Image Denoising by Wavelet Bayesian Network based on MAP Estimation

BHANUMATHI.V[1] and LAVANYA.S[2]
Department of Electronics and Communication Engineering,
Anna University, Regional Campus Coimbatore,
Coimbatore, Tamilnadu, INDIA
vbhanu_02@yahoo.com

Abstract: The objective of the work is to denoise the image and to provide better Peak Signal to Noise Ratio (PSNR) with edge preservation by using the hidden Bayesian network constructed from the wavelet coefficients. A Bayesian network which is also called as a directed acyclic graph is a graphical model with a set of conditional probabilities. Each node in the graph represents a random variable which is used to denote an attribute, feature and hypothesis. Bayesian network is constructed to model the priori probability of the original image for the image denoising problem, which involves removing white and homogeneous Gaussian noise with zero mean and known variance from an image. Two Maximum A Posteriori (MAP) techniques are used such as Bivariate Cauchy MAP (BCMAP) and Multivariate Cauchy MAP (MCMAP). From the simulation analysis, it is very clear that for various noise levels, the wavelet Bayesian network based on MAP estimation provides better PSNR value by preserving edges compared with the existing methods. For Lena image with noise variance of 15, the percentage increase in PSNR values are 2.08%, 4.16% and 7.38% for wavelet Bayesian, BCMAP and MCMAP compared with Bayesian Least Square Gaussian Scale Mixture (BLS-GSM) and for the same, the percentage increase in PSNR are 0.12%, 2.15% and 5.32% compared with Block Matching and 3-D filtering (BM3D).

Key-Words: - Denoising, Bayesian network, Wavelet coefficients, MAP, bivariate and multivariate.

1 Introduction

Visual information transmitted in the form of digital images is becoming a major method of communication in the modern age, but the image obtained after transmission is often corrupted with noise. The received image needs processing before it can be used in applications. Image denoising is nothing but reproducing an image with high quality by manipulating the image data. Different noise models including additive and multiplicative types are used such as Gaussian, salt and pepper, speckle and Brownian noise. Therefore, prior knowledge about the noise present in the image is needed in order to choose the appropriate denoising algorithm. The filtering approach has been proved to be the best when the image is corrupted with salt and pepper noise. It is also known from the literature that wavelet based methods are found to be the best for denoising the images that are corrupted with Gaussian noise.

The primary objective of the work is to construct a Bayesian network to model the priori probability of the original image for the image denoising problem, which removes white and homogeneous Gaussian noise with zero mean and known variance from an image. Bayesian networks [1-4] provide a means of parsimoniously expressing joint probability distributions over many interrelated hypotheses. These find place in many applications like medical sciences, image processing, economics, industrial and environmental engineering [5, 6]. A Bayesian network [7] consists of a Directed Acyclic Graph (DAG) and a set of local distributions. The graph and the local distributions together represent a joint distribution over the random variables denoted by the nodes of the graph.

In the modern image capturing devices, image denoising is particularly a serious problem because of the low signal-to noise (SNR). This is due to increase in the sensor's density per unit area of a chip. It also increases the capturing device's sensitivity to noise [8]. Most of the denoising algorithms follow Non-Local Means (NLM) approach [9, 10] exploiting the self-similarity and redundancy in an image.

A hidden markov model framework for statistical signal processing based on wavelet-domain (HMM's) is detailed in [11]. It is shown that this model paved a way to concisely model the non-Gaussian statistics of individual wavelet coefficients and also to capture statistical dependencies between coefficients. A statistical approach to wavelet based signal processing is adopted in which the signal and its wavelet coefficients are regarded as the random realizations from a family or distribution of signals. The coefficients have been modelled either as jointly Gaussian or as non-Gaussian but independent. However, Gaussian models are in conflict with the compression property, which implies that the wavelet transforms of most signals are sparse, resulting in a large number of small coefficients and a small number of large coefficients. Non-Gaussian models have also been formulated, but usually, the coefficients are assumed to be statistically independent of each other. The HMT model matches both the clustering and persistence properties of the wavelet transform.

