IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received August 1, 2017, accepted August 21, 2017, date of publication August 29, 2017, date of current version February 28, 2018.

Digital Object Identifier 10.1109/ACCESS.2017.2745903

Edge-Aware Spatial Denoising Filtering Based on a Psychological Model of Stimulus Similarity

JOSHIN MATHEW[1], AMIN ZOLLANVARI[2], (Member, IEEE), AND ALEX PAPPACHEN JAMES[2], (Senior Member, IEEE)

[1]Computer Vision Team, ARS Traffic & Transport Technology, Trivandrum 695 581, India
[2]Department of Electrical and Computer Engineering, Nazarbayev University, Astana 010000, Kazakhstan

Corresponding author: Alex James (apj@ieee.org)

ABSTRACT Noise reduction is a fundamental operation in image quality enhancement. In recent years, a large body of techniques at the crossroads of statistics and functional analysis have been developed to minimize the blurring artifact introduced in the denoising process. Recent studies focus on edge-aware filters due to their tendency to preserve image structures. In this paper, we adopt a psychological model of similarity based on Shepard's generalization law and introduce a new signal-dependent window selection technique. Such a focus is warranted, because blurring is essentially a cognitive act related to the human perception of physical stimuli (pixels). The proposed windowing technique can be used to implement a wide range of edge-aware spatial denoising filters, thereby transforming them into nonlocal filters. We employ simulations using both synthetic and real image samples to evaluate the performance of the proposed method by quantifying the enhancement in the signal strength, noise suppression, and structural preservation measured in terms of the peak signal-to-noise ratio (PSNR), mean square error (MSE), and structural similarity (SSIM) index, respectively. In our experiments, we observe that incorporating the proposed windowing technique in the design of mean, median, and nonlocal means filters substantially reduces the MSE while simultaneously increasing the PSNR and the SSIM.

INDEX TERMS Noise reduction, blurring artifact, non-local denoising, psychological model, generalization law.

I. INTRODUCTION

Image denoising is the process of noise reduction or removal in a corrupted image, and as such, it is an inevitable step in image quality enhancement. Recent advances in image processing have brought about a host of proposed denoising techniques based on various image statistics [1]–[4], different variations of partial differential equations [5]–[8], and structural correlation or similarity measures [9]–[12]. From a broader point of view, these techniques can be mainly categorized as operations in the spatial domain or in the transform domain [13], [14]. Spatial domain operations are further classified as local or nonlocal methods.

A method is called local if the intensity value of the pixel under consideration is determined by a filter with a support and weights that are determined by the spatial distance from the candidate pixel. On the other hand, a method is called nonlocal if the filter support and weights depend on the difference in the intensity value of the candidate pixel from other pixels within the support. In general, nonlocal methods achieve a better level of noise reduction at the cost of having a higher computational complexity [14]. Nevertheless, nonlocal methods that restrict the weight computation to a sub-image surrounding the candidate pixel show an improved computational efficiency [9].

Regardless of whether one uses a local or nonlocal method, a major drawback is the introduction of edge-blurring artifacts. Such artifacts lead to poor detection and localization as well as uncertain distinction between various regions [15]. The main cause of blurring artifacts is the presence of heterogeneous regions within the filter support. Adaptive window selection is a method to confine the estimation support to more homogeneous regions. This process generally restricts image blurring over boundaries and leads to an improved edge preservation [16]–[19]. At the same time, various other operations have been proposed that internally exhibit an edge-aware preservation property, for example, anisotropic diffusion, domain transform, and L_0 gradient minimization [15], [20], [21].

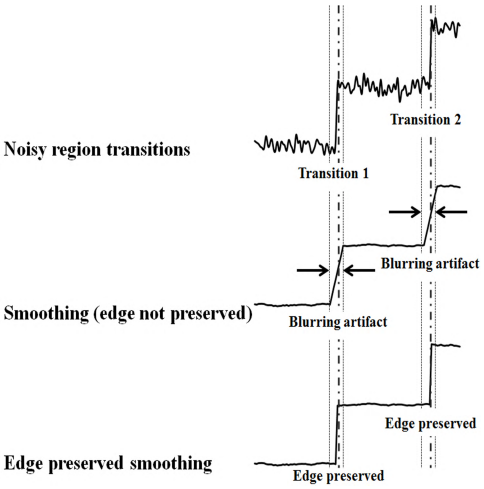

FIGURE 1. Demonstration of blurring artifact and edge-preserved denoising.

Figure 1 illustrates the basic concept of edge-preserving image smoothing or denoising. Whereas in conventional denoising the sharp transition (i.e., the edge) is over-smoothed, the goal of adaptive windowing is to determine the window shape and size to preserve the edge. The focus of adaptive methods is to determine the window shape, size, and orientations needed to deliver maximum denoising with minimal blurring artifacts. In this regard, image gradients, variance, and mean intensity level have been the key metrics employed in previous studies to adaptively select neighborhoods [16]–[19], [22], [23]. These metrics (described in greater detail in Section II) essentially serve as a measure of similarity to determine "neighboring" pixels.

The underlying hypothesis of the present study is that using a metric of pixel similarity analogous to the metric that the human cognitive system uses improves the quality of adaptive windowing in denoising applications. In other words, matching the algorithmic metric used for adaptive window selection with that of the human brain results in improved denoising performance with respect to typical similarity metrics such as gradient and image statistics. Thus, we adopt a psychological model of similarity based on Shepard's generalization law and introduce a new adaptive windowing technique. The proposed signal-dependent technique can then be used to implement a wide range of spatial denoising filters. In this regard, we demonstrate the application of the proposed technique in implementing a conventional mean filter [24], a median filter [24], and nonlocal means filters [11], [12]. The efficacy of the proposed technique is discussed using various metrics: signal strength (PSNR), noise suppression (MSE), and structural similarity preservation (SSIM).

Throughout this article, we denote an image by $\mathcal{I} \triangleq \{(x, y, I_{x,y})\}$, where $\{(x, y)\}$ is the set of pixel coordinates, $x = 1, .., M$, $y = 1, .., N$, and $I_{x,y}$ is the gray-level intensity value of the pixel at location (x, y).

II. BACKGROUND: ADAPTIVE WINDOWING TECHNIQUES

Typical adaptive window selection methods are either based on gradient vector or image statistics. This section provides a brief overview of these methods.

A. GRADIENT-BASED METHOD

Let g_x and g_y denote the gradient along the x and y directions, respectively. In this approach, for each pixel, g_x and g_y are used to determine the possible size and the orientation of a window that points to neighboring homogeneous regions [16]. In particular, let

$$W_x = \frac{\gamma}{(g_x + 1)}, \tag{1}$$

$$W_y = \frac{\gamma}{(g_y + 1)}, \tag{2}$$

where W_x and W_y denote the length of the window sides in the x and y directions, respectively, and γ is a proportionality constant determined either by the user or an external process to determine the amount of smoothing. The window is oriented such that its long (short) side is aligned in the direction of the minimum (maximum) gradient when the reference pixel is at an edge location. To mathematically formalize this process, let θ_0 denote the direction of the maximum gradient at pixel (x_0, y_0), and set $\alpha_0 \triangleq \theta_0 + \frac{\pi}{2}$. Consider a window with sides determined by (1) and (2), centered at the origin of an auxiliary coordinate system with coordinates denoted by (x_{aux}, y_{aux}). Then, the set of pixels in the original image coordinate system that are part of the window (denoted by (x, y)) can be found by solving the following system of equations:

$$x = x_{aux}\cos(\alpha_0) - y_{aux}\sin(\alpha_0) + x_0, \tag{3}$$

$$y = x_{aux}\sin(\alpha_0) + y_{aux}\cos(\alpha_0) + y_0. \tag{4}$$

In other words, for each entry (x_{aux}, y_{aux}) in the window, we may use this system of equations to find the corresponding pixel in the image coordinates.

B. IMAGE STATISTICS-BASED ADAPTIVE WINDOWING

In this approach, a local image variance is considered as a measure to determine the window shape and size [17]–[19]. The neighborhood selection criterion is defined to determine a region with the "maximum homogeneity". In this regard, the window $\mathcal{W}$ (representing the homogeneous region) centered at pixel (x_0, y_0, I_0) is determined by $\mathcal{W} = \{(x, y)|w_{(x,y)} = 1\}$, where

$$w_{(x,y)} = \begin{cases} 1, & \text{if } (1 - 2\sigma_0)\mu_0 \le I_0 \le (1 + 2\sigma_0)\mu_0 \\ 0, & \text{otherwise,} \end{cases} \tag{5}$$

with μ_0 and σ_0 being the mean and standard deviation of the gray-level intensity values within a fixed neighborhood of (x_0, y_0), respectively.

III. Shepard's GENERALIZATION LAW AND THE PROPOSED TECHNIQUE

In a seminal work, Shepard laid down the groundwork for mathematically characterizing the human perception of stimulus similarity [25]. Suppose that S_A and S_B are representations of stimuli A and B, respectively. Let $d(S_A, S_B)$ and $\delta(S_A, S_B)$ denote the *perceived* dissimilarity in a *psychological space* and the judged dissimilarity in the experimental space [26], respectively. Note that there is a distinction between perceived and judged dissimilarity. In experiments, when asking subjects to judge the similarity of stimuli, the measurable quantity is $\delta(S_A, S_B)$, not $d(S_A, S_B)$. In other words, the psychological space is inaccessible, and its properties are only revealed through $\delta(S_A, S_B)$. Thus, a general model that relates $d(S_A, S_B)$ and $\delta(S_A, S_B)$ is

$$\delta(S_A, S_B) = f[d(S_A, S_B)], \tag{6}$$

where f is a suitable monotonically nondecreasing function. Here, f is a measure of *dissimilarity* because it increases as the distance between S_A and S_B (i.e., $d(S_A, S_B)$) increases. In [25], Shepard attempts to derive a monotonic function $g[d(S_A, S_B)]$ that measures the *similarity* of stimuli using so-called generalization data. The term generalization here refers to the probability that a response learned to stimulus S_A results in the response (i.e., generalizes) to stimulus S_B and vice versa. He assumes the following: 1) the similarity function $g[.]$ has the same form for all types of stimuli (universality), and (2) the perceptual function $d(S_A, S_B)$ is a metric. Under these two assumptions, he proves that the similarity function (also known as the generalization function) takes on an exponential form [25]. In particular, if we assume a one-dimensional psychological space, then

$$g[d(S_A, S_B)] = e^{[-d(S_A, S_B)]} \tag{7}$$

such that

$$d(S_A, S_B) = |x_A - x_B|, \tag{8}$$

where x_A and x_B are the coordinates of stimuli S_A and S_B in the psychological space.

