CONTENT-ADAPTIVE LOW RANK REGULARIZATION FOR IMAGE DENOISING

Hangfan Liu[1], Xinfeng Zhang[2] and Ruiqin Xiong[1]

[1] Institute of Digital Media, Peking University, Beijing 100871, China
[2] Rapid-Rich Object Search (ROSE) Lab, Nanyang Technological University, Singapore 639798
Email: {liuhf, rqxiong}@pku.edu.cn, xfzhang@ntu.edu.sg

ABSTRACT

Prior knowledge plays an important role in image denoising tasks. This paper utilizes the data of the input image to adaptively model the prior distribution. The proposed scheme is based on the observation that, for a natural image, a matrix consisted of its vectorized non-local similar patches is of low rank. We use a non-convex smooth surrogate for the low-rank regularization, and view the optimization problem from the empirical Bayesian perspective. In such framework, a parameter-free distribution prior is derived from the grouped non-local similar image contents. Experimental results show that the proposed approach is highly competitive with several state-of-art denoising methods in PSNR and visual quality.

***Index Terms*—** Non-local similarity, low-rank, empirical Bayes, image denoising.

1. INTRODUCTION

As a classic image processing task, image denoising still remains an important application nowadays. For example, image signals captured by mobile-phone cameras are often affected by noise in low-light or high-speed conditions, hence an effective denoising approach is required to enhance the perceptual quality. Image denoising aims to recover an image $\mathbf{x}$ from its noisy observation $\mathbf{y}$, which is modeled as $\mathbf{y} = \mathbf{x} + \mathbf{n}$, with $\mathbf{n}$ being the additive noise. Such inverse problem is ill-posed in nature, thus it is necessary to exploit the prior knowledge of the image so as to regularize the solution space.

The most widely used priors are based on the observation that images can be sparsely presented in transform domain and separated from noise [1–4]. Since it would be difficult to efficiently present all kinds of patterns in images using a fixed basis. authors of [5] etc. propose to use content adaptive basis with the assistance of local principal component analysis. Some later works learn over-complete dictionaries to represent the image so that the sparsity is maximized [6–8].

This work was supported in part by the National Natural Science Foundation of China (61370114, 61421062), the National Basic Research Program of China (973 Program, 2015CB351800), Beijing Natural Science Foundation (4132039), Research Fund for the Doctoral Program of Higher Education (20120001110090) and also by the Cooperative Medianet Innovation Center.

The success of non-local means method [9] inspired extensive research works to utilize non-local similarity of natural images [10–21]. Among these structural clustering based schemes, one of the most well-known method is BM3D [10], which basically combines DCT coefficient thresholding and nonlocal block matching. Combining the ideas of structural clustering and dictionary learning, K-LLD [11] observes that similar patches should share similar subdictionaries and utilizes the subdictionaries for image modeling, and CSR [14] unifies local and nonlocal sparsity constraints. A recent line of research adopts low rank matrix approximation methods for image restoration [22–26]. After Wang *et al.* [22] validated the hypothesis that a matrix consists of nonlocal similar patches is of low rank and has sparse singular values, Ji *et al.* [24] employed the nuclear norm (NNM) method for video denoising, and Dong *et al.* [23] unified NNM and $l_{2,1}$-norm group sparsity for image restoration.

This paper aims to learn the optimal prior from the input noisy image, using non-local similar patches as source of data to study the prior distribution. To be specific, the proposed method is built upon the fact that matrixes formed by similar image patches are of low rank. Based on the low-rank regularization, we attempt to derive a parameter-free distribution model for the image signal from the empirical Bayesian perspective, rather than adopting some preset prior which usually requires tuning distribution parameters.

The remainder of this paper is organized as follows. Section 2 explains how the distribution model is developed based on the low-rank non-local similar image patches. Section 3 describes the proposed denoising algorithm. Section 4 shows experimental results and Section 5 concludes the paper.

2. DISTRIBUTION MODELING FOR LOW-RANK IMAGE DATA

2.1. Structural Clustering and Centralization

Non-local similar contents in an image are highly correlated, and matrixes stacked by vectorized version of such correlated contents are of low rank. Although shape-adaptive patches may further enhance the low rank property via selecting the most locally-correlated pixels to form a patch, this paper considers square patches of size $m \times m$ to reduce complexity.