In the existing Block Matching and 3-D filtering (BM3D) algorithm [10], 3-D groups are formed by combining similar 2-D patches that can be overlapped. Then by using, collaborative 3-D filtering technique, the system does the non-local filtering. In this, first the filtered blocks are sent back to their original positions, and then the weighted average of the estimates of the pixel in several different blocks is found out to estimate the pixel value. It is shown that it is better than the regular non-local means approach.

Plenty of denoising methods [12] exist, originating from various disciplines such as probability theory, statistics, and partial differential equations, linear and nonlinear filtering, spectral and multi resolution analysis. All these methods rely on some explicit or implicit assumptions about the true (noise free) signal in order to separate it properly from the random noise. In particular, the transform-domain denoising methods typically assume that the true signal can be well approximated by a linear combination of few basis elements. That is, the signal is sparsely represented in the transform domain. Hence, by preserving the few high-magnitude transform coefficients that convey mostly the true-signal energy and discarding the rest which are mainly due to noise, the true signal can be effectively estimated. The multi resolution transforms can achieve good sparsity for spatially localized details, such as edges and singularities.

The image modelling and estimation algorithm in [13] is a novel approach to non-local adaptive nonparametric filtering. It can be adapted to various noise models such as additive colored noise, non-Gaussian noise, etc., by modifying the calculation of coefficients variances in the basic and Wiener parts of the algorithm. It is a patch-based Wiener filter that exhibits the patch redundancy for image denoising. It utilizes both the shape wise characters and its perceived brightness to the human eye for the similar patches to estimate the different filter parameters. This method is comparable with nonlocal sparse model [14] and BM3D [10], even outperforming them in many cases where images exhibit higher levels of redundancy. NLM is a zeroth order kernel regression method, with a very specific choice of kernel. The existing is extended to higher orders of regression in order to approximate the image data locally by a polynomial or other localized basis of a given order. This provides better denoising in texture regions because of these extra degrees of freedom. Compared to the zeroth order method, higher order NLM shows better denoising in texture regions. Images captured by modern cameras are invariably corrupted by noise [12]. This denoising approach does not require parameter tuning and is practical, with the added benefit of a clean statistical motivation and analytical formulation. In a more practical setting where signal-dependent noise is observed, the clustering step needs to take into account color information as well. The noise in each cluster can be then assumed to be homogeneous, and the proposed filter can be independently applied in each cluster.

The Bayesian Least Square method based Gaussian Scale Mixture (BLS-GSM) given in [15] considers the distribution of wavelet coefficients in a 3×3 region, along with the center coefficient and it models this in the same orientation as a GSM. After the wavelet coefficient distribution, the center of the neighborhood system is found out using least square method based on Bayesian. In [16], an adaptive model which can make the Gaussian Scale Mixtures suitable for the image content called as Mixture GSM (MGSM) is detailed. Its working is similar to that of the BM3D algorithm.

The performance of the proposed is compared with that of other approaches like Block Matching 3D (BM3D), Bayes Least Square- Gaussian Scale

Mixture (BLS-GSM) and Wavelet Bayesian Network (WBN) and the demonstration yields a better Peak-Signal-to-Noise Ratio (PSNR) as well as better perceptual quality on the textured areas of an image. The proposed work is based on [17], which utilizes a hidden directed graph to model the prior probability of an image. The problems stated in this paper for the construction are as follows,

1) The coefficient and wavelet patch association problem, which involves associating subband coefficients with a wavelet patch.
2) The graph selection problem, i.e., determining the type of graph to construct.

To solve the above said problems, the heuristic procedure which is followed in the existing method is utilized. Assume that the wavelet patch is a matrix of $m \times m$ random variables. Let the size of a subband be $N \times N$ and let m divide N. The subband is partitioned into $(N/m)^2$ rectangular blocks, each of which contains $m \times m$ coefficients. Then, the coefficient at location (i, j) in each block is assigned as a realization of the random variable at location (i, j) in the wavelet patch. Thus, each random variable has $(N/m)^2$ sampled observations. For the second problem, the computational cost of a graph structure is analyzed for which the Maximum A Posteriori (MAP) solution can be derived efficiently by the standard Belief Propagation (BP) algorithm.

