Frame based Multiple Description for Multimedia Transmission over Wireless Networks

R. Bernardini*, R. Rinaldo*, A. Tonello*, A. Vitali†

*Università degli Studi di Udine - DIEGM
Via delle Scienze 208, Udine, Italy
e-mail: {bernardini,rinaldo,tonello}@uniud.it

†ST Microelectronics
Via C. Olivetti n. 2, 20041, Agrate Brianza - Italy
email: andrea.vitali@st.com

***Abstract*— In this paper, we propose a simplified Frame based Multiple Description (MD) coding system for video streams. The proposed system uses an overdetermined filter bank to generate multiple descriptions of the original video stream, and allows for exact signal reconstruction even in the presence of packet losses. Each description is coded using the recent H264 / AVC video coding standard. With no loss, the coded signal can be reconstructed by means of the dual synthesis filter bank. In the presence of packet losses, we adopt a "restoring stage" before the synthesis filter bank whose purpose is to recover of lost coefficients. This solution has limited computational complexity and does not introduce excessive delay in interactive applications. We compare the robustness and rate-distortion performance of the proposed solution with that of other MD solutions based on polyphase spatial subsampling. The experimental results show that the proposed scheme is competitive for relatively high bit-rates and loss probability.**

I. Introduction

Transmission of multimedia signals over wireless and Mobile Ad Hoc Networks (MANET) has recently attracted considerable attention. Due to the characteristics of the radio channel, these networks typically exhibit high error rate. This is especially true for MANET, due to the fact that they do not rely on pre-existing infrastructures and their topology can change very rapidly, thus making them particularly prone to errors and packet drops.

Robust transmission of multimedia streams over wireless networks is an important issue for many applications. The transmission of multimedia signals can make unpractical the usual mechanism of acknowledge and packet retransmission. Moreover, a return channel may not be available or inconvenient to use. The use of error correction techniques may also not help when packets are actually dropped or when one of the different paths used to send the bit stream totally fails, e.g, when a node is detached.

Multiple Description (MD) coding is a recently proposed solution for increasing the resilience of multimedia transmission to such problems. The idea is to send a redundant description of a single source to the receiver and be able to reconstruct the transmitted data when part of this information is lost. MD coding comprises a very wide range of techniques such as multiple description quantization [1], correlating transforms [2], [3], Forward Error Correction (FEC) coding [4], and redundant basis (frames) [5], [6], [7].

Among the different MD techniques known in the literature we consider here the use of redundant bases (frames). In frame-based MD, the input signal is analyzed with a redundant filterbank. The introduced redundancy allows one to recover the coded signal even in presence of packet loss.

In a sense, frame expansion is similar to MD Forward Error Correction coding, a recently proposed scheme where redundancy is added across packets. The possible advantage in using frames is that added information can be perceptively consistent, so that every description contains information that can help to reconstruct or approximate the original sequence. On the other hand, an error correcting code is helpful only if a certain minimum amount of information is correctly received, and if no exact recovery is possible, "parity check" packets are not useful.

In this paper, we analyze the performance of a simple frame based video coding scheme for transmission over wireless channels. In the proposed scheme the input signal is separated into its spatial polyphase components and an additional description, which acts like a visually consistent "parity check" sequence, is obtained by low-pass filtering and subsampling the original signal. In order to recover the coded signal in case of coefficient loss the error recovery algorithm of [8] is exploited. When exact reconstruction is not possible because of excessive errors, error concealment based on bilinear interpolation is used.

In all schemes, individual descriptions are coded using independent H264 / AVC video coders [9]. After error recovery or concealment, frames are copied onto the decoders' frame buffers at the receiver, in order to mitigate the effect of error propagation due to differential coding. The experimental results show that the proposed scheme is competitive for relatively high bit-rates and loss probability.

Section II provides an overview of frame theory and reviews the algorithm of [8]. Section III describes the proposed MD video coding scheme based on frames. In Section IV we present some experimental results evaluating the rate-distortion performance of the proposed scheme, comparing it with the other two MD schemes and with standard Single Description (SD) coding.

II. Oversampled filter banks and frames

In order to make this paper self-contained, in this section we briefly recall the main results of frame theory [10] and summarize the algorithm of [8].

