Image Noise: Detection, Measurement, and Removal Techniques

Zhifei Zhang
Department of Electrical Engineering and Computer Science
University of Tennessee, Knoxville, TN 37996, USA
zzhang61@vols.utk.edu

***Abstract*—The report surveys the evolution of image denoising techniques from the perspectives of detection, measurement, and removal. Since noise detection and measurement are intrinsically the same, we focus on the discussion of noise measurement and noise removal techniques. For noise measurement approaches, we survey filter-based, block-based, and wavelet-based, as well as other important achievements in the area. For noise removal approaches, we also survey from four stages in roughly chronological order from 1970s to the beginning of 2010s, i.e., spatial-domain, transform-domain, non-local, and recent achievements. Specifically, the recent approaches mainly consist of TV minimization, sparse coding, and deep learning.**

I. Introduction

As the first 100×100 CCD camera was invented in 1975, the study of digital image denoising started around the same time. Since then, the noise removal techniques have experienced prosperous development as CCD cameras are used widely in computer vision. There is a large amount of literature on image denoising. By contrast, the literature on noise detection and estimation is very limited [1]. Since noise measurement has implied the detection procedure, i.e., giving the noise level of a image implies whether there is noise in the image, we consider noise detection and measurement as the same process that is referred to as measurement. The scope of the report is to focus on noise measurement and removal techniques for natural images.

Since noise of a digital image is greatly related to the acquisition instrument, modeling the physical imaging process of a camera is an intuitive way to measure the noise level [2], [3]. In practice, however, noise modeling in images is also affected by data transmission media, discrete sources of radiation, etc. [4]. For simplicity, most of the natural images are simply assumed to be corrupted by additive random noise which is modeled as a zero-mean Gaussian distribution [5]–[7]. And most noise measurement works are filter-based or/and block-based approaches in either spatial or spectral domain. A brief review can be seen in [8] published in 2013.

Image denoising was first studied by Nasser Nahi at USC in early 1970s (though he used the name statistical image enhancement in his paper) [9]. In later 1970s, this problem was tackled by computer vision pioneers such as S. Zucker and A. Rosenfeld in their paper titled "Iterative enhancement of noisy images" [10]. In 1980, JS Lee published an important paper titled "Digital image enhancement and noise filtering by use of local statistics" [11]. Until then, most works applied statistical method in the spatial domain of the pixel array.

In late 1980s, the invention of wavelet transforms has led to dramatic progress in image denoising originated by D. Donoho [12] in 1995, E. Simoncelli and E. Adelson [13] in 1996. Wavelet transform gives a superior performance in image denoising due to properties such as sparsity and multiresolution structure. Since then, various wavelet-based image denoising algorithms were introduced. Ever since Donoho's wavelet-based soft-thresholding approach was published in 1995, researchers have published different algorithms to adaptively achieve the optimal threshold of wavelet coefficients [14], [15]. Probabilistic models [16] based on the statistical properties of the wavelet coefficients seems to outperform the thresholding techniques and gained ground like Gaussian scale mixture (GSM) [17]. The class of geometric wavelets such as curvelet transform [18] has also found promising application in image denoising.

In the recent decade, non-local denoising has been explored, and it potentially increases the performance of image denoising. [19] first presented the non-local means (NL-means) algorithm in 2005, which operated in the spatial domain. Unlike previous local smoothing filters, such as the Gaussian filtering [20], the anisotropic filtering [21], [22] and the neighborhood filtering [23], the NL-means not only compares the gray level in a single point but also the geometrical configuration in the whole neighborhood. Since then, the non-local method has been extended to the transformed domain and merged with other algorithms like sparse coding [24]. One of breakthroughs was made by the researchers from Finland in their block-matching 3D filtering (BM3D) framework [25] published in 2007. Then, its extensions, e.g., shape-adaptive BM3D [26] and BM3D-based deblurring [27] appeared in later years. Until now, the BM3D is still one of the most competitive filters.

In recent years, sparse coding becomes a hot topic in image denoising, where the learned simultaneous sparse coding (LSSC) [24] in 2009 and clustering-based sparse representation (CSR) [28] in 2011 are two milestones, whose denoising performance has shown convincing improvements over BM3D. With sparse coding gaining popularity in image denoising, related algorithms for dictionary learning and solving sparse problem are published, e.g., Bayesian dictionary learning [29], locally learned dictionaries [30] and the famous

K-SVD [31]. At the same time, total variation (TV) based algorithms [32] also receive enough attention, as well as deep learning methods [33].

This report will discuss the evolution of image noise measurement and removal techniques, respectively, in roughly chronological order. Due to the page limit, only the classic algorithms in each stage are described in detail. Those extensions will only be highlighted of their advantages and limitations. Section II briefly lists the algorithms of image noise measurement. Section III discusses the evolution of image denoising algorithms in roughly four stages: spatial-domain, transform-domain, non-local and recent denoising approaches.

II. Detection and Measurement of Image Noise

Noise level is an important parameter to many image processing applications such as denoising, segmentation, and so on. In 1993, S. Olsen [34] first gave a complete description and comparison of six early noise estimation algorithms. They are classified into two different approaches: filter-based (or smoothing-based) and block-based. In filter-based methods, the noisy image is first filtered by a low-pass filter to suppress the image noise. Then the noise variance is computed from the difference between the noisy image and the filtered image. In blocked-based methods, images are tessellated into a number of blocks. The noise variance is then computed from a set of homogeneous blocks. Since then, many different approaches were proposed for noise estimation. Basically they are still filter-based, bock-based, or the combination of both. The most common model for noise is the additive white Gaussian noise (AWGN). The goal of noise level estimation is to estimate the unknown standard deviation σ_n, given only a single observed noisy image. Section II-A and II-B discuss the filter-based and block-based approaches, respectively. After the invention of Wavelet transformation, noise estimation began to perform in the wavelet domain, as well as some other transformed domain, which is discussed in section II-C. Some other works that not belong to three mainstreams are discussed in section II-D.

A. Filter-based Approaches

In the filter-based approaches, there are two branches: 1) The processed image is convolved with a high-pass filter (e.g., Laplacian kernel [35]), so that the filtering result is assumed to contain only the noise, which allows direct estimation of the noise variance. 2) The difference of the raw image and the response of a low-pass filter (e.g., Gaussian kernel) is computed, and the difference is assumed to contain only the noise. A common step of the filter-based algorithms is the convolution between the raw image and a sliding mask (kernel), as shown in Fig. 1. This process is typically referred to as discrete convolution filtering or data masking (Gonzalez and Woods 1992) [36].

Later works mostly focused on how to design the kernel to achieve better and faster noise estimation. Two early kernels– the average and median kernel—were tested in [34], which

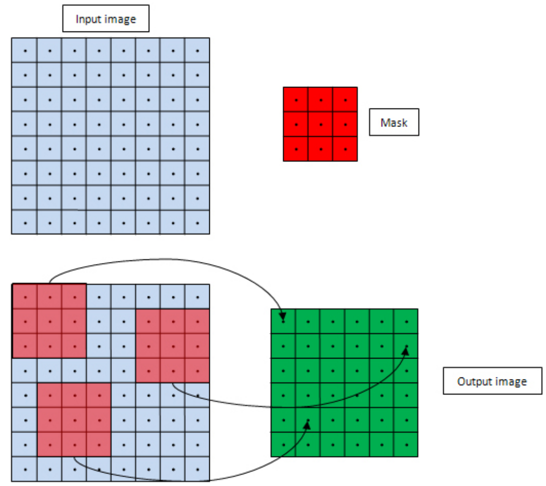

Fig. 1. A toy example of mask filtering. Convolution of a 8×8 image with a 3×3 filter to yield a 6×6 output image [37].

shows that these two kernel are unsuitable when there is fine texture in the image. [35] proposed the Laplacian kernel that is much faster to estimate the noise level because it directly extract high-frequency component (noise) and suppress the image structure. By the same token, [38], [39] applied a gradient related kernel. However, the edges or fine details are easy to be considered as noise by mistake. [36] employed Laplacian and gradient masks to remove edge structure from the noise. Similarity, [7] cooperated the Laplacian kernel with an edge detector. To further accurately extract noise, preprocessing on the raw image is performed to remove the influence of image structure like [40]. With the similar idea, [41] sharpened the raw image first to enhance the noise, and then the noise variance is estimated by analyzing the edge gradients after a smooth filter. The main difficulty of the filter-based approaches is that the filtered result is assumed to be the noise, but this assumption is not always true, especially for images with complex structure or fine details [42].

