Image Denoising Using Local Tangent Space Alignment

JianZhou Feng[a], Li Song[a], Xiaoming Huo[b],XiaoKang Yang[a] and Wenjun Zhang[a]

[a] Institute of Image Comm. & Information Proc., Shanghai Jiaotong University, 200240, Shanghai, China Shanghai Key Lab of Digital Media Processing and Transmission

[b] School of Industrial and Systems Engineering, Georgia Institute of Technology,Atlanta,GA 30332-0205, USA

ABSTRACT

We propose a novel image denoising approach, which is based on exploring an underlying (nonlinear) low-dimensional manifold. Using local tangent space alignment (LTSA), we 'learn' such a manifold, which approximates the image content effectively. The denoising is performed by minimizing a newly defined objective function, which is a sum of two terms: (a) the difference between the noisy image and the denoised image, (b) the distance from the image patch to the manifold. We extend the LTSA method from manifold learning to denoising. We introduce the local dimension concept that leads to adaptivity to different kind of image patches, e.g. flat patches have lower dimension. We also plug in a basic denoising stage to estimate the local coordinate more accurately. It is found that the proposed method is competitive: its performance surpasses the K-SVD denoising method.

Keywords: Image denoising, manifold, local tangent space alignment

1. INTRODUCTION

Image denoising plays a fundamental role in many applications. It is a typical inverse problem, and can be seen as a test bench for lots of models and techniques. Among the series of new methods, sparse representation has been exploited for image and video denoising. Interesting and inspiring results have been observed, e.g. Elad and coauthors studied image denoising problem with an overcomplete dictionary.[1] We review the formulation and key ideas in Elad's work, and propose an alternative. We will demonstrate that our new method extend Elad's approach from linear representations to nonlinear models. We consider an image model as follows

$$Y = X + \varepsilon, \tag{1}$$

where $Y \in \mathbb{R}^{n\times n}$ is an observed image, $X \in \mathbb{R}^{n\times n}$ is the underlying "true" noise-free image, and $\varepsilon \in \mathbb{R}^{n\times n}$ are small additive noises. When ε_{ij} (the (i,j) entry of ε) are i.i.d. with normal, the ε is a whitenoise. The above is suitable for a grayscale image. For videos, we can think of X, Y, ε as 3-dimensional tensors. For color images, one can treat X, Y, ε as a concatenation of three matrices. For simplicity, in the present paper, we focus on grayscale images.

We define a patch operator. Let $I = (a, b), 1 \le a, b \le n$, denote an index in an image. For fixed integer $w \ge 0$, we define a patch of Y as

$$Y_I = \{Y_{ij} : a - w \le i \le a + w, b - w \le j \le b + w\}. \tag{2}$$

A patch of X, denoted by X_I, is defined similarly. Recall that Elad's approach is to explore the sparse representation of X_I's: namely, suppose $\mathcal{D}$ corresponds to a large dictionary, whose elements form the columns of matrix $\mathbf{D}$. It is assume that $X_I = \mathbf{D} \cdot \beta_I$, where X_I should be viewed as a vector and β_I is a coefficient vector. The sparsity assumption is that the number of nonzero entries in β_I is small for near every I. Elad's paper[1] derives a complete numerical approach (named K-SVD) to explore the sparsity. We review their numerical approach at a high level, because it helps to explain our proposed method. Elad's approach iteratively minimizes an objective function that is of the following form:

$$\min_X \quad \lambda \|X - Y\|_F^2 + F(X), \tag{3}$$

Emails: {bravefjz,song_li}@sjtu.edu.cn, xiaoming.huo@isye.gatech.edu, {xkyang,zhangwenjun}@sjtu.edu.cn,

where $\|\cdot\|_F^2$ denote the Frobenius norm of a matrix, and $F(X)$ is a function which corresponds to the sparsity of the representations in an overcomplete dictionary. To avoid unnecessary repetition, we refer to the original paper for the definition of $F(X)$. A limitation of Elad's $F(X)$ is that it depends on a linear representation. Conceptually, each X_I must reside on a low-dimensional hyperplane that is spanned by a few columns of $\mathbf{D}$. The following scenario is more likely to be true, based on recently reported numerical studies in computer vision, X_I are located on a nonlinear low-dimentional manifold. This paper is to introduce an alternative to $F(X)$, such that the new $F(x)$ represent how close X_I's are to a manifold. We derive a numerical approach to utilize the above idea.

