2012 Data Compression Conference

Multi-scale Spatial Error Concealment via Hybrid Bayesian Regression

Xianming Liu[1], Deming Zhai[1], Guangtao Zhai[2], Debin Zhao[1], Ruiqin Xiong[2], Wen Gao[1,2]
[1]School of Computer Science and Technology, Harbin Institute of Technology, Harbin, China
[2]School of Electronic Engineering & Computer Science, Peking University, Beijing, China

Abstract: In this paper, we propose a novel multi-scale spatial error concealment algorithm to combine the modeling strengthes of the parametric and nonparametric Bayesian regression. We progressively recover missing blocks in the scale space from coarse to fine so that the sharp edges and texture in the finest scale can be eventually recovered. On one hand, in each scale, the nonparametric part of the methodology is used to exploit the intra-scale correlation, which relies on the data itself to dictate the structure of the model. In this procedure, the non-local self-similarity property is utilized as a fruitful resource for abstracting a priori knowledge of images. On the other hand, the parametric part is used to explicitly model the inter-scale correlation, in which the local structure regularity is thoroughly explored to recover the sharp edges and major texture features of images. It is not respected if only the nonparametric modeling is considering. We achieve the best of both worlds within a multi-scale framework. Experimental results on benchmark test images demonstrate that the proposed method achieves very competitive performance with the state-of-the-art error concealment algorithms.

I. Introduction

Transmitting image and video over the wireless network is a challenging problem since image and video may be seriously distorted due to fluctuating channel conditions, which appear in the form of blocks loss. This is especially true in multicast and broadcast situations where retransmissions are not permitted or are restricted. Due to this facts, it is useful to develop error resilience (ER) and error concealment (EC) techniques to control and recover from the errors in image and video transmission. 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. Alternatively, error concealment is a post processing technique, which conceals the errors utilizing the correctly received information at the decoder side without modifying source and channel coding schemes.

According to the information utilized, error concealment algorithms can be categorized into temporal approaches [1][2] and spatial approaches [3][4]. Temporal approaches restore the corrupted blocks by exploiting temporal correlation between successive frames. The spatial error concealment algorithms are normally applied in situations where information from neighboring frames is not available (images or intra coded frames), hence having to rely entirely on the information from intact neighboring macroblocks. In many situations, temporal error concealment is preferred to spatial error concealment since the temporal correlation between neighboring frames can be used to aid the blocks recovery.

1068-0314/12 $26.00 © 2012 IEEE
DOI 10.1109/DCC.2012.25

However, spatial error concealment is usually applied to conceal lost areas in intra-frames of hybrid video codecs like H.264/AVC, for which temporal prediction is not possible or prudent. Error concealment of these frames is very important in order to prevent error propagation as well as to have a good playback quality and user satisfaction. In this work, we will primarily be concerned with the recovery of missing blocks using spatial prediction alone.

From a statistical perspective, spatial error concealment is essentially ill-posed, and often can be formulated as an energy minimization or statistical inference problem in which either the optimal or most probable configuration is the goal. The key to this task is the prior knowledge about the properties that the target image should have. One common assumption for natural images is *intensity consistency*, which means: (1) nearby pixels are likely to have the same or similar intensity values; and (2) pixels on the same structure are likely to have the same or similar intensity values. Note that the first assumption means images are locally smooth, and the second assumption means images are with non-local self-similarity. Accordingly, how to choose statistical models that thoroughly explore such two prior knowledge directly determines the performance of the proposed spatial error concealment algorithms.

One approach to learn statistical regularities is to formulate a probabilistic model of how the data are generated and adapt its parameters to fit the observed distribution. Early heuristic observation about the local smoothness of image intensity field has been quantified by several linear parametric models, such as the piecewise autoregressive (PAR) image model. Moreover, the study of natural image statistics reveals that the second order statistics of natural images tends to be invariant across different scales, and those scale invariant features are shown to be crucial for visual perception [5][6]. This observation inspires us to learn and propagate the statistical features across different scales via parametric regression. On the other hand, the nonparametric kernel regression approach was found to be able to exploit the non-local self-similarity property of images, which in fact reflects the intra-scale correlation. Many follow-up studies have shown the potential of nonlocal sparse representations in other image processing tasks [7][8]. All those findings have strongly suggested intra-scale and inter-scale correlation are fruitful resources for abstracting reliable priori knowledge of photographic images [9].

