Wireless Multimedia Sensor Network Image De-Noising via a Detail-Preserving Sparse Model

Zhi Cui, Xian-pu Cui

School of Communication and Electronic Engineering, Hunan City University, Yiyang, China
Emails: zhicui@yeah.net cuixianpu@sohu.com

Abstract: In this paper, we propose a Detail-Preserving Sparse Model (DPSM) for de-noising of images that are usually interfered by noise on the Wireless Multimedia Sensor Network (WMSN). Specifically, based on the Structural SIMilarity (SSIM), the DPSM first incorporates a structural-preserving constraint, which enables the structure in the reconstructed image to be close to the ideal no-noise image. In addition, the DPSM adopts a residual ratio as the stopping condition of the sparse solution algorithm (e.g., Orthogonal Matching Pursuit), which enables the structures to be reconstructed under high noise conditions. The experimental results on several WMSN images have demonstrated the superiority of the proposed DPSM method over several well-known de-noising approaches in terms of PSNR and SSIM.

Keywords: Structural similarity, residual ratio, wireless multimedia sensor network, sparse representation, de-noising.

1. Introduction

A Wireless Multimedia Sensor Network (WMSN) is a distributed sensing monitoring network which is composed of new types of sensor nodes with video, audio and image perception sensing functions. It has broad applications, such as real-time traffic monitoring and industrial process control [1]. Due to the diversity of the objects it monitors also the complexity of the monitoring environments. A WMSN image usually contains abundant details. At the same time, WMSN images collected during bad weather usually contain high levels of noise, which seriously degrade the image quality [2]. Therefore, the de-noising of WMSN images is a hot topic [3, 4].

Transform domain based methods have been widely used for WMSN image
de-noising in the past [5-7]. It assumes that the noise in the image is mainly
distributed in the high frequency component, while the image’s contents are
distributed in the low frequency component. Therefore, the goal of de-noising can
be achieved by restraining or removing the high frequency component [8, 9].
However, a large number of experiments on WMSN images collected under real-
world conditions show that available information related to the image content could
be detected in the high frequency part of the spectrum, and that the noise
components also existed in the low frequency component [10-13]. This means that
when using transform domain-based de-noising methods, there is a risk that some
meaningful information located in the high frequency component might be filtered
out as noise, while the noise in the low frequency component is not always
efficiently removed [14, 15].

In recent years sparse representation has attracted increasing attention in the
field of image de-noising [16-18]. Compared with transform domain-based image
de-noising, the sparse representation theory assumes that the information available
in the images has specific structural features, strongly correlated to the structure of
the atoms of a dictionary, while the noise has not such a feature. Therefore, we can
separate the signal from noise according to whether the data has a sparse
representation in the dictionary or not [19, 20]. Recently, various approaches based
on sparse representation have been proposed to address the problem of image de-
noising. Zhang and Xie [21] proposed a de-noising method based on DCT and
sparse representation. In this method the image’s content could be efficiently
represented by an over-complete dictionary learned from the noisy image itself. But
the dictionary contains a large number of atoms that have nothing to do with the
image structure, which are causing reduction in the efficiency of image sparse
representation. In addition, Zhang et al. [22] proposed a de-noising method based
on adaptive sparse representation. In this method the K-SVD algorithm has been
used to learn an over-complete dictionary from the test image itself. This method is
sensitive to noise and therefore when the image contains large levels of noise, it can
cause fluctuations in the reconstructed image. Zhou and Luo [23] proposed a
novel K-Least Mean Square (K-LMS) algorithm for learning the over-complete
dictionary. Nevertheless, the fixed step length adopted in the decomposition process
of LMS algorithm could cause a large level of steady-state errors.

In most of the current sparse based de-noising methods, the reconstruction
error between the images before and after de-noising is considered as a fidelity term
and a hard threshold is adopted as iteration termination conditions. Due to the
complex work environment, the WMSN images contain rich details and low signal-
to-noise ratio. If we continue using the above criteria, it would be difficult to set a
threshold and the noise components introduced in the reconstructing process will
have a larger impact on the image reconstruction accuracy. In this paper we propose
a new sparsity based method for WMSN image de-noising. Motivated by the works
in [17], our method has two improvements. The first one is to replace the
reconstruction error with the Structural SIMilarity (SSIM) as a fidelity term. We
propose this because the structure similarity can conform to the characteristics of a
Human Visual System (HVS) and thus tend to produce better results that agree with
the HVS [24]. The second improvement is to adaptively tune the iteration number. Especially, instead of using a constant iteration number [17], the method proposed would determine the iteration number adaptively according to the residual ratio between the adjacent sparse decompositions.

