Computers and Electrical Engineering 70 (2018) 413–427

Contents lists available at ScienceDirect

Computers and Electrical Engineering

journal homepage: www.elsevier.com/locate/compeleceng

Image de-noising with subband replacement and fusion process using bayes estimators,☆,☆☆

Nazeer Muhammad[a], Nargis Bibi[b], Abdul Wahab[c,h], Zahid Mahmood[d], Tallha Akram[e], Syed Rameez Naqvi[e], Hyun Sook Oh[f], Dai-Gyoung Kim[g,*]

[a] Department of Mathematics, COMSATS Institute of Information Technology, G. T. Road, 47040, Wah Cantt., Pakistan
[b] Department of Computer Science, Fatima Jinnah Women University, The Mall, 46000, Rawalpindi, Pakistan
[c] Bio Imaging and Signal Processing Lab., Korea Advanced Institute of Science and Technology, Daejeon 34141, Korea
[d] Department of Electrical, COMSATS Institute of Information Technology, University Road, 22010, Abbotabad, Pakistan
[e] Department of Electrical Engineering, COMSATS Institute of Information Technology, University Road, 22010, Abbotabad, Pakistan
[f] Department of Applied Statistics, Gachon University, Seongnam-Si, 13120, Korea
[g] Department of Applied Mathematics, Hanyang University, Ansan, 15588, Korea
[h] Department of Mathematics, University of Education, Lahore, Attock Campus, Education Road, 43600, Attock, Pakistan

ARTICLE INFO

Article history:
Received 20 September 2016
Revised 19 May 2017
Accepted 19 May 2017
Available online 26 May 2017

Keywords:
Dictionary learning
Wavelet transform
Sub-band replacement
Fusion estimation
Factor analysis
Image de-noising

ABSTRACT

A hybrid image de-noising framework with an automatic parameter selection scheme is proposed to handle substantially high noise with an unknown variance. The impetus of the framework is to preserve the latent detail information of the noisy image while removing the noise with an appropriate smoothing and feasible sharpening. The proposed method is executed in two steps. First, the sub-band replacement and fusion process based on accelerated version of the Bayesian non local means method are implemented to enhance the weak edges that often result in low gradient magnitude and fade out during the de-noising process. Then, a truncated beta-Bernoulli process is employed to infer an appropriate dictionary of the edge enhanced data to obtain de-noising results precisely. Numerical simulations are performed to substantiate the restoration of the weak edges through sub-band replacement and fusion process. The proposed de-noising scheme is validated through visual and quantitative results using well established metrics.

1. Introduction

Restoration of images affected by inevitable noise is an important task in digital image processing. The main challenge in image de-noising is to eliminate noise effectively from a corrupted image by reconstructing an approximation of the clean image while preserving its latent information such as corners and edges [1–3]. The algorithms that fail to distinguish latent features from noise produce blurry outputs with de-noising artifacts [3]. The non-local means (NLM) based techniques, such

☆ This research was supported by the Korea Research Fellowship Program funded by the Ministry of Science, ICT and Future Planning through the National Research Foundation of Korea (NRF-2015H1D3A1062400).
☆☆ Reviews processed and recommended for publication to the Editor-in-Chief by Area Editor Dr. E. Cabal-Yepez.
* Corresponding author.
E-mail addresses: nazeermuhammad@ciitwah.edu.pk (N. Muhammad), nargis@fjwu.edu.pk (N. Bibi), wahab@kaist.ac.kr, abdul.wahab@ue.edu.pk (A. Wahab), zahid0987@ciit.net.pk (Z. Mahmood), tallha@ciitwah.edu.pk (T. Akram), rameeznaqvi@ciitwah.edu.pk (S.R. Naqvi), hoh@gachon.ac.kr (H.S. Oh), dgkim@hanyang.ac.kr (D.-G. Kim).

http://dx.doi.org/10.1016/j.compeleceng.2017.05.023
0045-7906/

Table 1
Comparison of image de-noising methods for House image based on PSNR values for mismatched estimations of σ.

σ	10/28.13			15/24.61			20/22.11			25/20.18		
Methods	σ_{+5}	σ_t	σ_{-5}	σ_{+5}	σ_t	σ_{-5}	σ_{+5}	σ_t	σ_{-5}	σ_{+5}	σ_t	σ_{-5}
OBNL	35.58	35.75	35.49	34.06	34.07	33.79	32.82	32.81	32.56	31.72	31.73	31.54
K-SVD	35.19	35.98	33.57	34.01	34.32	32.60	33.00	33.20	31.84	32.06	32.15	31.15
BM3D	35.60	36.71	30.42	34.44	34.94	30.92	33.52	33.70	31.30	32.72	**32.86**	31.26
BPFA	36.28	36.30	36.28	34.49	34.51	34.53	33.26	33.25	33.24	32.23	32.24	32.25
Proposed	**36.70**	**36.72**	**36.72**	**35.03**	**35.05**	**35.07**	**33.91**	**33.90**	**33.89**	**32.86**	**32.86**	**32.87**

as *Bayesian NLM filtering*, often perform well in maintaining sharp edges and corners (see for instance [4]). However, an important drawback of these algorithms is that the image is not processed if no similar patches are found. Moreover, the estimation of noise parameters is very delicate [5].

Optimization based methods have been effective in removing noise while preserving image structure (see, for example, [1,2]). Among those, *K-means singular value decomposition* (K-SVD) has demonstrated significant advantages [7]. It uses an over-complete learned dictionary matched to the images of interest. This works effectively on de-noising as the learned dictionaries render more adaptive priors for Bayesian estimation [4]. Unfortunately, the K-SVD algorithm is conditional since it applies only to the cases where the error parameter matches the *ground truth* [8]. In uncertainty an appropriate setting for finding K-SVD does not possess the adaptive dictionary of the similar pixel matching values in general. Therefore, the setting of either a sparsity level and a predefined dictionary size or an error threshold is required to determine the patch-specification. Other well known recent methodologies include *block-matching 3D filtering* (BM3D) [9], *learned simultaneous sparse coding* (LSSC) [10], *trainable nonlinear reaction diffusion* (TNRD) [11], *non-locally centralized sparse representation* (NCSR) [12], *weighted nuclear norm minimization* (WNNM) [14], *patch group prior based de-noising* (PGPD) [17], and *shrinkage fields* [18]. These methods show good performance in cases where an error parameter matches to the *ground truth* in the form of an exact noise variance [8]. However, their performance is significantly affected if the settings do not agree with the *ground truth*. The interested readers are referred, for instance, to [5] and Table 1 for further details.

In order to mitigate the aforementioned problem of matching the ground truth, a Bayesian non-parametric method, coined as *beta process factor analysis* (BPFA), has been suggested by Zhou et al. [5]. The beta process is derived by an underlying Poisson process, and its properties as a stochastic process for Bayesian modeling are well understood [19]. The purpose of using such a non-parametric Bayesian approach based on beta Bernoulli process is to infer the relative information in a non-parametric fashion. Besides, in BPFA method, the automatic inference of the sparsity level of the coefficients has been performed [5]. However, experimental results of the BPFA method in terms of *peak signal-to-noise ratio* (PSNR) do not correlate with outputs of related algorithms to achieve a state of the art performance [2]. Since the weak edges have low gradient magnitude and are similar to the background structure of noisy image, they are often corrupted by false or degraded edge formations during usual de-noising processes [2,8].

The aim of this article is to design a hybrid image de-noising framework with an integrated parameter selection scheme to handle substantially high noise having unknown variance. The main motivation behind this investigation is to design a hybrid Bayesian technique that benefits from the non-local property in keeping the latent features intact while availing the advantages of the dictionary learning with adaptive priors in removing the idiosyncratic noise from the corrupted image. The central idea is to effectively work in uncertainty of noise parameters where contemporary de-noising techniques often show compromised performance, and to substantiate that the dictionary redundancy performance of the dictionary learning system can be improved [7].

Towards this end, a two-step strategy is adopted. The first step consists of applying an optimized Bayesian NLM filtering using two different sets of parameters to noisy image followed by a sub-band replacement and fusion process. This hybrid application to the given noisy image provides the processed data with enhanced edges and corners. In the second step, dictionary learning is performed using adaptive patch size with a non-parametric Bayesian estimation.

The application of the NLM algorithm on noisy image, with appropriately tuned pairs of regularization parameters, patch sizes and search window sizes, furnishes two processed images, one with enhanced smooth regions (i.e. enhanced low frequency contents) and one with enhanced sharp regions (i.e. enhanced high frequency contents). Then a sub-band replacement is performed since the image with enhanced smooth regions (respectively sharp regions) may have degraded high (respectively low) frequency contents. These contents are replaced with those of the noisy image to obtain first stage processed images. The basic role of the proposed sub-band replacement is to preserve the salient features of the noisy image. The band splitting is achieved using multi-resolution wavelet transform [13,15,16]. The first stage processed images are then *fused* with the given noisy image for balancing the relevant information that may have been gotten out of place during the replacement procedure. The combining parameters in the fusion process are the weights estimated in terms of the *Shannon* entropy of the first stage processed images and the estimated noise variance of the noisy image. The process of fusion then provides two second stage processed images. We once again use NLM filtering to further refine the second stage images in order to remove the artifacts appeared because of sub-band replacement and fusion with the noisy image. The resulting images are then fused together furnishing the final stage processed data. In the second step of the de-noising framework a model based on truncated beta-process factor analysis [5] is invoked to separate the salient covariance configuration of the

data set from the idiosyncratic noise. This approach of dictionary learning is used to combine possibly weak predictors of an input noisy image in order to produce a unified predictor of strong edges with good noise reduction.

