Nonlocal Hierarchical Dictionary Learning using Wavelets and Gradient Histogram Preservation for Image Denoising: A Review

Manish Kumar Prajapati
Technocrats Institute of Technology,
Bhopal, India

Deepak Gyanchandani
Technocrats Institute of Technology,
Bhopal, India

ABSTRACT

Image denoising is an important image processing task, both as a process itself, and as a component in other processes. The main properties of a good image denoising model are that it will remove noise while preserving edges. Traditionally, linear models have been used. One common approach is to use a Gaussian filter, In spite of the great success of many denoising algorithms; they tend to smooth the fine scale image textures when removing noise, degrading the image visual quality. To address this problem we compare two methods in this paper. The Nonlocal Hierarchical Dictionary Learning using Wavelet (NHDLW) and Gradient Histogram Preservation (GHP),which is large success in denoising. Experimental result shows that the NHDLW get significantly better denoising results especially on an image denoising algorithms on higher noise levels.

Keywords

Image denoising, wavelets, sparse coding, multi-scale nonlocal, histogram specification, non-local, sparse representation.

1. INTRODUCTION

The image denoising, which aims to estimate the latent clean image A from its noisy observation B. it is a standard but yet still active topic in image processing and low level image. A widely used data observation model GHP is $B = A + v$, where v is additive white Gaussian noise (AWGN). One popular approach to image denoising is the variational method, where energy functional is minimized to search the desired estimation of A from its noisy observation B. The energy functional usually involves two terms: a data reliability term which depends on the image degeneration process and a regularization term which models the prior of clean natural images. In NHDLW method the input noisy image B is from a clean image A contaminated by additive noise with zero-mean. First it transforms B into the wavelet domain which contains several decomposition levels. Within each level, the wavelet coefficients are separated into overlapping patches of a fixed size and each patch is modeled as a vector variable. Then apply k-means clustering to the vectors. After that, in each cluster, a sub-dictionary is trained through a reweighted norm regularization process.

2. METHODOLOGY

2.1 Nonlocal Hierarchical Dictionary Learning in Wavelet Domain (NHDLWD)

As suggested in Ophir *et al.* work [1].The wavelet based dictionary learning can further utilize the sparsity between the wavelet coefficients:

$$\arg \min_{D^w, S^w} ||WY - D^w s^w||_2^2 \text{ s.t. } ||s_i^w||_0 < T, \ \forall_i \qquad (1)$$

Where $Y = [\boldsymbol{y_1}, \boldsymbol{y_2} \dots y_N]$the set of training samples, W is is the wavelet analysis operator, D^w denotes the learned dictionary in the wavelet domain, and S^w is the sparse code matrix. It also utilizes the sparsity in the wavelet domain but in a nonlocal way. This can be easily extended to the wavelet domain - the image self similarity exists in the spatial domain, so there is a strong possibility that the nonlocal characteristic also exists in the wavelet domain for sub bands from the same decomposition level. Therefore, for wavelet patches from the same scale, propose to train different sub-dictionaries in different underlying clusters. Then simply to obtain a light representation in the form that similar wavelet patches use the same dictionary elements but with different sparse codes [2], [3], [4].This would improve the coding efficiency and reduce artifacts. In NHDLWD, Use clustering to obtain training samples for each sub-dictionary. For each decomposition level q, the dictionary $D_{qj}^w \epsilon R^{M \times r_{qj}}$ for each cluster j is obtained from a regularizer with l_1 norms

$$l(S_{qj}^w, D_{qj}^w) = \arg \min_{D_{qj}^w, S_{qj}^w} ||Y_{qj}^w - D_{qj}^w S_{qj}^w||_2^2 + \lambda||S_{qj}^w||_1 (2)$$

