Image Denoising with Sparsity Distillation

SATOSHI KAWATA¹,a) NAO MISHIMA¹,b)

Received: March 13, 2015, Accepted: April 20, 2015, Released: July 27, 2015

Abstract: We propose a new image denoising method with shrinkage. In the proposed method, small blocks in an input image are projected to the space that makes projection coefficients sparse, and the explicitly evaluated sparsity degree is used to control the shrinkage threshold. On average, the proposed method obtained higher quantitative evaluation values (PSNRs and SSIMs) compared with one of the state-of-the-art methods in the field of image denoising. The proposed method removes random noise effectively from natural images while preserving intricate textures.

Keywords: image restoration, noise reduction, sparsity, weighted principal component analysis

1. Introduction

Denoising from captured digital images is a challenging problem in the field of digital image processing. Various image denoising methods [1], [2], [3], [4], [5], [6], [7], [8] have been proposed. Many denoising methods are called “filtering methods.” In filtering methods [1], [2], [3], denoising is achieved by a convolution of circumjacent pixel values. However, it is difficult to estimate a true (noise-free) signal from heavily noisy samples pixel-wise.

Recently, denoising methods using block projection to basis vectors have been reported [4], [5], [6], [7], [8]. In these methods, N-dimensional vectors of pixel values in an image block whose size is N pixels are projected to a space constructed by N basis vectors. Projection coefficients smaller than a shrinkage threshold are regarded as noise, and shrunk. We call these methods “shrinkage methods.”

In regard to shrinkage methods, two important questions arise. First, what kind of spaces should be used? Second, how should projection coefficients of noise be shrunk?

Donoho proposed the wavelet shrinkage method [4]. In wavelet shrinkage, blocks are transformed by 2D wavelet bases. The shrinkage process is soft-thresholding of projection coefficients.

Dabov et al. proposed the BM3D method [5]. In BM3D, blocks are transformed by a 3D array transform. The shrinkage process is composed of two steps. The first step is hard-thresholding of projection coefficients. The second step is Wiener filtering. BM3D is known as one of the state-of-the-art methods in the field of image denoising.

Yamauchi et al. proposed the nonlocal PCA method [6]. In nonlocal PCA, blocks are transformed by weighted principal component analysis (PCA) bases. The shrinkage process is composed of two steps. The first step is hard-thresholding using eigen values of weighted PCA. The second step is Wiener filtering.

In this paper, we propose a new shrinkage method for image denoising. In the proposed method, shrinkage is controlled by explicitly evaluated “sparsity” degrees in spaces of weighted PCA bases.

2. Proposed Method

2.1 Sparsity of Projection Coefficients

It is commonly known that projection coefficients of natural image blocks are concentrated on a small number of basis vectors. In other words, natural image blocks can be represented sparsely. On the other hand, coefficients of random noise are distributed to all basis vectors evenly. In other words, random noise is not sparse. Shrinkage methods are based on this phenomenon [4], [5], [6], [7], [8].

Extracting a few coefficients corresponding to image signals from corrupted coefficients, noise can be removed from an image. In the proposed method, blocks are projected to a space that makes coefficients sparser, and explicitly evaluated sparsity is used to control the gain of shrinkage.

2.2 Space for Denoising

BM3D [5] uses Discrete Cosine Transform (DCT) bases and Wavelet Transform (WT) bases for denoising space. These basis vectors represent an image block as a combination of geometric patterns. However, natural images have intricate textures. Hence, representation with geometric patterns is not necessarily optimal.

Figure 1 shows a conceptual diagram of projection image blocks to DCT space and the shrinkage process. For simplicity, let the size of a block be 2 pixels. Two axes in Fig. 1 show two 2-dimensional basis vectors. On the left in Fig. 1, yellow points show projection coefficients. On the right in Fig. 1, gray points show shrunk coefficients, and red points show coefficients larger than a shrinkage threshold. Basis vectors of DCT decomposes image blocks to frequency components. Hence, they cannot extract only image signals from the sample blocks buried in noise.
Fig. 1 Conceptual diagram of projection blocks to DCT basis vectors and shrinkage.

Fig. 2 Conceptual diagram of projection blocks to PCA basis vectors and shrinkage.

Therefore, noise remain, or image signals are lost with noise in DCT space.

PCA bases transform image blocks based on statistical distribution of image signals. Figure 2 shows a conceptual diagram of projection image blocks to PCA space and the shrinkage process. Basis vectors of PCA extract components that are plausible as image signals statistically. In the space of PCA bases, projection coefficients of image signals distribute sparsely. Hence, compared with DCT, space of PCA can reduce noise while preserving image signals. In the proposed method, space for noise reduction is generated by weighted PCA.

