Two-Stage Multiple Description Image Coders: Analysis and Comparative Study

Andrey Norkin [a,*], Atanas Gotchev [a], Karen Egiazarian [a], Jaakko Astola [a].

[a] *Institute of Signal Processing, Tampere University of Technology P.O.Box 553, FIN-33101 Tampere, FINLAND*

Abstract

A general two-stage multiple description coding (MDC) scheme using whitening transform is analyzed. It represents the original image in a form of a coarse image approximation and a residual image. The coarse approximation is subsequently duplicated and combined with the residual image further split into two descriptions using a chessboard block transform coefficients rearrangement. We identify the importance of a good coarse approximation and explore different approaches for changing its resolution and coding it. We also propose different approaches for coding the residual signal. The coder scheme is quite simple and yet achieves high performance comparable with other MDC methods.

Key words: Multiple description coding, error resilience, bit allocation.

1 Introduction

In the recent years, multiple description coding (MDC) has taken considerable attention as a method of communication over unreliable channels [1], [2], [3], [4]. MDC is a source-channel coding of information, which can be represented with different levels of quality. The source is encoded into several bitstreams

* Corresponding author.
Email addresses: andrey.norkin@tut.fi (Andrey Norkin), atanas.gotchev@tut.fi (Atanas Gotchev), karen.egiazarian@tut.fi (Karen Egiazarian), jaakko.astola@tut.fi (Jaakko Astola).

[1] This work was supported by Tampere Graduate School in Information Science and Engineering (TISE) and the Academy of Finland, project No. 213462 (Finnish Centre of Excellence program (2006 - 2011)).

(i.e. *multiple descriptions*) to be transmitted via *independent* channels. In the receiver, the source can be reconstructed by any single bitstream at lower but still acceptable quality. Higher quality is achieved by more bitstreams combined and the highest quality is achieved by all bitstreams received with no errors. By representing the source with different levels of quality MDC is similar to the layered coding. However, while the latter requires a correct reception of the base layer in order the enhancement layers to be useful, the former can reconstruct the source from any subset of bitstreams [1]. In order to achieve good source reconstruction from any description, all descriptions have to be similar to the source and, thus, similar to each other. However, having similar multiple descriptions is not an efficient way of representing a source. This makes the redundancy allocation an inherent issue of MDC. Under the assumption that some descriptions might be lost, reconstruction quality versus redundancy is the sought compromise in every MDC method.

One of the first practical MDC methods, called multiple description scalar quantization (MDSQ), was proposed by Vaishampayan in [5]. In this method, the source variable to be transmitted is quantized by two coarse quantizers with overlapping quantization cells. Together, those quantizers produce fine quantization with smaller quantization cells.

Another MDC method exploits a linear transform to introduce correlation between the transform coefficients [3], [4]. After a decorrelating block transform, variances of the transform coefficients are estimated along blocks. Coefficients with higher variances are paired with coefficients with lower variances and undergo a pair-wise correlating transform (PCT). This yields two cross-correlated descriptions. If one description is lost, the correlation introduced in a known manner allows estimating a lost coefficient from its counterpart in the pair. This requires knowledge, i.e. transmission of the coefficient variances to the decoder. To ensure more accurate variance estimation, transform blocks can be separated into classes (e.g. predominantly horizontal details, predominantly vertical details, smooth, etc.) and variances are calculated separately within each class [3]. However, this increases the amount of side information as the number of variances is multiplied by the number of block classes.

The above-mentioned problems have been addressed in [6] in the light of whitening the transform coefficients prior to PCT. Whitened coefficients have close variances that can be considered as equal. Correspondingly, there is no need to estimate and transmit those variances as they cancel in the estimator formula. The whitening transform is approximated by a subtraction of a downsampled and coarsely coded image from the original image [6]. Then, a PCT as in [7] is applied to the coefficients of the residual (whitened) image. The resulting two descriptions of the residual image are combined with the duplicated version of the coarse image (shaper).

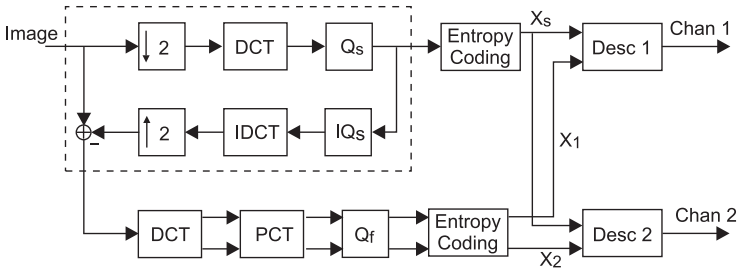


Fig. 1. General scheme of method proposed in [6].

However, for the variables with equal variances, PCT does not introduce any redundancy. Thus, side reconstruction cannot be lower than $\sigma^2/2$, where σ^2 is the coefficient variance. In this contribution, we adopt the idea of 2-stage image coding. However, we refuse from using PCT in the second stage of the coding, thus, eliminating any redundancy in representation of the residual signal. We also suggest modifications in the coarse approximation stage and in the residual image stage aimed at improving the quality for a given bit budget by a better redundancy management. We also present an algorithm, which optimizes the expected distortion based on channel conditions.

The paper is organized as follows. In Section 2, the general coder scheme is described. Next two sections present details about each of two stages: Section 3 deals with modifications in the coarse approximation coding stage while Section 4 deals with modifications in the residual image coding stage. Section 5 provides the analysis of the proposed scheme. Section 6 presents the numerical results and comparisons with other MDC methods, and Section 7 concludes the paper.

2 General coder structure

The general scheme of the method suggested in [6] is shown in Figure 1. The initial image is downsampled by two and then JPEG coded. Its decoded and interpolated version is subtracted from the initial image to approximate a whitening transform. DCT is applied to the residual image to get uncorrelated coefficients with approximately equal variance. They undergo pairwise correlating transform (PCT) outputting two bitstreams. The JPEG coded coarse approximation is called *shaper* and is included into both descriptions. The redundancy in this method is mostly determined by duplicating the shaper as PCT introduces little redundancy when applied to the variables with similar variance. The authors [6] claim their method to produce better results than the method in [3].

We modify the above-described method as shown in Figure 2. In our scheme the shaper (blocks bordered by the dashed line) is generated by decimation

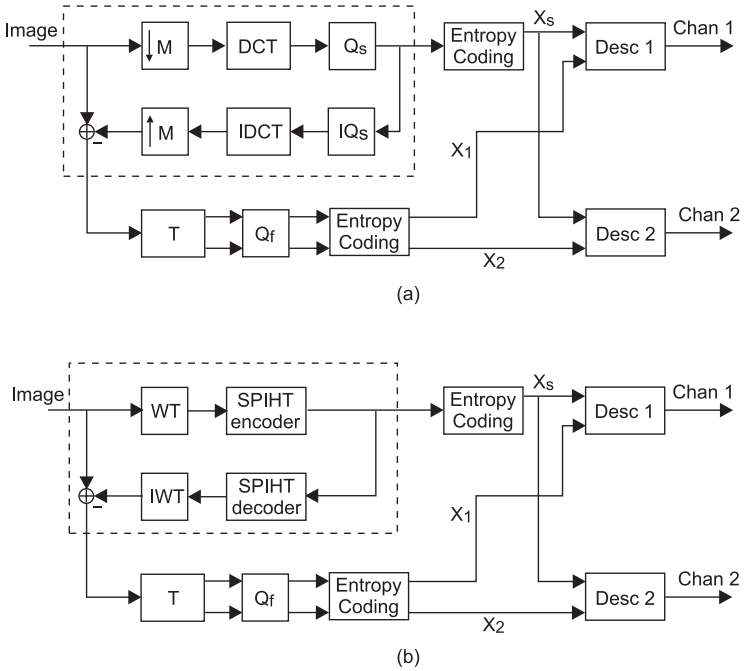


Fig. 2. Varieties of proposed scheme: (a) shaper is obtained by spline resizing and JPEG coding; (b) shaper is obtained by SPIHT coding.

with an arbitrary down-scaling factor of M followed by a JPEG coder (see Figure 2 (a)). We pay special attention to the way the image is decimated and interpolated. We favor a B-spline-based least square image resizing (biorthogonal projection) as it ensures a minimum loss of information [8]. Thus, most of image information is concentrated in the decimated image to be included in both descriptions. For the decimated image, a DCT-based coder is a reasonable choice. Alternatively, the shaper can be generated by a wavelet-based coder, e.g. SPIHT (Figure 2 (b)). In this case, the biorthogonal projection is inherently included in the scheme.

