Low Bit-Rate Compression Image Restoration through Subspace Joint Regression Learning*

Zongliang GAN[a], Member

SUMMARY In this letter, an effective low bit-rate image restoration method is proposed, in which image denoising and subspace regression learning are combined. The proposed framework has two parts: image main structure estimation by classical NLM denoising and texture component prediction by subspace joint regression learning. The local regression function are learned from denoised patch to original patch in each subspace, where the corresponding compression image patches are employed to generate anchoring points by the dictionary learning approach. Moreover, we extend Extreme Support Vector Regression (ESVR) as multi-variable nonlinear regression to get more robust regression results. Experimental results demonstrate the proposed method achieves favorable performance compared with other learning methods.

key words: compression image restoration, subspace regression learning, non-local means denoising, extreme support vector regression

1. Introduction

Block based lossy image compression is the most common approach for image storage and transmission, such as JPEG, JPEG2000, and WebP. It always has low-rate compression images in multimedia application, and unpleasing visual artifacts in high compression rate image can not be ignored. In general, there are three kinds of compression restoration approaches to deal with above problem, including decompression based methods, denoising based methods and learning based methods. For decompression based methods, quantization processing need be considered. Quantization noise sometimes need be modeled, and it can be used as constraint condition to remove compression artifacts [1]. Denoising method in video coding also is a kind of typical postprocessing method, related to quantization parameters. For denoising based methods, some image denoising method, such as [2], [3], can be used to improve compression image quality, in which the quantization noise in the compression image is treated as Gaussian noise. Recently, learning based compression image restoration methods are drawing more attention, and typical methods include the semi-local Gaussian processes-based solution (SLGP) method [4], adjusted an anchored neighborhood regression (A+) method [5] and deep learning based approach [6], [7].

In the following discussion, we denote \( f_{Iy} \) and \( X | Y \) as an original/compression image and patch, respectively. \( f_Y \) is feature of compression patch \( Y \). In training phase, it has three steps in A+ subspace learning based framework [5]. First, the feature \( f_Y \) of \( Y \) are got by 1st and 2nd order gradients, 99% energy maintained PCA dimension reduction and \( l_2 \) normalization. Then, the anchoring points, known as atoms \( D_i \), are obtained by dictionary learning with \( f_Y \). Last, local subspace linear regressors are trained by using ridge regression method. In test phase, after finding the nearest neighboring anchoring points by \( f_Y \) with dictionary \( D \), the patch is estimated through anchoring points regressors and the reconstructed patches are averaged in the overlapping areas to obtain final output image. More details can be found in [5] and the original version [8] for image super resolution.

At present, the main challenge of the A+ based method for low bit-rate compression image restoration is how to get more efficient regression model, since the mapping function from \( Y \) to \( X \) becomes difficult to be learned when the quantization noises very large. In order to reduce the impact of quantization noise and improve performance, the training images are compressed at 0.3 bits per pixel (BPP), and the test images are compressed at 0.1 BPP in [5]. However, the training data and test data may be mismatch in this case.

In this letter, we propose a subspace joint regression learning based image restoration for low bit-rate compression image. Compared to A+ method [4], the proposed method has been significantly extended in the following two aspects: (1) Subspace joint learning based restoration method is proposed, in which the non-local means (NLM) denoising method [9] is integrated into learning framework. Let \( Y_d \) represent the patch in denoised image by NLM and \( Y_d \) mainly represents the main structure component of \( X \). In each the anchoring points defined by A+ approach with \( f_Y \), the aim of regression function \( G(\cdot) \) is to get the relationship from \( Y_d \) to \( X \). In this way, for a given patch \( Y \), we recast the restoration problems to estimate the structure component \( Y_d \) and texture component \( G(f_{Y_d}) \). (2) We use Extreme Support Vector Regression (ESVR) [10] to get the multi-variable nonlinear regression function, and it can improve the mapping performance with fast learning speed. The contribution of this work can be summarized: (1) subspace partitioning is learned from the compression patch \( Y \), and the reconstruction function is got from \( Y_d \) and \( X \), rather than direct combination with denoising and subspace learning methods. (2) the multi-variable ESVR network structure is simplified through the common feature transform struc-
ture without degrading performance. Experimental results indicate that our approach is both quantitatively and qualitatively superior to other application-oriented compression image restoration methods.

