http://www.zju.edu.cn/jzus; E-mail: jzus@zju.edu.cn
ISSN 1009-3095 *Journal of Zhejiang University SCIENCE* V.5, No.1, P.55-61, Jan., 2004

Multiple descriptions based wavelet image coding

CHEN Hai-lin(陈海林)†, YANG Yu-hang(杨宇航)
(*Modern Communication Institute, Shanghai Jiaotong University, Shanghai 200030, China*)
†E-mail: heavyseatree@163.com

Received Oct.18, 2002; revision accepted Jan. 16, 2003

Abstract: We present a simple and efficient scheme that combines multiple descriptions coding with wavelet transform under JPEG2000 image coding architecture. To reduce packet losses, controlled amounts of redundancy are added to the wavelet transform coefficients to produce multiple descriptions of wavelet coefficients during the compression process to produce multiple descriptions bit-stream of a compressed image. Even if a receiver gets only parts of descriptions (other descriptions being lost), it can still reconstruct image with acceptable quality. Specifically, the scheme uses not only high-performance wavelet transform to improve compression efficiency, but also multiple descriptions technique to enhance the robustness of the compressed image that is transmitted through unreliable network channels.

Key words: Multiple descriptions, Discrete wavelet transform (DWT), JPEG2000
Document: A **CLC number:** TN919.81

INTRODUCTION

JPEG2000 (Lawson and Zhu, 2002) is a newly published still image-coding international standard. Its introduction meant that for the first time, the discrete wavelet transform (DWT) would be used for the decomposition and reconstruction of images together with an efficient coding scheme. In the JPEG2000 coder, before any wavelet decomposition is performed, the image is partitioned into non-overlapping tiles. These tiles are of equal size except possibly for those adjacent to the image boundary. Then wavelet transform is applied to transform the image into a multi-resolution representation on a tile-by-tile basis. After transformation, a uniform quantization matrix is applied, which is performed within each subband with different levels of quantization. After quantization, quantized wavelet coefficients are entropy encoded into bit-stream, packed, and transmitted to networks. Fig. 1 shows the block diagram of the JPEG2000 coding and decoding process.

Current JPEG2000 coding system typically generates bitstream in a progressive mode for network delivery purpose. Generally, low-resolution coefficients of multi-resolution representation within tiles are gathered, entropy encoded and filled into data packets, which are transmitted through networks first. Then the high-resolution coefficients are encoded, packed and transmitted following them.

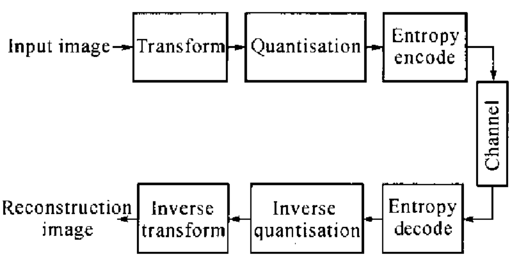

Fig.1 The JPEG2000 coding and decoding process

Most networks today are based on an exchange of packets. UDP is a main protocol that controls the transmission of the packets. In some cases, such as congestion and transmission errors, the packets will be possibly dropped at random and the networks will not take any action on this situation. We know that once packets containing low-resolution coefficients are lost, a successful reconstruction cannot be achieved at the receiver end at all even if the following packets containing the high-resolution coefficients are received safely. To guarantee receiving, some applications use TCP protocol to control the transmission of the packets. Unfortunately, also in congestion and error cases, TCP-based delivery may suffer from stalls because the delay in receiving retransmitted packets may be much longer than the normal inter-arrival time between received packets. Furthermore, that TCP transmission requires a feedback channel is not always true.

It seems that present networks cannot provide a satisfactory data delivery solution. We have to devise a new way to provide high efficiency and robust data delivery service.