B. Block-based Approaches

In stead of involving all pixels to estimate the noise variance, block-based approaches decompose the image into blocks with small standard deviation and change of intensity. Such blocks are usually referred to as homogeneous patches/blocks. The intensity variation of a homogeneous patch is assumed to equal the noise variance.

In [43], the blocks whose standard deviations of intensity close to the minimum standard deviation among decomposed blocks are selected, and then the noise level is computed from the selected blocks. [44] determined homogeneous blocks based on high-pass filter and special masks for corners to reject blocks with structure first and then estimate the noise variance based on those remained blocks. [45] supported the block-based approach in a more theoretical manner. [46] applied the Sobel edge detector on blocks to exclude structures or details from contributing to the estimation of noise variance. [47]

endeavored to speed up the searching of those optimal homogeneous blocks through trimming estimation samples. These methods are effective, but they tend to overestimate the noise level for small noise level cases and make underestimation under large noise level.

A breakthrough was made in [8], which showed that the noise variance can be estimated as the smallest eigenvalue of the image block covariance matrix, and it does not assume the existence of homogeneous areas in the input image, hence it can successfully process images containing only textures. [42] followed up to apply PCA on low-rank blocks without high frequency components. Actually, the similar algorithm was published about 20 years earlier in [48] (1996), titled “Noise estimation and filtering using block-based singular value decomposition”. But it did not make a relative complete comparison with 13 existing methods like [8] (2013).

C. Wavelet-based Approaches

The wavelet-based approaches usually assume that the wavelet coefficients at the finest decomposition level (subband HH1) corresponds to the noise. The process of wavelet transform is illustrated in Fig. 2, and Fig. 3 gives an example of three layer decomposition using discrete wavelet transform (DWT).

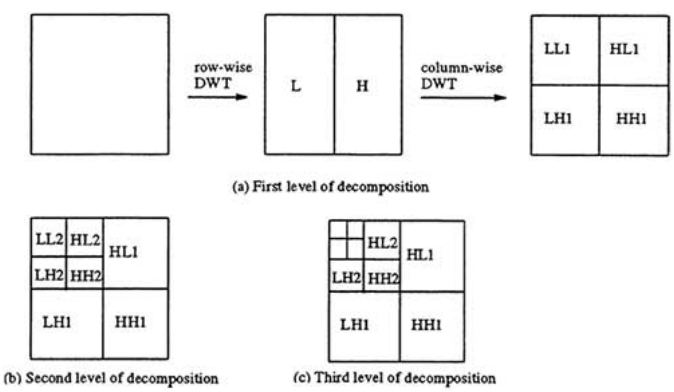

Fig. 2. Since two dimensional wavelet filters are separable functions, 2D DWT can be obtianed by first applying the 1D DWT row-wise (to produce L and H subbands in each row) and then column-wise as shown in (a). In the first level of decomposition, four subbands LL1, LH1, HL1 and HH1 are obtained. Repeating the same in the LL1 subband, it produces LL2, LH2, HL2 and HH2 and so on, as shown in (c). [49]

Since Donoho proposed the median absolute deviation (MAD) [5], [12] based on wavelet shrinkage (VisuShrink), his work was followed and improved by many researchers, most of whom focus on how to determine the threshold of wavelet coefficients to split noise. VisuShrink proposed a non-adaptive thresholding method, by contrast SureShrink [51] provided an adaptive one. [52] analyzed the noise variance on multiple subbands, rather than only on the HH1 subband. [13], [53] proposed the BayesShrink, an adaptive thresholding, which outperforms previous wavelet shrinkage methods. [6] proposed two alternative wavelet-based methods to estimate noise levels based on training samples. In the first method, the noise standard deviation estimation is computed as a linear

(a) Lena image (b) Three layer decomposition [50]

Fig. 3. The three layer decomposition of the ’Lena’ image using DWT

combination of normalized moments with learned coefficients. In the second method, the value of the cumulative distribution function (CDF) of local variances at a given point is computed for training images and stored in a lookup table against the noise variance. For a new image, the noise variance estimate is taken from the lookup table using the CDF value of this image.

However, such schemes tend to overestimate the noise standard deviation in applications where the SNR in the wavelet components is high because it assumes that the coefficients of the finest decomposition level is associated only to the noise. [54] proposed a residual autocorrelation power (RAP) approach based on BayesShrink, which does not require the above common assumption. The processed image is denoised by the BayesShrink algorithm using several values of the noise variance, and then the behavior of the residual autocorrelation in a range of noise variance values is analyzed in order to find the true noise variance. [55] cooperated block-based approach and MAD in the transformed domain—applying discrete cosine transform (DCT) on nonlocal similar blocks and then estimating the noise level through MAD.

D. Other Approaches

Some works estimate the noise level through fuzzy logic, modeling imaging process of the camera or noise statistics. Of course, the noise measurement algorithms do not limited to these mentioned ones. Here, only those representative methods are collected.

In camera modeling method, [2] gave an early attempt to build physical models for charged-coupled device (CCD) video cameras and material reflectance. It suggests that the variation in digitized pixel values is due to sensor noise and scene variation. Then, this model was simplified in [1]. [3] showed how to estimate an upper bound on the noise level from a single image based on a piecewise smooth image prior model and measured CCD camera response functions. It also suggests that noise level changes with brightness and uses Bayesian MAP inference to infer the noise level function.

An early literature using fuzzy method is [56], which requires a knowledge base contains a number of fuzzy sets decided by experts or derived from the histogram of a reference image. [57] applied the fussy system on chi-square statistics to decide the level of noise. [58] used fuzzy gradient values to determine if a certain pixel is corrupted with impulse noise or not.

For statistical methods, [59] modeled the image as a mixture of two Gaussian distributions, where the distributions are associated with the signal and noise, respectively. [60] suggested that the kurtosis of marginal bandpass filter response distributions should be constant for a noise-free image, which allows the construction of a kurtosis model for a noisy image, and the noise variance is assessed by finding the best parameters of this model. Last but not least, a Bayesian framework with a learned Markov random field prior (Fields of Experts [61]) for noise level estimation was proposed in [62].

III. Image Noise Removal Techniques

The ultimate goal of image noise measurement is alway to accurately remove the noise—denoising. Thus, most algorithms mentioned in section II can be directly applied in denoising. In this section, the image denoising algorithms will be arranged in the chronological order to illustrate the great evolution. Generally speaking, digital image denoising was performed in the spatial domain using statistical methods at the very beginning from the 1970s to 1980s (section III-A). As the invention of wavelet transform in late 1980s, a new era was started to dramatically improve the performance of image denoising, which will be discussed in section III-B. Until the year of 2005 when the non-local method was first proposed, the image noise removal technique entered another stage as illustrated in section III-C. Section III-D talks about some other original approaches, namely variational formulations, sparse coding and deep learning, which are mostly proposed in the last decade. An overview of the evolution of image denoising techniques is shown in Table I .

