MULTIPLE DESCRIPTION IMAGE CODER USING CORRELATING TRANSFORMS

Gianmarco Romano

Dipartimento di Ingegneria dell'Informazione,
Seconda Università di Napoli,
via Roma, 29, 81031 Aversa (CE), Italy.
Email: gianmarco.romano@unina2.it

Pierluigi Salvo Rossi

Dipartimento di Informatica e Sistemistica,
Università di Napoli "Federico II",
via Claudio, 21, 80125 Napoli, Italy.
Email: salvoros@unina.it

Francesco Palmieri

Dipartimento di Ingegneria dell'Informazione,
Seconda Università di Napoli,
via Roma, 29, 81031 Aversa (CE), Italy.
Email: francesco.palmieri@unina2.it

ABSTRACT

We propose a multiple description image coder which uses a linear transform to generate different correlated descriptions. This coding technique is useful when transform coefficients have to be transmitted over lossy channels, such as packet networks, since statistical dependencies among descriptions help the receiver to better estimate lost coefficients. The correlating transform is determined from a gradient search and assumes knowledge of the channel loss statistics as described in [1]. Simulation results show a qualitatively and quantitatively considerable performance increase over a standard DCT based image encoder.

1. INTRODUCTION

Multiple Description Coding (MDC) is a technique to encode real-time multimedia sources when transmitted over lossy channels such as packet networks. An MDC source encoder provides different representations of the source such that, when all descriptions are available at the receiver, an high-quality reconstruction of the source is possible, while if only a small number of descriptions is available, a reconstruction is still possible even though a lower quality is obtained. MDC is suitable in all those situations in which small transmission delays are important and long codes are not practical. Furthermore having more descriptions makes sense when channels are not reliable, such as best effort packet networks, and transmitted source signals allow different quality of reconstruction (distortions), as in the case of multimedia signals. It is the application to multimedia real-time transmission over packet networks that recently increased interest in MDC [1–6].

Transform coding may be used as a technique to generate different correlated descriptions. The statistical dependencies among descriptions helps the receiver to better estimate lost descriptions. This technique was first proposed by Wang *et al.* [4] and further developed by Goyal and Kovačević [3].

In this article a new compression scheme for images is proposed as a modification of widely adopted methods based on *Discrete Cosine Transform* (DCT). A correlating transform is applied to DCT coefficients in order to robustify transmission. The correlating transform used is determined with a numerical algorithm, [1,7], that relies on a gradient search and allows to compute an optimal transform applicable to any number of descriptions and any number of coefficients. Finally results of qualitatively and quantitatively improvements of reconstructed images over standard schemes are shown.

2. TRANSFORM CODING WITH ERASURES

Let's start considering the coding-decoding cascade model shown in Figure 1. With respect to the traditional transform coding we include an erasure mechanism that randomly cancels some of coefficients. The source is as an N-dimensional random vector $\mathbf{y} = (y_1, y_2, \ldots, y_N)^T$ and the $N \times N$ matrix $\mathbf{T}$ is the transform matrix. The N transform coefficients $\mathbf{z} = (z_1, \ldots, z_N)^T$, where $\mathbf{z} = \mathbf{T}^T\mathbf{y}$, are quantized, usually with N different scalar quantizers $\mathbf{z}_q = Q(\mathbf{z})$, and sent over the erasure channel. Say that N_e coefficient are lost and therefore only $N - N_e$ survivor coefficients are available. The receiver, that knows which coefficients have been erased[1], provides a linear reconstruction of $\mathbf{y}$ through a Wiener filter

$$\mathbf{C}_{ro}(\mathbf{e}) = E[\mathbf{z}_{qe}\mathbf{z}_{qe}^T]^{-1}E[\mathbf{z}_{qe}\mathbf{y}^T]$$

which minimizes the mean square error $\mathscr{E}(\mathbf{e}) = \frac{1}{N}E[\|\mathbf{y} - \mathbf{y}_r\|^2]$.

