Correspondence

Set Theoretic Compression with an Application to Image Coding

Nguyen T. Thao and Martin Vetterli

Abstract—We show that the complete information that is available after an image has been encoded is not just an approximate quantized image version, but a whole set of consistent images that contains the original image by necessity. From this starting point, we develop a set of tools to design a new class of encoders for image compression, based on a set decomposition and recombination of image features. As an initial validation, we show the results of an experiment where these tools are used to modify the encoding process of block discrete cosine transform (DCT) coding in order to yield less blocking artifacts.

I. INTRODUCTION

The problem of compression is the trade-off between description complexity and approximation quality. In traditional source coding, a given cost metric, e.g., mean square error (MSE), is minimized under a rate constraint. Attempts to include perceptual criteria have been made, but are difficult to model and include in optimization procedures. That is, "perceptual bit allocation" is more difficult than the standard bit allocation procedures. Furthermore, jointly optimizing several criteria like MSE and perceptual error can be a difficult problem.

In this correspondence, we propose a framework that defines a coded image as belonging to the intersection of several convex sets [1], and show how such sets can represent both traditional criteria as well as perceptually motivated ones. We thus call the method "set theoretic compression."

Using this framework, we investigate a classic coding problem, that of low bit rate discrete cosine transform (DCT) compression without blocking artifacts, and demonstrate improved performance using our method.

The basic idea is to efficiently describe convex sets containing the image. Of course, traditional quantization does just this. A hypercube with sides of size Δ is defined around the vector representing the image (see Fig. 1), and the rate distortion trade-off of such quantization is well studied for many sources of interest. A more sophisticated version is, of course, vector quantization [2], where the hypercubes are replaced by more general partitioning cells. For general convex sets, rate distortion trade-offs are not understood. Our main purpose here is to formalize a coding design framework in which convex sets can be targeted to represent selected perceptual features at low bit rate.

In Section II, we show that in any traditional coding scheme, the complete information that is provided by an encoder about its input image is always a convex set, whether this has been purposely thought by the designer or not (see Figs. 1 and 2). This set is

Fig. 1. Scalar quantization. The classical and the set theoretic interpretations of the output C are, respectively, represented by a dotted and a solid arrow.
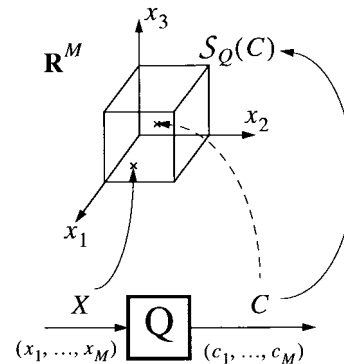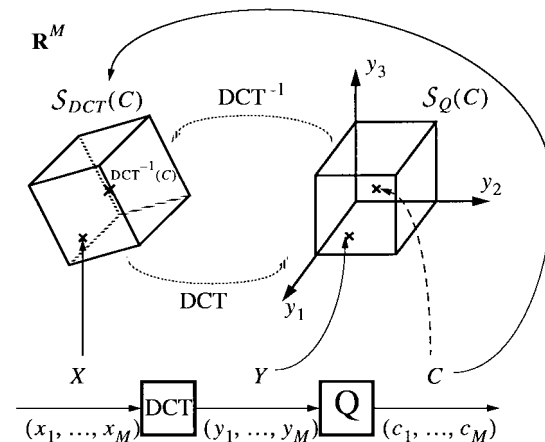

Fig. 2. DCT quantization. The classical and the set theoretic interpretations of the output C are, respectively, represented by a dotted and a solid arrow.

simply the set of all possible input images that provide the same coded output. We call it the *encoded set*. Decoding must at least recover an image from this set. It appears that decoding traditionally consists of systematically picking one particular element in this set (usually its center). However, at low bit rate, this simple decoding approach becomes insufficient as pointed out in [3]–[5] in the case of block DCT coding, and decoded image quality can be significantly improved by more carefully exploiting the complete set information.

While set-based coding provides a perspective for analyzing any type of coding algorithm, this work focuses on the case where the encoder is composed of several subencoders used in parallel. The global encoded set is mathematically described as the intersection of the sets provided by each subencoder. Suppose that the subencoders are specialized in capturing different types of features. By retrieving an estimate in the global encoded set, we will automatically reproduce an image that contains simultaneously all the targeted features. This opens new directions for flexible design of perceptually motivated encoders. A price to pay is, however, a more complex decoding where projections onto convex sets (POCS) [6], [7] have to be used to retrieve an element in an intersection (see Fig. 3). We give the tools to perform such a type of decoding in Section III.
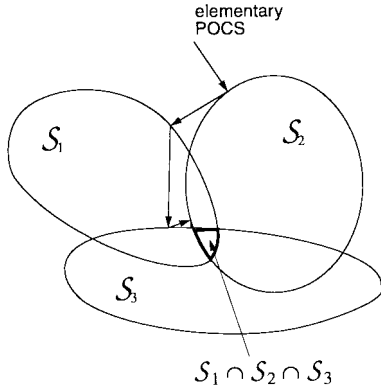
Fig. 3.   Alternating projection algorithm. Case of three convex sets.

