Path-Based Analysis for Structure-Preserving Image Filtering

Lijuan Xu[1], Fan Wang[1], Laura Dempere-Marco[2], Qi Wang[1], Yan Yang[1], Xiaopeng Hu[1*]

(*) Corresponding authoru: Xiaopeng Hu (xphu@dlut.edu.cn)
[1] School of Computer Science and Technology, Dalian University of Technology, Dalian 116024, China
[2] Department of Engineering, University of Vic-Central University of Catalonia, 08500, Vic, Barcelona, Spain

Abstract

Structure-preserving image filtering is an image smoothing technique that aims to preserve prominent structures while removing unwanted details in natural images. However, relevant studies mainly focus on small variances/fluctuations sup- pression and are vulnerable to separate pixels connected by some low-contrast edges or cluster pixels which exhibit strong differences between neighbors in highly textured region. Inspired by the fact that the human visual system significantly outperforms manually designed operators in extracting meaningful structures from natural scenes, we present an efficient structure-preserving filtering method which integrates similarity, proximity and continuation principles of human perception to accomplish high-contrast details (textures/noises) smoothing. Additionally, a Liebig's law of minimum-based distance transform is presented to seamlessly incorporate the three properties for the description of the filter kernel. Experiments demonstrate that our distance transform keeps a clustering-like manner of separating different image pixels and grouping similar ones with the awareness of structure. When integrating this affinity measure into the bilateral-filter-like framework, our method can efficiently remove high-contrast textures/noises while preserving major structures.

Keywords Structure-preserving image filtering, Distance transform, Gestalt principles of grouping

1. Introduction

In natural images, there exist various objects and back- grounds which truly represent rich visual information about the scenes of real world. However, noises, clutters and textures that will weaken the capability of image processing are also contained in these images. So as to eliminate these details without distracting the significant structure in the images, structure-preserving filtering method has attracted tremendous attention applied as a pre-processing step for many tasks in the field of computer vision and graphics, such as image abstraction/vectorization, image segmentation, image enhancement and content-aware image editing. Historically, Gaussian filter is the oldest and most commonly used method for filtering details. Nevertheless, it usually leads to blurs along sharp edges for only consider- ing the spatial relations between pixels at a constant rate in all directions. Bilateral filter [1] is then introduced to improve the blurring edges by controlling the kernel direction with the help of the partial image content. That is, the feature difference is utilized as a prior edge information to choose nearby similar pixels for averaging. As a result, dis- tinct edges are preserved in the filtering process. Bilateral- filter-like approaches, such as anisotropic diffusion filter [2], weighted least square filter [3] and L0 smoothing [4], commonly employ the contrast determined by the feature difference or gradient as the confidence of identifying edges and use this evidence to guide the filtering. However, textures/noises usually do not correspond to the low gradients/ contrast. In this case, the contrast-based definition of edges might fail to capture high-contrast textures that are related to fine-scale image details. In fact, the bilateral filter and its analogous methods cannot fully separate textured regions from the main structures as they regard the high-contrast texture as part of structure to be respected in the image.

To address it, the total variation regularization filter and its extending versions [5, 6] are proposed on the basis that the total variation of the high-contrast details is distinctly higher than that of the structures in the image. This kind of structure-preserving filtering technique can produce impressive results for highly textured images, but it is typically slow. Based on the assumption

that there exists a local linear model between the guidance image and the filtering out- put, He et al. [7] propose the guided filter to preserve the structures by taking the content of the guidance image into account. In addition, scale measurement [8–11] is introduced to distinguish structures from textures by observing that the two image components have showed large disparities within scales. However, designing a favorable scale-aware measurement is surprisingly difficult, due to the arbitrary scales, uncertain distributions of objects and textured background in natural scenes.
In the human visual system, it shows strong capability to effortlessly and efficiently separate meaningful structures blended with or formed by texture elements [12]. Encouraged by this mechanism, we believe that a further combination of the middle-level analysis and image filter- ing is a practical way to guide the decomposition of high- contrast details and prominent structures. Given that various approaches have been investigated for the awareness of visual content in recent years, for example, local-distribution-based filters [13–15] take the distribution or statistics of neighboring pixels into consideration, however, they may face some deviations from the original edges as a result of ignoring the global image geometric structures. Additional similarity metrics like geodesic or diffusion distance [2, 16, 17] instead of traditional Euclidean can enhance the ability of texture–structure separation to some extent; however, the two kinds of metrics may be disturbed in highly textured regions caused by the distances themselves.

In this paper, the Gestalt principles of grouping [18, 19] in human perception are introduced to better explore the structural relations for filtering the high-contrast details. Gestalt is a psychology term which refers to the principles of visual perception, and the gestalt theory attempts to describe how people reconstruct visual elements and organize them into groups or unified wholes with the use of certain laws, such as similarity, proximity, continuation, closure, symmetry and figure-ground. Specifically, the principle of similarity states that elements which share similar properties such as shape, color and orientation will be seen as belonging to each other, the principle of proximity means human beings tend to regard visual elements close to each other in spatial domain as a whole, and the principle of continuation holds that there is an innate tendency to perceive elements that are connected without abrupt changes as a perceptual whole in the human visual system. For bilateral-filter-like methods, they usually apply similarity and/or proximity principles to weight the filter kernel [1–4], and this may lead to some failures if there exist long-range texture regions in the image. The reason is that the kernel functions estimated by the similarity and/or proximity principles may miss some visual elements related to the continuation principle in the highly textured region for smoothing. Based on this, we propose a novel structure-aware method which includes similarity, proximity and continuation principles for filtering textures as well as small variations. In order to better explore these properties, we first regard the input image as an undirected graph, where the nodes are the pixels of the image, and the edges reflect the proximity between locally pairwise pix- els. By following the similarity, proximity and continuation principles between nodes on the graph, we perform analysis of the clustering path generation and its nonlinear distance transform to estimate the filter kernel. Experimental results demonstrate that our distance transform enables a cluster- ing-like-manner of minimizing intra-cluster dissimilarities and maximizing inter-cluster ones regardless of scales and distributions of the objects. When further integrated into a bilateral-filter-like framework, our filtering method is capable of extracting structures from natural scenes with highly textured regions.

In conclusion, the main contributions of our method are summarized as follows: (1) based on the similarity, proximity and continuation of the Gestalt principles of group- ing in human perception, we define the bottleneck distance transform based on the clustering path and highlight the visual relations between image elements. (2) We propose a structure-preserving filtering method, where the proposed minimum bottleneck distance transform is integrated into a bilateral-filter-like framework for defining the kernel weight. The experimental results demonstrate the performance of applying the proposed filtering method to remove high-contrast textures and noises, and preserve major structures in the image.

2. Related Works

The basic strategy of edge-preserving image decomposition is to perform image smoothing and edge detection jointly. The unified formulations, differing in how to define edges and how the edge prior guides the filtering, aim to decompose a given image into structure and details. According to the definition of the detail, the image filtering approaches can be categorized into low-contrast and high-contrast ones. The low-contrast details mainly contain small variations or clutters in the image, while the high-contrast ones correspond to textures and noises.

1. Low-contrast details filtering

Among all edge-preserving filtering methods proposed in recent years, the bilateral filter introduced by Tomasi and Manduchi [1] is likely to be the most popular one for its simplicity and effectiveness. Bilateral filter introduces both the spatial distance and color difference, respectively, estimated by the Gaussian kernel function, to determine the filtering weight. When the kernel is large, the brute-force implementation of the bilateral filter is very slow. Thus, various strategies [20–24] based on quantization, down-sampling or additional constraints on the spatial or range domain are proposed to accelerate the bilateral filter.

Theoretically, the bilateral filter is designed by a kind of filter kernel to measure the distance between two pixels in a local region, and this distance is then converted to the confidence of the edge awareness. Some studies attempt to improve this confidence by using better similarity metrics, such as geodesic [16] or diffusion [2, 17] distances instead of traditional Euclidean distances. More methods tend to pro- pose or introduce all kinds of theories to extract edges from the image, including anisotropic diffusion [2], weighted least square filter [3], edge-aware wavelets [25], guided image filter [7], non-local means filter [26], local Laplacian pyramid [27], L0 smoothing [4] and domain transform filter [28] and so on. Anisotropic diffusion [2] employs the partial differential equation (PDE)-based formulation in which pixel- wise diffusivities are estimated from image gradients. These diffusivities prevent smoothing at image edges and aid to preserve important structures while eliminating noise and fine details. To balance the degree of filtering by forcing the image to be filtered except at regions with large gradients, Farbman et al. [3] define an edge-preserving objective function using the image gradients and regularize it with the weighted least square optimization. Fattal [25] achieves a fast filtering approach by using edge-avoiding wavelets, but the results seem noisy in most cases. Paris et al. [27] perform the edge-preserving filtering operation based on local Laplacian pyramid and demonstrate that their method can avoid artifacts over edges. Xu et al. [4] introduce a global L0-based optimization framework which uses the number of image pixels with nonzero gradient magnitudes as a regularization constraint. Since these methods derive from image gradients and aim to remove details with small gradients, most salient edges can be preserved or even enhanced in images.

Although the filtering responses differ from each other, the aforementioned bilateral-filter-like studies can effectively filter out low-contrast details in the image. However, they are very sensitive to high-contrast details (noises or textures) since they commonly use feature (color, intensity and brightness) differences or gradient magnitudes to distinguish details from structures. For some tasks where high-contrast details need to be removed, such kind of operator usually results in unsatisfactory image decompositions.

2. High-contrast details filtering

The existing edge-detection-based filtering methods usually regard image edges signified by high image gradients as structures and may lead to the problem of falsely preserving textures or noises with large gradients. In order to filter such kind of high-contrast details from images, local-distribution- based filters [13–15] look into the distribution or statistics (local histogram, second-order statistics) of neighboring pixels rather than gradients or contrast. Although these filters can generally produce more smoothing results with sharp edges, they often face a problem of serious deviation from the original edges (especially at corners) since local histogram completely ignores image geometric structures.

The total variation model, one popular structure-preserving regularization framework, uses L1-norm-based regularization constraints to penalize large gradients and demonstrates fairly good

separation for structures from textures. Some studies extend this standard TV formulation with different norms. Buades et al. [5] propose a local total variation method and use a nonlinear low-pass and high-pass filter pair to decompose an image into structure and oscillatory texture. Based on the key observation that the aggregation result of signed gradient values in a local window often has a large absolute value for major edges than for textures, Xu et al. [6] propose a relative total variation (RTV) measure to better capture the difference between texture and structure and then develop an optimization system to extract main structures. Their method can produce impressive results for highly textured images (especially for mosaic images or graffiti on textured materials), but it may excessively smooth the natural images.