We denote the vectorized image patch at location i in $\mathbf{y}$ by $\mathbf{y}_i \in \mathbb{R}^{M\times 1}$, with $M = m^2$. For each $\mathbf{y}_i$, the structural clustering procedure is performed by gathering its N most similar patches within $\mathbf{y}$ (including $\mathbf{y}_i$ itself). The dissimilarity between two patches $\mathbf{y}_i$ and $\mathbf{y}_j$ is measured by the Euclidean distance of the pixel intensities:

$$d(i,j) = \frac{1}{M}\left\|\mathbf{y}_i - \mathbf{y}_j\right\|_2^2. \quad (1)$$

We record the index of these highly correlated patches in a set $\mathbb{L}_i$, and centralize the patches via

$$\bar{\mathbf{y}}_j = \mathbf{y}_j - \mu_i,\ \mu_i = \frac{1}{N}\sum_j \mathbf{y}_j \quad (2)$$

for all $j \in \mathbb{L}_i$. Then $Y_i = \{\bar{\mathbf{y}}_j\}_{j\in\mathbb{L}_i} \in \mathbb{R}^{M\times N}$ is the corresponding low-rank matrix, and X_i is the desired clean version of Y_i. With a slight abuse of notation, the rest of this paper omits the subscripts of X_i and Y_i for simplicity when the risk of confusion is little.

2.2. Low-Rank Regularization with Non-Convex Majorization - Minimization

Based on low-rank regularization, the estimation of patch group X given observation Y is formulated as

$$X = \arg\min_X \frac{1}{2\lambda}\left\|Y - X\right\|_2^2 + \text{Rank}(X), \quad (3)$$

where λ is the parameter balancing the contribution of the two competing terms. To conquer this NP-hard optimization problem, many existing works relax the rank penalty to convex terms such as nuclear norm, but it has been recognized that non-convex surrogates are potentially better approximation for rank minimization, hence we substitute the log-det term $\log\left|XX^{\mathrm{T}}\right|$ for the rank penalty in (3) as suggested in [27]:

$$X = \arg\min_X \frac{1}{2\lambda}\left\|Y - X\right\|_2^2 + \log\left|XX^{\mathrm{T}}\right|. \quad (4)$$

To tackle (4), we may employ a majorization-minimization method to derive a strict, convex upper-bound on the non-convex term [27]:

$$\log\left|XX^{\mathrm{T}}\right| = \min_\Phi \frac{1}{N}\text{Trace}\left(XX^{\mathrm{T}}\Phi^{-1}\right) + \log|\Phi| + C, \quad (5)$$

where $\Phi \in \mathbb{R}^{M\times M}$ is a positive semi-definite matrix of variational parameters and C is a constant. Combining (4) and (5), we arrive at the optimization problem

$$\min_{X,\Phi} \frac{1}{2\lambda}\left\|Y - X\right\|_2^2 + \frac{1}{N}\text{Trace}\left(XX^{\mathrm{T}}\Phi^{-1}\right) + \log|\Phi|. \quad (6)$$

Algorithm 1: Proposed Image Denoising Algorithm

Data: Noisy image $\mathbf{y}$;
for i = 1, 2, 3, … $PatchNumber$ **do**
  Calculate distances of non-local patches by (1);
  Group similar patches of $\mathbf{y}_i$ into Y_i;
  Centralize Y_i by (2);
  Initialize Φ_i as $\Phi_i^0 = Y_i Y_i^{\mathrm{T}}/N$;
  for k = 1, 2, 3, … K **do**
    Calculate X_i via (13);
    Update Φ_i by solving (12), i.e. $\Phi_i^{k+1} = \frac{1}{N}\left(X_iX_i^{\mathrm{T}} + \left(\Phi_i^k - \Phi_i^k\left(\Phi_i^k + \lambda I\right)^{-1}\Phi_i^k\right)\right)$;
  end
  Add back the mean of patches in X_i;
end
Aggregate estimated patches into $\mathbf{x}$;
Result: Denoised image $\mathbf{x}$.