In many applications one may not know the exact noise variance in the confidence of the estimate [5]. The proposed method allows to relax other assumptions that have been made for image de-noising [4]. It provides a wide-ranging framework to handle the significant details of noisy images and is better parameterized to be state-of-the-art image de-noising approach. The obtained results in terms of PSNR are encouraging from the point of view of recovering the natural and synthetic image signals.

The contents of this article are arranged in the following manner. The band splitting, parameter initialization and Bayesian NLM algorithm are discussed in Section 2. The sub-band replacement and fusion process with adaptive weights are elaborated in Section 3. Then the method of dictionary learning using the beta-Bernoulli process is presented in Section 4. A few numerical experiments are conducted in Section 6 to compare the proposed method with its counterparts and to substantiate its efficiency and advantages. Finally, the conclusions are drawn in Section 7.

2. Mathematical modeling

2.1. Parameter evaluation

Let $\Omega \subset \mathbb{Z}^2$ be a bounded rectangular grid. Let $U = \{u(i) \mid i \in \Omega\}$ and $V = \{v(i) \mid i \in \Omega\}$ be the true and realized images, respectively, such that

$$v(i) = u(i) + n(i), \qquad i = (i_1, i_2) \in \Omega, \tag{1}$$

where $u(i) \in \mathbb{R}_+$ and $v(i) \in \mathbb{R}_+$ are the intensities of gray level and $n(i)$ is an independent and identically different (i.i.d.) Gaussian random noise with zero mean and variance σ^2 at pixel $i \in \Omega$. The continuous image is interpreted as the Shannon interpolation of the discrete grid of samples $v(i)$ over Ω. The goal here is to *de-noise* V in order to reconstruct true image U.

The rest of this section is dedicated to initializing and debating de-noising parameters, and to recalling the NLM filtering procedure.

2.2. Band splitting and noise estimation

Let j be the finest resolution level given by the size of the image V. Then the stationary wavelet transform of V, denoted by $\mathcal{W}[V]$, provides four sub-band data sets at the resolution level $j-1$. Precisely, it renders *approximated data* $V_{j-1}^{(ll)}$, *horizontally detailed data* $V_{j-1}^{(lh)}$, *vertically detailed data* $V_{j-1}^{(hl)}$, and *diagonally detailed data* $V_{j-1}^{(hh)}$. In this article, the *Symmlet* wavelets are used due to their ability to preserve the continuity of weak edges in transformed domain [2].

The band splitting using wavelet operator $\mathcal{W}$ paves the way for noise variance estimation, which is usually unknown and needs to be estimated from observed image V. In this framework, the *median absolute deviation* based noise estimation function of Donoho and Johnstone [20] is implemented. Precisely, the noise standard deviation σ is approximated from diagonally detailed data $V_{j-1}^{(hh)}$ by

$$\widehat{\sigma} := 1.48258 \times \text{median}\left(\left|V_{j-1}^{(hh)}\right|\right). \tag{2}$$

2.3. Non-local means filtering

The NLM image de-noising algorithm was proposed by Buades et al. [3] which aims to exploit *self-similarities* in an image by comparing *patches* across the whole image. The NLM filtering of V at pixel i, $\text{NL}[V](i)$, is perceived as the weighted average of all the pixel values $v(j)$ for $j \in \Gamma_i$, where Γ_i is the search window centered at i and of size $q \times q$ with $q \in \mathbb{N}$. Therefore,

$$\text{NL}[V](i) := \sum_{j \in \Gamma_i} w(i,j)v(j), \tag{3}$$

where $w(i, j)$ is defined as

$$w(i,j) := \frac{1}{c_i} \exp\left(\frac{-\left\|V(N_i) - V(N_j)\right\|_{2,a}^2}{h^2}\right) \tag{4}$$

with

$$c_i := \sum_{j \in \Gamma_i} \exp\left(\frac{-\left\|V(N_i) - V(N_j)\right\|_{2,a}^2}{h^2}\right).$$

Here $\|\cdot\|_{2,a}$ is a Gaussian-weighted Euclidean norm with a^2 being the variance of the Gaussian, N_k denotes a square neighborhood of a fixed size and centered at the pixel k, $V(N_k) = \{v(j) \mid j \in N_k\}$, and c_i is the normalization factor. The parameter

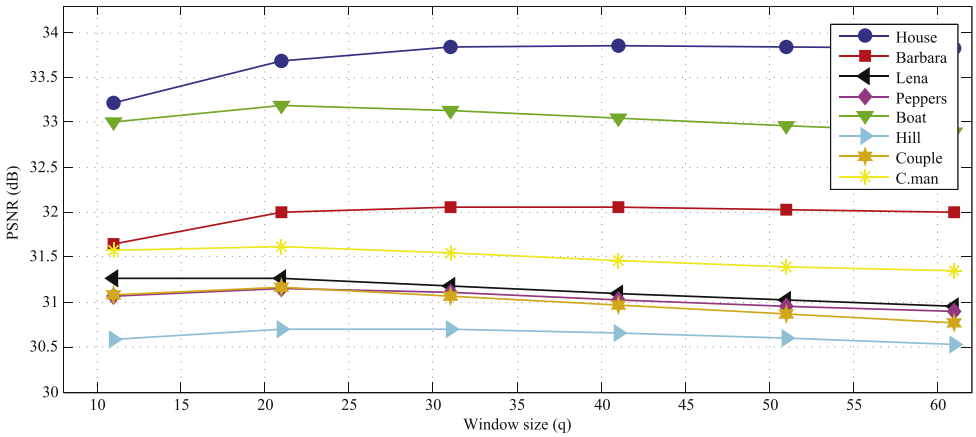

Fig. 1. PSNR versus window size for Bayesian NLM method with $\eta = 0.004$ and patch size $r = 7$, in the presence of additive white Gaussian noise with standard deviation $\sigma = 15$.

h controls the decay of the exponential function and is usually proportional to σ. In fact, h is estimated by

$$\widehat{h} := \widehat{\sigma}\,\eta, \qquad \eta > 0, \tag{5}$$

where η is a regularization parameter to retrieve the detail coherence from V. Choosing a small value of η tends to produce a noisy result, similar to V, whereas a very large value of η gives a smoothed image. Refer, for instance, to [2,4] for more details on tuning the parameters for the NLM filtering.

A better parametrized version of the NLM algorithm is suggested by Kervrann et al. [4] using Bayesian approach. They introduced local dictionaries and a new statistical distance measure to compare patches, whereas the amount of smoothing is directly determined by the noise variance given the patch size. A major drawback of the Bayesian NLM algorithm is the replacement of the unknown patches of U by V. Since V cannot be the exact replica of the true image U, the error thus incurred is not negligible. Therefore, it is not effective for de-noising the images corrupted by large noise variances. Consequently, a suitable parameter setting needs to be carefully considered. In the rest of this section the role of window and patch sizes on de-noising is discussed.

2.4. Window and patch sizes

Most of the NLM methods use a fixed window size $q \times q$ of search zone in which similar patches are explored. A small window size q (greater than patch size r) does not sufficiently reduce the noise. On the other hand, a very large q leads to *over-matching*. Consequently, the image details are degraded [2]. In practice, the window size is chosen empirically [21,22].

The effect of window size q on de-noising process in terms of PSNR using benchmark natural images is elucidated in Fig. 1. Two different groups of natural images based on uniform and textured regions are observed. The image data of group A (consisting of House and Barbara, having rich texture) have increasing PSNR for increasing values of q, whereas the image data of group B (Boat, Hill, Lena, Cameraman, Couple, and Peppers, having relatively more smooth regions) have decreasing PSNR for increasing q (greater than 21). In this article a varying window size strategy is proposed to separately handle smooth and oscillating regions.

At the first stage of the proposed de-noising framework, we choose a small patch size of 9×9 pixels (with $r = 9$). The combination of suitable parameters (η, q, r) for Bayesian NLM algorithm captures the local geometric patterns and textures without undesired distortions and renders a suitable processed noisy data for further edge enhancement.

3. Proposed method

In this section the Bayesian NLM algorithm is applied to the given noisy image V with two different sets of parameters $\Lambda_1 := (\eta_1, q_1, r)$ and $\Lambda_2 := (\eta_2, q_2, r)$ such that $q_1 < q_2$ and $\eta_2 < \eta_1$. This yields two processed images: F and G. In order to obtain the information of the smooth region of V, the large value of η is used while avoiding the artifacts or mismatched information of the smooth regions by choosing a small value of q. Therefore, the processed image F is expected to contain enhanced information associated with low frequencies. On the other hand, in order to capture fine details or local structures, small value of η is used while collecting more similar patches in the searching windows using a large value of q. Thus, G is expected to enhance information associated with high frequencies.