2. Proposed Method

2.1 Subspace Joint Learning Based Regression Model

Let \{X^{(i)}, Y^{(i)}, f^{(i)}_Y\}_{i=1,N} define as training samples. The dictionary \(D = \{D_1, \ldots, D_K\}\) is learned from \(\{f^{(i)}_Y\}\) by KSVD method, and there atoms \(\{D_i\}\) constitute the \(K\) anchoring points. For each training sample \(f^{(i)}_Y\) and its nearest neighboring atom with \(D_k\), let \(C_{i,j}\) define as \(C_{k,j} = 1\) and \(C_{i,j} = 0\), \((i \neq k)\). In \(A+\) method [5], the main task is to predict spatial quantization noise, and it can define as \(N^{(i)}_Q = X^{(i)} - Y^{(i)}\). In training processing, the regressors is learned by:

\[
\{P_i\} = \min \sum_{i=1}^{N} \sum_{j=1}^{N} C_{i,j}(N^{(i)}_Q - P_i f^{(i)}_Y)^2 \tag{1}
\]

where ridge regression is used to find the project matrix \(P_i\).

To block based image compression, the quantization noise is not only related to quantization parameters, but also related to the surrounding neighbor pixels. For low bit-rate compression images, the spatial quantization noise is very large and is not a stationary process. In this case, it is very difficult to estimate \(N_Q\) from (1).

It can be assumed that natural image \(I\) consists of two mainly parts: main structure component and texture component. For low bit-rate compression image \(I_c\), the main structure component are retained and most of texture component are discarded after quantization. Meanwhile, the artifacts in \(I_c\), known as compression noise, are added, and they often are high frequency information. Inspired by above observation, our method is mainly to solve two tasks: main structure and texture information estimation. Although quantization noise is not taken into account during image denoising algorithm design in most cases, it is able to recover the image main structure and partly reduce quantization noise. NLM denoising method [9] is used in this letter. The reason is that we want to make our idea more clear, since the performance of NLM is just acceptable. The denoising version \(I_d\) of \(I_c\) is defined as main structure component of the original image \(I\).

Let \(I_c(i)\) as the value of the \(i\)th pixel in \(I_c\), and explicit NLM filter output \(I_d(i)\) can be given by

\[
I_d(i) = \frac{\omega_{ij} \cdot I_c(j)}{\sum_{j} \omega_{ij}}, \tag{2}
\]

\[
\omega_{ij} = \exp\{-\sum_{u} [B(I_c(i), u) - B(I_c(j), u)] / num \cdot h^2 \} \tag{3}
\]

where \(B(I_c(i))\) is a small image patch centered on \(I_c(i)\), \(num\) is pixel number in patch and \(h\) controls the denoising strength. In this letter, patch size, windows size and \(h\) are set as \(7 \times 7, 25 \times 25\) and 0.9.

From now, for the training samples, it can be represented as \(\{X^{(i)}, Y^{(i)}, f^{(i)}_Y\}_{i=1,N}\), in which \(Y^{(i)}_d\) is the corresponding patches from denosing image \(I_d\). Here, we still use \(\{f^{(i)}_Y\}\) to get the anchor points, and it is the same way with the original algorithm [5]. In our tests, the anchor points by dictionary from \(f_d\) can not get desired performance. The most possible reason may be that the denosing processing may make some useful pattern information loss.

Let \(X^{(i)}_b = X^{(i)} - Y^{(i)}\), and the regression task is to predict the texture component of original patch. The anchored regressors can be solved by

\[
\{G_i\} = \min_{\{G_1, \ldots, G_k\}} \sum_{i=1}^{N} \sum_{j=1}^{N} C_{i,j}(X^{(i)}_b - G_j f^{(i)}_y)^2 \tag{4}
\]

where \(G_i\) is the \(i\)th anchor point regression function.

In test time, for given compression patch \(Y^{(i)}\), after obtaining \(Y^{(i)}_d\) and finding its nearest regressor \(G_k\) by \(f^{(i)}_y\), the restoration patch \(\hat{X}^{(i)}\) can be computed by:

\[
\hat{X}^{(i)} = Y^{(i)}_d + G_k f^{(i)}_y \tag{5}
\]

We call the proposed method as subspace joint learning regression, because the anchor point is got by the feature of compression patch \(f_d\) and regressor model is to get mapping function from the denosing patch feature \(f_d\) to original patch \(X\).

2.2 ESVR for Multi-Variable Regression

For each anchor point, we use the basic ESVR method [10] to get texture component regression function. ESVR method is a kind of single hidden layer feed forward network developed from Extreme Learning Machine (ELM) and Support Vector Machine (SVM). Hence ESVR has not only extremely learning speed as ELM, and but also a good generalization performance as SVM. More details of ESVR can be found in Ref. [10].