A. Spacial Approaches

Around 1980, those early ways of removing noise from digital image are to employ spatial filters—a sliding window performing as low pass filtering on groups of pixels with the assumption that the noise occupies the higher region of frequency spectrum. The Wiener filter [63] is a linear filter that is defined in terms of the signal (image) and noise autocorrelation functions. It seriously depends on the prior knowledge of the spectral properties of the original signal. [11] proposed the basic average filter that tends to blur the image. [10] presented a weighted average filter to overcome this drawback. The weights are determined depending on the presence of any edges or lines passing through the window. For the same purpose, a variety of weighted median filters [69], [70] were developed after the original median filter [64].

In the later decade, the development of spatial approach mainly lies in how to utilize the neighborhood. [20] filtered the image with a isotropic Gaussian kernel. By contrast, an anisotropic filters was proposed in [21], [22], which attempts to avoid the blurring effect of the Gaussian by convolving the image only in the direction orthogonal to the gradient. [67] proposed a steerable filter that changes the filtering direction according to the structure of neighborhood. [68] proposed a rank selection filter, which can be seen as a generalized median filter. It selects a value that is not necessary the median value as the output of the filter according to the feature within the sliding widow. [116] cooperates the median filter with a salt–pepper impulse noise detector. The bilateral filter was proposed in [71], which replaces the intensity value at each pixel in an image by a weighted average of intensity values from nearby pixels. The weights depend not only on Euclidean distance of pixels, but also on the range differences. This method effectively preserves sharp edges.

In the recent decade, it is more popular to consider statistical property of the neighborhood to suppress noise and keep details at the same time. [87] utilized the correlations between a pixel and its neighbors and derives the upper and lower bound of the homogeneity level that is defined for pixel values based on their global and local statistical properties. [88] introduced a local image statistic for identifying noise pixels in images corrupted with impulse noise of random values. The statistical values quantify how different in intensity the particular pixels are from their most similar neighbors. [89] performed an asymptotic analysis of neighborhood filters as the size of the neighborhood shrinks to zero. Usually, the development direction of spatial approach is how to design the sliding window to achieve better performance. [90] proposed an adaptive kernel regression method, in which the kernel can be Gaussian whose covariance matrix vary according to the condition of patches, e.g., texture, corner, edge or flat.

B. Transform-domain Approaches

The transformed domain mainly refers to the wavelet domain since most filtering algorithms are performed in the wavelet domain through the discrete wavelet transform (DWT). Of course, there also some works perform in other domains. [65] discussed a filter operated in the frequency domain using fast Fourier transform (FFT) with an adaptive cut-off frequency. [66] combined frequency and spatial domain information. Such method is restricted due to its limitations in providing sparse representation of data. Therefore, the frequency-domain approach is not considered as a main method in image denoising. In this section, we'll mainly discuss the algorithms in the wavelet domain.

The pioneer Donoho published the wavelet-based denoising method VisuShrink [12] in 1994, which is a non-adaptive thresholding method. The next year, he published a similar method but with soft thresholding (adaptive threshold) [5]. In [51], the SureShrink with adaptive threshold was proposed. Since then, the focus of image denoising was shifted from spatial and Fourier domain to the wavelet transform domain. And most literature focus on different ways to compute the parameters for the thresholding of wavelet coefficients. [72] improved the SureShrink by directly parameterizing the denoising process as a sum of elementary nonlinear processes with unknown weights. The thresholding/cut-off in the wavelet domain may cause artifacts such as pseudo-Gibbs phenomenon. [73] proposed the translation-invariant denoising to avoid such arti-

TABLE I
EVOLUTION OF IMAGE DENOISING TECHNIQUES

Time	spatial-domain	Transform-domain	Non-local	Others
1970s & 1980s	Wiener [63] Average [11] Weighted average [10] Median [64]	FFT [65] spatial-frequency [66]		
1990s	Steerable filter [67] Anisotropic filter [21], [22] Gaussian filter [20] Rank selection [68] Weighted median [69], [70] Bilateral filter [71]	DWT-VisuShrink [5], [12] DWT-SureShrink [51], [72] UDWT [73], [74] DWT-BayesShrink [13] Multiwavelet [75] DWT-MRF [76], [77] DWT-HMM [78] SB-TS [79], [80] DWT-Wiener [81], [82] SIWPD [83] CWT [84] ICA [85] Wedgelet transform [86]		
2000s	Local statistic [87], [88] Adaptive neighborhood [89] Kernel regression [90]	spatial-DWT [15] GSM [91] BLS-GSM [17] DWT-uHMT [92] Curvelet transform [18] Contourlet transform [93] DWT-BivariateShrink [94], [95] Bandelet transform [96] SA-DCT [97] FFT-DWT [98], [99] DWT-NeuralNetwork [100] DWT-trivariateShrink [101] Neighborhood DWT [99]	NL-means (spatial) [19] NL-means (Wavelet) [102] BM3D [25], [103] SA-BM3D [104] BM3D-SAPCA [26]	TV minimization [32] Sparse coding (SC) [31] K-SVD [105] K-LLD [30] LSSC [24]
2010s		PCA [106] DWT-TV [107]	BM4D [108]–[110]	CSR [28] SDAE [33] SSDA [113] MLP [114] AMC-SSDA [115]

facts. [74] proposed a similar work—shift invariant DWT, and uses non-orthogonal wavelet transform named undecimated discrete wavelet transform (UDWT). BayesShrink [13], [53], [117] minimizes the Bayes risk estimator function assuming generalized Gaussian prior and thus yields data adaptive threshold. BayesShrink outperforms SureShrink most of the times. [75] presented the multiwavelet—using multiple mother wavelet function. It performs better but increases the computation complexity.

The wavelet-based approaches are not limited to adaptive linear or non-linear thresholding, it is fast developed to statistical modeling or more novel ways to decompose the coefficients. [76], [77] modeled the wavelet coefficients using the Markov random field (MRF) which is efficient to capture intra-scale correlations. [78] developed a new framework for statistical signal processing based on wavelet-domain hidden Markov models (HMMs) that concisely models the statistical dependencies and non-Gaussian statistics encountered in real-world signals. Compared with MRF, the HMM models are efficient in capturing inter-scale dependencies. [79], [80] inspired a hierarchical interpretation of the wavelet decomposition based on the idea that wavelet coefficients of single will have large wavelet coefficients that persist along the branches of tree, otherwise the branches is noise and can be trimmed. This is the original work of optimal subband tree structuring (SB-TS). Unlike DWT, the coefficients are decomposed to create the full binary tree in SB-TS. During the same time, [83] proposed the shift invariant wavelet packet decomposition (SIWPD), which combine shift invariant wavelet transform and SB-TS. In, [16] the wavelet coefficients were modeled with zero-mean Gaussian random variables with high local correlation. The Gaussian scale mixture (GSM) was proposed in [91], which describes the joint densities of clusters of wavelet coefficients as a jointly Gaussian vector multiplied by a hidden scaling variable. By the same token, [17] proposed a model (BLS-GSM) in which neighborhoods of wavelet coefficients at adjacent positions and scales are described as GSM which is a product of a Gaussian random vector and an independent hidden random scalar multiplier. The latter modulates the local variance of the coefficients in the neighborhood, and is thus able to account for the empirically observed correlation between the coefficient amplitudes. [92] proposed the uHMT, a simplified hidden Markov tree (HMT), to release the computational burden of the training stage of HMM. [94] broke the independent assumption of wavelet coefficients and proposes a heavy-tailed bivariate PDF to model the statistics of wavelet coefficients. A simple nonlinear threshold function (shrinkage function) is derived from the