$$S.t. \ ||d_{qji}^w||_2^2 \leq b, \ \ i = 1,2, \dots \dots r_{qj}$$

$$J= 1, 2\dots\dots\dots k, \ \ \ q= 1, 2 \dots\dots L$$

Where matrix $\left[Y_{qj_i}^w, \dots \dots, Y_{qj N_q}^w\right] \epsilon R^{M \times N_q}$ represent the input wavelet coefficient patches in clusters j of decomposition level q, $S_{qj}^w = \left[S_{qj 1}^w, \dots \dots S_{qj N_q}^w\right] \epsilon R^{r_q \times N_q}$, and d_{qji}^w denotes i-th column of the dictionary. In terms of patch denoising after the dictionary is learned, two approaches are used in NHDLWD. The simple method is patch reconstruction, in each cluster using the trained sub-dictionary. However, due to the hierarchical structure of
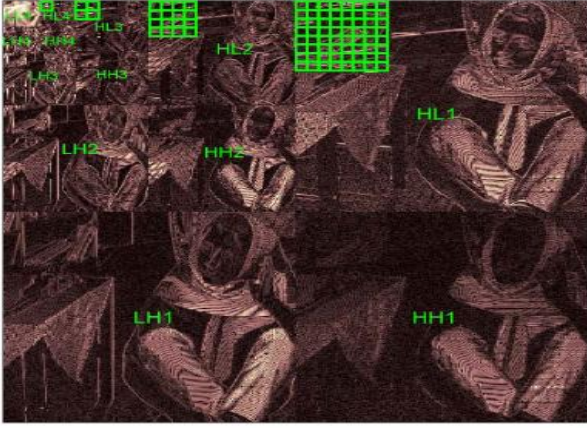
Fig. 1. Wavelet sub bands and quad-tree structure for a coefficient in subbandHL4. The highlighted patches in different scales depict the correlations between different wavelet decomposition levels.

The wavelet coefficients, the lower sub bands generally have far less training samples compared to the higher sub bands, and therefore the sub-dictionaries are usually under-complete. The other method is to link all the sub-dictionaries in the same scale (NHDLW), which identify the problem of insufficient training samples. The aim behind this is that the performance of the dictionary learning system improves with the dictionary redundancy. After the entity training within each cluster, the scale dictionary for each decomposition level q can be obtained by merging the above small dictionaries into an over-complete dictionary $D_q^w \epsilon R^{M \times r_q}$:

$$D_q^w = [D_{q1}^w, \dots . D_{qj}^w, \dots \dots . D_{qk}^w \quad (3)$$

However, along the same orientation, the lower sub bands and higher sub bands are correlated (Fig. 1). If a coefficient is small in the lower sub band, its descendants in the higher sub bands tend to be small accordingly. There are three orientations: LH, HL, and HH. In each orientation, patches with the size m × m are extracted from all the sub bands in all different scales with maximum overlapping (similar to the grouping that has been suggested in [1]). One benefit of this variation is that the number of the training samples in a single orientation is larger than that in a decomposition level, mainly for lower sub bands. However, the nonlocal grouping between patches from different scales.

2.2 Iterative Reweighted Regularizer

In regularization term, it is suggested in [3] that $l_1$norm can perform better in terms of dictionary training, and the l_0 norm can reconstruct the denoise image better. The accuracy of the clustering stage is mainly depends on how noisy the input wavelet coefficients are. From the previous work [3], [5] it can see that better grouping can be achieved with iterative schemes or pre-denoising before classification. The regularization parameter λ controls the balance between the fidelity term and the sparsity term. Assuming that

$A_{qj}^{w(t)} = \left[A_{qj_i}^{w(t)}, \dots \dots A_{qjN_q}^{w(t)} \right] \in R^{M \times N_q}$ Has already been denoised (M is the size of the patch vector and N_q is the number of the cluster members in decomposition level q), in the following iterations the regularization parameter between the fidelity term and the sparsity term should be changed with a more adaptive value θ. This can be expressed as:

$$\min_{S_{qj}^w \in R^{r_q}} \frac{1}{2} ||X_{qj_i}^{w(t)} - D_{qj}^{w(t)} S_{qj}^w||^2 + \gamma ||\theta^{(t)} S_{qj}^w||_1 \quad (4)$$

s.t. j = 1,2,….,k   q = 1,2,…..,j,   t = 1,2 …..,p

Where $D_{qj}^{w(t)} \in R^{M \times r_q}$, $\theta^{(t)}$ is the diagonal matrix with $\theta_1^{(t)}, \dots \dots \theta_{N_q}^{(t)}$ on the diagonal and zeros elsewhere. The above reweighting strategy has been proposed in the compressive sensing field [6] that the new parameter should be inversely proportional to the underlying signal magnitude. In this algorithm, adopt this as "the new parameter should be inversely proportional to the underlying sparse code":

$$\theta_1^{(t+1)} = \frac{1}{\left|S_{qj_i}^{w(t)}\right| + \varepsilon} \quad i = 1\dots\dots,r_q \quad (5)$$

Where ε is a very small constant. The intuition for this reweighted process is that a sparse code that has a small value after t iterations but is not exactly zero will have a large reweighting factor $\theta_1^{(t+1)}$ in the next iteration. This would be

2.3 Denoising Algorithm

In the wavelet domain, the lowest sub band (LL) contains the highest energy coefficients, and it is vigorous to noise. So simple soft-thresholding,[6] is applied on the LL coefficients. The rest of the sub bands are processed by Nonlocal Hierarchical Dictionary learning using Wavelets (NHDW) algorithm, as shown in Fig. 2.