The proposed scheme (Fig. 2) is similar to one-level Laplacian pyramid [9]. The major difference is that in our coder the quantization of the base layer is performed before obtaining the residual image. Hence, compression artifacts from the shaper are present in the residual image and affect its compression. Thus, the interpolation algorithm requires special attention.

In our modification, the residual image is coded by a JPEG-like coder using a block transform (denoted by T). It can be either DCT or lapped orthogonal transform (LOT). The transform coefficients are finely quantized by a uniform quantization step (Q_r). Then, transform blocks are directly split into two parts in a *chessboard* manner and entropy-coded. One part together with the shaper form *Description 1*, while the second part combined again with the shaper form *Description 2*. Thus, each description consists of the coarse image approximation and *half* of the transform blocks of the residual image.

Therefore, no extra redundancy is added in the residual image coding while generating two descriptions instead of one.

The obtained coder provides balanced descriptions both in terms of PSNR and bit rate. The amount of redundancy is also easily adjustable. The algorithm of optimal bit allocation subject to probability of a channel error is also provided. The following two sections explain in detail each stage of the coder. Also we give reasoning to use one or another method for each particular stage.

3 Coarse image approximation

The idea of this stage is to concentrate as much information as possible into the shaper within strict bit rate constraints. We would also like to reduce the artifacts and distortions appearing in the reconstructed coarse approximation. To realize this idea we explore two alternatives: 1) Least squares image resizing prior to JPEG coding; and 2) Wavelet-domain SPIHT coding.

3.1 Least squares spline-based resizing and JPEG coding

A JPEG coder with a limited bit budget would use a large quantization factor applied directly to the original image thus causing unacceptable blocking artifacts. A better alternative, especially for low bit-rate coding, is to decimate the image first and to apply JPEG with more moderate quantization factor. The original image resolution is reconstructed by interpolation as a post-processing step. It has been proven by an analytical model and numerical analysis that by this approach the bit budget is kept the same while the visual quality and PSNR are higher [10]. The method in [6] also makes use of this approach as follows. The decimation by a factor of two in each direction is achieved by averaging over two pixels along columns followed by the same operation along rows. The original resolution is reconstructed by nearest neighbor interpolation. This interpolation introduces blocking artifacts in the coarse approximation and as a result the residual image gets blocking artifacts as well.

In an attempt to concentrate more information in the coarse approximation and correspondingly to make the residual signal closer to white noise, we identify the need of a better interpolation and decimation method. Spline-based interpolation methods have shown their superiority in terms of quality and computational complexity [11], [12]. In the spline formalism, a continuous image model is fit over the discrete pixels, involving B-spline or other optimized piecewise-polynomial basis functions. It allows resampling the initial image at

any arbitrary finer grid. As far as the image decimation is concerned, it has to be performed using functions being biorthogonal to the chosen interpolation function. This is the biorthogonal projection or least squares paradigm, which ensures image decimation with a minimum loss of information [8], [12]. Our practical implementation makes use of a near least squares method for image decimation proved to be effective for a wide range of decimation ratios [13].

The redundancy in our coder is only determined by the size and quality of the shaper. Generally, there are two factors controlling the size of shaper (and hence, the redundancy). The first one is scaling (or interpolation) factor and the second one is the JPEG quantization factor. Using larger downsampling and quantization factors one can get lower level of redundancy, hence, lower quality of side reconstruction (reconstruction from only one description). Alternatively, using smaller downsampling and quantization factors, one can obtain higher quality side reconstruction. The quality of the two-channel reconstruction is determined mostly by quantization step used for quantization of LOT coefficients in the residual image.

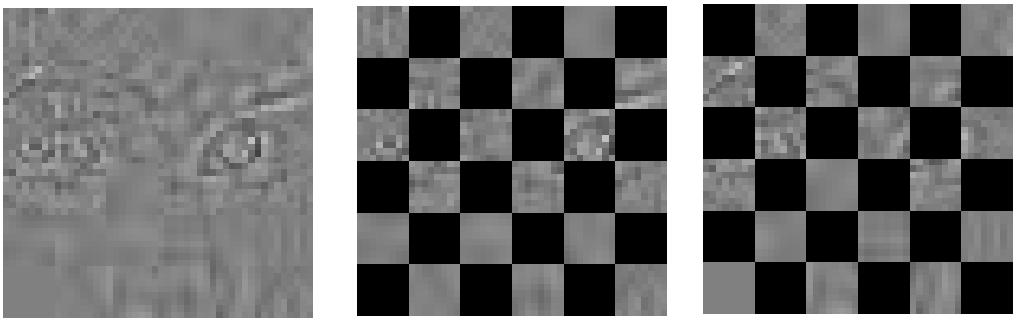
3.2 Wavelet-based coding

An alternative to JPEG coding in obtaining good low bit-rate image approximation is some wavelet-based coding scheme. In general, wavelets provide smooth reconstruction of compressed images even for low bit rates. As they are functions for multiresolution analysis, there is no need of a preliminary decimation step. In fact, the wavelet decomposition is precisely an orthogonal or biorthogonal projection into the space of synthesizing (reconstruction) wavelet functions. Moreover, the best wavelets for compression have been generated via splines, e.g. the famous 9/7 synthesis/analysis wavelet pair. In our scheme we have involved the SPIHT coding and quantization algorithm [14].

4 Residual image coding

Our approach relies on a quality versus bit budget compromise achieved into the coarse approximation brunch. We speculate that our coarse approximation is as good as possible for the given bit budget. Thus, the residual signal is less informative, and there is no need to introduce redundancy to this signal. Respectively, the total redundancy is added by only duplicating the base layer (shaper). We essentially aim at avoiding redundancy in the residual image coding.

The residual image coding in our method is done by a block transform, e.g.



(a) Residual image (b) Description 1 (c) Description 2

Fig. 3. Chessboard splitting of the residual image (in case of DCT).

blocks of 8×8 pixels are considered. In coding the residual image, no redundancy is added provided that a proper block transform has been chosen. To generate MDC, the blocks are simply split into two descriptions in a *chessboard* manner (see Figure 3).

We explore LOT and DCT as block transforms well suited for the residual image coding.

4.1 Coding of the residual signal with block DCT

The residual image is transformed using 8×8 DCT. Then, all transformed blocks are finely quantized with a scalar quantizer using a constant quantization step Q_f. The transform blocks are split between two descriptions in a chessboard manner and entropy coded separately.

4.2 Coding of the residual signal with lapped orthogonal transforms (LOT)

LOT is an alternative to DCT when the quality of the shaper is not good enough. In such cases some blocking artifacts can be encountered if the image reconstruction is based on one description only. LOT can efficiently smooth block borders based on the overlapping windows it uses.

By LOT, each signal block of size N is mapped into a set of N basis functions, each of them is longer than N samples, i.e. overlapping over adjacent blocks [15]. For the 2D case the LOT's are implemented in a separable manner.

In our coder we use Malvar's LOT [15]. The overlapped blocks of the size 16×16 in a spatial domain correspond to 8×8 blocks in the transform domain. Next steps, i.e. quantization by a uniform quantization step Q_r and

chessboard-like splitting into two parts are essentially the same like in the case of DCT block coding.

4.3 Reconstruction when one description is lost

When the decoder receives both descriptions, the reconstruction is straightforward. In case of one-channel reconstruction, the lost coefficients are just filled with zeros. Then, the inverse quantization and inverse transform are applied. The shaper can be obtained from the received description and added to the reconstructed residual image.