Joint source-channel coding is a useful solution. But due to its high cost and uncertainty of network conditions, it is also unsuitable for application. We have to turn to image source coding itself to solve the problem. It is known that packets erasure networks are often modeled through generalized multiple descriptions coding (MDC) (Wang *et al.*, 2000), in which the encoder produces multiple descriptions of the source data, each exactly filling certain packets. Using this concept, we developed an algorithm that adds controlled amounts of redundancy to the image wavelet transform coefficients to produce multiple descriptions of wavelet transform coefficients. Then these multiple descriptions transformed coefficients are regrouped, encoded, packed and delivered independently through diverse network channels. If all packets from all channels are received, a high-quality reconstruction can be achieved, while a lower but acceptable quality can be achieved from a subset of the packets. Obviously, the advantages of MDC are:

1. Robustness against loses: even if a receiver gets only parts of descriptions (other descriptions being lost), it can still reconstruct an image with acceptable quality;

2. Enhanced quality: if a receiver gets more descriptions, it can combine them together to produce a better reconstruction than that produced from their subset.

Fig.2 can clarify the above description. $X(k)$ is the input source, which is encoded as Description 1 and Description 2. When only Description 1 or Description 2 is received, $\tilde{X}_1(K)$ or $\tilde{X}_2(K)$ can provide low but acceptable reconstructions quality at the receiver end. When both Description 1 and Description 2 are received. $\tilde{X}_0(K)$ provides better reconstruction. Compared to nothing we can get in conventional systems, the proposed scheme is a better solution obviously.

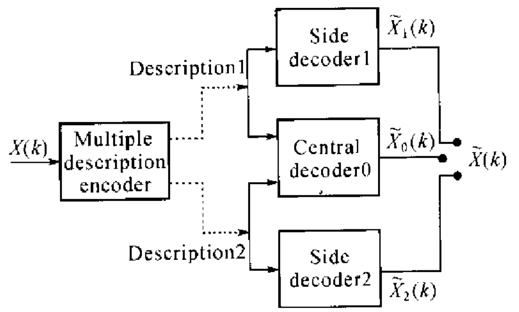

Fig.2 Multiple descriptions coding structure

MULTIPLE DESCRIPTION CODING

Multiple descriptions coding has been used before in the context of robust image coding. The first MD image coder that we are aware of was proposed by Vaishampayan and consisted of an extension of a JPEG coder using MD scalar quantizers (Vaishampayan, 1996). Wang *et al.* proposed another MD extension of a JPEG coder, using a class of pairwise correlating transforms to create the MD's (Wang *et al.*, 1997). In these documentations, MD scalar quantization and MD correlating transform were two typical MDC techniques. But it seemed that the MDC was conducted only in DCT domain, not considering wavelet domain. Sergio (Servetto *et al.*, 1998) studied MD scalar quantizers in wavelet domain, but did not consider correlating transforms in wavelet domain. So in this context, we mainly study the multiple descriptions correlating transform coding in wavelet domain, which is a blank in the research area at present.

Multiple descriptions correlating transform coding

In our proposed scheme, after wavelet transformation and quantization, each block of n nearly independent, zero-mean wavelet transform coefficients with different variances is transformed into a block of n transformed coefficients with an added statistical correlation between them. Then, the transformed coefficients are regrouped, encoded and distributed to different packets. So in the case of packet loses, the lost coefficients can be estimated from the received coefficients. The redundancy comes from the relative inefficiency of scalar entropy coding of correlated coefficients. For simplicity of analysis, we pick up only two coefficients with different variances to have a pairwise correlating transform at a time and produce two descriptions bitstreams in the end.