In this paper, we propose a novel multi-scale spatial error concealment algorithm to combine the modeling strengthes of the parametric and nonparametric models. We progressively recover missing blocks in the scale space from coarse to fine so that the sharp edges and texture in the finest scale can be eventually recovered. On one hand, in each scale, the nonparametric part of the methodology is used to exploit the intra-scale correlation, which relies on the data itself to dictate the structure of the model. In this procedure, the non-local self-similarity property is utilized as a fruitful resource for abstracting a priori knowledge of images. On the other hand, the parametric part is used to explicitly model the inter-scale correlation, in which the local structure regularity is thoroughly explored to preserve the sharp edges and major texture features of images. It is not respected if only the nonparametric modeling is considering. We achieve the best of both worlds within a multi-scale framework. By combining intra-scale and inter-scale correlation, we show that the proposed method can capture the local variation and the global trend of image texture in corrupted regions. Experimental results on benchmark test images demonstrate that the proposed method achieves more wonderful performance compared with the state-of-the-art error concealment algorithms.

The rest of this paper is organized as follows. Section II presents the framework of proposed spatial error concealment method. Section III details the used nonparametric and parametric regression technology in our method. Section IV presents some experimental results and comparative studies. And Section V concludes the paper.

II. The Framework of Proposed Scheme

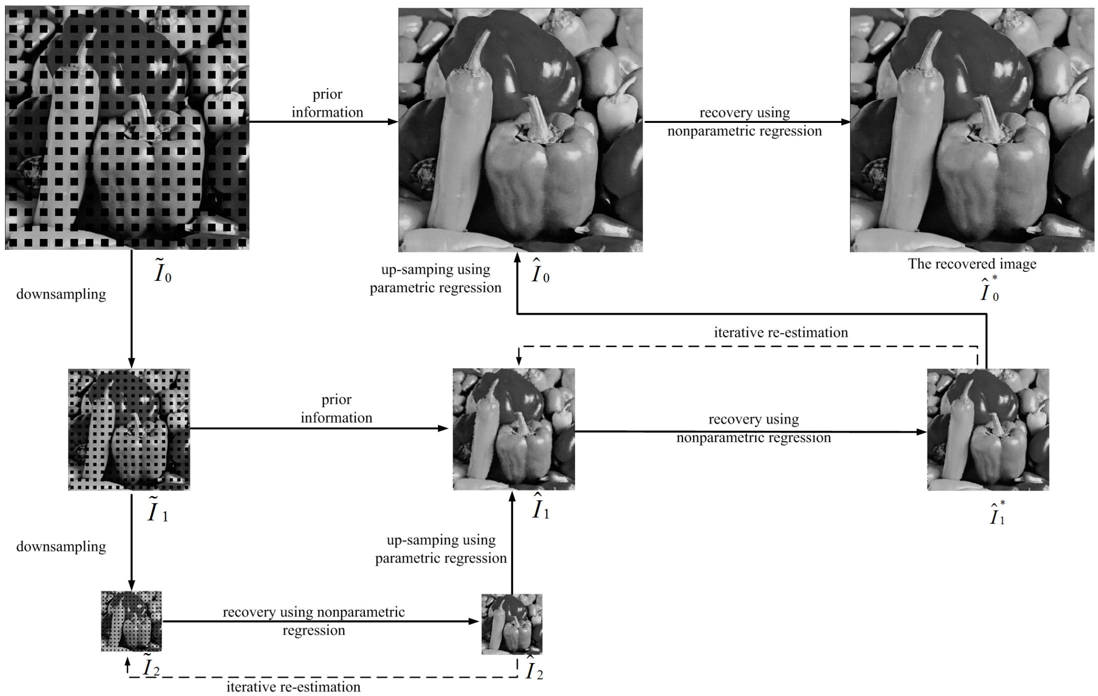

Fig. 1. Diagram of the proposed multi-scale spatial error concealment algorithm with hybrid Bayesian regression.

In this section, we present an overview of the proposed error concealment framework. We build a multi-scale pyramid of the full-resolution corrupted image. Our goal is to progressively recover missing blocks from coarse to fine scales (inter-scale), and progressively restore image details within every scale (intra-scale). To give a clear illustration, we summary the proposed multi-scale progressive error concealment algorithm in Fig.1, where 22% macroblocks in the test image *Lena* are regularly missing.

First, the error image is directly downsampled 2 times (from scale 0 to scale 2, denoted by $\widetilde{I}_0$, $\widetilde{I}_1$, $\widetilde{I}_2$) by averaging the existing pixels in a 2×2 neighborhood on higher resolution. At the coarsest scale 2, the missing blocks in $\widetilde{I}_2$ can be recovered via the nonparametric regression technique which will be detailed in Section III-A to get $\widehat{I}_2$. And this nonparametric estimation can be computed iteratively by feeding the processing results $\widehat{I}_2$ to the nonparametric estimator as a prior for computing the kernel distance. In practice, two iterations was found to be effective in improving the processing results in such type of operations.