The rest of the paper is organized as follows. The description of sparse representation theory is briefly presented in Section 2. The proposed method is described in Section 3. The experimental results and comparisons are given in Section 4. The conclusions are drawn in Section 5.

2. Sparse representing de-noising model

Research results have shown that the WMSN image contains a large number of detailed features, such as edges and mutation, while the local small image patch appears simple and has a consistent structure [25]. For this reason, we establish the de-noising model for local image blocks \( x \) instead of the whole image.

Let \( x \in \mathbb{R}^n \) be a \( \sqrt{n} \times \sqrt{n} \) \((n \ll N)\) image patch extracted from an image \((\text{of size } \sqrt{N} \times \sqrt{N})\) and let us define a redundant dictionary as \( D \in \mathbb{R}^{nk} \) \((n < k)\), the image patch can be represented as

\[
x = D\alpha,
\]

where \( \alpha \in \mathbb{R}^n \) is a matrix of the sparse coefficient that can be obtained by solving the following optimization problem:

\[
\hat{\alpha} = \arg \min_{\alpha} \|\alpha\|_0 \quad \text{s.t.} \quad \|D\alpha - x\|_2^2 \leq \varepsilon.
\]

Here \( \|\cdot\| \) is the \( l_0 \) norm, \( \|\alpha\|_0 \) is the number of the non-zero values, and \( \varepsilon \) represents the error tolerance. Equation (2) can also be changed to the following regularization optimization problem:

\[
\hat{\alpha} = \arg \min_{\alpha} \|\alpha\|_1 + r \|\alpha\|_1,
\]

where \( r \) is the regularization parameter. While solving (3) is an NP-hard problem, several strategies, such as basis pursuit [26] and OMP [27] are developed for approximating the solution of (3). The final de-noised model can be described as follows:

\[
\{\hat{\alpha}_y, \hat{Y}\} = \arg \min_{\alpha, Y} \lambda \|Y - X\|_2^2 + \sum_{i,j} \|\alpha_{ij}\|_1 + \sum_{i,j} \|D\alpha_{ij} - R_{ij} Y\|_2^2,
\]

where \( X \) is the noisy image, \( Y \) is the ideal image, \( \lambda \) is the Lagrange multiplier, \( r_{ij} \) is the regularization parameter, \( \alpha_{ij} \) is the sparse coefficient of the image with size of \( \sqrt{N} \times \sqrt{N} \) in \((i, j)\), and \( R_{ij} \) is a \( n \times n \) matrix to extract the image block in \((i, j)\) from the image of \( \sqrt{N} \times \sqrt{N} \). The first item in the right side represents the overall similarity degree between the noised image and the ideal image, the second item is the sparse constraint, and the third one is the reconstruction error with the constraint \( \|D\alpha_{ij} - R_{ij} Y\|_2^2 < c\sigma^2 \), where \( c \) is the noise factor associated with the dimensions of the image block and \( \sigma \) is the noise variance.
3. The detail-preserving sparse model for WMSN image de-noising

In this section, we describe the proposed method. The core of our method is to replace the reconstruction error with the structure similarity and make it a new fidelity term. In addition, instead of the residual hard threshold, we utilize the residual ratio as a new iteration stopping condition.

3.1. Structural similarity

SSIM is a new kind of standard for image visual quality evaluation. It measures the quality of the image in three aspects, such as brightness, contrast and structure. The SSIM can be described as follows:

\[
S(a, b) = \frac{(2\gamma_a\gamma_b + c_1)(2\sigma_a\sigma_b + c_2)}{(\gamma_a^2 + \gamma_b^2 + c_1)(\sigma_a^2 + \sigma_b^2 + c_2)},
\]

where \(\gamma_a, \gamma_b\) are the means of the noised image and the ideal image, \(\sigma_a, \sigma_b\) are the variances, \(c_1, c_2\) and \(c_3\) are the minimal positive constants related to the values of the pixel. SSIM measures the image quality for brightness, contrast and structure, which is more in line with the characteristics of the human visual system. It has a value between 0 and 1, the closer it is to 1, the more similar in structure are the noised and the de-noised images.