According to Wang's (Wang *et al.*, 1997) method, the MD character is achieved with a linear correlating transform that introduces statistical correlation between a pair of random variables. The pair of random variables $\boldsymbol{x} = [x_1, x_2]$ are assumed to be independent, zero-mean Gaussian components with variances σ_1^2 and σ_2^2 respectively. Without loss of generality, we assume $\sigma_1^2 \geqslant \sigma_2^2$. The pairwise correlating transform $\boldsymbol{T}$ produces descriptions y_1 and y_2 through $\boldsymbol{y} = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \boldsymbol{T}(x)$, where,

$$\boldsymbol{T} = \begin{bmatrix} a & b \\ c & d \end{bmatrix}, \text{ with det } \boldsymbol{T} = 1. \tag{1}$$

Then,

$$\begin{aligned} y_1 &= (Tx)_1 = ax_1 + bx_2 \\ y_2 &= (Tx)_2 = cx_1 + dx_2 \end{aligned} \tag{2}$$

After correlating transform, some statistical correlation is added to the two random variables, i.e. y_1 and y_2 have some certain statistical correlation between them. We can estimate one variable of y_1 or y_2 from another using the statistical correlations between them. Suppose y_1 is received and y_2 is lost, to minimize the MSE (mean-square error) distortion, the reconstruction at the first side decoder is $\hat{\boldsymbol{x}}^{(1)} = E[x/y_1]$. Since $\boldsymbol{x}$ and y_1 are jointly Gaussian, the conditional variable $\boldsymbol{x}/y_1$ is also Gaussian and computations of $\hat{\boldsymbol{x}}^{(1)}$ is simple. Noting that $[y_1, y_2]^{\mathrm{T}}$ is a linear function of $\boldsymbol{x}$, then

$$\hat{\boldsymbol{x}}^{(1)} = \boldsymbol{T}^{-1}\begin{bmatrix} y_1 \\ \hat{y}_2 \end{bmatrix}, \text{ where}$$

$$\hat{y}_2 = E[y_2/y_1] = \frac{E[y_1 y_2]}{E[y_1^2]} y_1 = \frac{ac\sigma_1^2 + bd\sigma_2^2}{a^2\sigma_1^2 + b^2\sigma_2^2} y_1. \tag{3}$$

Similarly, if y_1 is lost and y_2 is received, the reconstruction at the second side decoder is:

$$\hat{\boldsymbol{x}}^{(2)} = \boldsymbol{T}^{-1}\begin{bmatrix} \hat{y}_1 \\ y_2 \end{bmatrix}, \text{ where}$$

$$\hat{y}_1 = E[y_1/y_2] = \frac{E[y_1 y_2]}{E[y_2^2]} y_2 = \frac{ac\sigma_1^2 - bd\sigma_2^2}{c^2\sigma_1^2 + d^2\sigma_2^2} y_2. \tag{4}$$

If the correlating transform is balanced, with the condition of Eq. (1), we can get

$$\boldsymbol{T}_\alpha = \begin{bmatrix} \alpha & (2\alpha)^{-1} \\ -\alpha & (2\alpha)^{-1} \end{bmatrix} \tag{5}$$

In no erasure cases, the reconstruction uses

$$\boldsymbol{T}_\alpha^{-1} = \begin{bmatrix} (2\alpha)^{-1} & -(2\alpha)^{-1} \\ \alpha & \alpha \end{bmatrix} \tag{6}$$

In one description loss case, the reconstruction uses Eq. (3) or Eq. (4), which then becomes the following expressions respectively:

$$\hat{\boldsymbol{x}}^{(1)} = \frac{2\alpha}{4\alpha^4\sigma_1^2 + \sigma_2^2}\begin{bmatrix} 2\alpha^2\sigma_1^2 \\ \sigma_2^2 \end{bmatrix} y_1 \tag{7}$$

$$\hat{\boldsymbol{x}}^{(2)} = \frac{2\alpha}{4\alpha^4\sigma_1^2 + \sigma_2^2}\begin{bmatrix} -2\alpha^2\sigma_1^2 \\ \sigma_2^2 \end{bmatrix} y_2 \tag{8}$$

