SPATIAL ERROR CONCEALMENT VIA MODEL BASED COUPLED SPARSE REPRESENTATION

Deming Zhai[1], *Xianming Liu*[1*], *Jiantao Zhou*[2], *Debin Zhao*[1], *Wen Gao*[1,3]

[1]School of Computer Science and Technology, Harbin Institute of Technology, Harbin, China
[2]Department of CIS, University of Macau, Macau, China
[3]School of Electrical Engineering and Computer Science, Peking University, Beijing, 100871, China

ABSTRACT

In this paper, we propose a novel spatial error concealment algorithm through model-based coupled sparse representation. According to the non-local self-similarity property of natural images, we first collect two set of samples by template matching: one is called the latent set corresponding to the current missing patch and the other one is called the template set corresponding to the current template. Using these two sets of samples as the training data, we learn a dictionary pair and a linear prediction model simultaneously. The pair of dictionaries aims to characterize the two structural domains of the two sets, and the linear model is to reveal the intrinsic relationship between the sparse representations of the current missing patches and its template. Finally, we cast the non-local dictionary learning and local correlation model into a unified coupled sparse coding framework to obtain optimal sparse representation and further accurate estimation of the current missing patch. Experimental results demonstrate that the proposed method remarkably outperforms previous approaches.

***Index Terms*—** Spatial error concealment, adaptive dictionary learning, linear prediction model, coupled sparse representation

1. INTRODUCTION

Many popular image and video coding standards, such as JPEG and H.264/AVC, are based on partitioning the image or video into blocks. When a coded stream is transmitted over packet-loss networks, errors in the coded stream result in the loss of all the information related to the damaged block. This is especially true in multi-cast and broadcast situations where retransmissions are not permitted or are restricted. The effect of such information loss cannot be endured for the transport of images and videos because any damage to the compressed bit stream may lead to terrible visual distortion at the decoder.

Due to this fact, it is useful to develop error resilience (ER) and error concealment (EC) techniques to control and recover the errors in image and video transmission [1]. Error resilience is usually applied at the encoder side. The coding efficiency of an ER codec is lower than a normal codec, because the encoder needs to introduce some redundancy to the stream. Different from ER coding, the error concealment approach does not require any additional modification on source and channel coding schemes, but introduces post-processing modules at the decoder side to mask the visual effects of transmission errors by utilizing the correctly received information. In this work, we will primarily concern with the recovery of degraded images using spatial EC techniques.

In the literature, many spatial error concealment algorithms have been proposed to exploit image properties to conceal corrupted regions. Most of them are interpolation based, which take advantage of the fact that the image features are locally closely correlated. Kwok and Sun [2] used the multi-directional interpolation to mask corrupted pixel blocks and recover them with a set of available filters. This method can recover the major edges but may create false stripes in smooth image area. Li and Orchard [3] proposed the orientation adaptive interpolation (OAI) method, which provides flexible in-block direction due to its sequential operating manner. However, this method cannot restore textures well. Koloda *et al.* [4] developed a weighted template matching (WTM) algorithm, in which lost regions are recovered sequentially with a weighted combination of templates that are extracted from the available neighborhood. This method can recover large-scale edge and texture well but fails in reconstructing detail regions.

Another major group of spatial error concealment algorithms is to perform estimation in the transform domain. Guleryuz [5] proposed to use the DCT transform to provide sparse decompositions over missing regions; then adaptively determined the magnitude coefficients through thresholding to establish sparsity constraints, and finally estimated missing regions in images using information surrounding these regions. Zhai *et al.* [6] proposed a Bayesian DCT (BayesDCT) pyramid to progressively estimate the missing image block with the nearby correctly decoded pixels from DC to higher AC coefficients for the DCT matrices.

The performance of any EC technology depends on the

*xmliu@jdl.ac.cn

accuracy of the model to characterize the image source in a variety of block-loss situations. In this paper, we try to efficiently combine the properties of local smoothness and non-local self-similarity of natural images, and propose a novel spatial error concealment algorithm through model-based coupled sparse representation. According to the non-local self-similarity property of natural images, we collect two set of samples by template matching, which are corresponding to the current missing patch and its template, respectively. Using these two sets of samples as training data, we learn a dictionary pair and a mapping function simultaneously, and cast the non-local dictionary learning and local correlation model into a unified coupled sparse coding framework to obtain the accurate sparse representation of the current missing patch.