As a preliminary demonstration, we propose in Section IV a toy experiment of set theoretic design. In the context of block-based coding, we target two types of features to be encoded: i) the image details within the blocks and ii) the image smoothness across the block boundaries. The second type of feature is targeted to resolve the underlying uncertainty that causes the perceptually important "block artifacts" in decoded images. A standard DCT coder is used to represent feature i), while a special block boundary transition (BBT) encoder represents feature ii). By comparing our decoded image with that of a DCT encoder used alone at the fixed bit rate of 0.25 b/pixel (see Fig. 4), we show how, by set theoretic design, we can effectively manipulate the targeted features to produce a perceptually more pleasant image.

II. Encoded Sets

A. Image and Coding Formalization

We consider $N \times N$ black and white images. Every image is an M-dimensional vector $X = (x_1, x_2, \cdots, x_M) \in \mathbf{R}^M$ where $M = N^2$ and x_k is the luminance value of the kth pixel. In block-based image processing, an image X is divided into a partition of p nonoverlapping blocks B_i of size M_i, such that $M_1 + \cdots + M_p = M$. Using the cross product[1], such an image can be written as $X = B_1 \times \cdots \times B_p$. For a given block B of any size and geometry, $\|B\|^2$ designates the squared sum of its elements. For given subsets $\mathcal{S}_1, \cdots, \mathcal{S}_p$ of $\mathbf{R}^{M_1}, \cdots, \mathbf{R}^{M_p}$ respectively, $\mathcal{S} = \mathcal{S}_1 \times \cdots \times \mathcal{S}_p$ designates the set of images $X \in \mathbf{R}^M$ of the form $B_1 \times \cdots \times B_p$ where $B_i \in \mathcal{S}_i$ for all $i \in \{1, \cdots, p\}$.

An image encoder is an operator that maps an input image X into an encoded output C. The corresponding encoded set as defined in the introduction will be typically denoted by $\mathcal{S}(C)$. In the subsequent sections, we identify this set for various types of encoders.

B. Scalar Quantization

When an image X is quantized to give a new image $C = Q(X)$, every pixel x_i of X necessarily belongs to a certain interval $I_i(c_i)$ of $\mathbf{R}$. Typically, $I_i(c_i)$ is equal to $[c_i - \frac{\Delta_i}{2}, c_i + \frac{\Delta_i}{2}]$ where Δ_i is the given quantization step size for the ith pixel. In general, $I_i(c_i)$ is an interval of $\mathbf{R}$ that is uniquely determined by i and c_i. According to the notations of Section II-A, we have $C = Q(X)$ if and only if X belongs to the product set $\mathcal{S}_Q(C) = I_1(c_1) \times \cdots \times I_N(c_N)$. When the quantization is uniform, $\mathcal{S}_Q(C)$ is simply a hypercube of $\mathbf{R}^M$ (see Fig. 1). Otherwise, it is in general a rectangular parallelepiped.

---

[1]The cross product is such that $(x_1, x_2) \times (x_3, x_4) = (x_1, x_2, x_3, x_4)$.

C. Transform Quantization

A more general type of quantization consists in calculating $C = Q(T(X))$ where T is an isometry from $\mathbf{R}^M$ to $\mathbf{R}^M$. We recall that an isometry is a linear operator T that satisfies $\|T(Y) - T(X)\| = \|Y - X\|$. Such an operator is automatically invertible. In the case of 8×8 images, a typical example of isometry is the DCT operator. In general, we have $C = Q(T(X))$ if and only if $T(X) \in \mathcal{S}_Q(C)$, which is equivalent to $X \in T^{-1}(\mathcal{S}_Q(C))$. The encoded set is thus $\mathcal{S}_T(C) = T^{-1}(\mathcal{S}_Q(C))$ (see DCT example in Fig. 2). Because T is linear and conserves the distance associated with the norm $\|\cdot\|$, $T^{-1}(\mathcal{S}_Q(C))$ is also a rectangular parallelepiped.