Here, we take image pixels as the set of stimuli and consider their intensity values as coordinates of stimuli in the psychological space. Therefore, (7) determines the judged similarity between intensities I_A and I_B. For simplicity, we define the following new metric, referred to as Shepard's similarity function (SSF), to measure the judged similarity of intensities I_A and I_B:

$$\mathrm{SSF}(I_A, I_B) \triangleq g[d(I_A, I_B)] = e^{-|I_A - I_B|}. \tag{9}$$

SSF is a monotonic increasing function of $d(I_A, I_B)$ defined in (8) with a value of 1 for a perfectly judged similarity (i.e., at a perceived distance of 0) and approaches 0 as the perceived distance increases. Fig. 2 displays SSF versus intensity difference between -1 and 1. Note that the metric nature of $d(S_A, S_B)$ and the monotonic property of $g[d(S_A, S_B)]$ imply that SSF satisfies the following properties necessary for a similarity measure:

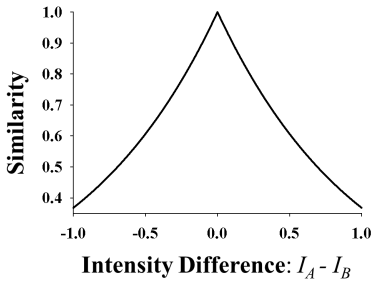

FIGURE 2. Judged similarity as a function of stimuli difference varying from −1.0 to 1.0.

- **Symmetry:** $d(I_A, I_B) = d(I_B, I_A) \Leftrightarrow \mathrm{SSF}(I_A, I_B) = \mathrm{SSF}(I_B, I_A)$.
- **Minimality:** if $d(I_A, I_B) > d(I_A, I_C) \Rightarrow \mathrm{SSF}(I_A, I_B) < \mathrm{SSF}(I_A, I_C)$.
- **Boundedness:** $\mathrm{SSF}(I_A, I_B) \leq 1$.
- **Unique maximum:** $\mathrm{SSF}(I_A, I_B) = 1 \Leftrightarrow I_A = I_B$.

In our approach, SSF is then employed as the measure to evaluate the similarity of neighboring pixels with respect to a candidate pixel. To determine the shape of the neighboring pixels (selected window), the image intensities are normalized to be between 0 and 1, and thus, the range of difference values will be between -1 and 1. This normalization is performed to make the window derivation feasible regardless of the input gray-level range. Nevertheless, the denoising procedure per se can be performed on the actual gray-level intensity values.

In the proposed framework, the adaptive window $\mathcal{W}$ centered at pixel (x_0, y_0, I_{x_0,y_0}) is formally determined as

$$\mathcal{W} = \{(x, y) | w_{(x,y)} = 1 \text{ and } (x, y) \in \mathcal{N}_{x_0,y_0}\}, \tag{10}$$

where $\forall x, y$,

$$w_{(x,y)} = \begin{cases} 1, & \text{if } \mathrm{SSF}(I_{x,y}, I_{x_0,y_0}) > \eta \\ 0, & \text{otherwise} \end{cases}, \tag{11}$$

with η denoting a similarity threshold and $\mathcal{N}_{x_0,y_0}$ being a given neighborhood of the candidate pixel at (x_0, y_0). From Fig. 2, the maximum similarity is 1 for a zero distance, and the minimum similarity value is 0.367. Therefore, η is a number in this range and provides a degree of freedom to control the trade-off between noise reduction and structure preservation. Unless otherwise stated, in all experiments hereafter, we set $\eta = 0.85$ (see Section VI-*B* for more details). For the neighborhood $\mathcal{N}_{x_0,y_0}$, we simply assume a square-shaped mask of size $m \times m$, in which m is provided by the user. In this case, the computational complexity of determining $\mathcal{W}$ is $\mathcal{O}(m^2)$.

The derived similarity window thus becomes a binary mask where neighboring locations with a value of 1 adapt to the shape and the size of the homogeneous region within the given neighborhood. This mask can then be used in the denoising process, where non-similar locations are ignored. In other words, one may easily employ the proposed adaptive windowing technique described here in conjunction with an

arbitrary filter by first determining the local neighborhood and then apply the filter.

Figure 3 demonstrates a comparison of window selection approaches applicable to non-adaptive (or conventional) and the proposed methods. In the conventional approach, all the pixels within a regular rectangular (square) region or neighborhood are used in noise reduction methods, whereas in the proposed technique based on the (judged) pixel-to-pixel similarity, the neighborhood of a pixel is selected.

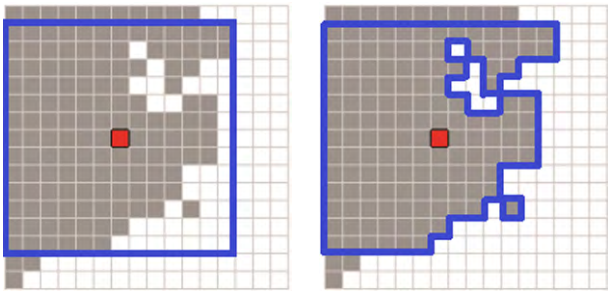

FIGURE 3. Comparison of window selection approaches. (a) non-adaptive window: a rectangular region is selected, (b) proposed window: the neighborhood shape is adapted to the homogeneous region based on Shepard's similarity function.

IV. ASSESSING THE PERFORMANCE OF DENOISING METHODS

In this section, we describe different quality measures to assess the performance of denoising techniques from various points of view.

A. METRICS RELATED TO ERROR POWER

The MSE is a quadratic form designed to predict the global energy of the discrepancy (error) between the original image and the processed image. Let $x = 1, .., M$, $y = 1, .., N$ and denote the original image (gold standard) and the denoised image by $\mathcal{I} = \{(x, y, I_{x,y})\}$ and $\mathcal{I}^D = \{(x, y, I^D_{x,y})\}$, respectively. Then,

$$\text{MSE} = \frac{1}{MN}\sum_{x=1}^{M}\sum_{y=1}^{N}(I^D_{x,y} - I_{x,y})^2. \tag{12}$$

Another measure related to the error power is the PSNR, which quantifies the maximum signal strength against the retained image noise [27]–[30]. The PSNR is defined as

$$\text{PSNR} = 10\log\left([\max_{(x,y)} I_{x,y}]^2 \times \frac{1}{\text{MSE}}\right). \tag{13}$$

Note that although a smaller MSE indicates a better denoising effect, in the case of the PSNR, the larger the metric, the better the denoising effect.

B. STRUCTURAL SIMILARITY

Blurring artifacts are common structural distortions. An ideal denoising technique should be capable of achieving signal enhancement and structural preservation simultaneously. Thus, when assessing the quality of enhancement processes, the structural similarity between the gold standard image and the processed image must be quantified. The Structural Similarity (SSIM) index is a widely accepted metric to predict the structural preservation in terms of comparing the local patterns of pixel intensities normalized for luminance and contrast. For an image $\mathcal{P}$ and a denoised image $\mathcal{P}^D$, the SSIM is defined as [31]–[33]

$$\text{SSIM}(\mathcal{I}, \mathcal{I}^D) = \frac{(2\mu_{\mathcal{I}}\mu_{\mathcal{I}^D} + c_1)(2\sigma_{\mathcal{I},\mathcal{I}^D} + c_2)}{(\mu^2_{\mathcal{I}} + \mu^2_{\mathcal{I}^D} + c_1)(\sigma^2_{\mathcal{I}} + \sigma^2_{\mathcal{I}^D} + c_2)}, \tag{14}$$

where $\mu_{\mathcal{I}}$ and $\sigma_{\mathcal{I}}$ are the mean and standard deviation of image $\mathcal{I}$, respectively. The parameter $\sigma_{\mathcal{I},\mathcal{I}^D}$ denotes the covariance of $\mathcal{I}$ and $\mathcal{I}^D$, and c_1 and c_2 are stabilizing constants.

C. ATTENTION-BASED QUALITY MEASURES

The preservation of distinct regions is essential in representing the content of an image. Simultaneously, the quality of the visual enhancement is related to the context of visual attention. The approach is to then compute signal quality benchmarks from *interest points* of visual attention. These are locations whereby a viewer immediately notices distinctive features. This approach helps creating a better inference on the visual quality enhancement because we examine content preservation rather than that of global statistics. The computer vision community refers to such locations as key feature points when calculating state-of-the-art image content descriptors [34]–[41].

To characterize the attention-based quality measures, we can take the signal strength and structural preservation metrics defined earlier (MSE, PNSR, and SSIM) and compute them over the interest points. We determine the interest points using the Harris corner detection technique [42]. Let $\mathcal{P} = \{(x, y, I_{x,y})\}$ and $\mathcal{P}^D = \{(x, y, I^D_{x,y})\}$ denote the set of interest points in the original image and the denoised image, with x and y being m and n integers from $\{1, ..., M\}$ and $\{1, ..., N\}$ that correspond to the coordinates of the interest points. Then, we can write the attention-based quality metrics as

$$\text{aMSE} = \frac{1}{mn}\sum_{(x,y)\in\mathcal{P}}(I^D_{x,y} - I_{x,y})^2, \tag{15}$$

$$\text{aPSNR} = 10\log\left([\max_{(x,y)\in\mathcal{P}} I_{x,y}]^2 \times \frac{1}{\text{aMSE}}\right), \tag{16}$$

$$a\text{SSIM}(\mathcal{P}, \mathcal{P}^D) = \frac{(2\mu_{\mathcal{P}}\mu_{\mathcal{P}^D} + c_1)(2\sigma_{\mathcal{P},\mathcal{P}^D} + c_2)}{(\mu^2_{\mathcal{P}} + \mu^2_{\mathcal{P}^D} + c_1)(\sigma^2_{\mathcal{P}} + \sigma^2_{\mathcal{P}^D} + c_2)}. \tag{17}$$

V. RESULTS

A. DENOISING PERFORMANCE COMPARED TO OTHER ADAPTIVE WINDOWING TECHNIQUES

Section II presented two popular adaptive windowing techniques: the gradient-based method and the image statistics-based method. In this section, we use a synthetic image to compare the performance of these methods with the proposed technique in a denoising application. In this regard, we apply