Here, we extend ESVR to multi-variable regression, as shown in Fig. 1. Overall, there are two parts in multi-variable ESVR. The first part is feature space random transform by the activation function \(\Phi(\Theta, f_d)\) and the second part

![Fig. 1 The Multi-variable ESVR network.](image-url)
can be recast as regularization regression. It should be noted that all predictors share the same feature transform structure, and this mechanism can reduce storage and speed up the algorithm almost without degrading performance.

The input of multi-variable ESVR are \( f_d \) with \( 1 \times m \), and the output are the texture component \( X_{hf} \) with \( 1 \times l \). Let \( A \) is data sample matrix of size \( n \times m \), where each row is one sample \( f_d \) and \( n \) is training sample number in a subspace. The \( \Theta \) is matrix of size \( m \times \tilde{m} \) and can be generated randomly. All the predictors share one transformation matrix \( \Theta \). \( H(\cdot) \) is mapping function with sigmoid function. Hence, \( \Phi(\Theta, f_d) = \text{sigmoid}(\Theta f_d) \): \( R_m \rightarrow R_m \) is feature transform function. Let \( e \) is a vector with \( n \times 1 \), \( W \) is the weight matrix with \( \tilde{m} \times l \), \( b \) is a weight vector with \( 1 \times l \) and \( X_{hf} \) with \( n \times l \) is the corresponding output of \( A \). The model of basic multi-variable ESVR is constructed as

\[
\min_{(w, b) \in \mathbb{R}^{n \times 1} \times \mathbb{R}^{1 \times 1}} = \frac{C}{2} ||e||_2^2 + \frac{1}{2} ||b||_2^2 + \sum_{i=1}^{l} \left( \theta \cdot (X_{hf} - \Phi(\theta, f_d) - b) \right)
\]

s.t. \( \Phi(\Theta, A)w - be - X_{hf} = \xi \) (6)

where \( C \) is the regularization factor. Let \( E = [\Phi(\Theta, A), -e] \in \mathbb{R}^{n \times (\tilde{m}+1)} \). By lagrangian method, when \( n > \tilde{m} + 1 \), it has close solution as

\[
[w, b] = \left( I + E^T E \right)^{-1} E^T X_{hf}
\]

In the test, \( \hat{X}_{hf}^{(i)} \) is the feature mapping \( \Phi(\Theta, f_d^{(i)}) \) feature transform of \( f_d^{(i)} \) for given \( f_d^{(i)} \). In our understanding, the feature mapping \( \Phi(\Theta, f_d) \) is the key idea of ESVR, which can lead to better generalization performances and alleviates the problem of over-fitting.

In practice, we set \( \tilde{m} = 3m \). For each anchor point, \( 2 \times 10^5 \) samples are collected, then \( 1 \times 10^5 \) samples are randomly taken to training the ESVR model. Let \( C = 0.4 \) as optimal parameter.

3. Results

In this letter, we compare with three typical methods: denosing approach BM3D [3], A+ [5] and ARCNN [6]. Two image sets are used, as shown in Fig. 2. One is set 16 in [5], and the other is Set 80, which are 80 images collected from web with medium size (around 500 × 500) for hoping to evaluate method performance in various cases\(^1\). All test images are compressed by JPEG 2000 Kakadu software at 0.1, 0.15, 0.2, 0.25 and 0.3 BPP, in which it is a challenge test environment for all compression restoration methods [5, 6].

Following A+ and ARCNN, only luminance channel (in YCrCb space) is used to restoration. For NLM and BM3D denoising, the initial noise standard variance \( \sigma^0 \) of every compression image is estimated by [11], and \( \sigma = 0.4\sigma^0 \) is used as optimal parameter. We use BSDS500 image set as the training image set for A+, ARCNN and our method. Both A+ and our method are held in 1024 regressors and \( 7 \times 7 \) pixels patch size. To learning based methods, the reconstruction mode is trained in each compression rate.

The diagrams in Fig. 3 show rate-distortion curves including PSNR (dB) and SSIM by different methods. Bjontegaard delta PSNR (BD-PSNR) [12, 13] is general algorithm, which can calculate average PSNR (in dB) difference over the whole range of bitrates. By Bjontegaard metric measurement, the average PSNR gain of four methods are given in Table 1. It can be observed that the proposed method outperform other methods. ARCNN has almost the same performance with A+, while BM3D is worse than A+.