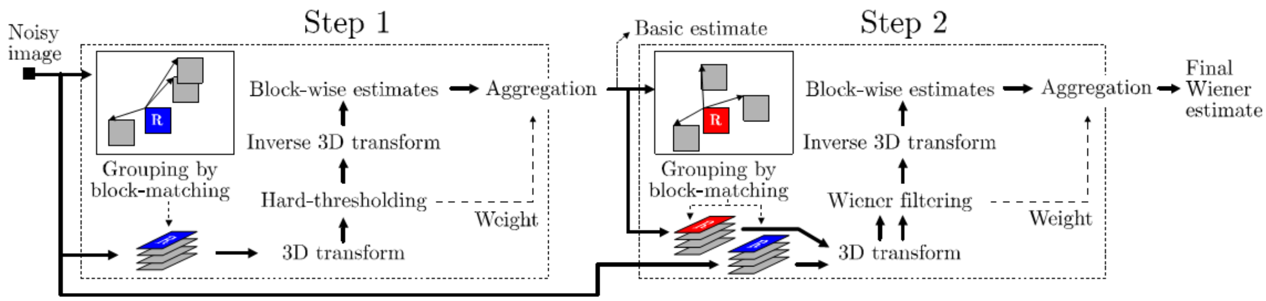

Fig. 4. Flowchart of the proposed image denoising algorithm. The operations surrounded by dashed lines are repeated for each processed block (marked with "R"). [25].

PDF using Bayesian estimation theory. [95] presented a locally adaptive denoising algorithm using the bivariate shrinkage function.

After the middle of 2000s, researchers tend to hybridize existing method to achieve better denoising performance. [15] presented an spatial adaptive wavelet thresholding, which selects threshold based on local spatial information (e.g., smooth or edge regions). [97] proposed the shape-adaptive discrete cosine transform (SA-DCT) transform, which defines the shape of the transforms support in a pointwise adaptive manner. The thresholded or attenuated SA-DCT coefficients are used to reconstruct a local estimate of the signal within the adaptive-shape support. Since supports corresponding to different points are in general overlapping, the local estimates are averaged together using adaptive weights that depend on the regions statistics. Transform-domain image denoising methods assume that the original signal can be sparsely represented in the transform domain, but none of the orthogonal transforms can achieve sparse representation for all images. [98] proposed a hybrid Fourier-wavelet denoising method to overcome this shortcoming. [100] introduced the subband-adaptive thresholding neural network to improve the efficiency of the denoising procedure. [101] incorporated a wavelet-based trivariate shrinkage filter with a spatial-based joint bilateral filter. A hybrid Fourier and neighborhood wavelet coefficients method was proposed in [99]. [107] combined wavelet-domain sparsity and total variation (TV) regularization, which will be discussed in section III-D.

Some other transform-domain methods are also investigated. [18] proposed the curvelet transform that cooperates Fourier and Radon transform with the wavelet transform to better recover the edges after denoising. Based on the similar idea, [93], [96] and [86] proposed contourlet, bandelet and wedgelet transform, representatively. Two representative works on statistical transform are ICA [85] and PCA [106] bases. The former shows that independent component analysis (ICA) is effective for sparse representation of natural-image patches and hence for image denoising, and the latter presents an image denoising scheme by using principal component analysis (PCA) with local pixel grouping (LPG). The PCA method yields exciting performance.

C. Non-local Approaches

A nonlocal filter exploits similarities between image blocks (patches) from various spatial locations, hence the name nonlocal. To the author's knowledge, the non-local means (NL-means) proposed by Buades [19] (2005) used for the first time the term nonlocal in the context of image denoising. The NL-means estimates a pixel as the weighted average of pixels with weights that depend solely on the similarity between neighborhoods centered at these pixels and the neighborhood centered at the estimated pixel. Fig. 5 illustrates the scheme of NL-means. It processes in the spatial domain.

Fig. 5. Scheme of NL-means strategy. Similar pixel neighborhoods give a large weight, w(p,q1) and w(p,q2), while much different neighborhoods give a small weight w(p,q3) [19].

In 2006, [102] extended the nonlocal spatial filter by Buades et al., associating with each pixel the weighted sum of data points within an adaptive neighborhood. It is competitive with the best transform-domain filters. Thus arises the question whether sparse transform-domain representations and nonlocal modeling can be combined so that the strengths of both techniques are preserved. The researcher from Finland, K. Dabov, showed that the answer to this question is positive [103], and he first named such algorithm as BM3D that is short for block-matching and 3-D filtering in [25] (2007). The good denoising result of the BM3D filter inspired applications of this denoising scheme to other image processing applications. The original BM3D scheme is shown in Fig. 4.

Since the invention of BM3D, most import developments based on the BM3D were proposed by the group involving K. Dabov [118] from Tampere University of Technology. They proposed extensions of the BM3D filter to shape adaptive (SA-BM3D) [104] and to shape-adaptive PCA representations (BM3D-SAPCA) [26]. A breakthrough in image deblurring occurred in [27] which also exploits the BM3D filter in iterative variational minimization with a prior on sparsity. [119] combines BM3D with transform-domain alpha-rooting in order to simultaneously sharpen and denoise the image. Then, the BM3D-based framework was extended to video or series frame denoising—BM4D [108]–[110].

D. Other Approaches

Beside above approaches, some other originals algorithm can also achieve the state-of-art performance like variational formulations, sparse coding and deep learning. These methods are mostly becoming hot topics in the recent decade.

1) Variational formulations: The original work of variational formulation was introduced by Rudin, Osher and Fatemi (ROF) in [120] as a regularization approach capable of handling properly edges and removing noise in a given image. [32] employed this method in image denoising. In later works, this method is always called total variation (TV) minimization. Usually, it has two terms—data fidelity term and regularization term which penalizes high frequency noise. Eq. 1 expresses the general formula of TV minimization [121].

$$\min_{\mathbf{u}} \ell(\mathbf{Ku}, \mathbf{b}) + \lambda\Omega(\nabla_x \mathbf{u}, \nabla_y \mathbf{u}) \tag{1}$$

where $\mathbf{u}$ is the unknown noise-free image and $\mathbf{K}$ represents a linear operator. $\ell(\mathbf{Ku}, \mathbf{b})$ measures the data fidelity between $\mathbf{Ku}$ and the observation $\mathbf{b}$. $\nabla_x \mathbf{u}$ and $\nabla_y \mathbf{u}$ compute the discrete gradients of the image $\mathbf{u}$ along the x-axis and y-axis, respectively. $\Omega(\nabla_x \mathbf{u}, \nabla_y \mathbf{u})$ is the regularizer on $\nabla_x \mathbf{u}$ and $\nabla_y \mathbf{u}$, and λ is a positive parameter used to balance the two terms for minimization. Most related works modify either the data fidelity term or the regularization term. A brief review is shown in Table II and III [121], representatively.