A family of signals $\Phi = \{\phi_k \in \ell^2(Z)\}_{k\in Z}$ constitutes a frame if for any signal $x \in \ell^2(Z)$ there exist two constants $A > 0$ and $B < \infty$ such that

$$A\|x\|^2 \leq \sum_{k=1}^{N} |\langle x, \phi_k\rangle|^2 \leq B\|x\|^2, \tag{1}$$

where $\langle f, g\rangle = \sum_n f[n]g^*[n]$ is the scalar product between f and g. In particular, the left-hand inequality guaranties that it is possible to reconstruct (in a robust way) the original signal x from the scalar products $y_k = \langle x, \phi_k\rangle$ and that it is possible to compute a set of *dual* signals $\tilde{\phi}_k$ such that

$$x = \sum_k \langle x, \phi_k\rangle \tilde{\phi}_k = \sum_k \langle x, \tilde{\phi}_k\rangle \phi_k. \tag{2}$$

In the framework of oversampled filter banks, one computes a vector of output coefficients in each channel $c = 1, ..., N$, via convolution, i.e.,

$$y_c[n] = \sum_{m\in Z} x(m)h_c[Mn - m]. \tag{3}$$

The right-hand side of Eq. (3) can be interpreted as the scalar product between the input and the analysis vector $\phi_k \triangleq h_c[Mn - \cdot]$. By appropriate filter design, the ϕ_k constitute a frame. Oversampling, i.e., choosing $N > M$, implies that there is redundancy in the coefficients $y_c[n]$ which can be exploited to reconstruct x even if some coefficients are lost.

For finite dimensional signals and FIR analysis filters, one can collect the input and the filter output coefficients in vectors $\mathbf{x}$ and

$\mathbf{y}$, respectively, and express the filtering operation via a matrix product $\mathbf{y} = \mathbf{F}\mathbf{x}$. One can recognize that the rows of matrix $\mathbf{F}$ correspond to ϕ_k and that they are the time-reversed and translated impulse responses of the analysis filters. Note that, because of oversampling, matrix $\mathbf{F}$ is a rectangular matrix with more rows than columns.

One can reconstruct $\mathbf{x}$ from $\mathbf{y}$ by means of the pseudo-inverse $\mathbf{F}^\dagger$ of matrix $\mathbf{F}$, i.e., $\mathbf{x} = \mathbf{F}^\dagger \mathbf{y}$. Considering (2), it turns out that the columns of $\mathbf{F}^\dagger$ correspond to the dual frame elements $\tilde{\phi}_k$. It is important to note that reconstruction via the dual frame is optimal even if $\mathbf{y}$ does not belong to Im($\mathbf{F}$). This is usually the case in applications where the coefficients in $\mathbf{y}$ are quantized into $\hat{\mathbf{y}}$ before coding and transmission. In such a case, the vector $\hat{\mathbf{x}} = \mathbf{F}^\dagger \hat{\mathbf{y}}$ is the minimum norm vector minimizing the squared error $||\mathbf{F}\hat{\mathbf{x}} - \hat{\mathbf{y}}||^2$, i.e., it is the one best describing the received coefficients.

In case of coefficient loss, one can pretend that the input was analyzed with a subset $\Phi_I = \{\phi_k\}_{k \in I}$ of the analysis functions. If the corresponding matrix $\mathbf{F}_I$ is still full rank, set Φ_I is indeed a frame and the input $\mathbf{x}$ can be recovered from the set of coefficients $\mathbf{y}_I$, namely $\mathbf{x} = \mathbf{F}_I^\dagger \mathbf{y}_I$. At the receiver, therefore, one needs to compute the pseudo-inverse of the original analysis matrix after some row cancellation, corresponding to the coefficient loss pattern. It turns out, unfortunately, that the pseudo-inverse operator does not have a filter bank structure anymore. Moreover, direct computation can be computationally demanding, due to the large dimensions of the involved matrix. In [8], an algorithm for the evaluation of the pseudo-inverse $\mathbf{F}_I^\dagger$ is presented. It puts in front of the original synthesis filter bank a "restoring" stage which recovers the lost coefficients from the received ones. For the sake of reference, we now briefly recall the results of [8].