It is quite clear that this kind of reconstruction is appropriate when using DCT for coding of the residual image. It was also found that it is the most appropriate way of the reconstruction when using LOT for the residual image coding.

In [16] and [17], it was shown that when reconstructing the original image from only one description, setting the lost coefficients equal to zero produces severe artifacts. Thus, [16] and [17] present methods for estimation of the lost coefficients. In [16], the lost LOT coefficients are estimated as the mean of corresponding coefficients in the neighboring blocks. In [17], it was proposed to use an iterative procedure using maximally smooth recovery method. Moreover, a family of LOT transforms with advanced reconstruction capabilities was presented in [16]. However, it was found that for coding of the residual zero-mean signal these methods work worse than just filling the lost coefficients with zeros. We suggest that this fact is connected with the high frequency nature of the residual signal that does not allow the estimation of the lost LOT block from the neighboring blocks.

4.4 Postprocessing for one-channel reconstruction

When DCT is used in the residual image coding, image reconstructed from one description consists of the blocks, which have different quality. Thus, low bitrate of the shaper causes visible blocking artifacts. There are basically two types of blocking artifacts: artifacts in flat areas of the image and artifacts in the areas with high-frequency content. Artifacts in the flat areas of the image look as the usual artifacts caused by coarse quantization. These artifacts are formed by the offset corresponding to the lost DC coefficient in the residual image and can be reduced by linear filtering across the block borders. In the areas with high-frequency content, high-quality blocks usually have texture while the blocks obtained solely from the shaper are flat. In these parts of the

image, visual quality can be improved by smoothing the border between the flat and textured blocks.

Postprocessing in our coder is based on the deblocking filter of MPEG-4 [18], [19], which we modify according to our needs. The deblocking filter operates in two separate modes depending on the pixel behavior around the block boundary. In each mode, one-dimensional filtering operations are performed across the block boundaries along horizontal and vertical directions (see Figure 4). In Figure 4, we assume that filtering is performed in horizontal direction across the vertical block boundary. The block on the left is reconstructed with higher quality, and the block on the right is reconstructed from the shaper and has lower quality.

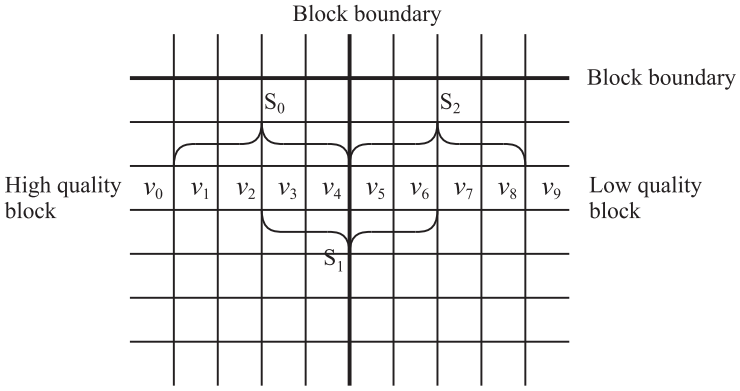


Fig. 4. Block boundaries.

The following procedure is used to find a smooth region with blocking artifacts due to the small DC offset caused by the loss of the DC coefficient in the residual image. The flatness of the region is examined by the following measurement [18]:

$$F(\mathbf{v}) = \sum_{i=0}^{8} \phi(v_i - v_{i+1}) \tag{1}$$

where

$$\phi(\delta) = \begin{cases} 1, \ |\delta| \leq T_1 \\ 0, \ \text{otherwise.} \end{cases} \tag{2}$$

If $F(\mathbf{v}) \geq T_2$, the block is classified as smooth and *smooth region mode* is applied. Otherwise, the block is non-smooth, and the *default mode* is applied. The thresholds are chosen as $T_1 = 2$ and $T_2 = 6$. In the smooth region mode, our filter is the same as the deblocking filter of MPEG-4. In this mode, 1-D nine-tap smoothing filter is applied across the block boundaries, as well as inside the block. Filtering is performed on eight pixels, four on each side of the boundary. The filter coefficients are $h(n) = \frac{1}{16}\{1, 1, 2, 2, 4, 2, 2, 1, 1\}$.

The default mode is modified compared to the MPEG-4 deblocking filter. In the default mode of MPEG-4 deblocking filter, two pixels are modified, each

one from its side of the border. In our coder, the pixels from higher-quality block should not be modified. Instead, we modify only pixel v_5 from the lower-quality block. This is done in the following way. Let us define $a_{0,1}$, $a_{1,1}$, $a_{2,1}$, and $a_{3,1}$ as the four-point DCT coefficients of the pixel array S_1 (see Figure 4). The high-frequency coefficient $a_{3,1}$ is the major factor, affecting the blocking artifact. It is found [18] that the adjustment of this term helps to reduce the blocking artifact. In our coder, the high frequency component is modified by a factor between 0 and 1, resulting in

$$a'_{3,1} = a_{3,1} \frac{\min(|a_{3,0}|, |a_{3,1}|, \frac{|a_{3,0}|+|a_{3,2}|}{2})}{|a_{3,1}|} \tag{3}$$

where $a_{3,0}$ and $a_{3,2}$ are defined similarly to $a_{3,1}$ for the pixels arrays S_0 and S_2, respectively. In (3), the expression $\frac{|a_{3,0}|+|a_{3,2}|}{2}$ is used rather than $|a_{3,2}|$ as in [18], because the $a_{3,2}$ corresponds to "over-smoothed" block of the image.

The coefficient $a_{3,1}$ can be found as

$$a_{3,1} = [k_3 \ -k_1 \ \ k_1 \ -k_3] \cdot [v_3 \ v_4 \ v_5 \ v_6]^T,$$

where

$$k_1 = \frac{1}{\sqrt{2}} \cos \frac{\pi}{8},$$
$$k_3 = \frac{1}{\sqrt{2}} \cos \frac{3\pi}{8}.$$

If the coefficient $a_{3,1}$ is modified to $a'_{3,1}$, pixel v_5 has to be modified as

$$v'_5 = v_5 + \frac{1}{k_1}(a'_{3,1} - a_{3,1}). \tag{4}$$

The clipping operation is applied to the value of v'_5 in order to keep it between 0 and 255. The filtering is performed separately in vertical and horizontal directions. The proposed filtering method improves both the subjective and objective quality of one-channel reconstruction of the image. The simulation results can be found in Section 6.

5 Scheme analysis

5.1 Optimization

For the proposed scheme, we present an algorithm, which optimizes bit allocation subject to probability of the description loss under bit rate constraints.

The proposed algorithm exploits DCT transform and quantization. Thus, it is difficult to achieve exact bit allocation as in the case of progressive coders (e.g. SPIHT). However, approximate bit allocation can be performed.

Let p denote probability of the description loss and R a target bit rate. R_c is the bitrate of the coarse image representation (shaper), and R_f is the bitrate of the refinement part (residual image). The central distortion is D_0, and the side distortions are D_1 and D_2. As we are considering balanced descriptions, $D_1 = D_2$. Then, our task is to minimize

$$2p(1-p)D_1 + (1-p)^2 D_0 \tag{5}$$

subject to

$$2R_c + R_f \leq R. \tag{6}$$

Consider chessboard DCT-based coding of the residual image. The side distortion (D_1) is formed by the blocks, half of which are coded with the distortion D_0, and another half is coded with the distortion of coarse image representation (shaper) D_c. Here we assume that all the 8×8 blocks of the image have the same expected distortion. Consequently,

$$D_1 = \frac{1}{2}(D_c + D_0). \tag{7}$$

Expression (7) can also be used in the case of LOT coding of the residual. As LOT is by definition an orthogonal transform, the mean-squared error distortion in spatial domain is equal to the distortion in transform domain. However, side distortion in transform domain is determined by the loss of half the transform coefficient blocks. Thus, the expression (7) is also valid in case of LOT.