3.1. Sub-band replacement

Some of the important information of V may be lost during this salient information enhancement process, which can be recovered from V using a sub-band replacement process. The method is based on the replacement of a particular sub-band information of F or G with that of V in wavelet domain. In fact, the direct manipulation of the wavelet coefficients sometimes develops ringing artifacts and wavelet shaped noise [2].

The processed image F tends to have a better low frequency information, yet its high frequency contents may have been degraded during NLM filtering. Therefore, in the wavelet decomposition of F, the high frequency sub-bands are replaced with those of noisy image V. Precisely, $\mathcal{W}[F]$ renders four sub-band data sets at the resolution level $j-1$ given by $F_{j-1}^{(ll)}$, $F_{j-1}^{(hl)}$, $F_{j-1}^{(lh)}$ and $F_{j-1}^{(hh)}$. The last three data sets consist of high frequency contents and are replaced by $V_{j-1}^{(hl)}$, $V_{j-1}^{(lh)}$ and $V_{j-1}^{(hh)}$ respectively. This yields a processed image $\mathcal{W}[\widetilde{F}]$ in the wavelet domain with sub-bands $F_{j-1}^{(ll)}$, $V_{j-1}^{(hl)}$, $V_{j-1}^{(lh)}$ and $V_{j-1}^{(hh)}$. Subsequently, the inverse wavelet transform, $\mathcal{W}^{-1}$, yields the first stage processed image $\widetilde{F}$.

Similarly, the processed image G tends to have a better high frequency information but its low frequency contents may have been degraded. Therefore, in the wavelet decomposition of G, the low frequency sub-band is replaced with that of noisy image V. In this case, only $G_{j-1}^{(ll)}$ is replaced by $V_{j-1}^{(ll)}$. This yields the processed image $\mathcal{W}[\widetilde{G}]$ in the wavelet domain with sub-bands $V_{j-1}^{(ll)}$, $G_{j-1}^{(hl)}$, $G_{j-1}^{(lh)}$ and $G_{j-1}^{(hh)}$. Subsequently, the inverse wavelet transform yields another first stage processed image $\widetilde{G}$.

The first stage processed images $\widetilde{F}$ and $\widetilde{G}$ have to be merged with the given noisy image V for balancing the relevant information that may have been gotten out-of-place during the replacement procedure. In the next section a fusion process is introduced for preserving the latent detail information of V by merging it with $\widetilde{F}$ and $\widetilde{G}$.

3.2. Fusion process

Let us consider the aggregation of $\widetilde{G}$ and V by

$$\Psi := \mu_{1\widetilde{G}}\widetilde{G} + \mu_{2\widetilde{G}}V, \tag{6}$$

where the weights $\mu_{1\widetilde{G}}$ and $\mu_{2\widetilde{G}}$ are chosen according to the *(Shannon) entropy* H and estimated noise standard deviation $\widehat{\sigma}$ of V. The entropy H of $|\nabla\widetilde{G}|$ is given by

$$H = -\sum_{k=0}^{M-1} p_k \log_2 p_k, \tag{7}$$

where $M = 2^m$ $(m = 8bit)$ indicates the number of the different shades of gray level, p_k is the probability of the gray scale k and the magnitude of the gradient $|\nabla\widetilde{G}|$ of $\widetilde{G}$ is based on the Sobel edge filter. The entropy H measures the detail information content in the image. The images with higher values of H contain richer details [2]. Intuitively, if the value of H is high, $\widetilde{G}$ is enriched with fine details. Thus, a large weight $\mu_{1\widetilde{G}}$ should be assigned to $\widetilde{G}$. On the other hand, if $\widehat{\sigma}$ is small, V merits a large weight $\mu_{2\widetilde{G}}$. In this regard, following weights are suggested

$$\mu_{1\widetilde{G}} = \frac{2H\log_2\left(\widehat{\sigma}^2+1\right)}{M}, \quad \text{and} \quad \mu_{2\widetilde{G}} = 1-\mu_{1\widetilde{G}}, \quad \text{such that} \quad 0 \le \mu_{1\widetilde{G}}, \mu_{2\widetilde{G}} \le 1. \tag{8}$$

Similarly, consider the aggregation of $\widetilde{F}$ and V by

$$\Phi := \mu_{1\widetilde{F}}\widetilde{F} + \mu_{2\widetilde{F}}V, \tag{9}$$

where the weights $\mu_{1\widetilde{F}}$ and $\mu_{2\widetilde{F}}$ are defined in terms of the entropy of $\widetilde{F}$ and noise standard deviation $\widehat{\sigma}$ as in (8).

We call the images Φ and Ψ as the second stage processed images. At the final stage of pre-processing image V, the Bayesian NLM filtering is applied to Φ and Ψ with parameters Λ_1 and Λ_2 respectively, thereby obtaining refined images $\widetilde{\Phi}$ and $\widetilde{\Psi}$ followed by a fusion process

$$Y := \mu_1\widetilde{\Phi} + \mu_2\widetilde{\Psi}, \tag{10}$$

considering a trade-off influence between $\widetilde{\Phi}$ and $\widetilde{\Psi}$ where μ_1 and μ_2 are computed in terms of $\widetilde{\Phi}$ and $\widetilde{\Psi}$. Here Y is the final processed image to be used as an input in the second step of the proposed imaging framework.

4. Beta process dictionary learning

In this section the second step of the proposed de-noising framework is executed. The method of non-parametric Bayesian dictionary learning using the beta-Bernoulli process is applied to the enhanced image data Y, following [5]. First, a vector form $\mathbf{y}$ of Y is considered, where the vector $\mathbf{y} \in \mathbb{R}^{|\Omega|}$ (with $|\Omega|$ being the cardinality of the grid Ω) is attained by lexicographic ordering. In BPFA, the dictionary and sparse coefficient vectors are perceived as random variables and the expectation of variables is inferred nonparametrically [2]. Then, the beta process prior in terms of non-parametric factor

analysis model is used to decompose image $\mathbf{y}$ (or more specifically overlapped patches of $\mathbf{y}$) in terms of a product of the *factor matrix* and the corresponding *feature vectors*. A feature vector is further decomposed into an element-wise Hadamard product of a binary vector (indicating the absence or the presence of a feature) and a weight vector (containing feature values). The binary vector is modeled using a Bernoulli process parameterized by a beta process. The annoying level of noise around the edges contained in $\mathbf{y}$ is then removed using the Bayesian estimators for dictionary learning based on the beta-Bernoulli process. The processed image $\mathbf{y}$ with enhanced edges is highly effective in terms of a mixture of the Bernoulli processes and the mixing measure is completely random [8].

Before further discussion, we recall the definition of a two-parameter beta process. The interested readers are referred to [2] for further details.

Definition. Let $a, b \in \mathbb{R}_+$ and $\mathcal{H}_0$ be a base measure. Then the two-parameter beta process is represented as $\mathrm{BP}(a, b, \mathcal{H}_0)$ and a draw $\mathcal{H} \sim \mathrm{BP}(a, b, \mathcal{H}_0)$ may be represented as

$$\mathcal{H}(\boldsymbol{\varphi}) = \sum_{k=1}^{K} \boldsymbol{\theta}_k \delta_{\boldsymbol{\varphi}_k}(\boldsymbol{\varphi}) \tag{11}$$

with

$$\boldsymbol{\theta}_k \sim \mathrm{Beta}(a/K, b(K-1)/K) \text{ and } \boldsymbol{\varphi}_k \sim \mathcal{H}_0, \tag{12}$$

where δ is the Kronecker's delta function, $\boldsymbol{\theta}_k$ represents a vector of K probabilities associated with $\boldsymbol{\varphi}_k$ drawn i.i.d. from $\mathcal{H}_0$, and $\mathcal{H}$ is a valid measure with $K \to \infty$. □

Let $\left\{\mathbf{x}_i \in \mathbb{R}^P \mid i = 1, \ldots, N\right\}$ correspond to a set of N overlapped patches, of size $R \times R$, extracted from refined noisy image $\mathbf{y}$; where $P = R^2$. We aim to express $\mathbf{x}_i$ in the form

$$\mathbf{x}_i = \mathbf{D}\mathbf{w}_i + \boldsymbol{\epsilon}_i, \qquad i \in \{1, \ldots, N\}, \tag{13}$$

where matrix $\mathbf{D} \in \mathbb{R}^{P \times K}$ is a dictionary of K atoms $\left\{\mathbf{d}_k \in \mathbb{R}^P \mid k = 1, \ldots, K\right\}$ representing the columns of $\mathbf{D}$, the vector $\mathbf{w}_i \in \mathbb{R}^K$ is the coefficient vector of patch $\mathbf{x}_i$ in the dictionary $\mathbf{D}$ such that $\mathbf{D}\mathbf{w}_i \approx \mathbf{x}_i$, and $\boldsymbol{\epsilon}_i \in \mathbb{R}^P$ represents the aggregate image noise and (relatively small) residual error.