The recovered image $\widehat{I}_2$ is then upsampled with the proposed parametric regression method, which will be introduced in detail in Section III-B, to get $\widehat{I}_1$ using the following steps: firstly, we collect basic function vectors $[\Phi(\mathbf{x}_1),\Phi(\mathbf{x}_2),\cdots,\Phi(\mathbf{x}_n)]$ for low-resolution pixels and arrange them into the matrix $\mathbf{\Phi}$, where Φ represents the operator to pick up the four 8-connected neighbors of the pixel from the context $\mathbf{x}_i$. And the observations

$[y_1, y_2, \cdots, y_n]$ are collected to construct the observation vector $\mathbf{y}$. Secondly, the kernel weights are computed for each available low-resolution pixels. Finally, the linear model parameters for lower resolution image $\widehat{I}_2$ are computed from the kernel ridge regression. With the derived model parameters, we construct the higher resolution image $\widehat{I}_1$.

The upsampled image $\widehat{I}_1$ can be used as a prior for calculating the weights in the nonparametric model towards a refined estimate $\widehat{I}_1^*$. The refined estimate can then be upconverted to $\widehat{I}_0$ by the parametric model with the steps listed above. And the refined estimate $\widehat{I}_0$ can be combined with $\widetilde{I}_0$ into another nonparametric recovery procedure towards the final results $\widehat{I}_0^*$. Using the above progressive recovery based on intra-scale and inter-scale correlation, we gradually recover an image with few artifacts.

III. Nonparametric and Parametric Bayesian Regression

From the overall description of the proposed framework in Section II, we can find nonparametric and parametric regression technologies play the central role in our method. Both parts can be casted into a probability framework, generally known as empirical Bayesian regression. In this section, a detailed description about the nonparametric and parametric empirical Bayesian inference technologies used in our method is presented.

A. Nonparametric Bayesian Regression

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 $\widehat{\mathbf{x}}(i)$, given the set of all the available pixels Ω. A subset of the correctly decoded pixels geometrically around $\mathbf{x}(i)$ 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 $\mathbf{x}$ according the information of $\mathbf{y}$.

To compute the optimal Bayesian estimator for the vector $\mathbf{x}(i)$, it is necessary to define an appropriate loss function $L\left(\mathbf{x}(i), \widehat{\mathbf{x}}(i)\right)$ which measures the estimation loss associated with choosing an estimator $\widehat{\mathbf{x}}(i)$. The optimal estimator $\widehat{\mathbf{x}}^*(i)$ is found by minimizing the posterior expected loss:

$$\mathrm{E}\left[L\left(\mathbf{x}(i), \widehat{\mathbf{x}}(i)\right)\right] = \sum_{\mathbf{x}(i)} L\left(\mathbf{x}(i), \widehat{\mathbf{x}}(i)\right) P\left(\mathbf{x}(i)|\mathbf{y}(i)\right). \tag{1}$$

We define the loss function $L\left(\mathbf{x}(i), \widehat{\mathbf{x}}(i)\right)$ is in a quadratic loss form, and the optimal Bayesian estimator can then be formulated as:

$$\widehat{\mathbf{x}}^*(i) = \arg\min_{\widehat{\mathbf{x}}(i)} \sum_{\mathbf{x}(i)} \left\|\mathbf{x}(i) - \widehat{\mathbf{x}}(i)\right\|^2 P\left(\mathbf{x}(i)|\mathbf{y}(i)\right). \tag{2}$$

Taking the derivative with respect to $\widehat{\mathbf{x}}(i)$ and set it to zero, we can get the well known Bayes least squares (BLS) estimator:

$$\widehat{\mathbf{x}}^*(i) = E_{X|Y}(\mathbf{x}|\mathbf{y}) = \sum_{\mathbf{x}(i)\in\Omega} \mathbf{x}(i) P(\mathbf{x}(i)|\mathbf{y}(i)), \tag{3}$$

which can be further written as:

$$\widehat{\mathbf{x}}^*(i) = \sum_{\mathbf{x}(i)\in\Omega} \mathbf{x}(i) \frac{P(\mathbf{x}(i), \mathbf{y}(i))}{P(\mathbf{y}(i))} = \frac{\sum\limits_{\mathbf{x}(i)\in\Omega} \mathbf{x}(i) P\left(\mathbf{y}(i)|\mathbf{x}(i)\right) P\left(\mathbf{x}(i)\right)}{\sum\limits_{\mathbf{x}(i)\in\Omega} P\left(\mathbf{y}(i)|\mathbf{x}(i)\right) P\left(\mathbf{x}(i)\right)}. \tag{4}$$

As indicated in Eq. (4), to compute $\widehat{\mathbf{x}}^*(i)$, we would like to know $P(\mathbf{y}(i)|\mathbf{x}(i))$ and $P(\mathbf{x}(i))$ for each block i. Ideally, $P(\mathbf{y}(i)|\mathbf{x}(i))$ and $P(\mathbf{x}(i))$ are obtained from a large number of repeated observations for each patch $\mathbf{x}(i)$. Unfortunately, we only have one image at our disposal, meaning that we have to adopt another way of estimating these *pdf*s. According to the self-similarity property of natural images, alternatively, we choose to collect some accurate observations of $\mathbf{x}(i)$ and $\mathbf{y}(i)$ from the correctly decoded pixel set Ω. More formally, we have a set of n posterior samples $\Psi_x = \{\mathbf{x}(i_1), \mathbf{x}(i_2), \cdots, \mathbf{x}(i_N)\}$ and $\Psi_y = \{\mathbf{y}(i_1), \mathbf{y}(i_2), \cdots, \mathbf{y}(i_N)\}$ taken in Ω.