Then, constrained minimization (5) can be transformed to the unconstrained minimization task.

$$J(R_c, R_f) = p(1-p)(D_c(R_c)+D_0(R_c, R_f))+(1-p)^2 D_0(R_c, R_f)+\lambda(2R_c+R_f-R) \tag{8}$$

For higher bit-rates, D_0 only weakly depends on R_c. It will be shown further in the simulation results. Hence, we can write $D_0(R_f)$. The minimization task is then simplified to

$$J(R_c, R_f) = p(1-p)(D_c(R_c)+D_0(R_f))+(1-p)^2 D_0(R_f)+\lambda(2R_c+R_f-R). \tag{9}$$

Generally, we are able to find experimentally the distortion-rate functions $D_0(R_f)$ and $D_c(R_c)$ to carry out the minimization task.

However, a closed-form solution can be found for i.i.d Gaussian random source. The distortion-rate function of a Gaussian source with variance σ^2 is $D(R) =$

$\sigma^2 2^{-2R}$. Hence,

$$D_c(R_c) = \sigma^2 2^{-2R_c} \tag{10}$$

It is shown [20] that Gaussian source is successively refinable in regard to the squared-error distortion measure. Thus, we can write

$$D_f(R_c, R_f) = \sigma^2 2^{-2(R_c+R_f)}. \tag{11}$$

Consequently, our task transforms to the unconstrained minimization of the function

$$J(R_c, R_f) = \sigma^2 p(1-p)(2^{-2R_c}+2^{-2(R_c+R_f)})+\sigma^2(1-p)^2 2^{-2(R_c+R_f)}+\lambda(2R_c+R_f-R). \tag{12}$$

We differentiate the Lagrangian with respect to R_c, R_f, and λ and solve the system of equations. The obtained optimal R_c and R_f are

$$\begin{aligned} R_c^* &= \frac{1}{2}R + \frac{1}{4}\log_2(p) \\ R_f^* &= -\frac{1}{2}\log_2(p) \end{aligned} \tag{13}$$

The rate allocation (13) is, however, sub-optimal because the results were obtained for the case of Gaussian random variable. The "true" optimal allocation must use rate-distortion characteristics of the particular image and to solve the optimization task (9) or (8). An interesting result is that optimal bit allocation for Gaussian source does not depends on the source variance σ^2. This has the meaning that bit allocation is the same for all Gaussian sources and does not depend on the source variance.

One can notice that the optimal bit-rate allocation (13) has the same form as that for the copies of polyphase components in [21] if we substitute $R_0 = R_c + R_f$, and $R_1 = R_c$. This result is quite expected. In fact, our approach and the approach from [21] are based on the same principle. When all the descriptions are received, all the pixels are reconstructed with fine quality. If some descriptions are lost, the pixels from corresponding spatial locations are reconstructed with lower quality. Thus, the results of the optimal bit rate allocation in terms of mean-squared error are expected to be the same.

5.2 Redundancy range

As it was mentioned earlier, the redundancy of the proposed method is $\rho = R_c$. Thus, optimal redundancy ρ^* is

$$\rho^* = R_c^* = \frac{1}{2}R + \frac{1}{4}\log_2(p) \tag{14}$$

We can see from (14) that optimal redundancy depends on the target bitrate R and the probability of description loss p. However, for the values $R \leq -\frac{1}{2}\log_2(p)$, optimal redundancy ρ^* is zero or negative. We interpret this result in a following way. For these values of R and p, one should not use the multiple description scheme. Instead, single description coding has to be used. In other words, when $p \leq 2^{2R}$, redundancy is set to zero, and single description coding is used. However, these results can differ from the actual values obtained by optimization using "true" RD characteristics of particular source. Thus, in practical implementation of the proposed bit allocation algorithm we use the following scheme. If the obtained value of R_c is less than 0.05, we fix $R_c = 0.05$, and $R_f = R - 2R_c$.

One can see from (14) that the upper limit for redundancy is obviously $R/2$. It corresponds to $R_f = 0$. Thus, all the bit budget is allocated to shaper, which is duplicated on both channels. This is achieved for $p = 1$, the situation when the channel is not functioning. Thus, when probability of the description loss is approaching unity, bit allocation is approaching the simple duplication of data in both descriptions.

5.3 Practical bit allocation

In the previous subsection we have solved the problem of optimal bit-rate allocation. It is easy to code the shaper to the given rate R_c with SPIHT or another progressive coder. However, it is difficult to achieve the exact target bit-rate when coding the shaper with JPEG coding. Fortunately, the approximate bit-rate can be estimated.

5.3.1 Coding the shaper

Down-scaling for better transform compression at low bit-rates was studied in [10] by Bruckstein et al. The authors presented an algorithm which finds the optimal down-scaling factor for the given bit-rate. The optimal down-scaling factor is found based on the target bitrate and the second-order image statistics. Thus, using the algorithm proposed in [10] one can find the appropriate down-scaling factor to code the given image with bit-rate R_c^*.

5.3.2 Coding the residual signal

It was mentioned earlier that subtracting a coarse image approximation has an effect of applying whitening transform to the original image. As the residual signal resembles white noise, the DCT coefficients of the residual image have approximately the same variance. Thus, the same quantization step is

used for all DCT coefficients. The compression ratio and the bit-rate of the residual signal are determined by the variance of the residual image σ_r^2 and the quantization step Q_f.

Let us model the DCT coefficients of the residual image as i.i.d. Gaussian random variables with variance σ_r^2. DCT is an orthogonal transform. Thus, variance of DCT coefficients σ_r^2 can be estimated as the variance of pixel values in the residual image.

The rate-distortion function of the Gaussian source is

$$R = \frac{1}{2} \log_2 \frac{\sigma^2}{D}. \tag{15}$$

Under the assumption of fine quantization (which is usually valid) the square-error distortion of the uniform quantization is

$$D = \frac{1}{12} Q^2, \tag{16}$$

where Q is the quantization step. Hence,

$$R_f = \frac{1}{2} \log_2 \frac{12\sigma_r^2}{Q_f} \tag{17}$$

and the quantization step for the residual image is chosen as

$$Q_f = \lceil \sqrt{12} \sigma_r 2^{-R_f} \rceil, \tag{18}$$

where $\lceil \cdot \rceil$ denotes rounding to the nearest integer towards positive infinity.

5.3.3 Basic coding algorithm

The image coding procedure is as follows. Having the probability of the description loss, we find the optimal rates R_c^* and R_f^* according to (13). Then, we estimate the optimal down-scaling factor using the procedure from [10]. Another alternative is to use SPIHT coder for bitrate R_c^*. Then, we decimate the original image with the estimated down-scaling factor and code the decimated image with JPEG to fit the shaper bit-rate (R_c^*). Then, we obtain the residual image as a difference between the original image and the shaper. From the residual image, we estimate the variance of its pixels σ_r^2. We apply (18) to find the quantization step Q_f, and code the residual image.

6 Simulation results

In this section, varieties of our method are explored and compared between themselves and with other MDC algorithms. For the evaluation, first, we applied our method to the test image Lena (512×512, 8 bpp). For most the experiments, we generated two rate-distortion curves. The first curve shows the reconstruction PSNR versus bitrate under the assumption that both descriptions are received. The second curve illustrates the case, when one description is lost. It is obtained by taking the mean result of two descriptions used separately to reconstruct the image.

6.1 Proper decimation and interpolation

In the first experiment, we compare different decimation and interpolation methods to produce the shaper. As for the residual image coding, we fix it to perform block transform coding, involving Malvar's LOT [15]. We apply three decimation/interpolation methods. The first is based on decimation by 2 by averaging over four nearest points and nearest neighbor interpolation, similarly to [6]. Second is DCT-based decimation and interpolation [22], and the third is a near least squares B-spline-based decimation and interpolation [13]. Those three approaches have been combined with the JPEG coder to get the coded shaper. Additionally, the shaper was obtained by a wavelet-domain SPIHT coding. Figure 5 shows the reconstruction results when both descriptions have been received and Figure 6 shows the results when only one description is received. As can be seen among JPEG methods proper anti-aliasing decimation and interpolation give substantial improvement. There, splines and DCT are quite competitive as pre- and post-processing functions. However, the spline-based method is computationally less costly. Among all methods, wavelet-based SPIHT gives superior results.