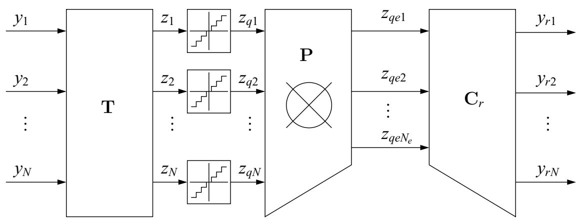

Figure 1: Transform coding with erasures

At each channel use, the erasures can be described by a random binary vector $\mathbf{e} = (e_1, e_2, ..., e_M)^T$, with $e_i = 0$, if the i-th component is erased and $e_i = 1$ otherwise. A compact description of the erasure process can be done by defining the residual vector $\mathbf{z}_e$ containing the $N - N_e$ survivor components kept in the same order and by defining an $(N - N_e) \times N$ matrix $\mathbf{P}(\mathbf{e})$ such that

$$\mathbf{z}_{qe} = \mathbf{P}(\mathbf{e})\mathbf{z}_q.$$

The structure of $\mathbf{P}(\mathbf{e})$ models the type of erasures that can happen on the channel: for example $\mathbf{P}(\mathbf{e})$ may have a block structure to model packet-wise (group) losses.

The minimum error for each error configuration $\mathbf{e}$ is

$$\mathscr{E}_o(\mathbf{e}) \quad = \quad \frac{1}{N}\mathrm{tr}\{\mathbf{R}_y - E[\mathbf{y}\mathbf{z}_{qe}^T]E[\mathbf{z}_{qe}\mathbf{z}_{qe}^T]^{-1}E[\mathbf{z}_{qe}\mathbf{y}^T]\}.$$

Assuming that the receiver acts optimally on each "erased" vector, the total mean squared error averaged over all erasure events is

$$\mathscr{E}_{to} = \frac{1}{N}\sum_{\mathbf{e}} \Pr\{\mathbf{e}\}\mathscr{E}_o(\mathbf{e}). \tag{1}$$

[1]In a packet communication link the receiver knows which packet has lost via a progressive number of the packets (such as in RTP).

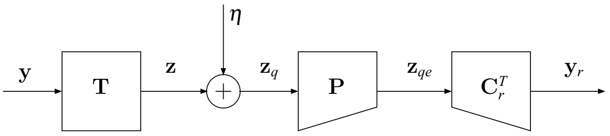

Figure 2: General scheme

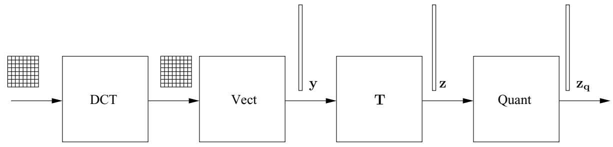

Figure 3: Forward Transform

In standard transform coding, KLT is known to be optimal in terms of rate-distortion performance, but in the presence of the erasures optimality is lost and a better transform can be determined. Our goal here is to search for a linear transform that minimizes $\mathscr{E}_{to}$ assuming that channels statistics are known. In looking for the best transform $\mathbf{T}$, we refer to the model shown in Fig. 2. Under fine quantization assumption, quantization is modeled as an additive noise η with zero mean and uncorrelated with the input signal $\mathbf{z}$ [8] . The correlation matrix of the quantization noise η is

$$\mathbf{R}_\eta = \mathbf{T}\mathbf{R}_y\mathbf{T}^T \odot \mathbf{B}, \tag{2}$$

where $\odot$ is the Hadamard product (element-by-element matrix product) and $\mathbf{B} = \mathrm{diag}(\beta_1, \ldots, \beta_N)$ where β_i has the form $a2^{-bN_{bit}}$ where N_{bit} is the number of bit allocated for the ith component and a and b depend on the type of quantizer and on the distribution of z_i [9, tab 4.4].

When erasures are taken into account, the role of the transform changes. KLT decorrelates input coefficients and its optimality relies on this property, but here the goal is to produce correlated coefficient so that lost coefficients can be better estimated at the receiver. Matrix $\mathbf{T}$ is therefore a *correlating transform* that has the role to protect transform coefficients against losses, increases the number of degrees of freedom in the coder design and could be considered a sort of "pre-emphasis" filter bank.