The rest of this paper is organized as follows. Section 2 introduces the sparse model and the proposed adaptive dictionary learning strategy. In Section 3, we detail the proposed model based coupled sparse coding scheme. Experimental results are given in Section 4. Section 5 concludes the paper.

2. SPARSE MODEL WITH ADAPTIVE DICTIONARY

The missing image block $\mathbf{x}_i$ indexed by i is defined as $\mathbf{x}_i := (x_k, x_k \in \Lambda) \in \mathcal{R}^d$ where Λ defines the $\sqrt{d} \times \sqrt{d}$ square neighborhood, and the pixels in $\mathbf{x}_i$ are ordered lexicographically. The problem of recovering the missing block is to find the best estimate of $\mathbf{x}_i$, denoted as $\mathbf{x}_i^*$, given the set of all the available pixels $\mathcal{S}$. A subset of the correctly decoded pixels geometrically close to $\mathbf{x}_i$ are formed into a vector $\mathbf{y}_i \in \mathcal{R}^l$, called the template vector. The key issue of spatial error concealment is how to infer the true value of $\mathbf{x}_i$ according to the information of $\mathbf{y}_i$.

In this paper, following the setting in [4], we define $\mathbf{x}_i$ as a 2×2 patch and let $\mathbf{y}_i$ include all the received or already recovered pixels within the 6×6 pixel neighborhood of $\mathbf{x}_i$, as illustrated in Fig. 1(a). The macroblock is sequentially recovered according to the order shown in Fig. 1(b), where the regions illustrated by brighter level are recovered first.

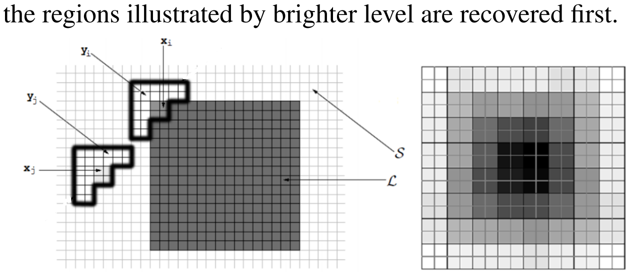

Fig. 1. [4] (a) Configuration for $\mathbf{x}_i$ and $\mathbf{y}_i$. (b) The recovered order, the regions illustrated by brighter level are recovered first.

2.1. Sparse Model

Research on image statistics suggests that image patches can be well-represented as a sparse linear combination of elements from an appropriately chosen dictionary. Inspired by this observation, we seek a sparse representation for each missing patch. Denoting $\mathbf{D} = [\mathbf{d}_1, \mathbf{d}_2, \cdots, \mathbf{d}_p] \in \Re^{d \times p}$ be the dictionary matrix, where each $\mathbf{d}_i$ represents a basis vector in the dictionary. A patch $\mathbf{x}_i$ can be represented as a linear combination of atoms in the dictionary $\mathbf{D}$ plus some perturbation ε, that is,

$$\mathbf{x}_i = \mathbf{D}\mathbf{a}_i + \varepsilon_i, \mathbf{a}_i \in \Re^{p \times 1}. \tag{1}$$

We say that the model is sparse if we can achieve $\|\varepsilon_i\|_2 \ll \|\mathbf{x}_i\|_2$ and $\|\mathbf{a}_i\|_0 \ll k$ simultaneously for all or most $i = 1, \cdots, n$. Seeking the sparsest representation is known to be a NP-hard problem. In order to obtain efficient algorithms, people usually relax the nonconvex ℓ_0 norm to the convex ℓ_1 norm, leading to the following form:

$$\mathbf{a}_i^* = \min_{\mathbf{a}_i} \{\|\mathbf{x}_i - \mathbf{D}\mathbf{a}_i\|^2 + \lambda \|\mathbf{a}_i\|_1\}, \tag{2}$$

where λ is a regularization parameter that controls the tradeoff between the quality of the reconstruction and the sparsity. The above approximation is known as the Lasso. Then, the optimal estimation can be obtained as:

$$\mathbf{x}_i^* = \mathbf{D}\mathbf{a}_i^*. \tag{3}$$

2.2. Adaptive Dictionary Learning

Choosing a good dictionary plays a key role in the above sparse model. The dictionary can either be pre-specified or designed by adapting its content to fit a given set of signal examples. The specified dictionaries, such as wavelet, DCT and curvelets, usually leads to fast algorithms for computing sparse representations. However, due to the fact they are signal-independent or not adaptive, their performance is usually poor. Another approach consists of learning the dictionary on a set of image examples from the offline training set. Unfortunately, in the current image recovery task, there is only one image at our disposal, meaning that we have to adopt another way of estimating an adaptive dictionary.

