IJCSET |September 2012 | Vol 2, Issue 9, 1386-1395 www.ijcset.net | ISSN:2231-0711

A Novel Sparse Representation Method for Image Restoration Applications

K.S.K.L.Priyanka, U.V.Ratna Kumari

Department of ECE,
Jawaharlal Nehru Technological University Kakinada (JNTUK),
Kakinada, India
priyanka.kskl@gmail.com,vinayratna74@gmail.com

***Abstract*—Sparse representation has been widely used in various image restoration applications. The quality of image restoration mainly depends on whether the used sparse domain can represent well the underlying image. Since the contents representing the underlying image can vary significantly across different images or different patches in an image, we propose to learn various sets of bases from a pre-collected dataset of example image patches and for processing a particular given patch, a suitable set of base is selected adaptively as local sparse domain. Here we introduce two adaptive regularization terms into the sparse representation framework. One is a set of auto regressive (AR) models are learned from the pre-collected dataset of example image patches and the best fitted AR model is adaptively selected for regularization. Second is image non-local self similarity regularization to regularize the image local structures. To make the sparse coding more accurate, a centralized sparsity constraint is introduced by exploiting the nonlocal image statistics. The local sparsity and the nonlocal sparsity constraints are unified into a variational framework for optimization. Extensive experiments on the proposed method of CSR method achieves convincing improvement over previous state-of-the-art methods in terms of PSNR and SSIM values.**

***Keywords*--------Image restoration, deblurring, super-resolution, sparse representation, regularization**

INTRODUCTION

Image restoration (IR) aims to recover a high-quality image from its degraded (e.g., noisy, blurred and/or downsampled) versions, which may be taken, for example, by a low-end camera and/or under limited conditions. For an observed image Y, the problem of IR can be expressed by

$$Y = DHX + v \qquad ..(1)$$

where X is the unknown image to be estimated, and H and Dare degrading operators and υ is additive noise. When H and D are identities, the IR problem becomes denoising; when D is identity and H is a blurring operator, IR becomes deblurring; when D is identity and H is a set of random projections, IR becomes compressed sensing[3]; when D is a down-sampling operator and H is a blurring operator, IR becomes (single-image) super-resolution. IR, a fundamental problem in image processing has been widely studied in the past three decades.In this paper, we focus on deblurring and single-image super-resolution.

Due to the ill-posed nature of IR[2], the solution to (1) with an l_2-norm fidelity constraint, i.e., $\hat{X} = \arg\min_X \|Y - DHX\|_2^2$ .is generally not unique. A prior knowledge of natural images can be used to regularize the IR problem for finding a better solution. One of the most commonly used regularization models is the total variation (TV) model

$$\hat{X} = \arg\min_X \{\|Y - DHX\|_2^2 + \lambda.|\nabla X|_1\} \qquad ,$$

where $|\nabla X|_1$ is the l_1-norm of the first-order derivative of X and λ is a constant.

The success of TV regularization validates the importance of good image prior models in solving the IR problems. In wavelet based image denoising the sparsity of wavelet coefficients can serve as good prior. This reveals the fact that many types of signals, e.g., natural images, can be sparsely represented (or coded) using a dictionary of atoms, such as DCT or wavelet bases, that is, denoting by Φ the dictionary, we have $X \approx \Phi\alpha$ and most of the coefficients in α are close to zero. With the sparsity prior, the representation of X over Φ can be estimated from its observation Y by solving the following l_0-minimization problem:

$$\hat{\alpha} = \arg\min_\alpha \{\|Y - DH\Phi\alpha\|_2^2 + \lambda.\|\alpha\|_0\} \qquad ,$$

where the l_0-norm counts the number of nonzero coefficients in vector α . Once we obtain $\hat{\alpha}$, X can then be estimated as $\hat{X} = \Phi\hat{\alpha}$. The l_0–minimization is an NP-hard combinatorial search problem, and is usually solved by greedy algorithms. The l_1-minimization, as the closest convex function to l_0-minimization, is then widely used as an alternative approach to solving the sparse coding problem: $\hat{\alpha} = \arg\min_\alpha \{\|Y - DH\Phi\alpha\|_2^2 + \lambda.\|\alpha\|_1\}$. For better IR results the l_1 –norm sparsity regularization term has been iteratively reweighted. Sparse representation has been successfully used in various image processing applications.

The main issue in sparse representation modeling is the determination of dictionary Φ. Many dictionary learning (DL) methods aim at learning a universal and over-complete dictionary to represent various image structures. However, sparse decomposition over a highly redundant

dictionary is potentially unstable and tends to generate visual artifacts. In this paper, we propose an adaptive sparse domain selection (ASDS) scheme for sparse representation. A set of compact subdictionaries is learned from high-quality example image patches. The example image patches are clustered into many clusters. Since each cluster consists of many patches with similar patterns, a compact subdictionary can be learned for each cluster. Particularly, for simplicity, we use the principal component analysis (PCA) technique to learn the subdictionaries. For an image patch to be coded, the best subdictionary that is most relevant to the given patch is selected. Since the given patch can be better represented by the adaptively selected subdictionary, the whole image can be more accurately reconstructed than using a universal dictionary, which will be validated by our experiments.

In addition to sparsity regularization, we introduce other regularization terms to further increase the IR performance. Here, we propose to use the piecewise autoregressive (AR) models, that are prelearned from the training dataset, to characterize the local image structures. For each given local patch, one or more AR models can be adaptively selected for regularization. On the other hand, considering the fact that there are often many repetitive image structures in an image, we introduce a nonlocal (NL) self-similarity constraint served as another regularization term, which preserves edge sharpness and suppresses noise.

We use an efficient iterative shrinkage (IS) algorithm to solve the l_1-minimization problem after adding ASDS and adaptive regularizations (AReg) into the sparse representation based IR framework. In addition, we adaptively estimate the image local sparsity to adjust the sparsity regularization parameters. In addition, we propose a centralized sparse representation (CSR) model to effectively reduce the sparse coding noise (SCN) (difference of original image and estimated image)and thus enhance the sparsity based IR performance. The basic idea is to integrate the image local sparsity constraint (i.e., a local patch can be coded by a few atoms sparsely selected from a dictionary) and the centralized sparsity constraint (i.e., the sparse coding coefficients should be close to their mean values) into a unified variational framework for optimization. Extensive experiments on IR are conducted, and the experimental results show that the proposed CSR algorithm outperforms significantly many state-of-the-art IR methods in terms of PSNR and SSIM values.

PRIOR ART

In recent years, sparse coding or sparse representation strategy has been widely studied to solve inverse problems, partially due to the progress of l_0-norm and l_1-norm minimization techniques .