2. REVISIT OF LTSA

Local tangent space alignment (LTSA)[2] was originally introduced as a nonlinear dimension reduction method. To motivate our definition of $F(X)$ in (3), we review the main idea of LTSA.

Consider $z_i \in \mathbb{R}^D$, $i = 1, 2, \cdots, n$. Here D is a large integer. The above z_i can be thought of as a vectorized X_I. Let $Z_i \in \mathbb{R}^{D\times(k+1)}$ denote a matrix whose columns are z_i itself and its k-nearest neighbors. LTSA is made of two steps. In the first step, for fixed i, we compute

$$\min_{L_i, A_i} \quad \|Z_i\overline{P_k} - L_iA_i\|_F^2, \tag{4}$$

where $\overline{P_k} = I_{k+1} - \frac{1}{k+1}\mathbf{1}\mathbf{1}^T$, and $\mathbf{1}^T = (1, \cdots, 1)$, and $L_i \in \mathbb{R}^{D\times d}$ is orthogonal ($L_i^TL_i = I_d$). It is known that L_i is made by the first d left singular vectors of $Z_i\overline{P_k}$. Matrix $A_i \in \mathbb{R}^{d\times(k+1)}$ are the coordinations of z_i and its k-nearest neighbors in a local tangent space with d dimensions. It is always assumed that d is much smaller than D ($d \ll D$). In the second step of LTSA, we solve

$$\min_{\Theta, \{T_i\}} \quad \sum_{i=1}^{n} \|\Theta_i\overline{P_k} - T_iA_i\|_F^2, \tag{5}$$

where $\Theta \in \mathbb{R}^{d\times n}$, $\Theta = (\theta_1, \cdots, \theta_n)$ and θ_i is a new representor of z_i, however with lower dimension: $\theta_i \in \mathbb{R}^d$. In addition, $\Theta_i \in \mathbb{R}^{d\times(k+1)}$ is defined as the representor of Z_i accordingly and A_i is inherited from the step 1. It seems to be more natural to require that $T_i \in \mathbb{R}^{d\times d}$ be orthogonal ($T_i^TT_i = I_d$) since L_i in (4) is orthogonal. But the optimization problem becomes too complicated under this restriction so that we will not consider the constraint in this paper. From the above, it is evident that the main motivation of LTSA is to retain the representations in local tangent spaces. Note that locally linear embedding[3] (LLE) adopts a similar idea, and has attracted a lot of attention. LLE retains the local convex representations, and has a limitation: the domain of the manifold must be convex. When the manifold is not convex, LTSA still applies, however LLE does not.

3. THE MANIFOLD-BASED IMAGE DENOISING ALGORITHM

3.1 The manifold-based cost function

Recall our objective is to define the new cost function $F(X)$. Recall in the second step of LTSA, the following quantity

$$\min_{T_i} \quad \|\Theta_i\overline{P_k} - T_iA_i\|_F^2, \tag{6}$$

can be treated as a measure on the closeness to a manifold structure at z_i. It is shown in Huo's report[4] that the following lemma is true.

LEMMA 3.1. *If V_i denote a matrix that is made by the first d right singular vectors of aforementioned $Z_i\overline{P_k}$, $1 \le i \le n$, we have*

$$\min_{T_i} \|\Theta_i\overline{P_k} - T_iA_i\|_F^2 = trace(\Theta_i(\overline{P_k} - V_i^TV_i)\Theta_i^T).$$

Moreover, the solution to the objective function in LTSA can be described as solution to the following problem:

$$\min_{\Theta} \quad trace(\Theta M\Theta^T),$$

where matrix $M \in \mathbb{R}^{n\times n}$ has the form

$$M = \begin{pmatrix} S_1 & S_2 & \cdots & S_n \end{pmatrix} \begin{pmatrix} \overline{P_k} - V_1^T V_1 & & & \\ & \overline{P_k} - V_2^T V_2 & & \\ & & \ddots & \\ & & & \overline{P_k} - V_n^T V_n \end{pmatrix} \begin{pmatrix} S_1^T \\ S_2^T \\ \vdots \\ S_n^T \end{pmatrix},$$

and S_i is a selection matrix, i.e., $\Theta_i = \Theta \cdot S_i$, and columns of $\Theta \in \mathbb{R}^{d\times n}$ are the low-dimensional θ_i's.