3.2. Residual ratio

In general, sparse coding is proceeded iteratively, so that the stop rule is important to keep useful information and de-noising. Recently, researchers mostly use the following two methods to determine the stop rules of OMP iteration. The first one is using a hard threshold, in which the maximum number of the iterations is chosen empirically and the residual is filtered as noise [28, 29]. It is hard to set the iterations number exactly. The smaller the process is, the more information is lost, otherwise it would add noise. The other method is based on the residual signal energy. The iterations would be stopped when the radius of the residual energy is below a certain threshold [16, 17, 30]. In [17] the threshold was set to \(cs^2\), where \(c\) is the noise factor, and \(s\) is the noise variance. However, when the signal dimension and the noise levels increase, the energy radius can enlarge quickly, the iteration will be prematurely terminated that might cause loss of some information. Considering that the above two methods cannot be appropriate for WMSN images, which are collected under condition of high intensity noise, we propose a new termination condition based on the residual ratio.

Assume that \(X_j\) is the \(j\)-th column of image \(X\), it can be expressed as follows when the decomposition is processed for \(m\)-th time:

\[
X_j = \bar{X}_j + R_m(X_j),
\]

where \(\bar{X}_j\) represents the useful information of the image, \(R_m(X_j)\) represents the residual of the \(m\)-th decomposition.
The residual ratio $\kappa$ can be described as follows:

\begin{equation}
\kappa = \frac{\| R_{n+1} X_j - \beta R_n X \|_2}{\beta R_n X \|_2},
\end{equation}

where $\beta$ is the normalization coefficient with the constraint $\beta = \sqrt{E(R_{n+1} X_j)^2 / E(R_n X_j)^2}$ and $E$ represents the solution of the expectation value.

It is clear from the foregoing, that the atoms which maximize $\| R_{n+1} X_j - \beta R_n X \|_2$ would be selected during the $m$-th iteration, i.e., the atoms which make the residual ratio approach to the greatest value would be selected. With the useful signal gradual extraction, the residual ratio of the maximum amplitude is gradually reduced. When the residual ratio reaches a minimum value, the residual signal would no longer contain any useful information and the OMP must be stopped in time.

3.3. The method proposed

On the basis of the above analysis, the Detailed-Preserving Sparse Model can be expressed as follows:

\begin{equation}
\hat{\alpha}_{ij} = \arg \min_{\alpha_{ij}} Y - X + \sum_{i,j} 1 - S(D\hat{\alpha}_{ij}, R_{ij} Y),
\end{equation}

where $\hat{\alpha}_{ij}$ is the coefficient of the sparse representation, the third item in the right side is the new fidelity term, $S(,)$ shows the calculation of the SSIM. The first term on the right side measures the overall proximity between $Y$ and $X$, the second part as a regular term reflects the sparse constraint, and the third one is the fidelity term which replaces the reconstruction error of K-SVD measuring the consistency between the reconstructed image and the ideal image. Here we have employed SSIM as a fidelity term.

From the above analysis, the specific process of our method is summarized in four stages.

(1) The initial stage

The over-completed DCT dictionary is adopted as an initial dictionary, satisfying $Y = X$.

(2) The stage of sparse coding

According to the proposed model, the structural similarity is incorporated into the OMP algorithm and each image patch is sparsely decomposed. The modified method which incorporates SSIM into the procedure is defined as follows:

\begin{equation}
\hat{\alpha}_{ij} = \arg \min_{\alpha} 1 - S(D\alpha_{ij}, R_{ij} Y).
\end{equation}

It can be seen that the role of $\|D\alpha_{ij} - R_{ij} Y\|_2$ is replaced by $1 - S(D\alpha_{ij}, R_{ij} Y)$. We solve the optimization problem by using the OMP, gathering one atom at a time, and stopping when $S(D\alpha_{ij}, R_{ij} Y)$ goes above $T$ which is given as follows:

\begin{equation}
T = 1 - \kappa,
\end{equation}
where $\kappa$ is the residual ratio. Subsequently the sparse coefficients vectors $\hat{\alpha}_y$ can be obtained.