D. Energy Quantization

In this correspondence, we will also use a type of encoder that is less common. For a given image X, we will be interested to measure the scalar encoded value $c = q(\|H(X)\|)$ where H is a given linear operator from $\mathbf{R}^M$ to $\mathbf{R}^{M'}$ and $q(\cdot)$ is a scalar quantization function of positive numbers such that $q(x) \geq x$. Because of this constraint, it is clear that X necessarily belongs to the set $\mathcal{S}_\mathcal{E}(c) = \{Y \in \mathbf{R}^M \mid \|H(Y)\| \leq c\}$. Because $\|H(\cdot)\|^2$ is a quadratic function, it is easy to show that $\mathcal{S}_\mathcal{E}(c)$ is a convex set. The set $\mathcal{S}_\mathcal{E}(c)$ will be considered as the encoded set[2].

E. Block-Based Quantization

Block-based quantization consists in splitting the input image X into a regular partition of blocks $B_1 \times \cdots \times B_p$, quantizing every block B_i into an encoded block D_i and forming the encoded output $C = D_1 \times \cdots \times D_p$. According to the previous paragraphs, the encoding of B_i into D_i implies that B_i must belong to a certain set $\mathcal{S}_i(D_i)$. According to Section II-A, it is then clear that X must belong to the encoded set $\mathcal{S}(C) = \mathcal{S}_1(D_1) \times \cdots \times \mathcal{S}_p(D_p)$. Scalar quantization is a trivial example of block-based quantization where the blocks are reduced to single pixels. A more important example is the block DCT coding where the image is split into 8×8 blocks on which DCT quantization is applied. Every set $\mathcal{S}_i(D_i)$ is in this case the rectangular parallelepiped $\text{DCT}^{-1}(\mathcal{S}_Q(D_i))$. Because the cross product is compatible[3] with the orthogonality in $\mathbf{R}^M$, the global encoded set $\mathcal{S}(C) = \mathcal{S}_1(D_1) \times \cdots \times \mathcal{S}_p(D_p)$ is also a rectangular parallelepiped.

III. Set Theoretic Decoding in Parallel Coding

As explained in the introduction, when an encoder is composed of n subencoders producing, respectively, n outputs $C_1, \cdots, C_n$ for a given input, the global encoded set is $\mathcal{S}(C_1, \cdots, C_n) = \mathcal{S}_1(C_1) \cap \cdots \cap \mathcal{S}_n(C_n)$ where $\mathcal{S}_1(C_1), \cdots, \mathcal{S}_n(C_n)$ are the sets provided by each subencoder. To extract an estimate from the intersection, we need to perform POCS on each of the individual sets (see Fig. 3).

Theorem and Definition III.1 ([7]): Let $\mathcal{S}$ be a convex subset of $\mathbf{R}^M$. For any given $X \in \mathbf{R}^M$, there exists a unique element $\hat{X}$ of the closure $\overline{\mathcal{S}}$ of $\mathcal{S}$ such that for all $X' \in \mathcal{S}$, $\|X' - X\| \geq \|\hat{X} - X\|$. The element $\hat{X}$ is called the projection of X onto the convex set $\mathcal{S}$.

---

[2]The encoded set should rigorously be $\mathcal{S}_\mathcal{E}(c) = \{Y \in \mathbf{R}^M \mid c' \leq \|H(Y)\| \leq c\}$ where c' is the nearest quantization level inferior to c. However, this set is not convex, and, in fact, will not be of practical use in decoding.

[3]By compatibility we mean the following property. In the cross-product decomposition $\mathbf{R}^{M_1} \times \cdots \times \mathbf{R}^{M_p}$ of $\mathbf{R}^M$, the two images $X = 0 \times \cdots \times 0 \times B_i \times 0 \times \cdots \times 0$ and $X' = 0 \times \cdots \times 0 \times B_i' \times 0 \times \cdots \times 0$ are orthogonal in $\mathbf{R}^M$ if and only if B_i and B_i' are orthogonal in $\mathbf{R}^{M_i}$, and the two images $X = 0 \times \cdots \times 0 \times B_i \times 0 \times \cdots \times 0$ and $Y = 0 \times \cdots \times 0 \times B_j \times 0 \times \cdots \times 0$ are automatically orthogonal when $i \neq j$.

(a)   (b)

Fig. 4.   DCT coding. (a) Original 512×512 image. (b) DCT decoded image at bit rate 0.25/pixel (PSNR = 30.15 dB).