In this paper, we give a brief proof of lemma 3.1.

Proof. A_i are the coordinations of Z_i in the local tangent space and L_i is made by the first d left singular vectors of $Z_i\overline{P_k}$ as mentioned. So

$$A_i = L_i^+ Z_i \overline{P_k} = L_i^+ L_i \Lambda_i V_i = \Lambda_i V_i,$$

where Λ_i is a diagonal matrix whose diagonal elements are the largest d singular values of $Z_i\overline{P_k}$.

Replacing A_i by $\Lambda_i V_i$ in (6) and using the fact that $V_i\overline{P_k} = V_i$, we get

$$\begin{aligned} &\min_{T_i} \|\Theta_i\overline{P_k} - T_i A_i\|_F^2 = \min_{T_i'} \|\Theta_i\overline{P_k} - T_i' V_i\|_F^2 \\ &= \|\Theta_i\overline{P_k}(I - V_i^T V_i)\|_F^2 = \|\Theta_i(\overline{P_k} - V_i^T V_i)\|_F^2 \\ &= trace(\Theta_i(\overline{P_k} - V_i^T V_i)\Theta_i^T). \end{aligned}$$

So the solution of (5) can be described as

$$\begin{aligned} &\min_{\Theta,\{T_i\}} \sum_{i=1}^{n} \|\Theta_i\overline{P_k} - T_i A_i\|_F^2 \\ =&\min_{\Theta} \sum_{i=1}^{n} trace(\Theta_i(\overline{P_k} - V_i^T V_i)\Theta_i^T) \\ =&\min_{\Theta} \sum_{i=1}^{n} trace(\Theta \cdot S_i(\overline{P_k} - V_i^T V_i)S_i^T \cdot \Theta^T) \\ =&\min_{\Theta} trace(\Theta M \Theta^T). \end{aligned}$$

Proof finished. □

The above motivates the following definition of the new cost function $F(X)$:

1. Vectorize each X_I to be a vector as z_i in LTSA.
2. Find the k-nearest neighbors for each z_i, compute for V_i and selection matrix S_i. Define M as in lemma 3.1.
3. For X, construct $\widetilde{X}$ such that each column of $\widetilde{X}$ is a vectorize X_I. We define

   $$F(X) = trace(\widetilde{X} M \widetilde{X}^T),$$

   where $\widetilde{X} = (R_1X, R_2X, \cdots, R_nX)$ and the matrix R_i is the selection matrix which extracts a patch from X. Note that we call both R_i and S_i selection matrix but their function are different, for R_i selects pixels in a block from the whole image while S_i selects $k+1$ patches from the total n patches set.

   It is evident that $F(X)$ is a manifold-based penalty:

   (a) If X_I's are on a manifold, then $F(X)$ is close to zero;

   (b) If X_I's are not on a manifold, one tends to have a large $F(X)$.

The new cost function $F(X)$ can also be represented as

$$\sum_{i=1}^{n} \|\widetilde{X} S_i(\overline{P_k} - V_i^T V_i)\|_F^2, \tag{7}$$

we shall use this equivalent form when solving the image denoising problem.

3.2 Image denoising algorithm using the new cost function

Under the manifold assumption, replacing $F(X)$ using (7) in (3), the new objective function is in the form of

$$\begin{aligned} &\min_X \lambda\|X - Y\|_F^2 + F(X) \\ =&\min_X \lambda\|X - Y\|_F^2 + \sum_{i=1}^{n} \|\widetilde{X} S_i(\overline{P_k} - V_i^T V_i)\|_F^2. \end{aligned} \tag{8}$$