Scale measurement is another way to distinguish the structure from texture/noises. Subr et al. [8] explicitly point out that details should be identified with the respect of the spatial scale, not the feature contrast. They then separate oscillations from the structure layer by constructing local extremal envelopes. However, simply employing the average of the maximal and minimal envelops to filter the natural image might suffer from great weakness when the image regions contain both textures and meaningful structures. Zhang et al. [9] propose a fast rolling guidance filter by designing one unique scale measure to control the level of details during filtering. This scale measure is useful when manual adjustment is required. Semantic filtering [10] seamlessly combines the recursive filtering and the learning- based edge classification together for fast scale-aware filter- ing regardless of the scales of the objects. Bao et al. [11] attempt to use the connectedness represented on a minimum spanning tree (MST) to distinguish small connected components (details) from large connected ones (major structures).

3. Path-Induced Distance Mapping

3.1. The Proposed Distance Transform

Given that distance transform (DT) [29, 30] has been widely used as an effective tool for analyzing object geometry and morphology, we introduce this notion to exploit the structural connection of the image in this paper.

3.1.1. Graph Construction

By treating an image I as a standard 4-connected, undirected grid (see Fig. 1b) with nodes being all the image pixels and edges between nearest neighboring pixels weighted by feature difference, a sparsely weighted graph G = (V, E) is constructed to represent the local relations between pixels [31, 32]. Specifically, the principles of proximity and similarity are coded in the graph by the edges and corresponding weights.

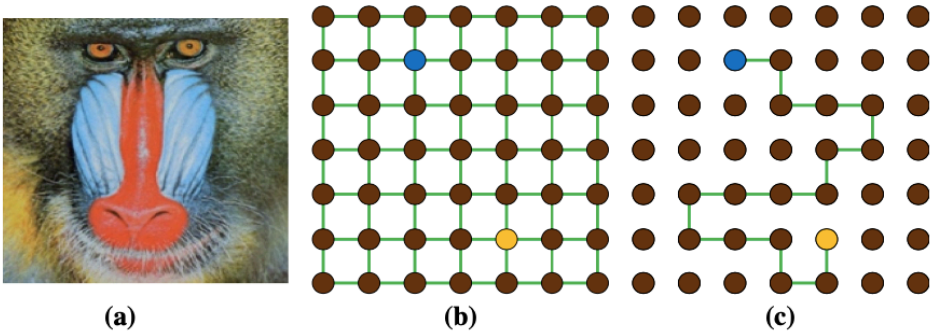


Fig. 1 (a) The image I, **(b)** shows its corresponding graph G, and **(c)** signifies one path on the graph with the end points highlighted by the blue and yellow color.

3.1.2. Distance Transform

In the field of image processing, the distance transforms commonly take the image data into consideration. For most methods, the distance between each pair of image elements (pixels, superpixels, regions) is usually defined as the minimum cost of all paths connecting them, and the path cost functions for the paths are task dependent and have achieved satisfactory performances in segmentation and detection related applications.

Let any sequence of vertices $\pi = <\pi(0), \pi(1), \pi(2), \ldots, \pi(k)>$ denote the path from s to v when $s = \pi(0)$ and $v = \pi(k)$ in the graph $G = (V, E)$ (see. Fig. 1c), and for each $i \in \{1,2,...,k\}$, $\pi(i) \in V$, $\{\pi(i-1), \pi(i)\} \in E$. Given a path cost function $f(\pi)$ associated with any path π in the graph, the corresponding distance transform is defined as

$$DT(s, v) = \min_{\pi \in \prod_{x,y}} f(\pi) \tag{1}$$

where $\prod_{S,v}$ denotes a familiy of all paths in G from seed pixel s to v. Usually, the definition of the path cost function $f(\pi)$ satisfies the following three properties.

a. Symmetry $f(u \to v) = f(v \to u)$
b. Nonnegative $f(u \to v) \geq 0$
c. Inequality $f(u \to v) \leq f(u \to t) + f(t \to v)$

where u, v and t are any nodes in the graph G. In fact, the path cost function $f(\pi)$ can be defined according to the weights of arcs or vertexes. The geodesic distance [16] accumulates the weights of edges for all traversed pairs of nodes on the path. To handle uncertainty caused by sampling artifacts, illumination inhomogeneities in the image representation, the distance transform defined on fuzzy subsets has gained a lot attention [33, 34]. The fuzzy distance [33] is a measure of traversing the graph with minimal "local material density," and the fuzzy connectedness [34] uses the weakest link of the path to define the path strength. The watershed segmentation algorithm [35] applies the maximal weight of the node on the path to define the path cost function. For the image foresting transform (IFT) [36], Falcao et al. state that the cost of a path in an oriented forest with all paths from the seed set to each pixel can be determined by any smooth path cost function (additive, max-arc, monotonic-incremental). Strand et al. propose the minimum barrier distance and formulate the barrier as the maximum of interval values among all nodes along the path [29, 30]. To sum up, different forms of edge or vertex weights have been studied to define the path cost functions, that is,

1. The arc method

For geodesic distance and IFT, the sum strategy is used,

$$f_{\text{sum,arc}}(\pi) = \sum_{i=0}^{k-1} |I(\pi(i+1) - I(\pi(i)|| \tag{2}$$

For fuzzy connectedness and IFT, they estimate the maximum weight of the edges,

$$f_{\text{max,arc}}(\pi) = \max_{i=0}^{k-1} |I(\pi(i+1) - I(\pi(i)|| \tag{3}$$

2. The vertex method

For watershed, the maximum of vertex value is used,

$$f_{\max,\text{vertex}}(\pi) = \max_{i=0}^{k} I(\pi(i)) \tag{4}$$

For fuzzy distance, the average of vertex values is applied

$$f_{\text{average,vertex}}(\pi) = \sum_{i=0}^{k-1} \frac{I(\pi(i+1) + I(\pi(i))}{2} \tag{5}$$

For minimum barrier distance, the maximal interval of vertex values is calculated,

$$f_{\max,\text{interval}}(\pi) = \max_{i=0}^{k} I(\pi(i) - \min_{i=0}^{k} I(\pi(i)) \tag{6}$$

The aforementioned path cost functions $f(\pi)$ have achieved varying degrees of success in measuring the connection values. However, they are sensitive to noises and textures in natural images. Under the geodesic and fuzzy distance frameworks, the path cost will strictly increase as the path grows, leading to false accumulation of small variations in cluttered regions. For the methods of watershed and the minimum barrier distance, the path cost may remain constant during the growth of the path until a new strong barrier is met on that path. Nevertheless, they may exaggerate the small difference in smooth but noisy regions since it they use the maximum value of vertex or the maximal interval of vertex values along the path.

3.1.3 The Clustering Path-Based Distance Measure

Based on the assumption in data clustering that two objects which are assigned to the same cluster are either similar (compactness) or with high probability there exists a path from s to v over other mediating objects of this cluster where the dissimilarities of two consecutive objects are small (connectedness) [37], we extend this intra-cluster path to explore the Gestalt principle of continuation hidden in the image and call it the clustering path (*c*-path), that is,

$$\exists \pi_{\text{cluster}} = < \pi(0), \pi(1), \pi(2), \ldots, \pi(k) > , \varepsilon \to 0$$

$$\text{for } \forall i \in \{0,1,...,k-1\} \tag{7}$$

$$\pi_{\text{cluster}} \in \prod_{s,v} \wedge \, | I(\pi(i+1) - I(\pi(i)) | < \varepsilon$$

Equation (7) formulates that the clustering path is deter- mined by the constrain $|I(\pi(i+1) - I(\pi(i))| < \varepsilon$, which indicates that the differences between successive objects of the path do not exceed a given threshold value [38]. This idea is equivalent to the single linkage clustering method [39] described in pattern recognition and then renamed as the ε-connected components in image partitioning [40] or the quasi-flat zones in mathematical morphology [41].

As indicated in [39, 40], Eq. (7) is a less restrictive constraint to represent the connectivity relation between two objects and may suffer from the chaining effect. That is, if two distinct objects are separated by one or more transitions going in steps with a weight less than or equal to ε, they may falsely appear within the same cluster. Since then, more strong connectivity constraints methods such as global range parameter [42], gray-scale blobs [43] and connectivity index [40] have been proposed to limit the connection relations from the global perspective in the connected component. If we can limit the differences between every pair of nodes on the path to be smaller than a given threshold, Eq. (7) becomes

$$\begin{aligned} &\exists \pi_{\text{cluster}} = < \pi(0), \pi(1), \pi(2), \ldots, \pi(k) > , \varepsilon \to 0 \\ &\forall i, j \in \{0,1,...,k\} \\ &\pi_{\text{cluster}} \in \prod_{s,v} \wedge \, | I(\pi(i) - I(\pi(j)) | < \varepsilon \end{aligned} \tag{8}$$

Equation (8) defines a more restrictive clustering path named *rc*-path. Inspired by the Liebig's law of minimum [44] in agricultural science that the capacity of a barrel with staves of unequal length is limited by the shortest stave, we apply the minimal similarities between the pair of nodes on the path to represent Eq. (8). Specifically, if the maximum edge weight of the path that connects two similar nodes can be limited to a certain value ε, we will get a favorable clustering path. In such way, Eq. (8) becomes

$$\begin{aligned} &\exists \pi_{\text{cluster}} = < \pi(0), \pi(1), \pi(2), \ldots, \pi(k) > , \varepsilon \to 0 \\ &\pi_{\text{cluster}} \in \prod_{s,v} \wedge (\max_{i=0}^{k-1} | I(\pi(i+1) - I(\pi(i)) |) \\ &< \varepsilon \wedge (\max_{i=0}^{k} I(\pi(i)) - \min_{i=0}^{k} I(\pi(i)) < \varepsilon \end{aligned} \tag{9}$$

It is obvious that the minimal clustering path (mc-path) described in Eq. (9) is determined by two constraints. The first one $(\max_{i=0}^{k-1} | I(\pi(i+1) - I(\pi(i)) |) < \varepsilon$ means that the differences between consecutive objects (the edge weight) cannot exceed a fixed value, and is used to represent the connectedness. The second one $(\max_{i=0}^{k} I(\pi(i)) - \min_{i=0}^{k} I(\pi(i)) < \varepsilon$ indicates that if π_{cluster} is within a cluster, the maximal differences between the objects should be lower than the pre-defined value ε.