1. Input: The noisy image B, standard deviation of the noise σ.

2. Parameters:

    *   The iteration times for the dictionary learning process P;

    *   The patch size of the wavelet coefficients M = m × m;

    *   The type of the wavelet transform;

    *   The number of dictionary atoms per dictionary for each Decomposition level r_l ;

    *   The K for k-means clustering

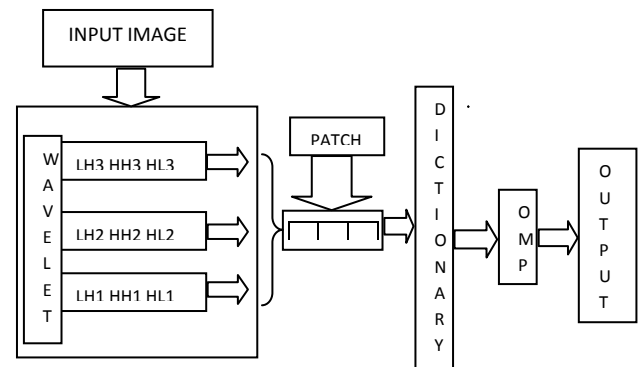   *   The initial γ for thel_1 regularization and the initial θ

Fig. 2. Flowchart of proposed method.

3. Initialization: The primary dictionaries can be learned from the online process or fixed basis transform matrices like Discrete Cosine Transform (DCT). It seems that, the pre learned dictionary (learned from the clean images) is a good initial dictionary for the denoising process [3]. However, in this method use DCT as the initial dictionary.

4. Wavelet decomposition: The 2-D wavelet transform is

Applied to the noisy input image B:

$$B^w = w_B = A^w + \eta^w \qquad (6)$$

For each decomposed level q, (q = 1, 2,….J), patches with the size $m \times m$ are extracted from with B^w maximum overlapping.

5. Dictionary Learning: In this process, for each wavelet decomposition level, repeat *P* times:

• K-means clustering: The wavelet coefficients patch matrix B_q^w will be clustered by k-means clustering into K clusters B_{qj}^w, (q = 1,2, ..… k)

$$\arg\min_{Hm} \ \sum_{j=1 \, H(B_{qj_i}^w)\in H}^{k} \sum_{mj, i=1,2....N_j} \ |H(B_{qj_i}^w) - \mu_j|^2 \quad (7)$$

Where $B_{qj_i}^w$ represents a noisy patch i belonging to cluster j decomposition level q, and m_j is the mean vector for the jth cluster H_{mj} Then, it can obtain K clusters $H_{m1}, H_{m2}, H_{mj}, \ldots, H_{mk}$. Each cluster H_{mj} is composed of N_j vectors.

• Sub-dictionary training: Within each cluster H_{mj}, (*j* = 1, 2, ..., *K)*, the optimization of the dictionary is between the dictionary and the sparse codes alternatively:

– Initialization: the sub-dictionary is initialized with DCT coefficients.

– repeat *P* times:

* Sparse Coding Stage - fix the dictionary, update the sparse codes: use the reweighted least angle regression (LARS) [7] to compute the sparse codes for the patches in each cluster.

∗ Dictionary Update Stage - fix the sparse codes, update the dictionary: compute dictionary $D_{qj}^{w(t)}$ using $D_{qj}^{w(t-1)}$ as the initial value. The learning process is the same as algorithm 2 in [8]

$$D_{qj}^{w(t)} = \arg\min_{D_{qj}^w} \sum_{t=1}^{t} \tfrac{1}{2}||A_{qj}^{w(t)} - D_{qj}^w S_{qj}^{w(t)}||_2^2 + \gamma||\theta^{(i)} S_{qj}^{w(i)}||_2 (8)$$

– Return the learned dictionary

Within each cluster rH_{mj}, (j = 1, 2, ..., K), the sub-dictionary $D_{qj}^w \in R^{m \times r_q}$ is trained by the iterative method.Afterwards, all the sub-dictionaries will be concatenated at each decomposition level. Now, there are over-completed dictionaries with different sizes for different decomposition levels $D_q^w \in R^{M \times r_l}$.

6. Image reconstruction:

Sparse coding: For the restoration stage, the dictionary for each decomposition level is known. The aim is to find the sparse representation $\hat{s}_{ij}$ for each location and the overall output wavelet coefficients. The Orthogonal Matching Pursuit (OMP) [9] is used for obtaining sparse codes:

$$\hat{s}_{ij}^w = \arg \min_{S_{ij}^w} ||s_{ij}^w||_0 \quad s.t. ||D_q^w s_{ij}^w - R_{ij}A_q^w||_2^2 \leq C\sigma \qquad (9)$$

Where s_{ij}^w indicates the sparse codes of wavelet coefficient located at (i, j) C is a constant, σ is the standard deviation of the noise, R_{ij} is the binary matrix that extracts the (i, j) patch from the wavelet coefficient matrix, and D_q^w is the combined dictionary for decomposition level q.