As α changes, different amounts of redundancy are added to the two source random variables. If the source base-coding rate is R^*, the average rate of output variables is R, then the redundancy is

$$\rho = R - R^*, \text{ where,}$$
$$R = (R_1 + R_2)/2, \ R_1 = H(y_1), \ R_2 = H(y_2). \tag{9}$$

Using Eq. (7) and Eq. (8), we can compute D_1 and D_2, which are the distortions per component in reconstructing from y_1 or y_2 alone. So, the average distortion of reconstruction per component at the receiver end is:

$$D = (D_1 + D_2)/2, \text{ where}$$

$$D_1 = \frac{1}{2}E[\| x - \hat{x}^{(1)} \|^2] \tag{10}$$

$$D_2 = \frac{1}{2}E[\| x - \hat{x}^{(2)} \|^2] \tag{11}$$

Simulation of two Gaussian random variables correlating transform

In order to show the performance of the correlating transform, we create a set of numerical computations. Fig. 3 considers two independent Gaussian random variables with $\sigma_1 = 1$, $\sigma_2 = 0.5$ and mean-values 0. Their base-coding rate is $R = 2$ bits/component. The robustness is measured by D, which is the average of side distortions at the receiver end.

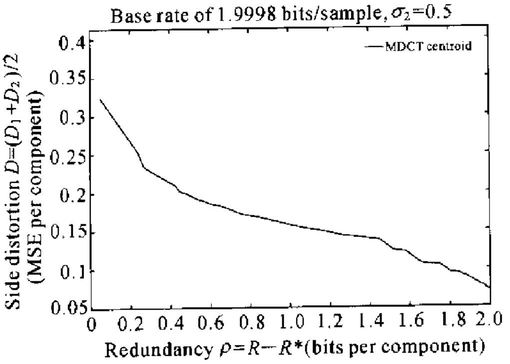

Fig. 3 Redundancy-distortion curve for Gaussian variables MDC

From the above plot, we can see that added small amounts of redundancy ρ, which is from 0 to 2 bits per component, the average of side distortions D diminishes considerably, which means that we can get acceptable reconstruction quality even if only one description is received.

MDC IN WAVELET DOMAIN

The discussion before assumes that random variables have Gaussian distribution, which is a base condition for using correlating transform. So, we have to be sure that the wavelet transform coefficients fit Gaussian distribution.

Under JPEG2000 wavelet image coding architecture, wavelet transform is applied to the original image on a tile-by-tile basis. After wavelet transformation, transform coefficients can be assured to be independent of each other, which is based on the primary properties of the wavelet transform and the fact that the wavelet transform "nearly" decorrelates a wide variety of signals and images. This property, combined with our view of the signal as a random realization from a family of distribution of signals, leads to the following simple model for an individual wavelet transform coefficient. Most wavelet transform coefficients contain very little signal information and hence have small, random values. A few wavelet transform coefficients have large values that represent significant signal information. Thus we can model each wavelet transform coefficient as being in one of two states: "high", corresponding to a wavelet component containing significant dominant contribution of signal energy, or "low", representing components with little signal energy. If we associate with each state a Gaussian probability density – say, a high-variance, zero-mean density for the "high" state and a low-variance, zero-mean density for the "low" state – the result is a Gaussian mixture model for each wavelet transform coefficient. In JPEG2000 wavelet image coding process, after wavelet transformation on tiles, before quantization, DC coefficients, which are "high" state, require difference computation. Then "high" state and "low" state can be combined together and modeled by a simple unified Gaussian distribution. In order to validate our analysis, we give their histograms in Fig. 4 and Fig. 5. From these histograms, we