TABLE II
DATA FIDELITY TERMS

$\ell(\mathbf{Ku}, \mathbf{b}) =$	Noise type
$\\|\mathbf{Ku} - \mathbf{b}\\|_2^2$	Gaussian [120]
$\langle \mathbf{Ku} - \mathbf{b} \cdot \log(\mathbf{u}), \mathbf{1} \rangle$	Multi-Poisson [122]
$\langle \log(\mathbf{Ku}) + \mathbf{b}/(\mathbf{Ku}), \mathbf{1} \rangle$	Multi-Gamma [123]
$\\|\mathbf{Ku} - \mathbf{b}\\|_1$	Laplace [124]
$\\|\mathbf{Ku} - \mathbf{b}\\|_\infty$	Uniform [125]
$\\|\mathbf{Ku} - \mathbf{b}\\|_0$	Impulse [121], [126]

TABLE III
REGULARIZATION TERMS

$\Omega(\mathbf{g}, \mathbf{h}) =$	Description
$\sum_{i=1}^{n} (\mathbf{g}_i^2 + \mathbf{h}_i^2)^{\frac{1}{2}}$	TV_2 (isotropic) [120]
$\sum_{i=1}^{n} \|\mathbf{g}_i\| + \|\mathbf{h}_i\|$	TV_1 (anisotropic) [124], [127]
$\sum_{i=1}^{n} \|\mathbf{g}_i\|_0 + \|\mathbf{h}_i\|_0$	TV_0 [128], [129]

Most TV related works focus on global restoration of the image. However, [130] employed adaptive median filter to identify pixels which are likely to be contaminated by noise, and then regularization method is applied only to those selected noise candidates. Very recently, the regularization term TV_0, which is based on the ℓ_0-norm, has received much attention. It has been shown to be particularly effective for image smoothing [128] and motion deblurring [129].

2) Sparse coding: In 2006, M. Elad proposed the initial sparse coding (SC) method for image denoising [31], in which the common expression of sparse coding is shown in Eq. 2.

$$\hat{\boldsymbol{\alpha}} = \arg\min_{\boldsymbol{\alpha}} \|\mathbf{D}\boldsymbol{\alpha} - \mathbf{y}\|_2^2 + \lambda\|\boldsymbol{\alpha}\|_0 \tag{2}$$

where considers image patches of size $\sqrt{n} \times \sqrt{n}$ pixels, vectorized as column vectors $\mathbf{y} \in \mathbb{R}^n$. The dictionary $\mathbf{D} \in \mathbb{R}^{n \times k}$ is learned from training patches through K-SVD [105]. $\boldsymbol{\alpha} \in \mathbb{R}^k$ denotes the sparse coefficient vector which is supposed to be sparse. It looks similar with the TV-based method. The significant difference between them is the second (regularization) term. TV minimization aims at removing high frequency part, while sparse coding tends to represent the raw image with a few atoms from the dictionary. Note that the ℓ_0-norm alway cause the optimization function unsolvable, so it is usually instituted by ℓ_1-norm.

Roughly, SC-based methods focus on two aspects: 1) dictionary learning and 2) sparse coding. To speed up the process of dictionary learning, [30] combined non-local method into dictionary learning and proposes the clustering-based method which makes use of locally learned dictionaries (K-LLD). [24] borrowed the idea of BM3D to construct the dictionary and uses grouped sparse coding. The author named such method as learned simultaneous sparse coding (LSSC). For different applications or better performance, later works added some other penalty term to the end of the original objective function (Eq. 2). [28] showed that the location uncertainty of nonzero sparse coefficients is often related to the nonlocal self similarity of image signals, and it propose the cluster-based sparse representation (CSR) to achieving higher sparsity by exploring the location related constraint. [112] (2014) employed the proximal method [111] to solve the ℓ_0-norm based dictionary learning, which used to be considered as unsolvable. There are many variants from the original SC, as well as different ways of solving the objective function. They cannot be exhausted in this report, but they all share the similar idea—representing an image patch as a linear combination of a few atoms chosen out from an over-complete dictionary.

3) Deep learning: [33] presented stacked denoising auto-encoders (SDAE) which is an original strategy for image denoising using deep learning in 2010. [113] combined sparse coding and SDAE to form stacked sparse denoising auto-encoders (SSDA). Multi layer perceptron (MLP) was applied in [114]. These methods are not robust to variation in noise types beyond what it has seen during training. To address this limitation, [115] presented the adaptive multi-column stacked sparse denoising autoencoder (AMC-SSDA). To my knowledge, there is not strong evidence to demonstrate that the deep learning based approaches perform better than BM3D or sparse coding.

IV. Conclusion

The main development direction of image noise measurement and removal techniques is from spatial to transformed domain, from local to non-local statistics, and from thresholding to optimization. From recent literature, the best performance is achieved through grouped sparse coding in the transformed domain (CSR) and shape-adaptive BM3D (BM3D-SAPCA). Actually, sparse coding and BM3D share the similar idea that the coefficients with relatively large value corresponds to the pure signal and such coefficients should present sparsity. No matter how denoising algorithm evolves, the ultimate goal of denoising has never changed—smoothing the image and preserving the details.

References

[1] Y. Tsin, V. Ramesh, and T. Kanade, "Statistical calibration of CCD imaging process," in *IEEE International Conference on Computer Vision*, vol. 1. IEEE, 2001, pp. 480–487.

[2] G. E. Healey and R. Kondepudy, "Radiometric CCD camera calibration and noise estimation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 16, no. 3, pp. 267–276, 1994.

[3] C. Liu, W. T. Freeman, R. Szeliski, and S. B. Kang, "Noise estimation from a single image," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 1. IEEE, 2006, pp. 901–908.

[4] M. C. Motwani, M. C. Gadiya, R. C. Motwani, and F. C. Harris, "Survey of image denoising techniques," in *Proceedings of GSPX*, 2004, pp. 27–30.

[5] D. L. Donoho, "De-noising by soft-thresholding," *IEEE Transactions on Information Theory*, vol. 41, no. 3, pp. 613–627, 1995.

[6] A. De Stefano, P. White, and W. Collis, "Training methods for image noise level estimation on wavelet components," *EURASIP Journal on Applied Signal Processing*, vol. 2004, pp. 2400–2407, 2004.

[7] S.-C. Tai and S.-M. Yang, "A fast method for image noise estimation using Laplacian operator and adaptive edge detection," in *3rd International Symposium on Communications, Control and Signal Processing*. IEEE, 2008, pp. 1077–1081.

[8] S. Pyatykh, J. Hesser, and L. Zheng, "Image noise level estimation by principal component analysis," *IEEE Transactions on Image Processing*, vol. 22, no. 2, pp. 687–699, 2013.

[9] X. Li. Reproducible research in computational science. [Online]. Available: http://www.csee.wvu.edu/~xinl/source.html

[10] A. Lev, S. W. Zucker, and A. Rosenfeld, "Iterative enhancemnent of noisy images," *IEEE Transactions on Systems, Man and Cybernetics*, vol. 7, no. 6, pp. 435–442, 1977.

[11] J.-S. Lee, "Digital image enhancement and noise filtering by use of local statistics," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, no. 2, pp. 165–168, 1980.

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

[13] E. P. Simoncelli and E. H. Adelson, "Noise removal via Bayesian wavelet coring," in *International Conference on Image Processing*, vol. 1. IEEE, 1996, pp. 379–382.

[14] S. G. Chang, B. Yu, and M. Vetterli, "Adaptive wavelet thresholding for image denoising and compression," *IEEE Transactions on Image Processing*, vol. 9, no. 9, pp. 1532–1546, 2000.

[15] ——, "Spatially adaptive wavelet thresholding with context modeling for image denoising," *IEEE Transactions on Image Processing*, vol. 9, no. 9, pp. 1522–1531, 2000.

[16] M. K. Mihcak, 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] 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.