If we denote with I^c the set of lost coefficients, one can write from (2)

$$\phi_k = \sum_{n \in Z} \phi_n \langle \phi_k, \tilde{\phi}_n \rangle = \sum_{n \notin I^c} \phi_n \langle \phi_k, \tilde{\phi}_n \rangle + \sum_{m \in I^c} \phi_m \langle \phi_k, \tilde{\phi}_m \rangle.$$

Taking the scalar product with x, we have

$$y_k = \langle x, \phi_k \rangle = \sum_{n \notin I^c} y_n \langle \tilde{\phi}_n, \phi_k \rangle + \sum_{m \in I^c} y_m \langle \tilde{\phi}_m, \phi_k \rangle.$$

In matrix form,

$$\boldsymbol{y}_m = \boldsymbol{M}\boldsymbol{y}_m + \boldsymbol{M}'\boldsymbol{y}_r$$

where $\boldsymbol{y}_m$ is the vector of *lost* coefficients and $\boldsymbol{y}_r$ the set of *received* coefficients. The lost coefficients can therefore be recovered by $\boldsymbol{y}_m = (\boldsymbol{I} - \boldsymbol{M})^{-1}\boldsymbol{M}'\boldsymbol{y}_r$. It is possible to show that using the restored coefficients as the input to the the original dual synthesis filter bank is equivalent to the application of the pseudo-inverse $\mathbf{F}_I^\dagger$. Moreover, if the original analysis and synthesis filter banks are FIR, matrix $(\boldsymbol{I} - \boldsymbol{M})$ has a block structure, and each block inversion can be computed as soon as the block is available, thus reducing the delay and the complexity of the procedure [8]. In the presence of excessive errors, it may happen that the resulting set Φ_I becomes incomplete. In this case, one could use the MSE optimal approximation $\boldsymbol{y}_m = (\boldsymbol{I} - \boldsymbol{M})^\dagger \boldsymbol{M}'\boldsymbol{y}_r$ [8]. In any case, the algorithm in [8] allows to recognize which parts of the input x could not be recovered.

III. Overview of the System

The results outlined above are used for the design of a simplified frame based MD video coder. The descriptions are generated using a one-dimensional filter bank applied to columns of every sequence frame. The filter outputs are subsampled by a factor 2. The analysis filters separate even and odd rows, i.e., the filter impulse responses are $h_0(n) = \delta_n$, $h_1(n) = \delta_{n+1}$. An additional low-pass filter $h_2(n)$ generates the third description. Thus, for an $N_r \times N_c$ input frame, the scheme originates 3 descriptions with dimension $N_r/2 \times N_c$ pixels. The filter $h_2(n)$ belongs to the family of the well-known wavelet Daubechies filters [10]. These filters are orthogonal to their factor 2 translations, and this grants that the synthesis filter bank is FIR, too. The effect of the filter length is briefly discussed below.

The descriptions are coded using three independent H264 / AVC standard coders, as shown in Fig. 1. The H264 / AVC coder divides the input frame into slices made from macroblocks (MB) organized into rows. Each slice is then sent over the network in a single packet. The reason for subsampling along the columns is that the loss of one slice results in a limited number of contiguous lost coefficients in each column, hopefully permitting error recovery.

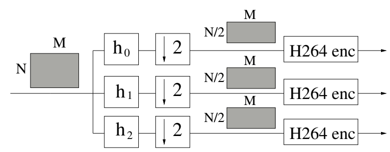


Fig. 1. Block diagram of the proposed frame based MD coder.

On the receiver side (Figure 2), the video streams are independently processed by H264 / AVC synchronised decoders. These decoders are connected to a restoring block $\mathbf{R}$, which recovers channel errors by implementing the algorithm outlined above. We remark that this solution needs only local information and does not introduce any relevant delay in the decoding process. In the case of unrecoverable errors, corresponding to an incomplete Φ_I, missing regions in lost descriptions are reconstructed using bilinear interpolation from the received ones. This still gives acceptable results, due to the high spatial correlation among descriptions in the proposed scheme. We found that this is indeed preferable to the use of the MSE optimal approximation $\boldsymbol{y}_m = (\boldsymbol{I} - \boldsymbol{M})^\dagger \boldsymbol{M}'\boldsymbol{y}_r$.

The result is a set of three recovered subframes with dimension $N_r/2 \times N_c$. These subframes are then fed into the synthesis filter bank, whose output is the original full size sequence. Recovered subframes for each description are copied into the corresponding decoder frame buffer, in order to prevent error propagation from reference frames due to interframe coding.