First, let us consider the case of the coarsest scale $\widetilde{I}_2$, where we only have a corrupted image as input. In this case, The collection can be realized through applying a threshold to the ℓ_2 distance between the center template vector $\mathbf{y}(i)$ and its observation $\mathbf{y}(j)$, that is

$$\begin{aligned} \Psi_x &= \{\mathbf{x}(j) \Big| ||\mathbf{m}(j)^T \cdot \mathbf{y}(j) - \mathbf{m}(i)^T \cdot \mathbf{y}(i)||_2^2 \leq \sigma^2\}; \\ \Psi_y &= \{\mathbf{y}(j) \Big| ||\mathbf{m}(j)^T \cdot \mathbf{y}(j) - \mathbf{m}(i)^T \cdot \mathbf{y}(i)||_2^2 \leq \sigma^2\}, \end{aligned} \tag{5}$$

where σ is the threshold, for which we set a large value (e.g., 100) to incorporate enough data samples in the observation set; $\mathbf{m}(i)$ and $\mathbf{m}(j)$ are masking vectors with 1s for available pixels and 0s for missing pixels. In this way, we allow part of the sampling vectors $\mathbf{x}(j)$ and $\mathbf{y}(j)$ to be missing pixels as long as the effective members of $\mathbf{y}(j)$ and $\mathbf{y}(i)$ are close enough to each other on average. This is very useful when block loss rates are high, such as consecutive block loss. Indeed, this is a procedure of template matching.

From this posterior sample set Ψ_x, we start by examining the prior distribution $P(\mathbf{x}(i))$. It is an immediate choice of using Markov random field (MRF), field of expert [10] (FoE) or other types of image models to capture interactions between pixels in the image patch, but the MRF framework involves the computationally intensive estimation of additional hyperparameters which must be likely adapted to each spatial position. For simplicity in both analysis and computation, we just use a uniform prior in this paper:

$$P(\mathbf{x}(i)) = \frac{1}{|\Psi_x|} = \frac{1}{N}. \tag{6}$$

This means there is no preference to choose a vector $\mathbf{x}(i)$ in the set Ψ_x which is assumed to be composed of N preliminarily selected similar patches. Therefore, we can propose a reasonable estimator for $\widehat{\mathbf{x}}(i)$:

$$\widehat{\mathbf{x}}^*(i) = \frac{\sum\limits_{\mathbf{x}(i)\in\Omega} \mathbf{x}(i) P(\mathbf{y}(i)|\mathbf{x}(i))}{\sum\limits_{\mathbf{x}(i)\in\Omega} P(\mathbf{y}(i)|\mathbf{x}(i))}. \tag{7}$$

However, the true set $\{\mathbf{x}(i)\}$ is not available but only a spatially varying dictionary $\Psi_x = \{\mathbf{x}(i_1), \mathbf{x}(i_2), \cdots, \mathbf{x}(i_N)\}$. Further changing Ω to Ψ_x we have the following reasonable estimation for $\widehat{\mathbf{x}}(i)$:

$$\widehat{\mathbf{x}}^*(i) = \frac{\sum\limits_{j=i_1}^{i_N} \mathbf{x}(j) P(\mathbf{y}(i)|\mathbf{x}(j))}{\sum\limits_{j=i_1}^{i_N} P(\mathbf{y}(i)|\mathbf{x}(j))}. \tag{8}$$

In the above equation, $P(\mathbf{y}(i)|\mathbf{x}(j))$ measures the probability of the current template sample set $\mathbf{y}(i)$ given a sampled image block vector $\mathbf{x}(j) \in \Psi_x$. Since $\mathbf{y}(i)$ and $\mathbf{x}(j)$ are with different shape and size, this probability is difficult to calculate directly. Alternatively, this probability can be approximated as $P(\mathbf{y}(i)|\mathbf{y}(j))$, $\mathbf{y}(j) \in \Psi_y$. Therefore, the above equation can be rewritten as

$$\widehat{\mathbf{x}}^*(i) = \frac{\sum_{j=i_1}^{i_N} \mathbf{x}(j) P(\mathbf{y}(i)|\mathbf{y}(j))}{\sum_{j=i_1}^{i_N} P(\mathbf{y}(i)|\mathbf{y}(j))}. \quad (9)$$

The accuracy of the final estimation mainly depends on $P(\mathbf{y}(i)|\mathbf{y}(j))$, which must be carefully examined. We propose the following definition for the likelihood:

$$P(\mathbf{y}(i)|\mathbf{y}(j)) = \exp\left\{\frac{-(||\mathbf{m}(i)^T \cdot \mathbf{y}(i) - \mathbf{m}(j)^T \cdot \mathbf{y}(j)||^2)}{\sigma_f^2}\right\} \cdot \exp\left\{\frac{-(||\mathbf{x}_i - \mathbf{x}_j||^2)}{\sigma_s^2}\right\}, \quad (10)$$

where $\mathbf{x}_i \in \mathcal{R}^2$ and $\mathbf{x}_j \in \mathcal{R}^2$ denote the center coordinates of the block i and j.