The standard implementation of the message passing algorithm in BP on $m \times m$ cliques runs in $O(N^2 k^{m\times m} T)$, where N^2 is the number of coefficients in a subband, k is the number of labels for each coefficient, and T is the number of iterations. Basically, computing each message takes $O(k^{m\times m})$ time and there are $O(N^2)$ messages per iteration.

The joint probability of a spanning tree $G = (V, E)$ can be formulated by the dependency structure in G as follows. Let $f(v_i \mid u_i)$ be the probability function associated with arc $u_i \rightarrow v_i$, where $u_i, v_j \in V$, and let u be the root of the tree with probability $f(u)$. The remainder of the paper comprise of the following. Second section gives an explanation of the construction of wavelet Bayesian networks. The third section details about the proposed methodology for image denoising based on MAP estimation. The fourth section analyzes the results. The fifth section concludes the proposed work.

2 Construction of Wavelet Bayesian Networks

Bayesian networks represent uncertainty using probabilities. Assigning a probability to an event gives an indication of how strongly we believe that the event will occur. The networks are used to compare information about a situation and make inferences. A network can provide information about the possible consequences of a situation, but can also provide information about the likely causes. An example of the Bayesian network is shown in Fig.1.

A Bayesian network, denoted as $B = (V, E, P)$, comprises a set of random variables and their conditional dependencies represented by a directed acyclic graph in which the nodes represent the elements in V. V represents the vertex set, E represents the edge set and P represents the probability model. Each edge element in E takes the form of a directed arc $x \rightarrow y$, where x and $y \in V$. The likelihood $p(y \mid x) \in P$ of an edge $x \rightarrow y \in E$ is the conditional probability of observing y given that x exists.

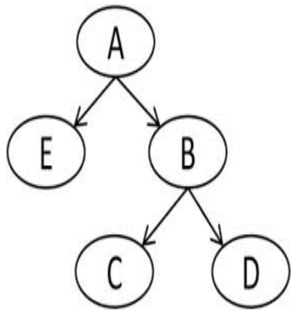

Figure. 1. Bayesian network - an example

One of the most important features is the fact that they provide an elegant mathematical structure for modeling complicated relationships among random variables while keeping a relatively simple visualization of these relationships. Initially, wavelet decomposition of an image F yields three images of wavelet coefficients with horizontal, vertical, and diagonal orientations respectively, and one approximate image of F. Let $W^h_j F(u, v)$,$W^v_j F(u, v)$and $W^d_j F(u, v)$ denote, respectively, the horizontal, vertical, and diagonal images of the wavelet coefficients at scale 2^j ; and let $A_j F(u, v)$ represent the approximated image at the same scale. If the undecimated DWT is decomposed J times, the wavelet coefficients $W^h_j F$, $W^v_j F$, and $W^d_j F$ are obtained with $j = 1, \ldots, J$ and the coarsest approximate image $A_J F$.

To construct a WBN, the subbands are grouped with the same orientation together to obtain a horizontal-group, a vertical-group, and a diagonal-group of wavelet coefficients. Then, we construct a Bayesian network B for each group. Let $B^h = (V^h, E^h, P^h)$, $B^v = (V^v, E^v, P^v)$, and $B^h = (V^d, E^d, P^d)$ denote the Bayesian networks constructed from the horizontal-group, vertical-group, and diagonal-group of wavelet coefficients respectively. The WBN $B = (V, E, P)$ is derived from B^h, B^v, and B^d by

$$V = V^h \cup V^v \cup V^d \quad (1)$$

$$E = E^h \cup E^v \cup E^d \quad (2)$$

$$P = P^h \cup P^v \cup P^d \quad (3)$$

There are two types of arcs in a Bayesian network B^u: 1) The inter-scale parent-child arc, which connects a node with its coarser-scale parent; and 2) the intra-scale sibling arc, which connects two nodes of the same scale. To obtain the probability inference, the probability function on each arc should be modelled. Then, the joint probability of the parent-child arc in $E^u{}_o$ is modelled as

$$f_o^u(x|x_p) = N(0; 2\omega x_p^2) \quad (4)$$

In order to obtain much better denoising and to preserve the texture the Bivariate Cauchy (BC) MAP and Multivariate Cauchy (MC) MAP estimation is introduced in the proposed work. It is seen that the MC distribution shown in [18, 19] is giving proven results in capturing the inter-channel dependencies by restoring the pure coefficients. It is also shown that it is possible with the help of a closed-form shrinkage function.