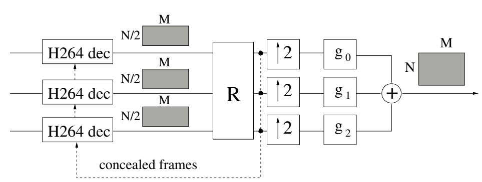
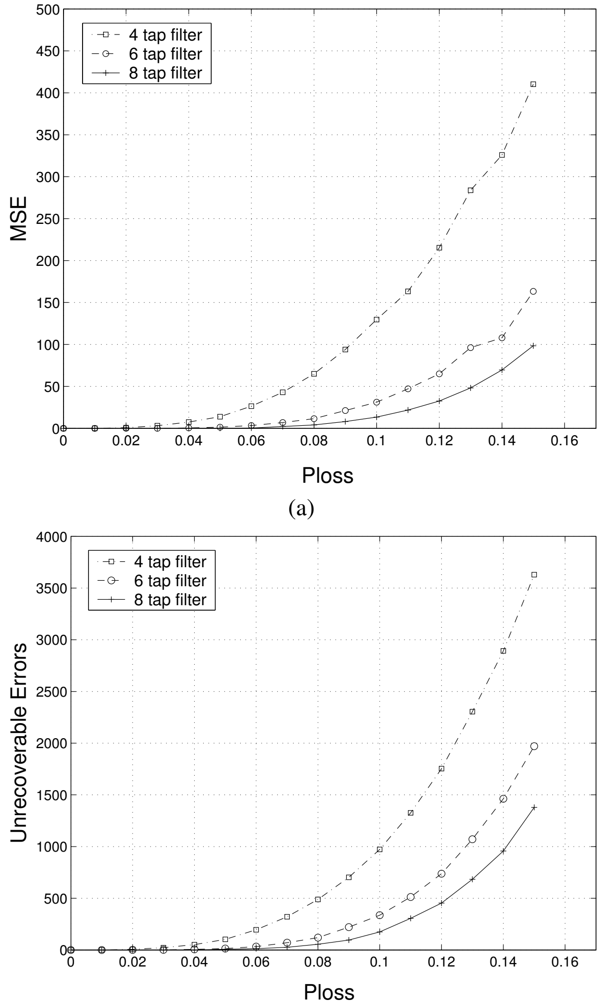


Fig. 2. Block diagram of the proposed decoder.

IV. Experimental Results

The first set of experiments aims at the evaluation of the intrinsic robustness of the proposed frame based scheme, without considering the quantization effect due to coding. Therefore, we first tested the proposed solution on uncoded still images. The following results are relative to the simulation of the analysis-transmission-recovery process for the 512 $\times$ 512 image *Lena*. The redundant analysis filter bank considered in the previous section is used, then each column coefficient in the three descriptions is lost independently with probability P_l. Since we are interested in the robustness of the proposed scheme, in case of unrecoverable errors lost coefficients are replaced by $\boldsymbol{y}_m = (\boldsymbol{I} - \boldsymbol{M})^\dagger \boldsymbol{M}'\boldsymbol{y}_r$. Results are averages of 50 independent trials. Figure 3.a shows the MSE between the original and reconstructed images as a function of P_l and for various choices of the Daubechies' filter length.

As it can be seen from the figure, error recovery capabilities increase with filter length. The 8 tap Daubechies filter allows for perfect reconstruction up to $P_l = 0.06$. The number of



(a)

(b)

Fig. 3. (a) Luminance MSE vs. coefficient loss probability P_l for the image *Lena*. (b) Number of unrecoverable errors vs. coefficient loss probability P_l.

unrecoverable pixels also increases with P_l, as depicted in Figure 3.b for different choices of the h_2 filter length.

We then evaluate the performance of the proposed video coding scheme presented in Section III. It is compared with other two MD systems based on spatial subsampling of original frames. The first one originates four descriptions from the spatial polyphase components of the original frame. Each description, whose dimension is 1/4 of that of the original frame, is compressed independently, packetized and sent over an error prone network. The second scheme has a similar structure, but only two descriptions are generated by separating the even and odd rows of the original frame. We compare the MD schemes with a standard Single Description H264 / AVC coder which includes basic error concealment as described in [11]. The coders use the H264 / AVC test model software version *JM6.0a*. To increase robustness to channel errors and make a fair comparison, the SD coder uses the Random Intra Macro Block Refresh coding option, i.e., 100 Macro Blocks for every CIF frame are coded in intra-mode. No Random Intra Macro Block Refresh coding option is activated in the MD schemes. Other coding options are the same for the SD and MD coders. In particular the GOP structure is I BBBB P BBBB P BBBB P BBBB I, and slices have a fixed 1000 byte dimension. Each slice is sent as a packet, and each packet is lost independently with probability P_l. For the two MD coders based on simple spatial subsampling, in case of errors in one or more descriptions appropriate error concealment via bilinear interpolation from correctly received descriptions is performed at the receiver. Similarly to our scheme, corrected subframes are copied into the corresponding receiver frame buffers. In case all the descriptions are lost, basic error concealment is applied as in [11] in all the MD coders, including the frame based one.