B. Parametric Bayesian Regression

With the result of nonparametric regression, we get a initial estimation of each scale. Then, we try to progressively recover missing blocks in the scale space from coarse to fine so that the sharp edges and texture in the finest scale can be eventually recovered.

In our scheme, we utilize the 2D piecewise autoregressive (2D-PAR) model to represent the inter-scale relationship. The PAR model is chosen for the following reasons. Firstly, the PAR model class is capable of fitting image waveforms ranging from smooth shades, periodic textures to transients like edges [11][12]. Secondly, the piecewise linearity of the PAR model makes the estimation problem convex, and hence, computationally amenable.

Let us assume that the data set from the previous scale to be interpolated is a set of pairs $X = \{(\mathbf{x}_1, y_1), \cdots, (\mathbf{x}_n, y_n)\}$, where $\mathbf{x}_i \in \mathcal{R}^2$ denotes the coordinates, $y_i \in \mathcal{R}$ is the corresponding intensity output, and $i = \{1, \cdots, n\}$ is a index running over the pairs. What we want to do is to estimate a model to approximate these sample points, and use this model to upsample the image of previous scale. In this way, statistical features are learned and propagated from the previous scale to the current scale. To define a linear parametric model, a set of k fixed basis function $\Phi = \{\phi_j(\mathbf{x}_i)\}$ is chosen, then the interpolated function $f(\mathbf{x}_i)$ is assumed to have the form:

$$f(\mathbf{x}_i) = \sum_{j=1}^{k} w_j \phi_j(\mathbf{x}_i) = \mathbf{w}^T \Phi(\mathbf{x}_i), \quad (11)$$

where $\mathbf{w}$ is the k-dimensional *model parameter vector*, which is estimated on the fly for each pixel using sample statistics of a local context. And the data set is modeled as deviating from this mapping under some additive noise process:

$$y_i = \mathbf{w}^T \Phi(\mathbf{x}_i) + v_i, \quad (12)$$

where v_i is the fitting error, which can be modeled as zero-mean Gaussian noise with standard deviation σ_v. Then the probability of the data given the parameters $\mathbf{w}$ is:

$$P(X|\mathbf{w}, \beta, \Phi) = \frac{\exp[-\beta E_X(X|\mathbf{w}, \Phi)]}{Z_X(\beta)}, \quad (13)$$

where $\beta = 1/\sigma_v^2$, $Z_X = (2\pi/\beta)^{k/2}$, and E_X is the data fitting term. $P(X|\mathbf{w}, \beta, \Phi)$ is called the likelihood. It is well known that finding the maximum likelihood parameters $\mathbf{w}_{ML}$ may be an ill-posed problem. Therefore, we need a prior R to express the sort of smoothness which we expect the interpolated function $f(\mathbf{x}_i)$ to have. We define the prior for the linear model with the following form:

$$P(f(\mathbf{x})|R, \alpha) = \frac{\exp[-\alpha E_f(f(\mathbf{x})|R)]}{Z_f(\alpha)}. \tag{14}$$

where E_f is a regularization function, $Z_f = \int \exp(-\alpha E_f) df$, and α is the regularization parameter to measure how smooth $f(\mathbf{x})$ is expected to be. Such a prior can further be written as a prior on the parameters $\mathbf{w}$:

$$P(\mathbf{w}|\Phi, R, \alpha) = \frac{\exp[-\alpha E_{\mathbf{w}}(\mathbf{w}|\Phi, R)]}{Z_{\mathbf{w}}(\alpha)}, \tag{15}$$

where $E_{\mathbf{w}}$ is commonly referred to as a regularizing function, and $Z_{\mathbf{w}} = \int \exp(-\alpha E_{\mathbf{w}}) d\mathbf{w}$.