From the perspective of the distance transform, Eq. (9) can be rewritten as

$$DT(s,v) = \min_{\pi \in \prod_{x,y}} \{ (\max_{i=0}^{k-1} | I(\pi(i+1)) - I(\pi(i)) | \odot (\max_{i=0}^{k} I(\pi(i)) - \min_{i=0}^{k} I(\pi(i)) \} \tag{10}$$

where $\odot$ is an operator to compute the cost on each path π. We can use multiply, sum or other complex operators to define it. In fact, Eq. (10), the distance transform based on the minimal clustering path, can be defined as the *mc*-DT. Usually, the chaining effect [39, 40] occurs when the transitions between consecutive objects go in steps toward the same gradient direction along the path. As a result, the difference between the endpoints of the path will be larger than all edges' weight due to the accumulation of the transitions between the mediating nodes on the path. To this point, it is with high

probability that the nodes with the maximal and minimal values will appear at the end point of the path ($\pi(0)$ or $\pi(k)$). Then, Eq. (10) can be simplified as

$$DT(s,v) = \min_{\pi \in \prod_{x,y}} \{(\max_{i=0}^{k-1} |I(\pi(i+1)) - I(\pi(i))|) \odot (|I(\pi(0)) - I(\pi(k))|\} \tag{11}$$

We can apply Eq. (11) (simplified *mc*-DT) to explore the Gestalt principle of connectedness hidden in the image for image filtering methods. If image pixels are similar, close and connected to each other, they will share low $DT(s,v)$ values. If two neighboring pixels s and v have similar features, the corresponding transform value $DT(s,v)$ will be small. If similar pixels s and v belong to disjoint regions, $DT(s,v)$ will be large. Given that the bilateral-filter-like framework [1, 11] has regarded $|I(\pi(0)) - I(\pi(k))|$ as an individual factor to measure the similarity between image pixels, so we only use the maximum edge strength of each path to define the connectivity in this paper:

$$DT(s,v) = min_{\pi \in \prod_{x,y}} \{(\max_{i=0}^{k-1} |I(\pi(i+1)) - I(\pi(i))|). \tag{12}$$

Due to the usage of the maximum strategy to define the distance, the path cost function in Eq. (12) is similar to the l_∞ distance between a pair of pixels or the function applied in fuzzy connectedness. For the notion of the maximum strength on a path, we redefine it as a bottleneck for describing the limitation specified by the Liebig's law of minimum. Consequently, we call the distance transform defined in Eq. (12) the minimum bottleneck distance in the rest of the paper, for representing the gestalt principle of connectivity between each pair of nodes in the graph.

3.2 Analysis

3.2.1 Computation Analysis

For generating a clustering path between a pair of nodes, the obvious way is to traverse the whole graph to find all paths that connect the two nodes, and then select one with the minimum largest edge strength. However, the computation load is high. As proved by J. Gower and G. Ross [39], the path for a pair of nodes on the MST is a single linkage clustering path for the two nodes on the graph. Thus, the problem of traversing all paths can be solved by extracting the MST for the graph [45].

To construct the corresponding MST on the graph, we use the priority queue data structure to efficiently conduct the Prim's algorithm [46], as indicated in [11]. The structure is based on the image with pixels changing from 0 to 255 and includes a bitset and 256 doubly linked lists. The bitset has a size of 256 with bit value 0 or 1 and is used for tracking the minimal edge cost on the graph. Each doubly linked list i saves the edges with weight i. By applying the structure, the Prim's algorithm runs in the linear time with O(|E| + |V|). By constraining the image to be 4-connected, undirected graph, the time complexity of the Prim's algorithm runs in O(|V|) and is linear to the number of pixels in the image.

3.2.2 DistanceAnalysis

To highlight the results of the proposed distance transform in measuring the topological connectivity regardless of noises and textures, we conduct the comparisons of different path cost functions on the MST instead of the weighted graph, due to the only one path between each pair of nodes on the MST. In this section, Prim's algorithm [46] is applied to construct the tree from the graph.

Taking efficiency and visualization into consideration, we use the superpixels [47] instead of pixels to construct the undirected graph for the image. In the graph, each super- pixel is only connected with its spatial neighbors, and each edge is weighted by the color difference between the mean values of the two superpixels in the CIE-Lab color space. Then, we can achieve distance transforms by applying the different path cost functions on the generated MST, and the resulted affinity matrix D represents the relations between each pair of superpixels in the image. In order to visualize these distance transforms, we introduce the MDS (multidimensional scaling) method to project the D into a three-dimensional

feature space (x, y, z), which is normalized into [0, 255]. The normalized coordinates (x, y, z) are then used as the RGB values (i.e., R=x,G=y and B=z) to fill each superpixel in the image. The pixels painted by similar colors in the image describe the close relations in the feature space. Algorithm 1 demonstrates the pseudo-code of generating the mapping images.

Algorithm 1: Pseudo-code for different distance transforms on MST

Input: Original image I

Output: the mapping image

---

(1) Use the SLIC algorithm [57] to segment the image I into N superpixel;

(2) //Construct the weighted graph G

for each superpixel i ($1 \le i \le N$)

Add edge $e_{i,j}$ between superpixel i and its spatial neighbor j;

Calculate the weight w_{ij} of edge e_{ij}, $w_{ij} = \|g(i)-g(j)\|_2$, $g(i)$ is the mean value of i in CIE-Lab space;

end for

(3) Apply the Prim's algorithm to build an MST from G;

(4) //Generate the distance matrix D for each pair of nodes on MST

for each pair of nodes A and B on the graph,

Find the path $\pi = \langle \pi(0), \pi(1), \pi(2), ..., \pi(k) \rangle$ on the MST between the two nodes;

Apply (2), (5) and (6) to calculate the geodesic, fuzzy and minimum barrier distance respectively;

$d_{tree}(A,B) = k$; $d_{bottleneck}(A,B) = \max_{i=0}^{k-1} w_{i,i+1}$

end for

(5) Use the MDS method to project D into three-dimensional feature space (x, y, z);

(6) **for** each superpixel i ($1 \le i \le N$),

Paint it with $R = x$, $G = y$, and $B = z$ in RGB space;

end for

In this section, the traditional Euclidean distance is applied as a comparison for estimating the dissimilarities between image elements, even though it only considers the similarity law. Figure 2 shows the visualization of the mapping results generated from the Euclidean, geodesic, fuzzy, minimum barrier, distance [11] and our minimum bottleneck distances on synthetic and natural images. By visual assessment, we note that the geodesic distance and fuzzy distance may accumulate the transitions in the homogenous image region, resulting in inconsistent description for the same or connected region (the synthetic texture image on the first row); the minimum barrier distance may describe disjoint regions with similar color (the two blue regions around the nose for the mandrill on the fifth row) due to the ignorance of the mediating variations along the path, the tree distance can lead to different representation for adjacent image elements with similar features on account of the excessively long tree path (the sky region on the eighth row). All in all, it is clear that the proposed distance transform follows the Gestalt grouping laws to represent the relationships between image elements regardless of textures and noises. That is, the minimum bottleneck distance can achieve consistent description for image elements in the same or connected region. Image elements that are similar and adjacent share the same representation, and similar image elements that belong to disjoint regions have different descriptions.

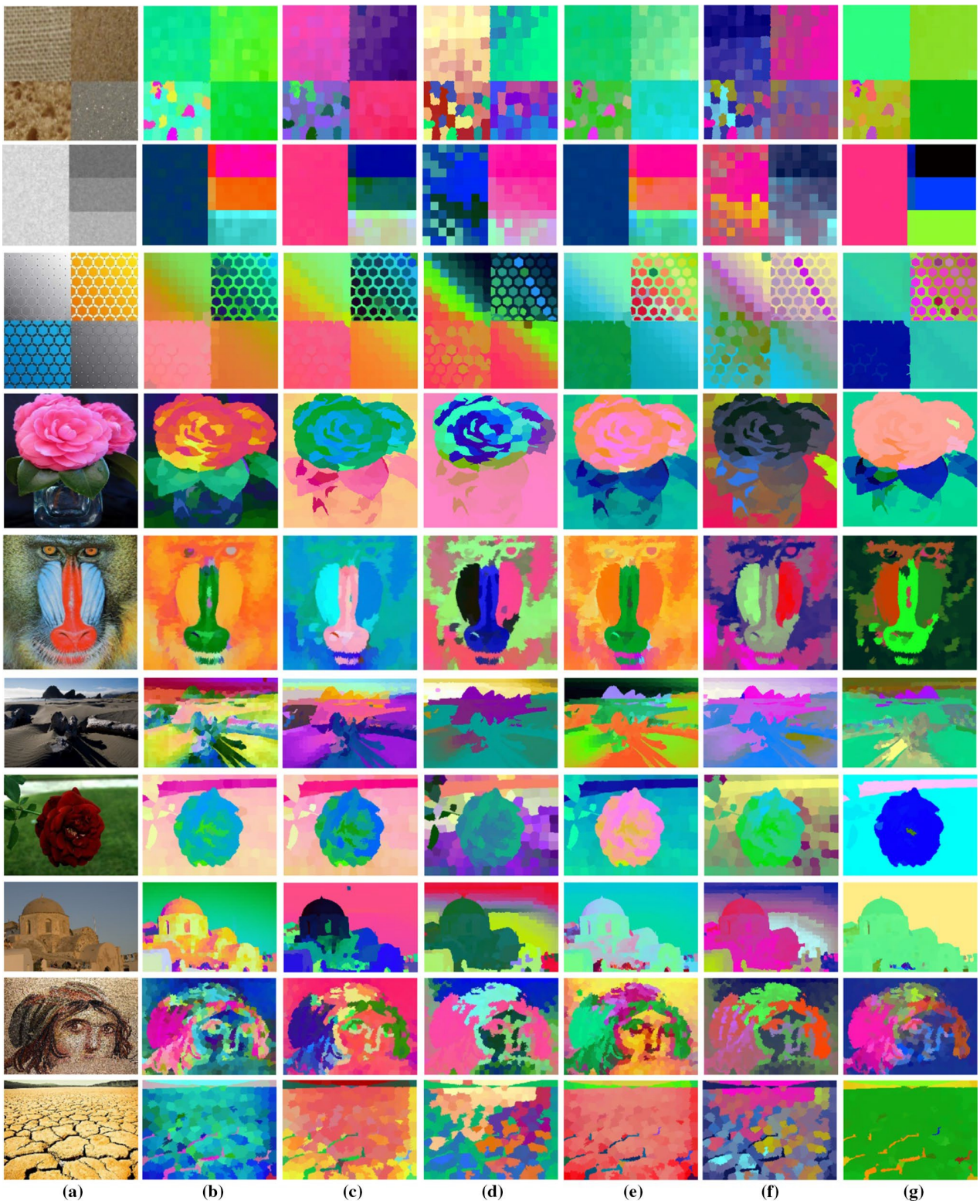


Fig. 2 The mapping results attained from different distance transforms. **a** is the original image, and **b–g,** respectively, show the description achieved by Euclidean distance, geodesic distance, fuzzy distance, minimum barrier distance, tree distance and our minimum bottleneck distance

3.2.3 Leak Problem of MST

Given the facts that MST can automatically drag away dis- similar pixels that are close in the spatial domain, and connect the small isolated region to its surrounding, Bao et al. [11] use the tree distance (the length of the path on the MST) to define the weight for perceive the structures from all edges in the image. However, this kind of edge-aware metric may face “false edges” (refers to large tree distance for pixels belonging to the same region) and “leak problem” (nearby dissimilar pix- els may have short tree distance due to the forced connection on the MST) [11, 48, 49].