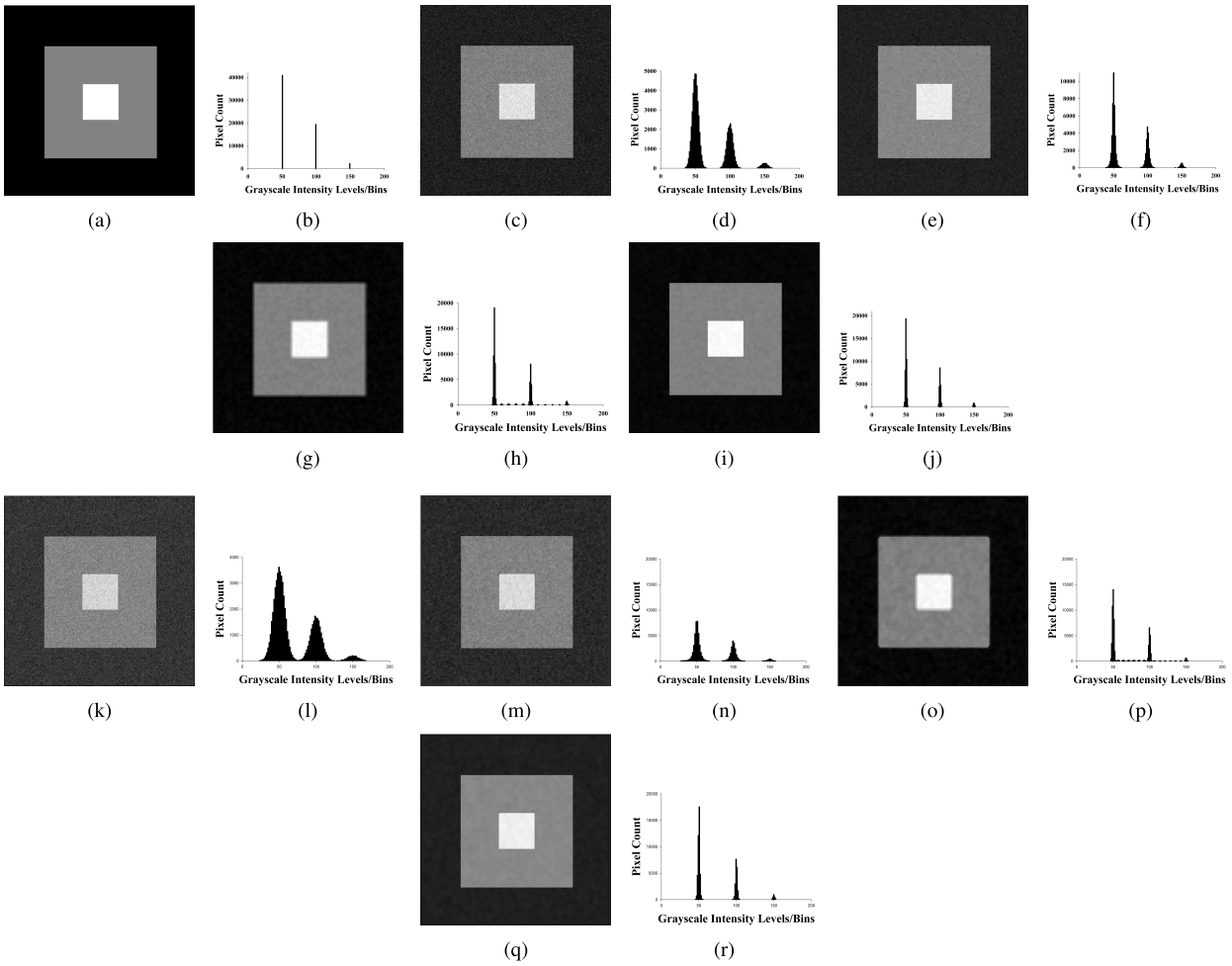

FIGURE 4. Comparisons of different methods based on the original image with the three black, gray, and white regions. The results are evaluated both qualitatively and quantitatively using filtered images and their histograms. The histogram plots indicate that by using the proposed technique, the standard deviation of the noise level becomes the lowest compared with the other adaptive windowing techniques. Plots (a), (c) and (k) show the original image, noise image with standard deviations 4 and 7.5 (which is the maximum σ examined in Fig. 5), respectively. The results of denoising using the gradient-based method, the statistics-based method, and the proposed technique are displayed in {(e), (m)}, {(g),(p)} and {(i),(q)} respectively. The histogram of the intensity levels for each image is displayed adjacent (right) to each image.

the mean filter (see [24] or Appendix I) to a window determined by each adaptive method.

Fig. 4(a) displays a synthetic 8-bit grayscale image with three regions (black, gray, and white), where the gray-level values for each region are $\gamma = 50$, 100, and 150, respectively. Fig. 4(c) presents the same image corrupted with a Gaussian noise with a mean of γ and $\sigma = 4, 7.5$, where σ denotes the standard deviation. Figs. 4(e), 4(g), and 4(i) show the result of the denoising process using the mean filter implemented using the gradient-based method, the statistics-based method, and the proposed window selection technique, respectively. In each case, we assume that the fixed neighborhood used to adaptively determine the window (e.g., $\mathcal{N}_{x_0,y_0}$ in (10)) is a 5×5 square window.

Comparing these figures with the original image in Fig. 4(a) shows that the proposed adaptive technique obtains a better denoising effect than do the other two algorithms. This is also verified by inspecting the histogram of grayscale intensities displayed adjacent to each image—the histogram in Fig. 4(j) is the closest histogram to the histogram of the original image. This is also quantitatively verified by evaluating the quality measures described earlier in Section IV (MSE, PSNR, and SSIM). The result presented in Table 1 uses the images in Fig. 4(a) and (c) indicates that the proposed technique achieves the lowest MSE and simultaneously results in the highest PSNR and SSIM compared to other algorithms.

The results presented here depend on the following two factors: 1) the variance of the noise σ^2 and 2) the size of the initial square window used in the method. Therefore, we further examined the performance of these methods by

- incrementally changing σ and
- incrementally changing the given window size.

In both cases, we evaluated all the quality measures mentioned earlier and plotted the results in Fig 5. This figure shows that in general the proposed technique outperforms the two other techniques in terms of the PSNR, SSIM, and MSE. In other words, the proposed method achieves a better

TABLE 1. The magnitude of region-wise standard deviation, mean squared error (MSE), peak signal-to-noise ratio (PSNR), and structural similarity index measure (SSIM) obtained for mean filter implemented using the three windowing techniques on noisy images of standard deviations, $\sigma = 4$ and $\sigma = 7.5$, respectively.

Filter Type	Noise level, σ = 4						Noise level, σ = 7.5					
	Region-wise σ			MSE	PSNR	SSIM	Region-wise σ			MSE	PSNR	SSIM
	Black	Region	Gray				Black	Region	Gray			
Noise-added image	4.4899	4.5175	4.696	3.5996	83.3	0.7718	7.5071	7.4497	7.4148	5.9725	35.76	0.5628
Gradient-based	2.4195	2.611	2.6869	1.7743	94.84	0.9226	4.4945	4.2147	4.3873	3.1339	38.56	0.8244
Image Statistics-based	2.7492	4.5509	6.1121	1.3953	91.32	0.9618	3.2929	5.4131	7.2025	1.8694	40.8	0.9475
The proposed	0.8838	0.9201	0.9689	0.6995	114.02	0.9929	2.4591	1.3115	1.3141	1.1331	42.97	0.9875

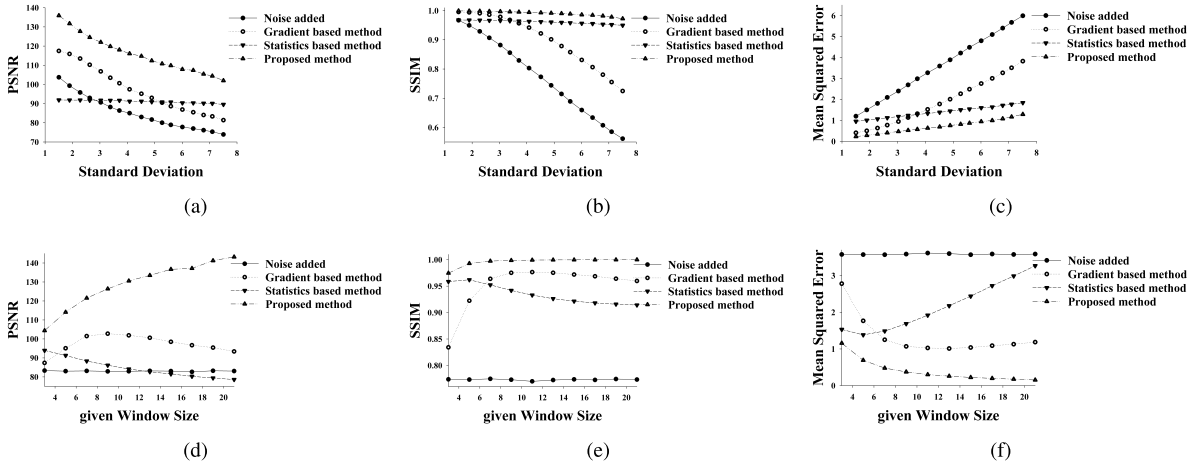

FIGURE 5. The performance of the three adaptive techniques as a function of the noise variance and the size of the initial given neighborhood (a square window). The performance is measured based on the PSNR, SSIM, and MSE. The proposed technique generally outperforms the other methods (based on image gradient and statistics) in terms of signal enhancement, structure preservation, and error reduction.

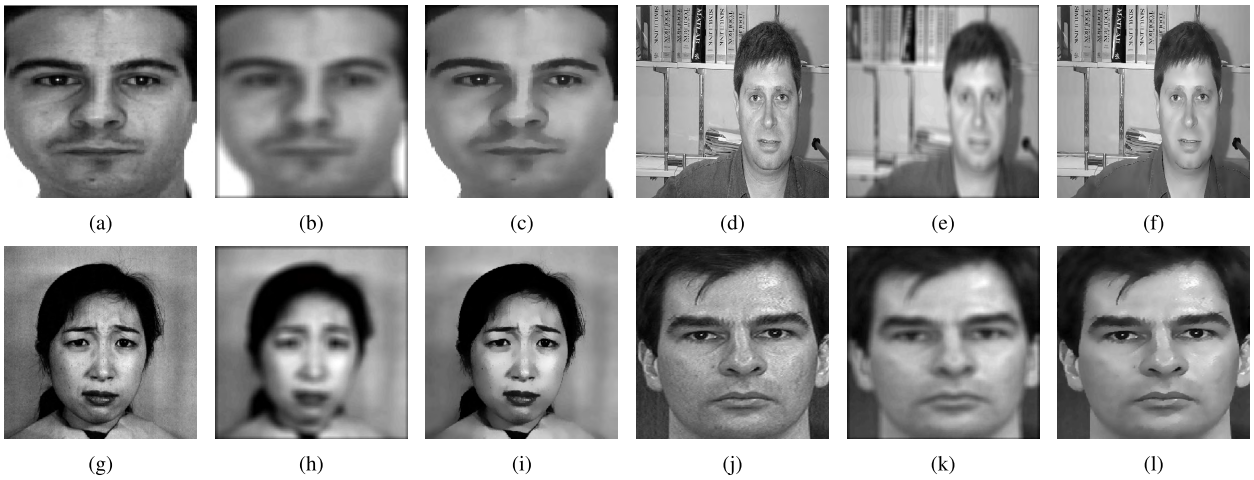

FIGURE 6. A comparison of the mean filter with and without the proposed adaptive window selection technique. (a), (d), (g) and (j) show the original images; (b), (e), (h) and (i) show the conventional mean filter; and (c), (f), (i) and (l) show the mean filter with the proposed adaptive window selection technique.

structural similarity, a higher level of maximum signal to noise ratio, and a reduced energy of the pixel-wise error.