As a result, the posterior probability of the parameters $\mathbf{w}$ could be expressed as:

$$P(\mathbf{w}|X, \alpha, \beta, \Phi, R) = \frac{P(X|\mathbf{w}, \beta, \Phi) P(\mathbf{w}|\Phi, R, \alpha)}{P(X|\alpha, \beta, \Phi, R)}. \tag{16}$$

Writing $M(\mathbf{w}) = \beta E_X + \alpha E_{\mathbf{w}}$, which can be further written as

$$M(\mathbf{w}) = E_X + \lambda E_{\mathbf{w}}, \lambda = \alpha/\beta, \tag{17}$$

the posterior becomes

$$P(\mathbf{w}|X, \lambda, \Phi, R) = \frac{\exp[-M(\mathbf{w})]}{Z_M(\lambda)}, \tag{18}$$

where $Z_M(\lambda) = \int \exp(-M) d\mathbf{w}$.

As stated above, the reconstruction process is converted to how to efficiently estimate the model parameters $\mathbf{w}$. To this end, there are some parts needed to complete, consisting of a choice of basis function Φ, a data fitting term E_X, a prior (regularizer) term $E_{\mathbf{w}}$, and the regularization parameter λ. Now let us consider how to define these elements:

- $\Phi(\mathbf{x}_i)$ is set as the intensity vector of the k-nearest neighbors of y_i in the context $\mathbf{x}_i$ from all surrounding directions. In this way, the uncertainty of y_i can be efficiently resolved by estimating along the edge orientation based on geometric constraint of edges.
- A reasonable assumption made with the natural image source is that it can be modeled as a locally stationary Gaussian process, *i.e.*, the neighboring samples maybe have the same or similar transformation function. Following this assumption, the optimal model parameter vector $\mathbf{w}$ is found by projecting the model onto the neighboring measured examples. Accordingly, we define E_X as the following moving least square (MLS) error form:

$$E_X = \sum_{j=1}^{n} \theta(\mathbf{x}_i, \mathbf{x}_j) \left\| y_j - \mathbf{w}^T \Phi(\mathbf{x}_j) \right\|^2, \tag{19}$$

where $\theta(\mathbf{x}_i, \mathbf{x}_j)$ is the similarity weight of the sample $\mathbf{x}_j$ for the current interpolation point $\mathbf{x}_i$. The kernel weight is driven by the distance

$$\theta(\mathbf{x}_i, \mathbf{x}_j) = \left\| \mathbf{m}^T \cdot (\mathbf{x}(i) - \mathbf{x}(j)) \right\|^2, \tag{20}$$

where $\mathbf{x}(i)$ denotes a local block centered on $\mathbf{x}_i$, $\mathbf{m}$ is a masking vector indicating existing pixels in $\mathbf{x}(i)$ and $\mathbf{x}(j)$. The moving least square can efficiently cancel the influence of statistical outlier and lead to a robust estimator of local image structure.

- To design more stable empirical minimization, we define the regularization term $E_{\mathbf{w}}$ as an additional estimator complexity penalty on E_X.

Finally, maximum the posterior probability is equal to minimize the combined objective function M:

$$M(\mathbf{w}) = \sum_{j=1}^{n} \theta(\mathbf{x}_i, \mathbf{x}_j) \left\| y_i - \mathbf{w}^T \Phi(\mathbf{x}_j) \right\|^2 + \lambda \left\| \mathbf{w} \right\|^2 . \tag{21}$$

This is also known as kernel ridge regression (KRR). The penalty term is stable because it does not depend on data. By choosing λ properly, an appropriate amount of ridge's stability translates into good statistical properties of the KRR estimator. It is easy to see that the solution of KRR estimator takes the form

$$\mathbf{w}^* = (\mathbf{\Phi}^T \mathbf{\Theta} \mathbf{\Phi} + \lambda \mathbf{I})^{-1} \mathbf{\Phi}^T \mathbf{\Theta} \mathbf{y}, \tag{22}$$

where $\mathbf{y} = (y_1, \cdots, y_n)^T$, $\mathbf{\Phi} = (\phi(\mathbf{x}_1), \cdots, \phi(\mathbf{x}_n))^T \in \Re^{n \times k}$, and $\mathbf{I}$ is the identity matrix, $\mathbf{\Theta} = diag(\theta(\mathbf{x}_i, \mathbf{x}_1), \cdots, \theta(\mathbf{x}_i, \mathbf{x}_n)) \in \Re^{n \times n}$ is a matrix with similarity weights on the diagonal line. For the regularization parameter λ, we generalize the previous work on spectral regression discriminant analysis [13] to the kernel version, and can derive a data-adaptively optimal solution.

$$\lambda^* = \frac{1}{L} Tr(\mathbf{\Phi}^T \mathbf{\Theta} \mathbf{\Phi}), \tag{23}$$

where L is a normalized parameter to keep λ in $(0, 1)$, which is set as 10^6 in the practice implementation; Tr is the operator of the trace.

IV. Experimental Results

In this section, experimental results are presented to verify the performance of the proposed spatial error concealment 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. Due to space limitation, here we choose to compare with two state-of-the-art methods in the literature: (1) the nonlinear approximation based image recovery technology using adaptive sparse reconstructions (SparseRec) proposed by Guleryuz [3]; (2) the kernel regression (KR) based methods proposed by Takeda *et al.* [4]. We select three popular 512×512 images, *Lena*, *Peppers* and *Boat*, as test ones.