3. Proposed MAP Estimation for Image Denoising

The input image is decomposed by using the wavelet decomposition and the corresponding horizontal vertical, diagonal, and approximation component are obtained. To construct the graph, a matrix of random variables is created from wavelet coefficients. A wavelet patch is formed from the coefficients. Each subband consists of variable nodes of specified length. Each random variable in the wavelet patch is associated with the variable node in the graph. Edge set represents the arcs that are used to connect the nodes.

There are two types of arcs-interscale arcs and intrascale arcs. The interscale arcs connect the vertices of different scales. The intrascale arcs connect the vertices of same scales. Recovering the Maximum A Posteriori (MAP) configuration of random variables in a graphical model is an important problem. The flow chart shown in Fig. 2 shows the detailed step by step procedure for the implementation of the proposed work.

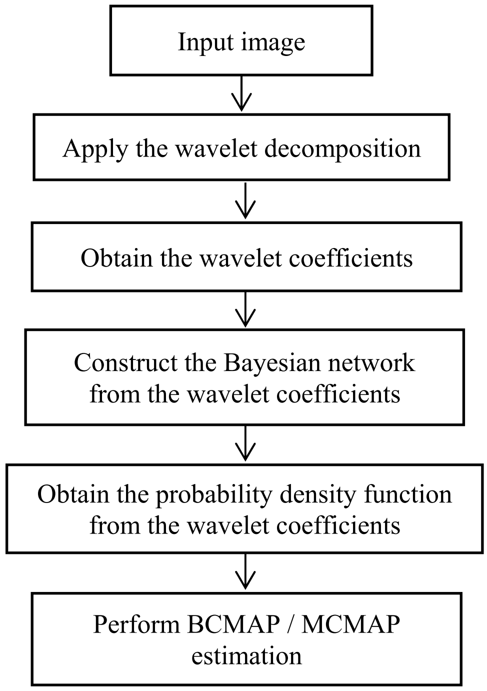

Figure. 2. Flow Chart showing the construction of WBN based on MAP estimation

After the construction of the wavelet coefficients, Bivariate Cauchy MAP (BCMAP) and then Multivariate Cauchy MAP (MCMAP) distributions are applied. The Bishrink Filter is a MAP filter that takes into account the interscale dependency of the wavelet coefficients considering the bivariate probability density function. This estimator requires prior knowledge of the noise variance and of the marginal variance of the noise-free image for each wavelet coefficient. To estimate the noise variance from the noisy wavelet coefficients, a robust median estimator from the finest scale wavelet coefficients obtained by applying the Bayes theorem is used. To make this estimation, σ2y represents the marginal variance of noisy observations $y1$ and $y2$. For the estimating the marginal variance of noisy observations, the local standard deviation of the useful component of the parent coefficients σ^2 in a given subband is interpolated by repeating each line and column.

The main drawback of Bishrink filter is that when the value of the estimation of the noise standard deviation is higher then, the performance of is poorer. Another very important parameter of the bishrink filter is the local estimation of the marginal variance of the noise-free image σ^n. The sensitivity of the estimation w^1 is a decreasing function of σ^n resulting in a reduction in the precision of the estimation on the use of the bishrink filter. It is noted that the bishrink filter treats the edges very well, the estimation of the textured regions must be corrected and the worst treatment corresponds to the homogeneous regions. The denoising quality of pixels with slightly different σ will be very different in the homogeneous regions. The sensitivity increases with the increasing of σ^n. So, the degradation of the homogeneous and textured zones of the noise-free image is amplified by an increase in the noise standard deviation. Secondly, the local variance of a pixel gives some information about the frequency content of the region to which the considered pixel belongs. If the pixels of a given region have low local variances then the considered region contains low frequencies and vice versa. The probability density function of the MC distribution given in [19(17)] is utilized and it is as follows,

$$P_X(X;\Sigma) = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{1}{2}\right)\pi^{0.5}|\Sigma|^{0.5}\left[1+X^T\Sigma^{-1}X\right]^{\frac{n+1}{2}}} \tag{5}$$

In the above equation, the term ‘n’ denotes dimensionality of the distribution (number channels) and ‘∑’ is used to represent the covariance matrix.