Let $X \in \Re^n$ is the target signal to be coded, and $\Phi = [\phi_1,...,\phi_m] \in \Re^{n \times m}$ is a given dictionary of atoms (i.e., code set). The sparse coding of X over Φ is to find a sparse vector $\alpha = [\acute{a}_1,...,\acute{a}_m]$ (i.e., most of the coefficients in α are close to zero) such that $X \approx \Phi\alpha$. If the sparsity is measured as the l_0-norm of α , which counts the nonzero coefficients in α , the sparse coding problem becomes $\min_{\alpha} \|X - \Phi\alpha\|_2^2$ s.t. $\|\alpha\|_0 \le T$, where T is a scalar controlling the sparsity. Alternatively, the sparse vector can also be found by

$$\hat{\alpha} = \arg\min_{\alpha} \{\|X - \Phi\alpha\|_2^2 + \lambda.\|\alpha\|_0\} \qquad (2)$$

where λ is a constant. Since the l_0-norm is nonconvex, it is often replaced by either the standard l_1-norm or the weighted l_1–norm to make the optimization problem convex.

The critical issue sparse representation modelling is the selection of dictionary Φ . Given a set of training image patches $S = [s_1,...,s_N] \in \Re^{n \times N}$, dictionary learning (DL) aims to jointly optimize the dictionary Φ and the representation coefficient matrix $\Lambda = [\alpha_1,...,\alpha_N]$ such that $s_i \approx \Phi\alpha_i$ and $\|\alpha_i\|_p \le T$, where $p = 0$ or 1. This can be formulated by the following minimization problem:

$$(\hat{\Phi}, \hat{\Lambda}) = \arg\min_{\Phi,\Lambda} \|S - \Phi\Lambda\|_F^2 \text{ s.t. } \|\alpha_i\|_p \le T, \forall_i \qquad (3)$$

where $\|\cdot\|_F$ is the Frobenius norm. The above minimization problem is nonconvex even when $p = 1$. To make it tractable, approximation approaches, including MOD and K-SVD ,have been proposed to alternatively optimizing Φ and Λ , leading to many state-of-the-art results in image processing.

Regularization has been used in IR for a long time to incorporate the image prior information. As a classic method, the autoregressive (AR) modeling has been successfully used in image compression [33] and interpolation. Recently, the AR model was used for adaptive regularization in compressive image recovery : $\min_X \sum_i \|X_i - \chi_i\alpha_i\|_2^2$ s.t. $Y = AX$, where χ_i is the vector containing the neighboring pixels of pixel X_i within the support of the AR model, and a_i is the AR parameter vector. In this paper, we propose a learning-based adaptive regularization, where the AR models are learned from high-quality training images, to increase the AR modeling accuracy.

In recent years, the nonlocal (NL) methods have led to promising results in various IR tasks, especially in image denoising. The idea of NL methods is very simple: the patches that have similar patterns can be spatially far from each other, and thus we can collect them in the whole image. This NL self-similarity prior was later employed in image deblurring and super-resolution as second regulariation term .

SPARSE REPRESENTATION WITH ADAPTIVE SPARSE DOMAIN SELECTION (ASDS)

In this section, we propose an ASDS scheme, in which a series of compact subdictionaries are learned and assigns adaptively each local patch a subdictionary as the sparse domain. With ASDS, a weighted l_1-norm sparse representation model will be proposed for IR tasks. Suppose that $\{\Phi_k\}$, $k=1,2,....,K$, is a set of orthonormal subdictionaries. Let X be an image vector, and , $X_i = R_i X, i=1,2,...,N$, be the i^{th} patch (size: $\sqrt{n}\times\sqrt{n}$) vector of X, where R_i is a matrix extracting patch X_i from X . For patch X_i ,suppose that a subdictionary Φ_{k_i} is selected for it. Then, X_i can be approximated as , $\hat{X}_i = \Phi_{k_i}\alpha_i, \|\alpha_i\|_1 \le T$, via sparse coding. The whole image X can be reconstructed by averaging all of the reconstructed patches $\hat{X}_i$, which can be mathematically written as

$$\hat{X} = (\sum_{i=1}^{N} R_i^T R_i)^{-1} \sum_{i=1}^{N} (R_i^T \Phi_{k_i} \alpha_i). \qquad (4)$$

In (4), the matrix to be inverted is a diagonal matrix, and hence the calculation of (4) can be done in a pixel-by-pixel manner. Obviously, the image patches can be overlapped to better suppress noise and block artifacts. For the convenience of expression, we define the following operator " $\circ$ ":

$$\hat{X} = \Phi \circ \alpha \overset{\Delta}{=} (\sum_{i=1}^{N} R_i^T R_i)^{-1} \sum_{i=1}^{N} (R_i^T \Phi_{k_i} \alpha_i) \qquad (5)$$

where Φ is the concatenation of all subdictionaries $\{\Phi_k\}$ and α is the concatenation of all α_i .

Let $Y = DHX + v$ be the observed degraded image, our goal is to recover the original image X from Y . With ASDS and the definition in (5), the IR problem can be formulated as follows:

$$\hat{\alpha} = \arg\min_{\alpha} \{\|Y - DH\Phi \circ \alpha\|_2^2 + \lambda\|\alpha\|_1\} \quad (6)$$

Clearly, one key procedure in the proposed ASDS scheme is the determination of Φ_{k_i} for each local patch. To facilitate the sparsity-based IR, we propose to learn offline the subdictionaries $\{\Phi_k\}$, and select online from $\{\Phi_k\}$ the best fitted subdictionary to each patch X_i .

A. Learning the Subdictionaries

In order to learn a series of subdictionaries to code the various local image structures, we need to first construct a dataset of local image patches for training. For this purpose, we collected a set of high-quality natural images and cropped from them a rich amount of image patches with size $\sqrt{n}\times\sqrt{n}$. A cropped image patch, denoted by s_i , will be involved in DL if its intensity variance $Var(s_i)$ is greater than a threshold Δ , i.e., $Var(s_i) > \Delta$. This patch selection criterion is to exclude the smooth patches from training and guarantee that only the meaningful patches with a certain amount of edge structures are involved in DL. Let M image patches $S = [s_1, s_2, ..., s_M]$ are selected. Now our goal is to learn K compact subdictionaries $\{\Phi_k\}$ from S so that, for each given local image patch, the most suitable subdictionary can be selected. For this, we cluster the dataset S into K clusters, and learn a subdictionary from each of the K clusters. Apparently, the K clusters are expected to represent the K distinctive patterns in S. To generate perceptually meaningful clusters, we perform the clustering in a feature space. In the hundreds of thousands patches cropped from the training images, many patches are approximately the rotated version of the others. Since the human visual system is sensitive to image edges, which convey most of the semantic information of an image, we use the high-pass filtering output of each patch as the feature for clustering. It allows us to focus on the edges and structures of image patches without taking into account the pixel intensities and helps to increase the accuracy of clustering

Consider $S_h = [s_1^h, s_2^h, ..., s_M^h]$ as the high-pass filtered dataset of S. We used the K-means algorithm to partition S_h into K clusters $\{C_1, C_2, ..., C_K\}$ and denote by μ_k the centroid of cluster C_k . Once S_h is partitioned, dataset S can then be clustered into K subsets $S_k, k=1,2,....,K$, and S_k is a matrix of dimension $n \times m_k$, where m_k denotes the number of samples in S_k .