In our experiments we have used linear splines for interpolation and their biorthogonal counterparts for decimation. Higher order splines would give better results in a pure decimation/interpolation setting. However, the JPEG quantization generates artifacts and the subsequent higher-order interpolation makes them better visible. Linear interpolation plays an additional smoothing effect to these artifacts. What is more important is the least squares setting where the image is properly decimated subject to the chosen interpolation method.

Figure 7 shows interpolation results for image Lena. The original 512×512 image Lena is downsampled with biorthogonal splines to the resolution 128×128. The downsampled image is then quantized and coded with a JPEG-like

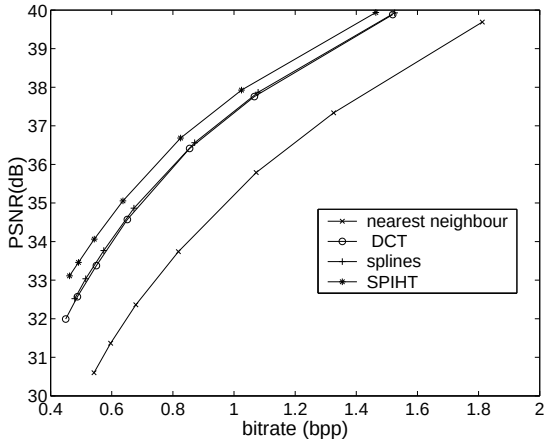


Fig. 5. Central PSNR of overall scheme using different interpolation methods.

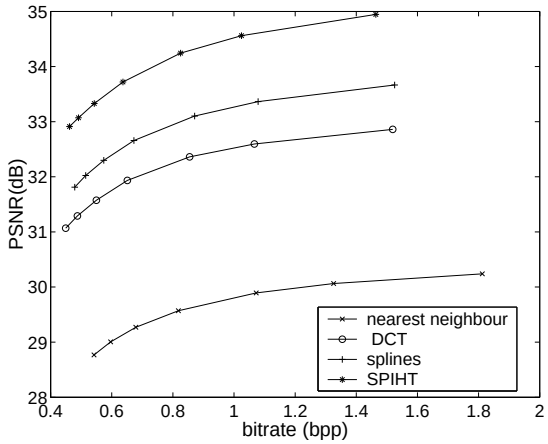


Fig. 6. Mean side PSNR of overall scheme using different interpolation methods.

algorithm. Then, quantized downsampled image is interpolated to the original resolution with linear splines. One can see that $R = 0.085$ allows to code this image with $D_c = 26.73$ dB.

Figures 8 and 9 compare the performance of the Spline-LOT method with that of the method introduced in [6] (denoted by WCT). Spline-LOT coder clearly outperforms WCT coder both for central and side reconstruction mainly due to the adequate decimation/interpolation.

6.2 Shaper scaling and quantization

Next, we explore how the shaper quality works on the total reconstruction quality. Again, our residual image coder is a LOT-based one, while the shaper coder is based on least squares spline decimation/interpolation and JPEG

Fig. 7. Image interpolation results. (a) Original image 512×512 pixels; (b) Image downscaled to 128×128, JPEG-coded and interpolated to the original resolution, $D = 26.73$ dB, $R = 0.085$ bpp.

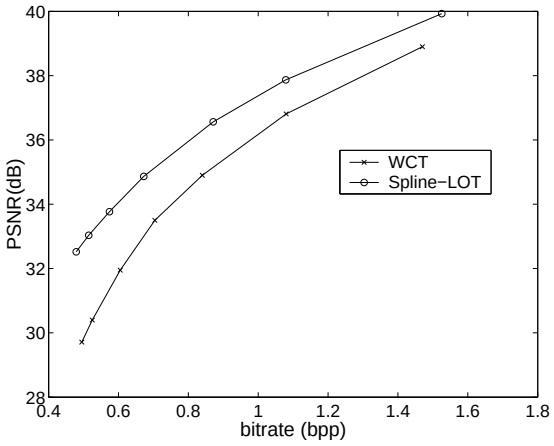


Fig. 8. Comparison of Spline-LOT and WCT coder if both descriptions are received (central PSNR). $Q_s = 0.7$.

with different quantization factor (denoted as Spline-LOT). The shaper quantization factor Q_s is determined as a multiplication factor applied to DCT coefficients before their quantization. Figure 10 shows the results for central reconstruction (from two descriptions) and Figure 11 shows the one-description reconstruction results.

One can see a higher shaper quantization factor slightly reduces the PSNR for central reconstruction but at the same time increases the PSNR when one description is lost. By a finer quantization we thereby provide more bit rate to the shaper. Thus, we introduce more redundancy that improves the side reconstruction.

In addition, the rate-distortion curves for central reconstruction have much steeper slope than rate-distortion curves for the side reconstruction. It evi-

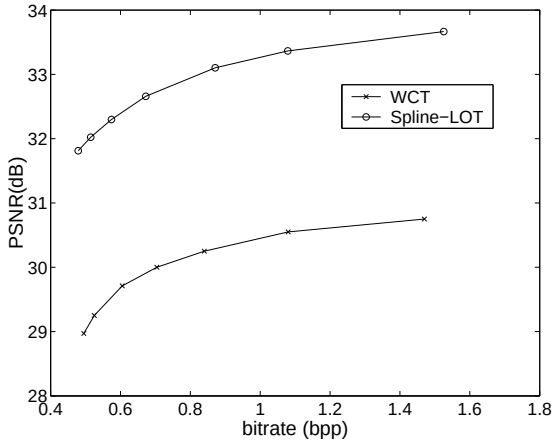


Fig. 9. Comparison of Spline-LOT and WCT coders if one description is received (mean side PSNR). $Q_s = 0.7$

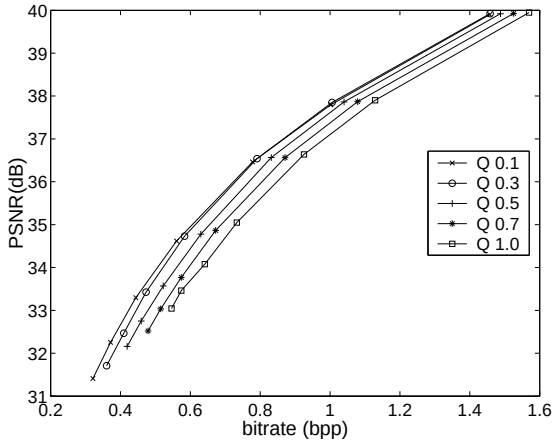


Fig. 10. Rate-distortion performance of Spline-LOT coder for different values of Q_s. Central PSNR.

dences that finer quantization of residual image results in better central reconstruction but has little influence on the single description reconstruction. This effect is cased by the form of the side distortion $D_1 = (D_0 + D_c)/2$. As $D_c \gg D_0$, D_c contributes more to D_1. Thus, finer quantization of the residual image will decrease D_0 but will not considerably decrease D_1. In order to decrease D_1, one has to allocate more bits to the shaper. This will decrease $D_c/2$ component in D_1.

The next algorithm uses variable down-scaling factor for coding the shaper signal. The results are parameterized by the shaper resolution and are shown in Figures 12 and 13. One can see that changing shaper resolution allows to achieve even smaller bitrates comparing with changing just quantization factor.

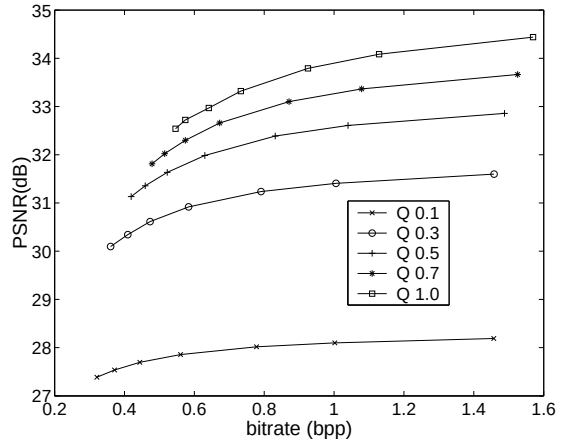


Fig. 11. Rate-distortion performance of Spline-LOT coder for different values of Q_s. Mean-side PSNR.