• Reconstruction in the wavelet domain:

The final reconstructed wavelet coefficients matrix $\hat{A}_q^w$ can be estimated by:

$$\hat{A}_q^w = \arg\min_{A_q^w} \lambda_1 ||A_q^w - B_q^w||_2^2 + \textstyle\sum_{ij} ||D_q^w \hat{s}_{ij}^w - R_{ij}A_q^w||_2^2$$

Where q = 1, 2,..., J . (10)

• Image reconstruction:

$$\hat{A} = W_s \hat{A}^w \ ;$$

Where W_s is the inverse wavelet transform. After obtain the denoised patches in spatial domain $\hat{A}$, then average the estimates of each pixel to reconstruct the last image.

2.4 The Texture Enhanced Image Denoising Framework

The noisy observation B of an unknown clean image A is usually modeled as

$$B = A + v \qquad (11)$$

Where v is the additive white Gaussian noise with zero mean and standard deviation σ. The aim of image denoising is to estimate the required image A from B. One popular method of image denoising is the variational method, in which the denoised image is procured by

$$\hat{A} = \arg\min_A \{\tfrac{1}{2\sigma^2}|| B - A||^2 + \lambda. R(A)\} \qquad (12)$$

Where *R(A)* denotes some regularization term and λ is a positive constant. The specific form of *R(*A*)* depends on the previous image. The common problem of image denoising methods is that the image fine scale details such as texture structures will be over-smoothed. An over-smoothed image will have much weaker gradients than the original image. Naturally, a good inference of A without smoothing too much the texture should have a similar gradient distribution to that of A. So in gradient histogram preservation (GHP) model for texture enhanced image denoising, whose construction is shown in Fig. 3.Assume that an estimation of the gradient histogram of A, denote by h_r. In order to make the gradient histogram of denoised image $\hat{A}$ nearly the same as the reference histogram h_r,the following GHP based image denoising model:

$$\hat{A} = \arg\min_A F\left\{\tfrac{1}{2\sigma^2}||B - A||^2 + \lambda R(A) + \mu||F(\nabla_A) - F\nabla_A||^2\right\}, s.t. h_F = h_r \quad (13)$$

Where *F* indicate an odd function which is monotonically non- cadent, . h_F indicate the histogram of the transformed gradient image |*F* (∇ A)|, ∇ denotes the gradient operator, and μ is a positive constant.GHP algorithm use the alternating optimization approach.
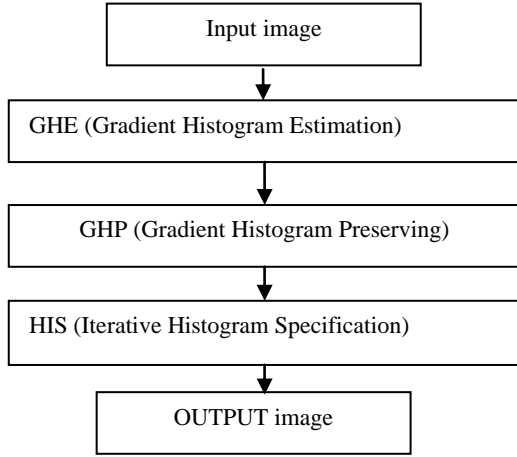
Fig. 3: Flowchart of the proposed texture enhanced image denoising framework.

Given A, it can revise F by the histogram specification based shrinkage operator. Thus, by insert F, it can easily include the gradient histogram constraint with any existing image Regularizer $R(A)$ One more issue in the GHP model is how to find the reference histogram h_r of unknown image A. In carry out, approximate h_r based on the noisy observation B. Once the reference histogram h_r is established, then the GHP algorithm is used for texture enhanced image denoising.