Now we focus on how to learn a subdictionary Φ_k from the cluster S_k such that all the elements in S_k can be faithfully represented by Φ_k . The design of Φ_k can be intuitively formulated by the following objective function:

$$(\hat{\Phi}_k, \hat{\Lambda}_k) = \arg\min_{\Phi_k, \Lambda_k} \{\|S_k - \Phi_k \Lambda_k\|_F^2 + \lambda\|\Lambda_k\|_1\} \qquad (7)$$

where Λ_k is the representation coefficient matrix of S_k over Φ_k Equation (7) is a joint optimization problem of Φ_k and Λ_k , and it can be solved by alternatively optimizing Φ_k and , Λ_k like in the K-SVD algorithm .

However, we do not directly use (7) to learn the subdictionary Φ_k based on the following considerations. First, the $l_2 - l_1$ joint minimization in (7) requires much computational cost. Second, we often assume that the dictionary Φ_k is over-complete. Nonetheless, here S_k is a subdataset after K-means clustering, which implies that, not only is the number of elements in S_k limited, but also these elements tend to have similar patterns. Therefore, it is not necessary to learn an over-complete dictionary Φ_k from S_k . In addition, a compact dictionary will decrease much the computational cost of the sparse coding of a given image patch. For the above said reasons, we propose to

learn a compact dictionary while trying to approximate (7). So we choose PCA as good solution for this problem.

PCA is a classical signal de-correlation and dimensionality reduction technique that is widely used in pattern recognition and statistical signal processing In this paper, we apply PCA to each subdataset S_k to compute the principal components, from which the dictionary Φ_k is constructed. Denote by Ω_k the co-variance matrix of dataset S_k . By applying PCA to Ω_k , an orthogonal transformation matrix P_k can be obtained. If we set P_k as the dictionary and let $Z_k = P_k^T S_k$, we will then have $\|S_k - P_k Z_k\|_F^2 = \|S_k - P_k P_k^T S_k\|_F^2 = 0$. In other words, the approximation term in (7) will be exactly zero, yet the corresponding sparsity regularization term $\|Z_k\|_1$ will have a certain amount because all the representation coefficients in Z_k are preserved.

For a better balance between the l_1 -norm regularization term and l_2 -norm approximation term in (7), we consider only the first r most important eigenvectors in P_k to form a dictionary Φ_r , i.e., $\Phi_r = [p_1, p_2, ..., p_r]$. Let $\Lambda_r = \Phi_r^T S_k$. Clearly, since not all of the eigenvectors are used to form Φ_r , the reconstruction error $\|S_k - \Phi_r \Lambda_r\|_F^2$ in (7) will increase with the decrease of r . However, the term $\|\Lambda_r\|_1$ will decrease. Therefore, the optimal value of r , denoted by $r_\circ$, can be determined by

$$r_\circ = \arg\min_r \{\|S_k - \Phi_r \Lambda_r\|_F^2 + \lambda \|\Lambda_r\|_1\} \quad (8)$$

Finally, the subdictionary learned from subdataset S_k is $\Phi_k = [p_1, p_2, ... p_{r_\circ}]$.

Applying the above procedures to all of the K subdatasets S_k , we could get K subdictionaries Φ_k , which will be used in the ASDS process of each given image patch.

B. Adaptive Selection of the Subdictionary

We have learned a dictionary Φ_k for each subset S_k in the above subsection. Meanwhile, we have computed the centroid μ_k of each cluster C_k associated with S_k . Therefore, we have K pairs $\{\Phi_k, \mu_k\}$, with which the ASDS of each given image patch can be accomplished.

In the proposed sparsity-based IR scheme, we assign adaptively a subdictionary to each local patch of X, spanning the adaptive sparse domain. Since X is unknown beforehand, we need to have an initial estimation of it. The initial estimation of X can be accomplished by taking wavelet bases as the dictionary and then solving (6) with the iterated shrinkage algorithm in [10]. Let $\hat{X}$ be the estimate of X and $\hat{X}_i$ be a local patch of $\hat{X}$. Recall that we have the centroid μ_k of each cluster available, and hence we could select the best fitted subdictionary to $\hat{X}_i$ by comparing the high-pass filtered patch of $\hat{X}_i$, denoted by $\hat{X}_i^h$,to the centroid μ_k . For example, we can select the dictionary for $\hat{X}_i$ based on the minimum distance between $\hat{X}_i^h$ and μ_k , i.e.,

$$k_i = \arg\min_k \|\hat{X}_i^h - \mu_k\|_2 \quad \ldots (9)$$

However, directly calculating the distance between $\hat{X}_i^h$ and μ_k may not be robust enough because the initial estimate $\hat{X}$ can be noisy. Here, we propose to determine the subdictionary in the subspace of μ_k . Let $U = [\mu_1, \mu_2, ...\mu_k]$ be the matrix containing all the centroids. By applying SVD to the co-variance matrix of U , we can obtain the PCA transformation matrix of U. Let Φ_c be the projection matrix composed by the first several most significant eigenvectors. We compute the distance between $\hat{X}_i^h$ and μ_k in the subspace spanned by Φ_c as

$$k_i = \arg\min_k \|\Phi_c \hat{X}_i^h - \Phi_c \mu_k\|_2 \quad (10)$$

Compared with (9), (10) can increase the robustness of adaptive dictionary selection. By using (10), the k_i th subdictionary Φ_{k_i} will be selected and assigned to patch $\hat{X}_i$. Then, we can update the estimation X of by minimizing (6) and letting $\hat{X} = \Phi \circ \hat{\alpha}$. With the updated estimate $\hat{X}$, the ASDS of X can be consequently updated. Such a process is iteratively implemented until the estimation $\hat{X}$ converges.

C. Adaptively Reweighted Sparsity Regularization

In (6), the parameter λ is a constant to weight the l_1 –norm sparsity regularization term .Here, we propose a new method to estimate adaptively the image local sparsity and then reweight the l_1 -norm sparsity in the ASDS scheme.