TABLE I
EXPRESSION OF POCS ON TYPICAL CONVEX SETS

case	set type	convex set $\mathcal{S}$	set parameters	projection of $X \notin \mathcal{S}$ on $\mathcal{S}$	projection parameters
(a)	interval	$[a, b]$	$M = 1, a < b$	a, if $X < a$ b, if $X > b$	
(b)	sphere	$\left\{ Y \in \mathbf{R}^M \,\middle\|\, \|Y\| \le c \right\}$	$c \in \mathbf{R}^+$	$c \frac{X}{\|X\|}$	
(c)	ellipsoid	$\left\{ Y \in \mathbf{R}^M \,\middle\|\, \|H(Y)\| \le c \right\}$	$c \in \mathbf{R}^+$ H is a linear operator	no algebraic closed form	
(d)	spherical cylinder	$\left\{ Y \in \mathbf{R}^M \,\middle\|\, \|Y \cdot U^\top\| \le c \right\}$	$c \in \mathbf{R}^+, M = N^2$ Y is expressed as an $N \times N$ matrix, U is a given N dimensional row vector	$X - (1 - \frac{c}{\|X \cdot U^\top\|}) X \cdot \frac{U^\top \cdot U}{U \cdot U^\top}$ (see proof in appendix)	
(e)	"rotated" convex set	$T^{-1}(\mathcal{S}')$	T is an isometry $\mathcal{S}'$ is a convex set	$T^{-1}(P(T(X)))$	P is the POCS on $\mathcal{S}'$
(f)	cross product set	$\mathcal{S}_1 \times \cdots \times \mathcal{S}_p$	$\mathcal{S}_1, ..., \mathcal{S}_p$ are convex sets of $\mathbf{R}^{M_1}, ..., \mathbf{R}^{M_p}$ respectively, with $M_1 + ... + M_p = M$	$P_1(B_1) \times \cdots \times P_p(B_p)$	$B_1 \times \cdots \times B_p$ is the decomposition of X in $\mathbf{R}^{M_1} \times \cdots \times \mathbf{R}^{M_p}$, P_i is the POCS on $\mathcal{S}_i$

As a consequence of this definition, the projections onto typical convex sets can be easily derived and their expressions are given in Table I. Using this table, it is easy to derive the projections on the sets encoded by the various quantization schemes presented in Section II. The case of scalar quantization is solved with the case (f) of the table, where every $\mathcal{S}_i$ is an interval, combined with the case (a). DCT quantization is solved with a combination of (e), (f) and (a). Indeed, from Section II-C, the encoded set in this case is of the type $\mathrm{DCT}^{-1}(\mathcal{S}_Q)$ [case (e)] and $\mathcal{S}_Q$ is the cross product of intervals [cases (f) and (a)]. As shown by case (b), we cannot perform the projection on the set encoded by any energy quantizer. However, in this correspondence, we will consider only energy quantizers where $H(\cdot)$ is such that $H(Y) = Y \cdot U^\top$, where Y is considered as an $N \times N$ matrix, and U is a fixed N-dimensional row vector. This case is covered by case (d) in Table I, and the projection result is proved in the Appendix. Finally, block-based encoding schemes in general are dealt with using case (f) of the table, where $\mathcal{S}_i$ is the set encoded within each block.

IV. EXPERIMENT OF SET THEORETIC ENCODER DESIGN

In this section, we describe our set theoretic design experiment as explained in the introduction. We first describe our block boundary transition (BBT) encoder in Section IV-A. We then give the coding experimental results in Section IV-B.

A. Design of the BBT Encoder

The BBT encoder is composed of two subencoders used in parallel, processing the vertical and the horizontal boundaries, respectively. Like the DCT encoder, each of them is an 8×8 block based encoder, but using a different image partitioning, as shown in Fig. 5. Let us concentrate on the vertical BBT encoder. In our design,
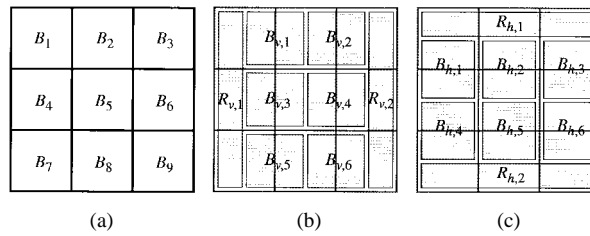
Fig. 5. An 8×8 image block partitioning. (a) DCT coding. (b) vertical BBT coding. (c) horizontal BBT coding.
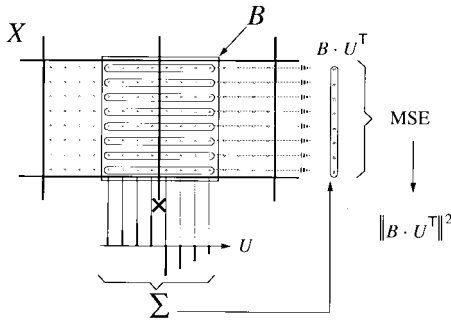

