Research Article

Multiscale Hybrid Nonlocal Means Filtering Using Modified Similarity Measure

Zahid Hussain Shamsi[1,2] and Dai-Gyoung Kim[1]

[1] *Department of Applied Mathematics, Hanyang University (ERICA), Ansan 426-791, Republic of Korea*
[2] *Department of Mathematics, University of the Punjab, Lahore 54590, Pakistan*

Correspondence should be addressed to Dai-Gyoung Kim; dgkim@hanyang.ac.kr

A new multiscale implementation of nonlocal means filtering (MHNLM) for image denoising is proposed. The proposed algorithm also introduces a modification of the similarity measure for patch comparison. Assuming the patch as an oriented surface, the notion of a normal vectors patch is introduced. The inner product of these normal vectors patches is defined and then used in the weighted Euclidean distance of intensity patches as the weight factor. The algorithm involves two steps: the first step is a multiscale implementation of an accelerated nonlocal means filtering in the discrete stationary wavelet domain to obtain a refined version of the noisy patches for later comparison. The next step is to apply the proposed modification of standard nonlocal means filtering to the noisy image using the reference patches obtained in the first step. These refined patches contain less noise, and consequently the computation of normal vectors and partial derivatives is more precise. Experimental results show equivalent or better performance of the proposed algorithm compared to various state-of-the-art algorithms.

1. Introduction

The phenomenon of image degradation is quite natural due to the digitization and quantization processes in signal acquisition devices. Although various image denoising techniques have been extensively studied and effectively employed in the last two decades, the preservation of texture, edges, and fine details during image denoising is an open problem and needs rigorous treatment.

In advance of the nonlocal methodology, a variety of variational [1–6] PDE [7–12] and wavelet based [13–16] approaches were proposed for image denoising that rely on the local features of image data. A major shift towards nonlocal filtering was initiated by the usage of bilateral filtering [17], which exploits spatial and intensity domains for image denoising. In this approach, spatially proximate pixels are given more weights in the similarity measure. Later on, the wavelet-based BLS-GSM [18] method provided the best results in terms of PSNR measure; however, these denoised images contain ringing artifacts and have low visual quality. More recently, Buades et al. [19] introduced nonlocal means filtering for image denoising. Although the PSNR of nonlocal means filtering was found to be less than the wavelet-based BLS-GSM [18], the notion of the patch-based approach combined with the idea of nonlocality has led to an entirely new way of attacking the problem. Earlier, approaches similar to nonlocal means were used for image inpainting [20] and texture synthesis [21]. Ever since the nonlocal means were proposed, more rigorous research for better estimation of parameters or finding suitable similarity measures has improved the performance of nonlocal means filtering for a variety of noise models in various image processing applications.

Kervrann et al. [22] provided a theoretical foundation for an intuitive nonlocal means approach using Bayesian statistics. In this approach, refined adaptive dictionaries of similar patches are obtained using Bayesian estimation, while the irrelevant patches are rejected. Afterwards, these learned dictionaries are used for patch-based comparison using a modified similarity measure. Elad and Aharon [23] proposed learned dictionaries of patches using K-means SVD algorithm and then employed sparse representation using L_0

pseudonorm over the learned dictionaries for construction of denoised image. Tasdizen [24] proposed using the similarity of patches in the principal component analysis domain. Extensive research in patch-based denoising has resulted in state-of-the-art algorithms [25, 26]. Dabov et al. [25] achieved enhanced sparsity for similar patches in the predesigned 3D dictionaries like discrete cosine transform (DCT) or discrete Haar wavelets transform. These dictionaries are separable and can be extended from two-dimensional to three-dimensional form using tensor product. Furthermore, Mairal et al. [26] extended the dictionary learning approach [23] using L_1 norm for learning phase and L_0 pseudonorm for the final construction phase. The major distinction between this approach and [25] is the use of learning phase to construct a dictionary rather than to use a predesigned dictionary. The learned dictionary is developed either from the given noisy image or from a database of patches extracted from large collection of natural images.

Several attempts have been made to relate nonlocal means filtering with the variational and PDE based image denoising techniques. Gilboa and Osher [27] provided an elegant interpretation of the nonlocal means approach as a generalization of variational and PDE-based formulations. Brox et al. [28] proposed a computationally efficient algorithm for nonlocal filtering that arranged the data in a cluster tree. This special arrangement further helps in preselection of similar patches. Also, an iterative version of nonlocal filtering based on the variational framework was suggested in [28]. Pizarro et al. [29] introduced a discrete version of variational formulation that exploits the nonlocal data and smoothness (NDS) constraints for a variety of generalized dissimilarity measures defined on the space of image patches. This formulation results in a new similarity measure that considers not only the patch similarity of the two selected pixels but also the similarity of the respective neighbors. Moreover, the formulation emphasized the connection between diffusion based approaches and NDS formulation. Yang and Jacob [30] proposed a unified variational framework for nonlocal regularization by introducing robust distance norm to determine interpatch distances. Furthermore, by using this formulation for inverse problems, theoretical justification is provided for heuristic iterative nonlocal means approaches. Tschumperlé and Brun [31] defined a patch space for implementing PDE-based diffusion or smoothing process.

In addition to the photometric similarity, which is used for patch-based comparison, the patches contain much more information that requires the attention of researchers. With this motivation, we propose the notion of a normal vectors patch corresponding to each intensity patch. By employing this new notion in the second step of our algorithm, we achieve remarkably better results than most of the state-of-the-art algorithms in the presence of moderate or severe noise. Inspired by the special treatment of the central patch in [32], we also employ a slightly higher weight than in the standard nonlocal means approach. The weight value in our approach is associated with the central patch empirically through experiments. This modification has further improved our results.

Earlier, Mahmoudi and Sapiro [33] have proposed the average gradient orientation difference in combination with nonlocal means filtering which seems to be similar to the second step of our proposed method. However, the approach of [33] significantly differs from the proposed approach due to two main reasons. Firstly, for [33], the average gradient information is employed only for preclassification of the intensity patches to increase the computational efficiency of standard nonlocal filtering without the modification of the classical similarity measure. Whereas, the proposed approach exploits the interactions of normal vectors patches for the modification of the similarity weights without any preclassification step. Secondly, the average gradient orientations of patches are used in [33] on the basis of the assumption that the average gradient direction is expected to be similar in the presence of additive white Gaussian noise. This assumption is somewhat arguable in the presence of severe noise level.

The rest of the paper is organized as follows. A brief review of methodologies used in this paper is provided in Section 2. The new multiscale algorithm and its implementation are explained in Section 3. Experimental results are described and discussed in Section 4, and conclusions are drawn in Section 5.

2. Preliminaries

2.1. Nonlocal Means Filtering.

Consider a noisy gray-scale intensity image

$$u = u_o + v\left(0, \sigma^2 \mathbf{I}\right), \tag{1}$$

where u_0 is the true noise free image to be recovered, v represents the zero mean additive white Gaussian noise of known variance σ^2 and $\mathbf{I}$ is the Identity matrix with the same dimensions as those of the given image. Instead of the nonlocal filtering, it is preferable to use the notion of semilocal filtering. It was mentioned in [32] that considering the whole image to search for similar patches has no major benefit, with the exception of periodic or quasiperiodic images. Also, searching the whole image for each pixel is computationally too expensive. Therefore, in the rest of this paper, the term nonlocal means refers specifically to semilocal filtering. The nonlocal means filtering is defined [19] as follows:

$$u_{\text{nlm}}(i) = \frac{1}{K(i)} \sum_{j \in \Delta_i} w(i, j) u(j), \tag{2}$$

where $u_{\text{nlm}}(i)$ denotes the denoised value at pixel location i and $w(i, j)$ is the weight obtained by determining similarity of noisy intensity patches $\mathbf{P}(i)$ and $\mathbf{P}(j)$ around the central pixel i and its neighboring pixel j, respectively, within the search window Δ_i. The normalization factor is given by $K(i) = \sum_{j \in \Delta_i} w(i, j)$.

The weights, $w(i, j)$, are obtained using the Gaussian weighted L_2 norm [19]:

$$w(i, j)_{i \neq j} = \exp\left(-\frac{\|\mathbf{P}(i) - \mathbf{P}(j)\|_{2,\alpha}^2}{h^2}\right), \tag{3}$$

where α is the standard deviation of Gaussian function to consider the spatial proximity of the central patch and its neighboring patches in the search window. The patches, in the search window, spatially closer to the central patch are assigned higher weights than those for the distant patches. The nonlocal means approach involves three parameters: the size of the patch $\mathbf{P}(i)$, the size of the search window Δ_i, and the filtering or smoothing parameter h. Detailed discussions on the suitable choices of these parameters can be found in [32, 34].

Note that the similarity measure, given in (3), trivially assigns the largest value of 1 to the patch when compared to itself, so that the central patch has the weight 1 as $\|\mathbf{P}(i) - \mathbf{P}(j)\|_{2,\alpha}^2 = 0$ for $i = j$. However, the weight value assigned to the central patch may not be reliable for the noisy image data. To avoid this unreliable self-similarity weight, Buades et al. [19] offered the maximum value of all the weights computed for $j \neq i$ around the central patch for the center self-similarity weight. The standard similarity measure can then be given by

$$w(i,j) = \begin{cases} \exp\left(-\dfrac{\|\mathbf{P}(i) - \mathbf{P}(j)\|_{2,\alpha}^2}{h^2}\right) & j \neq i \\ \max_{k \in \Delta_i; k \neq i} \{w(i,k)\} & j = i. \end{cases} \quad (4)$$

2.2. Accelerated Nonlocal Means Filtering.
To increase the computational efficiency of nonlocal means filtering, several approaches [28, 33, 34, 36] have been introduced. Buades et al. [34] suggested an accelerated version (block-based) of nonlocal means algorithm that, instead of using the central pixel, replaces the whole patch around the central pixel with a weighted average of patches around the neighboring pixels in the search window. Mathematically this filtering process is defined as [34]

$$\mathbf{P}_{\text{nlm}}(i) = \frac{1}{K(i)} \sum_{j \in \Delta_i} w(i,j) \mathbf{P}(j), \quad (5)$$

where the similarity measure, $w(i,j)$, is the same as defined in (3). Finally, the pixel value at central location i is recovered by averaging all the resulting estimators (patches) containing that pixel location. However, the cost for the computational improvement of the accelerated nonlocal means is a slight degradation in the visual quality of the denoised image.

Mahmoudi and Sapiro [33] proposed a fast implementation of nonlocal means filtering through preclassification step to remove unrelated patches prior to the computation of similarity weights. This preclassification step is implemented on the basis of either the comparison of the local average gray values computed from intensity patches or the comparison of average gradients computed from the gradient patches. In either case, due to the presence of noise, this preclassification step is followed by thresholding process. Finally, the weights are computed only with the preclassified patches. We employ the block-based nonlocal means filtering [34] in the first step of our algorithm exploiting its relative simplicity as compared to other fast implementations.