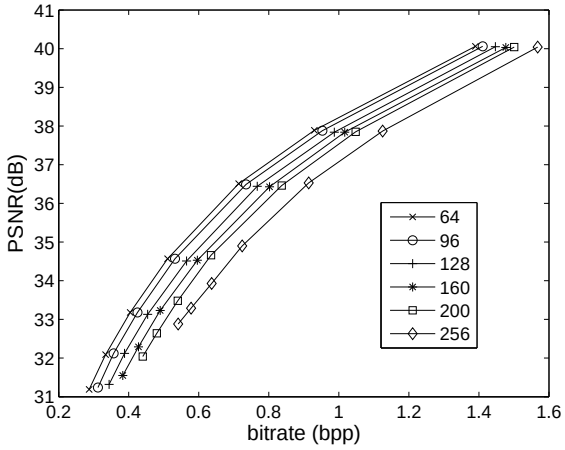


Fig. 12. Rate-distortion performance of Spline-LOT coder for different shaper resolution, Central PSNR.

6.3 Residual signal coding

In the next set of experiments, shaper is obtained by linear B-spline downsampling followed by JPEG-coding. The residual signal is coded as in the following. In the first coder, the residual image is coded by method from [6] involving DCT followed by PCT (this method is denoted Spline-PCT). The second coder exploits block-wise DCT followed by splitting the blocks between two descriptions (the method is denoted Spline-DCT). The third coder (Spline-LOT) uses LOT for coding the residual. For the Spline-DCT coder, mean side reconstruction results are obtained with and without postprocessing. Postprocessing is based on the deblocking filter described in Section 4.4.

Figures 14 and 15 compare the rate-distortion performance of the mentioned

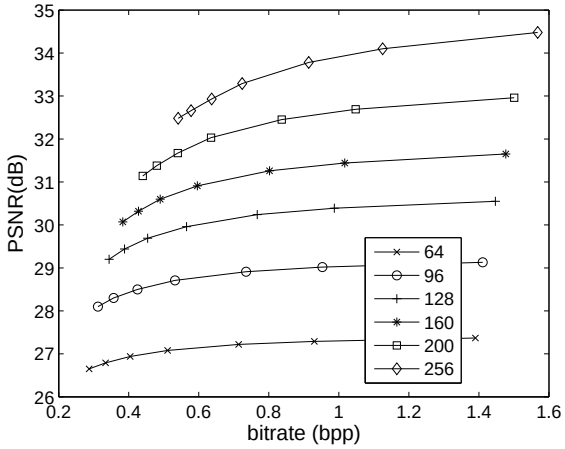


Fig. 13. Rate-distortion performance of Spline-LOT coder for different shaper resolution, Mean-side PSNR.

coders for central and mean side reconstruction for test image "Lena". Figures 16 and 17 show the simulation results for image "Stream and bridge".

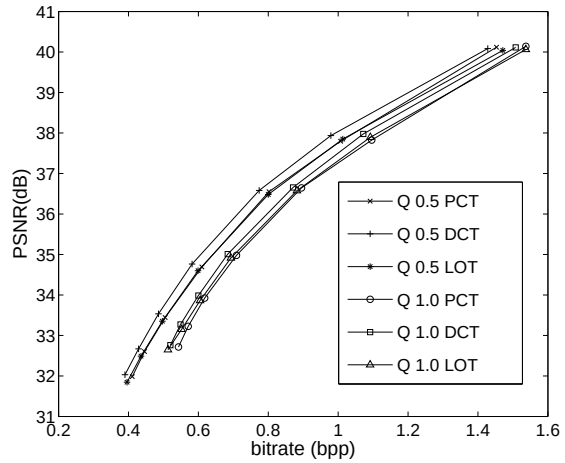


Fig. 14. Comparison of Spline-DCT, Spline-LOT, and Spline-PCT coders if both descriptions are received (central PSNR) for different values of Q_s. Image Lena.

One can see from the figures that for both low-frequency ("Lena") and high-frequency ("Stream and bridge") images, simple DCT shows slightly better performance than LOT and PCT. The results are similar for both low-frequency and high-frequency images. Surprisingly, in terms of PSNR, Spline-DCT coder is competitive and even better than expected to be superior, Spline-LOT coder. While the latter is showing less blocking artifacts, it is not as efficient as DCT in compressing the residual image which is high-frequency. Originally, LOTs have been optimized to compress low frequency signals [15]. One can speculate that using transforms which are optimized for higher frequency content images could give certain improvement in the presented scheme.

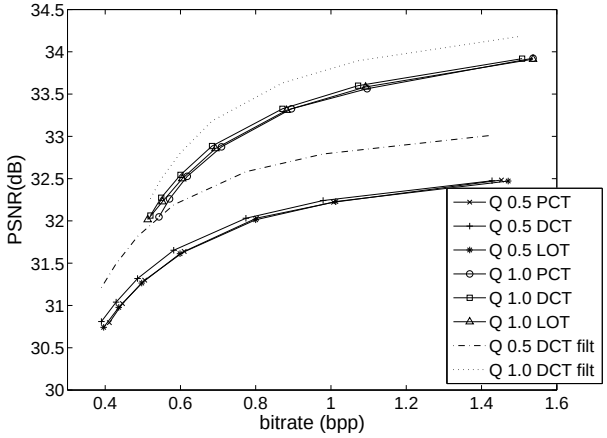


Fig. 15. Comparison of Spline-DCT, Spline-LOT, and Spline-PCT coders if one description is received (mean side PSNR) for different values of Q_s. Image Lena.

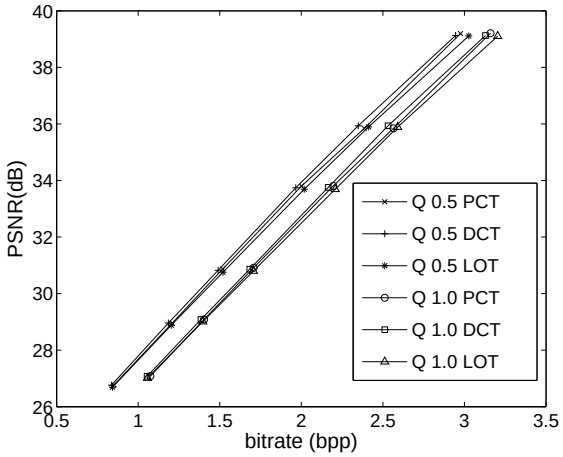


Fig. 16. Comparison of Spline-DCT, Spline-LOT, and Spline-PCT coders if both descriptions are received (central PSNR) for different values of Q_s. Image Stream and bridge.

The experiments with DCT in the residual image coding emphasize once again the importance of a good shaper coding. If we keep the quality of the shaper low to achieve smaller redundancy, the blocking (chessboard-like) artifacts are more visible. This is caused by reconstruction of neighboring blocks with different quality. However, if the shaper quality is high then, for most of the images, those kinds of artifacts are not visible. At least, they do not look visually more annoying than the artifacts caused by the coding of the residual image by LOT or PCT. Blocking artifacts due to DCT-based coding can also be reduced by postprocessing.