The reweighted l_1 -norm sparsity regularized minimization with ASDS can be formulated as follows:

$$\hat{\alpha} = \arg\min_\alpha \{\|y - DH\Phi \circ \alpha\|_2^2 + \sum_{i=1}^{N}\sum_{j=1}^{n} \lambda_{i,j} |\alpha_{i,j}|\} \quad (11)$$

where $\alpha_{i,j}$ is the coefficient associated with the jth atom of Φ_{k_i} and $\lambda_{i,j}$ is the weight assigned to $\alpha_{i,j}$. In [13], $\lambda_{i,j}$ is empirically computed as $\lambda_{i,j} = 1/(|\hat{\alpha}_{i,j}| + \varepsilon)$, where $\hat{\alpha}_{i,j}$ is the estimate of $\alpha_{i,j}$ and ε is a small

constant. Here, we propose a more robust method for computing $\lambda_{i,j}$ by formulating the sparsity estimation as a maximum a posterior (MAP) estimation problem. Under the Bayesian framework, with the observation Y the MAP estimation of α is given by

$$\hat{\alpha} = \arg\max_{\alpha}\{\log P(\alpha|Y)\}$$
$$= \arg\min_{\alpha}\{-\log P(Y|\alpha) - \log P(\alpha)\}. \quad (12)$$

By assuming Y is contaminated with additive Gaussian white noises of standard deviation σ_n, we have

$$P(Y|\alpha) = \frac{1}{\sigma_n\sqrt{2\Pi}}\exp(-\frac{1}{2\sigma_n^2}\|Y - DH\Phi\circ\alpha\|_2^2) \quad (13)$$

The prior distribution $P(\alpha)$ is often characterized by an i.i.d. zero-mean Laplacian probability model

$$P(\alpha) = \prod_{i=1}^{N}\prod_{j=1}^{n}\frac{1}{\sqrt{2}\sigma_{i,j}}\exp(-\frac{\sqrt{2}}{\sigma_{i,j}}|\alpha_{i,j}|) \quad (14)$$

where $\sigma_{i,j}$ is the standard deviation of $\alpha_{i,j}$. By plugging $P(Y|\alpha)$ and $P(\alpha)$ into (12), we could readily derive the desired weight in (11) as $\lambda_{i,j} = 2\sqrt{2}\sigma_n^2/\sigma_{i,j}$. For numerical stability,we compute the weights by

$$\lambda_{i,j} = \frac{2\sqrt{2}\sigma_n^2}{\hat{\sigma}_{i,j} + \varepsilon} \quad (15)$$

where $\hat{\sigma}_{i,j}$ is an estimate of $\sigma_{i,j}$ and ε is a small constant. Now, let us discuss how to estimate $\sigma_{i,j}$. Let $\hat{X}_i$ be the estimate of X_i and $\hat{X}_i^l, l = 1,2,....L$, be the nonlocal similar patches to $\hat{X}_i$. (The determination of nonlocal similar patches to $\hat{X}_i$ will be described in Section IV-C.) The representation coefficients of these similar patches over the selected subdictionary Φ_{k_i} is $\hat{\alpha}_i^l = \Phi_{k_i}^T\hat{X}_i^l$. Then, we can estimate $\sigma_{i,j}$ by calculating the standard deviation of each element $\hat{\alpha}_{i,j}$ in $\hat{\alpha}_i^l$.

SPATIALLY ADAPTIVE REGULARIZATION

In the above Section , we proposed to select adaptively a subdictionary to code the given image patch. To further improve the IR method we introduce two types of adaptive regularization (AReg) terms. A local area in a natural image can be viewed as a stationary process, which can be well modeled by the autoregressive (AR) models. Here, we propose to learn a set of AR models from the clustered high quality training image patches, and adaptively select one AR model to regularize the input image patch. Besides the AR models, which exploit the image local correlation, we propose to use the nonlocal similarity constraint as a complementary AReg term to the local AR models.

Training the AR Models

Recall that, in the above Section, we have partitioned the whole training dataset into K subdatasets S_k. For each S_k, an AR model can be trained using all of the sample patches inside it. Here we let the support of the AR model be a square window, and the AR model aims to predict the central pixel of the window by using the neighboring pixels. Considering that determining the best order of the AR model is not trivial, and a high order AR model may cause data over-fitting, in our experiments a 3×3 window (i.e., AR model of order 8) is used. The vector of AR model parameters, denoted by a_k, of the k^{th} subdataset S_k, can be easily computed by solving the following least square problem:

$$a_k = \arg\min_{a}\sum_{s_i\in S_k}(s_i - a^T q_i)^2 \quad (16)$$

where s_i is the central pixel of image patch s_i and q_i is the vector that consists of the neighboring pixels of s_i within the support of the AR model. By applying the AR model training process to each subdataset, we obtain a set of AR models $\{a_1, a_2, ..., a_K\}$ that are used for adaptive regularization.

A. Adaptive Selection of the AR Model for Regularization

The adaptive selection of the AR model for each patch X_i is similar to the selection of a subdictionary for X_i described in Section Adaptive Selection of the sub-dctionary. With an estimation $\hat{X}_i$ of X_i, we compute its high-pass Gaussian filtering output $\hat{X}_i^h$. Let $k_i = \arg\min_k\|\Phi_c\hat{X}_i^h - \Phi_c\mu_k\|_2$, and then the k_ith AR model a_{k_i} will be assigned to patch X_i. Denote by X_i the central pixel of patch X_i and by χ_i the vector containing the neighboring pixels of X_i within patch X_i. We can expect that the prediction error of X_i using a_{k_i} and χ_i should be small, i.e., $\|X_i - a_{k_i}^T\chi_i\|_2^2$ should be minimized. By incorporating this constraint into the ASDS-based sparse representation model in (11), we have a lifted objective function as follows:

$$\hat{\alpha} = \arg\min_{\alpha}\{\|Y - DH\Phi\circ\alpha\|_2^2 + \sum_{i=1}^{N}\sum_{j=1}^{n}\lambda_{i,j}|\alpha_{i,j}| + \gamma\cdot\sum_{X_i\in X}\|X_i - a_{k_i}^T\chi_i\|_2^2\} \quad (17)$$

where γ is a constant balancing the contribution of the AR regularization term. For the convenience of expression, we

write the third term $\sum_{X_i \in X} \left\| X_i - a_{k_i}^T \chi_i \right\|_2^2$ as $\left\| (I-A)X \right\|_2^2$, where I is the identity matrix and

$$A_{(i,j)} = \begin{cases} a_i, \\ 0, otherwise. \end{cases} if X_j is an element of \chi_i, a_i \in a_{k_i}$$

Then, (17) can be rewritten as

$$\hat{\alpha} = \arg\min_{\alpha} \{ \left\| Y - DH\Phi \circ \alpha \right\|_2^2 + \sum_{i=1}^{N} \sum_{j=1}^{n} \lambda_{i,j} \left| \alpha_{i,j} \right| + \gamma \cdot \left\| (I-A)X \right\|_2^2 \} \quad (18)$$

B. Adaptive Regularization by Nonlocal Similarity

The above discussed AR model-based AReg exploits the local statistics in each image patch. On the other hand, there are often many repetitive patterns throughout a natural image. Such nonlocal redundancy is very helpful to improve the quality of reconstructed images. As a complementary AReg term to AR models, we further introduce a nonlocal similarity regularization term into the sparsity-based IR framework.