2.3. Adaptive Similarity Measure Approaches.
In this section, we briefly discuss the recent approaches proposing effective modifications to the similarity measure used in the classical nonlocal means filtering. For further details of these modifications, we refer to [37, 38]. The weighted patch-wise photometric distance between patches $\mathbf{P}(i)$ and $\mathbf{P}(j)$ can be expressed in vectorial notation as

$$\|\mathbf{P}(i) - \mathbf{P}(j)\|^2 = (\mathbf{P}(i) - \mathbf{P}(j))^t \mathbf{G}(\mathbf{P}(i) - \mathbf{P}(j)), \quad (6)$$

where $\mathbf{G}$ denotes a fixed diagonal matrix which contains Gaussian weights. Despite the fact that the weighted L_2 norm represented by weight matrix $\mathbf{G}$ performs quite well, nonlocal means filtering results in oversmoothing in a certain area of a given image due to fixed weight matrix. One of the most effective approaches to incorporate the locally adaptive similarity weights is based on finding the local covariance matrix. The locally adaptive photometric distance is therefore defined as

$$\|\mathbf{P}(i) - \mathbf{P}(j)\|^2 = (\mathbf{P}(i) - \mathbf{P}(j))^t \mathbf{C}_{\mathbf{P}_j}^{-1} (\mathbf{P}(i) - \mathbf{P}(j)), \quad (7)$$

where the $\mathbf{C}_{\mathbf{P}_j}$ is the local covariance matrix for the patch centered at pixel j. However, it is not possible to obtain reliable covariance matrix from a single observation that contains noise as well. In order to overcome this problem, a certain redundancy in the given image is used as an implicit prior which is generally true in case of natural images. Exploiting this assumption, the noisy patches in the observed image similar to the reference patch are sampled and grouped to obtain an estimate of covariance matrix. Several methods [35, 39, 40] have been proposed to obtain robust and reliable estimates of locally adaptive covariance matrix which are briefly described below.

Dabov et al. [39] proposed further enhancement for the original BM3D method [25] by introducing shape adaptive transform basis (BM3D-SAPCA) instead of fixed DCT or wavelet transform basis. This shape adaptivity is achieved by incorporating the following modifications. Firstly, the shape adaptive 3D groups of patches similar to the reference patch are constructed as described in [41]. Secondly, the local covariance matrix is estimated from this group provided that the number of similar patches is large enough to ensure the reliable estimation of the local covariance matrix. Thirdly, PCA transform is obtained by the eigen-decomposition of the estimated covariance matrix and only those eigenvectors are retained as principal components whose eigenvalues are greater than a fixed threshold value. Next, the truncated PCA transform is applied on the 3D group of the shape adaptive patches with 1D orthogonal wavelet transform along the third direction of the group. All the remaining process is same as that of the original BM3D [25].

Zhang et al. [40] proposed LGP-PCA approach where PCA is performed on the local covariance matrix estimated by local pixel grouping (LPG) of the patches. Afterwards, the linear minimum mean square-error estimation (LMMSE) technique is used for coefficient shrinkage in the PCA transform domain obtained from diagonalization of the covariance matrix. This approach is similar to BM3D-SAPCA [39]

in the sense that it involves two steps denoising process. However, in contrast to the BM3D, it does not depend on basic estimate (oracle) obtained in first iteration. Recently, Lebrun et al. [35] has proposed NL-Bayes algorithm to merge the Fourier transform domain like methods with Bayesian framework. In this case, a reliable estimate of the covariance matrix is obtained by applying Bayes' rule for 3D groups of noisy patches followed by MAP (maximum *a posteriori*) estimation. The LPG-PCA method [40] can be realized as a special case of this more general approach.

As will be described in the next section, the proposed modification to the classical similarity measure significantly differs from the above mentioned approaches. It assigns the adaptive similarity weights based on the correlation of normal vectors patches in addition to the similarity of intensity patches. Also, unlike the above techniques, the proposed approach does not apply 2D or 3D grouping of the similar patches.

2.4. Discrete Stationary Wavelet Transform. Wavelets have numerous applications, in particular, in signal and image processing, thanks to the efficient time-frequency localization and multiresolution properties. The wavelet transform is generated by the convolution integral with a wavelet ψ which defines a Riesz basis for $L_2(\mathbb{R})$ with a two parameter family $\{\psi_{j,l}\}_{j,l\in\mathbb{Z}}$, where

$$\left\{\psi_{j,l}(x) = 2^{j/2}\psi\left(2^j x - l\right)\right\}_{j,l\in\mathbb{Z}}. \tag{8}$$

The wavelet decomposition of $f \in L_2(\mathbb{R})$ is given by

$$f(x) = \sum_{j,l\in\mathbb{Z}} d_{j,l}\psi_{j,l}(x) \quad \text{with } d_{j,l} = \langle f, \widetilde{\psi}_{j,l}\rangle, \tag{9}$$

where $\widetilde{\psi}_{j,l}$ is the dual wavelet to $\psi_{j,l}$. For the orthogonal wavelet case, $\widetilde{\psi}_{j,l} = \psi_{j,l}$. The wavelet decomposition provides the perfect reconstruction and multiresolution analysis of f, as well. In the practical situation, f is a signal with finite resolution scale. In this case, the wavelet decomposition has the following representation:

$$f(x) = \sum_l c_{J_0,l}\phi_{J_0,l} + \sum_{j=J_0}^{J}\sum_l d_{j,l}\psi_{j,l}(x) \tag{10}$$

$$\text{with } c_{J_0,l} = \langle f, \widetilde{\phi}_{J_0,l}\rangle,$$

where J is the finest resolution scale determined by the size of the signal f and $J_0 < J$ is a chosen coarsest resolution scale. Also, ϕ is the scaling function associated with the wavelet ψ. The standard discrete wavelet transform (DWT) decomposes a signal f into low and high pass subbands. At a given resolution scale j, $(c_{j,l})$ is the approximate (low-passed) data and $(d_{j,l})$ is the detail (high-passed) data of f. For the properties of the wavelet transform, we refer to [42].

The one-dimensional wavelet can be applied through separable extension to obtain a two-dimensional wavelet. One easy way to construct two-dimensional wavelet is to use the tensor product of one-dimensional wavelet. In this case, there are one scaling function ϕ and three kinds of wavelet, ψ^k ($k = 1, 2, 3$): ψ^1, ψ^2, and ψ^3 are horizontally, vertically, and diagonally oriented wavelets, respectively. For a given image data u, the wavelet transform provides the subband data through the wavelets and scaling function at the resolution scale j by

$$W_j^k(u) = \left(\langle u, \widetilde{\psi}_{j,l}^k\rangle\right), \quad k = 1, 2, 3, \tag{11}$$

$$S_j(u) = \left(\langle u, \widetilde{\phi}_{j,l}\rangle\right). \tag{12}$$

So, $S_j(u)$ represents the low-passed data and $W_j^k(u)$ are high-passed detail data: $W_j^1(u)$, $W_j^2(u)$, and $W_j^3(u)$ are horizontal detail, vertical detail, and diagonal detail subband data, respectively. Note that the standard DWT generates the downsampled data due to low and high pass filtering followed by the decimation in the nature of DWT. Also, even if DWT has the perfect reconstruction property, it is not shift invariant as the wavelet coefficients of the given data and its shifted version may not be same [43]. So, the transformed data loses some detail information.

To overcome this kind of disadvantage, the discrete stationary wavelet transform (DSWT) was proposed [42, 43]. One of the main ideas of DSWT is to perform low and high pass filtering without decimation. In this approach, multiresolution analysis is achieved by simply inserting zeroes between every adjacent pair of the coefficients in low and high pass filters associated with the wavelet. The insertion of zeroes makes the dimensions of filters at the resolution scale j same as those at the resolution scale $j+1$. Subsequently, the DSWT provides a redundant representation of a data so that the resulting wavelet decomposition is overcomplete. We notice the following advantages of the DSWT in the viewpoint of image denoising.

   (i) The redundant representation of data contains more information than the case of nonredundant one.

  (ii) The shift invariance representation has smooth regularity and also does not contain the error due to downsampling.

 (iii) The size of the subband data is same as that of the input image data, so that the spatial information of those data can be directly compared.

In the proposed method, we implicitly exploit these advantages to obtain the similarity measure of normal vectors patches. For our notational convenience, we will use the same notation for DSWT as the case of DWT. That is, $S_j(u)$, $W_j^k(u)$, and $k = 1, 2, 3$ are the subband data at the resolution scale j generated by the DSWT.

3. The Proposed Algorithm

In this section, we adopt a hybrid approach to nonlocal means filtering. Our scheme is composed of two steps, as shown in Figure 1. In the first step, we apply the accelerated (block-based) version in the wavelet transform domain to obtain a predenoised image as shown in Figure 2. In the second
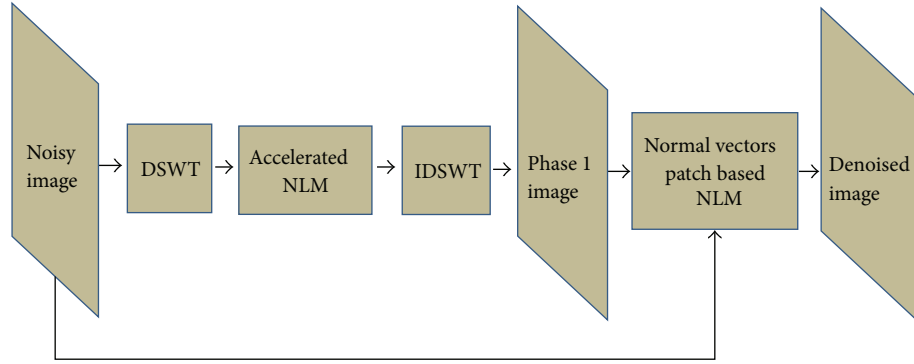
FIGURE 1: Schematic depiction of the proposed nonlocal means filtering algorithm. Phase 1 image is obtained in the first step. Afterwards, in the second step, phase 1 image is used in combination with the given noisy image to obtain a final denoised image.
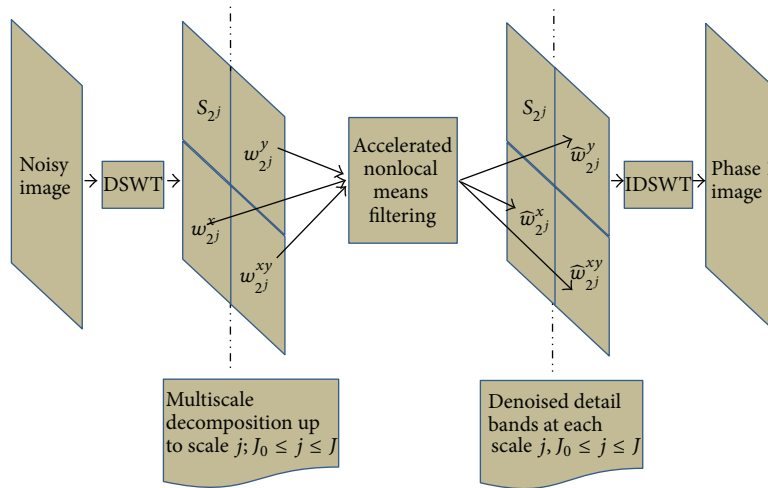

FIGURE 2: Schematic depiction of the first step of the proposed algorithm at the scale j with $J_0 \leq j \leq J$.

step, we employ a modified version of conventional nonlocal means filtering on the given noisy image. In this step, the weights are computed using the predenoised image obtained in the first step.