According to the non-local self-similarity property of natural images, alternatively, we choose to collect some accurate observations of $\mathbf{x}_i$ and $\mathbf{y}_i$ from the received or already recovered pixel set Ω. More formally, we have a set of n posterior samples $\Psi_x = \{\mathbf{x}_{i_1}, \mathbf{x}_{i_2}, \cdots, \mathbf{x}_{i_n}\}$ called the latent set, and $\Psi_y = \{\mathbf{y}_{i_1}, \mathbf{y}_{i_2}, \cdots, \mathbf{y}_{i_n}\}$ called the template set taken in Ω. The collection can be realized through applying a threshold to the ℓ_2 distance between the current template vector $\mathbf{y}_i$ and its observation $\mathbf{y}_j$, that is

$$\begin{aligned} \Psi_x = \{\mathbf{x}_j \,\big|\, \|\mathbf{y}_j - \mathbf{y}_i\|_2^2 / k \le \sigma^2\}; \\ \Psi_y = \{\mathbf{y}_j \,\big|\, \|\mathbf{y}_j - \mathbf{y}_i\|_2^2 / k \le \sigma^2\}, \end{aligned} \tag{4}$$

where σ is the threshold.

Through the above template matching procedure, for each missing patch, we collect samples to constitute two training data set. We stack the vectors of Ψ_x and Ψ_y into two matrixes denoted by $\mathbf{M}_x$ and $\mathbf{M}_y$, respectively. Then, we learn two adaptive dictionaries $\mathbf{D}_x$ and $\mathbf{D}_y$ that are most relevant to Ψ_x and Ψ_y using principle component analysis (PCA) [9].

Although we obtain an adaptive dictionary for the current missing patch, the optimization problem formulated in Eq.(2) still cannot be solved, since $\mathbf{x}_i$ is unknown. Following the similar work in image super-resolution [8], a direct idea is to enforce the similarity of sparse representations between the latent and template sample pair with respect to their own dictionaries. Therefore, the sparse representation of $\mathbf{y}_i$ can be applied with the observation dictionary $\mathbf{D}_x$ to generate a estimation for the current missing patch $\mathbf{x}_i$. However, since the dimensions of $\mathbf{y}_i$ and $\mathbf{x}_i$ are different, such a correspondence of sparse representations does not exist. Therefore, the problem of spatial EC is different from image super-resolution, we have to seek a new solution for the current task.

3. MODEL BASED COUPLED SPARSE REPRESENTATION

3.1. Local Correlation Model

Motivated by the Intra prediction in H.264/AVC, it is reasonable to assume there is strong spatial correlation between the current missing patch and its template. With the collection of observation-template pairs $(\mathbf{x}_i, \mathbf{y}_i), i = 1, \cdots, n$, we try to learn the optimal linear mapping $\mathbf{W}^*$ by addressing the following kernel ridge regression (KRR) problem:

$$\mathbf{W}^* = \min_{\mathbf{W}} \sum_{j=1}^{n} \theta_{ij} \left\| \mathbf{x}_j - \mathbf{W}^T \mathbf{y}_j \right\|_2^2 + \tau \left\| \mathbf{W} \right\|_F^2, \quad (5)$$

where $\|\mathbf{W}\|_F^2$ denotes the Frobenius norm of the matrix $\mathbf{W}$, θ_{ij} reflects the similarity between y_i and y_i, which can be computed as follows:

$$\theta_{ij} = \exp\left\{ -\frac{||\mathbf{y}_i - \mathbf{y}_j||^2}{\sigma_s^2} \right\}, \quad \sigma_s > 0. \quad (6)$$

It is easy to see that the solution of KRR estimator takes the form

$$\mathbf{W}^* = (\mathbf{Y}^T \boldsymbol{\Theta} \mathbf{Y} + \tau \mathbf{I})^{-1} \mathbf{Y} \boldsymbol{\Theta} \mathbf{X}^T, \quad (7)$$

where $\mathbf{Y} = \{\mathbf{y}_j\}_{j=1}^n, \mathbf{X} = \{\mathbf{x}_j\}_{j=1}^n, \boldsymbol{\Theta} = diag(\theta_{i1}, \cdots, \theta_{in})$ is a diagonal matrix and $\mathbf{I}$ is the identity matrix.