B. DENOISING PERFORMANCE OF ADAPTIVE AND NON-ADAPTIVE MEAN, MEDIAN, AND NON-LOCAL MEAN FILTERS

In this section, we compared the performance of three denoising filters implemented using the classical non-adaptive window selection technique to their implementation using the proposed adaptive window selection technique. We examine the following filters: 1) the mean filter [24], 2) the median filter [24], and 3) the non-local means filter [11], [12] (see Appendix I).

For the mean filter, samples from the following standard face databases are used: AR [43], CalTech [44], GeorgiaTech [45], and JAFFE [46] (see Appendix II for

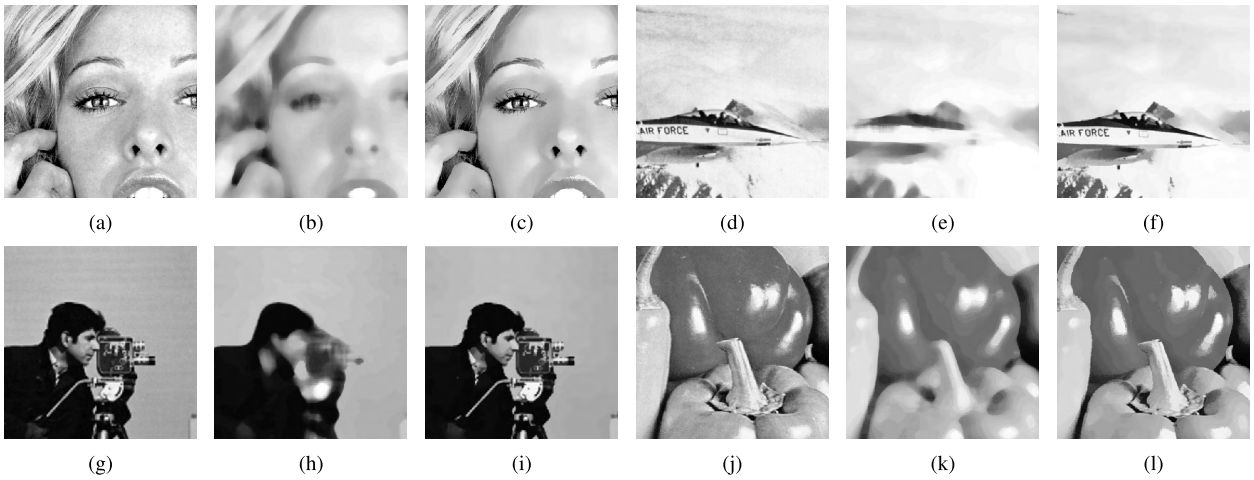

FIGURE 7. A comparison of the median filter with and without the proposed adaptive window selection technique. (a), (d), (g) and (j) show the original images; (b), (e), (h) and (i) show the conventional median filter; and (c), (f), (i) and (l) show the median filter with the proposed adaptive window selection technique.

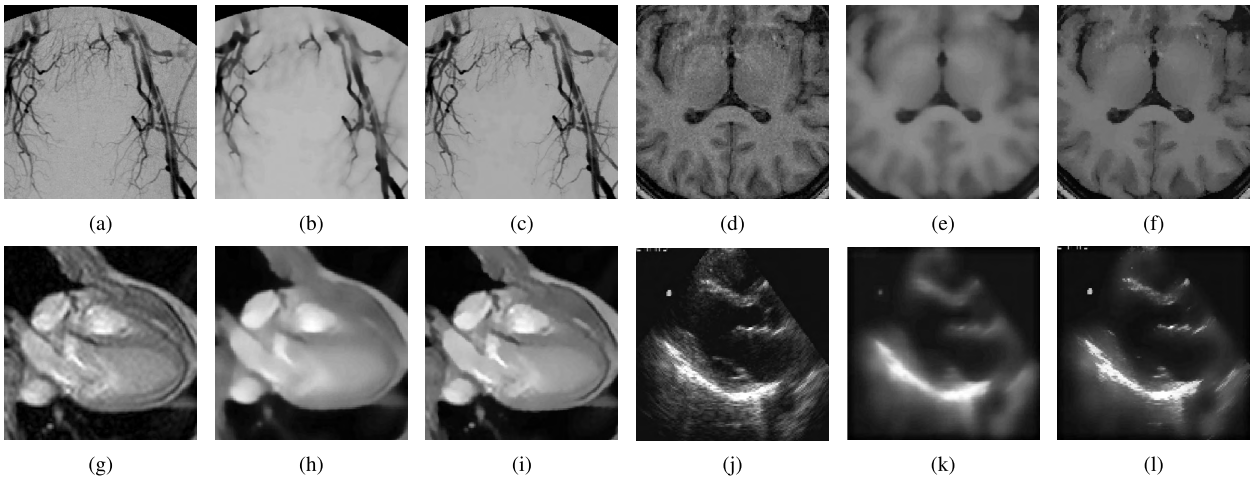

FIGURE 8. A comparison of the NLM filter with and without the proposed adaptive window selection technique. (a), (d), (g) and (j) show the original images; (b), (e), (h) and (i) show the conventional NLM filter; and (c), (f), (i) and (l) show the NLM filter with the proposed adaptive window selection technique.

more details about these datasets). For the median filter, we used standard test images [47] such as blonde woman, jetplane, cameraman, and gray peppers. In the case of the non-local means filter, a sample of medical images for OT-angiograms, MR-brain, MR-cardiac, and Ultrasound-Echocardiogram images is used [48].

Figures 6 to 8 present the comparison of the mean, median, and non-local means filters with and without the proposed adaptive windowing technique. These figures show that introducing the adaptive window selection technique to the implementation of these filters leads to a substantial improvement in the performance over classical non-adaptive techniques. This is further verified by Table 2, in which we tabulated the quality measures, MSE, PSNR and SSIM, and their attention-based counterparts, aMSE, aPSNR and aSSIM. As observed in this table, using the proposed technique in all filters leads to the lowest MSE (aMSE) and the highest PSNR (aPSNR) and SSIM (aSSIM).

C. COMPARISON TO OTHER EDGE-AWARE FILTERS

In this section, we present the result of experiments conducted to compare the performance of various edge-aware filters to a simple mean filter modified using the proposed window adaptation technique. These methods include Anisotropic Diffusion Filter [6], [8], [15], Bilateral Filter [49], [50], Domain Transform Filter [20], Guided Filter [51], and L0 (Gradient Minimization) Smoothing [21] (see Appendix III for a description of these techniques). The performances have been compared both visually (Fig. 9) and numerically (Table 3) on a sample image taken from BrainWeb Database. The metric used to numerically compare the binary images in this figure include MSE, PSNR, SSIM, aMSE, aPSNR, aSSIM, and an additional metric referred as Edge Match Measure (EMM).

EMM measures the amount in which the edge images from a filtered and a noise free image matches to each other. As the first step edge image is processed from a reference noise

TABLE 2. The performance comparison of the proposed adaptive technique with its conventional non-adaptive counterpart for mean, median, and non-local means filters.

Filter Used	Images used	Original method						Proposed approach					
		MSE	PSNR	SSIM	aMSE	aSNR	aSSIM	MSE	PSNR	SSIM	aMSE	aSNR	aSSIM
Mean	AR	8.8	88	0.82	22	76	0.78	3.2	99	0.92	3	88	0.94
	CalTech	15.4	83	0.58	21	69	0.57	4.1	96	0.92	5	81	0.92
	JAFFE	10.9	85	0.64	15	63	0.64	3.9	95	0.83	5	72	0.83
	Georgia Tech	7.4	89	0.71	12	68	0.71	3.9	95	0.85	5	75	0.85
Median	woman blonde	8.8	84	0.74	13	74	0.71	3.8	92	0.93	5	82	0.93
	jetplane	9.1	86	0.74	16	79	0.72	3.3	96	0.94	4	90	0.94
	camera man	8.0	89	0.75	12	67	0.75	3.4	98	0.91	4	76	0.91
	peppers gray	7.3	88	0.83	12	70	0.81	3.9	94	0.92	5	77	0.92
Non-local Means	Sample1 OT	3.2	94	0.92	6	80	0.90	1.6	100	0.97	2	85	0.97
	Sample2 MR	8.1	90	0.68	12	40	0.68	4.4	96	0.83	6	46	0.83
	Sample3 MR	4.5	95	0.82	8	38	0.81	2.9	99	0.89	5	41	0.88
	Sample4 US	9.3	88	0.67	20	52	0.65	6.6	91	0.79	10	57	0.80

TABLE 3. Comparison against various Edge-aware filters for noise reduction and edge preservation computed on the binary images in Fig. 9.

Filter Type	MSE	SSIM	PSNR	aMSE	aSSIM	aPSNR	EMM
Anisotropic Diffusion Filter	8.1757	0.7907	38.5869	0.00198	0.7949	46.2037	28.8079
Bilateral Filter	8.2843	0.8126	38.5295	0.00218	0.8378	47.4624	29.3789
Domain Transform Filter	8.9907	0.7942	38.1742	0.00246	0.8182	47.3569	30.7148
Guided Filter	9.8633	0.7684	37.7719	0.00283	0.7718	47.2934	27.6661
L0 Smoothing	10.3492	0.7389	37.5631	0.00272	0.7257	47.4622	28.1914
Modified Mean Filter (proposed)	6.9144	0.8592	39.3146	0.00169	**0.8956**	**47.5023**	13.9187
Modified Median Filter (proposed)	**6.7306**	**0.8600**	**39.4316**	**0.00158**	0.8922	47.2774	15.4031
Modified NLM Filter (proposed)	6.7682	0.8600	39.4074	0.00161	0.8932	47.3760	**13.1309**

free image and computes the pixel-wise difference against the filtered version of its noise added variants. Further counts the number of shifted, false, and missing edge pixels (such cases will reflect in the difference image) and compares against the pixel counts from noise free edge image (which is the ground truth).