3.1. Multiscale Accelerated Nonlocal Means Filtering. Inspired by various approaches [22, 23, 25], based on using predenoised images to obtain refined patches, we seek to obtain a predenoised image in our approach. However, in contrast to those techniques, we adopt a very simple approach as shown in Figure 2. We first decompose the image u using the two-dimensional discrete stationary wavelet transform [43] with a chosen coarsest level J_0,

$$\text{DSWT}(u) = \left(\left(\left(W_j^1 \right), \left(W_j^2 \right), \left(W_j^3 \right) \right)_{J_0 \leq j \leq J}, \left(S_{J_0} \right) \right) u, \quad (13)$$

where $W_j^1(u)$, $W_j^2(u)$, and $W_j^3(u)$ represent the horizontal, vertical, and diagonal detailed data, respectively, obtained with (11) at the scale j with $j = J_0, \ldots, J$. Also, $S_{J_0}(u)$ is the approximate data given by (12) at the coarsest resolution scale J_0.

The accelerated (block-based) version of nonlocal means filtering [34] is performed on each of the detail data $W_j^1(u)$, $W_j^2(u)$, and $W_j^3(u)$ at each scale j. To be precise, the estimated wavelet coefficients patch (or block) $\widehat{\mathbf{w}}_j^k u(t)$ at location t from the detailed data $W_j^k(u)$ is obtained by applying (5) as

$$\widehat{\mathbf{w}}_j^k u(t) = \frac{1}{K(r)} \sum_{s \in \Delta_t} w(t,s) \, \mathbf{w}_j^k u(s), \quad (14)$$

where the weights $w(t,s)$ are computed using wavelet coefficients patches $\mathbf{w}_j^k u(s)$ instead of using intensity patches $\mathbf{P}(s)$ in (4). Subsequently, the denoised wavelet coefficient $\widehat{w}_j^k u(t)$ at central location t is recovered by averaging all the resulting estimated blocks containing that location. This process is implemented on each detailed subband data independently and finally the predenoised image is obtained using the inverse discrete stationary wavelet transform. Thanks to overcomplete, shift invariant, and sparse representation of an image in the DSWT domain, the image can be efficiently denoised using accelerated nonlocal means filtering without

severe degradation of texture, edges, and fine details. This predenoised image will further be used as a reference image along with the given noisy image for computation and comparison of normal vectors patches in the subsequent section. It is worth noticing that several recent methods [44, 45] have implemented nonlocal means filtering in wavelet domain for super-resolution of images and video sequences.

3.2. Normal Vectors Patches and Weight Factors. Assuming the patches as surfaces, to each intensity patch of size $m \times m$ at pixel location i, we associate a normal vectors patch $\widetilde{\mathbf{N}}_i$ of size $2m \times 2m$. In order to construct this normal vectors patch, we first compute the gradient of the image at each pixel location and then form the patches of horizontal and vertical components of the gradient vectors around each pixel i. These partial derivative patches are denoted by $\mathbf{N}_{i,x}$ and $\mathbf{N}_{i,y}$, where the subscripts x and y represent partial derivatives in the horizontal and vertical directions, respectively. The size of $\mathbf{N}_{i,x}$ and $\mathbf{N}_{i,y}$ patches is the same as that of the intensity patch.

Definition 1. The normal vectors patch $\widetilde{\mathbf{N}}$ is defined as

$$\widetilde{\mathbf{N}} \equiv \left(\mathbf{N}_x\right)_{m \times m} \oplus \left(\mathbf{N}_y\right)_{m \times m} = \left(\begin{array}{c|c} \mathbf{N}_x & \mathbf{O} \\ \hline \mathbf{O} & \mathbf{N}_y \end{array}\right)_{2m \times 2m}, \quad (15)$$

where $\oplus$ denotes the direct sum of matrices and $\mathbf{O}$ is the zero matrix of size $m \times m$.

For a similarity measure between two normal vectors patches $\widetilde{\mathbf{N}}_i$ and $\widetilde{\mathbf{N}}_j$, we employ the inner product for square real matrices induced by the trace as follows:

$$\left\langle \widetilde{\mathbf{N}}_i, \widetilde{\mathbf{N}}_j \right\rangle = \mathrm{tr}\left(\Gamma_{ij}\right), \quad (16)$$

where we set the matrix Γ_{ij} as

$$\begin{aligned} \Gamma_{ij} &= \widetilde{\mathbf{N}}_i^t \widetilde{\mathbf{N}}_j = \left(\mathbf{N}_{i,x} \oplus \mathbf{N}_{i,y}\right)^t \left(\mathbf{N}_{j,x} \oplus \mathbf{N}_{j,y}\right) \\ &= \left(\mathbf{N}_{i,x}^t \mathbf{N}_{j,x} \oplus \mathbf{N}_{i,y}^t \mathbf{N}_{j,y}\right). \end{aligned} \quad (17)$$

Recall that for square real matrices A and B with same size, $\langle A, B \rangle = \mathrm{tr}(A^t B)$ is a well-defined inner product. The inner product defined above signifies the geometric correlation between the central normal vectors patch $\widetilde{\mathbf{N}}_i$ at central location i and the neighboring patch $\widetilde{\mathbf{N}}_j$ at the position j in the search window. The higher the absolute value of inner product the more the similarity between the normal vector patches and vice versa. In order to obtain the individual normal vector in the neighboring patch that may have the highest correlation with that of the central patch, we define the diagonal matrix as follows:

$$\Lambda_{ij} = \mathrm{diag}\left(\mathbf{N}_{i,x}^t \mathbf{N}_{j,x} + \mathbf{N}_{i,y}^t \mathbf{N}_{j,y}\right). \quad (18)$$

The maximum individual correlation, denoted by τ_{ij}, is defined as

$$\tau_{ij} = \left\|\Lambda_{ij}\right\|_\infty = \max_{1 \le p \le m} \sum_{q=1}^{n} \left|\lambda_{pq}\right|, \quad (19)$$

where $\|A\|_\infty$ norm denotes the maximum of absolute row sum of a matrix A. Since the matrix Λ_{ij} is the diagonal matrix, this norm yields the maximum absolute value of all diagonal entries of Λ_{ij}. In order to obtain the modified similarity measure between normal vectors patches in the next section, we define a weight factor as follows.

Definition 2. A weight factor $\eta(i, j)$ based on two normal vectors patches $\widetilde{\mathbf{N}}_i$ and $\widetilde{\mathbf{N}}_j$ is defined as

$$\begin{aligned} \eta(i, j)_{i \ne j} &= \exp\left(-\frac{\left\langle \widetilde{\mathbf{N}}_i, \widetilde{\mathbf{N}}_j \right\rangle}{\tau_{ij}\left(m^2 - 1\right)}\right) \\ &= \exp\left(-\frac{\mathrm{tr}\left(\Gamma_{ij}\right)}{\tau_{ij}\left(m^2 - 1\right)}\right). \end{aligned} \quad (20)$$

Remark 3. It can be easily verified that the weight factor η satisfies the following properties to measure the similarity between normal vectors patches.

(i) $\eta(i, j) > 0 \; \forall j \in \Delta_i$ (Positivity).

(ii) $\eta(i, j) = \eta(j, i) \; \forall j \in \Delta_i$ (Symmetry).

(iii) $\eta(i, j) \le M < \infty \; \forall j \in \Delta_i$ (Bounded).

It can be noticed that in case of more similar normal vectors patches, $\mathrm{tr}(\Gamma_{ij})/\tau_{ij}(m^2 - 1)$ is a large value resulting in the small value of $\eta(i, j)$ and vice versa. The intuitive motivation for introducing this factor is to acquire the degree of similarity of the original intensity patches, based on the similarity of the corresponding normal vectors patches, as shown in Figure 3. Note that the weight factor involves the computation of derivatives which is sensitive to the noise level in the image. Since in the predenoised image obtained in the first step, the noise level is much lower than in the given noisy image, the computation of partial derivatives is more reliable. Therefore, we use the weight factor in the second step of our algorithm.

3.3. Modified Nonlocal Means Filtering. After obtaining predenoised image as defined in Section 3.1, we perform the modified version of standard nonlocal means filtering on a given noisy image, based on the similarity of reference patches from the predenoised image.

Definition 4. The modified similarity measure between distinct intensity patches $\mathbf{P}(i)$ and $\mathbf{P}(j)$ is defined as

$$w(i, j)_{i \ne j} = \exp\left(-\frac{\left\|\mathbf{P}(i) - \mathbf{P}(j)\right\|_2^2 \eta(i, j)}{h^2}\right), \quad (21)$$

where the weight factor $\eta(i, j)$ is given in Definition 2.