It should be noted that the NLM denosing only take appropriate results in our method, worse than BM3D. However, it can provide suitable primary structure information to further estimate. By using multi-variable ESVR, the proposed method (NLM+ESVR) can get the best results. From Figs. 4–5, it also can be seen that the proposed method produce relatively sharp edges and suppress artifacts well.

There are two computation module in the test processing of proposed method, which are NLM denoising filter and ESVR based subspace joint regression. For a \( 512 \times 512 \) image, the proposed method requires about 40s by using Matlab code on Intel Core i7 3.5GHz processor, in which unoptimized NLM takes 32s and multi-variable ESVR takes about 8s. Contrast to ridge regression in A+, ESVR only add the feature transform processing.

4. Conclusion

In this letter, a novel sub-space joint regression learning based low-rate compression image restoration framework is proposed. It has two part: main structure component estimation by NLM filter and texture component prediction by multi-variable ESVR based regression. Experiment results show the advantage of our method. Furthermore, there are

\[\text{Table 1 Average PSNR gain by BD-PSNR (dB) for different methods}
\]

|          | BM3D [3] | A+ [5] | ARCNN [6] | Proposed |
|----------|----------|--------|------------|----------|
| Set 16   | 0.10     | 0.30   | 0.27       | 0.44     |
| Set 80   | 0.16     | 0.46   | 0.50       | 0.67     |

\(^1\)For more details including the codec, please visit our pages sites.google.com/site/compressionimagerestoration/
two possible approaches to improve the performance of proposed framework. One is that the more effective denoising method may lead to better initialization results. The other is that the hierarchical refinement structure also may get better results.

References

[1] M.A. Robertson and R.L. Stevenson, “DCT quantization noise in compressed images,” IEEE Trans. Circuits Syst. Video Technol., vol.15, no.1, pp.27–38, Jan. 2005.
[2] H. Takeda, S. Farsiu, and P. Milanfar, “Kernel regression for image processing and reconstruction,” IEEE Trans. Image Process., vol.16, no.2, pp.349–366, 2007.
[3] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, “Image denoising by sparse 3-D transform-domain collaborative filtering,” IEEE Trans. Image Process., vol.16, no.8, pp.2080–2095, Aug. 2007.
[4] Y. Kwon, K.I. Kim, J. Tompkin, J.H. Kim, and C. Theobalt, “Efficient learning of image super-resolution and compression artifact removal with semi-local gaussian processes,” IEEE Trans. Pattern Anal. Mach. Intell., vol.37, no.9, pp.1792–1805, 2015.
[5] R. Rothe, R. Timofte, and L. Van, “Efficient regression priors for reducing image compression artifacts,” Image Processing (ICIP), 2015 IEEE International Conference on, pp.1543–1547, IEEE, 2015.
[6] K. Yu, C. Dong, C.C. Loy, and X. Tang, “Deep convolution networks for compression artifacts reduction,” arXiv preprint arXiv:1608.02778, 2016.
[7] Z. Wang, D. Liu, S. Chang, Q. Ling, Y. Yang, and T.S. Huang, “D3: Deep dual-domain based fast restoration of jpeg-compressed images,” Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp.2764–2772, 2016.
[8] R. Timofte, V. De Smet, and L. Van Gool, “A+: Adjusted anchored neighborhood regression for fast super-resolution,” Asian Conference on Computer Vision, vol.9006, pp.111–126, Springer, 2014.
[9] A. Buades, B. Coll, and J.-M. Morel, “A non-local algorithm for image denoising,” Computer Vision and Pattern Recognition, 2005, CVPR 2005, IEEE Computer Society Conference on, pp.60–65, IEEE, 2005.
[10] W. Zhu, J. Miao, and L. Qing, “Robust regression with extreme support vectors,” pattern recognition letters, vol.45, pp.205–210, 2014.
[11] S.-M. Yang and S.-C. Tai, “Fast and reliable image-noise estimation using a hybrid approach,” Journal of Electronic Imaging, vol.19, no.3, pp.033007–033007, 2010.
[12] G. Bjontegard, “Calculation of average PSNR differences between RD-curves,” ITU-T SG16/Q6 VCEG 13th meeting Austin, Texas, USA, April 2001, Doc. VCEG-M33, 2001.
[13] G. Bjontegard, “Improvements of the BD-PSNR model,” ITU-T SG16/Q6 VCEG 35th meeting, Berlin, Germany, July 2008, Doc. VCEG-A111, 2008.