To solve the problem of leak issue, Danda et al. [48] introduce the UMST filter by extracting all MSTs, instead of one MST from the graph to define the weight, and prove that the UMST filter is the limit of filtering methods based on the shortest path (the geodesic filter [16], morphological amoebas [50]). However, the shortest distance for each pair of nodes is based on the accumulation of the weights of edges for the two nodes on the shortest path and may lead to inconsistent distance for nodes in the same region.

For illustrating the difference between the tree, the shortest and the minimum bottleneck distances in dealing with false edges and leak problem, we apply the ranking map, where the value of each node refers to the distance from the root node (the top-left image superpixel). To calculate the shortest distances between the root node and other nodes in the graph, we first generate the shortest path between them and then accumulate all edge’s weights on the shortest path. For the tree and the minimum bottleneck distances, we first extract the corresponding MST from the graph and regard the node on the top-left of the image as the root of the tree. For the paths between the root and other nodes on the MST, we, respectively, compute the number of edges and the maximum edge weight of the path. These values are then normalized into [0, 1] and assigned to each node in the image.

The ranking values can serve as a good visualization tool for inspecting the difference between tree distance, the shortest distance and the minimum bottleneck distance. As shown in Fig. 3, our proposed distance transform can deal with the false edges and leak problem caused by the tree distance without missing the advantages that the tree distance has demonstrated. For each pair of pixels in the same cluster with large tree distance, their minimum bottleneck distance is small as we do not accumulate the length of the path. For nearby dissimilar pixels connected by short tree distance, we use the feature affinity to estimate the distance and the result presents the large difference between them. Compared with the shortest distance, the proposed distance transform can produce consistent ranking values for regions with changing variations. In conclusion, our minimum bottleneck distance shows promising performance in edge-aware than the tree distance.

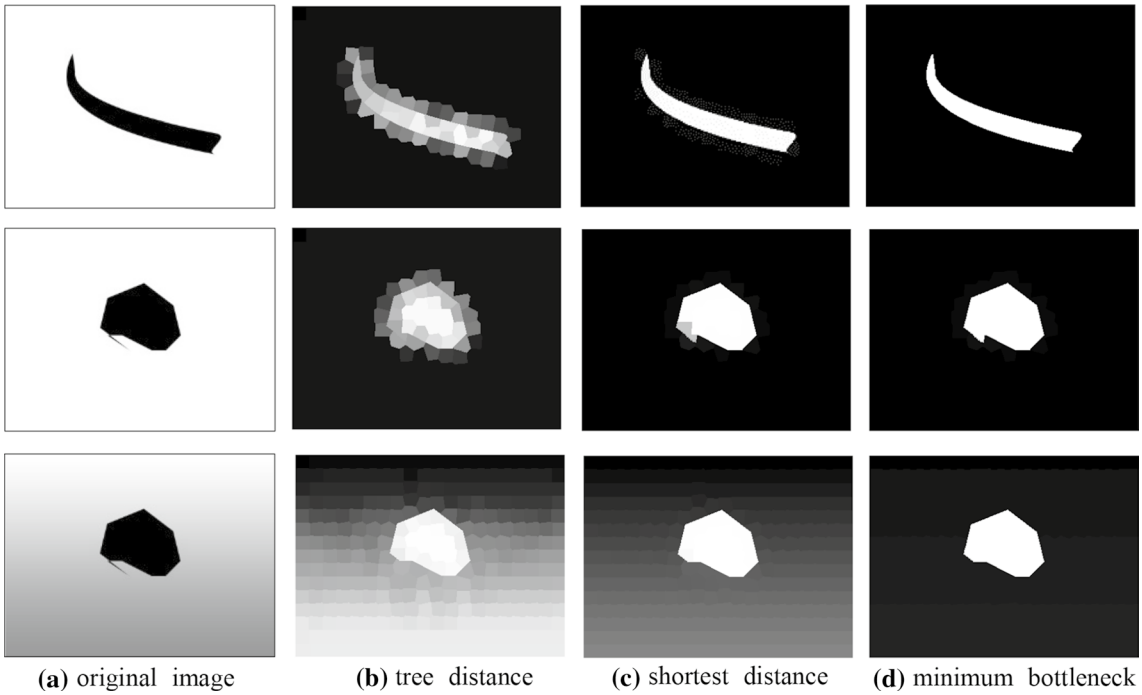


Fig. 3 (a) Three images with homogeneous image regions. The gray-scale value of a pixel in **(b)**, **(c)** and **(d)** shows its ranking value, a distance from the pixel to the top-left corner along a path. b–d are generated by applying the tree distance along the MST path, the geodesic distance along the shortest path and the minimum bottleneck distance along the MST path, respectively.

4 Structure-Preserving Image Filtering

Bilateral filter typically defines a simple monotonicity of the Euclidean distance between pixels in spatial and color spaces to calculate the similarity between low-contrast details and structures. However, this commonly used pairwise distance provides only partial information about the manner in which pixels should be clustered/ separated (highly textured regions should be clustered together for filtering, not separated), and does not work well when high-contrast details need to be smoothed. As demonstrated in Sect. 3, our minimum bottleneck distance, which accounts for gestalt grouping laws between the data points in the feature space, can serve as an effective structure-aware metric to estimate pixel affinity with a valuable global perspective of the clusters present in the feature space. Inspired by this, we aim to explore an automatically edge-aware filtering method by further integrating this high-level image understanding into the bilateral-filter-like manipulation. In order to reliably incorporate the minimum bottleneck distance into the definition of the filter kernel, we introduce the concept of trust mechanism in the wireless sensor network, where the behavior of a node is predicted by collecting preferences of other nodes with similar behaviors [51].

4.1 Definition

For each pixel i in image I, the output Si of the bilateral-filter- like method is calculated by

$$S_i = \sum_{j \in \Omega} w_i(j) I_j \tag{13}$$

In this paper, Ω is the set of all pixels in the whole image and $w_i(j)$ is the trust weight of pixel j contributing to i, which is defined as

$$w_i(j) = \sum_{k\in\Omega} b_j(k)c_k(i)$$

$$b_j(k) = \frac{G_{\sigma_s}(|j-k|)G_{\sigma_\tau}(|I_j - I_k|)}{\sum_{p\in\Omega} G_{\sigma_s}(|k-p|)G_{\sigma_\tau}(|I_k - I_p|)} \tag{14}$$

$$c_k(i) = \frac{G_\sigma(DT(k,i))}{\sum_{q\in\Omega} G_\sigma(DT(i,q))}$$

where $b_j(k)$ is the bilateral weight of pixel j estimated from its neighboring pixel k, $c_k(i)$ is the cluster weight of pixel k contributing to the center pixel i, and $G_{\sigma_s}(\cdot)$, $G_{\sigma_\tau}(\cdot)$ and $G_\sigma(\cdot)$ are three Gaussian weighting functions in spatial, range (color) and transform (minimum bottleneck distance) spaces, respectively.

In addition, it can be proved that sum of all colaborative filtering weights $\sum_{j\in\Omega} w_i(j)$ for each pixel i equals 1. According to Eqs. (13) and (14), we know that $\sum_{j\in\Omega} w_i(j) = \sum_{j\in\Omega}\sum_{k\in\Omega} b_j(k)c_k(i)$. Since $c_k(i)$ is uncorrelated with j, $\sum_{j\in\Omega}\sum_{k\in\Omega} b_j(k)c_k(i)$ can be rewritten as $\sum_{k\in\Omega} c_k(i)\sum_{j\in\Omega} b_j(k)$. In Eq. (14), Gaussian function is applied to define $c_k(i)$ and $b_j(k)$. In other words, we can get $\sum_{k\in\Omega} c_k(i) = 1$ and $\sum_{k\in\Omega} b_j(k) = 1$. Therefore, $\sum_{j\in\Omega} w_i(j) = \sum_{k\in\Omega} c_k(i)\times 1 = \sum_{k\in\Omega} c_k(i) = 1.$

4.2 Explanation

For the filtering weight from each pixel j to center pixel i, previous methods directly apply the relationship/affinity between the two pixels, and they are sensitive to noises or textures. In this paper, we apply a collaborative filtering method which uses as-many-as-possible neighbors of pixel j to jointly determine its contribution to the weight of center pixel i. Specifically, to compute the contribution from pixel j to center pixel i, we first choose neighbors of j. If j has close relations with its neighboring pixel k, we then adopt the contribution of k to i, and vice versa. Finally, the selected neighboring pixels k are joined together for measuring the weight of pixel j to center pixel i. The bilateral filtering approach is used to select nearby similar pixels, and this bilateral weight $b_j(k)$ is attenuated with the increase of either spatial or range distance. Cluster weight $c_k(i)$ is used to decide whether the neighbors of pixel j and the center pixel i follow the same geometric manifold. If they belong to the same cluster, the filtering confidence will be strong. As shown in Fig. 4, we note that the contribution from j to i is not simply achieved from pixel j itself, but indirectly deduced by combining all the decisions from its similar neighbors. The reason is that we will accept the final decision only when the majority of neighbors follow the same manifold with the center pixel.