For each i,

$$\begin{aligned} &\|\widetilde{X} S_i(\overline{P_k} - V_i^T V_i)\|_F^2 \\ =&\| \begin{pmatrix} R_i X & R_i^{(1)} X & \cdots & R_i^{(k)} X \end{pmatrix} (\overline{P_k} - V_i^T V_i)\|_F^2 \\ =&\sum_{j=1}^{k+1} \| \begin{pmatrix} R_i X & R_i^{(1)} X & \cdots & R_i^{(k)} X \end{pmatrix} \begin{pmatrix} w_{1j}^i \\ w_{2j}^i \\ \vdots \\ w_{(k+1)j}^i \end{pmatrix} \|_2^2 \\ =&\sum_{j=1}^{k+1} \|A_i^j X\|_2^2, \end{aligned} \tag{9}$$

where $A_i^j = w_{1j}^i R_i + w_{2j}^i R_i^{(1)} + \cdots + w_{(k+1)j}^i R_i^{(k)}$ and w_{st}^i is the (s,t) entry of matrix $W^i = \overline{P_k} - V_i^T V_i$. So objective function (8) is equivalent to

$$\min_X \lambda\|X - Y\|_F^2 + \sum_{i=1}^{n}\sum_{j=1}^{k+1} \|A_i^j X\|_2^2, \tag{10}$$

and the close form solution of is

$$\hat{X} = (\lambda I + \sum_{i=1}^{n}\sum_{j=1}^{k+1} ({A_i^j}^T A_i^j))^{-1} \lambda Y. \tag{11}$$

For calculation simplicity, using the fact that W^i is a symmetric projection matrix, we have

$$\begin{aligned} \sum_{j=1}^{k+1} ({A_i^j}^T A_i^j) &= \sum_{j=1}^{k+1} (w_{1j}^i R_i + w_{2j}^i R_i^{(1)} + \cdots + w_{(k+1)j}^i R_i^{(k)})^T (w_{1j}^i R_i + w_{2j}^i R_i^{(1)} + \cdots + w_{(k+1)j}^i R_i^{(k)}) \\ &= \sum_{j=1}^{k+1}\sum_{r=1}^{k+1}\sum_{s=1}^{k+1} w_{rj}^i {R_i^{(r-1)}}^T w_{sj}^i R_i^{(s-1)} \\ &= \sum_{r=1}^{k+1}\sum_{s=1}^{k+1} (\sum_{j=1}^{k+1} w_{rj}^i w_{sj}^i) {R_i^{(r-1)}}^T R_i^{(s-1)} \\ &= \sum_{r=1}^{k+1}\sum_{s=1}^{k+1} w_{rs}^i {R_i^{(r-1)}}^T R_i^{(s-1)}, \end{aligned} \tag{12}$$

Task: Denoising the given image Y knowing the noise power σ.

1. ***Basic denoising stage with*** d_i ***derived:***
   - Vectorize the image patch of Y to be z_i.
   - For each z_i, find k nearest neighbors and form Z_i.
   - Decompose $Z_i\overline{P_k}$ using singular value decomposition(SVD), $Z_i\overline{P_k} = U\Lambda V$. The singular values are $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_D \geq 0$.
   - Set $d_i = min_d\{d|\sum_{j=d}^{D}\lambda_j^2 < \alpha(k+1)D\sigma^2\}$.
   - Set $d_i' = min_d\{d|\sum_{j=d}^{D}\lambda_j^2 < \beta(k+1)D\sigma^2\}$.
   - Project $Z_i\overline{P_k}$ onto the first d_i' eigenvectors and seem the projections as the estimations of corresponding image patches.
   - Average the estimations at each pixel to get X'.
2. ***Manifold based denoising stage:***
   - Vectorize the image patch of X' to be z_i.
   - For each z_i, find k nearest neighbors and form Z_i.
   - Decompose $Z_i\overline{P_k}$ using singular value decomposition(SVD) $Z_i\overline{P_k} = U\Lambda V$. The singular values are $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_D \geq 0$.
   - Derive V_i to be the upper d_i rows of V.
   - Set $err_i = \sum_{j=d_i+1}^{D}\lambda_j^2$.
   - Set $\lambda = \gamma\frac{\text{mean}(err_i)}{(k+1)D\sigma^2}$.
   - $\hat{X} = (\lambda I + \sum_{i=1}^{n}\sum_{j=1}^{k+1}(A_i^{j^T}A_i^j))^{-1}\lambda Y$.