The system performance, averaged over all possible erasure events, becomes

$$\begin{aligned}\mathscr{E}_{to}(\mathbf{e}) \quad = \quad & \frac{1}{N}\mathrm{tr}(\mathbf{R}_y) \\ & -\frac{1}{N}E_{\mathbf{e}}[\mathrm{tr}(\mathbf{R}_y\mathbf{T}^T\mathbf{P}^T(\mathbf{P}(\mathbf{T}\mathbf{R}_y\mathbf{T}^T \\ & +\mathbf{R}_\eta)\mathbf{P})^{-1}\mathbf{P}\mathbf{T}\mathbf{R}_y)].\end{aligned}$$

The problem of non trivial optimal choice for $\mathbf{T}$ is

$$\begin{cases} \mathbf{T}_o & = & \mathrm{argmax}_{\mathbf{T}}\phi(\mathbf{T}) \\ \phi(\mathbf{T}) & = & E_{\mathbf{e}}[\mathrm{tr}(\mathbf{R}_y\mathbf{T}^T\mathbf{P}^T(\mathbf{P}(\mathbf{T}\mathbf{R}_y\mathbf{T}^T \\ & & +\mathbf{R}_\eta)\mathbf{P})^{-1}\mathbf{P}\mathbf{T}\mathbf{R}_y)] \end{cases}$$

where $\phi(\mathbf{T})$ is the optimization cost function (average distortion over all possible loss configurations).

The search for the optimal matrix $\mathbf{T}$ is based on a gradient ascent algorithm which uses the recursion

$$\mathbf{T}(n) = \mathbf{T}(n-1) + \mu\nabla_{\mathbf{T}}\phi(\mathbf{T}(n-1)), \tag{3}$$

where $\nabla_{\mathbf{T}}\phi$ is the gradient matrix (gradient flow) of ϕ with respect to $\mathbf{T}$ and μ is a scalar parameter. The gradient is computed by using techniques from matrix differential calculus [10] and a brief outline of derivation is reported in the appendix. It results to be

$$\nabla_{\mathbf{T}}\phi = 2E_{\mathbf{e}}[\mathbf{W}^T\mathbf{R}_y - \mathbf{W}^T\mathbf{W}\mathbf{T}\mathbf{R}_y - ((\mathbf{W}^T\mathbf{W}) \odot \mathbf{B})\mathbf{T}\mathbf{R}_y], \tag{4}$$

where

$$\mathbf{W} = \mathbf{R}_y\mathbf{T}^T\mathbf{P}^T(\mathbf{P}(\mathbf{T}\mathbf{R}_y\mathbf{T}^T + \mathbf{R}_\eta)\mathbf{P})^{-1}\mathbf{P}.$$

Application of the gradient ascent algorithm requires also knowledge of the correlation matrix of the input signal, $\mathbf{R}_y$, and the loss probability of the channel, $\mathbf{P}(\mathbf{e})$. At each iteration the gradient $\nabla\phi$ is evaluated, according to eq. (4), and a new matrix $\mathbf{T}$ is evaluated from eq. (3) where μ is a scalar constant parameter. Iterations stop when $\|\nabla\phi\|_2 < \varepsilon$, with ε in the order of 10^{-3}. Evaluation of the quantization noise correlation matrix $\mathbf{R}_\eta$ is required at each iteration since $\mathbf{R}_\eta$ depends on matrix $\mathbf{B}$ which in turn depends on the number of bit allocated to each coefficient. We maintain fixed the total number of bits to be allocated and therefore the bit-rate at the output of the quantizers. With such a budget we use a standard integer optimal bit allocation algorithm [9] to distribute bits among coefficients which depends on variances of coefficients after the correlating transform. Therefore allocation of bits has to be repeated at each iteration to reflect the fact that those variances change at each step. The search for the optimal matrix $\mathbf{T}$ requires a double iteration, i.e. at each step in which a new $\mathbf{T}$ is computed, an iteration to redistribute the total number of bits among coefficients is required and a new $\mathbf{R}_\eta$ is evaluated. Again the search for the optimal correlating transform is performed at constant bit-rate and therefore side and central distortions change to provide an overall better performance in the sense of the cost function (1).