Let $x = 1, .., M, y = 1, .., N$ and denote the original binary image (ground truth) and the denoised binary image by $\mathcal{I} = \{(x, y, I_{x,y})\}$ and $\mathcal{I}^F = \{(x, y, I^F_{x,y})\}$, respectively. Deviation in edges of the noisy filtered image $\mathcal{I}^F$ compared to reference edge noise free image $\mathcal{I}$ is calculated as $\mathcal{I}^{Diff} = |\mathcal{I} - \mathcal{I}^F|$. Then

$$\text{EMM} = 100 \times \frac{N_{Diff}}{N}, \tag{18}$$

where N_{Diff} and N denote the number of pixels in $\mathcal{I}^{Diff}$ and $\mathcal{I}$, respectively.

The results are computed using a sample noise free T1-weighted Magnetic Resonance image from BrainWeb Database [52]–[54] and a 3% noise added version of the same. Thus EMM is computed as the percentage of change in edge pixel locations with respect to a ground truth edge image. When metric other than EMM is considered, the proposed method generates denoised images with better or comparable performance compared to other edge-aware filters (see Table 3). On other hand, in terms of EMM the proposed method displayed a significant improvement in edge preservation (with minimum edge position shifts, lost edge regions, and false edges) against other methods. Table 3 also includes the results for modified versions of Median filter and Non-local Means filter using the proposed windowing technique. The trend is very similar to the modified simple mean filter.

VI. DISCUSSION

A. EFFECT ON SUBSEQUENT ANALYSES: AN APPLICATION TO EDGE DETECTION

Image denoising is generally applied as the first step in various applications, such as edge detection, artifact suppression, and image segmentation, and as such affects the performance of all subsequent analyses. In this section, we present an example to demonstrate the effect of the proposed adaptive technique in a succeeding edge detection application. The quality of edge detection based on the primal sketch model is determined with the help of sketchable and unsketchable edge responses [55]. Sketchability refers to desired edges, while unsketchabe edges refer to the false edges formed due to intra-region variability and noise.

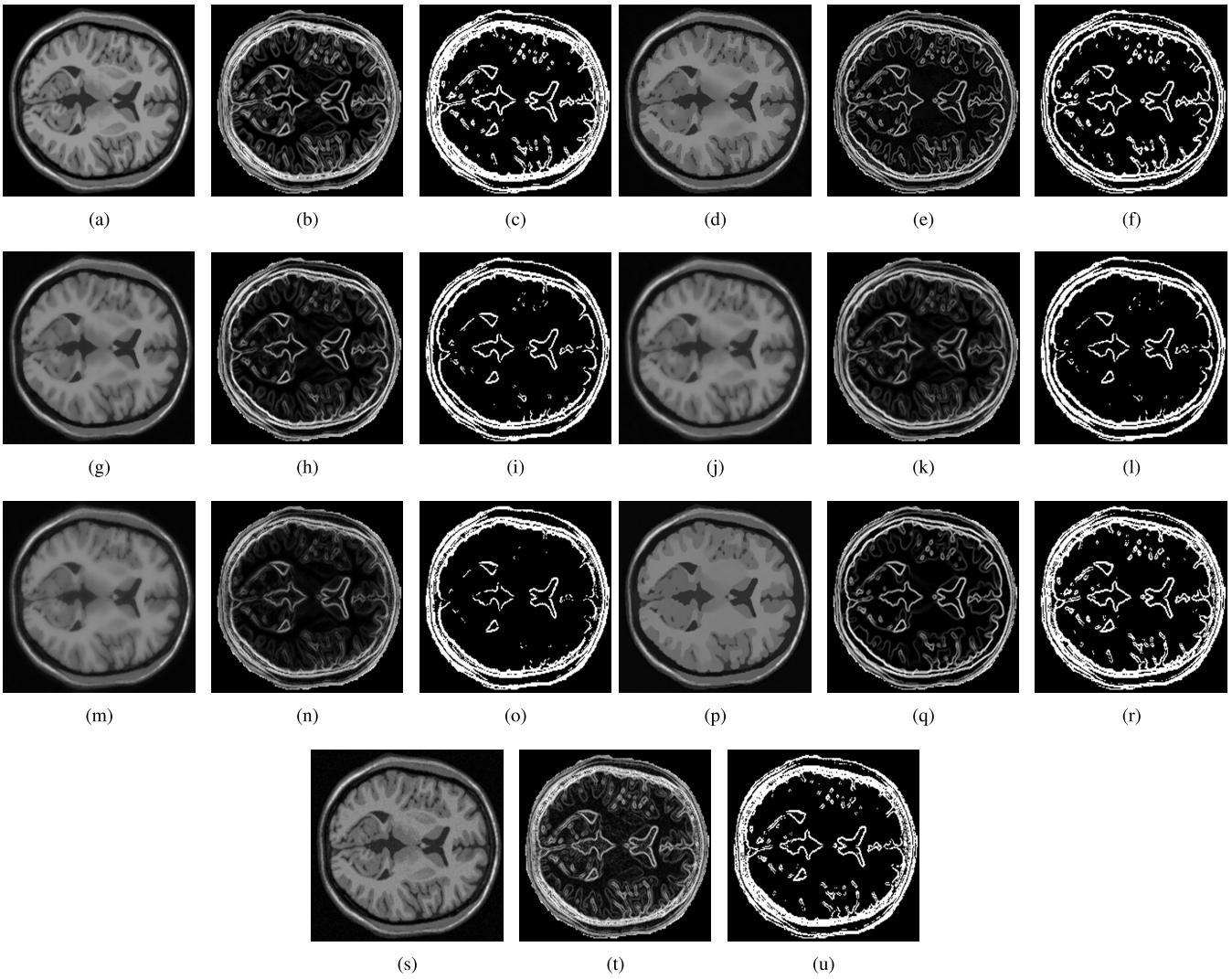

FIGURE 9. Edge preservation and removal of undesired edges are compared for modified mean filter (proposed) against various edge-aware filters Anisotropic Diffusion Filter, Bilateral Filter, Domain Transform Filter, Guided Filter, and L0 (Gradient Minimization) Smoothing. The order of sub-figures are Original image with 0% added noise, Anisotropic Diffusion Filter, Bilateral Filter, Domain Transform Filter, Guided Filter, and L0 (Gradient Minimization) Smoothing and proposed method (modified mean filter). There are three sub-figures per filter type which are grayscale image, edge transitions, and its binary versions; Noise free image (a), (b) and (c) Anisotropic Diffusion Filter (d), (e) and (f) Bilateral Filter (g), (h) and (i) Domain Transform Filter (j), (k) and (l) Guided Filter (m), (n) and (o) L0 (Gradient Minimization) Smoothing (p), (q) and (r) and the proposed method (s), (t) and (u).

Fig. 10 compares the quality of an edge detection algorithm applied to an image denoised using a mean filter. Similar to Section V-A, the mean filter has been implemented on a window determined by three adaptive techniques: the proposed algorithm, the gradient-based method, and the statistics-based method. As observed in this figure, the original image results in a significant number of spurious edges, whereas using adaptive denoising methods before edge detection produces more meaningful edge features. Nevertheless, the denoising process implemented using the proposed window selection technique leads to a binary image (presented in Fig. 10(h)) that presents the fewest spurious structures compared to the results of the other two methods (displayed in Figs. 10(f) and (g)). The position and thickness of edges in Fig. 10(h) are similar to those of the original image and demonstrate a substantial reduction in unsketchable edges and a good preservation of sketchable edges.

B. SIMILARITY THRESHOLD

A key parameter affecting the performance of the proposed technique is the similarity threshold η used in (11). To examine its effect, we conducted a set of simulations using a sample of a brain image taken from the BrainWeb database [52]–[54]. This database provides a simulation of magnetic resonance (MR) brain images with various noise levels: 1%, 3%, 5%, 7%, and 9%. The noise levels are defined as the standard deviation of an additive noise. The database contains different MR modalities such as proton density and T1- and T2-weighted images. Fig. 11 shows a plot of SSIM as a

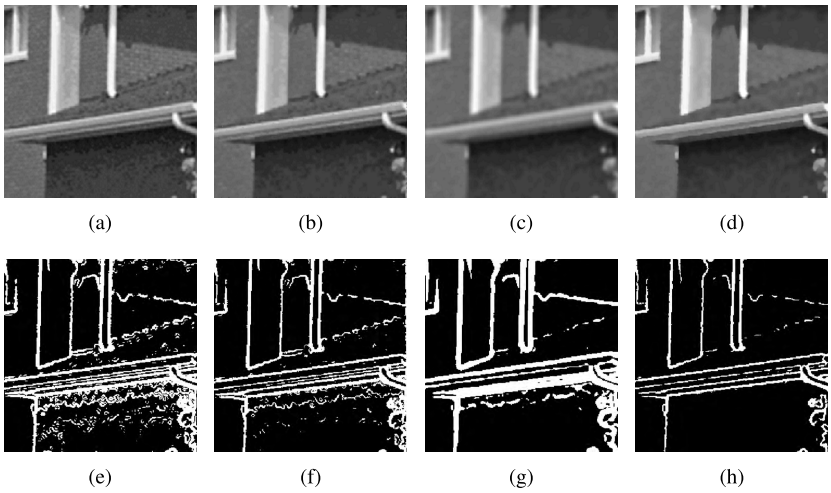

FIGURE 10. The performance of edge detection under the proposed method is compared against existing methods, gradient-based [16] and image statistics-based [16]–[19]. A given image (a) denoised under the mean filter modified using gradient-based method (b), statistics-based method (c) and the proposed method (d) is shown as (e), (f), (g), and (h), respectively. The original version resulted in both desired (sketchable) and undesired (unsketchable) edges (e), where the existing methods are found to fail in either completed noise reduction (f - spurious edges present) or structure preservation (g - thickening of edges). The proposed method (h) achieves a comparatively better trade off between structure preservation (desired edges similar to e) and the removal of spurious edges.

