MDL Context Modeling of Images with Application to Denoising

Guangtao Zhai∗†, Xiaolin Wu∗, Xiaokang Yang† and Wenjun Zhang†

Abstract—The lately popularized patch-based nonlocal (NL) image processing approach is cast into a framework of statistical context modeling, a thoroughly studied topic in data compression and information theory. The adaptation of image patch (context) to local waveform is crucial to the performance of NL-type of image processing but yet lacks a rigorous study. In this paper we propose a minimum description length (MDL) approach for choosing the size and spatial configuration of the context in which a degraded pixel is to be restored. The MDL criterion of context formation aims to strike an optimal balance between the variance and bias of the errors in fitting a 2D piecewise autoregressive (PAR) model to input image signal. To exemplify the use of the proposed context modeling technique in image processing, an MDL-guided context-based image denoiser is derived and its performance evaluated. Empirical results show that the new context-based denoiser is highly competitive against the current state of the art.

I. INTRODUCTION

Recently, the non-local (NL) approach of image processing has generated a great deal of enthusiasm and success, in particular for the application of image denoising [1] [2] [3]. Indeed, the NL type of image denoisers have produced some of the best results in the literature so far. In traditional image denoising techniques, a pixel is denoised (estimated) by taking a weighted average of neighboring pixels. The NL image denoisers depart from the previous practice in that a weighted average is taken over a set of pixels that have similar context (pixel patch) to that of the current pixel. The advantage of the NL denoising methods over their predecessors is in that they utilize relevant sample statistics gathered from the entire image rather than from a local neighborhood, hence the terminology of NL. As such the NL approach of image processing is essentially a methodology of statistical context modeling for two-dimensional image signals.

Statistical context modeling of images is the problem of estimating conditional probability \( P(x|C^k(x)) \), where random variable \( x \) is the pixel in question and \( C^k \) is a patch of \( k \) pixels related to \( x \), called an order-\( k \) context of \( x \). Context modeling is a well studied topic in data compression and information theory for its role in conditional entropy coding [4] [5] [6] [7]. These works, which predated the popularity of non-local image processing methods, use samples of similar context instances chosen from the entire sample population instead of only in a local window of \( x \) to estimate \( P(x|C^k(x)) \).

The main challenge for statistical context modeling is the conflict between the use of high order model (large context) and context dilution (or the penalty of model cost as quantified by Rissanen information theoretically [8]). The higher the order, the more expressive the model, but also are more samples needed to estimate the model parameters. In statistical inference this is the classic problem of bias versus variance. For example, in the design of context-based image denoisers, one needs to optimize the context model in terms of its order, spatial configuration (i.e., the size and shape of the patch used in NL methods) and the training set that provide samples to learn the model parameters. To the best of our knowledge, the above three issues have not yet been systematically studied in the literature on NL image processing.

In this paper we study the problem of context-based image processing in the framework of of minimum description length (MDL). The underlying context model for the image is that of two-dimensional piecewise autoregressive (PAR) process, and the context is the 2D support of the PAR process. An algorithm is developed to adaptively choose the order and the spatial configuration of the modeling context, as well as the training sample set used to estimate the PAR model parameters such that the resulting model produces the shortest description length of the image. The resulting MDL-guided context model is applied to image denoising and is able to improve the performance of the existing NL methods, providing empirical evidence for the importance and benefits of optimizing the 2D patch in NL image processing.

In the remainder of the paper, the problem of adaptive context formation in MDL principle is formulated and studied in Section II. As a natural outcome of the study, an MDL-guided context-based image denoiser is derived in Section III. Some experimental results are given in Section IV in comparison with the current state of the art. Section V concludes.

