UNIVERSITY
OF TRENTO

DEPARTMENT OF INFORMATION AND COMMUNICATION TECHNOLOGY

38050 Povo – Trento (Italy), Via Sommarive 14
http://www.dit.unitn.it

A SYNTAX-BASED APPROACH TO MULTIPLE DESCRIPTION VIDEO CODING

N. Conci, F. De Natale

December 2005

Technical Report DIT-05-070

A SYNTAX-BASED APPROACH TO MULTIPLE DESCRIPTION VIDEO CODING

Nicola Conci and Francesco G. B. De Natale

Dept. of Information and Communication Technology
University of Trento (Italy)
Contacts: conci@dit.unitn.it, denatale@ing.unitn.it

ABSTRACT

The paper describes an efficient method to achieve Multiple Description Coding (MDC) for video streaming applications. The idea is to use a JPEG-like syntax to represent an ordered sequence of the coefficients, alternately distributed among the two descriptors. The ordering procedure allows an efficient interpolation of the missing coefficients, and then an almost seamless reconstruction from a single descriptor. As compared to other approaches to MDC, the proposed strategy is characterized by an extremely low computational burden, while ensuring comparable overhead and good quality of single-descriptor video. We will consider the basic scheme, where only two descriptions are taken into account, bearing in mind that an extension to more descriptions is feasible with small effort, thanks to the extremely light nature of the encoder and decoder structure.

1. INTRODUCTION

The applications of internet video streaming are continuously growing due to many concurrent reasons, including the increasing diffusion of broadband networks (both wired and wireless), the reduction of connection costs, the growing awareness of users on the available technologies, the industrial push on the market of audio-visual systems. In this context, video streaming technologies are the key factor, for they determine the performance of the applications in terms of robustness, signal quality, bandwidth requirements, reliability, accessibility, or, in a single expression, quality-of-service.

One of the most challenging problems in video streaming is the need of achieving a robust transmission in the presence of unreliable networks characterized by high bit-error rates, strong delay jitters, packet losses. In particular, the streaming protocol should be able to guarantee a seamless decoding process even in the presence of large packet delays or losses, avoiding decoding errors, de-synchronization, and crashes. Rate control mechanisms can be effective solutions for point-to-point applications, but they turn out to be almost unfeasible in many practical cases. In particular, problems arise in multicast applications, where different users may want to access the same video content through networks and terminals with different capabilities. Furthermore, the unavailability of a return channel, typical of many streaming contexts, often hinders the possibility of implementing effective rate-control mechanisms.

In this context, the techniques known as "receiver-based rate control" can be effective solutions, making it possible for the decoder to select the quantity of information to retrieve, according to the (possibly time-varying) channel characteristics. This solution implies a suitable structuring of the video stream at the transmitter, which can be achieved in different ways including layering and MDC. The former structures the video in a base layer, representing a low-quality/low-bitrate version of the stream, and a number of enhancement layers, providing additional information to achieve a progressive increase of the quality. The latter structures the stream into two or more descriptors, individually decodable at lower quality, but providing maximum quality if decoded at once. It was demonstrated in [1] and [2] that MDC, combined with path diversity or server diversity, significantly improves the robustness of a real-time video application. An example of application of MDC to achieve path diversity is shown in Fig. 1.

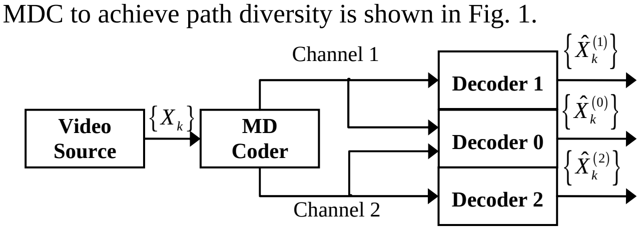

Fig. 1 Generic MDC scheme

Data is sent over two (generally *n*) independent paths and the MDC decoder must be able to reconstruct the original stream with an acceptable quality even if only one description (or a subset of *n*) is correctly received at the decoder side. The more descriptions the end user receives, the higher the quality of the decoded stream, which increases proportionally until full quality is achieved when all sub-streams are available.