3) The stage of updating a dictionary
Given all $\hat{\alpha}_y$, we can now fix them and turn to updating the initial dictionary column by a column using K-SVD. Assuming that $D$ is known similarly, $d_j$ is the $j$th column to be updated in the dictionary. Corresponding to $d_j$, $\alpha_j^T$ is the $j$th row in the sparse coefficient matrix. For each column $j = 1, 2, \ldots, k$ in $D$, it is updated by the following model:

$$\min_{\alpha_j}(1-S(Y, \sum_{j=1}^k d_j \alpha_j^T)) \quad \text{s.t.} \quad \alpha_j^T \in \omega.$$ (11)

After the above procedure, a new dictionary $\hat{D}$ will be formed, and along with this update, the representation coefficients change as well.

4) The stage of outputting results
After getting the new dictionary $\hat{D}$, we can get the approximate solution of the ideal image through Equation (12):

$$\hat{Y} = \arg \min_Y \lambda \|Y - X\|^2 + \sum_{i,j} \|D \alpha_y - R_y Y \|^2.$$ (12)

where $\hat{Y}$ is the output de-noised image. Solving it for the partial derivative of $Y$, we can get a closed-form solution of the form as follows:

$$\hat{Y} = (\lambda I + \sum_y R_y^T R_y)^{-1} (\lambda X + \sum_y R_y^T D \hat{\alpha}_y).$$ (13)

The graphic demonstration of the overall de-noising process is presented in Fig. 1.

4. Experimental results and analysis

In this section, in order to test the performance of the proposed method, we have used four different methods for comparison. The parameters of the proposed method are set as: $J = 10$, $\kappa = 0.03$ and $n = 8 \times 8$. As in reference [17], we set $\lambda$ and
c to be $30/\sigma$ and 1.15, respectively. For the compared methods, the parameters are set to the default value in order to reach their best results. In our experiments PSNR and SSIM were adopted as objective indices to evaluate the quality of de-noised images. All the experiments are implemented on a Core i5(R) 3.2 GHz PC with 8 GB RAM.

In the first experiment, we compared the SSIM result of the proposed method with those of Wavelet Transformations using a Statistical Model (WT-SM) [31], sparse representation using DCT dictionary (SR-DCT) [30], K-SVD [17] and the method proposed by Zhou and Luo [23]. The tests were performed on three images, which are given in Fig. 2: “Parking lot”, “Crossroad” and “Garage”. Each image has a size of 256×256 pixels, the horizontal and vertical resolution is 300 dpi and the bit depth is 8. White noise with variations of 25, 35 and 55 were added in the images. The SSIM results are given in Fig. 3.

![Fig. 2. The original images: Parking lot (a); crossroad (b); garage (c)](image)

![Fig. 3. The SSIM results obtained by five compared methods on three test images: Results on “Parking lot” image (a); results on “Crossroad” image (b); results on “Garage” image (c)](image)

As it can be observed in Fig. 3, the proposed method outperforms the other methods on all SSIM results for all three images. For instance, by adding white noise with variance of 35 to the Garage image, the SSIM result obtained by the proposed method is 0.809, with a gain of 0.125, 0.086, 0.064, and 0.063 respectively over the values obtained by the WT-SM, SR-DCT, K-SVD and the method in [23]. It can also be found that the results obtained by the proposed method are better than the others for different images, with an average rise at 0.111, 0.087, 0.053 and 0.044 in average, respectively in WT-SM, SR-DCT, K-SVD and
the method in [23]. In addition, when the images were de-noised by using the compared methods, some important details in the images would disappear together with the noise. Our method de-noises images by using a detailed-preserving sparse model, which can help to efficiently keep the structure characteristics, and meanwhile the useful information in WMSN images can be better retained.

In the second experiment we compared the PSNR result obtained by the proposed method with those of the other four methods. The tests were performed on the same images: “Parking lot”, “Crossroad” and “Garage”. AWGN is superimposed on them with 25, 35 and 55. The PSNR results are shown in Tables 1, 2 and 3. It can be found that the results obtained by the method proposed are better than those of the other methods for the same image. For instance, by adding white noise with variance of 35 to the “Garage”, the PSNR result obtained by the proposed method is 27.15 dB, with an increased value by 3.58 dB, 2.44 dB, 0.50 dB and 0.19 dB respectively compared to the values in WT-SM, SR-DCT, K-SVD and the method in [23]. We also found that the results obtained by the proposed method are better than the others for different images. For instance, the PSNR result obtained by the proposed method is increased by 3.51 dB, 2.56 dB, 0.53 dB and 0.36 dB on average, respectively to WT-SM, SR-DCT, K-SVD and the method in [23].