MC MAP estimation is to find the parameter values in a probabilistic model that best explain the data. The likelihood gives an explanation of the data in terms of the parameters. In a more comprehensive Bayesian approach, the goal is to characterize the full posterior distribution and not to simply find the mode of this distribution. The following are the steps for the proposed denoising algorithm based on MAP estimation.

Steps for WBN Denoising Algorithm with MAP Estimation

1) Calculate the wavelet transform of the noisy image and obtain the horizontal, vertical, and diagonal subbands.
2) Create the Vertex Set V: For each subband create the variable nodes for the wavelet coefficients according to the wavelet patch.
3) Create the Edge Set E: The inter scale and intra scale components are created to connect the nodes.
4) Create the WBN by connecting the vertices through the edges.
5) Reconstruct the denoised image from the wavelet coefficients.
6) Obtain the pixel value and calculate the probability density function.
7) Perform the BCMAP and MCMAP estimation using bayes theorem.

4. Results and Discussion

The main goal of this paper is to improve the PSNR value and also to maintain the texture of the input images. The input image is shown in Fig.3. For decomposing the original input image into different sub bands, DWT is applied. Then the process of interpolation is carried out for the high frequency subbands and input image. With the help of inverse DWT (IDWT), the subbands are combined to form a denoised image.

Figure. 3. Original input image of size 256x256

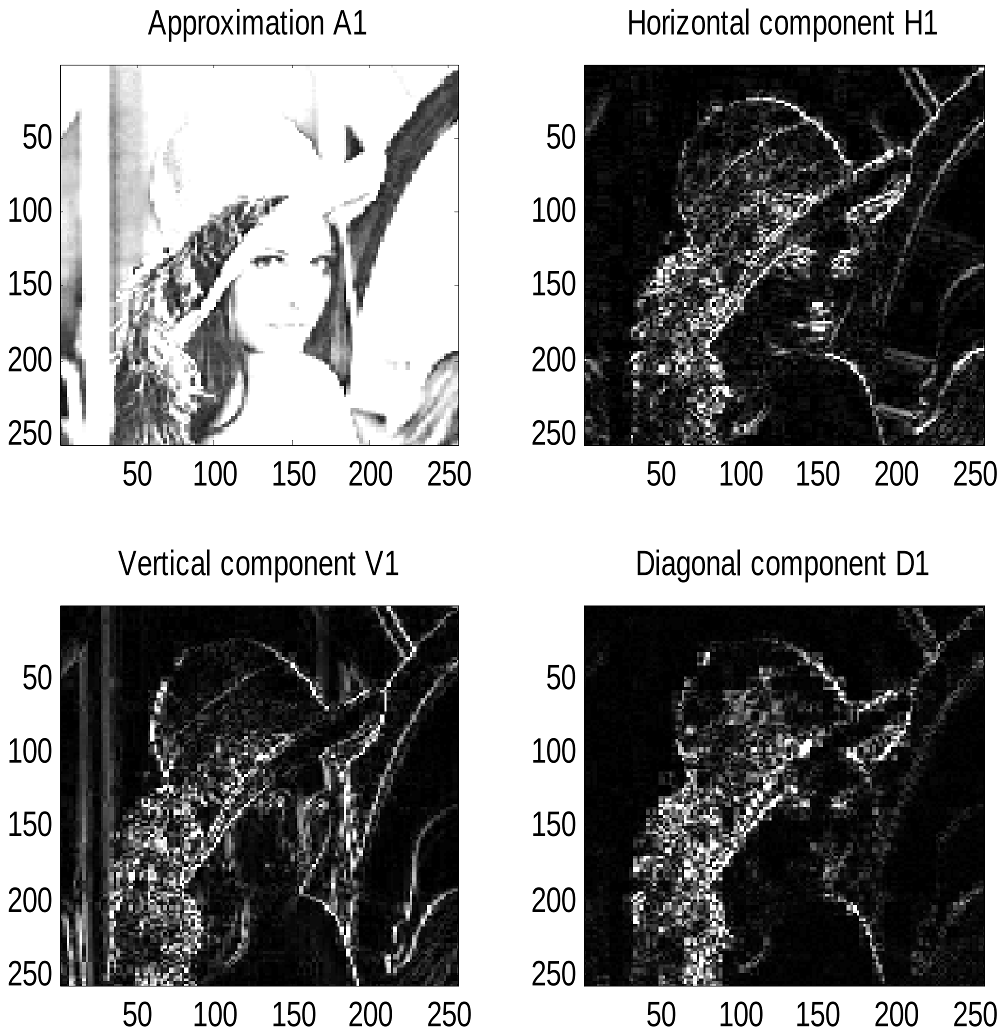

Figure. 4. Images obtained after Wavelet Decomposition

	
Figure 5 : Output Image of WBN from MCMAP Estimate (σ =10, PSNR value 38.92)	**Figure 6 :** Output Image of WBN from MCMAP Estimate (σ =15, PSNR value 36.63)

The approximation of the image, the vertical horizontal and diagonal components of the image is thus obtained after the wavelet decomposition and it is shown in Fig.4. The output image of the wavelet Bayesian network from the MCMAP estimate for different values of noise variance are found using the MATLAB simulation software. The output for the noise variance of 10 is shown in Fig.5 and for the noise variance of 15 is shown in Fig. 6. The corresponding PSNR values are noted down.