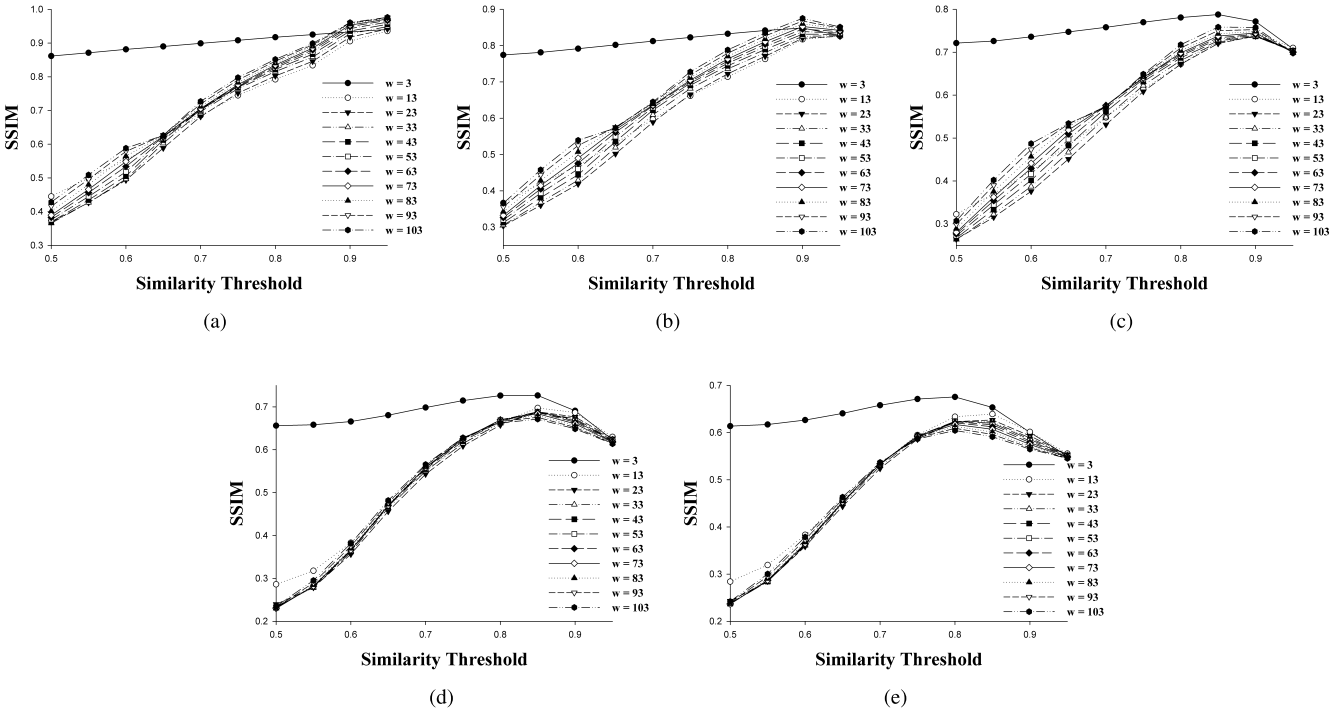

FIGURE 11. Choosing the optimal similarity threshold. The SSIM is plotted as a function of the similarity threshold for various window sizes (3 to 103 in steps of 10) using noisy (a) 1%, (b) 3%, (c) 5%, (d) 7%, and (e) 9% T2-weighted image samples from the BrainWeb database.

function of the similarity threshold for T2-weighted image samples. This result shows that the performance deteriorates as η decreases, and depending on the noise level, the optimal similarity threshold η varies in range from 0.8 to 0.95. The optimal similarity thresholds are tabulated in Table 4. The deteriorating effect of decreasing η is attributed to the participation of a larger number of heterogeneous intensity regions in the denoising process. This effect leads to an increased blurring artifact along with a larger structural distortion.

This result shows that depending on the application, there will be a possible range of acceptable similarity

TABLE 4. Noise level and corresponding optimal similarity threshold.

Noise Level	Optimal Similarity Threshold
1%	0.95
3%	0.90
5%	0.85
7%	0.85
9%	0.80

thresholds η. Nonetheless, a smaller η results in narrow low-contrast regions that are more washed out. Depending on the context, this occurrence could be a drawback or an advantage. For example, in certain computer vision applications, such as text extraction from old handwritten documents (using ink pen), it is desired to remove such narrow low-contrast details. A preprocessing step necessary in the extraction is image binarization. In this context, image binarization is applied before character localization and extraction. One of the main challenges in the binarization of scanned copies of such documents is the impression of text written on the other side of the paper. In image binarization, such impressions are reflected in the form of undesired character-like structures. The separation of actual details from undesired details becomes difficult once an image is binarized.

Consequently, removing the low-contrast narrow regions is thus an advantage because the desired characters and the impressions possess a high and low contrast, respectively. Fig. 12 shows this application, in which a mean filter implemented using the proposed adaptive mechanism at an appropriate threshold has been applied to remove undesired text impressions. The figure displays the binarized image before and after the application of the proposed method. Here, the trade-off is clearly displayed; smoothing successfully removed most of the undesired impressions and resulted in a clear binarized image, with the desired text being well preserved. Fig. 13 displays plots of the three denoising quality measures mentioned earlier (PSNR, SSIM, and MSE) as functions of η for the example presented in Fig. 11. This result shows that the performance abruptly deteriorates as η decreases, with $\eta = 0.85$ roughly being the tipping point. As mentioned earlier, decreasing η has a deteriorating effect that leads to an increased blurring artifact (low PSNR and high MSE) along with a large structural distortion (low SSIM).

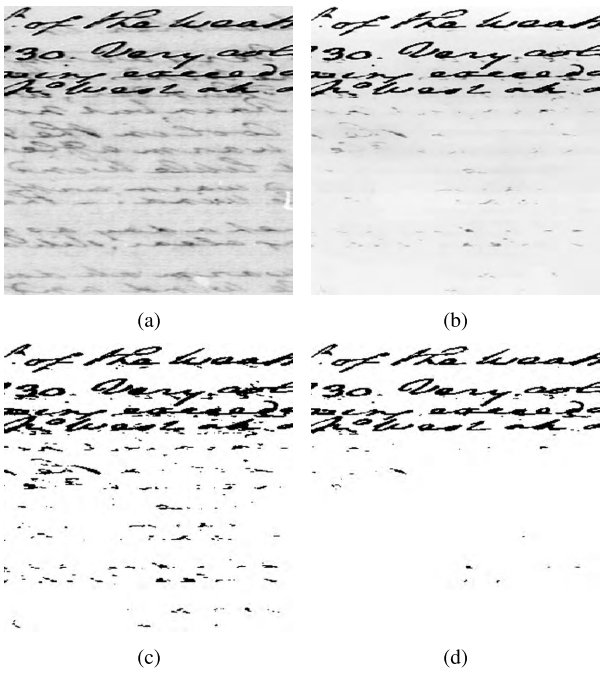

FIGURE 12. Removal of undesired text impressions from a scanned copy of old handwritten documents. a) Original scanned image with text impressions from the other side of the paper, b) denoised image by the mean filter implemented using the proposed adaptive technique, c) binarized version of the original image displayed in (a), and d) binarized version of the denoised image displayed in (b).

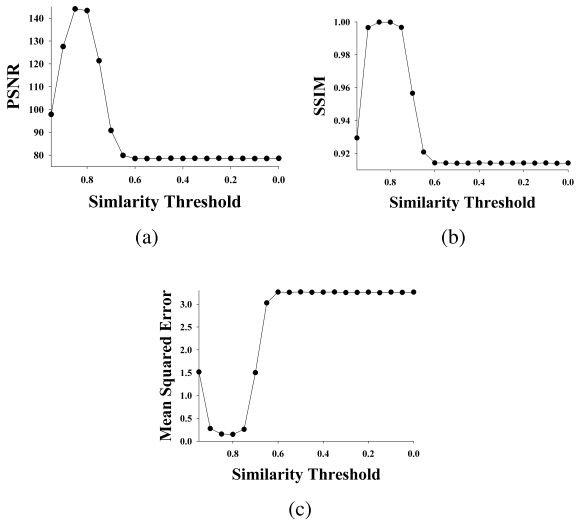

FIGURE 13. Various quality measures, PSNR, SSIM, and MSE, as functions of the similarity threshold. From the results, it is observed that for the given image, the proposed method achieves its best performance at a threshold of approximately 0.85.

Thus, we would like to conclude that the proposed method keeps a significant trade-off between image noise removal and blurring artifact.

VII. CONCLUSION

We have addressed the problem of adaptive window selection for image denoising using a psychological model suitable for quantifying the similarity between stimuli (here, pixels). The underlying hypothesis of the present study is that using a pixel similarity metric analogous to the metric that the human cognitive system uses improves the quality of adaptive window selection in denoising applications. For a given neighborhood of a candidate pixel, the method computes Shepard's similarity metric to determine the set of the most homogeneous neighboring pixels. One may then denoise the image by applying an arbitrary filter to the selected region. We examined the proposed method in various settings under both synthetic and real image samples. We observed that the proposed technique outperforms some well-known adaptive and non-adaptive denoising techniques. The improvement

is confirmed both visually and numerically. For the quality assessment metrics, we used the measures of signal-to-noise strength and structural preservation. This result demonstrates the efficacy of the proposed psychologically inspired approach to window selection and image denoising. The next natural step in this line of work is to extend and analyze the performance of the proposed technique for color image denoising. Furthermore, we will study the possibility of determining the similarity threshold η from the image content.

APPENDIX I
MEAN FILTER, MEDIAN FILTER, AND NON-LOCAL MEAN FILTER

Mean Filter is the simplest form of noise reduction, where the denoised pixel is expressed as the mean of the intensities within a given neighborhood. Let $\mathcal{N}_{x_0,y_0}$ and $I^D_{x_0,y_0}$ denote a given neighborhood of pixel (x_0, y_0, I_{x_0,y_0}) and the value of the intensity level of the pixel after applying the mean filter (i.e., denoised), respectively. Then,

$$I^D_{x_0,y_0} = \frac{1}{n_0} \sum_{(x,y)\in\mathcal{N}_{x_0,y_0}} I_{x,y} \tag{19}$$

where n_0 is the number of pixels in $\mathcal{N}_{x_0,y_0}$.

Median Filter is a non-linear filter, where for each pixel (x_0, y_0, I_{x_0,y_0}), the intensity level is replaced by the median value of pixels in a given neighborhood $\mathcal{N}_{x_0,y_0}$.