The simulations we present here are relative to 100 frames of the CIF sequence *News*. Results are averages of 50 independent transmission trials. Figure 4 shows the performance of the coders.

Despite the fact that the SD coder can exploit spatial redundancy more efficiently, it does not have the best performance even for $P_l = 0$, due to the intra refresh coding option. Nonetheless, the performance of the SD coder drops very rapidly for increasing P_l. We expect that the MD coder with four descriptions presents good robustness to errors from a subjective quality point of view, at the expense of some coding inefficiency. The proposed frame based coder adds 1.5 redundancy to the video stream, and has therefore a low coding efficiency but possibly good robustness to errors both from a subjective and objective quality point of view. Note that, with no coding, perfect reconstruction is still possible in this case even in the presence of errors. Finally, the MD coder with two descriptions has better coding efficiency but possibly worse performance in terms of subjective quality, since packet losses have to be corrected with the interpolation of entire rows.

It can be seen from Figure 4 that for loss probability $P_l = 0.05$, the proposed frame based MD scheme performs better than the SD scheme and the MD scheme with four descriptions. Moreover, at relatively high bit-rates, the proposed scheme has the best performance of all schemes. For $P_l = 0.1$, the advantage of the proposed solution is even more evident. To compare the visual quality performance of the proposed frame based scheme, we show in Fig. 5 (bottom) a detail of the reconstructed video stream *News* coded at about 1 Mb/s. In the same figure (top), we show the reconstructed frame for the MD scheme with two descriptions. We consider in both cases the effects of the loss of one single slice, corresponding to the region evidenced by the white rectangle in the figure (top). It can be seen from the figure that the MD scheme with two descriptions can originate annoying artifacts, especially along diagonal edges.

Fig. 5. Details of reconstructed frames for the two description MD system (top) and the frame-based MD system (bottom).