Fig. 6. Vertical BBT energy calculation.

for each block B of the vertical BBT partitioning [see Fig. 5(b)], the encoder quantizes the energy $\|B \cdot U^\top\|^2$ where B is viewed as an 8×8 matrix and U is a fixed row vector of length eight. When taking $U = [0 \; 0 \; 0 \; +1 \; -1 \; 0 \; 0 \; 0]$, $\|B \cdot U^\top\|^2$ gives the simplest measurement of block boundary discontinuity between two adjacent DCT blocks that one could think of, and that was already used in [5]. As can be seen in Fig. 6, this amounts to calculating the mean squared difference between the adjacent pixels of two neighboring DCT blocks along their separation boundary. The general energy function $\|B \cdot U^\top\|^2$ allows us to use more complex vectors U. We will see that better discontinuity measurements are obtained with $U = [+1 \; +2 \; +3 \; +4 \; -4 \; -3 \; -2 \; -1]$. Once the vector U has been chosen, the vertical BBT output code of an input image X, decomposed as $R_{v,1} \times B_{v,1} \times B_{v,2} \times \cdots \times B_{v,p} \times R_{v,2}$ according to the partitioning of Fig. 5(b), is $C_v = c_1 \times c_2 \times \cdots \times c_p$ such that $c_k = q(\|B_{v,k} \cdot U^\top\|)$. The blocks $R_{v,1}$ and $R_{v,2}$ are not coded. The horizontal BBT encoder is designed in a similar manner. The two BBT encoders define the encoded sets $\mathcal{S}_v(C_v)$ and $\mathcal{S}_h(C_h)$, respectively. Their description is given in Section II and their respective POCS are derived in Section III.

B. Decoding Experiments

The global encoded set is the intersection between $\mathcal{S}_v(C_v)$, $\mathcal{S}_h(C_h)$ and the DCT encoded set. The decoding can be performed by alternating the POCS respective to these three sets. In our experiment, we assign 0.21 b/pixel and 0.04 b/pixel to the DCT and the BBT encoders, respectively, thus maintaining the total bit rate of 0.25 b/pixel. We start the iteration from the purely decoded DCT image at 0.21 b/pixel shown in Fig. 7(a). As an initial experiment, we first assign an infinite bit rate to the BBT encoders. In other words, we quantize the energy coefficients with infinite precision. We show in Fig. 7(b) the decoded image in the case where $U = [0 \; 0 \; 0 \; +1 \; -1 \; 0 \; 0 \; 0]$. One can observe some reduction of the blocking artifacts compared to the purely DCT decoded image of Fig. 7(a). Much better deblocking results are obtained with $U = [+1 \; +2 \; +3 \; +4 \; -4 \; -3 \; -2 \; -1]$ as shown in Fig. 7(c). The following experiments all use this U vector.

Next, we propose to quantize the BBT energy coefficients using a geometric quantization scale. More precisely, we use the quantization function $q(e) = \Delta^m \cdot e_0$ such that m is the smallest integer such that $\Delta^m \cdot e_0 \geq e$ and where e_0 and Δ are predefined parameters. Note that this ensures that $q(e) \geq e$ as required in Section II-D. The coefficient e_0 is a normalization factor that we adjust for each BBT block. We take $e_0 = \|B \cdot U^\top\|$ where B is the block extracted from the purely DCT decoded image at the same BBT block location. For the sake of good data compression, we ignore the BBT encoded values of the blocks where the purely DCT decoded image is smoother than the original image according to the function $\|B \cdot U^\top\|$. Finally, the BBT quantized energy coefficients are compressed using a spatial run length coding method similar to the JPEG treatment of the AC coefficients of the DCT. After adjusting Δ so that the BBT encoder bit rate is 0.04 b/pixel, we obtain the decoded image of Fig. 7(d). Although some degradation can be observed compared to the infinite bit rate image of Fig. 7(c), the loss of quality is quite limited. This image is the result of our complete modified encoder at the fixed bit rate of 0.25 b/pixel. Compared with the original DCT image at the same bit rate shown in Fig. 4(b), we obtain a more pleasant image, where some detail information has been replaced by some information about the block boundary transitions.