2.3. Empirical Bayes Based Distribution Modeling

The proposed content-adaptive image prior distribution is based on (4) and (6). In Bayesian framework, we may interpret (4) as a maximum a posteriori (MAP) estimation solving

$$X = \arg\max_X p(X|Y) = \arg\max_X p(Y|X)p(X), \quad (7)$$

where

$$p(X) \propto \frac{1}{\left|XX^{\mathrm{T}}\right|} \quad (8)$$

describes the prior distribution, and

$$p(Y|X) \propto \exp\left\{-\frac{1}{2\lambda}\left\|Y - X\right\|_2^2\right\} \quad (9)$$

implies that the noise is white Gaussian with variance λ.

Unfortunately, there is no effective means to reach the global optimum of (7) to the best of our knowledge. Enlightened by the success of rank minimization approach described in [28], we approximate $p(X)$ by its strict lower bound $\hat{p}(X)$, which is formulated according to (5) as:

$$\hat{p}(X;\Phi) = \exp\left\{-\frac{1}{N}\text{Trace}\left(XX^{\mathrm{T}}\Phi^{-1}\right) - \log|\Phi|\right\}. \quad (10)$$

In this way, the approximate posterior distribution maintaining the original posterior mode is given by

$$\hat{p}(X|Y;\Phi) = \frac{p(Y|X)\hat{p}(X;\Phi)}{\int p(Y|X)\hat{p}(X;\Phi)dX}, \quad (11)$$

which has closed form first and second moments, converting the MAP estimation (7) to a tractable problem.

Table 1. PSNR Comparison of BM3D [10], LPG-PCA [29], CSR [14], LASSC [23] and the proposed method (Unit: dB).

σ_n	10					20				
Schemes	*BM3D*	*L.PCA*	*CSR*	*LASSC*	*Proposed*	*BM3D*	*L.PCA*	*CSR*	*LASSC*	*Proposed*
Airplane	35.98	35.57	35.93	35.81	**36.27**	32.71	32.06	32.62	32.45	**32.92**
Barbara	34.78	34.99	35.02	35.13	**35.32**	31.24	30.94	31.20	31.32	**31.44**
C.man	34.09	33.61	33.93	34.21	**34.34**	30.41	29.73	30.29	28.56	**30.66**
Lena	35.22	35.01	35.31	35.33	**35.50**	31.51	31.07	31.58	29.68	**31.65**
Monarch	34.19	34.09	34.52	34.77	**35.01**	30.42	30.09	30.67	30.19	**31.06**
R.R.Hood	35.13	34.88	35.17	32.56	**35.39**	32.25	31.72	32.19	29.41	**32.28**
Sailboats	36.46	36.16	36.38	34.92	**36.59**	33.11	32.49	32.92	31.83	**33.13**
Window	36.75	36.31	36.85	35.55	**37.13**	32.97	32.16	32.89	31.71	**33.17**
Baboon	29.55	29.40	29.37	29.50	**29.66**	25.57	25.39	25.53	25.57	**25.73**
Elaine	33.34	**33.82**	33.81	33.65	33.74	**31.46**	31.36	31.45	31.42	31.41
F.boat	33.89	33.62	33.83	33.51	**34.02**	30.81	30.24	30.72	30.71	**30.84**
House	36.71	36.17	36.82	36.65	**36.89**	33.78	33.05	33.87	31.87	**33.97**
Peppers	34.73	34.11	34.68	34.76	**34.94**	31.30	30.58	31.26	27.86	**31.49**
straw	30.93	31.35	31.49	31.57	**31.66**	27.08	27.08	27.42	27.19	**27.52**
Average	34.41	34.22	34.51	34.14	**34.75**	31.04	30.57	31.04	29.98	**31.23**
σ_n	30					40				
Schemes	*BM3D*	*L.PCA*	*CSR*	*LASSC*	*Proposed*	*BM3D*	*L.PCA*	*CSR*	*LASSC*	*Proposed*
Airplane	31.08	30.23	30.88	30.80	**31.21**	29.79	28.94	29.57	29.83	**29.95**
Barbara	29.02	28.46	28.83	29.13	**29.30**	27.14	26.76	26.76	**27.78**	27.51
C.man	28.65	27.83	28.54	27.38	**28.74**	27.20	26.52	27.12	26.84	**27.46**
Lena	29.45	28.79	29.44	27.96	**29.61**	27.87	27.24	28.03	27.65	**28.15**
Monarch	28.37	27.79	28.51	28.27	**28.82**	26.73	26.14	26.81	27.02	**27.27**
R.R.Hood	30.73	28.10	30.64	29.11	**30.77**	**29.62**	28.97	29.57	28.92	29.60
Sailboats	31.21	30.43	31.01	30.38	**31.25**	29.81	29.01	29.70	29.51	**29.85**
Window	30.81	29.76	30.63	30.05	**31.04**	29.21	28.13	29.06	29.26	**29.43**
Baboon	23.77	23.49	23.78	23.88	**24.05**	22.49	22.35	22.71	22.85	**22.94**
Elaine	**30.41**	30.05	30.37	30.34	30.33	29.48	29.02	**29.64**	29.55	29.37
F.boat	29.02	28.26	28.89	28.84	**29.03**	27.63	26.88	27.58	27.62	**27.67**
House	32.09	31.12	32.07	30.86	**32.40**	30.68	29.59	30.70	30.83	**31.09**
Peppers	29.30	28.47	29.24	27.56	**29.40**	27.70	26.93	27.76	27.05	**27.94**
straw	24.93	24.55	25.03	25.21	**25.38**	23.18	22.92	23.58	**23.89**	23.81
Average	29.20	28.52	29.13	28.56	**29.38**	27.75	27.10	27.80	27.76	**28.00**