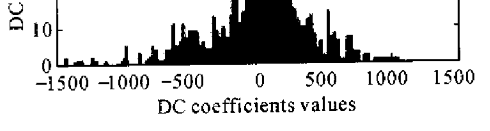

Fig. 4 Histogram of DC coefficients

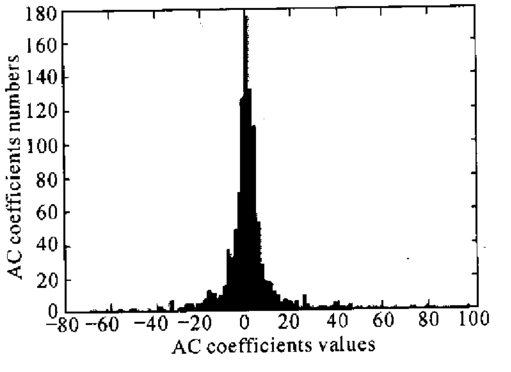

Fig. 5 Histogram of AC coefficients

can see that the difference DC coefficients and AC coefficients fit Gaussian distribution well.

Theoretically, Mallat (1989) proved that the histograms of wavelet transform coefficients of natural images could be modeled by a family of Gaussian distribution:

$$p(x) = Ke^{-(|x|/\alpha)^2} \tag{12}$$

where α models the width of the histogram peak (variance), K is a normalization constant to ensure that $\int p(x)\mathrm{d}x = 1$.

SIMULATIONS AND RESULTS

After we have proved the fitness of nearly Gaussian distribution of wavelet transform coefficients, we can use multiple descriptions correlating transform within wavelet domain in a strong base.

Our simulation is composed of the following steps:

1. Input a test source image and apply wavelet transform on a tile-by-tile basis on it.
2. Within each tile, a difference computation of DC wavelet coefficients is performed.
3. Using these transform coefficients as the samples of random variables, we compute the variance values and convert each coefficient to zero-mean. Then these coefficients can be seen as samples of corresponding variance Gaussian random variables.
4. Then, the coefficients are uniformly quantized at a given base rate.
5. Within each tile of image, we choose the coefficient corresponding to largest variance and the coefficient corresponding to smallest variance as source random variable and apply multiple descriptions correlating transform on the pair.
6. Output coefficients from each description are grouped together and entropy encoded independently. Go to 5th step until all coefficients within tiles are processed.
7. Each description bitstream is packed and sent independently through networks.

We discuss below our simulation results.

We input a $256 \times 256 \times 8$ Lena image and use Sym.5 wavelet base to apply a 3-level wavelet transform on a tile-by-tile basis. The tile size was chosen as 8×8. Fig.6 and Fig.7 give the original Lena image and the wavelet transformed image.

Fig.6 The original Lena image

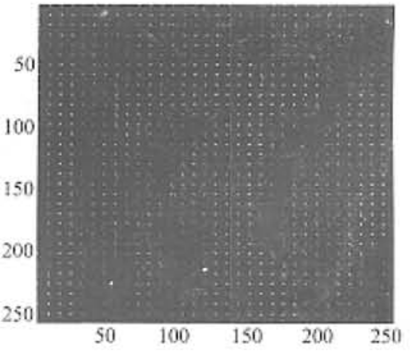

Fig.7 The wavelet transformed image.

After transformation, there is a DC coefficient within each transformed tile. According to the above coding process, DC coefficient difference computation can be performed. Then, each coefficient at a certain position within each tile can be seen as a sample of Gaussian random variable whose expectation mean and variance can be estimated from corresponding samples at the same relative positions within all tiles. To 256×256 image and 8×8 tiles, there are 1024 samples for each tile coefficient position, which can be seen as realizations of corresponding Gaussian random variables.

After DC coefficient difference computation, there are 64 coefficients within each tile. According to the above described step 5, we align the coefficients in decreasing variance order and pick up the largest variance sample and smallest variance sample to apply pairwise correlating transform on them. After transformation, the two output variables are separated as description 1 and description 2. This process repeats until all remaining coefficients within tiles are processed. At last, all output coefficients from the same description are grouped, entropy encoded, packed and transmitted independently through networks.

