Evolution-enhanced multiscale overcomplete dictionaries learning for image denoising

Shuyuan Yang \textsuperscript{a,}\textsuperscript{*}, Min Wang \textsuperscript{b}, Meirong Wei \textsuperscript{a}, Licheng Jiao \textsuperscript{a}

\textsuperscript{a} Key Lab of Intelligent Perception and Image Understanding of Ministry of Education, Xidian University, Xi’an 710071, China  
\textsuperscript{b} National Key Lab of Radar Signal Processing, Department of Electrical Engineering, Xidian University, Xi’an 710071, China

1. Introduction

Overcomplete (or redundancy) is important in transformation-based image denoising methods to have the shift invariance property (Coifman and Donoho, 1994; Bui and Chen, 1998). For example, with the growing realization of deficiencies of orthogonal wavelets in denoising images, some redundant multiscale transforms have been introduced, including undecimated Wavelets (Lang et al., 1996), Curvelet (Starck et al., 2002), Contourlet (Do and Vetterli, 2002), Wedgelet (Demaret et al., 2005), Bandelet (Zhang et al., 2010), Shearlet (Blu and Luisier, 2007) and so on. In the past decade, using spatial overcomplete representation and sparsity for images denoising has drawn much attention of researches (Elad et al., 2006; Elad and Aharon, 2006; Elad and Aharon, 2006). Its basic idea is that the sparse representation (SR) of images will help in automatically selecting the primary components in images while reducing the noise components, as long as the dictionary can well describe the characteristics of images. In more recent works (Aharon et al., 2006; Proctor and Elad, 2009; Chatterjee and Milanfar, 2009; Turek et al., 2010; Dong et al., 2011), image patches prove to well represent the statistical properties of the whole image, so a large number of image patches are taken as the examples from which a dictionary can be learned. The patches are taken from the noisy image and then sparsely represented and restored, which lead to state-of-the-art denoising result.

Although the SR-based denoising methods have proved to work well on the natural images (Aharon et al., 2006; Proctor and Elad, 2009; Chatterjee and Milanfar, 2009; Turek et al., 2010; Dong et al., 2011), the SR is all executed in the spatial domain. On the other hand, it has been aware that making avail of the multiscale properties of images will obtain better denoising result (Bui and Chen, 1998; Lang et al., 1996; Starck et al., 2002; Do and Vetterli, 2002; Demaret et al., 2005; Zhang et al., 2010; Blu and Luisier, 2007). Therefore, in this study we take both the overcomplete representations of images in spatial domain and transformation domain into account, and propose a multiscale dictionaries learning approach for image denoising. Some multiscale overcomplete dictionaries are learned from example patches, and then used to reduce the noise distributed in the different scales of images. We reduce the image denoising to an \( l_0 \)-norm minimization problem with multiple variables. The available optimization schemes for this NP-hard problem can be mainly divided into two catalogs: approximation method and relaxation method. The approximation method includes greedy algorithms and shrinkage algorithms. A greedy strategy abandons exhaustive search in favor of a series of locally optimal single-term updates. Its basic idea is to represent a signal as a weighted sum of atoms taken from a dictionary, such as matching pursuit (Mallat and Zhang, 1993), orthogonal matching pursuit (Tropp and Gilbert, 2007) and their variants (Donoho et al., 2006; Needell, 2009). The approximation method can correctly pick up atoms in the case of existing sparse solution and the selection rule is simple to understand. However, it is characteristics of heavy computation, slow convergence, and can only work well in the noiseless case. The shrinkage strategy iterates between shrinkage/thresholding operation and projection onto
perfect reconstruction, so they are characteristic of low computation complexity (Chen et al., 2001; Bioucas-Dias and Figueiredo, 2007). However, they commonly require much iteration when the support of the solutions cannot be determined. The relaxation method includes $l_1$-norm and $l_p$-norm methods. Basis pursuit (BP) (Blumensath and Davies, 2008) approximate the solution that minimizes $l_1$-norm and reduce the problem to a linear programming (LP) structure, which is solvable comparing to the $l_p$-norm minimization and is easy to be integrated into other variational model. However, it is a difficult optimization task and the tuning of parameter is not straightforward. Moreover, the equivalence of $l_p$-norm and $l_1$-norm minimization can only achieve under very strict assumption of the sparsity of signals (Candès and Wakin, 2008). Gradient based methods are discussed in paper (Figueiredo et al., 2007) and (Blumensath and Davies, 2008) to solve this problem. The $l_p$-norm ($0 < p < 1$) or the weak $l_p$-norm is a popular measure of sparsity used by the mathematical analysis community, so it is used to serve as a candidate function for $l_p$-norm (Candès et al., 2008). Although it is still non-convex, it is almost equivalent to $l_p$-norm and can be represented as a weighted $l_1$-norm form by the iterative-reweighed-least-squares (IRLS) method. However, this algorithm is very sensitive to the initialization of solution. Moreover, it is guaranteed to converge to a fixed-point that is not necessarily the optimal one.