3. PROPOSED DENOISING ALGORITHM

In order to decide a proper hyper-parameter Φ, we borrow the wisdom of [30] to solve

$$\Phi = \arg\min_{\Phi} \int p(Y|X) \left| p(X) - \hat{p}(X;\Phi) \right| dX, \qquad (12)$$

which minimizes the gap between $p(X)$ and $\hat{p}(X)$ iff the likelihood $p(Y|X)$ is significant. Given Φ, it is straightforward to calculate X by

$$X = \Phi(\Phi + \lambda I)^{-1} Y \qquad (13)$$

since the posterior distribution (11) is Gaussian. Here $(\Phi + \lambda I)$ can be seen as the covariance of every column in Y, $I \in \mathbb{R}^{M\times M}$ is the identity matrix. This is in fact assuming that non-locally searched similar image contents share the same statistical characteristics.

After obtaining all the denoised groups X_i, we add back the mean μ_i for each patch in group i, and then get the full image $\mathbf{x}$ by putting back the patches and averaging overlaps. So far, the main idea of this paper has been explained, we summarize the proposed denoising procedure in Algorithm 1.

4. EXPERIMENTAL RESULTS

In this section we evaluate the effectiveness of the proposed method by comparing it with several state-of-art denoising methods, including BM3D [10], LPG-PCA [29], CSR [14] and LASSC [23]. LASSC is a recent successful low-rank based denoising scheme, and the other three anchors, especially BM3D, are among the most famous anchors in the past

Fig. 1. Denoised fragments of *Cameraman*. From left to right: (a) Noisy image (σ_n = 20); (b) BM3D (PSNR = 30.41dB); (c) LPG-PCA (PSNR = 29.73dB); (d) CSR (PSNR = 30.29dB); (e) LASSC (PSNR = 28.56dB); (f) Proposed (PSNR = **30.66**dB).

Fig. 2. Denoised fragments of *Monarch*. From left to right: (a) Noisy image (σ_n = 30); (b) BM3D (PSNR = 28.37dB); (c) LPG-PCA (PSNR = 27.79dB); (d) CSR (PSNR = 28.51dB); (e) LASSC (PSNR = 28.27dB); (f) Proposed (PSNR = **28.82**dB).