A. NON-LOCAL MEANS FILTER

This filter computes a set of non-linear weights based on the similarity between a reference pixel and all neighboring pixels. For a pixel (x_0, y_0, I_{x_0,y_0}) and a neighborhood $\mathcal{N}_{x_0,y_0}$, the denoised pixel intensity is expressed as

$$I^D(x_0, y_0) = \frac{1}{z_0} \sum_{(x,y)\in\mathcal{N}_{x_0,y_0}} I_{x_0,y_0} \times f[I_{x_0,y_0}, I_{x,y}], \tag{20}$$

where

$$z_0 = \sum_{(x,y)\in\mathcal{N}_{x_0,y_0}} f[I_{x_0,y_0}, I_{x,y}], \tag{21}$$

and for a Gaussian weighting function,

$$f[I_{x_0,y_0}, I_{x,y}] = e^{-\frac{|I^D_{x_0,y_0} - I^D_{x,y}|^2}{h^2}}, \tag{22}$$

in which for any (x, y), $I^D_{x,y}$ is the mean value determined from (19) for some given neighborhood and h is a decay parameter controlling the degree of smoothing in NLM filtering.

APPENDIX II
DATABASES

The AR Face Database [43] provides face images under different image conditions such as facial expressions, illumination conditions, and occlusions (sunglasses and scarves). The CalTech face database [44] is a frontal face dataset that includes 450 face images with 896 × 592 pixels in JPEG format and approximately 27 unique people with different lighting/expressions/backgrounds. This database represents conditions such as variable and complex backgrounds, scale variations of the face, and varying face positions. The Georgia Tech (GTech) [45] face image database includes frontal and/or tilted faces with different facial expressions, lighting conditions and scale. The JAFFE database [46] contains 213 images of 7 facial expressions (6 basic facial expressions + 1 neutral) posed by 10 female Japanese models. Each image has been rated on 6 emotion adjectives by 60 Japanese subjects.

APPENDIX III
VARIOUS EDGE-AWARE FILTERS

B. ANISOTROPIC DIFFUSION

The anisotropic diffusion equation is given by [15],

$$\frac{tialI(x_0, y_0, t)}{tialt} = div\big[g(||\nabla I(x_0, y_0, t)||)\nabla I(x_0, y_0, t)\big] \tag{23}$$

in which $||.||$ denotes the L_2 norm, t is the time (scale) parameter, $g(.)$ is known as the conductance function, and div, ∇, $I(x_0, y_0, t)$ and $I(x_0, y_0, 0)$ denote the divergence operator, the gradient operator, the derived image pixel at scale parameter t, and the initial image pixel, respectively. A possible choice of the conductance function is

$$g(\nabla I) = e^{-(||\nabla I||/K)^2} \tag{24}$$

with K being a constant that controls the diffusion rate. In [15], Perona and malik proposed a discretized implementation of (23) in which

$$I^{D,j+1}_{x_0,y_0} = I^{D,j}_{x_0,y_0} + \frac{\lambda}{|\eta_{x_0,y_0}|} \sum_{i\in\eta_{x_0,y_0}} g(|\nabla_i I^{D,j}_{x_0,y_0}|)\nabla_i I^{D,j}_{x_0,y_0}, \tag{25}$$

where $\lambda \in [0, 1]$. j denotes the iteration index, $\eta_{x_0,y_0} = \{N, S, E, W\}$ denotes the 4-nearest-neighbors of pixel at (x, y), i.e., the pixels to the north (N), south (S), east (E), and west (W), and

$$\begin{aligned}
\nabla_N I_{x_0,y_0} &\triangleq I_{x_0-1,y_0} - I_{x_0,y_0},\\
\nabla_S I_{x_0,y_0} &\triangleq I_{x_0+1,y_0} - I_{x_0,y_0},\\
\nabla_E I_{x_0,y_0} &\triangleq I_{x_0,y_0+1} - I_{x_0,y_0},\\
\nabla_W I_{x_0,y_0} &\triangleq I_{x_0,y_0-1} - I_{x_0,y_0}.
\end{aligned} \tag{26}$$

The iterative process stops after a pre-determined number of iterations M. Thus, the value of intensity level of pixel (x_0, y_0, I_{x_0,y_0}) after applying the denoising anisotropic diffusion filter is $I^D_{x_0,y_0} \triangleq I^{D,M}_{x_0,y_0}$.

C. BILATERAL FILTER

In this method, the filtered image is obtained as a weighted average of pixels in the given image. In particular, for a given

pixel (x_0, y_0, I_{x_0,y_0}) and a neighborhood $\mathcal{N}_{x_0,y_0}$, the denoised pixel intensity is expressed as [50]

$$I^D_{x_0,y_0} = \frac{1}{W_{x_0,y_0}} \sum_{(x,y)\in\mathcal{N}_{x_0,y_0}} \Big[G_{\sigma_s}\left(||(x,y)-(x_0,y_0)||\right) \times G_{\sigma_r}\left(|I_{x,y}-I_{x_0,y_0}|\right) I_{x,y} \Big], \quad (27)$$

where $||.||$ denotes the L_2 norm, W_{x_0,y_0} is a normalization factor, and $G_\sigma(x)$ is the 2D Gaussian kernel given by

$$G_\sigma(x) = \frac{1}{2\pi\sigma^2} e^{\left(\frac{-x^2}{2\sigma^2}\right)}. \quad (28)$$

D. DOMAIN TRANSFORM FILTER

Let I_x be a 1D signal that maps elements of the spatial domain Ω to real line $\mathbb{R}$. The idea of domain transform is to obtain a transform T, that satisfies [20],

$$|T(x_i, I_{x_i}) - T(x_{i+1}, I_{x_{i+1}})| = ||(x_i, I_{x_i}) - (x_{i+1}, I_{x_{i+1}})|| \quad (29)$$

in which x_i, $i = 1, ..., N$ denotes a sample of Ω, $|.|$ is the absolute value, and $||.||$ is some metric. Let $t(x) \triangleq T(x, I_x)$. Taking $||.||$ as the L_1 norm, and $t(x)$ as a monotonic increasing function, it can be shown that [20],

$$t'(x) = 1 + |I'_x|, \quad (30)$$

where $t'(x)$ and I'_x denote the derivative of $t(x)$ and I_x with respect to x, respectively. This transformation reduces the $\mathbb{R}^2$ domain of (x, I_x) to $\mathbb{R}$ in which a 1D filtering operation is applied. Denoting the variances of the filter in the signal's original spatial domain and range by σ_s^2 and σ_r^2, respectively, one may encode these values in the transformation to obtain the final form of domain transform as [20]

$$t'(x) = 1 + \frac{\sigma_s}{\sigma_r}|I'_x|. \quad (31)$$

In 2D signals (an image), the transformation and filtering is performed along each row, and then along each column. Various forms of 1D filters including normalized convolution, interpolated convolution, and recursive filtering have been proposed in [20]. In our experiments, we use the normalized convolution filter.

E. GUIDED FILTER

For a given pixel (x_0, y_0, I_{x_0,y_0}) and a neighborhood $\mathcal{N}_{x_0,y_0}$, this method assumed that the intensity level of the denoised pixel is given by an affine mapping of the input pixel intensity as follows [51]:

$$I^D_{x_0,y_0} = a_{\mathcal{N}_{x_0,y_0}} I_{x_0,y_0} + b_{\mathcal{N}_{x_0,y_0}}, \quad (32)$$

where $a_{\mathcal{N}_{x_0,y_0}}$ and $b_{\mathcal{N}_{x_0,y_0}}$ are some coefficients assumed to be constant in $\mathcal{N}_{x_0,y_0}$. To determine these coefficients, it is then desired to minimize the following energy function:

$$E(a_{\mathcal{N}_{x_0,y_0}}, b_{\mathcal{N}_{x_0,y_0}}) = \sum_{(x,y)\in\mathcal{N}_{x_0,y_0}} \left((a_{\mathcal{N}_{x_0,y_0}} I_{x,y} + b_{\mathcal{N}_{x_0,y_0}} - I_{x,y})^2 + \epsilon a^2_{\mathcal{N}_{x_0,y_0}} \right), \quad (33)$$

where ϵ is a regularization parameter that penalizes large values of $a_{\mathcal{N}_{x_0,y_0}}$. Similarity of (33) to the linear ridge regression problem leads to a closed form solution given in [51] for $a_{\mathcal{N}_{x_0,y_0}}$ and $b_{\mathcal{N}_{x_0,y_0}}$.

F. L_0 (GRADIENT MINIMIZATION) SMOOTHING

This smoothing method is designed to enhance highest-contrast edges and, at the same time, eliminating low-amplitude structures. Let $C(\mathcal{I})$ count the number of amplitude changes along the x and y direction (non-zero gradients); to wit,

$$C(\mathcal{I}) = \#\{(x,y) \mid |\nabla_x I_{x,y}| + |\nabla_y I_{x,y}| \neq 0\} \quad (34)$$

where

$$\nabla_x I_{x,y} = I_{x+1,y} - I_{x,y}, \quad \nabla_y I_{x,y} = I_{x,y+1} - I_{x,y}. \quad (35)$$

Then the denoised image is estimated using the following objective function, which controls the trade-off between the structure similarity and sharpening major edges:

$$\min_{\mathcal{I}^D} \left\{ \sum_{(x,y)} (I_{x,y} - I^D_{x,y})^2 + \lambda C(\mathcal{I}) \right\}. \quad (36)$$

with $\lambda > 0$ being a weight controlling the trade-off. An iterative process has been proposed in [21] to solve (36).

REFERENCES

[1] C. Tang, X. Yang, and G. Zhai, "Noise estimation of natural images via statistical analysis and noise injection," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 25, no. 8, pp. 1283–1294, Aug. 2015.

[2] Y. Wu, B. Tracey, P. Natarajan, and J. P. Noonan, "Probabilistic non-local means," *IEEE Signal Process. Lett.*, vol. 20, no. 8, pp. 763–766, Aug. 2013.

[3] S. Roth and M. J. Black, "Fields of experts: A framework for learning image priors," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, vol. 2. Jun. 2005, pp. 860–867.

[4] J. Portilla, V. Strela, M. J. Wainwright, and E. P. Simoncelli, "Image denoising using scale mixtures of Gaussians in the wavelet domain," *IEEE Trans. Image Process.*, vol. 12, no. 11, pp. 1338–1351, Nov. 2003.

[5] J. Mohana, V. Krishnaveni, and Y. Guo, "A survey on the magnetic resonance image denoising methods," *Biomed. Signal Process. Control*, vol. 9, pp. 56–69, Jan. 2014.

[6] J. Weickert, *Anisotropic Diffusion in Image Processing*, vol. 1. Stuttgart, Germany: Teubner, 1998.

[7] K. Chen, "Adaptive smoothing via contextual and local discontinuities," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 10, pp. 1552–1567, Oct. 2005.

[8] M. J. Black, G. Sapiro, D. H. Marimont, and D. Heeger, "Robust anisotropic diffusion," *IEEE Trans. Image Process.*, vol. 7, no. 3, pp. 421–432, Mar. 1998.