2.5 Denoising With Gradient Histogram Preservation

1. The Denoising Model

This denoising model is a patch based method. Let $A_i = R_i A$ be a patch extracted at position i, $i = 1,2, \ldots \ldots, N$ where R_i the patch extraction operator and N are the number of pixels in the image. Given a dictionary D, it meagerly encodes the patch A_i over D, resulting in a thin coding vector α_i. Once the coding vectors of all image patches are obtained, the whole image A can be reconstructed by [9]:

$$A = D \circ \alpha \triangleq \left(\sum_{i=1}^{N} R_i^T R_i \right)^{-1} \sum_{i=1}^{N} R_i^T D\alpha_i \quad (14)$$

Where $\boldsymbol{\alpha}$ is the concatenation of all α_i. A proper integration of different priors could further get better the denoising performance. For example, the methods in [10], [11], and [12] integrate image local sparsity prior with nonlocal NSS prior, and they have shown promising denoising results. In the proposed GHP model, the following sparse nonlocal regularization term used in the non -locally centralized sparse representation (NCSR) model [11]:

$$R(A) = \sum_i || \alpha_i - \beta_i ||_1 \quad , \text{ s.t. } A = D \circ \alpha \quad (15)$$

Where β_i is defined as the weighted average of α_i^q:

$$\beta_i = \sum_q w_i^q \alpha_i^q, \quad (16)$$

And α_i^q is the coding vector of the qth adjacent patch (denoted by A_i^q) to x_i. The weight is defined as

$$w_i^q = \frac{1}{w} \exp\left(-\frac{1}{h} ||\hat{A}_i - \hat{A}_i^q||^2 \right)$$

($\hat{A}_i$ And. $\hat{A}_i^q$ denoted the current estimated of A_i and A_i^q, respectively), where h is a predefined constant and W is the normalization factor. By using the above $R(A)$ into Eq. (13), then the GHP model can be formulated as:

$$\hat{A} = \arg\min_A, F\{ \frac{1}{2\sigma^2} || B - A ||^2 + \lambda \sum_i || \alpha_i - \beta_i ||_1 + \mu ||F(\nabla_A) - \nabla_A ||^2 \}$$

$$\text{s.t. } A = D \circ \alpha, \ h_F = h_r \quad (17)$$

In the GHP model with sparse nonlocal regularization, if the histogram regularization parameter μ is high, the function F (∇_A) will be close to ∇_A Since the histogram h_F of $|F\nabla_A|$ is

required to be the same as h_r the histogram of ∇_A will be comparable to h_r leading to the preferred gradient histogram preserved image denoising.

2. Region-based GHP

The histogram constraint in Eq. (17) is universal. If the image consists of different region with different textures, GHP may produce some false textures in the less textured areas. To address this problem, split the noisy image into different area and use the reference gradient histogram of each area, and then apply GHP to each area for denoising. As shown in Fig. 3, using the two method to partition the first method (Fig. 3(a)), namely S-GHP, is to use k-means clustering method to nearly partition the image into K homogeneous regions, while the second method (Fig. 3(b)), namely B-GHP, simply partitions the noisy image into $k = \sqrt{k} \times \sqrt{k}$ blocks with equal size. indicate by $\{\Omega_K, \ldots \Omega_K, \ldots \Omega_K\}$ the partitioned regions. Each region Ω_K has the equivalent reference gradient histogram $h_{r,k}$ and have a function F_k to process the pixels within region Ω_k:

$$\min_{F_k} \sum_{(i,j) \in \Omega_k} (F_k (\nabla_A) ij)^2$$

$$\text{s.t. } hF_k = h_{r,k} \quad (18)$$

Define an indicator function

$$1_{\Omega_c}(i,j) = \begin{cases} 1, \text{ if } (i,j) \in \Omega_k \\ 0, \quad \text{ else} \end{cases} \quad (19)$$

The F (∇_A) for S-GHP/B-GHP can then be defined as

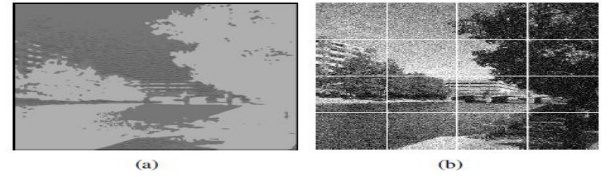$$F(\nabla A) = \sum_k F_k (\nabla A) 1_{\Omega_k} \quad (20)$$

Fig. 3: Two image partition method. (a) The noisy image is partitioned into K homogeneous regions by k-means clustering. (b) The noisy image is partitioned into $\sqrt{k} \times \sqrt{k}$ blocks.