Evolutionary algorithms (EAs) provide a general and global searching approach for solving combinatorial and NP-hard optimization tasks (De Jong, 2006), so in this paper we use EAs to solve the $l_p$-norm minimization problem discussed above. Genetic Algorithm (GA) is one of the effective EAs that simulate natural evolution (crossover, mutation and selection) over populations of candidate solution (Goldberg, 1989). However, GA is characteristic of slow convergence (Alberto and Carlos, 2003). The memetic algorithm (MA) (Alberto and Carlos, 2003; Badillo et al., 2011; Amaya et al., 2010; Krasnogor and Smith 2005) makes an improvement on GA by combining GA with a local searching operation, and proves to perform much better than GA in terms of the quality of solution and computational cost. In the MA, GA is used for coarse search, while the subsequent local improvement is then used to refine the GA. Its superior performance over GA has been found for various applications, such as combinatorial optimization problems (Tang et al., 2007), control design (Caponio et al., 2007), VLSI design (Tang and Yao, 2007), image segmentation (Jiao et al., 2010) and so on.

In order to tune multiple variables in the optimization problem discussed above, in our study a MA based alternate optimization strategy is employed to optimize the dictionaries and denoise the multiscale images. In the algorithm, two-dimensional individual is adopted to represent a dictionary. The individuals are used to perform a global search, followed by a local search operator, singular value decomposition (SVD), to further reduce the objective function. The dictionaries and sparse coefficients are alternately updated until the stop condition is satisfied. Some experiments are taken on some benchmark natural images to investigate the performance of our proposed method.

The rest of this paper is organized as follows. In Section 2, we addressed the classic image denoising problem, and depicted the memetic algorithm-enhanced multiscale dictionaries learning algorithm. In Section 3, some simulation experiments are taken to illustrate the efficiency and superiority of our proposed method to its counterparts. Finally some conclusions are drawn in Section 4.

2. Evolution-enhanced multiscale dictionaries learning

Considering the classic image denoising problem: an image is measured in the presence of an additive zero-mean white and homogeneous Gaussian noise, with standard deviation $\sigma$. Thus the measured image is $Y = X + n$, and the goal of image denoising is to recover the clean image $X$ from the noisy image $Y$.

2.1. A multiscale overcomplete dictionaries learning

Consider a redundant transformation on the noisy image, and denote $\mathcal{R}$ as the transformation dictionary. The noisy image can be written as,

$$ Y = \mathcal{R}\beta + n $$

(1)

where $\beta$ and $\mathcal{R}$ are the transformation coefficients of the noisy image $Y$ and clean image $X$, respectively. When the transformation has the redundancy and multiscale property, $\mathcal{R}$ is often determined by the frame theory, such as undecimated wavelets frame, or some frame composed by cascade orthogonal bases.

Considering the multiscale property of $\mathcal{R}$, we can reformulate (1) as,

$$ Y = [\mathcal{R}_1, \ldots, \mathcal{R}_N] \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_N \end{bmatrix} + n $$

(2)

where $\mathcal{R}_1, \ldots, \mathcal{R}_N$ are the $N$ multiscale dictionaries and $\beta_1, \ldots, \beta_N$ are the corresponding coefficients. Inspired by the example-based denoising scheme (Chatterjee and Milanfar, 2009), we extract the patches from the multiscale coefficient images $\beta_j (j = 1, \ldots, N)$ (the patches are processed in raster-scan order in $\beta_j$, from left to right and top to bottom). Then we use these patches to train a dictionary that is representative of all the image patches and used to recover $\beta_j$ from $\beta_j^*$. Let $\beta_j^* \in \mathbb{R}^K$ denote the $i$th ($i = 1 \times 2$) local patch vector extracted from the multiscale coefficients matrix $\beta_j$ at the spatial location $i$: $\beta_j^* = [\beta_{ij}]$, where $P_i$ denotes a rectangular windowing operator and the overlapping is allowed. $\beta_j^*$ can represent the $i$th patch in $\beta_j$ with its coefficients being ordered lexicographically as column vector. Assume each patch vector $\beta_j^*$ belongs to the Sparseland signal (Elad, 2010), i.e., $\beta_j^*$ can be represented sparsely under a redundant dictionary $D_j \in \mathbb{R}^{K \times K}$ that contains $K$ prototype signal-atoms for columns $\{d_{ij}\}_{i=1}^K$, that is, $\beta_j^* = D_j x_j \in \mathbb{R}^K$ and $\|x_j\|_0 < K$, the MAP estimator for denoising this coefficient patch is built by solving (Elad, 2010),

$$ \min_{|x_j|_0, \beta_j^*} \sum_{i=1}^K \| \beta_{ij}^* \|_0 \\ s.t. \sum_{i=1}^K \| D_j x_j - P_i \beta_{ij}^* \|_2^2 \leq \varepsilon $$