The gradient expression requires evaluation of the average over all the possible erasure events and needs to know, or estimate, the probability of each one of them. This is exponentially complex if the number of loss configurations is large. In our application it is likely that the coefficient set $\mathbf{z}_{qe}$ are divided into a small number of subgroups (packets, in our experiments only three), that exhibit a manageable number of loss configurations.

3. COMPRESSION SCHEME

Our image encoder is based on DCT transform which is widely adopted transform in image compression such as JPEG [11]. The forward transform scheme is shown is Fig 2. Each image is divided in square blocks of dimension 8×8 and all samples of the block are dc-level shifted from unsigned integers with range $[0, 2^p - 1]$ to signed integers with range $[-2^{p-1}, 2^{p-1} - 1]$ ($p = 8$). Each block is then transform coded independently and DCT coefficients are rearranged in a one-dimensional vector according a zig-zag pattern that has the advantage of placing low frequency coefficients in first positions. Since DCT transform has the property to pack considerable amount of image energy in the first coefficients, compression can be achieved selecting a subset of first N transform coefficients. Our choice of N is based on the resulting PSNR (*peak signal-to-noise ratio*). We choose $N = 30$ because, with the quantization scheme described in the following, a PSNR of about 30dB is obtained, a threshold value for an image of visually acceptable quality.

A correlating transform $\mathbf{T}$ is then applied to the N survivor DCT coefficients, i.e. to the vector $\mathbf{y}$. With such a scheme, correlation matrix is applied to $N = 30$ coefficients, rather than to 64 coefficient, and therefore a matrix of size 30×30 instead of 64×64 has to be found. Matrix $\mathbf{T}$ is determined by using a gradient ascent algorithm described in the previous section. Note that gradient evaluation assumes that autocorrelation matrix $\mathbf{R}_y$ and the number of descriptions are known. The former is estimated from real data, i.e. from DCT coefficients of test images, and the latter is chosen to be equal to 3, a practical value that reduces the average over all the possible erasure events to a manageable number of loss configurations.

Prior transmission, each coefficient is quantized with a different uniform scalar quantizer which use optimal allocated number of bits that results from evaluation of optimal $\mathbf{T}$. Finally, quantized