Table 1. Performance of the de-noised methods of “Parking lot” by PSNR (dB)

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>25</td>
<td>28.93</td>
<td>30.07</td>
<td>32.45</td>
<td>32.45</td>
<td>32.65</td>
</tr>
<tr>
<td>35</td>
<td>28.45</td>
<td>29.52</td>
<td>31.69</td>
<td>31.86</td>
<td>32.12</td>
</tr>
</tbody>
</table>

As a final experiment, we compared the de-noised effects achieved by the proposed method, K-SVD and the method in [23], because they use similar techniques. Figs 4 and 5 both offer visual comparison between them. From the visual results one can see that the proposed method can preserve the structural features better than the other methods.

Table 2. Performance of the de-noised methods of “Crossroad” by PSNR (dB)

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>25</td>
<td>26.87</td>
<td>27.95</td>
<td>30.18</td>
<td>30.30</td>
<td>30.69</td>
</tr>
<tr>
<td>35</td>
<td>26.01</td>
<td>27.04</td>
<td>29.17</td>
<td>29.26</td>
<td>29.63</td>
</tr>
<tr>
<td>55</td>
<td>21.44</td>
<td>22.16</td>
<td>24.14</td>
<td>24.19</td>
<td>24.68</td>
</tr>
</tbody>
</table>

Table 3. Performance of the de-noised methods of “Garage” by PSNR (dB)

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>25</td>
<td>25.04</td>
<td>26.20</td>
<td>28.46</td>
<td>28.54</td>
<td>28.95</td>
</tr>
<tr>
<td>35</td>
<td>23.57</td>
<td>24.71</td>
<td>26.65</td>
<td>26.96</td>
<td>27.15</td>
</tr>
<tr>
<td>55</td>
<td>19.58</td>
<td>19.87</td>
<td>21.27</td>
<td>21.65</td>
<td>22.13</td>
</tr>
</tbody>
</table>
Fig. 4. Visual comparison of the de-noising results on three images (Parking lot, Crossroad and Garage), with $\sigma = 75$. The first row: De-noising results by using WT-SM. The second row: De-noising results by using SR-DCT. The third row: De-noising results by using K-SVD. The fourth row: De-noising results by using the method in [23]. The final row: De-noising results obtained by the proposed method.
Fig. 5. Detail comparison of the de-noising results on three images (Parking lot, Crossroad and Garage), with $\sigma = 75$. The first row: De-noising results by using WT-SM. The second row: De-noising results by using SR-DCT. The third row: De-noising results by using K-SVD. The fourth row: De-noising results by using the method in [23]. The final row: De-noising results obtained by the proposed method.
The reasons giving our proposed method better performance than the other four algorithms can be summarized as follows: First, the de-noising based on sparse representation can efficiently distinguish the useful information and the noise. Second, since SSIM can accurately evaluate the perception quality of the image, the proposed method uses SSIM as a new fidelity term, which has some advantages in terms of the image structure characteristics. Third, the proposed method uses the residual ratio as an iteration termination condition and therefore avoids the defect of a hard threshold which can introduce noise in the process of iterations. Because of this it achieves a better de-noising effect than the other methods.

5. Conclusion

In this paper we have proposed an improved de-noising method for WMSN image de-noising. The proposed method can overcome the drawbacks of the traditional methods based on sparse representation. We take SSIM as a fidelity term and utilize the residual ratio as an iteration termination condition, so that the de-noised image can be efficiently reconstructed by this Detail-Preserving Sparse Model. The experimental results have shown that the method suggested has better de-noising performance than some other methods in terms of SSIM and PSNR. In the future, we will focus on the dictionary learning method and the low-complexity sparse recovery method to improve the efficiency of the proposed method.

Acknowledgments: This work is supported by the Scientific Research Fund of Hunan Provincial Education Department under Grant No 13C122, and the Natural Science Foundation of the Program in Hunan Province of China under Grant No 13JJ6072. The authors would like to thank the anonymous reviewers for their insightful comments and suggestions, which have greatly improved this paper.

References