Figure 1. Manifold based image denoising algorithm

where $R_i^{(r-1)^T}R_i^{(s-1)}$ is a sparse matrix with all of its nonzero elements being 1 and the position of these 1s are easy to calculate.

Since LTSA is not designed for image denoising, we modified the LTSA algorithm from the following two aspects.

1. LTSA assumes that the manifold has a global dimension d while experiments show that the local space of different kind of image patches, i.e. different points on the manifold, are different. For example, flat image patches lie in a lower dimensional space. So we introduce the local dimension to extend the LTSA method. In our extension, d_i is the local dimension of z_i's local tangent space so that $V_i \in \mathbb{R}^{d_i\times(k+1)}$ can have size adaptive rows. Setting a proper value for each d_i become a task of our new algorithm.
2. LTSA requires the noise free image X for calculating V_i while we only have badly defected image Y in real application. One approximate solution is using Y to derive V_i, but experiment results is poor, especially when the noise level is high. So in our proposed algorithm, we add a basic denoising stage which construct a basic denoised image X' for deriving more accurate V_i.

In the manifold based image denoising algorithm, as shown in Fig. 1, we apply a two stage denoising strategy. The basic denoising stage is actually projecting the noisy image patches onto a low dimension space and average the projections over each pixel. It is related to the local linear projection.[5] In this stage, we also calculate d_i, the local dimension estimation of the manifold at z_i. We choose $\alpha(k+1)D\sigma^2$ to be the threshold because the power of noise contained in Z_i is $(k+1)D\sigma^2$ and α is an adjustable factor. We use another factor β to adjust d_i', the

Table 1. Summary of the denoising PSNR results in decibels. In each cell, two denoising results are reported and the higher one is bolded. Left: Results of Elad etc.[1] Right: Results of the method proposed in this paper. We set $k = 100$, $\alpha = 0.5$, $\beta = 0.8$, $\gamma = 7700$. The factors are chosen through experiment results.

σ/PSNR	Lena		Barb		Boats		House		Peppers	
2/42.14	42.62	42.62	42.84	**43.00**	42.60	42.60	45.14	**45.20**	**45.32**	45.13
5/34.12	38.42	**38.45**	37.17	**37.47**	36.15	**36.20**	39.19	**39.95**	41.94	**42.83**
10/28.08	35.59	**35.75**	34.42	**34.77**	**31.72**	31.64	38.43	**38.86**	33.61	**34.07**
15/24.58	36.11	**36.53**	32.62	**32.81**	31.64	**31.81**	**39.37**	39.27	33.89	**34.51**
20/22.06	30.83	**31.48**	29.90	**30.58**	29.81	**30.07**	**33.82**	33.69	35.22	**35.27**
25/20.19	30.51	**30.82**	30.73	**31.70**	28.76	**29.45**	30.53	**30.84**	31.28	**31.65**

target dimension of the space that we shall project Z_i onto. In the manifold based denosing stage, we estimate the noise power err_i contained in the basic denoised image X' and use it as an estimation of the distance between X and the learnt manifold to adjust λ. At last, we use (11) to obtain the denoised image $\hat{X}$.

4. EXPERIMENT RESULTS

In this section, we demonstrate the results achieved by applying the above method on several test images. The tested images and noise levels are the same as reported in Elad's work.[1] As our proposed method is highly computational demanding currently, we just denoise a randomly chosen 64×64 block from each original image or denoise the downsampled version of the image. The size of image patch used in both denosing algorithm is 8×8 so that the dimension $D = 64$.

We first compare the results on randomly chosen blocks. Table 1 show us the PSNR improvements of our method and Fig. 2 show the visual results. Then we do the same experiments on the downsampled images. All the images are downsampled to 64×64. The results are shown in Fig. 3. Further experiments such as denoising the normal size images will be done in future research.

5. CONCLUSION AND FUTURE WORK

This work has presented a novel manifold based image denoising method, leading to state-of-the-art performance and surpassing the the K-SVD denoising method. The proposed method extend the LTSA by introducing local dimension and adding basic denoising step into it. The results in the paper is preliminary and we have a lot to do in the future, e.g. reduce the computation complexity so that it can be used for large images. More analysis of the image properties can make the performance even better.