Let us examine the performance of compared methods on random block losses, where the test images are subjected to 50% 8×8 blocks loss. At such a high block loss rate, the error concealment task is more difficult since many missing blocks cluster together. Table I 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 gain compared with SparseRec is up to 1.31dB. Table I also gives processing times results on a typical computer (2.5GHz Intel Dual Core, 4G Memory) of compared algorithms. KR and the proposed method are run on Matlab, while SparseRec are run on C since there is no Matlab source code available for it. Even running on C, the SparseRec method needs about fifty minutes to process an image. This is due to the fact that the SparseRec method is related to a nonconvex optimization problem. It has to rely on a iterative procedure to achieve convergence results, which is computationally expensive. For the proposed method, we only need about 70 seconds to recover one image.

TABLE I
QUANTITATIVE COMPARISON OF THREE ALGORITHMS ON RECOVERY OF 50% RANDOM BLOCK LOSSES

Images	Lena			Peppers			Boat		
	PSNR(dB)	SSIM	TIME(s)	PSNR(dB)	SSIM	TIME(s)	PSNR(dB)	SSIM	TIME(s)
KR	26.81	0.8676	108.47	26.11	0.8559	96	24.26	0.8281	108.19
SparseRec	26.40	0.8702	3037.89	26.68	0.8667	3037.89	24.85	0.8408	3653.63
Ours	**27.71**	**0.8782**	**72.13**	**26.92**	**0.8711**	**71.95**	**25.09**	**0.8443**	**74.47**

We also test the subjective quality comparison. As illustrated in Fig. 2, our method produces the most visually pleasant results among all comparative studies. Even under such 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 edge along the shoulder in *Lena*, the edge along the pepper in *Peppers*, and the edges along the mast in *Boat*. It is easy to find that the edge across the region of heavy consecutive block loss cannot be well recovered with the KR and SparseRec method. This further demonstrates the power of the proposed multi-scale progressive error concealment algorithm. The strength of the proposed approach comes from its full utilization of the intra-scale and inter-scale correlations, which are neglected by the currently available single-scale methods. And many large-scale structures can be well recovered based upon the progressively computed low-level results, which is impossible for traditional single level error concealment algorithms.

V. CONCLUSION

In this paper, we present a novel spatial error concealment algorithm by combining the modeling strengths of the parametric and nonparametric regression. The framework we use in the proposed method is a multi-scale pyramid, where the intra-scale relationship can be modeled with the nonparametric approach while the inter-scale dependency can be learned and propagated with the parametric model. In this way, both local and nonlocal regularity constrains are exploited to capture the local variation and the global trend of image texture in corrupted regions. Experimental results over a wide range of test images demonstrate that our method achieves very competitive recovery performance compared the state-of-the-art methods in both objective and subjective visual quality.

VI. ACKNOWLEDGEMENT

This work was supported in part by the National Science Foundations of China (60736043, 61073083, 61001145), in part by the Major State Basic Research Development Program of China (973 Program 2009CB320905).

REFERENCES

[1] Y. Zhang, X. Xiang, S. Ma, D. Zhao, W. Gao, “ Auto Regressive Model and Weighted Least Squares Based Packet Video Error Concealment,” IEEE Data Compression Conference 2010, pp. 455-464, Snowbird, Utah, USA.

[2] Mengyao Ma, Oscar C. Au, Shueng-Han Gary Chan, Ming-Ting Sun, “ Edge-Directed Error Concealment,” IEEE Trans. Circuits Syst. Video Techn. 20(3): 382-395 (2010).

[3] 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)