V. SUMMARY AND DISCUSSION

With the set theoretic interpretation of image compression, we have introduced new tools for designing an encoder. In a preliminary experiment, we applied these tools to show how the set theoretic approach can effectively be used to manipulate the perceptual encoded features for improving block DCT coding. But the general motivation is to present a new technique for designing image encoders. This technique seems particularly suitable to very low bit rate compression, where the features to be preserved become more specific and more object oriented but also more heterogeneous. For example, the lower the bit rate is, the more one has to selectively extract perceptual features such as edges, texture, smooth areas, etc. In classical encoding schemes, the separation and reconstruction of these different features is usually performed by either linear decomposition or image segmentation, which have known limitations. In the set theoretic approach, the reconstructed estimate is an image that simultaneously reproduces the required features (to a certain precision) with no need for linear separation and no artificial image segmentation.

Philosophically speaking, our set theoretic technique is intermediate between vector quantization—which typically uses the notion of encoded set (cell) in a very optimal manner (from an information theoretic point of view)—and object-oriented coding, which performs feature extraction in a very targeted manner (from a direct perceptual point of view). While our technique may not have the strong points of these two extreme techniques, it avoids their respective drawbacks (lack of universal modeling function of visual perception, and complex decoding in the first case, too restrictive image coding range and very irregular coding algorithms not suitable for VLSI implementation in the second case).
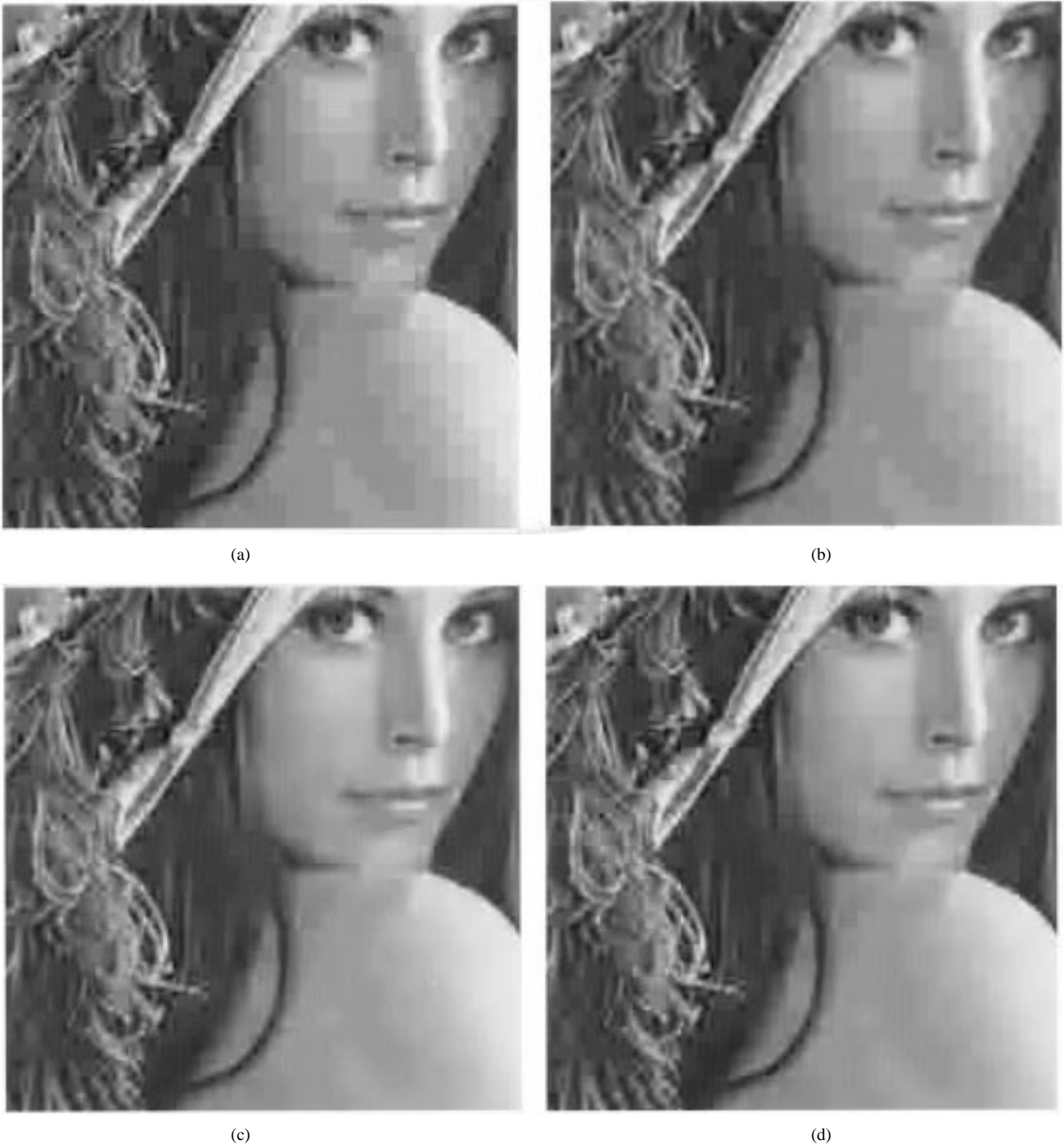
(a)