II. MDL-GUIDED CONTEXT FORMATION

For ease and concreteness of our discussions, the problem of adaptive context formation is stated in the interest of image denoising. However, the following developments can be easily extended to other image processing tasks, such as restoration and coding. The pixels of a clean image are denoted by \( x_i \), and the noisy observation of \( x_i \) by \( y_i \),

\[
y_i = x_i + n
\]
where \( n \) is an additive white noise of zero mean. To avoid cluttered notations we use a single lexicographical index to identify the 2D location of a pixel. At the heart of our context-based image denoiser is a 2D piecewise autoregressive (PAR) model of images. With a PAR model of order \( K \) a pixel \( x_0 \) can be expressed as

\[
x_0 = C^K(x_0)\alpha^T + \epsilon.
\]

(2)

where \( C^K(x_0) = (x_1, x_2, \ldots, x_K) \) is a set of \( K \) pixels on which \( x_0 \) is statistically dependent, called an order-\( K \) context of \( x_0 \). The vector \( \alpha = (a_1, a_2, \ldots, a_K) \) is the set of PAR model parameters. The term \( \epsilon \) is a zero-mean Gaussian perturbation.

For the application of image denoising, one only has the noisy observation of the context \( C^K(x_0) \), denoted by \( C^K(y_0) = (y_1, y_2, \ldots, y_K) \). Given the \( K \)-point spatial configuration of \( C^K(y_0) \), the PAR model parameters \( \alpha = (a_1, a_2, \ldots, a_K) \) are estimated by

\[
\alpha = \arg \min_{\alpha \in \mathbb{R}^K} E_{C^K(y_0)|S}[\|y - C^K(y)\alpha^T\|_2]
\]

(3)

where \( C^K(y) \) stands for the 2D context of the same spatial configuration as \( C^K(y_0) \) but with respect to \( y \). The expectation is taken over a set of noisy context samples \( S \).

It follows from our problem formulation that the performance of the context-based denoiser heavily depends on the choices of context \( C^K(y_0) \) and sample set \( S \). The issue also arises for the patch used in the NL-type of image denoising techniques. The prevailing practice is quite ad hoc: a patch \( C^K(y_i) \) of fixed size and shape is empirically chosen, and kept the same over all locations \( i \). The shape of patch \( C^K(y_i) \) is made isotropic and centered at the current pixel position \( i \). In this design the underlying assumption is that the closer a pixel is to \( x_i \), the more related \( x_i \) it is. But this assumption is invalid wherever the local image waveform is anisotropic (e.g., edges and textures). In such cases, a signal-independent patch is clearly problematic because it brings irrelevant pixels and miss relevant ones in context modeling of \( x_i \), creating estimation biases. Therefore, we need to make the spatial configuration of \( C^K(y_i) \) adaptive to the local image waveform involving \( y_i \).

Two other design decisions also affect the performance of context-based denoisers: the order of the underlying context model \( K \) and the sample set \( S \) in (3) used in the solution of the estimation problem. They are concerned with the tradeoff between the estimation bias and variance, or the prevention of data overfitting. As such all above design decisions fall naturally into the framework of minimum description length (MDL) [4]. Specifically, we strike an optimal balance between the bias and variance of the estimation errors by minimizing

\[
L(C^K(y_0), S_k) = H(C^K(y_0), S_k) + \frac{k}{2|m|S_k} \log |S_k|.
\]

(4)

The variables in the MDL problem are the 2D context \( C^K(y_0) = \{y_1, y_2, \ldots, y_k\} \) for the current pixel \( y_0 \) and the sample set \( S_k \) from which the PAR model parameters \( \alpha = (a_1, a_2, \ldots, a_K) \) are learnt. The first term \( H(C^K, S_k) \) is the expected code length (measured by empirical entropy) of the error of fitting the PAR model to the observed waveform at \( y_0 \). The second term of (4) is the so-called model cost, and it quantifies the risk of data overfitting when the order \( k \) of the context \( C^K(y_0) \) is too high relative to the size of the sample set \( |S_k| \). The MDL design of context-based denoiser is to minimize (4) over all possible \( k \), \( C^K(y_0) \), and \( S_k \). Note that \( k \), \( C^K(y_0) \), and \( S_k \) are determined with respect to \( y_0 \).