At the receiver end, if both descriptions are received, we combine them together and use Eq.(6) to give a complete inverse transformation, which provides best reconstruction quality. If one description is lost, we have to use Eq.(3) or Eq.(4) to estimate another description from the received one, using the statistical correlation between them, i.e using Eq.(7) or Eq.(8) to compute partial inverse transformation, which provides low but acceptable reconstruction quality.

For evaluating the coding system performance, we code the image in base rate of 2 bits/sample and add different increasing amounts of redundancy to each pair of wavelet transform coefficients (we choose ρ from 0.5 to 2 per component). At the receiver end, we evaluate its robustness by its average side distortions values D when only one description is received. Fig.8 gives the simulation results.

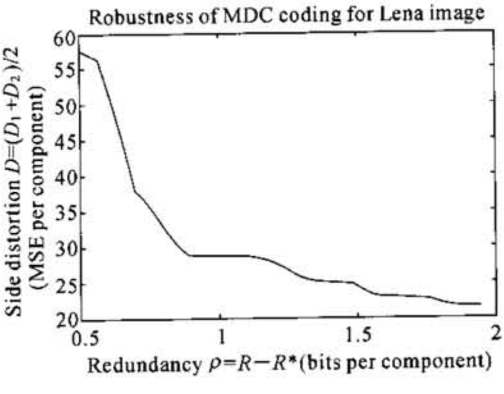

Fig.8 Redundancy-distortion plot for Lena image (at base rate of $R = 2$bpp, $\rho = 0.5 - 2$)

From Fig.8, we can see that given some little amounts of redundancy, MDC coding system provides great improvement of image reconstruction quality, when only one description is received.

Fig.9 and Fig.10 give the image reconstruction from one description with two different redundancies. It is clear that the Lena reconstruction image looks better as redundancy increases.

If two descriptions are received, they are combined and give better reconstruction, as shown in Fig.11.

Fig.9 Image reconstruction from one description of MDC (at base rate $R = 2$bpp, $\rho = 0.6$)

Fig.10 Image reconstruction from one descrip-tion of MDC (at base rate $R = 2$bpp, $\rho = 0.9$)

Fig.11 Image reconstruction from two descrip-tions of MDC (at base rate $R = 2$bpp)

CONCLUSIONS

In this work, we developed a multiple descriptions based wavelet image coding scheme under JPEG2000 image coding architecture. The scheme use not only high-performance wavelet transform to improve compression efficiency, but also multiple descriptions technique to enhance the robustness of the compressed image that is transmitted through unreliable networks. Through some experiments, it was found that even if a receiver gets only one description (other descriptions being lost), it could still reconstruct image with low but acceptable quality.

References

Lawson, S. and Zhu, J., 2002. Image compression using wavelets and JPEG2000: a tutorial. *Electron. Commun. Eng. J*, **6**: 112 – 121.

Mallat, S., 1989. A theory for multiresolution signal decomposition: the wavelet representation. *IEEE Trans. Patt. Anal. Mach.. Intel*, **11**(7): 674 – 693.

Servetto, S. D., Ramchandran, K., Vaishampayan, V. and Nahrstedt, K., 1998. Multiple-Description Wavelet Based Image Coding. *In*: Proc. Int. Conf. Image Processing '98, Chicago, IL.

Vaishampayan, V. A., 1996. Application of Multiple Description Codes to Image and Video Transmission over Lossy Networks. *In*: Proc. 7th Int. Workshop Packet Video, Brisbane, Australia.

Wang, Y., Orchard, M. T. and Reibman, A., 1997. Multiple Description Image Coding for Noisy Channels by Pairing Transform Coefficients. *In*: Proc. 1st Workshop IEEE Signal Processing Soc. Multimedia Signal Processing, Princeton, NJ.

Wang, Y., Orchard, M. T., Vaishampayan, V. and Reibman A. R., 2000. Multiple description coding using pairwise correlating transforms. *IEEE Trans. Image Processing*, **9**: 813 – 825.