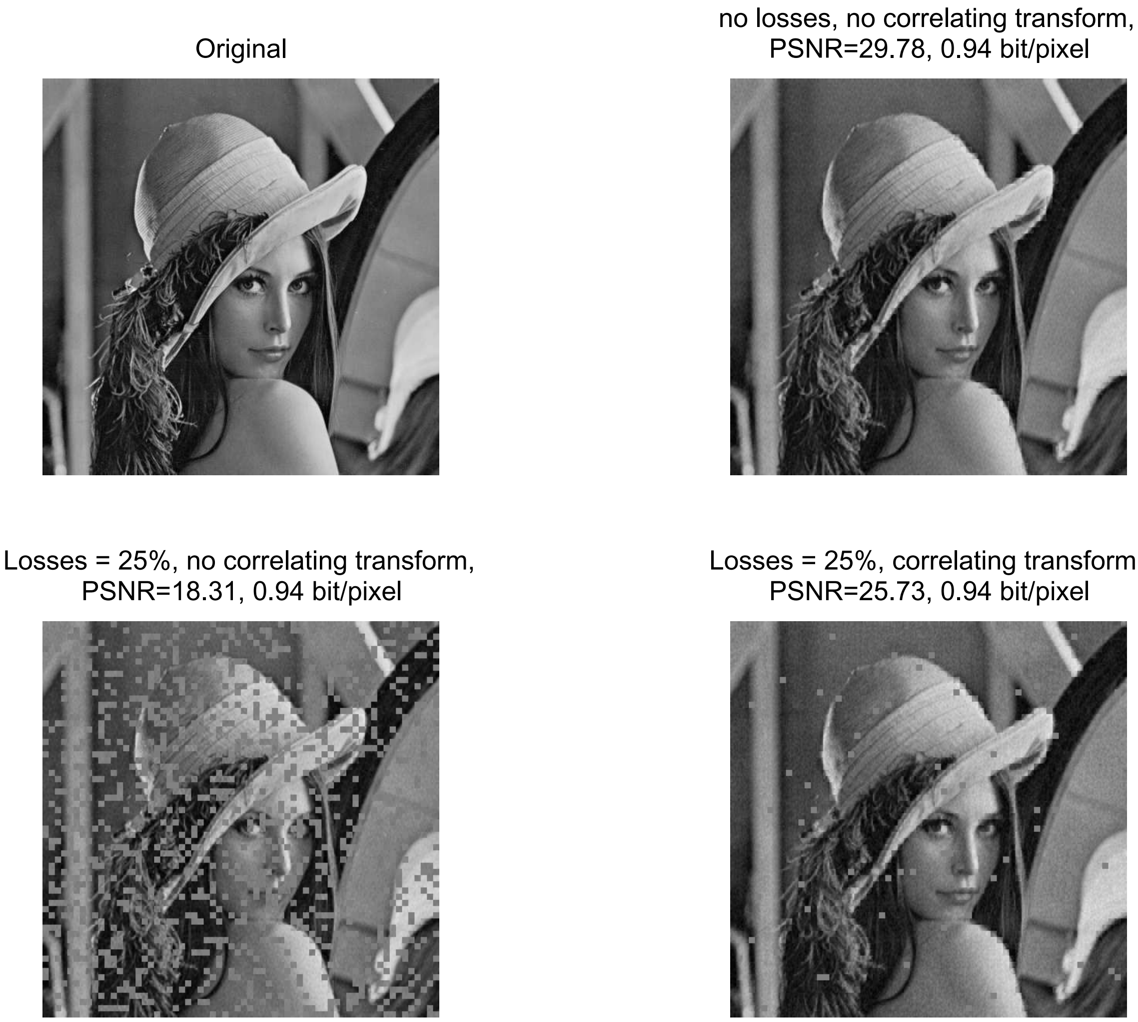

Figure 4: Lena

coefficients are partitioned into the chosen number of descriptions to be transmitted independently over erasure channels.

Scalar uniform quantization is clearly a non optimum choice here because of the presence of correlation among coefficients and because in general the correlating transform might be non orthogonal. As already pointed out in [12] when a non orthogonal linear transform is used and scalar quantization of transform coefficients applied a larger mean-squared error is obtained for a given rate because quantizer cells in signal space become non cubic. In order to reduce the resulting MSE both Orchard *et al.* and Goyal and Kovačević have chosen to quantize before the correlating transform is applied. This has the advantage to maintain quantizer cells cubic and eventually to apply quantization to uncorrelated coefficient. The correlating transform has to be an integer-to-integer (I2I) transform and an error is introduced when approximating the I2I transform with the linear continuous transform determined by the analysis performed by Orchard and Goyal. An high rate analysis of central distortion shows that an I2I transform does not change central distortion which depends only on quantization (i.e. does not depend on transform). Therefore one can maintain to a fixed relatively small value the central distortion and try to minimize side distortion searching for an optimal correlation transform with a constraint on the rate.

Here we follow a different approach and maintain scalar quantization after the correlating transform. This give us the possibility to control through the correlating transform both central and side distortions and obtain an optimal tradeoff given the channel model. Therefore if we transmit over a very lossy channel, on which the probability of getting all descriptions together is small, we prefer to reduce central distortion and increase the side distortions maintaining fixed the overall rate, trying to make a better use of the bits. This is the meaning of the cost function (1). Therefore we consider scalar quantization a suboptimal choice which maintains simplicity because our algorithm optimize also respect to quantization noise since variances change at each iteration. In order to get better results one should use vector quantization, but the resulting coder would be too complex.

At the receiver reverse operations are needed. A Wiener filter provides the best MSE reconstruction of vector $\mathbf{y}$, i.e. survivor DCT coefficients after compression, and all reconstructed samples are dc level shifted to unsigned integers. A reverse vectorization operation provides reconstructed 8×8 image blocks.

Our encoder is still based on a linear transform that consists of a cascade of DCT and the correlating transform $\mathbf{T}$ in the forward path and a cascade of Wiener filter and inverse DCT in the reverse path. At the receiver reconstruction filter has size $N \times (N - N_e)$ and the resulting vector has size $N = 30$ which is then used to provide a block of 8×8 to the inverse DCT.