(3)

where $P_i$ is a patch extraction operator, $Q$ is the number of patches extracted from $\beta_j^*$ and $\varepsilon$ and $\delta$ are dictated by $\sigma$.

Assume the number of the example patches at each scale take the same value: $Q_1 = \ldots = Q_K = Q$, and denote the sparse coefficients of patches in $\beta_j$ under the $j$th multiscale dictionary $D_j \in \mathbb{R}^{Q_j \times K}$ as $x_j = [x_{j1}, x_{j2}, \ldots, x_{jQ_j}] \in \mathbb{R}^{Q_j \times K}$. Let $\Phi = [P_1, P_2, \ldots, P_Q]$, the optimization problem can then be reduced to,

$$ \min_{|x_j|_0, \beta_j^*} \sum_{i=1}^Q \| x_i \|_0, \| \beta_{ij}^* \|_0 \\ s.t. \sum_{i=1}^K \| D_j x_j - \Phi \beta_{ij}^* \|_2^2 \leq \varepsilon $$

$$ \| \beta_j^* - Y \|_2^2 \leq \delta $$

(4)

where $\| x_i \|_0, \| \beta_{ij}^* \|_0$ is the $l_0/l_1$-norm of the matrix $x_i$ and represent the
summarization of the $b_0$—norm of the columns in the matrix $z_i$ and $\beta=[\beta_1,\ldots,\beta_N]^T$. In the formula, there are four variables to be determined. To simplify the problem, we assume that $W$ is an orthonormal matrix (such as standard orthogonal wavelet transform) and $\mathfrak{R}$ is composed of the shifted version of matrix $W$: $\mathfrak{R}=[R_1,\ldots,R_N]=[W,FW,\ldots,F^{N-1}W]$ with $F$ representing the shift operation. So the formula (4) can be rewritten as,

$$\min_{\{\alpha_i D_i, \beta\}} \sum_{j=1}^{N} \|D_i z_j - \Phi \beta\|_2^2 + \lambda \|\beta - \mathfrak{R}^j\|_2^2 + \mu \sum_{j=1}^{N} \|z_j\|_{0,1} \quad (5)$$

KSVD algorithm has been used to solve this form of optimization problem. However, it performs a local searching and cannot assure the convergence of algorithm [Aharon et al., 2006]. In this study, we proposed a memetic algorithm-based optimization method for (5). We use the individuals to code the multiscale dictionaries $\{D_i\}$, the variables are tuned using the steps shown in the Pseudocode of the proposed algorithm. In the initialization, we randomly generate the dictionary population $\text{POP}(0)=\{I_1(0),I_2(0),\ldots,I_M(0)\}$ with $M$ individuals. Each $I_i=[I_{i1}(0),I_{i2}(0),\ldots,I_{iN}(0)]$ is corresponding to the multiscale dictionaries $\{D_i\}$, and $[I_{ij}(0)]_{1\leq i \leq M \leq N}$ is corresponding to $D_i$. In the evolution, define the fitness function of individuals as the reciprocal of the objective function in (5). In the mutation, for each individual, we randomly generate a number between 0 and 1, if it is smaller than the mutation probability $P_{ini}$, then perform a mutation operation on the population using

$$I'(t+1)=c_1 \times I'(t)+c_2 \times \text{Best}(t) \quad 0 \leq c_1, \quad c_2 \leq 1, \quad c_1+c_2=1 \quad (6)$$

where the parameter $c_2$ controls the degree of the individual moving towards the best individual; else perform the following random mutation,

$$I'(t+1)=I'(t)+\gamma \times \text{rand} \quad (7)$$

where $\gamma$ controls the mutation intensity. The mutation in (6) emphasizes the guidance of population by the best individual, and the mutation in (7) emphasizes the randomness of evolution. Experimentally we found that a combination of these two schemes outperforms the sole one. For combining the global searching of evolutionary optimization with the local searching of SVD and OMP algorithm, the optimized multiscale dictionaries can produce more sparse representation of images and better denoising result. The flowchart of the algorithm is shown in Fig. 1.