2.3 Shrinkage Using Sparsity

As conceptually shown in Fig. 2, image blocks projected to PCA space are expected to be decomposed into image signals and noise. When image signals are dominant within a block (in a region where an image structure such as an edge exists), projection coefficients of an image signal are concentrated on a small number of basis vectors. On the other hand, in a flat region, projection coefficients are distributed evenly. From here onwards, evaluating how sparsely coefficients are concentrating, it can be judged whether an image signal is dominant in a block. Furthermore, when coefficients are concentrated on a small number of basis vectors, components plausible as image signals can be extracted.

When a certain block is given and projection coefficients are concentrated on a small number of basis vectors, components plausible as image signals can be extracted. When a certain block is given and projection coefficients are concentrated on a small number of basis vectors, coefficients of noise fall relatively. In a flat region where image texture signals do not exist, in order to ensure concentration of coefficients cannot occur, all basis vectors take the same coefficients. In the proposed method, a shrinkage threshold is made a low value in a sparse region, and it is made a high value in a region that is not sparse. This enables noise to be reduced efficiently without loss of image texture signals.

2.4 Algorithm

We consider an observed pixel value \( v(x_i, y_i) \) at pixel \( i \) as

\[
    v(x_i, y_i) = z(x_i, y_i) + n
\]

(1)

where \( z \) is a true (noise-free) signal, and \( n \) is an additive Gaussian noise with standard deviation \( \sigma \). \((x_i, y_i)\) is a 2D spatial coordinate of pixel \( i \) that belongs to the image domain. We define an image block of fixed \( N \times N \) size containing pixel \( i \) at top-left as an \( N^2 \)-dimensional column vector

\[
    V(i) = [z(x_i, y_i), \ldots, z(x_i+N-1, y_i+N-1)]^T
\]

(2)

where \( T \) denotes transposition of vector or matrix.

2.4.1 Reference Blocks Setup

Figure 3 is a diagram showing how blocks are set up in the proposed method. At first, some reference blocks are set up in an input image by \( N_{\text{step}} \) pixels. In Fig. 3, reference blocks are shown as red blocks. A reference block at pixel \( i \) (yellow point in Fig. 3) is defined as

\[
    V_{\text{ref}}(i) = [z(x_i, y_i), \ldots, z(x_i+N-1, y_i+N-1)]^T.
\]

(3)

Reference blocks are focal point blocks of each process unit.
2.4.2 Weighted PCA

Circumjacent blocks (blue blocks in Fig. 3) are collected in a search region $\Omega(i)$ (purple region in Fig. 3) centered at pixel $i$, and similarities $w(i, j)$ (green arrows in Fig. 3) between a reference block $V_{ref}(i)$ and each circumjacent block $V(j)$ as,

$$w(i, j) = \exp \left( -\frac{||V_{ref}(i) - V(j)||^2}{2\sigma^2} \right)$$

(4)

where $||\cdot||$ denotes a Sum of Square Difference (SSD) between 2 blocks. If a reference block $V_{ref}(i)$ and a circumjacent block $V(j)$ are completely homologous, a similarity $w(i, j)$ is set to 1. Weighted covariance matrix $C(i)$ is computed as weighted average of autocovariance matrices of circumjacent blocks as,

$$C(i) = \frac{1}{W(i)} \sum_{j \in \Omega(i)} w(i, j)V(j)V^T(j)$$

(5)

$$W(i) = \sum_{j \in \Omega(i)} w(i, j).$$

(6)

$C(i)$ is a matrix with $N^2 \times N^2$ size. Eigenvalue decomposing $C(i)$, $N^2$ set of basis vectors $B_k(i)$ whose dimensions are $N^2$ and corresponding eigenvalues $\lambda_k(i)$ are obtained $(1 \leq k \leq N^2)$. Basis vectors $B_k(i)$ are used to space for image denoising.