[9] M. Mahmoudi and G. Sapiro, "Fast image and video denoising via nonlocal means of similar neighborhoods," *IEEE Signal Process. Lett.*, vol. 12, no. 12, pp. 839–842, Dec. 2005.

[10] Q. Chen and D. Wu, "Image denoising by bounded block matching and 3D filtering," *Signal Process.*, vol. 90, no. 9, pp. 2778–2783, 2010.

[11] A. Buades, B. Coll, and J.-M. Morel, "A non-local algorithm for image denoising," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, vol. 2. San Diego, CA, USA, Jun. 2005, pp. 60–65.

[12] B. K. S. Kumar, "Image denoising based on non-local means filter and its method noise thresholding," *Signal, Image Video Process.*, vol. 7, no. 6, pp. 1159–1172, 2013.

[13] V. Katkovnik, A. Foi, K. Egiazarian, and J. Astola, "From local kernel to nonlocal multiple-model image denoising," *Int. J. Comput. Vis.*, vol. 86, no. 1, p. 1, 2010.
[14] L. Shao, R. Yan, X. Li, and Y. Liu, "From heuristic optimization to dictionary learning: A review and comprehensive comparison of image denoising algorithms," *IEEE Trans. Cybern.*, vol. 44, no. 7, pp. 1001–1013, Jul. 2014.
[15] P. Perona and J. Malik, "Scale-space and edge detection using anisotropic diffusion," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 12, no. 7, pp. 629–639, Jul. 1990.
[16] A. Goshtasby and M. Satter, "An adaptive window mechanism for image smoothing," *Comput. Vis. Image Understand.*, vol. 111, no. 2, pp. 155–169, 2008.
[17] S. Ali and R. Burge, "New automatic techniques for smoothing and segmenting SAR images," *Signal Process.*, vol. 14, no. 4, pp. 335–346, 1988.
[18] D. Sheet, S. Pal, A. Chakraborty, J. Chatterjee, and A. K. Ray, "Image quality assessment for performance evaluation of despeckle filters in optical coherence tomography of human skin," in *Proc. IEEE EMBS Conf. Biomed. Eng. Sci. (IECBES)*, Dec. 2010, pp. 499–504.
[19] D. Sheet, S. Pal, A. Chakraborty, J. Chatterjee, and A. K. Ray, "Visual importance pooling for image quality assessment of despeckle filters in optical coherence tomography," in *Proc. Int. Conf. Syst. Med. Biol. (ICSMB)*, Dec. 2010, pp. 102–107.
[20] E. S. L. Gastal and M. M. Oliveira, "Domain transform for edge-aware image and video processing," *ACM Trans. Graph.*, vol. 30, no. 4, p. 69, 2011.
[21] L. Xu, C. Lu, Y. Xu, and J. Jia, "Image smoothing via L0 gradient minimization," *ACM Trans. Graph.*, vol. 30, no. 6, p. 174, 2011.
[22] C. Kervrann, "An adaptive window approach for image smoothing and structures preserving," in *Proc. Eur. Conf. Comput. Vis.*, 2004, pp. 132–144.
[23] C. Kervrann and J. Boulanger, "Local adaptivity to variable smoothness for exemplar-based image regularization and representation," *Int. J. Comput. Vis.*, vol. 79, no. 1, pp. 45–69, Aug. 2008.
[24] F. Y. Shih, *Image Processing and Pattern Recognition: Fundamentals and Techniques*. Hoboken, NJ, USA: Wiley, May 2010.
[25] R. N. Shepard, "Toward a universal law of generalization for psychological science," *Science*, vol. 237, no. 4820, pp. 1317–1323, 1987.
[26] F. G. Ashby and N. A. Perrin, "Toward a unified theory of similarity and recognition," *Psychol. Rev.*, vol. 95, no. 1, p. 124, 1988.
[27] A. M. Eskicioglu and P. S. Fisher, "Image quality measures and their performance," *IEEE Trans. Commun.*, vol. 43, no. 12, pp. 2959–2965, Dec. 1995.
[28] R. Sakuldee and S. Udomhunsakul, "Objective performance of compressed image quality assessments," *Int. J. Comput. Sci.*, vol. 2, no. 4, pp. 258–267, 2007.
[29] A. Hore and D. Ziou, "Image quality metrics: PSNR vs. SSIM," in *Proc. 20th Int. Conf. Pattern Recognit. (ICPR)*, Aug. 2010, pp. 2366–2369.
[30] Q. Huynh-Thu and M. Ghanbari, "Scope of validity of PSNR in image/video quality assessment," *Electron. Lett.*, vol. 44, no. 13, pp. 800–801, Jun. 2008.
[31] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, Apr. 2004.
[32] A. Rehman and Z. Wang, "Reduced-reference SSIM estimation," in *Proc. 17th IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2010, pp. 289–292.
[33] A. Loza, L. Mihaylova, N. Canagarajah, and D. Bull, "Structural similarity-based object tracking in video sequences," in *Proc. 9th Int. Conf. Inf. Fusion*, 2006, pp. 1–6.
[34] A. Baumberg, "Reliable feature matching across widely separated views," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, vol. 1. Jun. 2000, pp. 774–781.
[35] M. Brown and D. G. Lowe, "Invariant features from interest point groups," in *Proc. BMVC*, vol. 4. pp. 23.1–23.10, Sep. 2002.
[36] H. Bay, T. Tuytelaars, and L. van Gool, "SURF: Speeded up robust features," in *Proc. Eur. Conf. Comput. Vis.*, 2006, pp. 404–417.
[37] A. E. Johnson and M. Hebert, "Recognizing objects by matching oriented points," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, Jun. 1997, pp. 684–689.
[38] D. G. Lowe, "Object recognition from local scale-invariant features," in *Proc. IEEE Int. Conf. Comput. Vis.*, vol. 2. Sep. 1999, pp. 1150–1157.
[39] K. Mikolajczyk and C. Schmid, "Indexing based on scale invariant interest points," in *Proc. 8th IEEE Int. Conf. Comput. Vis. (ICCV)*, vol. 1. Jul. 2001, pp. 525–531.
[40] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. J. Comput. Vis.*, vol. 60, no. 2, pp. 91–110, 2004.
[41] K. Mikolajczyk and C. Schmid, "A performance evaluation of local descriptors," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 10, pp. 1615–1630, Oct. 2005.
[42] C. Harris and M. Stephens, "A combined corner and edge detector," in *Proc. Alvey Vis. Conf.*, 1988, vol. 15. no. 50, pp. 5210–5244.
[43] A. M. Martinez, "The ar face database," Central Vigilance Commission, New Delhi, India, Tech. Rep. 24, 1998.
[44] M. Weber and M. Weber, "Caltech frontal face database," California Inst. Technol., Pasadena, CA, USA, Tech. Rep., 1999.
[45] L. Chen, H. Man, and A. V. Nefian, "Face recognition based on multi-class mapping of fisher scores," *Pattern Recognit.*, vol. 38, no. 6, pp. 799–811, 2005.
[46] M. Lyons, S. Akamatsu, M. Kamachi, and J. Gyoba, "Coding facial expressions with Gabor wavelets," in *Proc. 3rd IEEE Int. Conf. Autom. Face Gesture Recognit.*, Apr. 1998, pp. 200–205.
[47] *Image Processing Place: Repository*. Accessed on Sep. 30, 2016. [Online]. Available: http://www.imageprocessingplace.com/downloads_V3/root_downloads/ image_databases/standard_test_images.zip
[48] *Medical Images Repository*. Accessed on Sep. 30, 2016. [Online]. Available: http://www.barre.nom.fr /medical/samples/
[49] D. Barash and D. Comaniciu, "A common framework for nonlinear diffusion, adaptive smoothing, bilateral filtering and mean shift," *Image Vis. Comput.*, vol. 22, no. 1, pp. 73–81, 2004.
[50] M. Elad, "On the origin of the bilateral filter and ways to improve it," *IEEE Trans. Image Process.*, vol. 11, no. 10, pp. 1141–1151, Oct. 2002.
[51] K. He, J. Sun, and X. Tang, "Guided image filtering," in *Proc. Eur. Conf. Comput. Vis.*, 2010, pp. 1–14.
[52] R. K. S. Kwan, A. C. Evans, and G. B. Pike, "MRI simulation-based evaluation of image-processing and classification methods," *IEEE Trans. Med. Imag.*, vol. 18, no. 11, pp. 1085–1097, Nov. 1999.
[53] R. K.-S. Kwan, A. C. Evans, and G. B. Pike, "An extensible MRI simulator for post-processing evaluation," in *Visualization in Biomedical Computing (VBC)* (Lecture Notes in Computer Science), vol. 1131. Springer-Verlag, 1996, pp. 135–140.
[54] D. L. Collins *et al.*, "Design and construction of a realistic digital brain phantom," *IEEE Trans. Med. Imag.*, vol. 17, no. 3, pp. 463–468, Jun. 1998.
[55] J. J. Mathew and A. P. James, "Spatial stimuli gradient sketch model," *IEEE Signal Process. Lett.*, vol. 22, no. 9, pp. 1336–1339, Sep. 2015.

JOSHIN MATHEW received the M.S. degree in computer science and information technology from the Indian Institute of Information Technology and Management, India. He is currently an Engineer with the Computer Vision Team, ARS T & TT, India. His research interests include image enhancement, feature extraction, medical imaging, object detection, and classification problems.

AMIN ZOLLANVARI (M'10) received the Ph.D. degree in electrical engineering from Texas A&M University, College Station, TX, in 2010. He held a post-doctoral position at the Harvard Medical School and the Brigham and Women's Hospital, Boston, MA, from 2010 to 2012, and then joined the Department of Statistics, Texas A&M University, as an Assistant Research Scientist from 2012 to 2014. He is currently an Assistant Professor with the Department of Electrical and Electronic Engineering, Nazarbayev University. His research interest includes signal processing, machine learning, and bioinformatics.

ALEX PAPPACHEN JAMES (SM'13) received the Ph.D. degree from the Griffith School of Engineering, Griffith University. He is currently the Chair of the Electrical Engineering Department, Nazarbayev University. He is currently involved in brain-inspired circuits, memristor circuits, and algorithms and systems. He has a sustained experience of managing industry and academic projects in board design, VLSI and pattern recognition algorithms, and semiconductor industry. He is currently the Chair of the IEEE Kazakhstan subsection. He is also an Associate Editor of *Human-Centric Computing and Information Sciences*, IEEE Access, IEEE Transactions on Emerging Topics in Computational Intelligence (special issue), and *IET Cyber-Physical Systems* (special issue). He is a Senior Fellow of the Higher Education Academy, U.K.