It can be noticed that the Gaussian weight G_α in (3) is replaced by the weight factor $\eta(i, j)$ obtained with (20). The Gaussian weights were used to take into account the spatial

(a) Intensity based patches
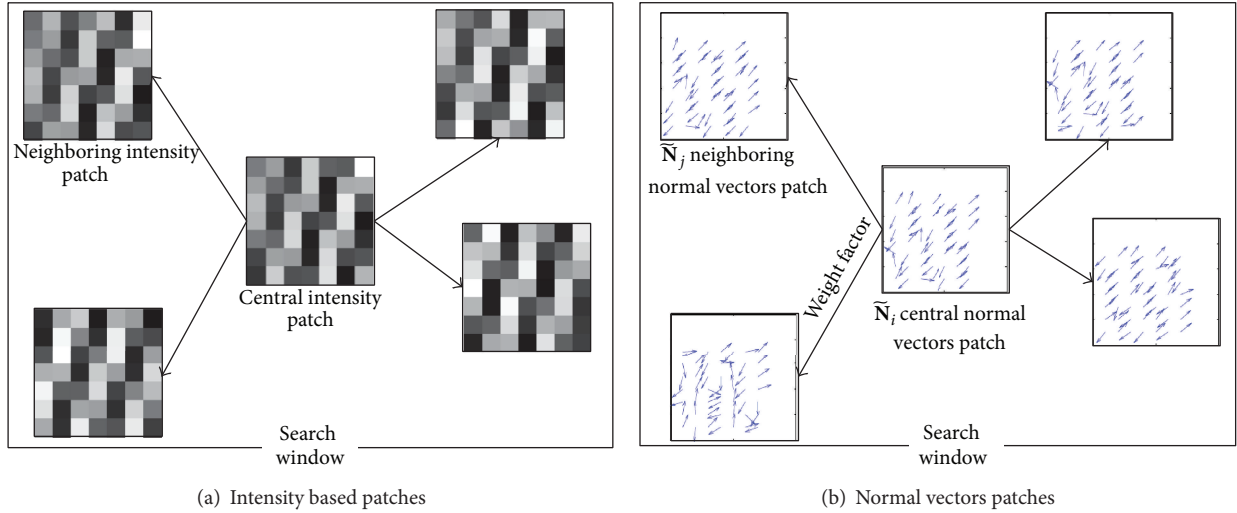

(b) Normal vectors patches

FIGURE 3: Schematic depiction for normal vectors patches, corresponding to the respective intensity patches, extracted from Barbara image.

proximity of the patches around the central patch. However, the proposed weight factor is based upon the similarity or correlation of the mean normal orientation of the central patch with other patches in the search window. The smaller the value of the weight factor $\eta(i, j)$, the higher the similarity weight $w(i, j)$ assigned to the corresponding intensity patch and vice versa. The proposed metric performs better than the standard L_2 norm. This is due to the fact that the standard metric, using spatial proximity weights, G_α, may assign a small weight value to a distant but similar intensity patch in the search window. On the other hand, the new metric assigns more weight to the distant but similar intensity patch based on the similarity of normal vectors patches.

Our second modification to standard nonlocal means filtering is made in the assignment of self-similarity weight. The standard nonlocal means method assigns the maximum of all the computed weights as obtained in (4) for $j = i$. This choice of weight for self-comparison is intuitive and arbitrary. Salmon [32] discussed various choices for assigning self-similarity weight. One of the choices assumed was based on Stein unbiased risk estimation (SURE) for the central patch weight that does not affect the remaining similarity weights. It has been shown in [32] that SURE choice of self-similarity weight is at par with the arbitrary choice [19] for higher levels of noise. However, we follow the approach adopted in [19] with a little modification based on experimental results for various arbitrary values. For low or medium noise levels in the image, we assign a slightly higher self-similarity weight than that in (4), that is, the proposed self-similarity weight is defined by

$$w(i, i) = \frac{4}{3} \max_{j \in \Delta_i; j \neq i} \{w(i, j)\}. \tag{22}$$

The intuitive justification for assigning higher self-similarity weight is based on the assumption that the low or medium levels of noise do not produce a significance effect on self-similarity. Therefore this choice yields slightly better results

for low levels of noise. However, for severe noise ($\sigma > 20$), the performance of the self-similarity weights defined in (4) and (22) is found to be the same. We use (22) in our simulations for both the low and high levels of noise. The second modification of self-similarity measure is then optional. The distinction in visual appearance with or without this modification is imperceivable.

3.4. Summary of the Proposed Algorithm.
The flow chart of the proposed denoising algorithm is shown in Figure 4 and it can be summarized as follows.

(i) Construct multiscale data in the wavelet domain by decomposing the noisy image using discrete stationary wavelet transform up to the coarsest scale J_0.

(ii) Perform the accelerated nonlocal means filtering on the multiscale data, as described in Section 3.1, to obtain the denoised wavelet transform components.

(iii) Reconstruct the predenoised image using inverse of the DSWT. This image will be used as a reference in the next step for patch-based comparison.

(iv) Denoise the given noisy image using the modified nonlocal means filtering as proposed in Section 3.3. The patch comparison is performed on the refined patches obtained in the previous step.

3.5. Extension to Color Image Denoising.
Although the proposed algorithm is primarily designed for gray-scale images, a possible extension to color images is also concerned here. For this purpose, we consider a natural color image in *RGB* color domain with additive i.i.d. zero mean Gaussian noise and variances $\sigma_R^2 = \sigma_G^2 = \sigma_B^2 = \sigma^2$ in each of *RGB* channels. Nonlocal means filtering can be extended in a straightforward fashion to joint or simultaneous filtering of all color channels without disturbing the inherent correlations of color channels. However, due to 2D DSWT transform in the first step of
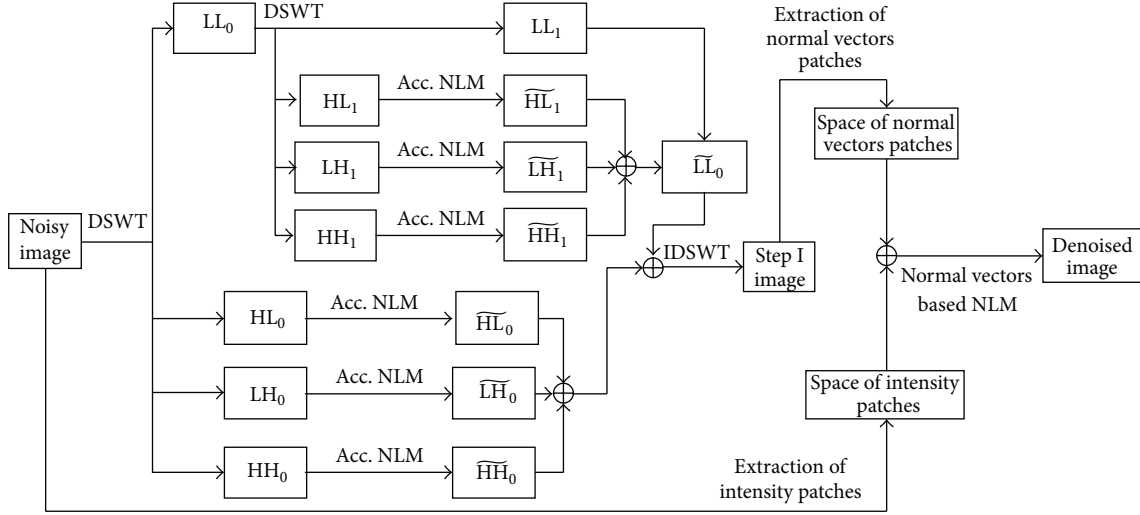
FIGURE 4: The flow chart of the proposed MHNLM algorithm. The given image is decomposed up to two scale levels using DSWT and accelerated nonlocal means [34] is applied on the detail data (HL, LH, and HH) at each scale. Afterwards, the predenoised image is obtained using IDSWT. Finally, the modified similarity measure based nonlocal means filtering is employed on the given noisy image using reference normal vector patches extracted from the predenoised image.

the proposed algorithm, we follow the approach as described in [25, 35, 41]. For the first step only, The given noisy image is transformed to *YUV* color domain using the transformation defined by

$$A_{YUV} = \begin{pmatrix} 0.30 & 0.59 & 0.11 \\ -0.15 & -0.29 & 0.44 \\ 0.61 & -0.51 & -0.10 \end{pmatrix}, \tag{23}$$

where *Y* denotes the luminance channel and *U* and *V* are two chrominance channels. The resulting channels are sufficiently decorrelated. Further, due to this transformation, the luminance channel has higher SNR than the other two chrominance channels. The noise variances σ_Y^2, σ_U^2, and σ_V^2 can be determined using the relation [41]:

$$\begin{bmatrix} \sigma_Y^2 & \sigma_U^2 & \sigma_V^2 \end{bmatrix} = \begin{bmatrix} \sigma_R^2 & \sigma_G^2 & \sigma_B^2 \end{bmatrix} T, \tag{24}$$

where *T* denotes the transpose of A_{YUV} with element wise square of its entries.

The first step of the algorithm, described in Section 3.1, is applied to each channel independently. Consequently, a predenoised image in *YUV* color space is transformed back to original *RGB* color space. For the second step of the proposed algorithm, let the normal vectors patch, $\widetilde{\mathbf{N}}_i$, centered at pixel *i* in *RGB* color space be denoted by

$$\widetilde{\mathbf{N}}_i := \left\{ \widetilde{\mathbf{N}}_i^{(c)} \right\}_{c=1}^3, \tag{25}$$

where the superscript $c = 1, 2, 3$ denotes the individual color channel *R*, *G*, *B*, respectively, and $\widetilde{\mathbf{N}}_i^{(c)}$ represents the normal vectors patch in that color channel defined by (15).

The similarity of normal vectors patches, denoted by $d_{ij}^{(N)}$, is performed jointly on all the color channels as

$$d_{ij}^{(N)} = \frac{1}{3(m^2 - 1)} \sum_{c=1}^{3} \frac{\left\langle \widetilde{\mathbf{N}}_i^{(c)}, \widetilde{\mathbf{N}}_j^{(c)} \right\rangle}{\tau_{ij}^{(c)}}$$

$$= \frac{1}{3(m^2 - 1)} \sum_{c=1}^{3} \frac{\mathrm{tr}\left(\Gamma_{ij}^{(c)}\right)}{\tau_{ij}^{(c)}}, \tag{26}$$

where $\Gamma_{ij}^{(c)}$ and $\tau_{ij}^{(c)}$ are defined for each channel using the relations (17) and (19), respectively. The weight factor, η_{ij}, used in the proposed similarity measure can then be obtained with

$$\eta_{ij} = \exp\left(-d_{ij}^{(N)}\right). \tag{27}$$

Finally, the modified photometric similarity measure can be defined as

$$w(i, j) = \exp\left(-d_{ij}^2 \eta_{ij}\right), \tag{28}$$

where d_{ij} denotes the classical nonlocal means distance for *RGB* color patches $\mathbf{P}(i)$ and $\mathbf{P}(j)$ given by

$$d_{ij}^2 = \frac{1}{3} \sum_{c=1}^{3} \frac{\left\| \mathbf{P}^{(c)}(i) - \mathbf{P}^{(c)}(j) \right\|^2}{\left(h^{(c)}\right)^2}. \tag{29}$$

4. Results and Discussion

4.1. Gray-Scale Image Denoising. To evaluate the results of the proposed method, the benchmark gray-scale intensity images of Lena, Barbara, Boats, Peppers, and House are

considered. We use 9×9 patches, 15×15 search windows, and $h = 6\sigma$ as the filtering parameter in the first step. In the second step, we select 7×7 patches with filtering parameter $h = 4\sigma$. The filtering parameter is empirically reduced to 4σ from 6σ to avoid the oversmoothing phenomenon. The finest resolution scale for 512×512 gray-scale image is $J = 9$. The proposed algorithm, denoted by MHNLM, is compared with various state-of-the-art algorithms: those by Dabov et al. [25] (BM3D), Lebrun et al. [35] (NL-Bayes), Portilla et al. [18] (BLS-GSM), Buades et al. [19] (NLM), Elad and Aharon [23] (K-SVD), and Kervrann et al. [22] (BNLM) (Table 3). The denoising results for NL-Bayes and the standard NLM methods are obtained using source codes [46, 47], respectively.