### 3. Experimental results

In this section, some experiments are taken to investigate the performance of our proposed method. All experiments are executed on a 2.4-GHz Pentium-IV PC with 1-GB RAM. The test natural images are all of size $512 \times 512$ and the test SAR images are of size $256 \times 256$.

#### 3.1. Experiment 1: denoised result of our method on Lena image

Firstly we take $W$ as the wavelet transform (‘haar’ wavelets [Elad, 2010]) to denoise the Lena image with additive zero-mean Gaussian noise (the variance $\sigma=20$). In our experiment, we let $N=10$, $p=64$, $K=150$, $M=20$, $c_1=0.3$, $c_2=0.7$, $\gamma=\mu=0.1$, $Q=10000$, $\lambda=0.1$, $P_{ini}=0.7$. Fig. 2 compares the multiscale coefficients $I_i(0)$ of the noisy image with the multiscale coefficients $I_i(0)$ of the image recovered by our proposed method.

In Fig. 2, $l_1$, $H$, $V$, $D_1$ and $D_2$ represent the low-frequency, horizontal high-frequency, vertical high-frequency and diagonally high-frequency components at the first scale, respectively. From these we can see that the noises in the four components are all reduced remarkably by our method, especially the $D_1$ component. When the noises existed in the high-frequency and low-frequency

---

**Pseudocode of the proposed algorithm:**

1. **Initialize** $\beta = \hat{\beta}$;
2. **Perform** a local searching to update the individual
   1. **Repeat** the following steps
      1. **Minimize** the objective function
         $$\sum_{j=1}^{N} \|D_i \alpha_j - \Phi \beta\|_2^2 + \lambda \|\beta - \mathfrak{R}^j\|_2^2 + \mu \sum_{j=1}^{N} \|\alpha_j\|_{0,1}$$
         using the pursuit algorithm and SVD to obtain $\{\alpha_j\}$ and $\{D_i\}$;
      2. **Update** $\beta$: when $\{\alpha_j\}$ and $\{D_i\}$ are determined, the objective function becomes a quadratic and convex function, so we estimate $\beta$ by setting the derivative of
         $$\sum_{j=1}^{N} \|D_i \alpha_j - \Phi \beta\|_2^2 + \lambda \|\beta - \mathfrak{R}^j\|_2^2$$
         as 0.
   3. **Until** the stop condition is satisfied
3. **Evaluate** the fitness function of the population and save the best individual;
4. **Mutate** the population;
5. **Until** the stop condition is satisfied

---

**Fig. 1.** Flowchart of the memetic algorithm-enhanced optimization.
components of the noisy image are efficiently removed by the learned dictionaries, the $\mathbf{b}_1, \ldots, \mathbf{b}_N$ can be correctly recovered.

3.2. Experiment 2: Comparisons on benchmark natural images

In this experiment, we change the noise level and the performance of our proposed method is compared with that of Lee filtering, Bayesian Soft threshold (BST) (Abramovich et al., 1998), Non-local means (NLM) (Buades and Morela, 2005), redundant DCT (RDCT) dictionary and KSVD dictionary learning based methods (Aharon et al., 2006). The number of atoms in RDCT dictionary and KSVD dictionary are the same with that of our method, and the number of training samples in KSVD and our method are the same. The code of KSVD algorithm and RDCT dictionary based denoising.
The denoised result of Lena image when \( s = 20 \) is shown in Fig. 3. Fig. 3(a)–(f) give the denoised images by Lee, BST, NLM, RDCT, KSVD dictionary learning method (Aharon et al., 2006) and our proposed method respectively. From the result we can see that Lee filtering outputs the worst result, and BST cannot reduce the noise completely. The denoised image by NLM is too smooth. RDCT, KSVD dictionary learning and our proposed methods can achieve better result than NLM. Compared with RDCT and KSVD, our proposed method can utilize both the advantages of multiscale denoising and sparse representation based spatial denoising schemes, so resulting higher quality images that well preserve the edges, contours and textures. The amplifications of the local regions in the denoised images are shown in Fig. 4.

Fig. 4(a)–(d) show the original image, the recovered images recovered by RDCT, KSVD and our method respectively. From it we can see our proposed method outperforms the other methods in preserving the details and textures. The edges of the hat, the contours of the eye, and the texture of the hair in the image recovered by our method are clearer than that of other methods.

The Barbara image and Peppers image are also tested, and the PSNRs (dB) of the denoised images are shown in Table 1, with the bold indicated the best result among the four methods. From it we can see that our proposed method has an improvement over KSVD dictionary learning algorithm Table 2.