Fig. 3. Denoised *Red Riding Hood*. From left to right: (a) Noisy fragment (σ_n = 40); (b) BM3D (PSNR = 30.73dB); (c) LPG-PCA (PSNR = 28.10dB); (d) CSR (PSNR = 30.64dB); (e) LASSC (PSNR = 29.11dB); (f) Proposed (PSNR = **30.77**dB).

few years. The noise is modeled as white Gaussian with known variance σ_n^2, which is used as an input parameter for all the five competing methods. The parameter settings of the anchor source codes are not changed.

This paper uses peak signal-to-noise ratio (PSNR) as the objective assessment for denoising performance. We display the PSNR scores of 15 denoised images in Table 1. The bold highlight the best scores. It is obvious that, in most cases, the proposed method achieves the highest PSNR scores, and enjoys evident gain over the four anchors in average at different noise levels. More importantly, the proposed method produces restored images of better perceptual quality. As can be seen from Fig. 1-3, the output of the proposed scheme tends to be clearer with more details and less artifacts.

5. CONCLUSION

In order for a prior that reflects the statistical characteristics of image contents, this paper utilizes the data from the input noisy image to learn the distribution. Taking advantage of self-similarity of natural images, the proposed scheme searches for highly correlated non-local contents within the image to form low-rank matrixes. Based on the variational bound of the low-rank penalty, we derive the prior distribution in empirical Bayesian framework, leading to an effective posterior approximation. Experimental results demonstrate that the proposed scheme achieves better performance than state-of-art denoising methods, including BM3D and LASSC, in terms of both objective and subjective qualities.

6. REFERENCES

[1] D. Donoho, “De-noising by soft-thresholding,” *IEEE Transactions on Information Theory*, vol. 41, no. 3, pp. 613–627, May 1995.

[2] S. Chang, B. Yu, and M. Vetterli, “Adaptive wavelet thresholding for image denoising and compression,” *IEEE Transactions on Image Processing*, vol. 9, no. 9, pp. 1532–1546, 2000.

[3] L. Sendur and I. Selesnick, “Bivariate shrinkage functions for wavelet-based denoising exploiting interscale dependency,” *IEEE Transactions on Signal Processing*, vol. 50, no. 11, pp. 2744–2756, 2002.

[4] G. Yu and G. Sapiro, “DCT image denoising: a simple and effective image denoising algorithm,” *Image Processing On Line*, vol. 1, 2011, http://dx.doi.org/10.5201/ipol.2011.ys-dct.

[5] D. Muresan and T. Parks, “Adaptive principal components and image denoising,” in *IEEE International Conference on Image Processing*, vol. 1, Sept 2003, pp. 101–104.

[6] M. Elad and M. Aharon, “Image denoising via sparse and redundant representations over learned dictionaries,” *IEEE Transactions on Image Processing*, vol. 15, no. 12, pp. 3736–3745, Dec 2006.

[7] J. Mairal, M. Elad, and G. Sapiro, “Sparse representation for color image restoration,” *IEEE Transactions on Image Processing*, vol. 17, no. 1, pp. 53–69, 2008.

[8] R. Rubinstein, A. Bruckstein, and M. Elad, “Dictionaries for sparse representation modeling,” *Proceedings of the IEEE*, vol. 98, no. 6, pp. 1045–1057, 2010.

[9] A. Buades, B. Coll, and J. Morel, “A non-local algorithm for image denoising,” in *IEEE Conference on Computer Vision and Pattern Recognition*, June 2005.

[10] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, “Image denoising by sparse 3-D transform-domain collaborative filtering,” *IEEE Transactions on Image Processing*, vol. 16, no. 8, pp. 2080–2095, Aug 2007.

[11] P. Chatterjee and P. Milanfar, “Clustering-based denoising with locally learned dictionaries,” *IEEE Transactions on Image Processing*, vol. 18, no. 7, pp. 1438–1451, July 2009.

[12] M. Lebrun, A. Buades, and J.-M. Morel, “Implementation of the ”non-local bayes” (nl-bayes) image denoising algorithm,” 2013, http://dx.doi.org/10.5201/ipol.2013.16.

[13] H. Liu, R. Xiong, S. Ma, X. Fan, and W. Gao, “Gradient based image transmission and reconstruction using non-local gradient sparsity regularization,” in *IEEE International Conference on Multimedia and Expo*, Jul. 2014, pp. 1–6.