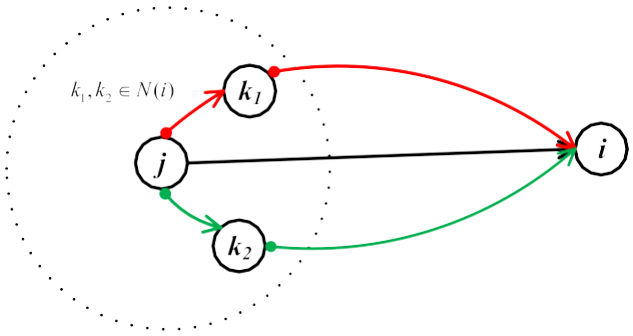


Fig. 4 Illustration of the collaborative filtering that pixel *j* contributes to center pixel *i*. *N*(*i*) is the set of neighbors for pixel *j*

From the perspective of filtering high-contrast details, $w_i(j)$ chooses the nearby, similar and consecutive pixels generated from the Gestalt principles of grouping to filter the center pixel. Similar to the bilateral filter, our method can produce favorable results in filtering noises without blurring edges in the image, by taking the similarity between image pixel into consideration. As indicated in Eq. (14), the range distance, the spatial distance and the minimum bottleneck distance are integrated for assisting the construction of the bilateral-like filter kernel. For example, assuming that pixel i is located on the edge, if the difference between i and its neighboring pixel j is large, the filtering weight that j contributes to i will be small. By this means, the proposed method can smooth noises while preserving edges.

In addition, the proposed filter outperforms the tree filter and UMST filter in removing high-contrast textures. For pixels in a highly textured region, they usually exhibit strong differences between neighbors. As indicated in Eq. (12), the Gestalt-grouping-based minimum bottleneck distance is determined by the largest edge weight along the path. That is, it will ignore other smaller edges' weights on the path when estimating the minimum bottleneck distance for pixels in the highly textured region. As a result, we can achieve consistent distance description between all pixels inside textured regions. In other words, small edge weights are filtering out. As a result, the proposed filter, where minimum bottleneck distances play an important role, is able to smooth pixels in textured regions by following the Gestalt principles of grouping. However, the tree filter (path length) and UMST filter (geodesic distance) will take different pixels inside the same textured region as different, leading to large distances between pixels. As described in Sect. 3.2.2, we note that the tree distance and geodesic distance can yield inconsistent description for similar image elements in a large homogenous region due to the excessively long path on the tree and the accumulation of the transitions in the region.

4.3 Analysis

In fact, the calculation of the filtering results can be rewritten as

$$S_i = \sum_{j\in\Omega} w_i(j)I_j = \sum_{j\in\Omega}\sum_{k\in\Omega} b_j(k)c_k(i)I_j = \sum_{k\in\Omega} c_k(i)\sum_{j\in\Omega} b_j(k)I_j. \tag{15}$$

Set

$$B_k = \sum_{j\in\Omega} b_j(k)I_j \tag{16}$$

Then,

$$S_i = \sum_{k\in\Omega} c_k(i)B_k. \tag{17}$$

Thus, the computation of the filtering outputs can be divided into two phases. First, B_k is the typical bilateral filter technique and has many fast approximation versions. In this paper, we employ one simple and fast implementation proposed by [22], where the down-sampling of the convolution computation is processed in high-dimensional space to reduce the complexity. As for the estimation of the minimum-bottleneck-distance-based filtering process in (16), we follow the work [52] by traverses the whole tree from leaf nodes to root node and then from root to leaves. The computational complexity of this approach is linear to the number of image pixels. Algorithm 2 demonstrates the pseudo-code of generating the filtering output.

Algorithm 2: Pseudo-code for filtering

Input: Original image I

(1) Construct the 4-connected graph G for the image I;

(2) Apply the fast Prim's algorithm to build an MST from G;

(3) Apply (16) to calculate the bilateral filtering output B_k for each node in G;

(4) Assign the value B_k to each node on MST;

(4) Set a root node on MST, and scan the tree from leaf nodes to root node, then root node to leaf nodes to compute the minimum-bottleneck-distance based filtering output by (17);

Output: the filtering output S_i for each node

5. Experimental Results

5.1 ParametersSetting

In this paper, there are three parameters: σ_s, σ_r and σ, required to be predefined, respectively, for calculating the filtering kernel. The σ_s determines the selection of neighboring pixels in the spatial domain, and the σ_r chooses similar pixels in the feature domain. As stated in the bilateral filter [1], σ_s is measured by the pixel number and σ_r is a real number between 0 and 1. The σ controls the choice of pixels that belong to the same cluster in the manifold space. Similar to the Euclidean distance in feature space, the measurement of the minimum bottleneck distance also satisfies the basic principles in the transform domain. For this reason, the σ is also set between 0 and 1.

However, it is still difficult to simultaneously balance all three parameters for achieving satisfactory filtering results. In order to select pixels as similar as possible for collaborative filtering, we fix the range parameter σ_r to a small value since a large σ_r may undesirably assign large weight to some dissimilar pixels. As a result, σ_r is fixed to 0.05 for producing all the experiments results in this paper. Then, the spatial variable σ_s is tuned with the cluster variable σ to achieve different degrees of smoothing.

5.2 Experimental Comparisons

5.2.1 Results Under Different Parameters

Figure 5 shows the filtering results of the lamp image achieved by changing the parameters σ from 0.01 to 0.8 and σ_s from 2 to 8. Given a fixed value to the cluster parameter σ (the same column from top to

down), we note that larger spatial parameter σ_s can produce more smoothing results than smaller ones. The reason is that more faraway pixels are allocated to higher weights for smoothing (similar to Gaussian filter). However, some prominent edges are overly filtered, resulting in blurry edges. On the contrary, if a small σ_s is chosen (less pixels participate in the filtering process), some textured details will be retained. In fact, the trade-off of σ_s controls the choice between detail-filtering and structure-preserving. As described above, higher σ_s means more details filtering, while lower σ_s indicates more edges preserving. In this paper, σ_s is set to 3 for the following comparisons.

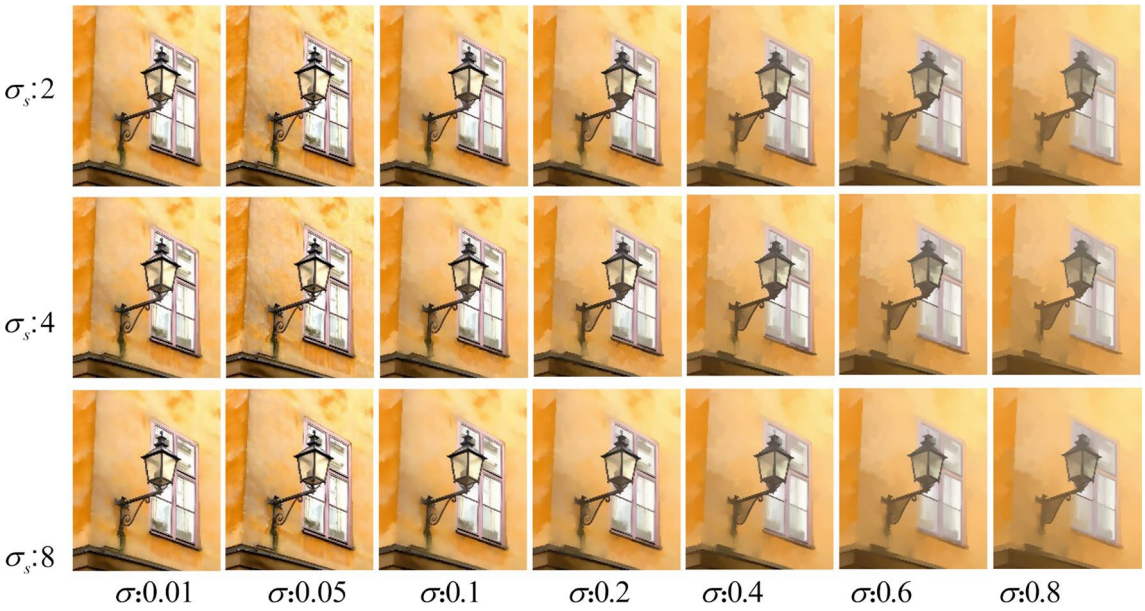


Fig. 5 Filtering results generated from varying parameters σ and σ_s

As indicated in Fig. 5, the larger the cluster parameter σ, the smoother the regions are. This is because a larger σ allows more neighbor pixels inside a region to interact with each other for filtering. Consequently, the homogeneous region may be falsely filtered by involving more dissimilar pixels outside the region. In this case, the cluster information does not work, and largely blurring regions may occur in the whole image. However, if σ is set too small, insufficient pixels in the cluster will be chosen for filtering textured regions. Hence, the selection of σ may control the number of pixels required for smoothing. As demonstrated above, higher σ means large-scale filtering, while lower σ indicates more structures respecting. In this paper, σ is set to 0.1 for all the following experiments.

5.2.2 Qualitative Comparisons of Different Methods

As the evaluation of the structure-preserving filtering is subjective, relevant researches are usually assessed by the visual evaluation. For comparison, we choose bilateral filter, bilateral-filter-like methods (anisotropic diffusion filter [1], L0 smoothing filter [4]), local-histogram-based filter (region covariance filter [15], local-extrema-based method [8]), structure-preserving methods (relative total variance [6]) and non-local-geometry-based methods (non-local means method [26], tree filter [11]) in this paper. Figure 6 visually shows the comparisons of the proposed structure-preserving filter with the above state-of-the-art ones on a baboon image, which is full of high-contrast details that need to be smoothed out. Since the bilateral filter and its analogous ones apply the contrast/gradient to distinguish detail from structures, the highly textured regions are usually regarded as edges and not removed (see Fig. 6b–d). The local-histogram-based filter, which relies on local statistics of an area to separate textures and edges, generates smoothing results in textured regions. However, there exist serious deviations near the sharp edges (the edge along the red nose in Fig. 6e) since the topology information is totally

neglected by this strategy. Instead of using the direct distribution in a local window, the local-extreme-based method exploits both the maximum and minimum values to smooth the details, also leading to distinct deviations across prominent edges (shown in Fig. 6f). In Fig. 6g, we note that highly textured regions have been consistently filtered by the relative total variation measure. Nevertheless, the edges are excessively flattened, especially at the corners. As for the tree filter, leak problems still happen near the sharp edges since the tree distance con- siders nearby dissimilar pixels with hard edges having short tree distance on the tree (demonstrated in Fig. 6h, the small region between nose and beard has been contaminated by the red color). Figure 6i shows the results produced by the non-local means filter, which explores global similar patches for smoothing. Although it presents favorable performance in saving important edges, the textured regions cannot be handled effectively. It is clear that our filter not only removes the details in textured region, but also successfully avoids the problem of leaking distance.