In order to find the de-noised version $\widehat{\mathbf{x}}_i := \mathbf{D}\mathbf{w}_i$ of a given imaging patch one should first learn the dictionary $\mathbf{D}$ and search for appropriate $\mathbf{w}_i$ such that $\mathbf{x}_i$ by itself is not exactly equal to $\widehat{\mathbf{x}}_i$. Towards this end, sparse vectors $\mathbf{w}_i$ are sought, so that $\mathbf{D}$ renders a compressed representation of $\mathbf{x}_i$. The assumption of sparsity dictates that $\mathbf{w}_i$ can be further decomposed as $\mathbf{w}_i = \mathbf{z}_i \odot \mathbf{s}_i$, where $\mathbf{z}_i \in \{0, 1\}^K$ contains binary valued assignments $z_{ik} \sim \mathrm{Bernoulli}(\theta_k)$ with $\theta_k \sim \mathrm{Beta}(a/K, b(K-1)/K)$ and $\mathbf{s}_i \sim \mathcal{N}\left(0, \gamma_s^{-1}\mathbf{I}_K\right)$ contains the corresponding weights $s_{ik} \in \mathbb{R}_+$. The operator $\odot$ represents the element-wise product of vectors and $\mathbf{I}_K$ is $K \times K$ identity matrix, which indicates that a same precision γ_s is used for all s_{ik} where $\gamma_s \sim \mathrm{Gamma}(c, d)$ is a non-informative gamma hyper-prior with hyper-parameters c, d initialized by 10^{-6}. As $K \to \infty$ the expected sparsity level is drawn from $\mathrm{Poisson}(a/b)$ whereas $\mathrm{Beta}(a/K, b(K-1)/K)$ approximates *improper beta distribution* with parameters a and b.

In this construction, the dictionary atoms $\mathbf{d}_k$ are modeled using multivariate zero-mean Gaussian distribution with variance $P^{-1}\mathbf{I}_P$, that is, $\mathbf{d}_k \sim \mathcal{N}\left(0, P^{-1}\mathbf{I}_P\right)$ and the error $\boldsymbol{\epsilon}_i$ is drawn from $\mathcal{N}\left(0, \gamma_\epsilon^{-1}\mathbf{I}_P\right)$ with the precision $\gamma_\epsilon \sim \mathrm{Gamma}(e, f)$ where hyper-parameters e and f are also initialized by 10^{-6} [5].

In a nutshell, the BPFA model can be expressed as (see, for instance, [5])

$$\begin{cases} \mathbf{x}_i = \mathbf{D}\mathbf{w}_i + \boldsymbol{\epsilon}_i, \\ \mathbf{w}_i = \mathbf{z}_i \odot \mathbf{s}_i, \\ \mathbf{d}_k \sim \mathcal{N}\left(0, P^{-1}\mathbf{I}_P\right), \\ \mathbf{s}_k \sim \mathcal{N}\left(0, \gamma_s^{-1}\mathbf{I}_K\right), \\ \mathbf{z}_i \sim \Pi_{k=1}^{K}\mathrm{Bernoulli}(\theta_k), \\ \theta_k \sim \mathrm{Beta}(a/K, b(K-1)/K), \\ \boldsymbol{\epsilon}_i \sim \mathcal{N}\left(0, \gamma_\epsilon^{-1}\mathbf{I}_P\right), \\ \gamma_s \sim \mathrm{Gamma}(c, d), \\ \gamma_\epsilon \sim \mathrm{Gamma}(e, f). \end{cases} \tag{14}$$

Since the elements in BPFA model belong to the conjugate exponential family, the posterior inference is implemented via *Gibbs sampling method* with analytic update equations [8]. The final de-noised image $\widehat{\mathbf{u}}$ ($\widehat{U}$) is finally obtained from the processed noisy image data $\mathbf{y}$ by implementing BPFA model and the inference executed by Gibbs sampling.

5. Hierarchy of the proposed framework

The flow of the image processing in the proposed scheme is elaborated in Fig. 2. and is summarized below.

Step 1. Input: Noisy image V. Output: processed noisy image Y.

a. Find $\widehat{\sigma}$ of input noisy image using (2).

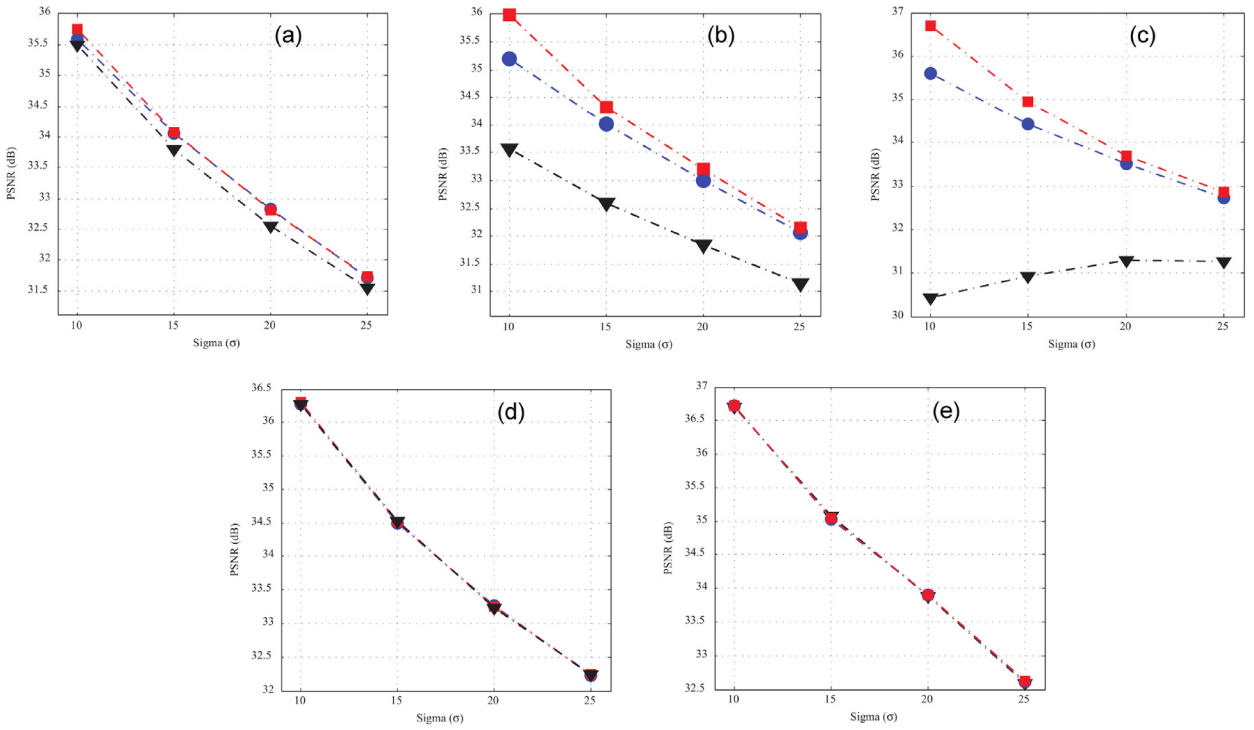

Fig. 2. Comparison of image de-noising PSNR versus σ_t (squares), σ_{-5} (triangles) and σ_{+5} (circles) for gray scale House image of size 256 × 256. (a) OBNL (b) K-SVD (c) BM3D (d) BPFA (e) Proposed method.

b. Apply Bayesian NLM filtering on V with $\Lambda_1 = (\eta_1, q_1, r)$ to get F and with $\Lambda_2 = (\eta_2, q_2, r)$ to get G.
c. Perform sub-band replacement on F and G.
   i. Find $\mathcal{W}[F]$, $\mathcal{W}[G]$ and $\mathcal{W}[V]$.
   ii. Replace high frequency data sets in $\mathcal{W}[F]$ with those of $\mathcal{W}[V]$ to get $\mathcal{W}[\widetilde{F}]$.
   iii. Replace low frequency data set in $\mathcal{W}[G]$ with that of $\mathcal{W}[V]$ to get $\mathcal{W}[\widetilde{G}]$.
   iv. Apply $\mathcal{W}^{-1}$ to get first stage refined images $\widetilde{F}$ and $\widetilde{G}$.
d. Perform fusion process on $\widetilde{F}$ and $\widetilde{G}$.
   i. Define weights $\mu_{1\widehat{G}}$ and $\mu_{2\widehat{G}}$ satisfying (8) and evaluate second stage processed image Ψ as the aggregation of $\widetilde{G}$ and V using (6).
   ii. Define weights $\mu_{1\widehat{F}}$ and $\mu_{2\widehat{F}}$ satisfying (8) and evaluate second stage processed image Φ as the aggregation of $\widetilde{F}$ and V using (9).
   iii. Apply Bayesian NLM filtering to Φ with parameters Λ_1 to get $\widetilde{\Phi}$ and to Ψ with parameters Λ_2 to get $\widetilde{\Psi}$.
   iv. Choose weights μ_1 and μ_2 and find the processed image Y as the aggregate of $\widetilde{\Phi}$ and $\widetilde{\Psi}$ using (10).