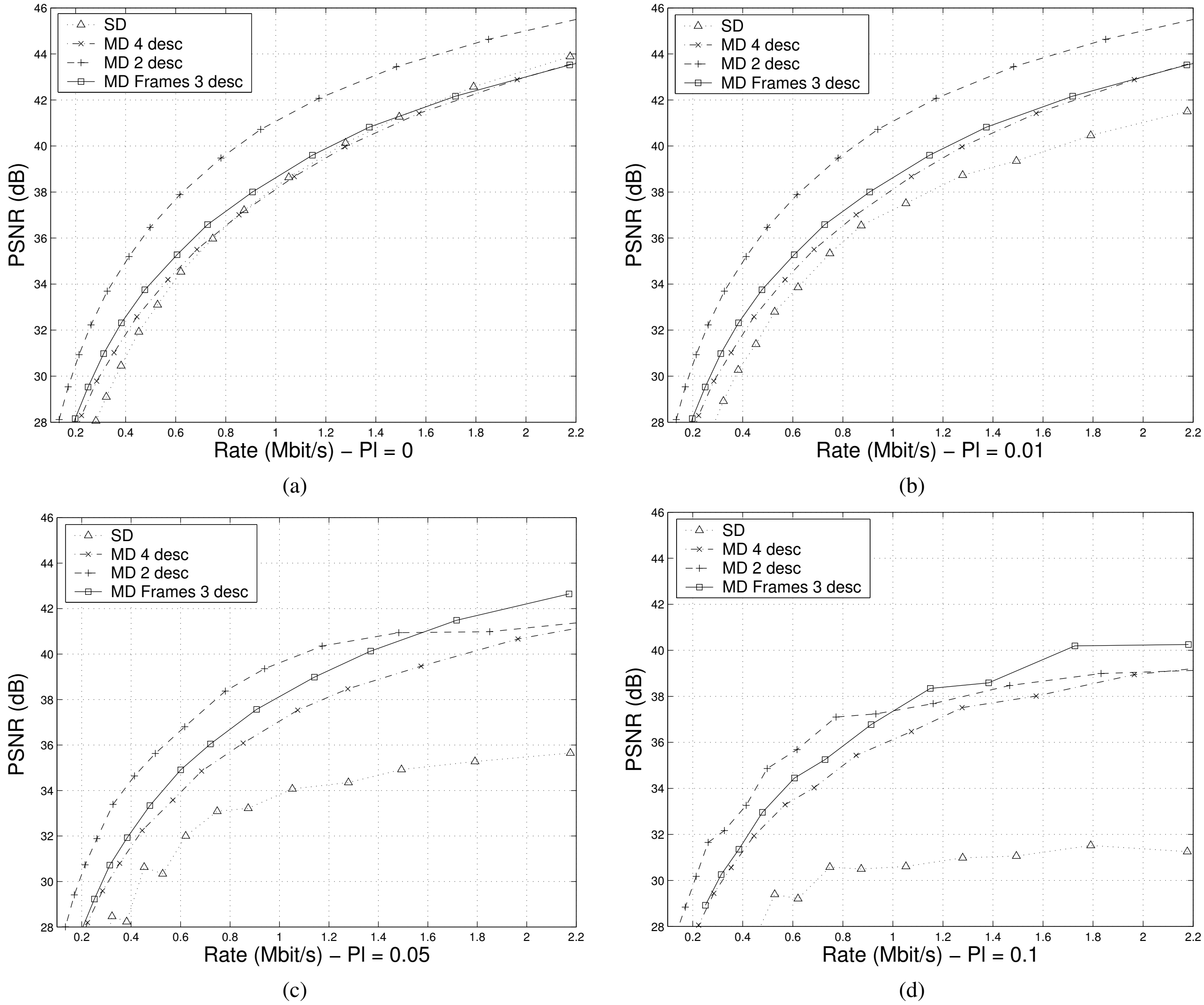


Fig. 4. Rate-distortion comparison of SD and MD schemes for different values of P_l.

References

[1] S. D. Servetto, V. A. Vaishampayan, and N. J. A. Sloane, "Multiple description lattice vector quantization," in *Proc. IEEE Data Compression Conference*, (Snowbird, UT), pp. 13–22, IEEE, Mar. 1999.

[2] Y. Wang, M. T. Orchard, and A. R. Reibman, "Multiple description image coding for noisy channels by pairing transform coefficients," in *Proc. IEEE Workshop on Multimedia Signal Processing*, (Princenton, NJ), pp. 419–424, June 1997.

[3] V. K. Goyal and J. Kovačević, "Generalized multiple description coding with correlating transforms," *IEEE Transactions on Information Theory*, vol. 47, Sept. 2001.

[4] A. E. Mohr, E. A. Riskin, and R. E. Ladner, "Unequal loss protection: Graceful degradation of image quality over packet erasure channel through forward error correction," *IEEE Journal on Selected Areas in Communications*, vol. 18, pp. 819–828, June 2000.

[5] V. K. Goyal, M. Vetterli, and N. T. Thao, "Quantized overcomplete expansions in R^N: Analysis, synthesis and algorithms," *IEEE Transactions on Information Theory*, vol. 44, pp. 16–31, Jan. 1998.

[6] X. Yang and K. Ramchandran, "Optimal subband filter banks for multiple description coding," *IEEE Transactions on Information Theory*, vol. 46, pp. 2477–2490, Nov. 2000.

[7] P. Dragotti, J. Kovačević, and V. K. Goyal, "Quantized oversampled filter banks with erasures," in *Proc. IEEE Data Compression Conference*, (Snowbird, UT), pp. 173–182, Mar. 2001.

[8] R. Bernardini, M. Durigon, and R. Rinaldo, "Low-delay reconstruction of punctured frame-coded streams," in *Asilomar Conference on Signals, Systems, and Computers*, (Pacific Grove, CA), Nov. 2003.

[9] Joint Video Team of ITU-T and ISO/IEC JTC 1, "Draft ITU-T Recommendation and Final Draft International Standard of Joint Video Specification (ITU-T Rec. H.264 — ISO/IEC 14496-10 AVC)." Joint Video Team (JVT) of ISO/IEC MPEG and ITU-T VCEG, JVT-G050, Mar. 2003.

[10] I. Daubechies, *Ten Lecturs on Wavelets*. Philadelphia: SIAM, 1992.

[11] T. Stockhammer, M. M. Hannuksela, and T. Wiegand, "H.264/AVC in wireless environments," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 13, no. 7, 2003.

Proceedings of International Symposium on Wireless Personal Multimedia Communications 2004, WPMC 04
Abano Terme, Italy - September 12-15, 2004

Vol. 2, pp. 529-533