4.2. Computational Complexity Comparison. In what follows, we present the comparison of computational complexity in terms of the number of operations required. In our proposed algorithm, the computational time is proportional to $N^2 n^2 m^2 (1 + 3(J - J_0)/m^2)$, where N^2, n^2, and m^2 are the sizes of image, search window, and patch, respectively. The first term in this expression refers to the number of operations required in classical nonlocal means filtering. The second term refers to complexity of the accelerated nonlocal mean filtering on three detail subbands at each scale. The coarsest level of multiscale decomposition in DSWT domain is $J_0 = 7$. In our simulations, we employ biorthogonal spline wavelet transform for DSWT decomposition. The time complexity of the nonlocal means filtering is $N^4 n^2$ where the search window is set as the whole image. However, for semilocal version of the standard NLM [19], compared here, the time complexity is given by $N^2 n^2 m^2$. In case of BNLM [22], the computational time depends upon the sub-sampling size L of the image and is given by $N^2 n^2 m^2 / L^2$. In the absence of sub-sampling scheme ($L = 1$), the computational complexity of BNLM is the same as that of NLM. For BM3D [25], the time complexity, excluding the 3D transformation overhead, is given by $2(n_1^2 + n_2)m^2 N^2 / L^2$, where the sum, $n_1^2 + n_2$ represents the number of operations required for grouping the patches in 3D transform domain. The multiplication factor, 2, is used to indicate the two step image denoising nature of BM3D algorithm. The sub-sampling parameter, L is the sliding window size used for block matching and is set to 1 when the grouped patches do not overlap. The computational complexity of NL-Bayes [35] is similar to that of BM3D. This is due the fact that NL-Bayes algorithm employs Bayes' rule to the the 3D groups obtained with BM3D approach. Finally, the computational cost of K-SVD [23] is given by $(4Km^2L + 2K(L + m^2) + K^3)N^2$ [48], where K denotes the desired sparsity level and L represents the number of atoms (columns) in the dictionary. The remaining terms, m^2 and N^2 have the usual meanings of patch and image size, respectively. Among all the methods compared here, K-SVD algorithm is the most expensive due to the sequential update of each of the L atoms (columns) in the dictionary. From the above analysis, it can be noticed that the computational complexity of the proposed algorithm,

somehow, lies between those of the classical nonlocal means filtering and BM3D.

4.3. Visual Quality Comparison. Here, we discuss and compare the denoising capability of the proposed algorithm with that of the above mentioned algorithms. The peak signal-to-noise ratio (PSNR) is used for comparing the denoising capability of each scheme. For an original gray-scale image u_o of size $M \times N$, let $\hat{u}$ be the denoised image obtained as its estimate using certain denoising algorithm. The PSNR is then defined as

$$
\begin{aligned}
&\text{PNSR}(\hat{u}) \\
&= 10 \log \left(\frac{(255)^2}{(MN)^{-1} \sum_{i=1}^{M} \sum_{j=1}^{N} \left[u_o(i,j) - \hat{u}(i,j) \right]^2} \right).
\end{aligned}
\tag{30}
$$

Table 1 provides the PSNR comparison values for the above cited approaches; the bold-faced values represent the best performance among the last five columns. The results of BM3D and NL-Bayes are quite competitive regardless of the noise level and are better than the rest of the compared algorithms. The reason for these outstanding results is the enhanced sparse representation obtained with 3D transform domain and collaborative filtering as discussed earlier. NL-Bayes is similar to BM3D in its structure. However, it exploits Bayesian framework rather than hard thresholding or Wiener filtering used in BM3D. For the rest of the algorithms, KSVD performance is better when the noise level is low. On the other hand, the proposed method yields better results in the presence of moderate and severe noise ($\sigma \geq 20$). It can be noticed that the results of the proposed algorithm are always better than the classical nonlocal means filtering. Our results can also be confirmed according to the average of the performances at various noise levels. Table 2 indicates that, on average, our algorithm outperforms KSVD, BLS-GSM, and BNLM in the presence of moderate or severe noise. However, the performance of the proposed method is slightly lower than those of the KSVD, BLS-GSM, and BNLM for low level of noise ($\sigma = 10$) as depicted in Figure 8. Therefore, the proposed algorithm is more suitable in the scenarios where the noise levels are high.

Apart from the quantitative analysis with the PSNR values, the quality of visual appearance for the test images can also be assessed in terms of fine details and textures. For instance, in case of Lena's face shown in Figure 5(d) weak fine marks on the chin are preserved at the moderate noise level of $\sigma = 20$. Barbara image is a well-known example that contains a lot of texture. As shown at the magnified scale in Figures 6(d) and 7(d), our algorithm efficiently retains the texture on the table cover, scarf, and chair. Similarly, the noise free image of house, shown in Figure 7(e) contains very fine regular pattern of bricks that is quite intermingled with the noise. The proposed algorithm has managed to recover the brick structure while removing the noise effectively. Lastly, the boats image is the representative of the benchmark images with very fine details, sharp edges, and discontinuities. Although, our algorithm effectively removes

TABLE 1: PSNR comparison of the proposed and various state-of-the-art algorithms for gray-scale images.

Image	σ	BM3D [25]	NL-Bayes [35]	GSM [18]	NLM [19]	KSVD [23]	BNLM [22]	MHNLM
Lena	10	35.93	35.88	**35.61**	34.35	35.47	35.25	35.40
	15	34.27	34.02	33.90	32.10	33.70	33.68	**33.93**
	20	33.05	32.88	32.66	31.56	32.38	32.63	**32.75**
	25	32.05	31.89	31.69	30.44	31.32	31.55	**31.76**
	50	28.86	28.83	**28.61**	27.35	27.79	27.51	28.47
Barbara	10	34.98	34.95	34.03	33.18	**34.42**	33.83	33.80
	15	33.11	32.70	31.86	30.77	**32.37**	32.21	32.35
	20	31.78	31.55	30.32	30.27	30.83	30.88	**31.13**
	25	30.72	30.39	29.13	29.01	29.60	29.77	**30.07**
	50	27.17	26.84	25.48	25.62	25.47	24.91	**26.41**
Boats	10	33.92	33.93	33.58	32.93	**33.64**	33.18	32.94
	15	32.14	31.93	31.70	30.70	**31.73**	31.45	31.52
	20	30.88	30.71	30.38	29.73	30.36	30.16	**30.42**
	25	29.91	29.68	29.37	28.56	29.28	29.11	**29.43**
	50	26.64	26.40	**26.38**	25.25	25.95	25.13	26.22
Peppers	10	34.68	34.77	33.77	33.51	**34.28**	33.87	33.53
	15	32.70	32.53	31.74	31.07	**32.22**	32.06	32.03
	20	31.29	31.31	30.31	30.04	**30.82**	30.75	**30.82**
	25	30.16	30.16	29.21	28.68	29.73	29.77	**29.78**
	50	26.41	26.49	25.90	24.17	26.13	23.84	**26.28**
House	10	36.71	36.18	35.35	34.91	**35.98**	35.67	35.78
	15	34.94	34.33	33.64	32.70	**34.32**	34.23	34.23
	20	33.77	33.33	32.39	32.43	33.20	**33.24**	33.11
	25	32.86	32.48	31.40	31.18	32.15	**32.30**	32.15
	50	29.37	29.04	28.26	27.62	27.95	27.64	**28.38**

TABLE 2: Average PSNR comparison for various noise levels.

σ	BM3D [25]	NL-Bayes [35]	GSM [18]	NLM [19]	KSVD [23]	BNLM [22]	MHNLM
10	35.24	35.14	34.47	33.78	**34.76**	34.36	34.29
15	33.43	33.10	32.57	31.47	**32.87**	32.73	32.81
20	32.15	31.96	31.21	30.81	31.52	31.53	**31.65**
25	31.14	30.92	30.16	29.58	30.42	30.50	**30.63**
50	27.69	27.52	26.93	26.00	26.66	25.81	**27.15**

the noise, very thin ropes and strings have not been fully recovered. Moreover, at the magnified scale, there are very weak oscillations across the sharp discontinuities such as edges of ropes and rods. This oscillatory phenomenon may be inherent to wavelet based filtering which we have used in the first step of our algorithm.

An appropriate approach to ascertain the visual quality is the method noise mechanism suggested by Buades et al. [19]. The method noise is defined as the difference between the given noisy image and its denoised version which depicts the noise removed by a specific denoising operator or algorithm. Buades et al. [19] emphasized that the method noise should be as closer to random additive noise as possible to justify the efficacy of certain denoising algorithm. It provides a way to visualize fine geometrical features or details such as texture and edges which may not be preserved by certain denoising algorithm and may be removed along with the noise. Unfortunately, due to complex nature of several state-of-the-art algorithms compared here, the closed form of mathematical description is difficult to be obtained for the method noise comparison. Therefore the method noise will be estimated using the difference images only. It is worth noticing that the visual comparison of the method noise mechanism relies on the suitable selection of noise level of the additive white Gaussian noise. A reasonable value of standard deviation of the noise is 5 or 10 as very
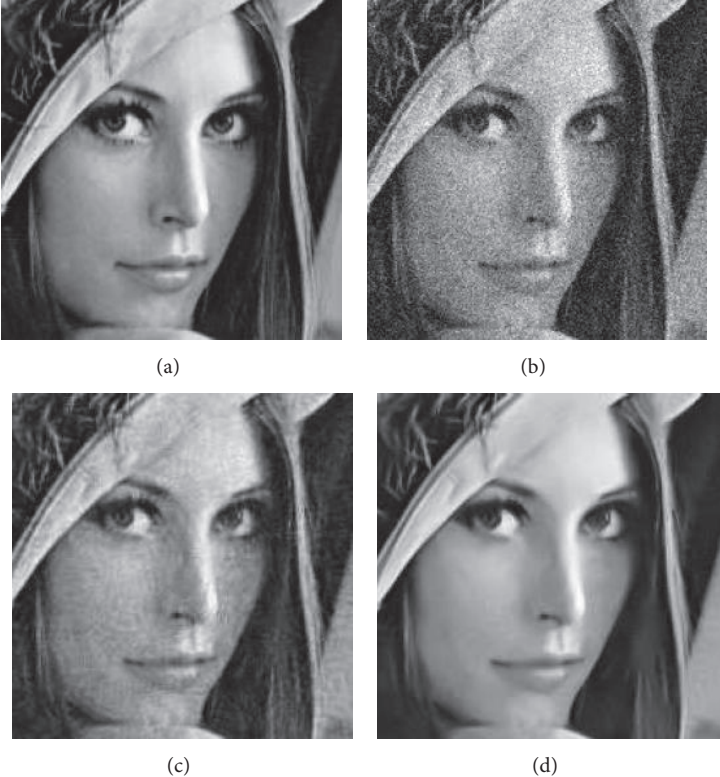
(a)

(b)

(c)

(d)

FIGURE 5: (a), (b), (c), and (d) represent the magnified original, noisy ($\sigma = 20$), predenoised, and denoised subimages from Lena image, respectively.