In the MDL optimization of the size \( k \) and the shape of the modeling context \( C^K(y_0) \), we sequentialize the pixels in \( C^K(y_0) \) so that it grows spatially in \( k \). To adapt \( C^K(y_0) \) to the local waveform around \( y_0 \), we order the pixels \( y_i \in C^K(y_0) \) by ranking their correlations to current pixel \( y_0 \), instead of ranking \( y_i \) by the Euclidean distance between \( y_i \) and \( y_0 \). This creates an ordered set

\[
C^K(y_0) = \{y_1, y_2, \ldots, y_k\} \rho_{1,0} \geq \rho_{2,0} \geq \cdots \geq \rho_{k,0}
\]

(5)

where \( \rho_{k,0} \) is the correlation coefficient between \( y_0 \) and \( y_k \). The resulting context \( C^K(y_0) \) can have an arbitrary 2D shape, and its spatial configuration may not even be connected in texture areas. Contexts \( C^K \) of increasing order \( k \) are nested, i.e., \( C^{k_1} \subset C^{k_2} \) for any \( k_1 < k_2 \). The geometry of our context \( C^K(y_0) \) has a much greater degree of freedom than adaptive NL patches in recent literature: ellipse-shaped patch with the major axis aligned with the direction of local edge [9], or anisotropic patch created by shape-adaptive DCT [10].

To estimate the correlations \( \rho_{k,0} \) between the current pixel \( y_0 \) and a neighboring pixel \( y_k \), we need to judiciously collect samples since most natural images have nonstationary second order statistics. To this end we create a seed patch \( C_0(y) = (c_1, c_2, \ldots, c_M) \), \( M > K \), such that \( c_m \) is the \( m \)-th closest pixel to \( x \). We form a sample set \( T \) of patches \( C_0(y) \) that sufficiently match the seed patch \( C_0(y) \)

\[
T = \{C_0(y) \parallel C_0(y) - C_0(y)\| \leq \tau_T\}
\]

(6)

where \( \tau_T \) is a threshold. Then we estimate

\[
\rho_{m,0} = \frac{\sum_{C_0(y) \in T} c_m y - \frac{1}{|T|} \sum_{C_0(y) \in T} y \sum_{C_0(y) \in T} c_m}{\sqrt{\sum_{T} y^2 - \left(\sum_{T} y\right)^2} \sqrt{\sum_{T} c_m^2 - \left(\sum_{T} c_m\right)^2}}
\]

(7)

using the sample vectors in \( T \). With these estimated correlation coefficients \( \rho_{m,0}, 1 \leq m \leq M \), adaptive nested 2D contexts \( C^K(y_0) = \{y_1, y_2, \ldots, y_k\}, k = 1, 2, \ldots \), are obtained by (5).

The set \( S_k \) in (4) supplies the samples to estimate the PAR model parameters \( \alpha = (a_1, a_2, \ldots, a_K) \) in (3), and it interacts with the model order \( k \) in the MDL objective function (4). Similarly to the tree splitting technique of Rissanen’s algorithm Context [11], we refine the sample set \( S_k \) by matching the higher order statistics of the current context \( C^K(y_0) \) as \( k \) increases. Specifically, we recursively define the
set $S_k$

$$S_k = \{ C^k(y) \mid \| C^k(y) - C^k(y_0) \| \leq \tau_S, C^k(y) \in S_k \}$$

$k > 1$;

$$S_1 = T$$

(8)

where $\tau_S$ is a threshold. It follows from (8) that $S_1 \supseteq S_2 \supseteq \cdots \supseteq S_k \supseteq \cdots$, reflecting the fact that fewer samples are available to learn the PAR model of higher order, or the inherent conflict of estimation bias and variance.

Given the context $C^K(y_0)$ and the sample set $S_k$, the PAR model coefficients $a$ are given by (3), and thus the model fit errors $e$ in $S_k$ can be measured. From the error histogram $P(e)$ in $S_k$ we can compute the empirical entropy term $H(C^K, S_k)$ in (4):

$$H(C^K, S_k) = -\sum_{-255 \leq e \leq 255} P(e) \log P(e).$$

(9)

In practice, $P(e)$ is commonly modeled as Laplacian, hence $P(e)$ can be determined by the variance $\sigma^2$ of $e$, specifically

$$P(e) = \begin{cases} 
1 - e^{-}\frac{|e|}{\sigma} & e = 0 \\
\frac{1}{2} \left( e^{-\frac{|e|+1}{\sigma^2}} - e^{-\frac{|e|+2}{\sigma^2}} \right) & 0 < |e| < 255 \\
\frac{1}{2} e^{-\frac{|e|+2}{\sigma^2}} & |e| = 255 
\end{cases}$$

(10)