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

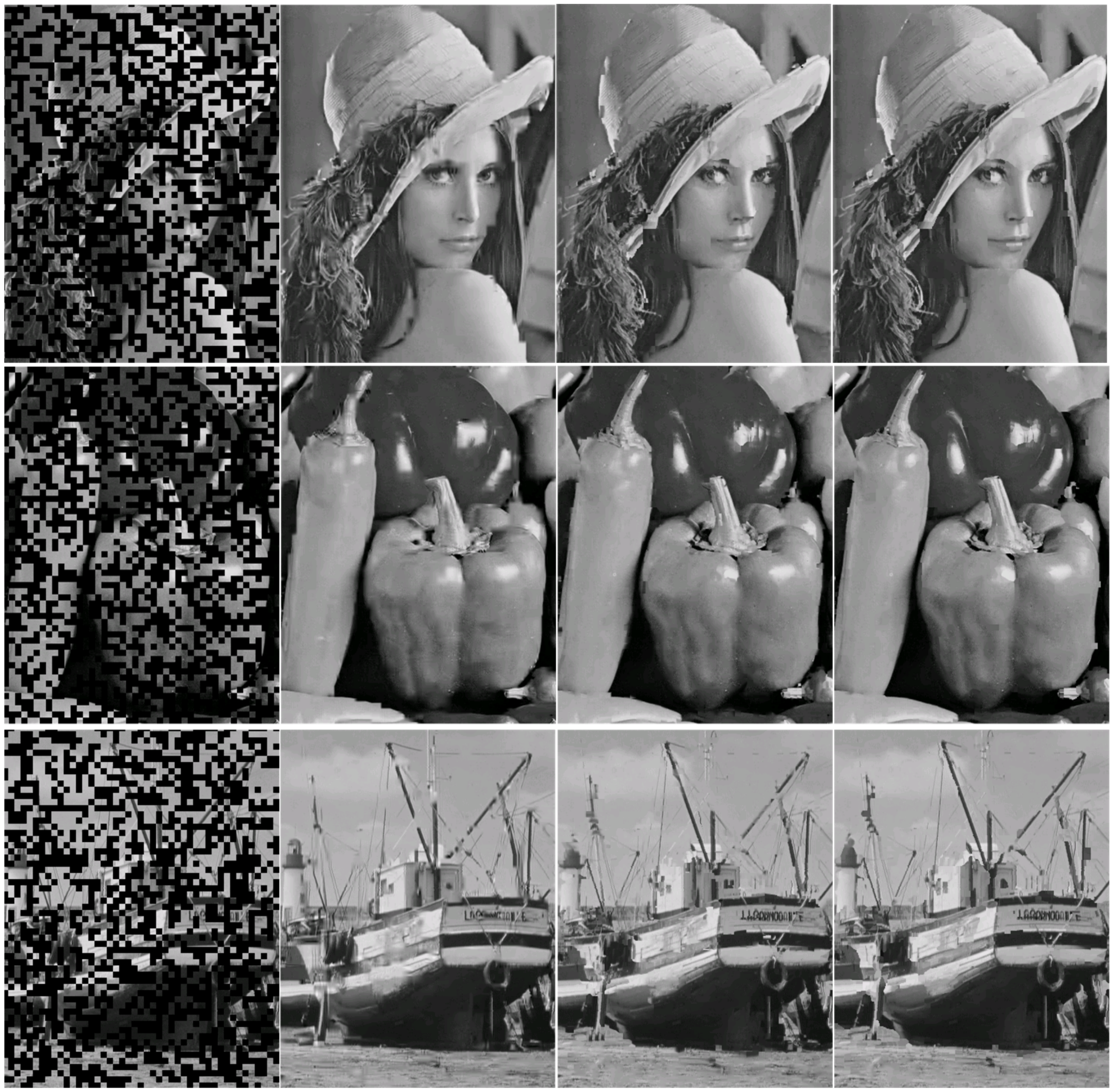

Fig. 2. Subjective quality comparison of reconstructed images from 50% random block loss. From left to right: corrupted images, SparseRec, KR and our method.

[5] E. P. Simoncelli and B. A. Olshausen, "Natural image statistics and neural representation," Annual Review of Neuroscience, vol. 24, pp. 1193-1216, 2001.

[6] A. Srivastava, A. B. Lee, E. P. Simoncelli, and S.-C. Zhu, "On advances in statistical modeling of natural images," Journal of Mathematical Imaging and Vision, vol. 18, no. 1, pp. 17-33, 2003.

[7] A. Buades, B. Coll, and J. Morel, "Nonlocal image and movie denoising," International Journal of Computer Vision, vol. 76, no. 2, pp. 123-139, 2008.

[8] Daniel Glasner, Shai Bagon, Michal Irani, " Super-resolution from a single image," ICCV 2009, pp. 349-356.

[9] G. Zhai and X. Yang, "Image Reconstruction from Random Samples with Multiscale Hybrid Parametric and Nonparametric Modeling," 2011 IEEE Visual Communications and Image Processing, pp.1-4, Tainan, Taiwan.

[10] S. Roth and M. J. Black, "Fields of experts: A framework for learning image priors," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., vol. 2. 2005, pp. 860-867.

[11] X. Liu, D. Zhao, R. Xiong, S. Ma, W. Gao, and H. Sun, "Transductive Regression with Local and Global Consistency for Image Super-Resolution," IEEE Data Compression Conference 2011, pp. 173-182, Snowbird, Utah, USA.

[12] X. Liu, D. Zhao, R. Xiong, S. Ma, W. Gao and H. Sun, " Image Interpolation Via Regularized Local Linear Regression," IEEE Trans. on Imag. Process., vol.20, no.12, pp.3455-3469, Dec.2011

[13] W. Chen, C. Shan, G. Haan, "Optimal Regularization Parameter Estimation for Spectral Regression Discriminant Analysis," IEEE Trans. Circuits Syst. Video Technol., (19)12: 1921-1926 (2009)