(a)

(b)

(c)

(d)

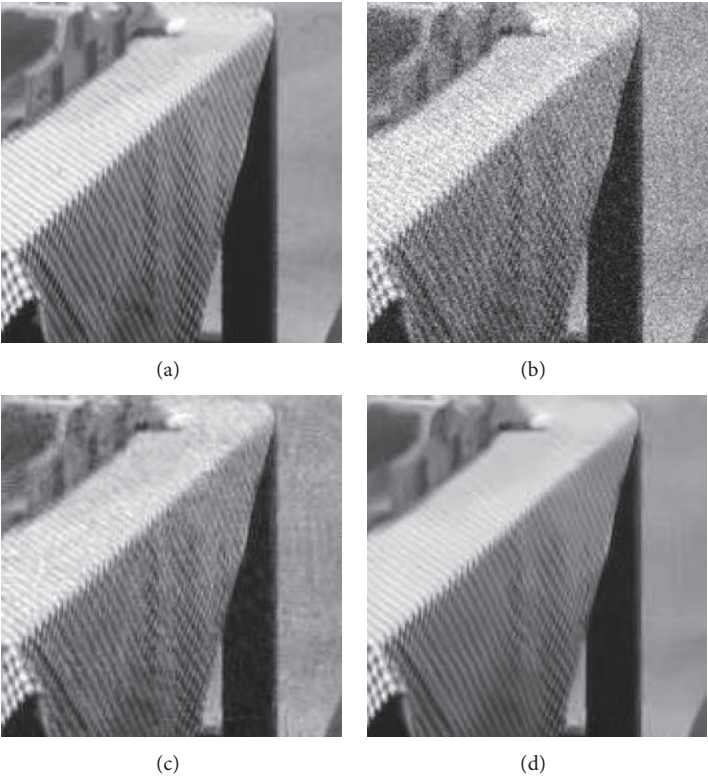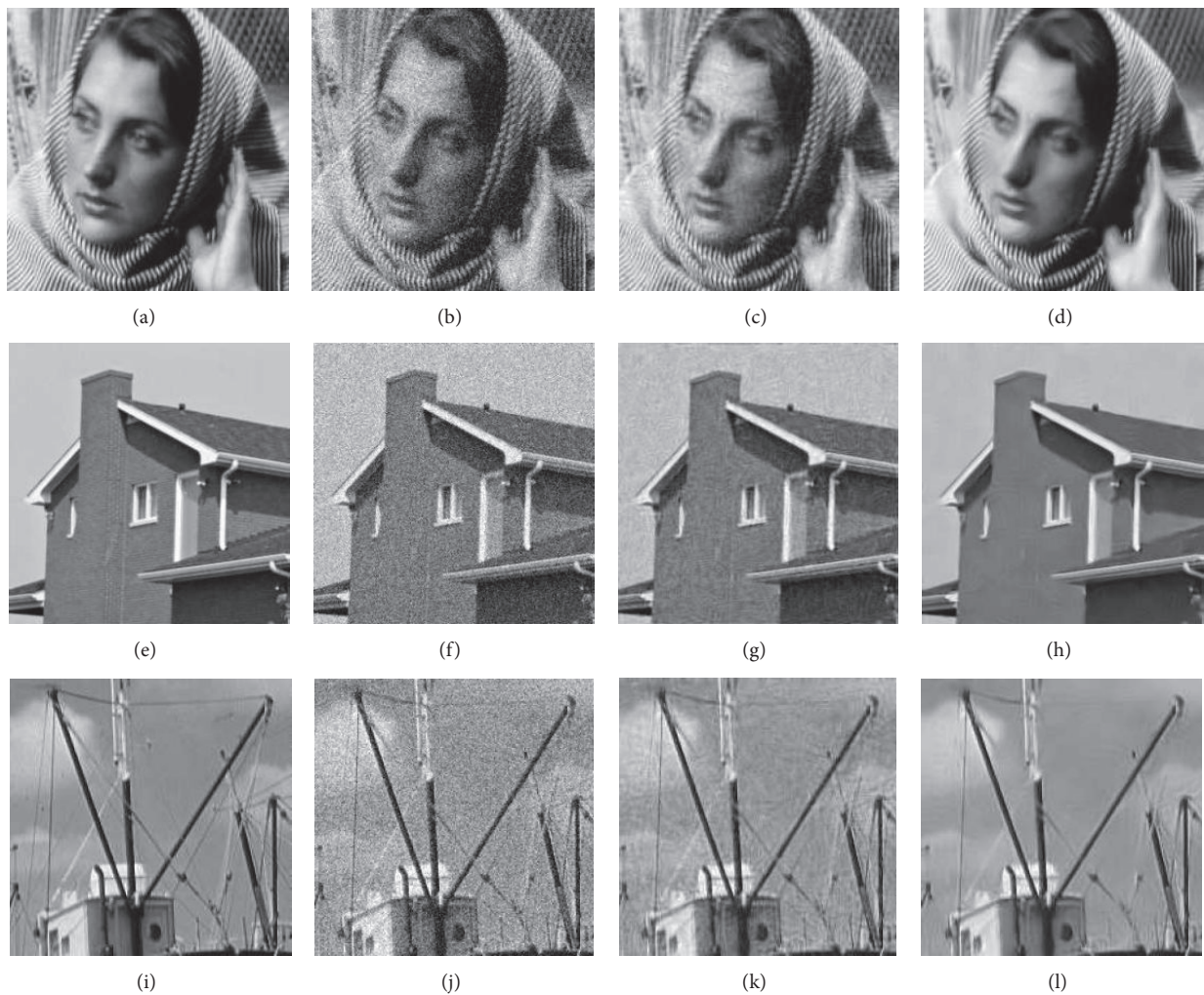FIGURE 6: (a), (b), (c), and (d) represent the magnified original, noisy ($\sigma = 20$), predenoised, and denoised subimages from Barbara image, respectively.

TABLE 3: PSNR comparison of the proposed and various state-of-the-art algorithms for color images.

Image	σ	BM3D	NL-Bayes	NLM	MHNLM
Lena	10	35.22	35.26	34.77	34.52
	15	33.94	33.75	33.15	33.15
	20	33.02	32.95	31.92	32.18
	25	32.27	32.08	30.96	31.48
	50	29.88	29.89	27.89	28.41
House	10	36.23	35.82	35.22	35.02
	15	34.85	34.90	33.75	33.55
	20	33.84	33.28	32.55	32.55
	25	33.03	32.42	31.44	31.84
	50	30.47	30.20	27.50	28.33
Peppers	10	33.78	33.92	33.74	33.20
	15	32.60	32.60	32.28	32.27
	20	31.83	31.84	31.24	31.27
	25	31.20	31.10	30.35	30.66
	50	28.93	28.98	27.14	28.02

| (a) | (b) | (c) | (d) |

| (e) | (f) | (g) | (h) |

| (i) | (j) | (k) | (l) |

FIGURE 7: Columns from left to right represent the magnified original, noisy ($\sigma = 20$), predenoised, and denoised subimages from Barbara, house, and boat images, respectively.
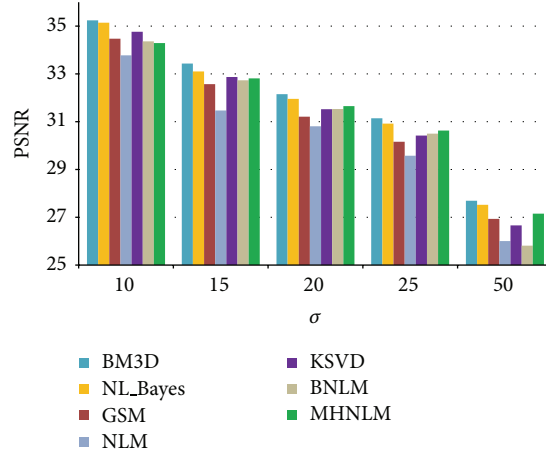
FIGURE 8: Average PSNR versus noise level σ comparison of BM3D, NL-Bayes, GSM, NLM, KSVD, BNLM, and MHNLM.

nicely explained in [37]. This is due to the fact that when the standard deviation of the additive noise is higher than the image contrast, features even removed during denoising process will not be perceivable as they may be buried in the severe noise removed from the image. In such situation, the empirical visual assessment of the method noise may not be reliable.

The method noise of the BM3D [25], NL-Bayes [35], and classical NL-means [19] are compared with that of the proposed algorithm. Each of the gray-scale images of Barbara, Lena, boats, and house is contaminated with additive white Gaussian noise of standard deviation $\sigma = 10$. In order to obtain proper visualization, the difference images are rescaled as described in [37]. The method noise results of the compared algorithms, including the proposed one, are similar to the white Gaussian noise as shown in Figure 9. In case of classical nonlocal means, the amplitude of the noise removed is uniform all over the image regardless of the structures like textures or edges present in the image as can be seen in Figures 9(c), 9(g), 9(k), and 9(o). For the rest of the algorithms, the amplitude of the removed noise varies with the geometrical features of the underlying image. Moreover the magnitudes of those differences obtained through classical nonlocal means are comparatively greater than the respective results for the rest of the approaches. The method noise result of the proposed algorithm is similar to those of the BM3D and NL-Bayes as shown in Figure 9. It is worth noticing that BM3D performs thresholding in 3D transform domain and NL-Bayes exploits the Bayesian framework for shrinkage of 3D groups of similar patches. In contrast to those two approaches, the proposed algorithm does not perform any thresholding. Instead, it applies the accelerated version of the classical nonlocal means filtering on the 2D transform domain. Even then, the noise method is much similar to BM3D and NL-Bayes. This similarity of the method noise indicates that multiscale nonlocal means filtering in transform domain is empirically equivalent to thresholding or shrinkage in transform domain. Further, like BM3D and NL-Bayes, magnitude of the noise removed by the proposed

algorithm depends upon the variation of geometry of the underlying image. Due to this fact, the proposed method also removes very fine and sharp discontinuities like thin ropes in boat image as shown in Figure 9(l).