6. ACKNOWLEDGEMENT

This work is supported by National Natural Science Foundation of China (60702044, 60932006, 60625103) and 973 Program (2010CB731401).

REFERENCES

1. M. Elad and M. Aharon, "Image denoising via sparse and redundant representations over learned dictionaries," *IEEE Transactions on Image Processing*, vol. 15, no. 12, December 2006.
2. Z. Zhang and H. Zha, "Principal manifolds and nonlinear dimensionality reduction via tangent space alignment," *Society for Industrial and Applied Mathematics*, vol. 26, no. 1, pp. 313–338, 2004.
3. S. Roweis and L. Saul, "Nonlinear dimensionality reduction by locally linear embedding," *Science*, vol. 290, pp. 2323–2326, 2000.

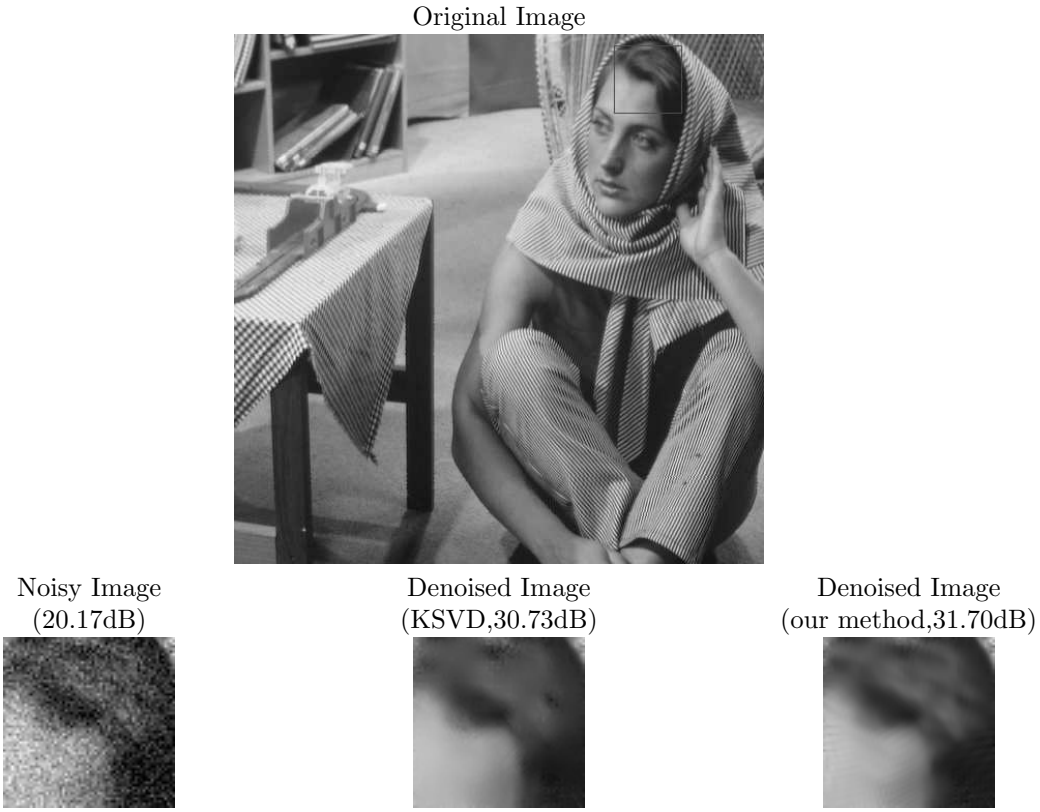

Figure 2. Example of the denoising results for the image "Barbara" with $\sigma = 25$, this patch is extracted from the head of the woman.

4. X. Huo and A. K. Smith, "Performance analysis of a manifold learning algorithm in dimension reduction," *http://www2.isye.gatech.edu/statistics/papers/06-06.pdf*.
5. X. Huo and J. Chen, "Local linear projection (LLP)," in *First IEEE Workshop on Genomic Signal Processing and Statistics (GENSIPS)*, Raleigh, NC, October 2002.

Figure 3. Original image, noisy image, denoised image of K-SVD and our proposed algorithm($\sigma = 15$).