2.6 Reference Gradient Histogram Estimation

To use this system in Eq. (17), First of all know the reference gradient histogram h*r* of original image A. In this part, use a regularized deconvolution method to estimate the histogram h_r.Let the pixels in gradient image ∇A are independent and identically distributed (i.i.d.) Analysis them as the samples of a scalar variable, denoted by A. The normalized histogram of ∇_A can be regarded as a discrete approximation of the probability density function (PDF) of A. For the AWGN v, its elements as the samples of an i.i.d. variable, denoted by

V, since $v \sim N(0, \sigma^2)$ and let $\varepsilon = \nabla v$, can then be well approximated by the i.i.d. Gaussian with PDF [16]

$$\rho_\varepsilon = \frac{1}{2\sqrt{\pi}\sigma} \exp(-\frac{\varepsilon^2}{4\sigma^2}) \quad (21)$$

Since B = A + v, and $\nabla_B = \nabla_A + \nabla_v$. It is ready to model ∇_B as an i.i.d. variable, denoted by B, and we have $B = A + \varepsilon$. Let p_A be the PDF of A, and p_B are the PDF of B. Since A and ε are independent, the joint PDF p (A; ε) is

$$P(A, \varepsilon) = p_A \times p_\varepsilon \quad (22)$$

Then the PDF p_B is

$$p_B(B = t) = \int_a (A = a) \times p_\varepsilon(\varepsilon = (t - a)) da \quad (23)$$

use the normalized histogram h_A and h_B to approximate p_A and p_B ,modify Eq. (23) in the discrete domain as:

$$h_B = h_A \otimes h_\varepsilon \qquad (24)$$

Where $\otimes$ denotes the convolution operator. Clearly, the estimation of h_A can be commonly modeled as a deconvolution problem:

$$h_r = \arg\min_{h_r}\{||h_B - h_A \otimes h_\varepsilon||^2 + c\,.\,R\,(h_A)\} \qquad (25)$$

Where c is a constant and $R(h_A)$ is some regularization term based on the previous information of natural image's gradient histogram. Consider two kinds of constraints on h_A. First, it has been shown that p_A (i.e., the continuous counterpart of h_A) can be approximated by hyper-Laplacian distribution [15],[16],[17]. Considering that the real h_A strength turn from the hyper-Laplacian provides to some extent, it only need that h_A should be close to the hyper-Laplacian distribution:

$$p_A \approx C.\exp(-k|A|^\gamma) \qquad (26)$$

Where C is the normalization factor γ and k are the two parameters of the hyper-Laplacian distribution. More specifically, let $k \in [0:001; 3]$ and $\gamma \in [0.02, 1.5]$. Second, each element of h_A should be nonnegative. Based on these two constraints, gradient histogram guess can be formulated as the following regularized deconvolution problem:
$h_r =$
$\text{argmin}_{h_A}, c, k, \gamma ||h_B - h_A \otimes h_\varepsilon||^2 +$
$c||h_A - C.\exp(-k|A|^\gamma)||^2$

$$\text{s.t.}\,h_A \geq 0,$$

(27)

which can be re- written as:

$$h_r = \text{argmin}_{h_A, h'_A} c, k, \gamma \left\{ \begin{array}{c} ||h_B - h_A \otimes h_\varepsilon||^2 \\ +c||h_A - C.\exp(-k|A|^\gamma)||^2 \\ +\eta||h_A - h'_A||^2 \end{array} \right\}$$

(28)

$$\text{s.t.}\,h'_A \geq 0$$

By iteratively update h_A, h'_A, C, $\gamma$and k alternatively. Let

$h_0 = C \cdot \exp(-k|A|^\gamma)$, h_A is updated by

$$h_A = \frac{\overline{\text{FFT}\,(h_\varepsilon)} \bullet \text{FFT}\,(h_B) + c\text{FFT}\,(h_0) + \eta\text{FFT}(h'_A)}{\text{FFT}\,(h_\varepsilon) \bullet \overline{\text{FFT}\,(h_\varepsilon)} + c + \eta} \qquad (29)$$

Where "•" denotes the element-wise multiplication, " $\frac{*}{*}$ " denotes the element-wise division, and "$\overline{*}$" denotes the complex conjugate operator. h'_A is simplified by

$$h'_A(i) = \max\,(h_A\,(i),0). \qquad (30)$$

C is simplified by

$$C = \frac{\sum_i \exp(-k|i|^\gamma)}{\sum_i h_A(i)} \qquad (31)$$

γ And k updated based on gradient decent

$$K^{(t+1)} = k^{(t)} + \tau \sum_i C\,|i|^\gamma \exp\big(-k^{(t)}|i|^\gamma\big)\,(C.\exp\mathbb{Z}\!-\!k^{(t)}|i|^\gamma - h_A(i)) \qquad (32)$$

$$\gamma^{(t+1)} = \gamma^{(t)} + \rho \sum_{|i\neq0|} \left\{ \begin{array}{l} c_{k^{(t)}}|i|^\gamma \text{ in } |i|)\exp\mathbb{Z}\!-\!k^{(t)}|i|^\gamma \\ .\,(C.\exp\mathbb{Z}\!-\!k^{(-t)}|i|^\gamma - h_A(i)) \end{array} \right\} \qquad (33)$$

In region based B-GHP and S-GHP, the regularized deconvolution scheme can be directly applied to each area to estimate the corresponding reference gradient histogram.