4.4. Color Image Denoising. To evaluate the results of the proposed method for *RGB* color images, we consider the color images of Lena, House, and Peppers. In the first step of the algorithm, we use 9×9 patches, 15×15 search windows, and $h = 5\sigma$ as the filtering parameter. In the second step, we select 7×7 patches with filtering parameter $h = 4\sigma$. The search window size is selected as 15×15 for noise levels with $\sigma \leq 25$ and is changed to 19×19 otherwise. The proposed algorithm (MHNLM) is compared with various state-of-the-art algorithms: those by Dabov et al. [25] (BM3D), Lebrun et al. [35] (NL-Bayes), and Buades et al. [19] (NLM). The denoising results for NL-Bayes and standard NLM methods are obtained using source codes [46, 47], respectively (Figure 10). The source codes are available at the public domain IPOL. The PSNR for *RGB* color images is defined as

$$
\begin{aligned}
&\text{PNSR}\,(\widehat{u}) \\
&= 10 \log \left(\frac{(255)^2}{(3\,|X|)^{-1} \sum_{c=1}^{3} \sum_{x \in X} \left[u_o^{(c)}\,(x) - \widehat{u}^{(c)}\,(x) \right]^2} \right),
\end{aligned}
\tag{31}
$$

where $c = 1, 2, 3$ corresponds to the color channel *R*, *G*, *B*, respectively. Also, $|X|$ denotes the size of 2D image in the individual channel and u_o and $\widehat{u}$ represent the original and denoised *RGB* color images, respectively.

The results of BM3D and NL-Bayes are almost similar and clearly better than the proposed algorithm. Both the algorithms exploit the 3D grouping strategy, which is not employed in the proposed algorithm. Therefore, it would be interesting to examine the combined effects of 3D collaborative filtering and normal vectors patch comparison on denoising capability. For moderate and severe levels ($\sigma \geq 20$)

(a) Barbara (b) (c) (d)

(e) Lena (f) (g) (h)
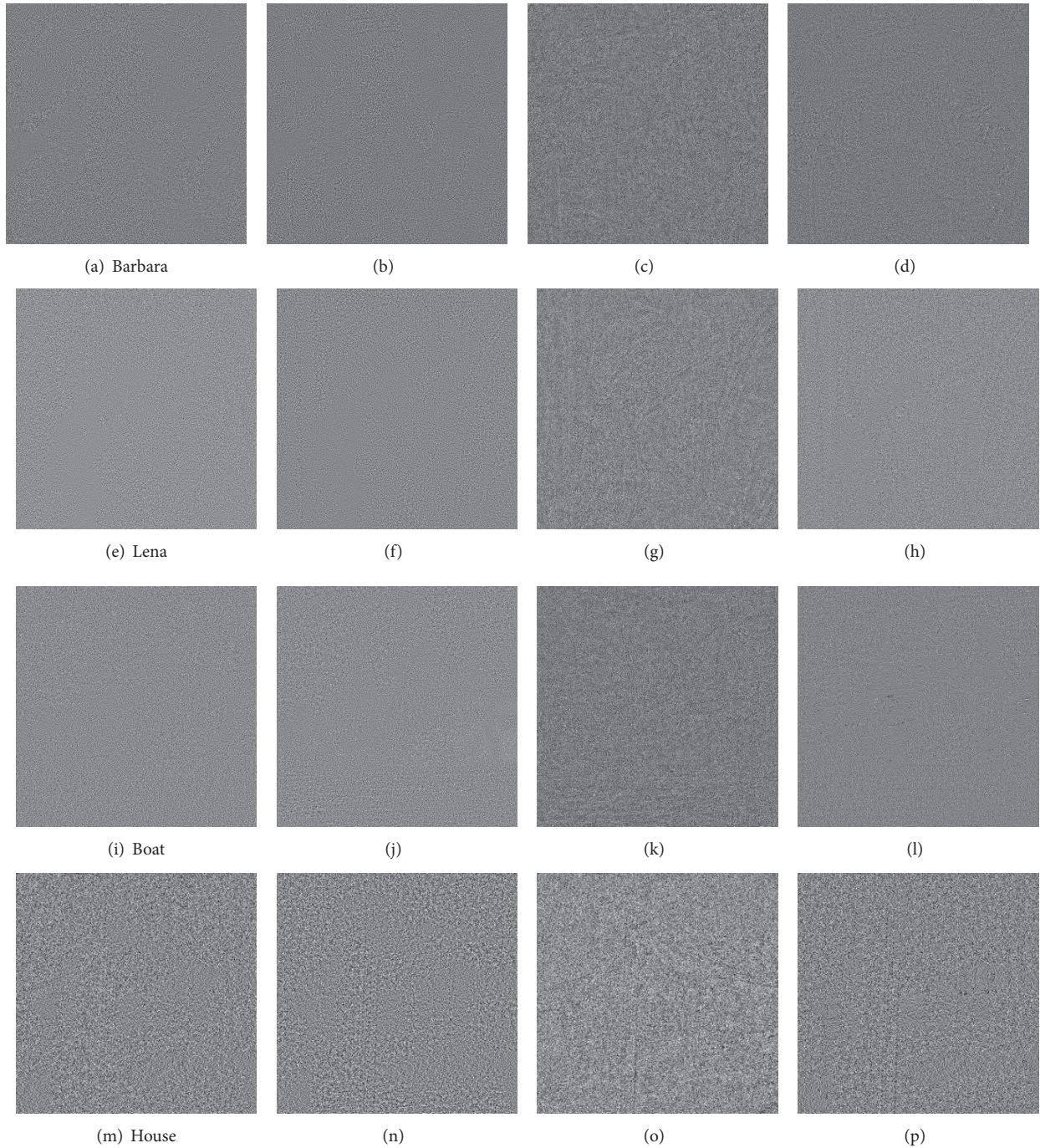
(i) Boat (j) (k) (l)

(m) House (n) (o) (p)

FIGURE 9: Columns from left to right represent difference of the noisy ($\sigma = 10$) and the denoised images using BM3D [25], NL-Bayes [35], NLM [19], and the proposed MHNLM algorithms, respectively. The rows from top to bottom indicate the results on the gray-scale images of Barbara, Lena, boat, and house, respectively.

of noise, the proposed algorithm yields better results than the classical nonlocal means in terms of PSNR. However, in the presence of low noise ($\sigma \leq 15$) the performance of the classical nonlocal means provides better results over the proposed algorithm. The performance of the proposed algorithm may be enhanced by certain modifications. For instance, 3D grouping used in BM3D or NL-Bayes can be combined with the proposed similarity measure.

5. Conclusions

A new definition of normal vectors patch is introduced to acquire more information about the similarity of intensity patches combined with a multiscale implementation of the accelerated nonlocal means filtering. The experimental results demonstrate the effectiveness of the proposed algorithm using the notion of normal vectors patch. In future,
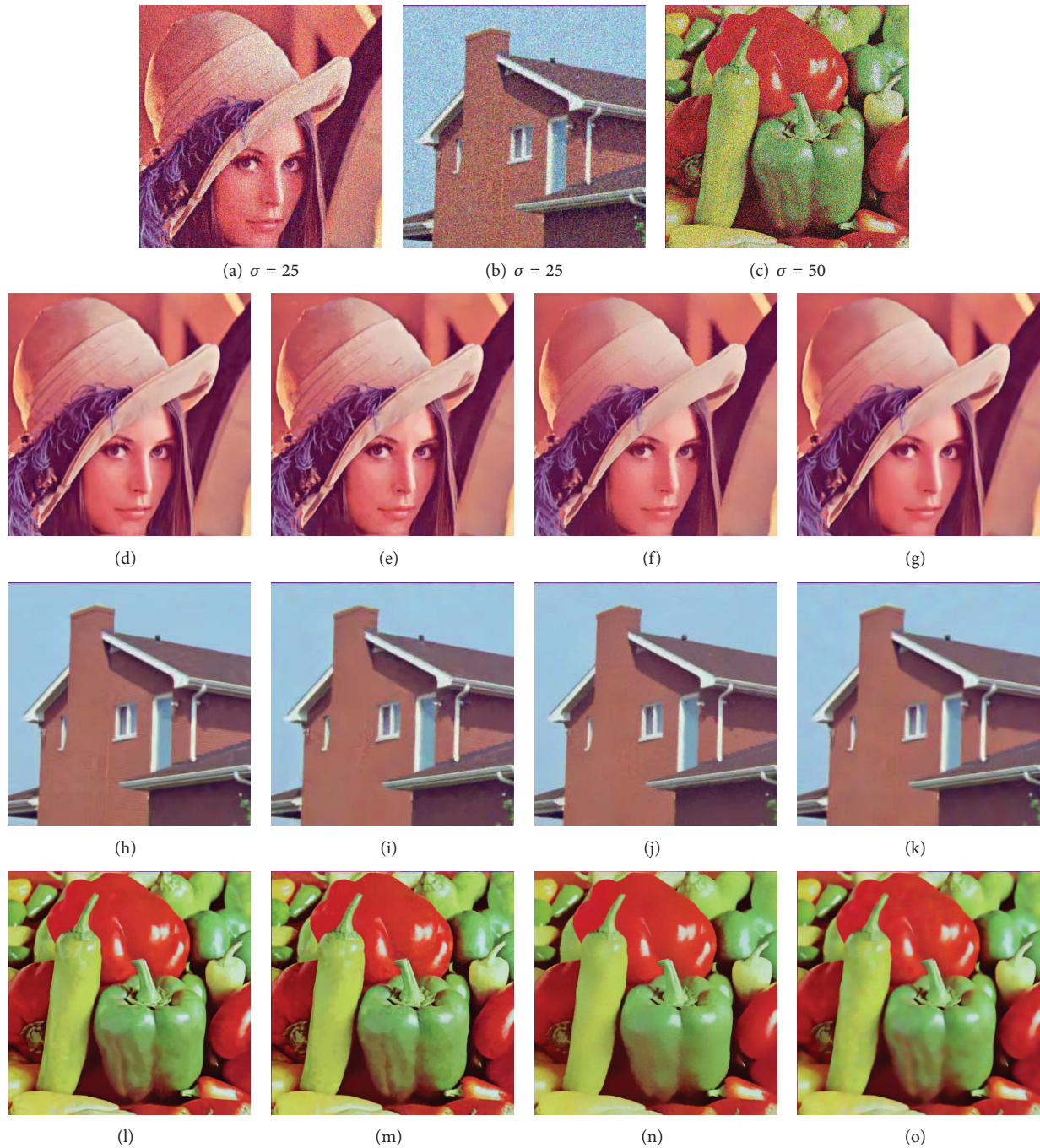
FIGURE 10: In the first row, the noisy *RGB* color images of Lena, house, and peppers are shown from left to right with standard deviations $\sigma = 25$, $\sigma = 25$, and $\sigma = 50$, respectively. Columns from left to right in the remaining rows show the denoising results of BM3D [25], NL-Bayes [35], NLM [19], and the proposed MHNLM algorithms, respectively.

we plan to extend this approach in 3D transform domain that may achieve enhanced sparsity level to further improve the image denoising capability.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

References

[1] L. I. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Physica D: Nonlinear Phenomena*, vol. 60, no. 1-4, pp. 259–268, 1992.