### 3.3. Experiment 3: comparison result of MA with GA

In this test, we take the Lena image as an example to investigate the efficiency of the memetic algorithm optimization in our method. We let the population size \( M = 1, 20, 50 \), respectively, and compare the performance of memetic algorithm with GA. The mutation and crossover probabilities are set as 0.3 and 0.7, respectively. Both the denoised result and the consumed time are considered. Considering the randomness of the initialization, ten independent experiments are taken and the average result is shown in Table 3. From it we can see that MA outperforms GA in PSNRs because GA is often trapped into local minimums. Moreover, MA combines the local searching operator with the global searching of EAs, so achieving rapid convergence than GA. Moreover, experimentally we found that when \( M \) is larger than 50, no remarkable improvement can be observed for MA and GA.

### Table 2

Comparison result of the PSNRs (dB).

<table>
<thead>
<tr>
<th>Methods</th>
<th>Population</th>
<th>( \sigma = 20 )</th>
<th>( \sigma = 30 )</th>
<th>( \sigma = 40 )</th>
<th>( \sigma = 50 )</th>
<th>Average time(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>MA</td>
<td>( M = 1 )</td>
<td>32.9301</td>
<td>30.8104</td>
<td>29.3775</td>
<td>27.7016</td>
<td>199</td>
</tr>
<tr>
<td></td>
<td>( M = 20 )</td>
<td>33.0439</td>
<td>30.9836</td>
<td>29.5582</td>
<td>28.3947</td>
<td>4513</td>
</tr>
<tr>
<td></td>
<td>( M = 50 )</td>
<td>33.0524</td>
<td>31.0070</td>
<td>29.5622</td>
<td>28.3998</td>
<td>10283</td>
</tr>
<tr>
<td>GA</td>
<td>( M = 1 )</td>
<td>32.2328</td>
<td>30.1942</td>
<td>28.6923</td>
<td>27.2426</td>
<td>784</td>
</tr>
<tr>
<td></td>
<td>( M = 20 )</td>
<td>32.8171</td>
<td>30.7455</td>
<td>29.2661</td>
<td>28.1005</td>
<td>17253</td>
</tr>
<tr>
<td></td>
<td>( M = 50 )</td>
<td>32.8826</td>
<td>30.7931</td>
<td>29.2662</td>
<td>28.1025</td>
<td>43833</td>
</tr>
</tbody>
</table>

### Table 3

PSNR (dB) of the denoised image by different mutation strategies.

<table>
<thead>
<tr>
<th>Number of iteration</th>
<th>Best strategy</th>
<th>Random strategy</th>
<th>Our mutation</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>32.8246</td>
<td>32.4585</td>
<td>32.8202</td>
</tr>
<tr>
<td>20</td>
<td>32.9193</td>
<td>32.7409</td>
<td>32.9433</td>
</tr>
<tr>
<td>30</td>
<td>33.0258</td>
<td>33.0111</td>
<td>33.0171</td>
</tr>
<tr>
<td>40</td>
<td>32.9001</td>
<td>33.0362</td>
<td>33.0392</td>
</tr>
</tbody>
</table>
ones, which are denoted as best strategy and random strategy, respectively. The denoised result after some iteration is shown in Table 3. From the result we can see that the “Best strategy” converge rapidly but is reliable to be trapped into local minimum, “Random strategy” is good at exploiting the possible solution space. A combination of the two schemes will be helpful in obtaining a stable and rapid convergence in the evolution.

4. Conclusions

In this paper we proposed a new evolution-enhanced multiscale overcomplete dictionary learning method for image denoising. The main contribution of the paper can be summarized as,

1) We combined the multiscale image denoising method with the recent developed SR-based spatial denoising approach, and proposed a multiscale overcomplete dictionaries learning approach for image denoising. Then we reduce the dictionary learning to a NP-hard $l_0$-norm minimization problem with multiple variables.

2) In order to solve the NP-hard optimization problem, we proposed a memetic algorithm-enhanced method to alternately optimize the variables. The algorithm has the capability of global searching, so outperforms the available optimization problems in reducing the noise existed in images.

3) Some experiments are taken on investigating our proposed method using some benchmark images, and the denoised images are better than the state-of-the-art result in PSNR.

Acknowledgments

The authors would like to thank the anonymous reviewers for their constructive comments. This work was supported by the National Science Foundation of China under Grant nos. 61072108, 60601029 and 60971112, the Program of NCET-10-0668, the Basic Science Research Fund in Xidian University under Grant no. JY10000902041, the Science Foundation of Shaanxi Province under Grant no. 2006F25 and the Scientific Research Project Funded by Baoji University of Arts and Sciences (Grant no. ZK09167).

References