3. RESULT ANALYSIS AND DISCUSSION

Experimental results of this section are obtained by applying the proposed method on real images. All test images of size 512×512, were corrupted with the additive white Gaussian noise of various standard deviation σ .It compares the both the NHDLW and GHP method with the several state-of-the-art methods. To verify the performance of the proposed NHDLW and GHP based image denoising method, apply it to natural images with various texture structures, whose scenes are shown in Fig.4 & 5.All the test images are gray-scale images with gray level ranging from 0 to 255. Finally, experiments are conducted to validate its performance in comparison with the state-of-the-art denoising algorithms. In the following experiments set the AWGN standard deviation i.e. σ from 20 to 40 with step length. Use the PSNR as the objective measure of the denoising quality, as it is widespread throughout the image processing community. As PSNR is proved to be inconsistent with the human eye perception, it also use more profound structural similarity index measure (SSIM) [13] to additionally validate the proposed method, in comparison to the baseline algorithms. The resulting images for the baseline algorithms were chosen so that the best PSNR is reached, as it is usually done in the text. Results presented show that the proposed method significantly outperforms all baseline methods in the means of both, the PSNR and the SSIM measure. It can be mention that in the actual exploitation the referent, i.e., original image is not available. Thus, one has to fix the number of iterations of the used diffusion algorithm. This possibly causes the stronger over smoothing effect on the target image. So the actual result would be even more in favor of the proposed method. In Figure 4.a original test images are presented, while the corresponding noisy images are presented in Figure 4b,c,d for $\sigma = 25$.
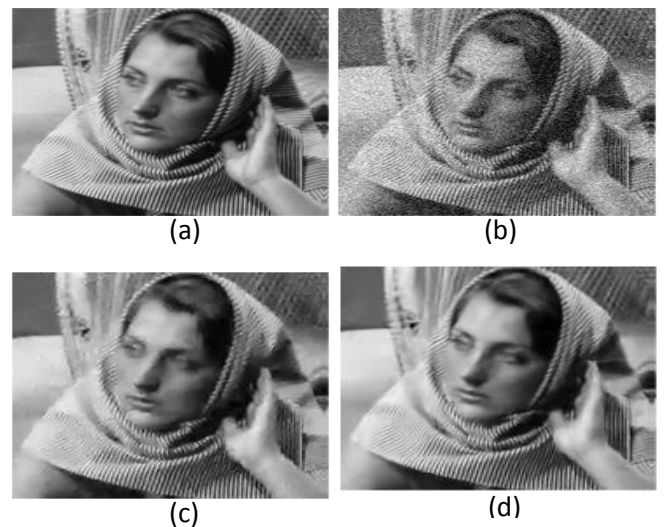

(a)

(b)

(c)

(d)

Fig.4. Denoising results for the image "Barbara" with σ = 25 (Gaussian noise).

Tested the proposed algorithm on "Barbara", and "car - scene" test Images, with three different values of Gaussian white noise standard deviation σ.
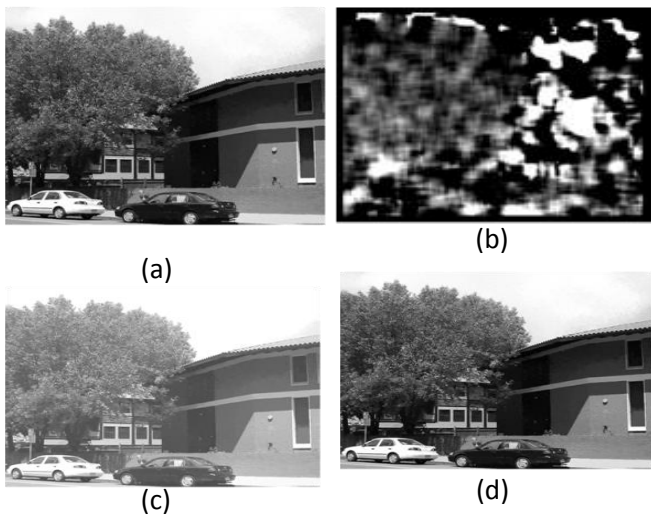

(a)

(b)

(c)

(d)

Fig. 5: Car scene From left to right and from top to bottom: original image, the difference PSNR map.