(b)

(c)

(d)

Fig. 7. Coding experiments on Lenna image. (a) DCT reconstructed image at bit rate 0.21 b/pixel (PSNR = 29.03 dB). (b) Projection of image (a) on $\mathcal{S}_v(C_v)$ and on $\mathcal{S}_h(C_h)$ with vector $U = [0 \ 0 \ 0 \ 1 \ -1 \ 0 \ 0 \ 0]$ (PSNR = 29.91 dB). (c) Projection of image (a) on $\mathcal{S}_v(C_v)$ and on $\mathcal{S}_h(C_h)$ with vector $U = [1 \ 2 \ 3 \ 4 \ -4 \ -3 \ -2 \ -1]$ (PSNR = 29.97 dB). (d) Same as (c) but at bit rate 0.04 b/pixel for the BBT encoding; the total bit rate is 0.25 b/pixel (PSNR = 29.67 dB).

Appendix
Proof of Table I, Case (d)

Because $X \notin \mathcal{S}$, its projection X' on $\mathcal{S}$ is such that $\|X' - X\|$ is minimized subject to the equality constraint $\|X' \cdot U^\top\| = c$. Consider the space $\mathcal{V}$ of $N \times N$ matrices W such that $W \cdot U^\top = 0$. It can be shown that the orthogonal space $\mathcal{V}^\perp$ of $\mathcal{V}$ in $\mathbf{R}^M$ is spanned by the $N \times N$ matrices of the type $V^\top \cdot U$ where V is any N-dimensional row vector. Let $X = V^\top \cdot U + W$ and $X' = V'^\top \cdot U + W'$ be, respectively, the orthogonal decomposition of X and X' in $\mathcal{V} \oplus \mathcal{V}^\perp$. We necessarily have $W' = W$. Otherwise, by taking $X'' = V'^\top \cdot U + W$, we would obtain $\|X' - X\|^2 = \|X'' - X\|^2 + \|W' - W\|^2 > \|X'' - X\|^2$ and $\|X'' \cdot U^\top\| = c$, which would contradict the fact that X' is the projection of X on $\mathcal{S}$. As a result $X' - X = (V' - V)^\top \cdot U$, which implies that $\|X' - X\|^2 = \|V' - V\|^2 \cdot \|U\|^2$. Now, we also have $c = \|X' \cdot U^\top\| = \|V'^\top \cdot U \cdot U^\top + W' \cdot U^\top\| = \|V'^\top \cdot (U \cdot U^\top)\| = (U \cdot U^\top)\|V'\|$. Therefore V' is such that $\|V' - V\|$ is minimized subject to $\|V\| = \frac{c}{U \cdot U^\top}$. Therefore, V' is the projection of V onto the sphere of radius $\frac{c}{U \cdot U^\top}$, which is equal to $\frac{c}{U \cdot U^\top} \frac{V}{\|V\|}$. Then, $X' - X = (\frac{c}{(U \cdot U^\top)\|V\|} - 1)V^\top \cdot U$.

Because $X \cdot U^{\top} = V^{\top} \cdot U \cdot U^{\top} + W \cdot U^{\top} = V^{\top} \cdot (U \cdot U^{\top})$, we have $\|X \cdot U^{\top}\| = (U \cdot U^{\top})\|V\|$ and $V^{\top} = \frac{X \cdot U^{\top}}{U \cdot U^{\top}}$. This leads to $X' = X - (\frac{c}{\|X \cdot U^{\top}\|} - 1)X \cdot \frac{U^{\top} \cdot U}{U \cdot U^{\top}}$.

References

[1] N. T. Thao, K. Asai, and M. Vetterli, "Set theoretic compression with an application to image coding," in *Proc. IEEE Int. Conf. Image Processing*, Nov. 1994, vol. II, pp. 336–340.

[2] A. Gersho and R. M. Gray, *Vector Quantization and Signal Compression*. Boston, MA: Kluwer, 1992.