It is observed from the above Figures 5 and 6 that there is an improvement in the PSNR values of the output image for the different noise vaiances and also the texture regions are recovered by the proposed MAP estimation.

Table 1 consists of the PSNR values for two different images for the noise variances of 10, 15, 20, 25 and 30 for the existing and proposed image denoising methods.

Table 1 Comparison of PSNR values for different noise levels

Image	Noise Variance (σ)	PSNR				
		BLS-GSM [16]	BM3D [10]	WBN [17]	BCMAP	MCMAP
Lena (256x256)	10	35.91	36.59	36.63	37.31	38.92
	15	34.11	34.78	34.82	35.53	36.63
	20	33.12	33.81	33.94	34.23	35.01
	25	32.01	32.60	32.75	33.61	34.21
	30	31.32	31.92	31.98	32.52	33.36
Cameraman (256x256)	10	33.12	33.94	34.01	35.45	36.91
	15	31.01	31.28	31.63	32.71	33.80
	20	29.96	30.05	30.21	31.51	32.12
	25	29.01	29.81	29.97	30.86	31.59
	30	27.93	28.26	28.81	30.21	30.79

It is estimated from the Table.1 that there is a significant improvement in PSNR values for the proposed BCMAP and MCMAP estimation in compared with the existing BLS-GSM, BM3D and Wavelet Bayesian for the two different images. Compared to the proposed BCMAP, MCMAP performs much better in terms of PSNR value and text preservation. For example, it is noted that for Lena image with noise variance of 15, the percentage increase in PSNR values are 2.08%, 4.16% and 7.38% for wavelet Bayesian, and the proposed BCMAP and MCMAP in comparison with BLS-GSM and for the same, PSNR values are 0.12%, 2.15% and 5.32% for wavelet Bayesian, BCMAP and MCMAP in comparison with BM3D. For Cameraman image, for the same noise variance of 15, with BLS-GSM, the increase in percentage of PSNR values are 1.99%, 5.48% and 8.99% and with BM3D, the increase of PSNR values are 1.11%, 4.57% and 8.05% respectively.

Table 2. Average PSNR improvement of MCMAP in comparison with BCMAP

Noise Variance (σ)	PSNR improvement (%)		Average (%)
	Lena (256x256)	Cameraman (256x256)	
10	4.30	4.11	4.21
15	3.09	3.33	3.21
20	2.28	1.93	2.11
25	1.79	2.36	2.07
30	2.58	1.91	2.25

The above Table.2 shows the PSNR improvement of the proposed MCMAP estimate in comparison with the proposed BCMAP estimate for the Lena and Cameraman images. It is evident that the MCMAP technique preserves the textures and gives an improved PSNR value compared to the BCMAP technique.