[2] S. Osher, A. Sol, and L. Vese, "Image decomposition and restoration using total variation minimization and the H^1," *Multiscale Modeling Simulation*, vol. 1, no. 3, pp. 349–370, 2003.

[3] G. Gilboa, N. Sochen, and Y. Y. Zeevi, "Variational denoising of partly textured images by spatially varying constraints," *IEEE Transactions on Image Processing*, vol. 15, no. 8, pp. 2281–2289, 2006.

[4] T. Chan, A. Marquina, and P. Mulet, "High-order total variation-based image restoration," *SIAM Journal on Scientific Computing*, vol. 22, no. 2, pp. 503–516, 2000.

[5] Q. Chen, P. Montesinos, Q. S. Sun, P. A. Heng, and D. S. Xia, "Adaptive total variation denoising based on difference curvature," *Image and Vision Computing*, vol. 28, no. 3, pp. 298–306, 2010.

[6] A. Chambolle, "An algorithm for total variation minimization and applications," *Journal of Mathematical Imaging and Vision*, vol. 20, no. 1-2, pp. 89–97, 2004.

[7] P. Perona and J. Malik, "Scale-space and edge detection using anisotropic diffusion," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 12, no. 7, pp. 629–639, 1990.

[8] Y.-L. You, W. Xu, A. Tannenbaum, and M. Kaveh, "Behavioral analysis of anisotropic diffusion in image processing," *IEEE Transactions on Image Processing*, vol. 5, no. 11, pp. 1539–1553, 1996.

[9] G. Steidl, J. Weickert, T. Brox, P. Mrázek, and M. Welk, "On the equivalence of soft wavelet shrinkage, total variation diffusion, total variation regularization, and sides," *SIAM Journal on Numerical Analysis*, vol. 42, no. 2, pp. 686–713, 2004.

[10] M. J. Black, G. Sapiro, D. H. Marimont, and D. Heeger, "Robust anisotropic diffusion," *IEEE Transactions on Image Processing*, vol. 7, no. 3, pp. 421–432, 1998.

[11] M. Lysaker, A. Lundervold, and X.-C. Tai, "Noise removal using fourth-order partial differential equation with applications to medical magnetic resonance images in space and time," *IEEE Transactions on Image Processing*, vol. 12, no. 12, pp. 1579–1589, 2003.

[12] T. Brox, J. Weickert, B. Burgeth, and P. Mrázek, "Nonlinear structure tensors," *Image and Vision Computing*, vol. 24, no. 1, pp. 41–55, 2006.

[13] S. Mallat and W. L. Hwang, "Singularity detection and processing with wavelets," *IEEE Transactions on Information Theory*, vol. 38, no. 2, pp. 617–643, 1992.

[14] J. Zhong and H. Sun, "Wavelet-based multiscale anisotropic diffusion with adaptive statistical analysis for image restoration," *IEEE Transactions on Circuits and Systems. I. Regular Papers*, vol. 55, no. 9, pp. 2716–2725, 2008.

[15] D. L. Donoho and I. M. Johnstone, "Ideal spatial adaptation by wavelet shrinkage," *Biometrika*, vol. 81, no. 3, pp. 425–455, 1994.

[16] M. K. Mihçak, I. Kozintsev, K. Ramchandran, and P. Moulin, "Low-complexity image denoising based on statistical modeling of wavelet coefficients," *IEEE Signal Processing Letters*, vol. 6, no. 12, pp. 300–303, 1999.

[17] C. Tomasi and R. Manduchi, "Bilateral filtering for gray and color images," in *Proceedings of the 6th International Conference on Computer Vision*, pp. 839–846, January 1998.

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

[19] A. Buades, B. Coll, and J. M. Morel, "A non-local algorithm for image denoising," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '05)*, vol. 2, pp. 60–65, 2005.

[20] A. Criminisi, P. Pérez, and K. Toyama, "Region filling and object removal by exemplar-based image inpainting," *IEEE Transactions on Image Processing*, vol. 13, no. 9, pp. 1200–1212, 2004.

[21] A. A. Efros and T. K. Leung, "Texture synthesis by non-parametric sampling," in *Proceedings of the 7th IEEE International Conference on Computer Vision (ICCV '99)*, vol. 2, pp. 1033–1038, September 1999.

[22] C. Kervrann, J. Boulanger, and P. Coup, "Bayesian non-local means filter , image redundancy and adaptive dictionaries for noise removal," in *Scale Space and Variational Methods in Computer Vision*, F. Sgallari, A. Murli, and N. Paragios, Eds., vol. 4485 of *Lecture Notes in Computer Science*, pp. 520–532, Springer, Berlin, Germany, 2007.

[23] M. Elad and M. Aharon, "Image denoising via sparse and redundant representations over learned dictionaries," *IEEE Transactions on Image Processing*, vol. 15, no. 12, pp. 3736–3745, 2006.

[24] T. Tasdizen, "Principal neighborhood dictionaries for nonlocal means image denoising," *IEEE Transactions on Image Processing*, vol. 18, no. 12, pp. 2649–2660, 2009.

[25] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, "Image denoising by sparse 3-D transform-domain collaborative filtering," *IEEE Transactions on Image Processing*, vol. 16, no. 8, pp. 2080–2095, 2007.

[26] J. Mairal, F. Bach, J. Ponce, G. Sapiro, and A. Zisserman, "Nonlocal sparse models for image restoration," in *Proceedings of the 12th International Conference on Computer Vision (ICCV '09)*, pp. 2272–2279, Kyoto, Japan, October 2009.

[27] G. Gilboa and S. Osher, "Nonlocal operators with applications to image processing," *Multiscale Modeling & Simulation*, vol. 7, no. 3, pp. 1005–1028, 2009.

[28] T. Brox, O. Kleinschmidt, and D. Cremers, "Efficient nonlocal means for denoising of textural patterns," *IEEE Transactions on Image Processing*, vol. 17, no. 7, pp. 1083–1092, 2008.

[29] L. Pizarro, P. Mrázek, S. Didas, S. Grewenig, and J. Weickert, "Generalised nonlocal image smoothing," *International Journal of Computer Vision*, vol. 90, no. 1, pp. 62–87, 2010.

[30] Z. Yang and M. Jacob, "Nonlocal regularization of inverse problems: a unified variational framework," *IEEE Transactions on Image Processing*, vol. 22, no. 8, pp. 3192–3203, 2013.

[31] D. Tschumperlé and L. Brun, "Non-local image smoothing by applying anisotropic diffusion PDE's in the space of patches," in *Proceedings of the IEEE 16th International Conference on Image Processing (ICIP '09)*, pp. 2957–2960, November 2009.

[32] J. Salmon, "On two parameters for denoising with non-local means," *IEEE Signal Processing Letters*, vol. 17, no. 3, pp. 269–272, 2010.

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

[34] A. Buades, B. Coll, and J. M. Morel, "A review of image denoising algorithms, with a new one," *Multiscale Modeling and Simulation*, vol. 4, no. 2, pp. 490–530, 2005.

[35] M. Lebrun, A. Buades, and J. M. Morel, "A nonlocal Bayesian image denoising algorithm," *SIAM Journal on Imaging Sciences*, vol. 6, no. 3, pp. 1665–1688, 2013.

[36] Y.-L. Liu, J. Wang, X. Chen, Y.-W. Guo, and Q.-S. Peng, "A robust and fast non-local means algorithm for image denoising," *Journal of Computer Science and Technology*, vol. 23, no. 2, pp. 270–279, 2008.

[37] M. Lebrun, M. Colom, A. Buades, and J. M. Morel, "Secrets of image denoising cuisine," *Acta Numerica*, vol. 21, pp. 475–576, 2012.

[38] P. Milanfar, "A tour of modern image filtering," *IEEE Signal Processing Magazine*, vol. 30, no. 1, pp. 106–123, 2013.

[39] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, "BM3D image denoising with shape-adaptive principal component analysis," in *Proceedings of the Workshop on Signal Processing with Adaptive Sparse Structured Representations (SPARS'09)*, 2009.

[40] L. Zhang, W. Dong, D. Zhang, and G. Shi, "Two-stage image denoising by principal component analysis with local pixel grouping," *Pattern Recognition*, vol. 43, no. 4, pp. 1531–1549, 2010.

[41] A. Foi, V. Katkovnik, and K. Egiazarian, "Pointwise shape-adaptive DCT for high-quality denoising and deblocking of grayscale and color images," *IEEE Transactions on Image Processing*, vol. 16, no. 5, pp. 1395–1411, 2007.

[42] S. Mallat, *A Wavelet Tour of Signal Processing: The Sparse Way*, Academic Press, New York, NY, USA, 3rd edition, 2008.

[43] G. Nason and B. Silverman, "The stationary wavelet transform and some statistical applications," in *Wavelets and Statistics*, A. Antoniadis and G. Oppenheim, Eds., vol. 103 of *Lecture Notes in Statistics*, pp. 281–299, Springer, New York, NY, USA, 1995.

[44] H. Zheng, A. Bouzerdoum, and S. L. Phung, "Wavelet based nonlocal-means super-resolution for video sequences," in *Proceedings of the 17th IEEE International Conference on Image Processing (ICIP '10)*, pp. 2817–2820, Hong Kong, September 2010.

[45] M. Z. Iqbal, A. Ghafoor, and A. M. Siddiqui, "Satellite image resolution enhancement using dual-tree complex wavelet transform and nonlocal means," *IEEE Geoscience and Remote Sensing Letters*, vol. 10, no. 3, pp. 451–455, 2013.

[46] M. Lebrun, A. Buades, and J.-M. Morel, "Implementation of the "Non-Local Bayes" (NL-Bayes) image denoising algorithm," *Image Processing On Line*, vol. 3, pp. 1–42, 2013.

[47] A. Buades, B. Coll, and J.-M. Morel, "Non-local means denoising," *Image Processing On Line*, vol. 1, 2011.

[48] R. Rubinstein, M. Zibulevsky, and M. Elad, "Double sparsity: learning sparse dictionaries for sparse signal approximation," *IEEE Transactions on Signal Processing*, vol. 58, no. 3, part 2, pp. 1553–1564, 2010.

Advances in
Operations Research

Advances in
Decision Sciences

Journal of
Applied Mathematics

Algebra

Journal of
Probability and Statistics

The Scientific
World Journal

International Journal of
Differential Equations

International Journal of
Combinatorics

Advances in
Mathematical Physics

Hindawi

Submit your manuscripts at
http://www.hindawi.com

Journal of
Complex Analysis

Journal of
Mathematics

Mathematical Problems
in Engineering

Abstract and
Applied Analysis

Discrete Dynamics in
Nature and Society

International
Journal of
Mathematics and
Mathematical
Sciences

Journal of
Discrete Mathematics

Journal of
Function Spaces

International Journal of
Stochastic Analysis

Journal of
Optimization