[18] J.-L. Starck, E. J. Candès, and D. L. Donoho, "The curvelet transform for image denoising," *IEEE Transactions on Image Processing*, vol. 11, no. 6, pp. 670–684, 2002.

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

[20] M. Lindenbaum, M. Fischer, and A. Bruckstein, "On Gabor's contribution to image enhancement," *Pattern Recognition*, vol. 27, no. 1, pp. 1–8, 1994.

[21] L. Alvarez, P.-L. Lions, and J.-M. Morel, "Image selective smoothing and edge detection by nonlinear diffusion. II," *SIAM Journal on numerical analysis*, vol. 29, no. 3, pp. 845–866, 1992.

[22] 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.

[23] L. P. Yaroslavsky and L. Yaroslavskij, "Digital picture processing. An introduction." *Springer Seriesin Information Sciences, ISBN 0-387-11934-5 (USA).*, vol. 1, 1985.

[24] J. Mairal, F. Bach, J. Ponce, G. Sapiro, and A. Zisserman, "Non-local sparse models for image restoration," in *IEEE 12th International Conference on Computer Vision*. IEEE, 2009, pp. 2272–2279.

[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] ——, "BM3D image denoising with shape-adaptive principal component analysis," in *SPARS'09-Signal Processing with Adaptive Sparse Structured Representations*, 2009.

[27] A. Danielyan, V. Katkovnik, and K. Egiazarian, "BM3D frames and variational image deblurring," *IEEE Transactions on Image Processing*, vol. 21, no. 4, pp. 1715–1728, 2012.

[28] W. Dong, X. Li, L. Zhang, and G. Shi, "Sparsity-based image denoising via dictionary learning and structural clustering," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, 2011, pp. 457–464.

[29] M. Zhou, H. Chen, J. Paisley, L. Ren, L. Li, Z. Xing, D. Dunson, G. Sapiro, and L. Carin, "Nonparametric Bayesian dictionary learning for analysis of noisy and incomplete images," *IEEE Transactions on Image Processing*, vol. 21, no. 1, pp. 130–144, 2012.

[30] P. Chatterjee and P. Milanfar, "Clustering-based denoising with locally learned dictionaries," *IEEE Transactions on Image Processing*, vol. 18, no. 7, pp. 1438–1451, 2009.

[31] 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.

[32] S. Kindermann, S. Osher, and P. W. Jones, "Deblurring and denoising of images by nonlocal functionals," *Multiscale Modeling & Simulation*, vol. 4, no. 4, pp. 1091–1115, 2005.

[33] P. Vincent, H. Larochelle, I. Lajoie, Y. Bengio, and P.-A. Manzagol, "Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion," *The Journal of Machine Learning Research*, vol. 11, pp. 3371–3408, 2010.

[34] S. I. Olsen, "Estimation of noise in images: an evaluation," *Graphical Models and Image Processing*, vol. 55, no. 4, pp. 319–323, 1993.

[35] J. Immerkaer, "Fast noise variance estimation," *Computer Vision and Image Understanding*, vol. 64, no. 2, pp. 300–302, 1996.

[36] B. Corner, R. Narayanan, and S. Reichenbach, "Noise estimation in remote sensing imagery using data masking," *International Journal of Remote Sensing*, vol. 24, no. 4, pp. 689–702, 2003.

[37] Image convolution using OpenCL. [Online]. Available: http://developer.amd.com/tools-and-sdks/opencl-zone/opencl-resources/programming-in-opencl/image-convolution-using-opencl/

[38] K. Rank, M. Lendl, and R. Unbehauen, "Estimation of image noise variance," *IEE Proceedings-Vision, Image and Signal Processing*, vol. 146, no. 2, pp. 80–84, 1999.

[39] R. Bilcu and M. Vehvilainen, "New method for noise estimation in images," in *Nonlinear Signal and Image Processing*. IEEE, 2005, p. 25.

[40] A. Liu, "A fast method of estimating Gaussian noise," in *ICISE*. IEEE, 1899, pp. 441–444.

[41] F. Russo, "Gaussian noise estimation in digital images using nonlinear sharpening and genetic optimization," in *Instrumentation and Measurement Technology Conference Proceedings*. IEEE, 2007, pp. 1–5.

[42] X. Liu, M. Tanaka, and M. Okutomi, "Single-image noise level estimation for blind denoising," *IEEE Transactions on Image Processing*, vol. 22, no. 12, pp. 5226–5237, 2013.

[43] D.-H. Shin, R.-H. Park, S. Yang, and J.-H. Jung, "Block-based noise estimation using adaptive Gaussian filtering," *IEEE Transactions on Consumer Electronics*, vol. 51, no. 1, pp. 218–226, 2005.

[44] A. Amer and E. Dubois, "Fast and reliable structure-oriented video noise estimation," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 15, no. 1, pp. 113–118, 2005.

[45] S. Aja-Fernández, G. Vegas-Sánchez-Ferrero, M. Martín-Fernández, and C. Alberola-López, "Automatic noise estimation in images using local statistics. Additive and multiplicative cases," *Image and Vision Computing*, vol. 27, no. 6, pp. 756–770, 2009.

[46] S.-M. Yang and S.-C. Tai, "Fast and reliable image-noise estimation using a hybrid approach," *Journal of Electronic Imaging*, vol. 19, no. 3, pp. 033 007–033 007, 2010.

[47] M. Ghazal and A. Amer, "Homogeneity localization using particle filters with application to noise estimation," *IEEE Transactions on Image Processing*, vol. 20, no. 7, pp. 1788–1796, 2011.

[48] K. Konstantinides, B. Natarajan, and G. S. Yovanof, "Noise estimation and filtering using block-based singular value decomposition." *IEEE transactions on image processing: a publication of the IEEE Signal Processing Society*, vol. 6, no. 3, pp. 479–483, 1996.

[49] P. P. I. P. Ltd. Implementation of lifting DWT technique. [Online]. Available: https://www.pantechsolutions.net/spartan-3-edk-sourcecode/implementation-of-lifting-dwt-technique

[50] S. Kumar. Classifying image data. [Online]. Available: http://www.debugmode.com/imagecmp/classify.htm

[51] D. L. Donoho and I. M. Johnstone, "Adapting to unknown smoothness via wavelet shrinkage," *Journal of the american statistical association*, vol. 90, no. 432, pp. 1200–1224, 1995.

[52] J.-L. Starck and F. Murtagh, "Automatic noise estimation from the multiresolution support," *Publications of the Astronomical Society of the Pacific*, vol. 110, no. 744, pp. 193–199, 1998.

[53] H. A. Chipman, E. D. Kolaczyk, and R. E. McCulloch, "Adaptive Bayesian wavelet shrinkage," *Journal of the American Statistical Association*, vol. 92, no. 440, pp. 1413–1421, 1997.

[54] M. Hashemi and S. Beheshti, "Adaptive noise variance estimation in BayesShrink," *IEEE Signal Processing Letters*, vol. 1, no. 17, pp. 12–15, 2010.

[55] A. Danielyan and A. Foi, "Noise variance estimation in nonlocal transform domain," in *International Workshop on Local and Non-Local Approximation in Image Processing*. IEEE, 2009, pp. 41–45.

[56] C.-S. Lee, Y.-H. Kuo, and P.-T. Yu, "Weighted fuzzy mean filters for image processing," *Fuzzy Sets and Systems*, vol. 89, no. 2, pp. 157–180, 1997.

[57] M. Salmeri, A. Mencattini, E. Ricci, and A. Salsano, "Noise estimation in digital images using fuzzy processing," in *International Conference on Image Processing*, vol. 1. IEEE, 2001, pp. 517–520.