For each local patch X_i, we search for the similar patches to it in the whole image X(in practice, in a sufficiently large area around X_i). A patch X_i^l is selected as a similar patch to X_i if $e_i^l = \left\| \hat{X}_i - \hat{X}_i^l \right\|_2^2 \le t$, where t is a preset threshold, and $\hat{X}_i$ and $\hat{X}_i^l$ are the current estimates of X_i and X_i^l, respectively, or we can select the patch $\hat{X}_i^l$ if it is within the first L(L=10 in our experiments) closest patches to $\hat{X}_i$. Let X_i be the central pixel of patch X_i, and X_i^l be the central pixel of patch X_i^l. Then, we can use the weighted average of , X_i^l i.e., $\sum_{l=1}^{L} b_i^l X_i^l$, to predict X_i, and the weight b_i^l assigned to X_i^l is set as $b_i^l = \exp(-e_i^l / h) / c_i$, where h is a controlling factor of the weight and $c_i = \sum_{l=1}^{L} \exp(-e_i^l / h)$ is the normalization factor. Considering that there is much nonlocal redundancy in natural images, we expect that the prediction error $\left\| X_i - \sum_{l=1}^{L} b_i^l X_i^l \right\|_2^2$ should be small. Let b_i be the column vector containing all the weights b_i^l and β_i be the column vector containing all X_i^l. By incorporating the nonlocal similarity regularization term into the ASDS based sparse representation in (11), we have

$$\hat{\alpha} = \arg\min_{\alpha} \{ \left\| Y - DH\Phi \circ \alpha \right\|_2^2 + \sum_{i=1}^{N} \sum_{j=1}^{n} \lambda_{i,j} \left| \alpha_{i,j} \right| + \eta \cdot \sum_{X_i \in X} \left\| X_i - b_i^T \beta_i \right\|_2^2 \} \quad (19)$$

where η is a constant balancing the contribution of nonlocal regularization. Equation (19) can be rewritten as

$$\hat{\alpha} = \arg\min_{\alpha} \{ \left\| Y - DH\Phi \circ \alpha \right\|_2^2 + \sum_{i=1}^{N} \sum_{j=1}^{n} \lambda_{i,j} \left| \alpha_{i,j} \right| + \eta \cdot \left\| (I-B)\Phi\alpha \right\|^2 \} \quad (20)$$

where I is the identity matrix and

$$B(i,l) = \begin{cases} b_i^l, if X_i^l is an element of \beta_i, b_i^l \in b_i \\ 0, otherwise \end{cases}$$

SUMMARY OF THE ALGORITHM

Afetr including both the local AR regularization and the nonlocal similarity regularization into the ASDS-based sparse representation in (11), we have the following ASDS-AReg-based sparse representation to solve the IR problem:

$$\hat{\alpha} = \arg\min_{\alpha} \{ \left\| Y - DH\Phi \circ \alpha \right\|_2^2 + \gamma \cdot \left\| (I-A)\Phi \circ \alpha \right\|_2^2 + \sum_{i=1}^{N} \sum_{j=1}^{n} \lambda_{i,j} \left| \alpha_{i,j} \right| + \eta \cdot \left\| (I-B)\Phi \circ \alpha \right\|_2^2 \} \quad (21)$$

In (21), the first l_2 -norm term is the fidelity term, guaranteeing that the solution $\hat{X} = \Phi \circ \hat{\alpha}$ can well fit the observation Y after degradation by operators H and D; the second l_2 -norm term is the local AR model-based adaptive regularization term, requiring that the estimated image is locally stationary; the third l_2 -norm term is the nonlocal similarity regularization term, which uses the nonlocal redundancy to enhance each local patch; and the last weighted l_1 -norm term is the sparsity penalty term, requiring that the estimated image should be sparse in the adaptively selected domain. Equation (21) can be rewritten as

$$\hat{\alpha} = \arg\min_{\alpha} \left\| \begin{bmatrix} Y \\ 0 \\ 0 \end{bmatrix} - \begin{bmatrix} DH \\ \gamma \cdot (I-A) \\ \eta \cdot (I-B) \end{bmatrix} \Phi \circ \alpha \right\|_2^2 + \sum_{i=1}^{N} \sum_{j=1}^{n} \lambda_{i,j} \left| \alpha_{i,j} \right| \quad (22)$$

By letting

$$\tilde{Y} = \begin{bmatrix} Y \\ 0 \\ 0 \end{bmatrix} \qquad K = \begin{bmatrix} DH \\ \gamma \cdot (I-A) \\ \eta \cdot (I-B) \end{bmatrix} \quad (23)$$

(22) can be rewritten as

$$\hat{\alpha} = \arg\min_{\alpha} \{ \left\| \tilde{Y} - K\Phi \circ \alpha \right\|_2^2 + \sum_{i=1}^{N} \sum_{j=1}^{n} \lambda_{i,j} \left| \alpha_{i,j} \right| \} \quad (24)$$

This is a reweighted l_1 -minimization problem, which can be effectively solved by the iterative shrinkage algorithm [8]. We outline the iterative shrinkage algorithm for optimizing (24) in Algorithm 1.

Algorithm 1

1) Initialization:

a) By taking the wavelet domain as the sparse domain, we can compute an initial estimate, denoted by $\widehat{X}$, of X by using the iterated wavelet shrinkage algorithm [8];

b) With the initial estimate $\widehat{X}$, we select the subdictionary Φ_{k_i} and the AR model a_i using (10), and calculate the nonlocal weight b_i for each local patch $\widehat{X}_i$;

c) Initialize A and B with the selected AR models and the nonlocal weights;

d) Preset γ, η, P, e and the maximal iteration number, denoted by Max_Iter;

e) Set k=0.