5. Conclusion

A constructive data-adaptive procedure that derives a directive acyclic graph structure from the wavelet coefficients based on MAP estimation is proposed. The graph is then used to model the prior probability of the original image for denoising purposes. From the experimental analysis it is clear that for various noise levels, the wavelet Bayesian network based on MCMAP estimation provides better PSNR values when compared to other methods. The proposed method is very efficient and reduces the computational cost. It recovers the texture element. As a future work it has been decided to construct the Curvelet Bayesian Network to suppress Gaussian noise and provide better PSNR values than the present proposed method.

References:

1. Cooper G.F, Herskovits E (1992) A Bayesian method for the induction of probabilistic networks from data, Mach. Learn., 9(4):309–347.
2. Heckerman D, Geiger D, Chickering D.M (1995) Learning Bayesian networks: The combination of knowledge and statistical data, Mach. Learn., 20(3):197– 243.
3. Srivastava A, Lee, A.B, Simoncelli E.P, Zhu S.C (2003), On advances in statistical modeling of natural images, *Journal of Mathematical Imaging and Vision*, 18(1) : 17–33.
4. Goossens B, Pizurica A, Philips W (2009), Image denoising using mixtures of projected Gaussian scale mixtures, *IEEE Trans. Image Process.,* 18(8):1689–1702.
5. Money E.S, Reckhow K.H, Wiesner M.R (2012) The use of Bayesian networks for nanoparticle risk forecasting: model formulation and baseline evaluation, Science of the Total Environment, 42(6) : 436–445.
6. Madadgar S, Moradkhani H (2014) Spatio-temporal drought forecasting within Bayesian networks, Journal of Hydrology, 512(6) :134–146.
7. Heckerman D (2008) A tutorial on learning Bayesian networks, Innovations in Bayesian Networks : 33-82.
8. Chatterjee P, Milanfar P (2010) Is denoising dead? IEEE Trans. Image Process., 19(4): 895–911.
9. Kervrann C, Boulanger J (2006) Optimal spatial adaptation for patch based image denoising, IEEE Trans. Image Process., 15(10): 2866–2878.
10. Dabov K, Foi A, Katkovnik V, Egiazarian K (2007) Image denoising by sparse 3-D transform-domain collaborative filtering, IEEE Trans. Image Process., 16(8):2080–2095.
11. Crouse M.S, Nowak R.D, Baraniuk R.G (1998) Wavelet-based statistical signal processing using hidden Markov models, IEEE Trans. Signal Process., 46(4): 886–902.
12. Milanfar P (2013) A tour of modern image filtering: New Insights and methods, both practical and theoretical, IEEE Signal Process. Mag., 30(1):106–128.
13. Chatterjee P and Milanfar P (2012) Patch-based near-optimal image denoising, IEEE Trans. Image Process., 21(4):1635–1649.
14. Chatterjee P and Milanfar P (2008), A generalization of non-local means via kernel regression in *Proc of SPIE conf. on Computational Imaging*.
15. Mairal J, Bach F, Ponce J, Sapiro G, Zisserman A (2009) Non-local sparse models for image restoration, in Proc. 12th IEEE Int. Conf. Comput. Vision, Kyoto, Japan, 2272–2279.
16. Portilla J, Strela V, Wainwright M.J, Simoncelli E.P (2003) Image denoising using scale mixtures of Gaussians in the wavelet domain, IEEE Trans. Image Process., 12(11) : 1338–1351.
17. Ho J and Hwang W-L (2013) Wavelet Bayesian network image denoising, IEEE Trans. Image Process., 22(4): 1277-1290.
18. Sadreazami H, Omair Ahmad M, Swamy M.N.S (2015), “A robust multiplicative watermark detector for color images in sparse domain, IEEE Transactions on Circuits and Systems II: Express Briefs, 62(12):1159-1163.
19. Sadreazami H, Omair Ahmad M, Swamy M.N.S (2016) Color image denoising using multivariate cauchy PDF in the contourlet domain, In proc. of IEEE Canadian Conference on Electrical and Computer Engineering (CCECE), 1- 4.