With the learned linear mapping, we can derive the linear relationship between sparse representations of $\mathbf{x}_i$ and $\mathbf{y}_i$:

$$\mathbf{D}_x \mathbf{a}_{\mathbf{x}_i} \doteq \mathbf{W}^T \mathbf{D}_y \mathbf{a}_{\mathbf{y}_i} \Rightarrow \mathbf{a}_{\mathbf{x}_i} \doteq \mathbf{D}_x^T \mathbf{W}^T \mathbf{D}_y \mathbf{a}_{\mathbf{y}_i}. \quad (8)$$

We denote $\mathbf{M} = \mathbf{D}_x^T \mathbf{W}^T \mathbf{D}_y$.

3.2. Coupled Sparse Representation

After learning the dictionaries $\mathbf{D}_x$ and $\mathbf{D}_y$ and the linear mapping $\mathbf{M}$, we can derive optimal sparse representation $\mathbf{a}_{x_i}$ of the current missing patch $\mathbf{x}_i$ by addressing the following coupled sparse coding problem:

$$\min_{\{\mathbf{a}_{\mathbf{x}_i}, \mathbf{a}_{\mathbf{y}_i}\}} \left\{ \begin{array}{l} \|\widetilde{\mathbf{x}}_i - \mathbf{D}_x \mathbf{a}_{\mathbf{x}_i}\|_2^2 + \lambda_1 \|\mathbf{a}_{\mathbf{x}_i}\|_F^2 + \|\mathbf{y}_i - \mathbf{D}_y \mathbf{a}_{\mathbf{y}_i}\|_2^2 \\ +\lambda_1 \|\mathbf{a}_{\mathbf{y}_i}\|_F^2 + \gamma \|\mathbf{a}_{\mathbf{x}_i} - \mathbf{M} \mathbf{a}_{\mathbf{y}_i}\|_2^2 \end{array} \right\} \quad (9)$$

where $\widetilde{\mathbf{x}}_i$ is the initial estimation of $\mathbf{x}_i$, which can be obtained as:

$$\widetilde{\mathbf{x}}_i = \mathbf{W}^T \mathbf{y}_i. \quad (10)$$

The above optimization problem is jointly convex in both $\mathbf{a}_{\mathbf{x}_i}$ and $\mathbf{a}_{\mathbf{y}_i}$. Therefore, it is convenient to use an alternating optimization procedure to accurately compute the solution by iteratively optimizing $\mathbf{a}_{\mathbf{x}_i}$ when $\mathbf{a}_{\mathbf{y}_i}$ is fixed, and then optimizing $\mathbf{a}_{\mathbf{y}_i}$ when $\mathbf{a}_{\mathbf{x}_i}$ is fixed. Finally, the current missing patch $\mathbf{x}_i$ can be reconstructed as:

$$\mathbf{x}_i^* = \mathbf{D}_x \mathbf{a}_{\mathbf{x}_i}^*. \quad (11)$$

4. EXPERIMENTAL RESULTS

In this section, experimental results are presented to verify the performance of the proposed algorithm. Given the fact that numerous error concealment algorithms have been developed during the last two decades, it would be virtually impossible for us to perform a thorough comparative study of the proposed error concealment algorithm. Here we choose to compare with some representative work in the literature. More specifically, four approaches are included in our comparative study: (1) the OAI algorithm [3]; (2) the kernel regression (KR) based method [7]; (3) BayesDCT [6]; (4) WTM [5]. We select three popular 512×512 images, *Lena*, *Peppers* and *Boat*, as test ones, and try to recover them from regular block missing with 50% loss rate.