4. SIMULATIONS

We apply our MD encoder to a set of gray images and transmission of different descriptions on an erasure channel is simulated. We report also results of a comparison with a coding scheme without correlating transform. Results of reconstructed image are shown in Fig. 4 and show the visual improvement on reconstructed image that corresponds to a significant increase in PSNR. More specifically, in Fig. 4 the upper-left corner shows the original (not compressed) image. On the right a DCT compressed version of the same image is shown for comparison. This is compressed with a number of DCT coefficients to be transmitted equal to 30. Each coefficient is quantized by means of an uniform scalar quantizer and each one of them uses an optimally allocated number of bits. We used an integer bit allocation algorithm which provides the best distribution of bits based on distribution of variances of each coefficient.

In the bottom row we show reconstructed images after transmission over an erasure channel. In lower left corner the reconstructed image without use of correlating transform shows the disruptive effect of cancellations that causes a decrease of more 10dB in PSNR. On lower right corner the reconstructed image with application of the correlating transform $\mathbf{T}$ shows the improvement of about 7dB in PSNR and how our scheme mitigates the effect of lost descriptions.

In our simulations we fix the total number of bits available for transform coefficients to be transmitted. In other words we maintain constant the value of the number of bit per pixel, i.e. the rate, and use our degrees of freedom with the choice of the correlating transform to change central and side distortions to get better overall system performance. The value of 0.94 bits/pixel is obtained with a budget of 60 bits to be allocated for $N = 30$ coefficients.

5. CONCLUSIONS

When a signal is transmitted over a lossy channels the effect of erasures may be disruptive at the receiver if some countermeasures are not adopted. The scheme proposed here copes with losses mitigating the effect of erasures in such a way that a acceptable quality of reconstruction is still possible. This goal is achieved through a simple modification to standard transform coding applied to images, such as DCT, and simulations show considerable improvements (about 7dB in PSNR) under heavy losses condition. These results also confirm our theoretical simulation performed in earlier works [1].

Further work is being carried out to find better algorithm to determine optimal correlating transform. We are considering more realistic and accurate channel models in order to improve robustness to source-channel changes. Concerns are also about the structure of the correlating transform to reduce the number of free parameters to optimize and to allow more efficient implementation of coding and decoding operations.

A. DERIVATION OF THE GRADIENT

The derivation of the gradient (4) is based on technique of differential matrix calculus which means that all derivatives are not calculated element by element, but respect to the whole matrices. Derivatives, as suggested in [10], are computed from differentials, therefore we first derive the differential of the term inside the expectation in the cost function $\phi(\mathbf{T})$

$$\mathrm{d}(\mathrm{tr}(\mathbf{R}_y\mathbf{T}^T\mathbf{P}^T(\mathbf{P}(\mathbf{T}\mathbf{R}_y\mathbf{T}^T+\mathbf{R}_\eta)\mathbf{P})^{-1}\mathbf{P}\mathbf{T}\mathbf{R}_y)). \tag{5}$$

respect to the matrix $\mathbf{T}$. Note the there is also an implicit dependence from $\mathbf{T}$ through $\mathbf{R}_\eta$ according to eq. (2). Manipulation of (5) is performed using the following properties

- $\mathrm{d}(\mathrm{tr}(\mathbf{CD})) = \mathrm{tr}(\mathrm{d}\mathbf{C}\mathbf{D} + \mathbf{C}\mathrm{d}\mathbf{D})$,
- $\mathrm{d}(\mathbf{F}\mathbf{G}\mathbf{F}^T) = \mathbf{F}\mathrm{d}\mathbf{G}\mathbf{F}^T$, where $\mathbf{F}$ is a constant matrix [10, p. 174, eq. (5)],
- $\mathrm{d}(\mathbf{G}^{-1}) = -\mathbf{G}^{-1}\mathrm{d}\mathbf{G}\mathbf{G}^{-1}$ [10, p. 183, eq. (17)],
- $\mathrm{tr}(\mathbf{A}^T(\mathbf{B}\odot\mathbf{C})) = \mathrm{tr}((\mathbf{A}^T\odot\mathbf{B}^T)\mathbf{C}) = \mathrm{tr}((\mathbf{A}^T\odot\mathbf{C}^T)\mathbf{B})$, [10, p. 46, Th. 7].