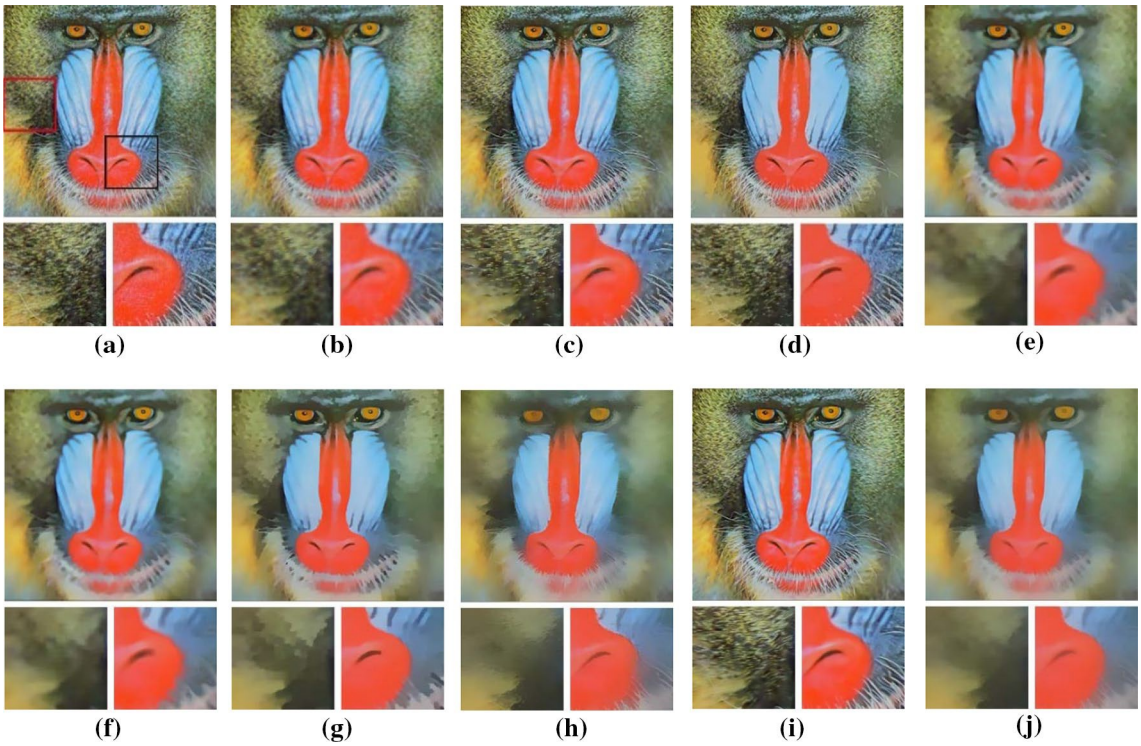


Fig. 6 Filtering results generated by comparing the proposed filter with the state-of-the-art ones under the default parameter settings published by authors. **(a)** Original image. b Bilateral filter. **(c)**, **(d)** Bilateral-filter-like methods, anisotropic diffusion filter and L0 smoothing filter. **(e)**, **(f)** Local-distribution-based methods, region covariance filter and local extrema filter. **(g)** Relative total variance filter. **(h)**, **(i)** Non-local-geometry-based methods, tree filter and non-local means filter. **(j)** Our approach

Recently, the learning-based filter [10] has been employed as a prior to preserve major structures while removing details and achieved high-quality filtering results for textured images. Figure 7 shows more comparative results among the bilateral, relative total variance, tree, learning-based and our filters on other two images. It is obvious that both the relative total variance and the learning-based filters can produce flattened regions in highly textured areas. However, the meaningful edges have been extremely filtered too (see Fig. 7c, e). As demonstrated in Fig. 7d, the tree filter fails to uniformly smooth the hair region because the tree distances are incorrectly accumulated in a large scale of region caused by the false edges problem. We observe that our method not only filters the high-contrast textures but also preserves significant edges in the image.

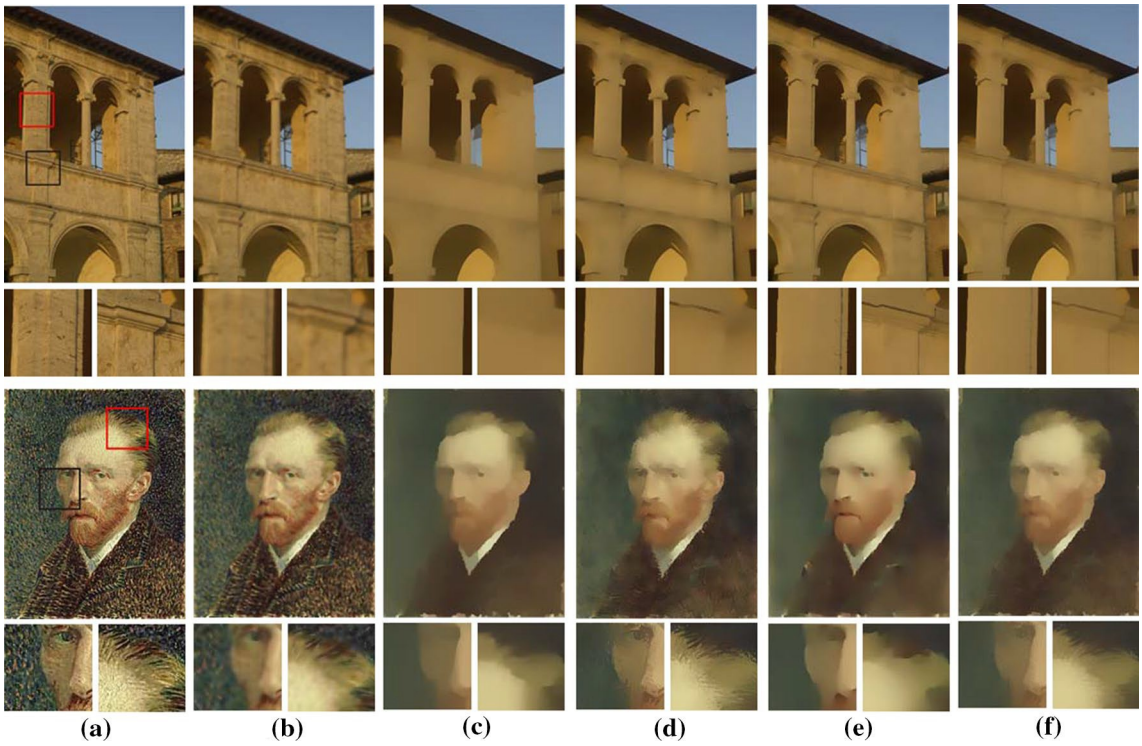
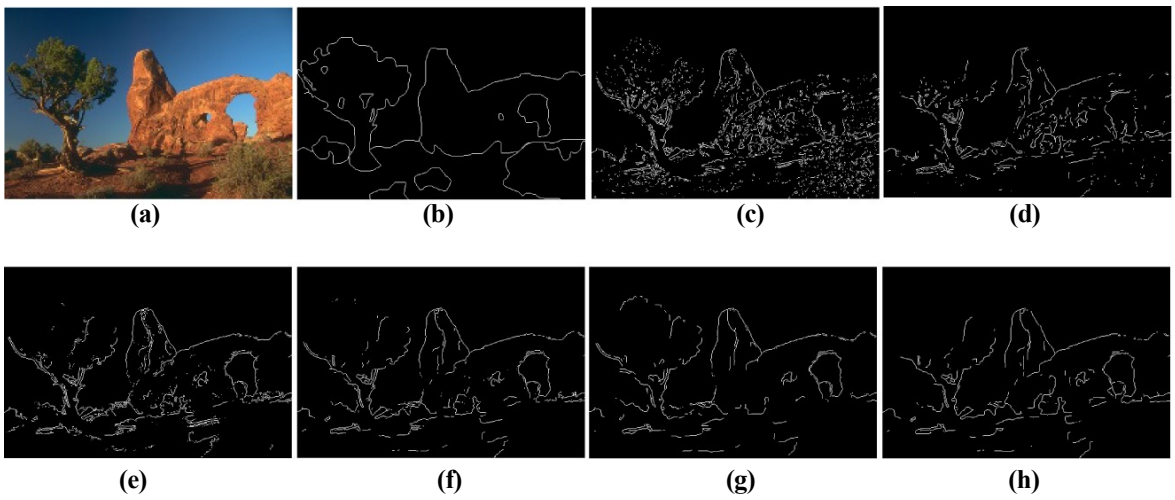
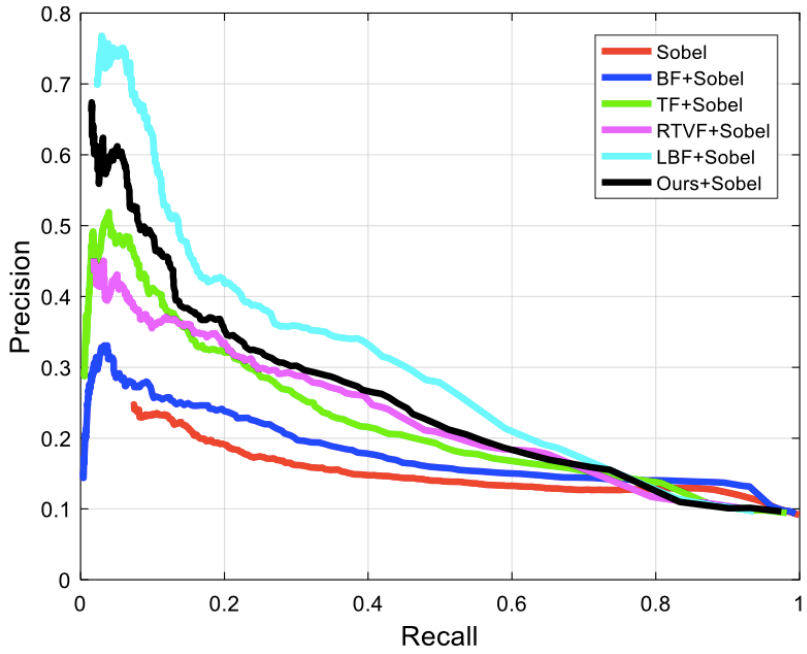


Fig. 7 Filtering results on two other images. **(a)** Original image, **(b)** bilateral filter, **(c)** relative total variance filter, **(d)** tree filter, **(e)** learn-based filter and **(f)** ours

5.2.3 Quantitative Comparisons of Different Methods

To quantitatively validate the effectiveness of the proposed filtering method in removing textures and preserving structures in the image, we apply the Sobel operator to extract the edges from the original and filtered images, respectively, as shown in Fig. 8c–h. The bilateral filter, tree filter, relative total variance filter and learned- based filter are chosen for filtering the original image. It is clear that the textures have been filtered in different degrees. Then, we compare the detected edges with the labeled edges from human beings to generate the precision–recall curve, for further evaluating the performance of different filters in assisting the boundary detection. As shown in Fig. 8k, we note that the learning-based filter has achieved the best performance for boundary detection, and the proposed method gets the second-best one. One possible reason is that the learning-based method is formulated by applying the pre-detected edges.