Figures 15 and 17 show that postprocessing is beneficial for wide range of bitrates and different redundancies, especially for low-frequency images like "Lena". For high-frequency images, postprocessing gives some advantage when

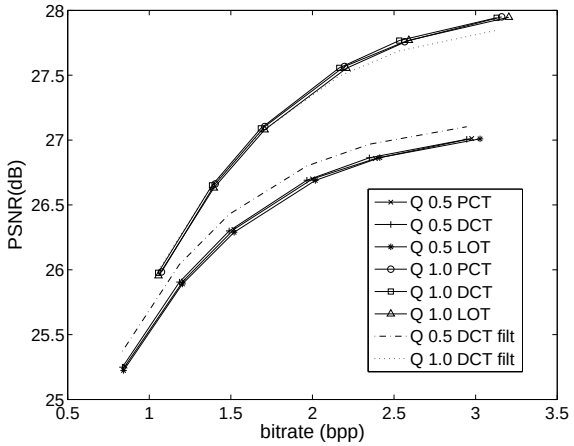


Fig. 17. Comparison of Spline-DCT, Spline-LOT, and Spline-PCT coders if both descriptions are received (mean-side PSNR) for different values of Q_s. Image Stream and bridge.

the shaper quality is low. However, postprocessing does not improve the side reconstruction of the high-frequency image when the shaper has higher quality. This result is expected, as it is difficult to reconstruct high-frequency content of the block from its neighborhood.

Figures 18 and 19 show the images reconstructed from one and two descriptions by the Spline-DCT coder together with the images reconstructed from one and two descriptions by the Spline-LOT coder. One can notice that DCT and LOT produce different visual artifacts when the image is reconstructed from a single description. In particular, DCT produces blocking artifacts caused by the different reconstruction quality of neighboring blocks. LOT does not produce blocking artifacts. Instead, it produces artifacts that look similar to ringing.

Figure 20 shows the effect of postprocessing for one-channel reconstruction of the Spline-DCT coder. As one can see from Figure 20, postprocessing increases both the subjective and objective quality of the low-frequency image reconstructed from one description. The results of postprocessing for one-channel reconstruction for wider range of redundancies can be found in Figure 23 and Tables 1 and 2. The results show that postprocessing is beneficial for wide range of redundancies.

6.4 Bit allocation

This section give the example of bit allocation algorithm proposed in subsection 5.3. Test image is "Boat" (256×256, 8 bpp). The bit allocation algorithm is given the desired bit rate and the probability of the description loss. Bit allo-

Fig. 18. Reconstructed image "Lena", spline interpolation and DCT coding of shaper. DCT coding of the residual ($R = 0.623$ bpp, $\rho = 28\%$): (a) reconstruction from both descriptions, $D_0 = 35.24$ dB; (b) reconstruction from Description 1, $D_1 = 31.84$ dB. LOT coding of the residual ($R = 0.632$ bpp, $\rho = 27.5\%$): (c) reconstruction from both descriptions, $D_0 = 34.80$ dB; (d) reconstruction from Description 1, $D_1 = 30.78$ dB.

cation procedure exploits the closed-form solution (13). One has to notice that this bit allocation is suboptimal as it does not use the "true" rate-distortion characteristics of the source.

Figures 21 and 22 show the RD curves obtained with the bit allocation algorithm for different probabilities of description loss. The coder in these experiments exploits linear splines for decimation and interpolation, and DCT for residual signal coding. These figures show the desired behavior of the rate allocation algorithm. In particular, higher probability of description loss results in bit allocation, which produces higher side PSNR and lower central PSNR.

Fig. 19. Reconstructed image "Stream and bridge", spline interpolation and DCT coding of shaper. DCT coding of the residual ($R = 1.700$ bpp, $\rho = 14.7\%$): (a) reconstruction from both descriptions, $D_0 = 32.10$ dB; (b) reconstruction from Description 1, $D_1 = 26.52$ dB. LOT coding of the residual ($R = 1.705$ bpp, $\rho = 14.7\%$): (c) reconstruction from both descriptions, $D_0 = 31.58$ dB; (d) reconstruction from Description 1, $D_1 = 25.63$ dB.

(a) (b)

Fig. 20. Effect of postprocessing on one-description reconstruction; $D_0 = 35.813$ dB, 0.636 bpp: (a) Not filtered, $D_1 = 28.784$ dB; (b) Filtered, $D_1 = 29.883$ dB;

In the above, we compared our coder with the WCT coder of [6]. Figures 8 and 9 prove that our coder clearly outperforms the WCT coder. We suggest that this is due to proper decimation/interpolation.

Here we compare our 2-stage coder with other MD image coders based on JPEG. The 2-stage coder exploits B-splines and JPEG for coding of shaper. The residual image is coded with DCT. Two modifications of this coder are used for comparison: with and without postprocessing described in Section 4.4. These coders are called 2-stage+post-filt and 2-stage, respectively. We compare our 2-stage coder with two MD coders presented in [3]. One of those coders is a JPEG-based MDTC image coder. Correlating transform is applied to the pairs of DCT coefficients. The correlation added by this transform allows to estimate the value of the lost coefficient from the received coefficient. Another coder (MDSQ) is based on applying multiple description scalar quantization [5] to DCT coefficients of the JPEG coder. Test image Lena (512×512, 8 bpp) is used for comparison.

The central distortion for MDTC and MDSQ coders is $D_0 = 35.78$ dB. The central distortion for our 2-stage coder ranges from 35.80 dB to 36.00 dB. Figure 23 represents the RD curves for one-channel reconstruction for the 2-stage, 2-stage+post-filt, MDTC, and MDSQ coders. Different operating points for the 2-stage coder are obtained by varying the down-sampling factor for the shaper.

Figure 23 demonstrates that even without postprocessing, 2-stage coder substantially outperforms MDTC coder for the whole range of redundancies. The difference is even larger for low redundancies. We suggest that the superior performance of the 2-stage coder in the low redundancy region is due to the down-sampling before JPEG coding of shaper. This down-sampling before compression allows to obtain higher PSNR for low bit rates compared to the conventional JPEG compression. The 2-stage coder also performs better or comparable with MDSQ coder for higher redundancies. In the middle range of redundancies our coder works better than MDSQ. Moreover, our coder achieves smaller redundancies than the coders based on MDSQ and MDCT. One can notice that 2-stage coder is able to produce meaningful side reconstruction even with the redundancy less than 5%. Figure 23 shows that postprocessing can improve both the subjective and objective quality, especially for low bitrates. The increase in PSNR due to the postprocessing is up to 1 dB.

The simulation results for the 2-stage coder with and without postprocessing for image “Lena” are given in Table 1. The simulation results for image

"Stream and bridge" can be found in Table 2.

Bit rate (bpp)	Mean side PSNR (dB)	Mean side PSNR (post-filtering) (dB)	Central PSNR (dB)	Shaper PSNR (dB)	Redundancy (%)
0.617	27.053	28.383	35.834	24.340	4.6
0.636	28.714	29.776	35.813	26.149	8.8
0.663	29.998	30.832	35.792	27.600	13.7
0.694	30.984	31.658	35.828	28.751	19.1
0.737	32.138	32.559	35.839	30.169	26.4
0.807	33.458	33.743	35.983	31.871	38.6
0.878	34.006	34.104	35.953	32.667	45.3

Table 1
Performance of the 2-stage coder. Image "Lena" (512×512, 8 bpp).

Bit rate (bpp)	Mean side PSNR (dB)	Mean side PSNR (post-filtering) (dB)	Central PSNR (dB)	Shaper PSNR (dB)	Redundancy (%)
1.577	24.026	24.630	32.383	21.345	3.1
1.612	25.271	25.661	32.391	22.703	6.5
1.659	25.895	26.152	32.396	23.400	9.6
1.748	26.538	26.656	32.396	24.131	14.3
1.833	26.954	26.995	32.403	24.612	18.1
1.911	27.25	27.228	32.415	24.957	21.3
1.950	27.373	27.323	32.419	25.101	22.7

Table 2
Performance of the 2-stage coder. Image "Stream and bridge" (512×512, 8 bpp).

7 Conclusion

We have developed a practical MDC method that improves the two-stage scheme proposed previously in [6]. The first stage of our coder employs spline interpolation to obtain the image with lower resolution, which is then coded and sent to both channels. This coarse image is coded in a way to have a lower bit rate, yet being smooth and providing satisfactory quality. Then, properly interpolated, this image is subtracted from the original one, yielding a residual (details) image. We spend no redundancy in coding two descriptions out of it. To achieve this, a chessboard splitting of block transform coefficients is applied.