2.4.3 Sparsity Evaluation

Eigenvalue $\lambda_k(i)$ means a statistical contribution degree of each basis vector. When projection coefficients are sparse, the eigenvalues corresponding to a small number of basis vectors take relatively large values. In a not sparse region, all eigenvalues come close to $\sigma^2$. Hence, by calculating a deviation of eigenvalues, a sparsity can be evaluated. Contribution ratio $\rho(i)$ of each basis vector is calculated as

$$\rho_k(i) = \frac{\lambda_k(i)}{\Lambda(i)}$$

(7)

$$\Lambda(i) = \sum_{k=1}^{N^2} \lambda_k(i)$$

(8)

where $\Lambda(i)$ means summation of eigenvalues. Contribution ratio $\rho_k(i)$ is accumulated in descending order until an accumulation value $\rho_{sum}(i)$ reaches a threshold $\rho_h$

$$\rho_{sum}(i) = \sum_{k=1}^{d(i)} \rho_k(i) \leq \rho_h$$

(9)

where $d(i)$ denotes a number of accumulated contribution ratios when $\rho_{sum}(i)$ reaches $\rho_h$. Threshold $\rho_h$ decides the boundary where signals are dominant to noise. In a noise-free image, it is preferable to set $\rho_h = 1$ because all basis vectors represent image signals. In a heavily noisy image, it is preferable to lower $\rho_h$ from 1 because eigenvalues are influenced by noise. It is desirable that $\rho_h$ be determined by the ratio of each power spectrum of noise-free signals and observed signals. However, noise-free signals are unavailable. So in the proposed method, $\rho_h$ is determined using a maximum dynamic range of image, as

$$\rho_h = \frac{255^2}{255^2 + \sigma^2}$$

(10)

when an input image is composed of 8-bit data. The sparsity $s(i)$ is defined as

$$s(i) = \frac{N^2}{d(i)}.$$  

(11)

and $s(i)$ is evaluated for each reference block respectively.

2.4.4 Shrinkage Using Sparsity

Circumjacent blocks $V(j)$ of pixel $i$ are collected again, and these are projected to an $N^2$-dimensional space with $N^2$ basis vectors $B_k(i) (1 \leq k \leq N^2)$. Obtained $N^2$ projection coefficients are denoted as $a_k(j)$. To represent image blocks sparsely for denoising, projection coefficients that are less than or equal to a threshold value are shrunk as,

$$\hat{a}_k(i) = \begin{cases} 0, & \text{for } a_k^2(j) \leq \tau(i) \\ a_k(j), & \text{otherwise.} \end{cases}$$

(12)

where $\hat{a}_k(i)$ denotes denoised coefficient. In the proposed method, the threshold value for shrinkage $\tau(i)$ is set low so as not to lose image texture in sparse regions, and it is set high to reduce noise sufficiently in regions that are not sparse. $\tau(i)$ is determined by standard deviation of noise and sparsity as

$$\tau(i) = \frac{\sigma^2}{s(i)}.$$  

(13)

2.4.5 Image Reconstruction

Denoised blocks are reconstructed with linear combination of basis vectors using denoised coefficients as,

$$\hat{V}(j) = \sum_{k=1}^{N^2} \hat{a}_k(i)B_k(i).$$

(14)

Replacing $V(j)$ to $\hat{V}(j)$, circumjacent blocks are denoised.

Iterating the above process, random noise is removed from an input image.
Table 1  Comparison of PSNRs and SSIMs. Best results are in bold.