[3] R. Rosenholtz and A. Zakhor, "Iterative procedures for reduction of blocking effects in transform image coding," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 2, pp. 91–95, Mar. 1992.

[4] S. J. Reeves and S. L. Eddins, "Comments on 'Iterative procedures for reduction of blocking effects in transform image coding'," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 3, pp. 439–440, Dec. 1993.

[5] Y. Yang, N. P. Galatsanos, and A. K. Katsaggelos, "Regularized reconstruction to reduce blocking artifacts of block discrete cosine transform compressed images," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 3, pp. 421–432, Dec. 1993.

[6] D. C. Youla and H. Webb, "Image restoration by the method of convex projections: Part 1—Theory," *IEEE Trans. Med. Imag.*, vol. MI-1, pp. 81–94, Oct. 1982.

[7] P. L. Combettes, "The foundations of set theoretic estimation," *Proc. IEEE*, vol. 81, pp. 1175–1186, Feb. 1993.

Best Neighborhood Matching:
An Information Loss Restoration Technique for Block-Based Image Coding Systems

Zhou Wang, Yinglin Yu, and David Zhang

Abstract—**Imperfect transmission of block-coded images often results in lost blocks. A novel error concealment method called *best neighborhood matching* (BNM) is presented by using a special kind of information redundancy—blockwise similarity within the image. The proposed algorithm can utilize the information of not only neighboring pixels, but also remote regions in the image. Very good restoration results are obtained by experiments.**

Index Terms—**Error concealment, image coding, image restoration.**

I. Introduction

Recently, many image coding algorithms have been developed to reduce the bit rate for digital image and video representation and transmission. Among them, block-based techniques have proved to be the most practical and are adopted by most existing image and video compression standards such as the Joint Photographers Expert Group

(JPEG) [1], Motion Picture Expert Group (MPEG) [2], and H.261 [3]. Since real-world communication channels are not error free, the coded data transmitted on them may be corrupted. Block-based image coding systems are vulnerable to transmission impairment. Loss of a single bit often results in loss of a whole block and may cause consecutive block losses.

Error concealment is aimed at masking the effect of these missing blocks to create subjectively acceptable images. So far many error concealment methods have been proposed [4]–[10]. Some of them have the capability of detecting damaged blocks before recovering them [7] while the others must be supported by an appropriate transform format and/or an error detection algorithm that helps to identify damaged blocks [6], [8], [9]. Some of them are developed for DCT-based image coding methods [7], [10] while some others can be applied to any block-based approaches [5], [8], [9]. Some of them only deal with still images [8], [9], while the others can be used on image sequences or on both still images and image sequences [4], [7]. In this correspondence, we introduce a new error concealment algorithm that

1) assumes we know which information is received correctly and which is not;
2) can be applied to any block-based image coding system, i.e., independent of block encoding approach;
3) only recovers lost blocks for gray scale still images.

Although there are many variations in previously published methods, they can be categorized into one framework. First, all of them used the information of only neighboring pixels as the source to generate pixel values of the missing blocks. Second, they rely on some predefined constraints on the spectra or structures of the lost blocks. The reconstructed blocks should be smoothly connected with adjacent regions either in spatial or in transform domain. Consequently, a well-recovered lost block is often a smooth block, a sharp edge block, or a stripe block with very consistent directions. The concealment algorithm is a lowpass filter or directionally lowpass filter in nature.

The technique proposed in this correspondence is fundamentally different from the previous framework. It is developed by making use of a special kind of *a priori* information—blockwise similarity within the image. By taking advantage of such kind of information redundancy, the fractal block coding (FBC) technique, a very promising technique for high compression ratios, was developed [11], [12]. The FBC algorithm introduced a practical way to find blockwise self-similarities within natural images. In this correspondence, we will use a similar way to find blockwise similarities. However, our algorithm can not be called *fractal* because the word "fractal" in "fractal image compression" means self-similarity at every different scale while we are trying to utilize blockwise similarities at the same scale.

II. Best Neighborhood Matching (BNM) Algorithm

Let x_{ij} and x_{ij}^{new} represent the pixel values at position (i,j) in the damaged image and the restored image, respectively. Since we know which blocks in the image are received correctly and which are not, it is easy to give each pixel a binary flag f_{ij} indicating whether it is lost, i.e., $f_{ij} = 1$ means (i,j) is missing and $f_{ij} = 0$ means it is good. The goal of our error concealment algorithm is to give each missing pixel with $f_{ij} = 1$ a new value that can comply with our visual sense. In this section, we first introduce a class of blockwise luminance transformations. The BNM algorithm is then developed by making use of such kind of transformations.