Table 1 Comparison Of Nhdlw And Ghp Image Denoising Methods

S.NO	METHOD	PSNR For σ=20	PSNR For σ=25	PSNR For σ=30
1	NHDLW	31.12	30.71	29.32
2	GHP	30.83	29.66	28.75

4. CONCLUSION

This paper compares NHDLW and GHP image denoising methods. The NHDLW method, takes advantages of the sparse coding framework, nonlocal grouping and wavelet transform, leading to the state of-the-art denoising performance. It builds a nonlocal hierarchical sparse dictionary on the wavelet coefficients of a noisy image. Once the dictionaries are trained, they are combined as a whole to represent the entire decomposition level. In GHP model, its associated algorithm was presented to estimate the reference gradient histogram from the noisy image, and an efficient iterative histogram specification algorithm was developed to implement the GHP model. One limitation of GHP is that it cannot be directly applied to non-additive noise removal, such as multiplicative Poisson noise and signal-dependent noise [14]. It would be interesting and valuable to study more general models and algorithms for non-additive noise removal with texture enhancement. Thus, the NHDLW method gives the better performance at different σ as compare to GHP method, so it conclude NHDLW method is more preferable over GHP method. The computation time can be the matter for future work.

5. REFERENCES

[1] B. Ophir, M. Lusting, and M. Elad, "Multi-scale dictionary learning using wavelets," *IEEE J. Sel. Topics Signal Process.*, vol. 5, no. 5, pp. 1014–1024, Sep. 2011.

[2] P. Chatterjee and P. Milanfar, "Clustering-based denoising with locally learned dictionaries," *IEEE Trans. Image Process.*, vol. 18, no. 7, pp. 1438–1451, Jul. 2009.

[3] J. Mairal, F. Bach, J. Ponce, G. Sapiro, and A. Zisserman, "Non-local sparse models for image restoration," in *Proc. IEEE Int. Conf. Comput. Vis.*, Kyoto, Japan, Sep./Oct. 2009, pp. 2272–2279.

[4] K. Vladimir, F. Alessandro, E. Karen, and A. Jaakko, "From local kernel to nonlocal multiple-model image denoising," *Int. J. Comput.Vis.*, vol. 86, no. 1, pp. 1–32, Jan. 2010.

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

[6] E. Candes, M. Wakin, and S. Boyd, "Enhancing sparsity by reweighted L1 minimization," *J. Fourier Anal. Appl.*, vol. 14, pp. 877–905, Nov. 2008.

[7] B. Efron, T. Hastie, I. Johnstone, and R. Tibshirani, "Least angle regression," *Ann. Stat.*, vol. 32, no. 2, pp. 407–409, 2004.

[8] J. Mairal, F. Bach, J. Ponce, and G. Sapiro, "Online dictionary learning for sparse coding," in *Proc. IEEE Signal Process. Lett*, Montreal, QC Canada, Aug. 2009, pp. 689–696.

[9] M. Elad and M. Aharon,"Image denoising via sparse and redundant representations over learned dictionaries," *IEEE Trans. Image Process.*, vol. 15, no. 12, pp. 3736–3745, Dec. 2006.

[10] J. Mairal, F. Bach, J. Ponce, G. Sapiro, and A. Zisserman, "Non-local sparse models for image restoration," in *Proc. Int. Conf. Comput. Vis.*, Sep./oct. 2009, pp. 2272–2279.

[11] W. Dong, L. Zhang, G. Shi, and X. Li, "Nonlocally centralized spars representation for image restoration," *IEEE Trans. Image Process.*, vol. 22, no. 4, pp. 1620–1630, Apr. 2013.

[12] J. Jancsary, S. Nowozin, and C. Rother, "Loss-specific training of nonparametric image restoration models: A new state of the art," in *Proc.12th ECCV*, 2012, pp. 112–125.

[13] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE* Trans. Image Process., vol. 13, no. 4, pp. 600–612, Apr. 2004.

[14] B. Zhang, J. M. Fadili, and J. L. Starck, "Wavelets, ridgelets, and curvelets for Poisson noise removal," *IEEE Trans. Image Process.*vol. 17, no. 7, pp. 1093–1108, Jul. 2008.

[15] D. Krishnan and R. Fergus, "Fast image deconvolution using hyper-Laplacian priors," in *Proc. NIPS*, 2009, pp. 1033–1041.

[16] T. S. Cho, C. L. Zitnick, N. Joshi, S. B. Kang, R. Szeliski, and W. T. Freeman, "Image restoration by matching gradient distributions,"IEEE. Trans. Pattern Anal. *Mach. Intell.*, vol. 34, no. 4, pp. 683–694,Apr. 2012.

[17] T. S. Cho, N. Joshi, C. L. Zitnick, S. B. Kang, R. Szeliski, and W. T. Freeman, "A content-aware image prior," in *Proc. Int. Conf. CVPR*, Jun. 2010, pp. 169–176.