<table>
<thead>
<tr>
<th>Test image</th>
<th>Noise $\sigma$</th>
<th>PSNR [dB]</th>
<th>SSIM [%]</th>
<th>PSNR [dB]</th>
<th>SSIM [%]</th>
<th>PSNR [dB]</th>
<th>SSIM [%]</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Input</td>
<td>CBM3D [5]</td>
<td>Proposed</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(A) Dog.</td>
<td>15</td>
<td>24.64</td>
<td>67.39</td>
<td>31.79</td>
<td>90.79</td>
<td>32.09</td>
<td>91.45</td>
</tr>
<tr>
<td></td>
<td>20</td>
<td>22.19</td>
<td>58.16</td>
<td>30.33</td>
<td>87.85</td>
<td>30.56</td>
<td>88.59</td>
</tr>
<tr>
<td></td>
<td>25</td>
<td>20.31</td>
<td>51.06</td>
<td>29.24</td>
<td>85.16</td>
<td>29.47</td>
<td>86.08</td>
</tr>
<tr>
<td></td>
<td>30</td>
<td>18.81</td>
<td>45.57</td>
<td>28.32</td>
<td>82.47</td>
<td>28.54</td>
<td>83.38</td>
</tr>
<tr>
<td>(B) Bridge.</td>
<td>15</td>
<td>24.70</td>
<td>60.79</td>
<td>33.33</td>
<td>90.81</td>
<td>33.32</td>
<td>91.25</td>
</tr>
<tr>
<td></td>
<td>20</td>
<td>22.28</td>
<td>51.49</td>
<td>31.92</td>
<td>87.96</td>
<td>31.93</td>
<td>88.48</td>
</tr>
<tr>
<td></td>
<td>25</td>
<td>20.44</td>
<td>44.92</td>
<td>30.91</td>
<td>85.61</td>
<td>30.86</td>
<td>85.33</td>
</tr>
<tr>
<td></td>
<td>30</td>
<td>18.97</td>
<td>40.19</td>
<td>30.02</td>
<td>83.41</td>
<td>29.91</td>
<td>83.59</td>
</tr>
<tr>
<td>(C) Beach.</td>
<td>15</td>
<td>24.81</td>
<td>73.01</td>
<td>30.27</td>
<td>92.13</td>
<td>30.40</td>
<td>92.34</td>
</tr>
<tr>
<td></td>
<td>20</td>
<td>22.36</td>
<td>65.33</td>
<td>28.45</td>
<td>89.30</td>
<td>28.65</td>
<td>89.64</td>
</tr>
<tr>
<td></td>
<td>25</td>
<td>20.52</td>
<td>59.10</td>
<td>27.01</td>
<td>86.09</td>
<td>27.20</td>
<td>86.50</td>
</tr>
<tr>
<td></td>
<td>30</td>
<td>19.02</td>
<td>53.91</td>
<td>25.77</td>
<td>82.16</td>
<td>26.09</td>
<td>83.32</td>
</tr>
<tr>
<td>(D) Sea.</td>
<td>15</td>
<td>24.74</td>
<td>50.31</td>
<td>35.82</td>
<td>92.58</td>
<td>36.10</td>
<td>92.97</td>
</tr>
<tr>
<td></td>
<td>20</td>
<td>22.27</td>
<td>41.41</td>
<td>34.51</td>
<td>90.78</td>
<td>34.63</td>
<td>90.94</td>
</tr>
<tr>
<td></td>
<td>25</td>
<td>20.40</td>
<td>35.81</td>
<td>33.39</td>
<td>89.09</td>
<td>33.39</td>
<td>88.94</td>
</tr>
<tr>
<td></td>
<td>30</td>
<td>18.88</td>
<td>31.96</td>
<td>32.42</td>
<td>87.52</td>
<td>32.29</td>
<td>86.97</td>
</tr>
<tr>
<td>(E) Island.</td>
<td>15</td>
<td>24.69</td>
<td>71.77</td>
<td>30.85</td>
<td>90.47</td>
<td>31.06</td>
<td>91.16</td>
</tr>
<tr>
<td></td>
<td>20</td>
<td>22.27</td>
<td>62.50</td>
<td>29.27</td>
<td>86.65</td>
<td>29.51</td>
<td>87.77</td>
</tr>
<tr>
<td></td>
<td>25</td>
<td>20.42</td>
<td>55.09</td>
<td>28.11</td>
<td>83.16</td>
<td>28.29</td>
<td>84.48</td>
</tr>
<tr>
<td></td>
<td>30</td>
<td>18.93</td>
<td>49.28</td>
<td>27.16</td>
<td>79.74</td>
<td>27.30</td>
<td>80.99</td>
</tr>
<tr>
<td>(F) Sunset.</td>
<td>15</td>
<td>24.95</td>
<td>40.20</td>
<td>35.58</td>
<td>91.49</td>
<td>35.68</td>
<td>91.54</td>
</tr>
<tr>
<td></td>
<td>20</td>
<td>22.48</td>
<td>40.66</td>
<td>33.87</td>
<td>89.20</td>
<td>33.94</td>
<td>89.09</td>
</tr>
<tr>
<td></td>
<td>25</td>
<td>20.57</td>
<td>35.21</td>
<td>32.53</td>
<td>87.36</td>
<td>32.53</td>
<td>86.98</td>
</tr>
<tr>
<td></td>
<td>30</td>
<td>19.05</td>
<td>31.51</td>
<td>31.38</td>
<td>85.76</td>
<td>31.35</td>
<td>85.03</td>
</tr>
<tr>
<td>(G) Squirrel.</td>
<td>15</td>
<td>24.67</td>
<td>55.94</td>
<td>35.14</td>
<td>92.75</td>
<td>35.26</td>
<td>93.03</td>
</tr>
<tr>
<td></td>
<td>20</td>
<td>22.21</td>
<td>46.33</td>
<td>33.87</td>
<td>91.02</td>
<td>33.96</td>
<td>91.20</td>
</tr>
<tr>
<td></td>
<td>25</td>
<td>20.32</td>
<td>39.78</td>
<td>32.85</td>
<td>89.43</td>
<td>32.93</td>
<td>89.67</td>
</tr>
<tr>
<td></td>
<td>30</td>
<td>18.80</td>
<td>35.21</td>
<td>31.99</td>
<td>88.01</td>
<td>32.02</td>
<td>88.17</td>
</tr>
<tr>
<td>(H) Mountain.</td>
<td>15</td>
<td>24.88</td>
<td>71.82</td>
<td>31.54</td>
<td>91.46</td>
<td>31.74</td>
<td>92.11</td>
</tr>
<tr>
<td></td>
<td>20</td>
<td>22.46</td>
<td>62.14</td>
<td>29.82</td>
<td>87.46</td>
<td>29.98</td>
<td>88.45</td>
</tr>
<tr>
<td></td>
<td>25</td>
<td>20.61</td>
<td>54.36</td>
<td>28.54</td>
<td>83.38</td>
<td>28.67</td>
<td>84.74</td>
</tr>
<tr>
<td></td>
<td>30</td>
<td>19.13</td>
<td>48.31</td>
<td>27.48</td>
<td>79.53</td>
<td>27.60</td>
<td>80.98</td>
</tr>
<tr>
<td>Average.</td>
<td>15</td>
<td>24.76</td>
<td>62.53</td>
<td>33.04</td>
<td>91.56</td>
<td>33.21</td>
<td>91.98</td>
</tr>
<tr>
<td></td>
<td>20</td>
<td>22.32</td>
<td>53.50</td>
<td>31.50</td>
<td>88.78</td>
<td>31.64</td>
<td>89.27</td>
</tr>
<tr>
<td></td>
<td>25</td>
<td>20.48</td>
<td>46.92</td>
<td>30.32</td>
<td>86.16</td>
<td>30.42</td>
<td>86.59</td>
</tr>
<tr>
<td></td>
<td>30</td>
<td>18.95</td>
<td>41.99</td>
<td>29.32</td>
<td>83.57</td>
<td>29.39</td>
<td>84.05</td>
</tr>
</tbody>
</table>