2) Iterate on k until $\left\|\widehat{X}^{(k)} - \widehat{X}^{(k+1)}\right\|_2^2 / N \le e$ or $k \ge Max\_Iter$ is satisfied.

a) $\widehat{X}^{(k+1/2)} = \widehat{X}^{(k)} + K^T(\widetilde{Y} - K\widehat{X}^{(k)}) = \widehat{X}^{(k)} + (UY - U\widehat{X}^{(k)} - V\widehat{X}^{(k)})$, where $U = (DH)^T DH$ and $V = \gamma^2 (I-A)^T (I-A) + \eta^2 (I-B)^T (I-B)$;

b)Compute $\alpha^{(k+1/2)} = [\Phi_{k_1}^T R_1 \widehat{X}^{(k+1/2)}, ..., \Phi_{k_N}^T R_N \widehat{X}^{(k+1/2)}]$, where N is the total number of image patches;

c) $\widehat{\alpha}_{i,j}^{(k+1)} = soft(\alpha_{i,j}^{(k+1/2)}, \tau_{i,j})$, where $soft(\cdot, \tau_{i,j})$ is a soft thresholding function with threshold $\tau_{i,j}$;

d) Compute $\widehat{X}^{(k+1)} = \Phi \circ \alpha^{(k+1)}$ using (5), which can be calculated by first reconstructing each image patch with $\widehat{X}_i = \Phi_{k_i} \alpha_i^{(k+1)}$ and then averaging all the reconstructed image patches;

e) If mod(k,P) = 0 , update the adaptive sparse domain of X and the matrices A and B using the improved estimate $\widehat{X}^{(k+1)}$.

In Algorithm 1, e is a prespecified scalar controlling the convergence of the iterative process, and Max_Iter is the allowed maximum number of iterations. The thresholds $\tau_{i,j}$ are locally computed as $\tau_{i,j} = \lambda_{i,j} / r$ [8], where $\lambda_{i,j}$ are calculated by (15) and r is chosen such that $r > \left\|(K\Phi)^T K\Phi\right\|_2$. Since the dictionary Φ_{k_i} varies across the image, the optimal determination of r for each local patch is difficult. Here, we empirically set r=4.7 for all of the patches .P is a preset integer, and we only update the subdictionaries Φ_{k_i} , the AR models a_i and the weights b_i in every P iterations to save computational cost. With the updated a_i and b_i, A and B can be updated, and then the matrix V can be updated.

Centralised Sparse Representation:

The equation for CSR model can be written as:

$$\alpha_Y = \arg\min_Y \{\left\|Y - H\Phi \circ \alpha\right\|_2^2 + \lambda\left\|\alpha\right\|_1 + \gamma \sum_{i=1}^{N} \left\|\alpha_i - \mu_i\right\| l_p$$

$$\mu_i = \sum_{j \in C_i} w_{i,j} \alpha_{i,j}$$

where (where $w_{i,j}$ is the weight, $w_{i,j}$ can be set to be inverse proportional to the distance between patches i and j: $w_{i,j} = \exp(-\left\|\widehat{X}_i - \widehat{X}_{i,j}\right\|_2^2 / h) / W$ where $\widehat{X}_i = \Phi\widehat{\alpha}_i$ and $\widehat{X}_{i,j} = \Phi\widehat{\alpha}_{i,j}$ are the estimates of patches i and j, W is a normalization factor and h is a predetermined scalar.), l_p is l-norm and p=0 or 1, From above Eq we can more clearly see that the CSR model unifies the local sparsity (i.e. $\left\|\alpha\right\|_1$) and nonlocal similarity induced sparsity (i.e. $\left\|\alpha_i - \mu_i\right\| l_p$) into a unified variational formulation. By exploiting both the local and nonlocal redundancy,better IR results can be expected.

EXPERIMENTAL RESULTS

To illustrate the robustness of the proposed method to the training dataset, we use two different sets of training images in the experiments, each set having five high-quality images as shown in Fig. 1. We can see that these two sets of training images are very different in contents. We use $Var(s_i) > \Delta$ with $\Delta = 16$ to exclude the smooth image patches, and a total amount of 727 615 patches of size 7× 7 are randomly cropped from each set of training images. In the experiments of deblurring, two types of blur kernels, a Gaussian kernel of standard deviation 3 and a 9× 9 uniform kernel, were used to simulate blurred images. Additive Gaussian white noises with standard deviations $\sqrt{2}$ was then added to the blurred images, respectively. In the experiments of super-resolution, the degraded LR images were generated by first applying a truncated 7 ×7 Gaussian kernel of standard deviation 1.6 to the original image and then down-sampling by a factor of 3. In both of the deblurring and super-resolution experiments, 7 × 7 patches (for HR image) with 5-pixel- width overlap between adjacent patches were used in the proposed methods. For color images, all of the test methods were applied to the luminance component only because human visual system is more sensitive to luminance changes, and the bi-cubic interpolator was applied to the chromatic components. Here we only report the PSNR and SSIM [12] results for the luminance component. To examine more comprehensively the proposed approach, we give three results of the proposed method: the results by using only ASDS (denoted by ASDS), by using ASDS plus AR regularization (denoted by ASDS-AR), and by using ASDS with both AR and nonlocal similarity regularization (denoted by ASDS-AR-NL),CSR.

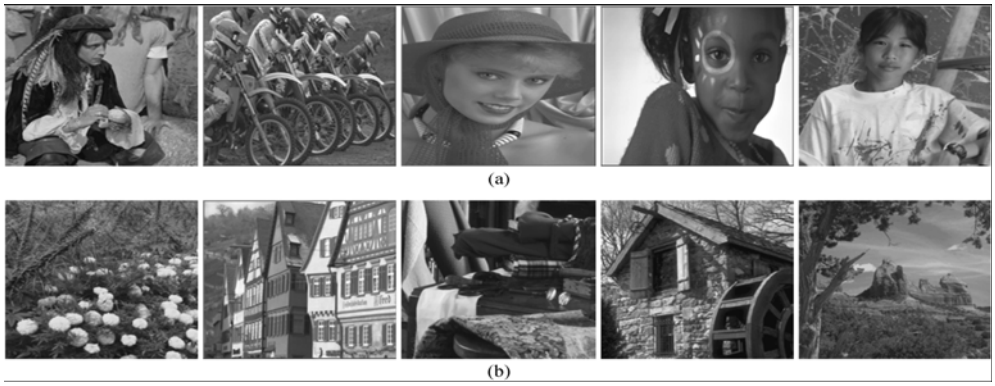

Fig 1 Two sets of high-quality images used for training subdictionaries and AR models. (a) Training dataset 1. (b) Training dataset 2.We see that the two trainingdatasets are very different in the contents.

Fig:2 Comparison of the deblurred images on *barbara* by different methods on training dataset 1(Gaussian blur kernel and $\sigma_n = \sqrt{2}$). Original, degraded,ASDS method
(PSNR=26.85dB,SSIM=0.7791), the ASDS_AR method (PSNR= 27.14dB, SSIM=0.7934), the ASDS_AR_NL method (PSNR=27.92dB,SSIM=0.8263), the CSR method(PSNR=27.98 dB,SSIM=0.8216).

Fig:3 Comparison of the deblurred images on *hat* by different methods on training dataset 2(uniform blur kernel and $\sigma_n = \sqrt{2}$). Original, degraded,ASDS method
(PSNR=31.54dB,SSIM=0.8675), the ASDS_AR method (PSNR= 31.47dB, SSIM=0.8660), the ASDS_AR_NL method (PSNR=31.55dB,SSIM=0.8659), the CSR method(PSNR=31.95 dB,SSIM=0.8742).