We obtain

$$\begin{aligned}\mathrm{d}(\cdot) &= 2\mathrm{tr}(\mathbf{R}_y\mathbf{W}\mathrm{d}\mathbf{T} - \mathbf{R}_y\mathbf{T}^T\mathbf{W}^T\mathbf{W}\mathrm{d}\mathbf{T})\\ &\quad -\mathrm{tr}(((\mathbf{W}^T\mathbf{W})\odot\mathbf{B})(\mathrm{d}\mathbf{T}\mathbf{R}_y\mathbf{T}^T)\\ &\quad +((\mathbf{W}^T\mathbf{W})\odot\mathbf{B})(\mathbf{T}\mathbf{R}_y(\mathrm{d}\mathbf{T})^T)\\ &= 2\mathrm{tr}(\mathbf{R}_y\mathbf{W}\mathrm{d}\mathbf{T} - \mathbf{R}_y\mathbf{T}^T\mathbf{W}^T\mathbf{W}\mathrm{d}\mathbf{T}\\ &\quad -\mathbf{R}_y\mathbf{T}^T((\mathbf{W}^T\mathbf{W})\odot\mathbf{B})\mathrm{d}\mathbf{T}),\end{aligned}$$

and using the identification table [10, p. 176, Tab. 2] and observing that

$$\mathrm{d}(\cdot) = \mathrm{tr}(\mathbf{C}\mathrm{d}\mathbf{T}) = (\mathrm{vect}\mathbf{C}^T)^T\mathrm{d}(\mathrm{vect}\mathbf{T}),$$

we get

$$\frac{\partial\phi}{\partial\mathbf{T}} = \mathbf{C}^T,$$

which leads to the gradient expression (4).

REFERENCES

[1] G. Romano and F. Palmieri, "An algorithm for transform coding on lossy packet networks," in *Proc. of Conference on Information Sciences and Systems*. The Johns Hopkins University, Mar. 12–14, 2003.

[2] V. Goyal, J. Kovačević, and J. Kelner, "Quantized frame expansions with erasures," *Journal of Appl. and Comput. Harmonic Analysis*, vol. 10, no. 3, pp. 203–233, May 2001.

[3] V.K. Goyal and J. Kovačević, "Generalized multiple description coding with correlating transforms," *Information Theory, IEEE Transactions on*, vol. 47, no. 6, pp. 2199–2224, Sep 2001.

[4] Yao Wang, M.T. Orchard, V. Vaishampayan, and A.R. Reibman, "Multiple description coding using pairwise correlating transforms," *Image Processing, IEEE Transactions on*, vol. 10, no. 3, pp. 351–366, Mar 2001.

[5] V.A. Vaishampayan, "Design of multiple description scalar quantizers," *Information Theory, IEEE Transactions on*, vol. 39, no. 3, pp. 821–834, May 1993.

[6] V.K. Goyal, "Multiple description coding: compression meets the network," *IEEE Signal Processing Magazine*, vol. 18, no. 5, pp. 74–93, Sep 2001.

[7] G. Romano and F. Palmieri, "Numerical evaluation of optimal correlating transforms for erasure channels," *in preparation*, 2003.

[8] B. Widrow, I. Kollar, and Ming-Chang Liu, "Statistical theory of quantization," *Instrumentation and Measurement, IEEE Transactions on*, vol. 45, no. 2, pp. 353–361, Apr 1996.

[9] A.K. Jain, *Fundamentals of Digital Image Processing*, Prentice Hall, 1989.

[10] J.R. Magnus and H. Neudecker, *Matrix Differential Calculus with Applications in Statistics and Econometrics*, John Wiley, 1988.

[11] G. K. Wallace, "The JPEG still picture compression standard," *Consumer Electronics, IEEE Transactions on*, vol. 38, no. 1, pp. xviii–xxxiv, Feb 1992.

[12] M.T. Orchard, Y. Wang, V. Vaishampayan, and A.R Reibman, "Redundancy rate-distortion analysis of multiple description coding using pairwise correlating transforms," in *Image Processing, 1997. Proceedings., International Conference on*, Oct. 26–29, 1997, vol. 1, pp. 608–611.