Fig. 5  Comparison of resultant images for (A) Dog with noise $\sigma = 25$.

Fig. 6  Comparison of resultant images for (B) Bridge with noise $\sigma = 25$. 
3. Experiments

We experimented with denoising from noisy input images. On the extreme left in Table 1, test images used in the experiments are shown. Gaussian noises are added to these images for input images. Input images are denoised by CBM3D (BM3D for color image) [5] and the proposed method.

Figure 4 shows an example set of an input image with noise $\sigma = 20$ and calculated sparsity degrees. As shown in Fig. 4, in regions where some sort of image structure exists (for example, edges), coefficients are sparse.

Table 1 shows comparisons of PSNR (Peak Signal to Noise Ratio) [dB] and SSIM (Structure Similarity) [%] [9] values between noise-free images and each resultant image. In Table 1, best results are marked as bold. On average, the proposed method obtained higher PSNR and SSIM values. However, for some images, CBM3D obtained higher PSNR or SSIM values.

Figure 5 shows a comparison of resultant images for (A) Dog with noise $\sigma = 25$. With respect to (A) Dog, the proposed method obtained higher PSNR and SSIM values compared with CBM3D.

As shown in Fig. 5, CBM3D blurs textures of images. The proposed method removes noise from an input image without causing blur.

Figure 6 shows a comparison of resultant images for (B) Bridge with noise $\sigma = 25$. With respect to (B) Bridge, CBM3D obtained higher PSNR and SSIM values. As shown in Fig. 6, the proposed method cannot remove noise sufficiently in flat regions. The proposed method is a process to extract image structure signals from noisy blocks by degrees. However, in flat regions, image structures do not exist. Therefore, it is considered that the proposed method does not work well. This is an issue to be addressed in future work.

4. Conclusion

We proposed a new effective shrinkage method for image denoising in this paper. In the proposed method, image blocks are projected to a space that makes projection coefficients sparse, and explicitly evaluated sparsity degree is used to control the shrinkage threshold. On average, the proposed method obtained higher PSNRs or SSIMs compared with one of the state-of-the-art methods in the field of image denoising. An issue was also identified concerning flat regions in images, but the proposed method effectively removes random noise from natural images with intricate textures.

References


(Communicated by Yasuhiro Mukaigawa)