Fig:4 Comparison of the super-resolution images on *plants* by different methods on training dataset 1(scaling factor=3). Original, degraded,ASDS method
(PSNR=33.07dB,SSIM=0.9025), the ASDS_AR method (PSNR=33.02dB, SSIM=0.9019), the ASDS_AR_NL method (PSNR=33.40dB,SSIM=0.9072), the CSR method(PSNR= 34.11dB,SSIM=0.9217).

Fig:5 Comparison of the super-resolution images on noisy girl by different methods on training dataset 2(scaling factor=3). Original, degraded,ASDS method
(PSNR=33.39dB,SSIM=0.8190), the ASDS_AR method (PSNR=33.40dB, SSIM=0.8192), the ASDS_AR_NL method (PSNR=33.47dB,SSIM=0.8200), the CSR method(PSNR= 33.68dB,SSIM=0.8258).

The above figures gives the comparision results of deblurring and super-resolution reconstructed by using ASDS,ASDS_AR,ASDS_AR_NL and CSR methods and their PSNR,SSIM values.The tables below gives the results of deblurring and super-resolution for different images in terms of PSNR and SSIM values on Training dataset 1 and Training dataset 2 respectively.

Images	ASDS-TD1	ASDS-TD2	ASDS-AR-TD1	ASDS-AR-TD2	ASDS-AR-NL-TD1	ASDS-AR-NL-TD2	CSR
Barbara	26.85 0.7791	26.80 0.7765	27.14 0.7934	27.09 0.7925	27.92 0.8263	27.90 0.8258	27.98 0.8216
Bike	25.57 0.8119	25.53 0.8103	25.47 0.8089	25.43 0.8072	25.54 0.8095	25.50 0.8075	25.85 0.8218
Parrots	31.27 0.8997	31.23 0.8998	31.15 0.8988	31.13 0.8988	31.20 0.8984	31.20 0.8984	32.05 0.9111
Baboon	21.41 0.5845	21.41 0.5840	21.51 0.5835	21.50 0.5824	21.58 0.5740	21.58 0.5741	21.56 0.5824
Pentagon	25.66 0.7257	25.65 0.7242	25.91 0.7343	25.91 0.7332	26.61 0.7561	26.60 0.7548	26.69 0.7587
Boats	28.84 0.8033	28.80 0.8010	29.36 0.8233	29.33 0.8221	30.78 0.8677	30.76 0.8671	31.10 0.8816
Cameraman	27.19 0.7860	27.15 0.7838	27.37 0.8103	27.37 0.8137	28.12 0.8546	28.10 0.8548	28.54 0.8561
Hat	31.61 0.8678	31.54 0.8675	31.54 0.8666	31.47 0.8660	31.58 0.8655	31.55 0.8659	31.95 0.8742
Peppers	28.09 0.7670	28.10 0.7672	28.47 0.7883	28.47 0.7892	29.44 0.8344	29.44 0.8350	29.66 0.8406
Straw	22.32 0.6681	22.31 0.6658	22.41 0.6667	22.41 0.6642	22.54 0.6570	22.52 0.6550	22.76 0.6705

Table:1PSNR and SSIM results of Deblurred Images (Uniform Blur Kernel ,Noise Level $\sigma_n = \sqrt{2}$)

Images	ASDS-TD1	ASDS-TD2	ASDS-AR-TD1	ASDS-AR-TD2	ASDS-AR-NL-TD1	ASDS-AR-NL-TD2	CSR
Barbara	23.82 0.6598	23.81 0.6593	23.82 0.6602	23.81 0.6597	23.85 0.6637	23.84 0.6632	27.77 0.8252
Bike	22.66 0.6727	22.63 0.6707	22.65 0.6728	22.63 0.6708	22.82 0.6843	22.81 0.6831	27.59 0.8842
Parrots	27.78 0.8629	27.75 0.8620	27.84 0.8631	27.81 0.8622	27.95 0.8676	27.91 0.8668	33.45 0.9361
Baboon	20.11 0.3880	20.09 0.3874	20.10 0.3886	20.08 0.3880	20.12 0.3950	20.11 0.3933	21.55 0.5956
Pentagon	23.71 0.5990	23.71 0.5978	23.70 0.5995	23.71 0.5987	23.75 0.6103	23.74 0.6088	27.79 0.8284
Boats	26.91 0.7677	26.93 0.7675	27.01 0.7675	27.01 0.7672	27.12 0.7697	27.13 0.7696	31.43 0.8941
Cameraman	24.07 0.7596	24.06 0.7603	24.04 0.7584	24.04 0.7595	24.21 0.7600	24.21 0.7605	28.29 0.8561
Hat	28.79 0.8022	28.83 0.8034	28.82 0.8016	28.83 0.8030	28.94 0.8051	28.94 0.8064	33.09 0.9026
Peppers	26.29 0.7822	26.28 0.7865	26.25 0.7874	26.26 0.7864	26.38 0.7890	26.40 0.7880	30.20 0.8673
Straw	20.74 0.4757	20.74 0.4749	20.76 0.4781	20.76 0.4775	20.86 0.4881	20.85 0.4873	25.47 0.8383

Table:2 PSNR and SSIM results of Deblurred Images (Gaussian Blur Kernel ,Noise Level $\sigma_n = \sqrt{2}$)

Images	ASDS(TD1)	ASDS(TD2)	ASDS-AR(TD1)	ASDS-AR(TD2)	ASDS-AR-NL(TD1)	ASDS-AR-NL(TD2)	CSR
Parrot	29.75 0.9071	29.75 0.9066	29.80 0.9070	29.80 0.9065	30.10 0.9099	30.09 0.9100	30.63 0.9181
Hat	30.91 0.8694	30.82 0.8681	30.84 0.8684	30.76 0.8671	31.01 0.8716	30.99 0.8713	31.35 0.8738
Bike	24.33 0.7855	24.60 0.7959	24.28 0.7835	24.31 0.7844	24.61 0.7962	24.26 0.7826	24.72 0.8029
Butterfly	26.74 0.8939	26.69 0.8923	26.65 0.8909	26.59 0.8893	27.35 0.9057	27.30 0.9048	28.24 0.9219
Flower	28.90 0.8386	28.89 0.8377	28.90 0.8385	28.88 0.8375	29.17 0.8464	29.17 0.8462	29.55 0.8586
Girl	33.41 0.8193	33.39 0.8190	33.40 0.8193	33.40 0.8192	33.47 0.8201	33.47 0.8200	33.68 0.8258
Leaves	26.25 0.8947	26.22 0.8940	26.17 0.8927	26.17 0.8924	26.92 0.9094	26.92 0.9095	27.59 0.9273
Parthenon	26.74 0.7322	26.75 0.7326	26.73 0.7317	26.75 0.7326	26.89 0.7367	26.89 0.7366	27.23 0.7524
Plants	33.07 0.9025	33.05 0.9020	33.02 0.9019	33.00 0.9014	33.40 0.9072	33.40 0.9070	34.11 0.9217
Raccoon	29.13 0.7650	29.11 0.7633	29.14 0.7653	29.10 0.7636	29.25 0.7669	29.22 0.7654	29.30 0.7677

