, 2017, *7*, 81–92.

Likas, A., Vlassis, N., & Verbeek, J. J. (2003). The global k-means clustering algorithm. *Pattern Recognition, 36*(2), 451–461.

Lindeberg, T. (1994). Scale-space theory: A basic tool for analyzing structures at different scales. *Journal of Applied Statistics, 21*(1-2), 225–270.

Lindeberg, T. (1994). *Scale-space theory in computer vision* (Vol. 256). New York: Springer Science & Business Media.

Liu, Q., Hong, X., Zou, B., Chen, J., Chen, Z., & Zhao, G. (2017). Hierarchical contour closure-based holistic salient object detection. *IEEE Transactions on Image Processing, 26*(9), 4537–4552.

Marr, D., & Nishihara, H. K. (1978). Representation and recognition of the spatial organization of three-dimensional shapes. *Proceedings of the Royal Society of London. Series B. Biological Sciences, 200*(1140), 269–294.

Newman, M. (2018). *Networks (2nd ed.)*. Oxford: Oxford university press.

Otsu, N. (1979). A threshold selection method from gray-level histograms. *IEEE Transactions on Systems, Man, and Cybernetics, 9*(1), 62–66.

Pablo, A., Michael, M., Charless, F., & Jitendra, M. (2011). Contour detection and hierarchical image segmentation. *IEEE Transactions on Pattern Analysis & Machine Intelligence, 33*(5), 898–916.

Palmer, S. E. (1977). Hierarchical structure in perceptual representation. *Cognitive Psychology, 9*(4), 441–474.

Papari, G., & Petkov, N. (2005). Algorithm that mimics human perceptual grouping of dot patterns.

Downloaded from abstracts.iovs.org on 04/19/2024

Proceeding of the International Symposium on Brain Vision and Artificial Intelligence, 3704, 497–506.

Peng, B., Zhang, L., & Zhang, D. (2013). A survey of graph theoretical approaches to image segmentation. *Pattern Recognition, 46*(3), 1020–1038.

Peters, J. F., & Wasilewski, P. (2012). Tolerance spaces: Origins, theoretical aspects and applications. *Information Sciences, 195*, 211–225.

Pun, C. S., Xia, K., & Lee, S. X. (2018). Persistent-homology-based machine learning and its applications–A survey. arXiv preprint arXiv:1811.00252 [math.AT].

Rand, W. M. (1971). Objective criteria for the evaluation of clustering methods. *Journal of the American Statistical Association, 66*(336), 846–850.

Ronneberger, O., Fischer, P., & Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical Image Computing and Computer-assisted Intervention.* Munich, Germany (pp. 234–241), Springer, Cham.

Rosenholtz, R., Twarog, N. R., Schinkel-Bielefeld, N., & Wattenberg, M. (2009). An intuitive model of perceptual grouping for hci design. In *Proceedings of the Sigchi Conference on Human Factors in Computing Systems.* Honolulu, Hl (pp. 1331–1340), ACM.

Saxena, A., Prasad, M., Gupta, A., Bharill, N., Patel, O. P., & Tiwari, A. et al. (2017). A review of clustering techniques and developments. *Neurocomputing, 267*, 664–681.

Shi, J., & Malik, J. (2000). Normalized cuts and image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 22*(8), 888–905.

Sossinsky, A. B. (1986). Tolerance space theory and some applications. *Acta Applicandae Mathematica, 5*(2), 137–167.

Treisman, A. (1982). Perceptual grouping and attention in visual search for features and for objects. *Journal of Experimental Psychology: Human Perception and Performance, 8*(2), 194.

Van Oeffelen, M. P., & Vos, P. G. (1982). Configurational effects on the enumeration of dots: Counting by groups. *Memory & Cognition, 10*(4), 396–404.

Van Oeffelen, M. P., & Vos, P. G. (1983). An algorithm for pattern description on the level of relative proximity. *Pattern Recognition, 16*(3), 341–348.

Vinh, N. X., Epps, J., & Bailey, J. (2010). Information theoretic measures for clusterings comparison: Variants, properties, normalization and correction for chance. *Journal of Machine Learning Research, 11*(Oct.), 2837–2854.

Wagemans, J. (2015). *The Oxford handbook of perceptual organization*. New York: Oxford University Press.

Wagemans, J. (2018). Perceptual organization. In The Stevens' handbook of experimental psychology and cognitive neuroscience: Vol. 2. Sensation, perception & attention: 803–872. Wixted, J.T. & Serences, J. (Eds.). Hoboken: John Wiley & Sons, Inc.

Wagemans, J., Elder, J. H., Kubovy, M., Palmer, S. E., Peterson, M. A., & Singh, M. et al. (2012). A century of gestalt psychology in visual perception: I. perceptual grouping and figure–Ground organization. *Psychological Bulletin, 138*(6), 1172.

Watt, R., Ledgeway, T., & Dakin, S. C. (2008). Families of models for gabor paths demonstrate the importance of spatial adjacency. *Journal of Vision, 8*(7), 23–23.

Wertheimer, M. (1938). Laws of organization in perceptual forms. In W. D. Ellis (Ed.), A source book of Gestalt psychology (p. 71–88). London: Kegan Paul, Trench, Trubner & Company.

Wertheimer, M., & Riezler, K. (1944). Gestalt theory. *Social Research, 11*(1), 78–99.

Yang, K.-F., Li, H., Li, C.-Y., & Li, Y.-J. (2016). A unified framework for salient structure detection by contour-guided visual search. *IEEE Transactions on Image Processing, 25*(8), 3475–3488.

Zeeman, E. C. (1962). The topology of the brain and visual perception. *Topology of 3-Manifolds and Related Topics, 3*, 240–256.

Zhang, J., & Sclaroff, S. (2015). Exploiting surroundedness for saliency detection: A boolean map approach. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 38*(5), 889–902.

Zhao, H.-K., Osher, S., & Fedkiw, R. (2001). Fast surface reconstruction using the level set method. In *Proceedings IEEE Workshop on Variational and Level Set Methods in Computer Vision Vancouver, BC, Canada* (pp. 194–201), IEEE.

Zomorodian, A. (2010). Fast construction of the vietoris-rips complex. *Computers & Graphics, 34*(3), 263–271.

Downloaded from abstracts.iovs.org on 04/19/2024