Step 2 Input: Vector form $\mathbf{y}$ of noisy data Y. Output: Denoised image $\widehat{\mathbf{u}}$.

a. Estimate adaptive patch size $\widehat{P}$ from $|\Omega|$ (refer to (15)) and derive basic estimation of $\widehat{\sigma}$.
b. Construct the patches $\mathbf{x}_i$ from $\mathbf{y}$
c. Initialize the dictionary $\mathbf{D}$ with $\mathbf{d}_k$, $\mathbf{z}_i$, $\mathbf{s}_i$, $\mathbf{w}_i$, $\boldsymbol{\epsilon}_i$ as defined in BPFA model.
d. Evaluate the output image $\widehat{\mathbf{u}}$ as the average of all $\widehat{\mathbf{x}}_i = \mathbf{D}\mathbf{w}_i$.

6. Numerical results and discussion

This section is dedicated to the numerical validation of the proposed de-noising scheme. The quantitative comparison is performed in terms of PSNR (in dB), which quantifies the recovery by comparing the ground truth of the input clean image and the de-noised image. The clean image is used only for the PSNR calculation at the final stage. The computational speed of the proposed framework is roughly twice as compared to the BPFA [5].

The performance of the proposed de-noising algorithm is compared with that of K-SVD [7], optimized Bayesian NLM algorithm (OBNL) [4], BPFA [5] and BM3D filtering [9]. The recent de-noising algorithms take account of BM3D filtering [9] for comparative analysis. However, the comparison of the proposed framework with BM3D algorithm is not appropriate since BM3D assumes an à priori known noise variance which is not the case here and the crux of the proposed framework is to infer the noise variance automatically. Nevertheless, the comparison is done for the completeness sake and to establishing

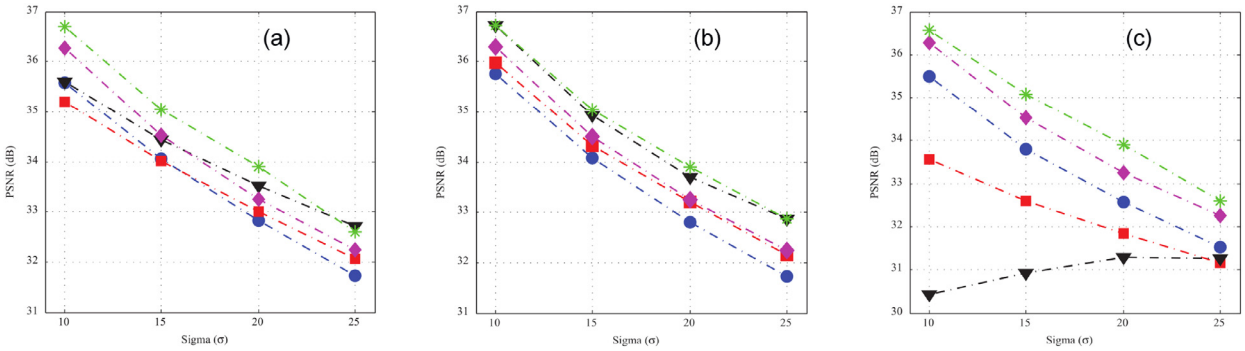

Fig. 3. Comparison of image de-noising PSNR versus different noise variance estimations using OBNL (circles), K-SVD (squares), BPFA (diamonds), BM3D (triangles) and proposed method (asterisks). (a) σ_{+5} (b) σ_t (c) σ_{-5}.

the validity of the proposed framework for both the cases of a known or an unknown noise variance. It is generally predicted that BM3D does not perform effectively in the sense of de-noising accuracy with respect to errors in the estimated noise level [8].

The OBNL algorithm uses a Bayesian framework to derive a NLM filter, adaptive to a relevant noise model, for preserving edges and texture details of noisy image [4]. The K-SVD algorithm learns the sparsity pattern of the coding vector using the orthogonal matching pursuit algorithm [7]. The sparsity level is fixed and each dictionary element and dimension of sparse matrix are jointly updated by a rank one approximation to the residual. The BPFA algorithm updates the sparsity pattern by generating the weights and learning dictionary from Gaussian posterior distributions. The BM3D de-noising algorithm processes blocks within the image in a sliding manner and utilizes the block-matching concept by searching for blocks which are similar to the processed ones [9]. The proposed algorithm first sharpens the edges by improving the essential details using Bayesian NLM framework along with application of stationary wavelet transform, sub-band replacement and fusion process, and then exploits beta Bernoulli process for de-noising. The OBNL image noise reduction code is taken from http://www.bic.mni.mcgill.ca/PersonalCoupepierrick/HomePag , the K-SVD dictionary learning algorithm code is taken from http://cs.technion.ac.il/~elad/software , BPFA non-parametric Bayesian dictionary learning code is taken from http://people.duke.edu/~mz31/Results/BPFAImage/ and BM3D code is taken from http://www.cs.tut.fi/~foi/GCF-BM3D/index.html#ref_software.

6.1. Parameter selection

The true value of a noisy pixel i is deduced using its neighborhood with 8×8 patch size in general, whereas most of the articles focus on de-noising results for $\sigma \in [10, 40]$. However, the size of the patches should be adaptive to the nature and amount of noise level. In fact, a patch size 8×8 (with $P = 8$) may appear to be insufficient for a very large σ, however the same might be better for a small σ [19]. Towards this end, the patch size can be decided according to the estimated value $\widehat{\sigma}$ of noise parameter and the size $|\Omega|$ of the given noisy image as

$$\widehat{P} = \begin{cases} \left\lfloor \sqrt{\widehat{\sigma}\log(|\Omega|+1)} \right\rfloor, & \left\lfloor \sqrt{\widehat{\sigma}\log(|\Omega|+1)} \right\rfloor < 16, \\ 16, & \left\lfloor \sqrt{\widehat{\sigma}\log(|\Omega|+1)} \right\rfloor > 16. \end{cases} \tag{15}$$

The minimum patch size for least amount of the estimated $\widehat{\sigma}$ is taken as 3×3. It gradually increases up to maximum patch size of 16×16 for a large $\widehat{\sigma}$ and in direct proportion to $|\Omega|$. The dictionary size is fixed at $K = 1024$ for all images. Moreover, $r = 9$, $\eta_1 = 0.0048$, $\eta_2 = 0.0016$, $q_1 = 25$, and $q_2 = 63$ are used to construct $\Lambda_1 = (\eta_1, q_1, r)$ and $\Lambda_2 = (\eta_2, q_2, r)$.

6.2. Validity of automatic noise estimation

In order to validate the automatic noise estimation technique, the PSNR values of the proposed framework are compared with those of OBNL, K-SVD, BM3D and BPFA for de-noising gray scale House image in Figs. 2 and 3, and Table 1. Three different scenarios are considered:

a. When the estimated noise parameter matches the ground truth, denoted by σ_t.
b. When the noise parameter is over-estimated by 5, denoted by σ_{+5}.
c. When the noise parameter is under-estimated by 5, denoted by σ_{-5}.

In Fig. 2 the performance of each one of the listed de-noising algorithms is evaluated in terms of PSNR versus σ_t, σ_{+5} and σ_{-5}. Unsurprisingly, the performance of K-SVD and BM3D suffers when the estimated noise differs from the ground truth. The performance of OBNL is less affected by the mismatch. On the other hand, BPFA and the proposed framework show

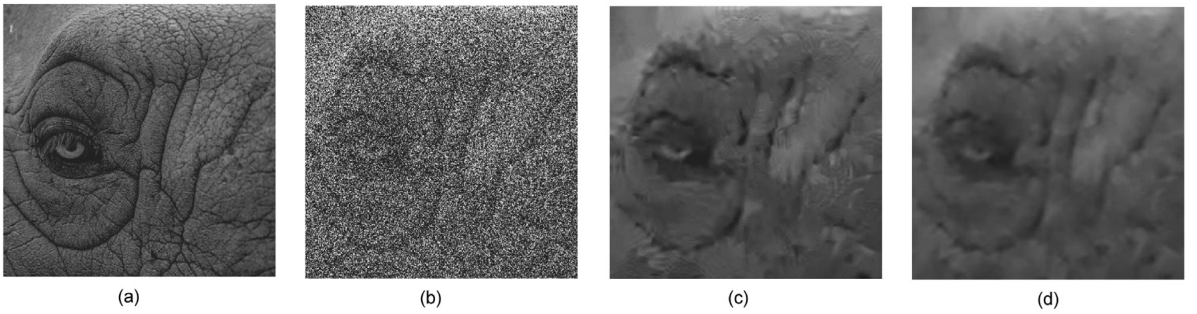

Fig. 4. De-noising experience for zoomed-in fragment of elephant eye along skin image (the facts are better perceived by digital zooming). (a) Clean image (b) Noisy image ($\sigma = 100$) (c) BM3D (PSNR = 22.48) (d) Proposed method (PSNR = 22.53).