[58] S. Schulte, M. Nachtegael, V. De Witte, D. Van der Weken, and E. E. Kerre, "A fuzzy impulse noise detection and reduction method," *IEEE Transactions on Image Processing*, vol. 15, no. 5, pp. 1153–1162, 2006.

[59] P. Wyatt and H. Nakai, "Developing nonstationary noise estimation for application in edge and corner detection," *IEEE Transactions on Image Processing*, vol. 16, no. 7, pp. 1840–1853, 2007.

[60] D. Zoran and Y. Weiss, "Scale invariance and noise in natural images," in *IEEE 12th International Conference on Computer Vision*. IEEE, 2009, pp. 2209–2216.

[61] S. Roth and M. J. Black, "Fields of experts," *International Journal of Computer Vision*, vol. 82, no. 2, pp. 205–229, 2009.

[62] U. Schmidt, K. Schelten, and S. Roth, "Bayesian deblurring with integrated noise estimation," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, 2011, pp. 2625–2632.

[63] W. K. Pratt, "Generalized Wiener filtering computation techniques," *IEEE Transactions on Computers*, vol. 100, no. 7, pp. 636–641, 1972.

[64] T. S. Huang, G. J. Yang, and G. Y. Tang, "A fast two-dimensional median filtering algorithm," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 27, no. 1, pp. 13–18, 1979.

[65] A. K. Jain, *Fundamentals of digital image processing*. Prentice-Hall, Inc., 1989.

[66] A. N. Hirani and T. Totsuka, "Combining frequency and spatial domain information for fast interactive image noise removal," in *Proceedings of the 23rd annual conference on Computer graphics and interactive techniques*. ACM, 1996, pp. 269–276.

[67] W. T. Freeman and E. H. Adelson, "The design and use of steerable filters," *IEEE Transactions on Pattern Analysis & Machine Intelligence*, no. 9, pp. 891–906, 1991.

[68] R. C. Hardie and K. E. Barner, "Rank conditioned rank selection filters for signal restoration," *IEEE Transactions on Image Processing*, vol. 3, no. 2, pp. 192–206, 1994.

[69] R. Yang, L. Yin, M. Gabbouj, J. Astola, and Y. Neuvo, "Optimal weighted median filtering under structural constraints," *IEEE Transactions on Signal Processing*, vol. 43, no. 3, pp. 591–604, 1995.

[70] A. B. Hamza, P. L. Luque-Escamilla, J. Martínez-Aroza, and R. Román-Roldán, "Removing noise and preserving details with relaxed median filters," *Journal of Mathematical Imaging and Vision*, vol. 11, no. 2, pp. 161–177, 1999.

[71] C. Tomasi and R. Manduchi, "Bilateral filtering for gray and color images," in *Sixth International Conference on Computer Vision*. IEEE, 1998, pp. 839–846.

[72] F. Luisier, T. Blu, and M. Unser, "A new SURE approach to image denoising: Interscale orthonormal wavelet thresholding," *IEEE Transactions on Image Processing*, vol. 16, no. 3, pp. 593–606, 2007.

[73] R. R. Coifman and D. L. Donoho, *Translation-invariant de-noising*. Springer, 1995.

[74] M. Lang, H. Guo, J. E. Odegard, C. S. Burrus, and R. O. Wells Jr, "Nonlinear processing of a shift-invariant discrete wavelet transform (DWT) for noise reduction," in *SPIE's 1995 Symposium on OE/Aerospace Sensing and Dual Use Photonics*. International Society for Optics and Photonics, 1995, pp. 640–651.

[75] T. D. Bui and G. Chen, "Translation-invariant denoising using multiwavelets," *IEEE Transactions on Signal Processing*, vol. 46, no. 12, pp. 3414–3420, 1998.

[76] M. Malfait and D. Roose, "Wavelet-based image denoising using a Markov random field a priori model," *IEEE Transactions on Image Processing*, vol. 6, no. 4, pp. 549–565, 1997.

[77] M. Jansen and A. Bultheel, "Empirical Bayes approach to improve wavelet thresholding for image noise reduction," *Journal of the American Statistical Association*, vol. 96, no. 454, 2001.

[78] M. S. Crouse, R. D. Nowak, and R. G. Baraniuk, "Wavelet-based statistical signal processing using hidden Markov models," *IEEE Transactions on Signal Processing*, vol. 46, no. 4, pp. 886–902, 1998.

[79] J. Lu, J. B. Weaver, D. M. Healy Jr, and Y. Xu, "Noise reduction with a multiscale edge representation and perceptual criteria," in *Proceedings of the IEEE-SP International Symposium on Time-Frequency and Time-Scale Analysis*. IEEE, 1992, pp. 555–558.

[80] R. G. Baraniuk, "Optimal tree approximation with wavelets," in *SPIE's International Symposium on Optical Science, Engineering, and Instrumentation*. International Society for Optics and Photonics, 1999, pp. 196–207.

[81] H. Choi and R. Baraniuk, "Analysis of wavelet-domain Wiener filters," in *Proceedings of the IEEE-SP International Symposium on Time-Frequency and Time-Scale Analysis*. IEEE, 1998, pp. 613–616.

[82] V. Strela, "Denoising via block Wiener filtering in wavelet domain," *Progress in Mathematics-Boston-*, vol. 202, pp. 619–626, 2001.

[83] I. Cohen, S. Raz, and D. Malah, "Translation-invariant denoising using the minimum description length criterion," *Signal Processing*, vol. 75, no. 3, pp. 201–223, 1999.

[84] N. Kingsbury, "Image processing with complex wavelets," *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, vol. 357, no. 1760, pp. 2543–2560, 1999.

[85] P. Hoyer *et al.*, "Independent component analysis in image denoising," *Master's Thesis, Helsinki University of Technology*, 1999.

[86] D. L. Donoho *et al.*, "Wedgelets: Nearly minimax estimation of edges," *The Annals of Statistics*, vol. 27, no. 3, pp. 859–897, 1999.

[87] G. Pok, J.-C. Liu, and A. S. Nair, "Selective removal of impulse noise based on homogeneity level information," *IEEE Transactions on Image Processing*, vol. 12, no. 1, pp. 85–92, 2003.

[88] R. Garnett, T. Huegerich, C. Chui, and W. He, "A universal noise removal algorithm with an impulse detector," *IEEE Transactions on Image Processing*, vol. 14, no. 11, pp. 1747–1754, 2005.

[89] A. Buades, B. Coll, and J.-M. Morel, "Neighborhood filters and PDEs," *Numerische Mathematik*, vol. 105, no. 1, pp. 1–34, 2006.

[90] H. Takeda, S. Farsiu, and P. Milanfar, "Kernel regression for image processing and reconstruction," *IEEE Transactions on Image Processing*, vol. 16, no. 2, pp. 349–366, 2007.

[91] V. Strela, J. Portilla, and E. P. Simoncelli, "Image denoising using a local Gaussian scale mixture model in the wavelet domain," in *International Symposium on Optical Science and Technology*. International Society for Optics and Photonics, 2000, pp. 363–371.

[92] J. K. Romberg, H. Choi, and R. G. Baraniuk, "Bayesian tree-structured image modeling using wavelet-domain hidden Markov models," *IEEE Transactions on Image Processing*, vol. 10, no. 7, pp. 1056–1068, 2001.

[93] M. N. Do and M. Vetterli, "Contourlets: A directional multiresolution image representation," in *International Conference on Image Processing*, vol. 1. IEEE, 2002, pp. I–357.