Two block-transform coders were compared for coding of the residual image. The simpler DCT-based coder showed competitive results to the LOT-based one. While the latter was expected to yield reconstructed images with less blocking artifacts, the good results for the former prove that we have achieved a residual image as high-frequency (noisy-like) as possible and correspondingly better compressible by DCT. The improved performance is due to the adequate decimation/interpolation scheme we have applied based on biorthogonal projection (either spline or wavelet). The postprocessing in the DCT-based coder reduces blocking artifacts and increases side reconstruction quality.

Our MDC method shows better performance comparing to the method in [6] both for reconstruction from one and two descriptions. It also outperforms the MD coders from [3].

The further development of this coder may employ using suitable wavelet transforms for coding the residual signal. An application of this method for video coding is also to be considered.

References

[1] V. Goyal, Multiple description coding: compression meets the network, IEEE Signal Processing Mag. 18 (2001) 74–93.

[2] L. Ozarow, On a source-channel coding problem with two channels and three receivers, Bell Syst. Tech. J. 59 (10) (1980) 1909–1921.

[3] Y. Wang, M. Orchard, V. Vaishampayan, A. Reibman, Multiple description coding using pairwise correlating transforms, IEEE Trans. Image Processing 10 (3) (2001) 351–366.

[4] V. Goyal, J. Kovacevic, Generalized multiple description coding with correlating transforms, IEEE Trans. Inform. Theory 47 (6) (2001) 2199–2224.

[5] V. Vaishampayan, Design of multiple description scalar quantizers, IEEE Trans. Inform. Theory 39 (3) (1993) 821–834.

[6] K.-P. Choi, K.-Y. Lee, An efficient multiple description coding using whitening transform, IEICE Trans. Fundamentals E86-A (6) (2003) 1382–1389.

[7] M. Orchard, Y.Wang, V. Vaishampayan, A. Reibman, Redundancy rate distortion analysis of multiple description image coding using pairwise correlating transforms, in: in Proc. Int. Conf. Image Processing, Santa Barbara, CA, 1997, pp. 608–611.

[8] A. Munos, T. Blu, M. Unser, Least squares image resizing using finite differences, IEEE Trans. Image Processing 10 (2001) 1365–1378.

[9] P. Burt, E. Adelson, The laplacian pyramyd as a compact image code, IEEE Trans. Communications 31 (1983) 532–540.

[10] A. Bruchstein, M. Elad, R. Kimmel, Down-scaling for better transform compression, IEEE Trans. Image Processing 12 (9) (2003) 1132–1144.

[11] P. Thevenaz, T. Blu, M. Unser, Interpolation revisited, IEEE Trans. Medical Imaging 19 (2000) 739–758.

[12] A. Gotchev, Spline and wavelet based methods for signal and image processing, Doctor of technology thesis, Tampere University of Technology (Sept. 2003).

[13] A. Gotchev, K. Egiazarian, G. Marchokov, T. Saramäki, A near least squares method for image decimation, in: in Proc. Int. Conf. Image Processing (ICIP'03), Barcelona, Spain, 2003.

[14] A. Said, W. Pearlman, A new, fast, and efficient image codec based on set partitioning in hierarchical trees, IEEE Trans. Circuits Syst. Video Technol. 6 (3) (1996) 243–250.

[15] H. S. Malvar, D. H. Staelin, The LOT: transform coding without blocking effects, IEEE Trans. Acoustics, Speech, and Signal Processing 37 (1989) 553–559.

[16] S. S. Hemami, Reconstruction-optimized lapped orthogonal transforms for robust image transmission, IEEE Trans. Circuits Syst. Video Technol. 6 (1996) 168–181.

[17] D.-M. Chung, Y. Wang, Multiple description image coding using signal decomposition and reconstruction based on lapped orthogonal transforms, IEEE Trans. Circuits Syst. Video Technol. 9 (1999) 895–908.

[18] S. D. Kim, J. Yi, H. M. Kim, J. B. Ra, A deblocking filter with two separate modes in block-based video coding, IEEE Trans. Circuits Syst. Video Technol. 9 (1999) 156–160.

[19] MPEG-4 video verfification model v.9.1, ISO/IEC JTC/SC 29/WG11/N2502a.

[20] W. Equitz, T. Cover, Successive refinement of information, IEEE Trans. Inform. Theory 37 (2) (1991) 269–275.

[21] P. Sagetong, A. Ortega, Optimal bit allocation for channel-adaptive multiple description coding, in: in Proc. Image and Video Commun. and Processing, San Jose, CA, 2000, pp. 53–63.

[22] L. Yaroslavsky, A. Happonen, Y. Katiyi, Discrete signal sinc-interpolation in DCT domain: fast algorithms, in: in Proc. Int. TICSP Workshop on Spectral Methods and Multirate Signal Processing, SMMSP'02, Toulouse, France, 2002, pp. 179–185.

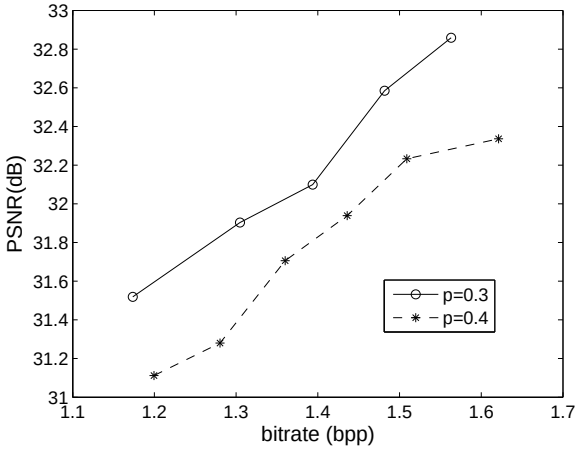


Fig. 21. Image Boat. Bit allocation algorithm for different values of p (probability of description loss). Central PSNR.

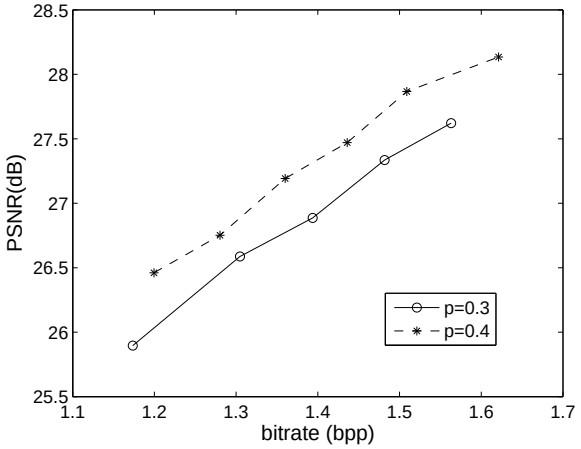


Fig. 22. Image Boat. Bit allocation algorithm for different values of p (probability of description loss). Mean side PSNR.

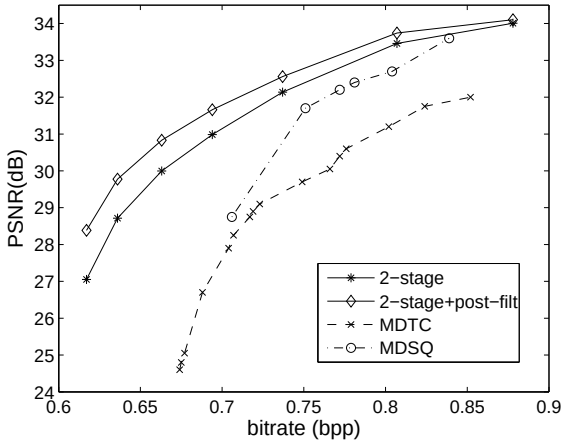


Fig. 23. RD performance of different coders; image Lena (512×512). Reconstruction from a single description. For MDTC and MDSQ, $D_0 = 35.78$ dB. For 2-stage and 2-stage with post-filtering, $D_0 = 35.80 \div 36.00$).