[14] W. Dong, X. Li, D. Zhang, and G. Shi, “Sparsity-based image denoising via dictionary learning and structural clustering,” in *IEEE Conference on Computer Vision and Pattern Recognition*, 2011, pp. 457–464.

[15] H. Liu, R. Xiong, S. Ma, X. Fan, and W. Gao, “Gradient based image/video SoftCast with grouped-patch collaborative reconstruction,” in *IEEE Visual Communications and Image Processing Conference*, Dec. 2014, pp. 141–144.

[16] X. Zhang, R. Xiong, X. Fan, S. Ma, and W. Gao, “Compression Artifact Reduction by Overlapped-Block Transform Coefficient Estimation With Block Similarity,” *IEEE Transactions on Image Processing*, vol. 22, no. 12, pp. 4613–4626, Dec. 2013.

[17] H. Liu, R. Xiong, X. Zhang, Y. Zhang, S. Ma, and W. Gao, “Non-local gradient sparsity regularization for image restoration,” *IEEE Transactions on Circuits and Systems for Video Technology*, 2016.

[18] C. Kervrann, “PEWA: Patch-based exponentially weighted aggregation for image denoising,” *Advances in Neural Information Processing Systems*, vol. 3, pp. 2150–2158, 2014.

[19] J. Zhang, D. Zhao, R. Xiong, S. Ma, and W. Gao, “Image restoration using joint statistical modeling in a space-transform domain,” *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 24, no. 6, pp. 915–928, 2014.

[20] X. Liu, D. Zhai, D. Zhao, G. Zhai, and W. Gao, “Progressive image denoising through hybrid graph laplacian regularization: A unified framework,” *IEEE Transactions on Image Processing*, vol. 23, no. 4, pp. 1491–1503, April 2014.

[21] H. Liu, R. Xiong, J. Zhang, and W. Gao, “Image denoising via adaptive soft-thresholding based on non-local samples,” in *IEEE Conference on Computer Vision and Pattern Recognition*, Jun. 2015, pp. 484–492.

[22] S. Wang, L. Zhang, and Y. Liang, “Nonlocal spectral prior model for low-level vision,” in *11th Asian Conference on Computer Vision*, Nov. 2012, pp. 231–244.

[23] W. Dong, G. Shi, and X. Li, “Nonlocal image restoration with bilateral variance estimation: a low-rank approach,” *IEEE Transactions on Image Processing*, vol. 22, no. 2, pp. 700–711, 2013.

[24] H. Ji, C. Liu, Z. Shen, and Y. Xu, “Robust video denoising using low rank matrix completion,” in *IEEE Conference on Computer Vision and Pattern Recognition*, June 2010.

[25] S. Gu, L. Zhang, W. Zuo, and X. Feng, “Weighted nuclear norm minimization with application to image denoising,” in *IEEE Conference on Computer Vision and Pattern Recognition*, vol. 2, 2014, pp. 2862–2869.

[26] J. Zhang, R. Xiong, C. Zhao, Y. Zhang, S. Ma, and W. Gao, “CONCOLOR: Constrained non-convex low-rank model for image deblocking,” *IEEE Transactions on Image Processing*, vol. 25, no. 3, pp. 1246–1259, 2016.

[27] K. Mohan and M. Fazel, “Iterative reweighted least squares for matrix rank minimization,” in *Forty-Eighth Annual Allerton Conference*, Sep. 2010, pp. 653–661.

[28] D. Wipf, “Non-convex rank minimization via an empirical bayesian approach,” 2014, http://arxiv.org/abs/1408.2054.

[29] L. Zhang, W. Dong, D. Zhang, and G. Shi, “Two-stage image denoising by principal component analysis with local pixel grouping,” *Pattern Recognition*, vol. 43, no. 4, pp. 1531–1549, 2010.

[30] D. P. Wipf, B. D. Rao, and S. Nagarajan, “Latent variable bayesian models for promoting sparsity,” *IEEE Transactions on Information Theory*, vol. 57, no. 9, pp. 6236–6255, 2011.