(i)

Fig. 8 Boundary detection. (a) Original image, (b) ground truth, (c) Sobel, (d) bilateral filter + Sobel, (e) tree filter + Sobel, (f) relative total variance filter + Sobel, (g) learn-based filter + Sobel, (h) ours + Sobel and (i) shows the PR curves for different filtered based boundary detection methods.

Moreover, we add synthetic structured noise and random noise including Gaussian, salt and pepper, Poisson and speckle noises of various scales into the Lenna image (see Fig. 9), and then calculate the SSIM values to evaluate the performance of denoising and preserving structures. The experimental results are shown in Table 1. For the bilateral filter, it can attain good SSIM values for Gaussian, Poisson and speckle noise, but its performance drops sharply as the variance σ of the salt and pepper noise and the amplitude A and angular frequency ω of periodic noise increase.

Table 1 SSIM values for different methods under different noise of various scales

	BF	TF	RTVF	LBF	Ours
Random noise					
Gaussian (σ)					
.02	.9658	.9342	.9621	.9396	.9638
.05	.9607	.9284	.9573	.9341	.9582
.1	.9440	.9107	.9415	.9167	.9431
Salt and pepper (σ)					
.02	.8832	.9590	.9610	.9483	.9594
.05	.7732	.9519	.9575	.9386	.9530
.1	.6347	.9371	.9486	.9249	.9413
Speckle (σ)					
.02	.9698	.9333	.9624	.9461	.9608
.05	.9480	.9200	.9588	.9382	.9536
.1	.8758	.9060	.9488	.9274	.9402
Poisson					
	.9742	.9514	.9638	.9500	.9677
Structured noise					
Periodic ($y = Asin(\omega x)$, A, ω)					
20, 30	.9762	.9500	.9479	.9622	.9685
60, 30	.7698	.9046	.9126	.9282	.9233
60, 60	.7398	.8971	.9228	.9412	.9404

Compared with the tree filter, the proposed method is capable of improving the performance of denoising. In addition, it can achieve favorable results on the random and structured noises at various scales. For the structured noise, we note that the learning-based filter can achieve promising results. The reason is that the learning-based method applies the pre-detected edges to accomplish the structure preserving and noise filtering (Table 1).

Fig. 9 Different noises added into the Lenna image. (a) Lenna image, (b) Gaussian noise with σ = 0.05, (c) salt and pepper noise with σ = 0.05, (d) speckle noise with σ = 0.05, (e) Poisson noise and d periodic noise with amplitude A = 60, angular frequency ω = 30

5.3 Applications

1. Scene simplification

According to [12], human beings primarily tend to perceive the overall structural information, not the individual details in natural scenes. Stimulated by this, the authors [53] have obtained enhanced results of detecting salient regions in images by applying the relative total variance filter to work as pre-processing step in saliency detection. In fact, the scene simplification can be used as an effective tool for many applications in the field of computer vision. For simplicity, we compare the detected saliency maps (the saliency detection approach based on discriminative regional feature integration [54] is introduced here) generated by utilizing the original image and its filtered version, respectively. Figure 10 presents the results applying the filter or not. It is clear that the saliency map generated by the filtering operation is much more uniform and consistent. The reason is that the main idea of estimating the saliency is based on the feature contrast and the filtering assists in removing most of the trivial details which will decrease the effectiveness of calculating the contrast.

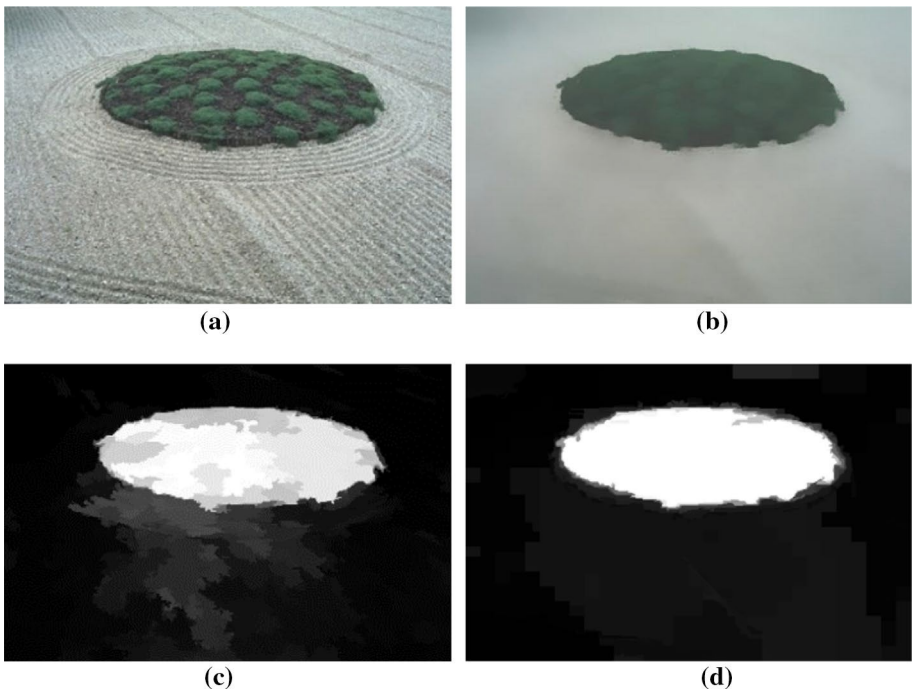


Fig. 10 Saliency maps produced by using the filtering process or not. **(a)** Original image, **(b)** filtered image, **(c)** and **(d)**, respectively, present the saliency maps generated from **(a)** and **(b)**.

2. Texture editing and image composition

Apart from simplifying the natural images for further process, we can also modify the texture layers to create different visual impressions. Two examples of texture editing are included in Fig. 11a. The top image shows the result of adding the structure layer and contrast-enhanced texture layer together. The bottom one illustrates the result of replacing the original grid texture layer by integrating the new rain drops texture into the structure layer. We note that the texture pat- terns are nicely separated from the images by our method, and the newly created images can yield favorable visual effects.

Due to the incompatibility between the source and target textures, paintings, drawings and graffiti images sometimes cannot be seamlessly recomposed by directly cloning these source images into the target ones in image composition. In this paper, we apply the mixing-gradient Poisson cloning method, which locally selects the maximum gradient from the source and target image for fusion [55]. As shown in the top row of Fig. 11b, just compositing with the source image does not produce visually reasonable effects. The reason is that the two textures are incompatible with different texture patterns. It is evident that the results are more natural and favorable if we only merge the structure layer of the source image into the target one (see the bottom row in Fig. 11b).

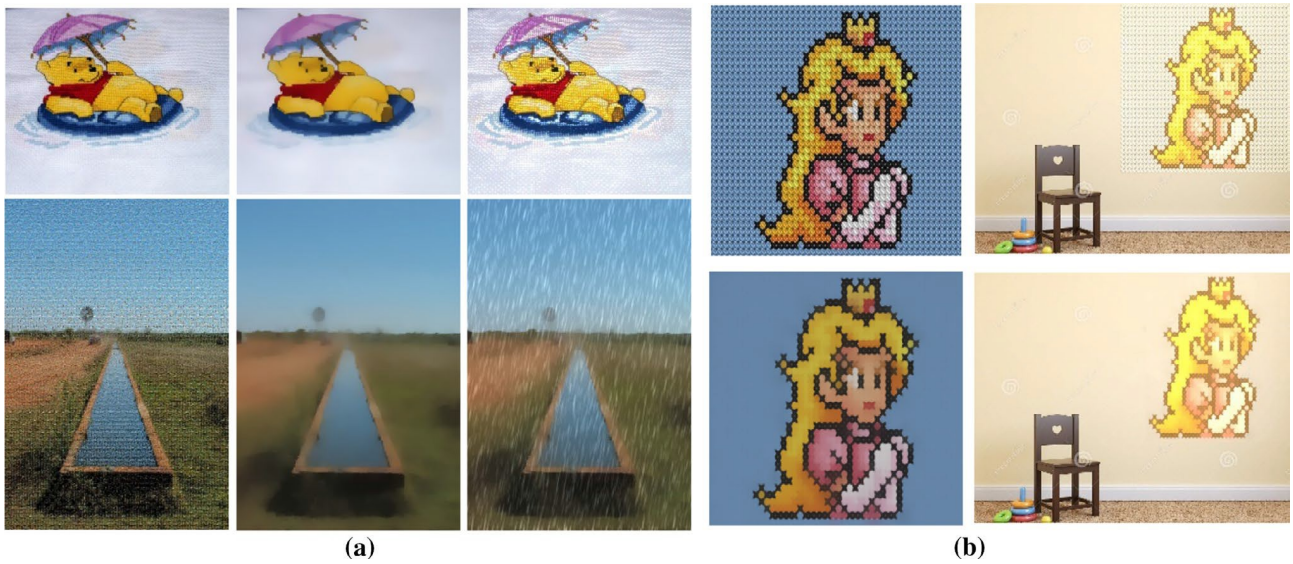


Fig. 11 The illustrations of the texture editing and image composition results. The left two images in **(a)** demonstrate the texture enhancement and texture replacement, respectively (original image, filtered image and created image). The images in **(b)** show the image composition results. The top two are original image and its composition one, and the bottom two correspond to the filtered image and its composition.

6 Conclusion

Observing the phenomenon that dissimilar regions may be connected with low-contrast edges and nearby pixels with similar pattern show large disparities in highly textured regions, it is particularly difficult to devise a favorable structure-aware image filter since there is not a perfect solution for obtaining all ideal edges from natural images of high complexity. For this reason, we propose a novel cluster kernel which takes the similarity, proximity and continuation principles of human perception into consideration for describing the structural relations between pixels. Our filter kernel shows strong strength in allowing those pixels with high consistency to be grouped together for averaging without simply relying on the feature contrast or gradients. In order to better integrate this cluster prior into the traditional bilateral-filter-like framework, we follow the concept of the trust-based mechanism for wireless sensor network in this paper. It is a favorable combination method since more decisions from similar neighbors are collected and used for filtering. The experimental results demonstrate the advantage of applying the collaborative model in preserving the meaningful structures and filtering the trivial details.