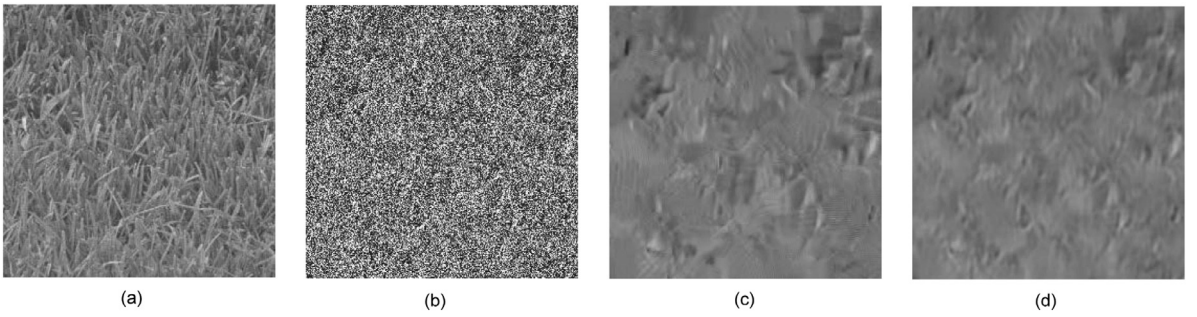

Fig. 5. De-noising experience for zoomed-in fragment of gross image (the facts are better perceived by digital zooming). (a) Clean image (b) Noisy image ($\sigma = 100$) (c) BM3D (PSNR = 21.39) (d) Proposed method (PSNR = 21.58).

consistent results. Indeed, they are adaptive to inferring the noise variance which leads to an improvement and consistency in de-noising.

In Fig. 3 the outputs of different methods are compared for σ_t, σ_{+5} and σ_{-5}. The performances of BPFA and K-SVD are comparable for σ_t (exact match). However, BPFA outperforms K-SVD for a mismatch. It is clearly demonstrated that the proposed method outperforms OBNL, K-SVD, BM3D and BPFA when the noise estimation deviates from the ground truth. Moreover, the results of BM3D and the proposed method are comparable for the case of matched ground truth.

A detailed comparison of image de-noising methods based on PSNR values against σ_t, σ_{+5} and σ_{-5} for varying noise levels is displayed in Table 1. It can be observed that for mismatched noise estimation of the proposed method outperforms all the other de-noising methods under discussion. Note that BM3D shows comparable results in some cases where the relative error in the noise estimation is very small but the PSNR values obtained by the proposed method are still comparable with those of BM3D. Since, one does not know whether the ground truth for estimated noise is matched or not in practice, the proposed method is potentially very stable in the sense of PSNR outputs.

6.3. De-noising results for benchmark natural images

Fig. 4. shows only a few disturbing artifacts for relatively high level of noise with $\sigma = 50$. The proposed framework attains good preservation of sharp details (hair and hat edges), smooth regions (the cheeks), and oscillatory patterns (around the eye brows, eye balls, and lips); whereas most of the other algorithms fail to preserve latent detail information.

Fig. 5. validates the proposed method for its capability to preserve textures. A fragment of noisy Barbara image with $\sigma = 100$ is de-noised. Although the true signal is almost fully corrupted, the clothes stripes are effectively reinstated by the proposed framework and BM3D. Once again, a higher PSNR value is achieved by the proposed framework than BM3D despite the fact that a known noise parameter is used for de-noising. It is interesting to note that all the other methods fail to recover fine structure details.

Table 2 displays that across different levels of noise, the proposed method consistently and significantly performs better than all the other listed methods (except BM3D in certain cases) since it benefits from both the non local property and the wavelet multi-resolution advantages [6,23,24].

It is favored by the strategy of restoring the weak edges for obtaining processed data and then using the beta Bernoulli process along with adaptive patch size. The results of BM3D appear to be superior at number of occasions since a known noise parameter is used to evaluate PSNR values, whereas the strength of the proposed method is to work with uncertainty of noise estimation. The proposed scheme shows excellent performance as compared to BM3D for images containing rich textures (Barbara and House) at high level of noises due to its integrated edge enhancement approach. In order to substan-

Table 2
Comparison of PSNR values for different methods at various noise levels (σ_t known).

σ	Methods	C.Man	House	Peppers	Lena	Barbara	Boat	Hill	Man	Couple
15/24.61	BM3D	**31.91**	34.94	**32.70**	**34.27**	**33.11**	**32.14**	31.86	**31.93**	**32.11**
	OBNL	30.24	34.07	31.67	30.24	32.22	31.32	31.19	31.16	31.24
	K-SVD	31.41	34.32	32.22	33.70	32.37	31.74	31.48	31.53	31.46
	BPFA	31.35	34.51	32.45	33.92	32.62	31.98	31.64	31.74	31.74
	Proposed	31.50	**35.05**	32.55	34.22	**33.11**	32.02	**31.89**	31.92	31.92
20/22.11	BM3D	**30.48**	33.77	**31.29**	**33.05**	31.78	**30.88**	**30.72**	30.59	**30.76**
	OBNL	29.14	32.81	30.29	29.14	30.88	29.94	29.88	29.74	29.82
	K-SVD	29.91	33.20	30.82	32.38	30.83	30.37	30.20	30.14	30.02
	BPFA	30.04	33.25	31.10	32.66	31.10	30.71	30.46	30.43	30.34
	Proposed	30.11	**33.90**	31.22	33.03	**31.80**	30.81	**30.72**	**30.61**	30.65
25/20.18	BM3D	**29.45**	**32.86**	30.16	32.08	30.72	**29.91**	**29.85**	**29.62**	**29.72**
	OBNL	28.28	31.73	29.15	28.28	29.73	28.86	28.85	28.67	28.61
	K-SVD	28.85	32.15	29.73	31.32	29.61	29.29	29.20	29.10	28.91
	BPFA	29.00	32.24	29.99	31.63	29.88	29.71	29.58	29.45	29.29
	Proposed	29.12	**32.86**	**30.17**	**32.10**	**30.78**	29.85	29.84	**29.62**	29.65
50/14.16	BM3D	26.12	29.69	26.68	29.05	27.23	26.78	27.19	26.81	**26.46**
	OBNL	25.09	27.51	25.16	25.09	25.76	25.36	25.76	25.52	24.76
	K-SVD	25.73	27.95	26.13	27.79	25.45	25.96	26.28	26.09	25.32
	BPFA	25.67	28.49	26.47	28.30	26.04	26.51	26.82	26.53	25.94
	Proposed	**26.24**	**29.79**	**26.71**	**29.07**	**27.31**	**26.89**	**27.22**	**26.84**	26.41
100/8.14	BM3D	**23.07**	**25.87**	**23.39**	25.95	23.62	23.97	24.58	**24.22**	**23.51**
	OBNL	21.17	23.24	21.23	21.17	22.13	22.31	23.27	22.82	21.99
	K-SVD	21.69	23.71	21.75	24.46	21.88	22.81	23.99	23.43	22.60
	BPFA	21.93	24.38	22.74	24.96	22.14	23.33	24.23	23.76	23.02
	Proposed	22.88	25.75	23.33	**25.99**	**23.74**	**23.99**	**24.64**	**24.22**	23.43

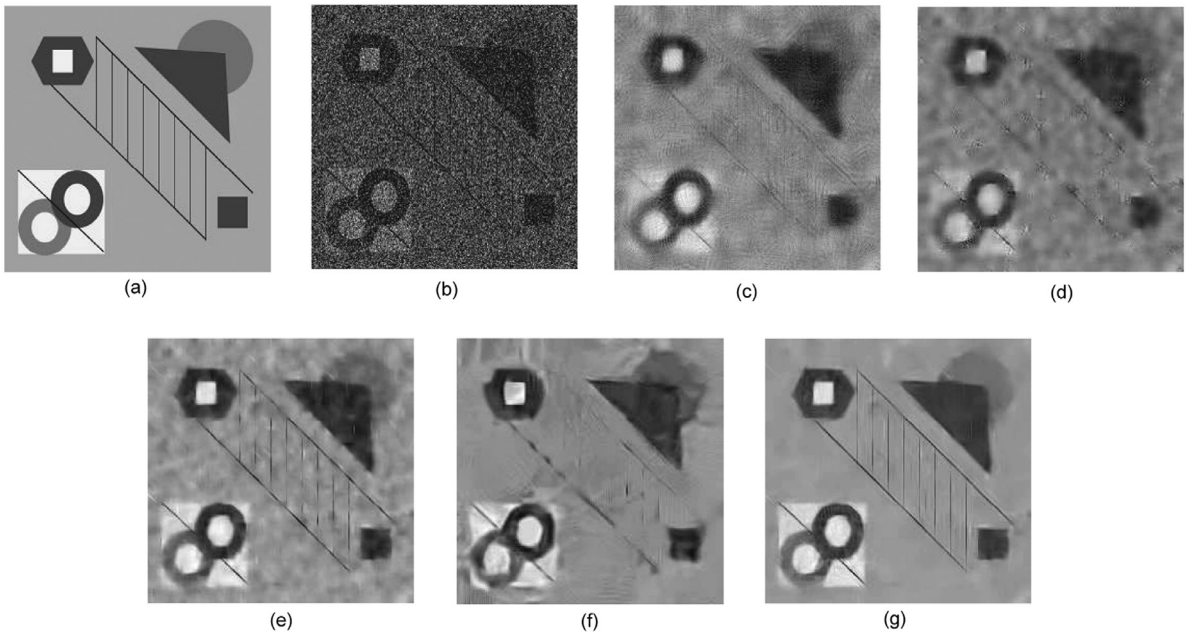

Fig. 6. De-noising experience for synthetic image (the facts are better perceived by digital zooming). (a) Clean image (b) Noisy image ($\sigma = 100$) (c) OBNL (PSNR=21.36) (d) K-SVD (PSNR = 21.86) (e) BPFA (PSNR = 22.95) (f) BM3D (PSNR = 24.42) (g) Proposed method (PSNR = 25.52).