[94] L. Sendur and I. W. Selesnick, "A bivariate shrinkage function for wavelet-based denoising," in *IEEE International Conference on Acoustics Speech and Signal Processing*, vol. 2. Citeseer, 2002, pp. II–1261.

[95] L. Şendur and I. W. Selesnick, "Bivariate shrinkage with local variance estimation," *IEEE Signal Processing Letters*, vol. 9, no. 12, pp. 438–441, 2002.

[96] E. Le Pennec and S. Mallat, "Sparse geometric image representations with bandelets," *IEEE Transactions on Image Processing*, vol. 14, no. 4, pp. 423–438, 2005.

[97] 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.

[98] S. Jiang and X. Hao, "Hybrid Fourier-wavelet image denoising," *Electronics Letters*, vol. 43, no. 20, pp. 1081–1082, 2007.

[99] J. Cheng and S. Lei, "Image denoising based on hybrid fourier and neighborhood wavelet coefficients," in *2nd International Conference on Electronic & Mechanical Engineering and Information Technology*. Atlantis Press, 2012.

[100] M. Nasri and H. Nezamabadi-pour, "Image denoising in the wavelet domain using a new adaptive thresholding function," *Neurocomputing*, vol. 72, no. 4, pp. 1012–1025, 2009.

[101] H. Yu, L. Zhao, and H. Wang, "Image denoising using trivariate shrinkage filter in the wavelet domain and joint bilateral filter in the spatial domain," *IEEE Transactions on Image Processing*, vol. 18, no. 10, pp. 2364–2369, 2009.

[102] C. Kervrann and J. Boulanger, "Optimal spatial adaptation for patch-based image denoising," *IEEE Transactions on Image Processing*, vol. 15, no. 10, pp. 2866–2878, 2006.

[103] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, "Image denoising with block-matching and 3D filtering," in *Electronic Imaging*. International Society for Optics and Photonics, 2006, pp. 606 414–606 414.

[104] ——, "A nonlocal and shape-adaptive transform-domain collaborative filtering," in *Proc. Int. Workshop on Local and Non-Local Approx. in Image Process*, 2008.

[105] M. Aharon, M. Elad, and A. Bruckstein, "K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation," *IEEE Transactions on Signal Processing*, vol. 54, no. 11, pp. 4311–4322, 2006.

[106] 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.

[107] Y. Ding and I. W. Selesnick, "Artifact-free wavelet denoising: non-convex sparse regularization, convex optimization," *IEEE Signal Processing Letters*, vol. 22, no. 9, pp. 1364–1368, 2015.

[108] M. Maggioni, G. Boracchi, A. Foi, and K. Egiazarian, "Video denoising using separable 4D nonlocal spatiotemporal transforms," in *IS&T/SPIE Electronic Imaging*. International Society for Optics and Photonics, 2011, pp. 787 003–787 003.

[109] M. Maggioni and A. Foi, "Nonlocal transform-domain denoising of volumetric data with groupwise adaptive variance estimation," in *IS&T/SPIE Electronic Imaging*. International Society for Optics and Photonics, 2012, pp. 82 960O–82 960O.

[110] M. Maggioni, V. Katkovnik, K. Egiazarian, and A. Foi, "Nonlocal transform-domain filter for volumetric data denoising and reconstruction," *IEEE Transactions on Image Processing*, vol. 22, no. 1, pp. 119–133, 2013.

[111] J. Bolte, S. Sabach, and M. Teboulle, "Proximal alternating linearized minimization for nonconvex and nonsmooth problems," *Mathematical Programming*, vol. 146, no. 1-2, pp. 459–494, 2014.

[112] C. Bao, H. Ji, Y. Quan, and Z. Shen, "ℓ_0 norm based dictionary learning by proximal methods with global convergence," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, 2014, pp. 3858–3865.

[113] J. Xie, L. Xu, and E. Chen, "Image denoising and inpainting with deep neural networks," in *Advances in Neural Information Processing Systems*, 2012, pp. 341–349.

[114] H. C. Burger, C. J. Schuler, and S. Harmeling, "Image denoising: Can plain neural networks compete with bm3d?" in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, 2012, pp. 2392–2399.

[115] F. Agostinelli, M. R. Anderson, and H. Lee, "Adaptive multi-column deep neural networks with application to robust image denoising," in *Advances in Neural Information Processing Systems*, 2013, pp. 1493–1501.

[116] Z. Wang and D. Zhang, "Progressive switching median filter for the removal of impulse noise from highly corrupted images," *IEEE Transactions on Circuits and Systems II: Analog and Digital Signal Processing*, vol. 46, no. 1, pp. 78–80, 1999.

[117] F. Abramovich, T. Sapatinas, and B. W. Silverman, "Wavelet thresholding via a Bayesian approach," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 60, no. 4, pp. 725–749, 1998.

[118] M. Maggioni, E. Snchez-Monge, A. Foi, A. Danielyan, K. Dabov, V. Katkovnik, and K. Egiazarian. Image and video denoising by sparse 3D transform-domain collaborative filtering—Block-matching and 3D filtering (BM3D) algorithm and its extensions. [Online]. Available: http://www.cs.tut.fi/~foi/GCF-BM3D/

[119] M. Mäkitalo and A. Foi, "Spatially adaptive alpha-rooting in BM3D sharpening," in *IS&T/SPIE Electronic Imaging*. International Society for Optics and Photonics, 2011, pp. 787 012–787 012.

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

[121] G. Yuan and B. Ghanem, "ℓ_0TV: A new method for image restoration in the presence of impulse noise," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, 2015, pp. 2272–2279.

[122] T. Le, R. Chartrand, and T. J. Asaki, "A variational approach to reconstructing images corrupted by poisson noise," *Journal of Mathematical Imaging and Vision*, vol. 27, no. 3, pp. 257–263, 2007.

[123] G. Aubert and J.-F. Aujol, "A variational approach to removing multiplicative noise," *SIAM Journal on Applied Mathematics*, vol. 68, no. 4, pp. 925–946, 2008.

[124] J. Yang, Y. Zhang, and W. Yin, "An efficient TVL1 algorithm for deblurring multichannel images corrupted by impulsive noise," *SIAM Journal on Scientific Computing*, vol. 31, no. 4, pp. 2842–2865, 2009.

[125] C. Clason, "L^∞ fitting for inverse problems with uniform noise," *Inverse Problems*, vol. 28, no. 10, p. 104007, 2012.

[126] M. Nikolova, "A variational approach to remove outliers and impulse noise," *Journal of Mathematical Imaging and Vision*, vol. 20, no. 1-2, pp. 99–120, 2004.

[127] Y. Wang, J. Yang, W. Yin, and Y. Zhang, "A new alternating minimization algorithm for total variation image reconstruction," *SIAM Journal on Imaging Sciences*, vol. 1, no. 3, pp. 248–272, 2008.

[128] L. Xu, C. Lu, Y. Xu, and J. Jia, "Image smoothing via l_0 gradient minimization," in *ACM Transactions on Graphics (TOG)*, vol. 30, no. 6. ACM, 2011, p. 174.

[129] L. Xu, S. Zheng, and J. Jia, "Unnatural l_0 sparse representation for natural image deblurring," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, 2013, pp. 1107–1114.

[130] R. H. Chan, C.-W. Ho, and M. Nikolova, "Salt-and-pepper noise removal by median-type noise detectors and detail-preserving regularization," *IEEE Transactions on Image Processing*, vol. 14, no. 10, pp. 1479–1485, 2005.