### III. CONTEXT-BASED DENOISING

To denoise the current pixel $y_0$, we compute the MDL context $C^K(y_0)$, the MDL sample set $S_K$, and the corresponding PAR model parameters $a$ for estimating $x_0$ in context $C^K(y_0)$, as described in the proceeding section. Now considering $y_i$ such that $C^K(y_i) \in S_K$, it follows from (2) that

$$x_0 = x_i + [C^K(x_0) - C^K(x_i)]a^T + \epsilon$$

$$= C^K(y_i)a^T + [C^K(x_0) - C^K(x_i)]a^T + \epsilon$$

(11)

where $\epsilon$ is the error of the PAR estimator $\hat{x}_i = C^K(y_i)a^T$. Letting $Y_K = \{ y_i \mid C^K(y_i) \in S_K \}$, we finally derive a MDL-guided context-based estimator of $x_0$

$$\hat{x}_0 = E_{Y_K} \left[ C^K(y_i)a^T + [C^K(x_0) - C^K(x_i)]a^T + \epsilon \right]$$

$$= E_{Y_K} \left[ C^K(y_i)a^T \right]$$

(12)

in which we use the facts that $\epsilon$ and $\epsilon$ are zero mean, and

$$E_{Y_K} \left[ C^K(x_0) - C^K(x_i) \right] = E_{Y_K} \left[ C^K(y_0) - C^K(y_i) \right] = 0$$

(13)

Since the context-based denoiser operates on noisy samples $y_i$, we can also adopt the TLS technique in estimating the PAR model parameters. Specifically, the TLS estimates of the PAR model parameters are

$$a = \arg \min_{a} \Delta S_K, \Delta Y_K \| \Delta S_K, \Delta Y_K \|_F$$

subject to

$$\Delta S_K, \Delta Y_K \|_F$$

(14)

where $\Delta S_K$ and $\Delta Y_K$ are the perturbations to account for the influence of noise. The subscript $F$ stands for Frobenius norm.

### IV. EXPERIMENTAL RESULTS

To demonstrate the power of the proposed MDL context formation technique, we compare our context-based image denoiser with the most competitive methods: NL-means filter [1], K-SVD [12] and BM3D [3]. Fig. 1 shows the results of these methods on a portion of test image Cameraman corrupted by additive zero-mean Gaussian noise of $\sigma = 20$. Note that along the high-contrast edges the context-based MDL denoiser has the least ringing artifacts. In Fig. 2 we also compare our method with BM3D, which is widely regarded the best performer so far, on test image Lena that has real scanner noises of unknown model. In this case our method appears to preserve fine details better than BM3D (see eyelashes). This comparison is significant because it reveals how the two methods measure up when applied to real physical instead of synthetic noises. These two methods are also compared on two synthetic images corrupted by synthetic noises. One is a test pattern corrupted by zero-mean Gaussian noise of $\sigma = 5$ as shown in Fig. 3; the other is a ray-traced image corrupted by a compound noise whose model is $y = x + (5 + 0.05x)\eta$, $\eta \sim N(0,1)$, as shown in Fig. 4. In both cases, our method wins in PSNR as well as visual quality. Note that the experiment of Fig. 4 is designed to simulate reality, because the object model (a car headlight) used in image synthesis is highly realistic and it achieves photo realism, and because the compound noise model $y = x + (5 + 0.05x)\eta$ is more realistic than additive white Gaussian noises, including both additive and multiplicative noises.

### V. CONCLUSION

This paper sheds a light of statistical context modeling on nonlocal image processing. An MDL-guided context (patch) formation algorithm is proposed to strike an balance between the bias and variance of the estimation methods encountered in image processing. This work leads to a highly competitive context-based image denoiser.

### REFERENCES


Fig. 1. Denoising results on a portion of Cameraman. 1st row, from left to right: original image, noisy image and denoised with nonlocal means filter; 2nd row, from left to right: denoised with K-SVD, BM3D and the proposed method.

Fig. 2. Denoising results on a portion of Lena. From left to right: original image with scanner noises, denoised with the proposed method and BM3D.

Fig. 4. Denoising results on a ray tracing image. From left to right: noisy image (PSNR=22.56 dB), original image, denoised with the proposed method (34.98 dB) and BM3D (29.29 dB).