Table 1. Quantitative comparison of five algorithms on recovery of 50% regular block losses

Images	Lena		Peppers		Boat	
	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
OAI	23.71	0.7943	22.09	0.7086	21.49	0.6912
KR	25.45	0.8476	24.53	0.8358	24.21	0.8227
BayesDCT	27.54	0.8117	26.09	0.7882	25.24	0.7674
WTM	27.24	0.8712	26.37	0.8592	24.90	0.8356
Ours	**29.05**	**0.8952**	**27.57**	**0.8761**	**25.86**	**0.8579**

Table 1 illustrates the quantitative comparison on these the test images. It can be observed that the proposed algorithm achieves the highest PSNR and SSIM values for all test images. The PSNR gains of the proposed method over the second best method are 1.51dB, 1.2dB and 0.62dB for *Lena*, *Peppers* and *Boat*, respectively.

We also test the subjective quality comparison. The performance of OAI is too poor, here we omit its results. As illustrated in Fig. 2, our method produces the most visually pleasant results among all comparative studies. It is easy to find that the edge across the region of heavy consecutive

Fig. 2. Objective performance comparison among competing schemes on *Lena* and *Boat* with 50% regular block losses

block loss cannot be well recovered with the KR method. BayesDCT reconstructs edges well but the detail regions are over-smoothed. WTM produces large-scale edges well but it cannot give satisfactory results on local regions that change rapidly, such as the eyes in *Lena*. Even under such a high block loss rate, the proposed algorithm is still capable of restoring major edges and repetitive textures of the images. It is noticed that the proposed method can more accurately recover global object contours with severe losses, such as the edges along the mast in *Boat*. These results demonstrate the power of the proposed spatial error concealment algorithm. The strength of the proposed approach comes from its full utilization of constraints of the non-local similarity by adaptive dictionary and local smoothness by local correlation model.

5. CONCLUSION

In this paper, we presented an efficient sparse coding based spatial error concealment algorithm by combining adaptive dictionary learning and linear prediction model. The procedure of dictionary learning is actually to exploit the non-local self-similarity of natural images, while the linear prediction model is to explore the local smoothness among adjacent samples. Therefore, the nature of our method is to propose a coupled sparse coding framework to consider both local and global information contained in image samples. Experimental results demonstrate the proposed method outperforms state-of-the-art spatial error concealment algorithms.

6. ACKNOWLEDGEMENT

This work is supported in part by the Major State Basic Research Development Program of China (973 Program) under Grant 2009CB320905 and by the National Science Foundations of China under Grant 61272386, the Start-up Research Grant (SRG) SRG023-FST13-ZJT of the University of Macau, and the new Ph.D researcher award of Chinese Ministry of Education.

7. REFERENCES

[1] Y. Wang and Q. Zhu, “Error control and concealment for video communication: A review,” Proc. IEEE, vol. 86, no. 5, pp. 974-997, May 1998.

[2] W. Kwok and H. Sun, “Multi-directional interpolation for spatial error concealment,” IEEE Trans. Consumer Electron., vol. 39, no. 3, pp. 455-460, Jun. 1993.

[3] X. Li, M. T. Orchard, “Novel sequential error-concealment techniques using orientation adaptive interpolation,” IEEE Trans. Circuits Syst. Video Techn. 12(10): 857-864 (2002)

[4] J. Koloda, *et al.*, “Sequential Error Concealment for Video/Images by Weighted Template Matching,” in Proceeding of *DCC*, 2012, pp. 159-170.

[5] O. G. Guleryuz, “Nonlinear approximation based image recovery using adaptive sparse reconstructions and iterated denoising-part II:Adaptive algorithms,” IEEE Trans. Image Process., 15(3): 555-571 (2006)

[6] G. Zhai, X. Yang, W. Lin, and W. Zhang, “Bayesian error concealment with DCT pyramid for images,” IEEE Trans. Circuits Syst. Video Technol., vol. 20, pp. 1224-1232, September 2010.

[7] H. Takeda, S. Farsiu, and P. Milanfar, “Kernel Regression for Image Processing and Reconstruction,” IEEE Trans. Image Process., 16(2):349-366 (2007).

[8] J. Yang, J. Wright, T. Huang, and Y. Ma, “Image super-resolution via sparse representation,” IEEE. Trans. Image Process., vol. 19, no. 11, pp. 2861-2873, Nov. 2010.

[9] W. Dong, L. Zhang, G. Shi, and X. Wu, “Image deblurring and super-resolution by adaptive sparse domain selection and adaptive regularization,” IEEE Trans. On Image Process., vol. 20, no. 7, pp. 1838-1857, July 2011.