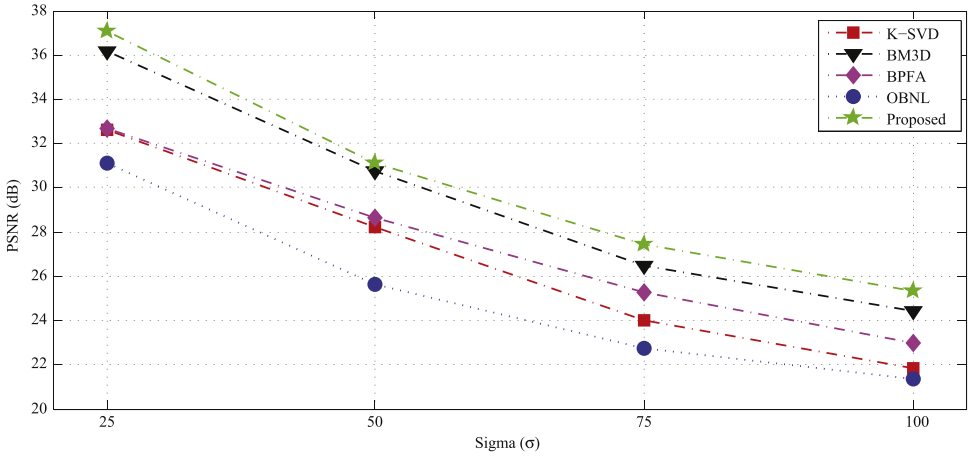

Fig. 7. De-noising experience for synthetic image (the facts are better perceived by digital zooming). (a) Clean image (b) Noisy image ($\sigma = 100$) (c) OBNL (PSNR = 21.36) (d) K-SVD (PSNR = 21.86) (e) BPFA (PSNR = 22.95) (f) BM3D (PSNR = 24.42) (g) Proposed method (PSNR = 25.52).

tiate this inference, we consider in Figs. 4 and 5 two segments of rich texture images (elephant eye along skin image and grass image) corrupted by heavy noise ($\sigma = 100$) and compare the de-noising results of the proposed algorithm with those of BM3D. These textural images show only limited retrievals. However, the proposed method retains the salient information with pleasant structure of de-noised image in comparison with the benchmark method BM3D and achieves better PSNR values.

6.4. De-noising results for synthetic image

Besides the aforementioned results, the proposed algorithm is further tested in Figs. 6 and 7 on synthetic image which contains different geometric structures (pentagon, triangle, circles, and lines etc.) and is contaminated by an additive white Gaussian noise. Four different noise levels with $\sigma = 25$, 50, 75, and 100 are considered to show the significant improvements by proposed algorithm and the comparison with listed algorithms is presented in Table 3.

In Fig. 6 the visual restoration results for the noisy synthetic image with $\sigma = 100$ are presented. It is noticed that the proposed method still recovers some spurious structures in the smooth regions of the synthetic image that appear to be

Table 3
Image de-noising PSNR comparison of synthetic image with different techniques.

σ	25/20.15	50/14.13	75/10.60	100/08.10
BM3D	36.16	30.76	26.46	24.42
OBNL	31.10	25.63	22.74	21.36
K-SVD	32.63	28.23	24.02	21.86
BPFA	32.70	28.66	25.26	22.95
Proposed	**36.87**	**31.32**	**27.49**	**25.52**

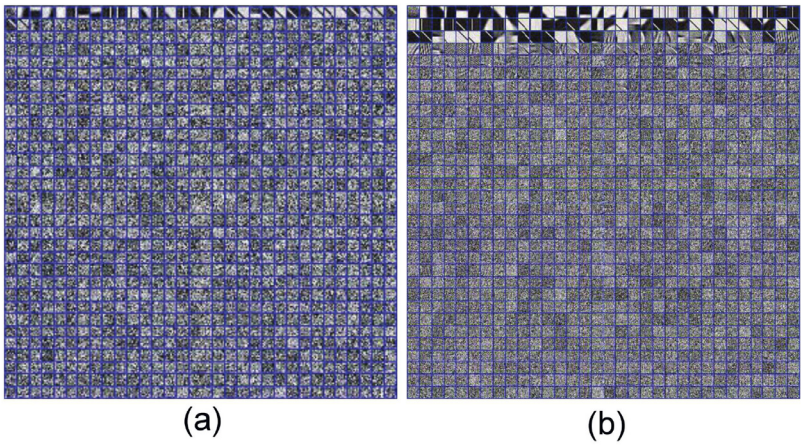

Fig. 8. Learned dictionaries for synthetic image with $\sigma = 100$. (a) BPFA (b) Proposed method.

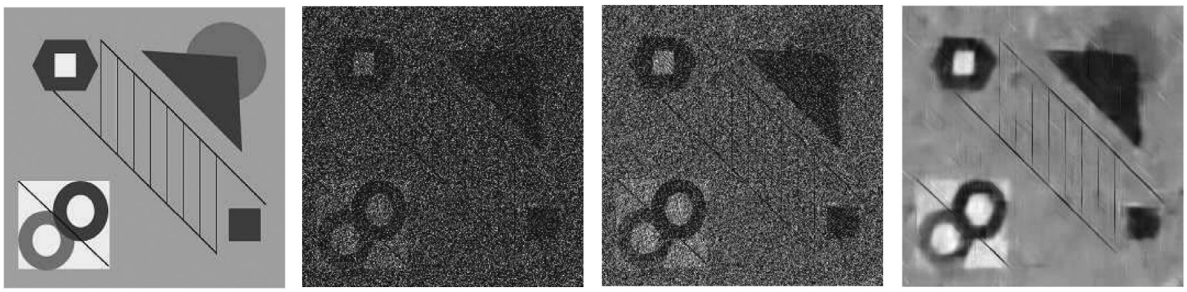

Fig. 9. Validation of the edge enhancement strategy in terms of PSNR values. (a) U (b) V ($\sigma = 100$) (c) Y (PSNR = 13.05) (d) $\hat{\mathbf{u}}$ (PSNR = 25.52).

artifacts in low-frequency regions as shown by other listed techniques. However, sharp lines, corners and edges are retained, from the coarse scale to the fine scale. Moreover, it generates fewer artifacts as compared to BPFA.

A comparison of the algorithms in terms of PSNR versus σ is presented in Fig. 7 and the quantitative results are shown in Table 3. It is interesting to remark that the BPFA method fails to produce good results when the noise level is high because of the erroneous overlapping of weak edges. Finally, constructing dictionaries on processed data is better than the direct application of BPFA, which is demonstrated in Fig. 8.

6.5. Validation of edge enhancement strategy

In order to elucidate the image de-noising and latent detail enhancement process at different stages of the proposed framework, a step-wise performance is evaluated in terms of PSNR and the de-noising results are presented in Fig. 9. Moreover, the validity of the edge enhancement strategy is ascertained in Fig. 10 using the Sobel edge filtering [2]. The results clearly show the appositeness of the proposed algorithm for preserving the texture and salient features during the de-noising process. In fact, it is established that the first step of the algorithm is felicitous to enhancing the salient feature, thereby improving the redundancy performance of the dictionary learning.

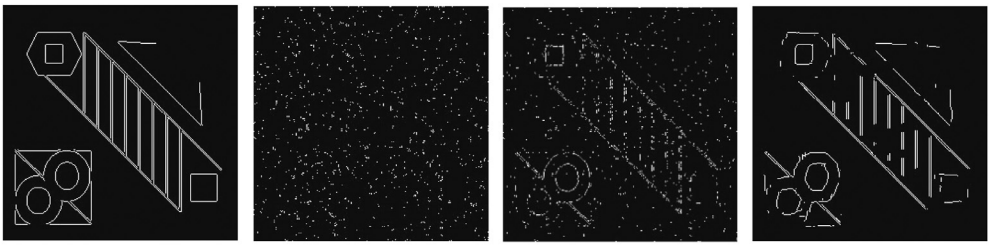

Fig. 10. (a) U (b) V (c) Y (d) $\hat{\mathbf{u}}$.