Furthermore, we propose the minimum bottleneck distance to quantitatively define the structural relations which includes similarity, proximity and continuation principles. It outperforms some existing distance transforms in minimizing intra-cluster dissimilarities regardless of noises and textures. Based on this, it can be applied in many applications involving affinity definition, such as clustering, image segmentation and image representation and so on.

Acknowledgements

This research was supported by the National Key Research and Development Program of China (No.2018YFA0704605), the National Key Project of Science and Technology of China (No.2017ZX05064), National Natural Science Foundation of China (No. 61272523) and China Scholarship Council (CSC).

References

1. Tomasi, C., Manduchi, R.: Bilateral filtering for gray and color images. In: IEEE Computer Vision, pp. 839–846 (1998)

2. Perona, P., Malik, J.: Scale-space and edge detection using anisotropic diffusion. IEEE Trans. Pattern Anal. Mach. Intell. 12(7), 629–639 (1990)

3. Farbman, Z., Fattal, R., Lischinski, D., Szeliski, R.: Edge-preserving decompositions for multi-scale tone and detail manipulation. ACM Trans. Graph. (TOG) 27(3), 67 (2008)

4. Xu, L., Lu, C., Xu, Y., Jia, J.: Image smoothing via L0 gradient minimization. ACM Trans. Graph. 30(6), 174:1–174:12 (2011)

5. Rudin, L.I., Osher, S., Fatemi, E.: Nonlinear total variation-based noise removal algorithms. Phys. D 60(1–4), 259–268 (1992)

6. Xu, L., Yan, Q., Xia, Y., Jia, J.: Structure extraction from texture via natural variation measure. ACM Trans. Graph. (SIGGRAPH Asia) 31(6), 1–10 (2012)

7. He, K., Sun, J., Tang, X.: Guided image filtering. In: ECCV, pp. 1–14 (2010)

8. Subr, K., Soler, C., Durand, F.: Edge-preserving multiscale image decomposition based on local extrema. ACM Trans. Graph. 28(5), 147:1–147:9 (2009)

9. Zhang, Q., Shen, X., Xu, L., Jia, J.: Rolling guidance filter. In: ECCV (2014)

10. Yang, Q.: Semantic filtering. In: CVPR, pp. 4517–4526 (2016) 11. Bao, L., Song, Y., Yang, Q., Yuan, H., Wang, G.: Tree filtering: efficient structure-preserving smoothing with a minimum spanning tree. IEEE Trans. Image Process. 23(2), 555–569 (2014)

12. Arnheim, R.: Art and Visual Perception: A Psychology of the Creative Eye. University of California Press, Berkeley (1956)

13. Felsberg, M., Forssén, P., Scharr, H.: Channel smoothing: efficient robust smoothing of low-level signal features. IEEE Trans. Pattern Anal. Mach. Intell. 28(2), 209–222 (2006)

14. Kass, M., Solomon, J.: Smoothed local histogram filters. ACM Trans. Graph. 29(4), 100:1–100:10 (2010)

15. Karacan, L., Erdem, E., Erdem, A.: Structure-preserving image smoothing via region covariances. ACM Trans. Graph. (TOG) 32(6), 176:1–176:11 (2013)

16. Criminisi, A., Sharp, T., Rother, C., Perez, P.: Geodesic image and video editing. ACM Trans. Graph. (TOG) 29(5), 134 (2010)

17. Farbman, Z., Fattal, R., Lischinski, D.: Diffusion maps for edge- aware image editing. ACM Trans. Graph. (TOG) 29(6), 145:1– 145:10 (2010)

18. Wolters, A., Koffka, K.: Principles of Gestalt Psychology, pp. 502–504. Routledge, London (1936)

19. Desolneux, A., Moisan, L., Morel, J.-M.: From Gestalt Theory to Image Analysis: A Probabilistic Approach. Springer, Berlin (2008)

20. Chen, J., Paris, S., Durand, F.: Real-time edge-aware image pro- cessing with the bilateral grid. ACM Trans. Graph. (TOG) 26(3), 103:1–103:10 (2007)

21. Porikli, F.: Constant time O(1) bilateral filtering. In: IEEE CVPR, pp. 1–8 (2008)

22. Paris, S., Durand, F.: A fast approximation of the bilateral filter using a signal processing approach. Int. J. Comput. Vision 81(1), 24–52 (2009)

23. Yang, Q., Tan, K., Ahuja, N.: Real-time O(1) bilateral filtering. In: IEEE CVPR, pp. 557–564 (2009)

24. Yang, Q.: Recursive bilateral filtering. In: ECCV, pp. 399–413 (2012)

25. Fattal, R.: Edge-avoiding wavelets and their applications. ACM Trans. Graph. 28(3), 22:1–22:10 (2009)

26. Buades, A., Coll, B., Morel, J.M.: A non-local algorithm for image denoising. Comput. Vis. Pattern Recognit. 2, 60–65 (2005)

27. Paris, S., Hasinoff, S., Kautz, J.: Local Laplacian filters: edge-aware image processing with a Laplacian pyramid. ACM Trans. Graph. 30(4), 68:1–68:12 (2011)

28. Gastal, E., Oliveira, M.: Domain transform for edge-aware image and video processing. ACM Trans. Graph. 30(4), 69:1–69:12 (2011)

29. Strand, R., Ciesielski, K., Malmberg, F., Saha, P.: The minimum barrier distance. Comput. Vis. Image Underst. 117(4), 429–437 (2013)

30. Ciesielski, K., Strand, R., Malmberg, F., Saha, P.: Efficient algorithm for finding the exact minimum barrier distance. Comput. Vis. Image Underst. 123, 53–64 (2014)

31. Lezoray, O., Grady, L. (eds.): Image Processing and Analysis with Graphs. CRC Press, Boca Raton (2012)

32. Sanfeliu, A., Alquézar, R., Andrade, J., et al.: Graph-based representations and techniques for image processing and image analysis. Pattern Recognit. 35(3), 639–650 (2002)

33. Saha, P., Wehrli, F., Gomberg, B.: Fuzzy distance transform: theory, algorithms, and applications. Comput. Vis. Image Underst. 86(3), 171–190 (2002)

34. Udupa, J., Samarasekera, S.: Fuzzy connectedness and object definition: theory, algorithms, and applications in image segmentation. Graph. Models Image Process. 58(3), 246–261 (1996)

35. Couprie, M., Najman, L., Bertrand, G.: Quasi-linear algorithms for the topological watershed. J. Math. Imag. Vis. 22(2–3), 231–249 (2005)

36. Falcao, A., Stol, J., Alencar Lotufo, R.: The image foresting transform: theory, algorithms, and applications. IEEE PAMI 26(1), 19 (2004)

37. Fischer, B., Zöller, T., Buhmann, J. M.: Path based pairwise data clustering with application to texture segmentation. In: Computer Vision and Pattern Recognition, pp. 235–250 (2001)

38. Nagao, M., Matsuyama, T., Ikeda, Y.: Region extraction and shape analysis in aerial photographs. Comput. Graph. Image Process. 10(3), 195–223 (1979)

39. Gower, J., Ross, G.: Minimum spanning trees and single linkage cluster analysis. Appl. Stat. 18(1), 54–64 (1969)

40. Soille, Pierre: Constrained connectivity for hierarchical image partitioning and simplification. IEEE Trans. Pattern Anal. Mach. Intell. 30(7), 1132–1145 (2008)

41. Meyer, F., Maragos, P.: Nonlinear scale-space representation with morphological levelings. J. Vis. Commun. Image Represent. 11(3), 245–265 (2000)

42. Hambrusch, S., He, X., Miller, R.: Parallel algorithms for gray- scale image component labeling on a mesh-connected computer. In: Proceedings of the Fourth Annual ACM Symposium on Paral- lel Algorithms and Architectures, pp. 100–108 (1992)

43. Braga-Neto, U., Goutsias, J.: Grayscale level connectivity: theory and applications. IEEE Trans. Image Process. 13(12), 1567–1580 (2004)

44. Liebig, J.: 1840, Salisbury, Plant Physiology, 4th edn. Wads- worth, Belmont (1992)

45. Najman, L., Cousty, J., Perret, B.: Playing with kruskal: algo- rithms for morphological trees in edge-weighted graphs. In: International Symposium on Mathematical Morphology and Its Applications to Signal and Image Processing. Springer, pp 135–146 (2013)

46. Prim, R.C.: Shortest connection networks and some generaliza- tions. Bell Syst. Tech. J. 36(6), 1389–1401 (1957)

47. Achanta, R., Shaji, A., Smith, K., Lucchi, A., Fua, P., Susstrunk, S.: SLIC superpixels compared to state-of-the-art superpixel methods. IEEE Trans. Pattern Anal. Mach. Intell. 34, 2274– 2282 (2012)

48. Danda, S., Challa, A., Daya Sagar, B., Najman, L.: Some theo- retical links between shortest path filters and minimum spanning tree filters. J. Math. Imag. Vis. 61(6), 745–762 (2018)

49. Miranda, P.A.V., Falcao, A.X.: Image Segmentation by the Image Foresting Transform. XX Concurso de Teses e Disser- taçoes, pp. 2043–2047 (2007)

50. Lerallut, R., Decenciere, E., Meyer, F.: Image filtering using morphological amoebas. Image Vis. Comput. 25(4), 395–404 (2007)

51. Huang, L., Li, L., Tan, Q.: Behavior-based trust in wireless sen- sor network. In: Asia-Pacific Web Conference. Springer, Berlin, pp. 214–223 (2006)

52. Yang, Q.: A non-local cost aggregation method for stereo matching. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1402–1409 (2012)

53. Xu, L., Yan, Q., Xia, Y., Jia, J.: Structure extraction from texture via relative total variation. ACM Trans. Graph. (TOG) 31(6), 139 (2012)

54. Jiang, H., Wang, J., Yuan, Z., et al.: Salient object detection: a discriminative regional feature integration approach. In: Pro- ceedings of the IEEE Conference on Computer Vision and Pat- tern Recognition, pp. 2083–2090 (2013)

55. Pérez, P., Gangnet, M., Blake, A.: Poisson image editing. ACM Trans. Graph. (TOG) 22(3), 313–318 (2003)