Table:3 PSNR and SSIM results of reconstructed HR images (Noise Level $\sigma_n = 0$)

Images	ASDS(TD1)	ASDS(TD2)	ASDS-AR(TD1)	ASDS-AR(TD2)	ASDS-AR-NL(TD1)	ASDS-AR-NL(TD2)	CSR
Noisy Parrot	28.92 0.8684	28.90 0.8679	28.81 0.8633	28.78 0.8633	28.74 0.8653	28.72 0.8655	29.50 0.8775
Noisy Hat	29.61 0.8087	29.61 0.8097	29.50 0.8070	29.50 0.8084	29.57 0.8131	29.57 0.8147	29.96 0.8249
Noisy Bike	23.60 0.7234	23.58 0.7222	23.48 0.7197	23.48 0.7186	23.53 0.7208	23.52 0.7197	23.77 0.7363
Noisy Butterfly	25.76 0.8417	25.78 0.8406	25.57 0.8367	26.02 0.8604	26.01 0.8600	26.01 0.8603	26.89 0.8887
Noisy Flower	27.58 0.7685	27.54 0.7665	27.64 0.7705	27.62 0.7700	27.67 0.7744	27.66 0.7742	28.07 0.7936
Noisy Girl	31.69 0.7574	31.72 0.7591	31.70 0.7578	31.72 0.7591	31.76 0.7585	31.77 0.7589	32.04 0.7642
Noisy Leaves	25.22 0.8471	25.14 0.8455	25.09 0.8430	25.05 0.8425	25.47 0.8628	25.50 0.8637	26.24 0.8949
Noisy Parthenon	26.09 0.6842	26.12 0.6855	26.06 0.6819	26.08 0.6833	26.07 0.6803	26.08 0.6814	26.42 0.7008
Noisy Plants	30.94 0.8288	30.95 0.8292	30.89 0.8285	30.91 0.8291	31.06 0.8348	31.06 0.8344	31.81 0.8606
Noisy Raccoon	27.99 0.6883	27.98 0.6882	28.00 0.6880	27.98 0.6871	28.00 0.6815	27.99 0.6814	28.02 0.6812

Table:4 PSNR and SSIM results of reconstructed HR images (Noise Level $\sigma_n = 5$)

CONCLUSION

Image restoration (IR) is a fundamental topic in image processing and computer vision applications, and it has been widely studied. In this paper, we investigated IR with the sparse coding techniques. A novel sparse representation-based image deblurring and (single-image) super-resolution method using adaptive sparse domain selection (ASDS) and adaptive regularization (AReg),Centralised Sparse Representation(CSR). Since the optimal sparse domains of natural images can vary significantly across different images and different image patches in a single image, we selected adaptively the dictionaries that were prelearned from a dataset of high-quality example patches for each local patch. To further improve the quality of reconstructed images (reconstructed by using ASDS), we introduced two AReg terms into the ASDS based image restoration framework. One is a set of autoregressive (AR) models were learned from the training dataset and were used to regularize the image local smoothness. The image nonlocal similarity was incorporated as another regularization term to exploit the image nonlocal redundancies. An iterated shrinkage algorithm was proposed to implement the proposed ASDS algorithm with AReg. To further enhance the quality of the reconstructed images we introduced the concept of sparse coding noise (SCN), and it was empirically found that SCN follows Laplacian distributions. To suppress SCN and thus improve the quality of IR, the centralized sparse representation (CSR) model was proposed by exploiting the image nonlocal self-similarity. The experimental results on natural images showed that the proposed approach can significantly outperform other leading IR methods in both PSNR and visual quality(SSIM).

REFERENCES

[1] Weisheng Dong, Lei Zhang, Guangming Shi, Xiaolin Wu, "Image Deblurring and Super-Resolution by Adaptive Sparse Domain Selection and Adaptive Regularization", *IEEE TRANSACTIONS ON IMAGE PROCESSING, VOL. 20, NO. 7, JULY 2011*

[2] M. Bertero and P. Boccacci, *Introduction to Inverse Problems in Imaging*. Bristol, U.K.: IOP, 1998.

[3] D. Donoho, "Compressed sensing," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1289–1306, Apr. 2006.

[4] E. Candès, J. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *IEEE Trans. Inf. Theory*, vol. 52, no. 2, pp. 489–509, Feb. 2006.

[5] M. R. Banham and A. K. Katsaggelos, "Digital image restoration," *IEEE Trans. Signal Process. Mag.*, vol. 14, no. 2, pp. 24–41, Mar. 1997.

[6] L. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Phys. D*, vol. 60, pp. 259–268, 1992.

[7] T. Chan, S. Esedoglu, F. Park, and A. Yip, "Recent developments in total variation image restoration," in *Mathematical Models of Computer Vision*, N. Paragios, Y. Chen, and O. Faugeras, Eds. New York: Springer-Verlag, 2005.

[8] I. Daubechies, M. Defriese, and C. DeMol, "An iterative thresholding algorithm for linear inverse problems with a sparsity constraint," *Commun. Pure Appl. Math.*, vol. 57, pp. 1413–1457, 2004.

[9] S. D. Babacan, R. Molina, and A. K. Katsaggelos, "Total variation super resolution using a variational approach," in *Proc. Int. Conf. Image Process.*, Oct. 2008, pp. 641–644.

[10] D. L. Donoho, "De-noising by soft thresholding," *IEEE Trans. Inf. Theory*, vol. 41, no. 3, pp. 613–627, May 1995.

[11] X. Wu, K. U. Barthel, and W. Zhang, "Piecewise 2-D autoregression for predictive image coding," in *Proc. Int. Conf. Image Process.*, Oct. 1998, vol. 3, pp. 901–904.

[12] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error measurement to structural similarity," *IEEE Trans. Image Process.*, vol. 3, no. 4, pp. 600–612, Apr. 2004.

[13] E. Candes, M. B. Wakin, and S. P. Boyd, "Enhancing sparsity by reweighted minimization," *J. Fourier Anal. Applic.*, vol. 14, pp. 877–905, Dec. 2008.

[14] Lei Zhang a,_, WeishengDong a,b, DavidZhang a, GuangmingShi b," Two-stage image denoising by principal component analysis with local pixel grouping", L. Zhang et al. / Pattern Recognition 43 (2010) 1531–1549