7. Conclusion

The proposed method develops the statistics of the de-noised image following the statistical regularities of the natural images. It offers several advantages not attained in earlier methodologies that have sought point estimates in general.

In this work, a hybrid two-stage Bayesian non-parametric algorithm is presented for image de-noising to handle substantially high noise while automatically inferring noise parameter and preserving latent detail information. At first stage, the detail features are enhanced using Bayesian non-local means filtering blended with a sub-band replacement and fusion process. Then, a dictionary learning approach based on beta process factor analysis model is used to remove idiosyncratic noise from refined image data.

The proposed framework benefits from non-local property, wavelet multi-resolution and beta process dictionary learning advantages. Its performance in terms of quantitative results and visual quality was compared with state-of-the-art alternatives. The experimental results substantiate that it performs better than the listed algorithms and maintains excellent visual quality for high noise levels, especially when the estimated noise parameter differs from the ground truth and the image contains rich texture. However, the computational speed of the proposed algorithm is relatively high due to its multistage paradigm wherein the emphases is on visual quality. Towards this end, a neural network based robust multistage framework with desired improvements will be the subject of a future study.

References

[1] Elad M, Aharon M. Image denoising via sparse and redundant representations over learned dictionaries. IEEE Trans Image Process 2006;15(12):3736–45.
[2] Muhammad N, Bibi N, Jahangir A, Mahmood Z. Image denoising with norm weighted fusion estimators. Pattern Anal Appl 2017:1–10.
[3] Buades A, Coll B, Morel JM. A review of image de-noising algorithms, with a new one. Multiscale Model Simul 2006;4(2):490–530.
[4] Kervrann C, Boulanger J, Coupé P. Bayesian non-local means filter, image redundancy and adaptive dictionaries for noise removal. In: Proceedings of the 1st international conference on scale space and variational methods in computer vision, SSVM 07, 30 May 2007–02 June 2007, Ishia, Italy. Berlin: Springer; 2007. p. 520–32.
[5] Zhou M, Chen H, Paisley J, Ren L, Li L, Xing Z, et al. Nonparametric bayesian dictionary learning for analysis of noisy and incomplete images. IEEE Trans Image Process 2012;21(1):130–44.
[6] Muhammad N, Nargis B, Zahid M, Tallha A, Syed RN. Reversible integer wavelet transform for blind image hiding method. PLoS ONE 2017a;12(5):1–17.
[7] Aharon M, Elad M, Bruckstein A. K-SVD: an algorithm for designing overcomplete dictionaries for sparse representation. IEEE Trans Signal Process 2006;54(11):4311–22.
[8] Zhou M, Chen H, Paisley J, Ren L, Sapiro G, Carin L. Non-parametric bayesian dictionary learning for sparse image representations. In: Advances in neural information processing systems 22, NIPS 2009, 7–10 December 2009, Vancouver, BC, Canada. New York: Curran Associates, Inc.; 2009. p. 2295–303.
[9] Dabov K, Foi A, Katkovnik V, Egiazarian K. Image de-noising by sparse 3-d transform-domain collaborative filtering. IEEE Trans Image Process 2007;16(8):2080–95.
[10] Mairal J, Bach F, Ponce J, Sapiro G, Zisserman A. Non-local sparse models for image restoration. In: IEEE 12th international conference on computer vision, 2009, 29 September 2009–02 October 2009, Kyoto, Japan. New York: IEEE; 2009. p. 2272–9.
[11] Chen Y., Pock T.. Trainable nonlinear reaction diffusion: a flexible framework for fast and effective image restoration. arXiv2013;[Preprint] 2013. Available from: https://arxiv.org/abs/1503.05768.
[12] Dong W, Zhang L, Shi G, Li X. Nonlocally centralized sparse representation for image restoration. IEEE Trans Image Process 2013;22(4):1620–30.
[13] Muhammad N, Nargis B. Digital image watermarking using partial pivoting lower and upper triangular decomposition into the wavelet domain. IET Image Proc 2015;9(9):795–803.
[14] Gu S, Zhang L, Zuo W, Feng X. Weighted nuclear norm minimization with application to image denoising. In: IEEE conference on computer vision and pattern recognition, CVPR 2014, 23–28 June 2014, Columbus, OH, USA. New York:IEEE; 2014. p. 2862–9.
[15] Bushra M, Muhammad N, Muhammad S, Tanzila S, Amjad R. Extraction of breast border and removal of pectoral muscle in wavelet domain. Biomed Res 2017;28(10):1–3.
[16] Muhammad N, Nargis B, Iqbal Q, Adnan J, Zahid M. Digital watermarking using hall property image decomposition method. Pattern Anal Appl 2017b;1(1):1–16.
[17] Xu J, Zhang L, Zuo W, Zhang D, Feng X. Patch group based nonlocal self-similarity prior learning for image denoising. In: IEEE international conference on computer vision, ICCV 2015, 7–13 December 2015, Santiago, Chili. New York: IEEE; 2015a. p. 244–52.
[18] Schmidt U, Roth S. Shrinkage fields for effective image restoration. In: IEEE conference on computer vision and pattern recognition, CVPR 2014, 23–28 June 2014, Columbus, OH, USA. New York: IEEE; 2014. p. 2862–9.
[19] Lebrun M, Colom M, Buades A, Morel JM. Secrets of image denoising cuisine. Acta Numerica 2012;21(1):475–576.
[20] Donoho DL, Johnstone IM. Adapting to unknown smoothness via wavelet shrinkage. J Am Stat Assoc 1995;90(432):1200–24.

[21] Louchet C, Moisan L. Total variation as a local filter. SIAM J Imaging Sci 2011;4(2):651–94.
[22] Lebrun M, Buades A, Morel JM. A nonlocal bayesian image denoising algorithm. SIAM J Imaging Sci 2013;6(3):1665–88.
[23] Yan R, Shao L, Liu Y. Nonlocal hierarchical dictionary learning using wavelets for image denoising. IEEE Trans Image Process 2013;22(12):4689–98.
[24] Mahmood Z, Muhammad N, Bibi N, Ali T. A review on state-of-the-art face recognition approaches. Fractals 2017;25(02):1750025.

Nazeer Muhammad received a degree in Applied Mathematics from Hanyang University, South Korea in 2015. He received the prestigious Pakistan Government higher education commission (HEC) scholarship award for MS and Ph.D. Currently, he is Assistant Professor at the Department of Mathematics, COMSATS Institute of Information Technology, Wah Cantt, Pakistan. His interests are digital signal and image processing.

Nargis Bibi received a Ph.D. degree in Computer Science the School of Computer Science, University of Manchester, UK in 2014. Currently, she is Assistant Professor at the Department of Computer Science, Fatima Jinnah Women University (FJWU), Rawalpindi, Pakistan. She is currently employed in FJWU as Assistant Professor. Her interests are digital signal processing, OFDM, coding theory and information theory.

Abdul Wahab received his Ph.D. in Applied Mathematics in 2011 from University Paris 7. Currently, he is Research Assistant Professor at Bio Imaging and Signal Processing Lab., Department of Bio and Brain Engineering, Korea Advanced Institute of Science and Technology, Korea.

Zahid Mahmood received a B.S. degree in Computer Engineering from CIIT Pakistan, in 2007 and MS degree in Electrical Engineering from South Korea in 2011, and a Ph.D. degree from NDSU, USA in 2015. Zahid Mahmoods research expertise encompasses in Image Processing and Signal Processing. He is the recipient of Government of Pakistan scholarship award for MS and Ph.D.

Tallha Akram received his MSc in Embedded Systems and Control Engineering in 2007, and Ph.D. in Computer Vision and Machine Learning in 2014 from University of Leicester, UK and Chongqing University, P.R. China respectively. He is currently working as Assistant Professor with the Department of Electrical Engineering at COMSATS Institute of Information Technology, Wah, Pakistan.

Syed Rameez Naqvi received his MSc in Electronic Engineering in 2007, and Ph.D. in Informatics in 2013 from University of Sheffield, UK and Vienna University of Technology, Austria respectively. He is currently working as Assistant Professor with the Department of Electrical Engineering at COMSATS Institute of Information Technology, Wah, Pakistan.

Hyun Sook Oh received her Ph.D. (1993) in Statistics from Purdue University. Currently, she is a professor in the department of Applied Statistics, Gachon University, Korea.

Dai-Gyoung Kim received his Ph.D. (1994) in Applied Mathematics from Purdue University, Post Doctor, IMA, University of Minnesota, 1995. Visiting Scholar, Image Processing Group, UCLA, 2003–2004. Currently, he is a professor in the department of Applied Mathematics, Hanyang University, Korea.