In this paper we propose a new approach to MDC suitable for motion-compensated transformed coders, which are the basis of current video coding standards. The innovative characteristic of such method consists in the fact that it derives its efficiency from a careful definition of the code syntax, relaying for the generation of the different descriptors on a very simple and time-effective ordering and interpolation of coefficients. This peculiarity makes the proposed technique particularly suitable for the implementation of real-time MDC-based video streaming tools. The proposed approach has been tested on an M-JPG codec in order evaluate the performance on a very critical context, where all frames are coded in I-mode, generating a very high bitrate and at the same time a high sensitivity to packet losses. Nevertheless, it is to be pointed out that the method can be implemented on any kind of codec that employs transform coding.

The paper is structured as follows: in section 2 a brief overview on MDC is sketched, in section 3 the proposed approach is explained in detail, in section 4 experimental results are presented and discussed, while in section 5 conclusion are drawn and further directions highlighted.

2. A SHORT OVERVIEW ON MDC

The problem of describing an information source with more separate descriptions has been studied since the end of the ‘70 for several different purposes [1]. The main idea of MDC is to subdivide the information into equal sub-streams, in order to make the reconstruction possible even when only part of the original data is available. This operation has obviously a cost to pay, and the target is to introduce in each descriptor the minimum amount of *redundancy* (ρ) needed to recover the best possible visual content. Several valuable methods have been proposed for MDC. In [2], multiple state coding and path diversity is implemented: each substream has its own decoding information and is sent over different paths to achieve an “average” path behaviour at the receiver. In [3] the descriptors are defined in the Wavelet domain to achieve higher flexibility in terms of space-frequency resolution. An interesting approach [4][8] is the application of the so called “Matching Pursuit” (MP). In MP, the prediction error can be represented with a number M of basic functions (called atoms). The L most informative atoms are shared over the whole set of descriptors, while the remaining $M–L$ atoms are distributed among the descriptors. At the decoder, the sequence is reconstructed from the available atoms, according to the number of descriptors correctly received. The trade-off between quality and efficiency is decided by L, which determines the overhead. Another interesting approach was proposed by Orchard et al. in [5][6], using the so called Correlating Transforms (CT). In their proposal, two descriptors are generated by correlating pairs of DCT coefficients through a Pairwise Correlating Transform (PCT). Each pair of PCT coefficients [*C D*] is obtained by a pair of DCT coefficients [*A B*] by applying the linear matrix transform *T* in Eq. 1.

$$T = \begin{bmatrix} \sqrt{\dfrac{\cot \theta_1}{2}} & \sqrt{\dfrac{\tan \theta_1}{2}} \\ -\sqrt{\dfrac{\cot \theta_1}{2}} & \sqrt{\dfrac{\tan \theta_1}{2}} \end{bmatrix} \tag{1}$$

Then, the PCT-transformed coefficients [*C D*] are split between the two descriptors so that, whenever a description is lost, the missing coefficients can be recovered from the correlated ones through a linear prediction. The process is applied only to a part of DCT coefficient, selecting the pairs [*A B*] a-priori on the basis of their statistical properties and of the quantization step. The remaining coefficients are simply split into two groups in an odd-even manner. If both descriptions are received, the original coefficient can be restored by applying the inverse transform T^{-1} to the [*C D*] pairs and alternating the remaining coefficients. The main drawbacks of this method are in the complexity and in the overhead deriving from the association rule, which usually couples most significant coefficients to least significant ones, which are often associated to zeros (then, no cost) in the original description.

3. THE PROPOSED METHOD

The main purpose in creating a new approach to MDC was to define a scheme suitable for low-cost, real-time implementation, and with nearly standard characteristics. In this respect, we focused the attention onto two aspects: (i) a simple and reliable strategy for the distribution of the coefficients among the descriptors and the interpolation of the complementary information at the decoder, and (ii) an optimized syntactic structure of the stream able to take full advantage of the above strategy. For the sake of simplicity, the presentation of the system is targeted to two descriptors only, although the light nature of the encoding procedure makes it extensible to three or more descriptors, as we will mention in the following.

As far as the first point is concerned, the creation of each descriptor relies on a sorting and alternate selection of the quantized coefficients. The selection allows reducing the number of coefficients to be transmitted to one half, thus achieving a sharp bitrate reduction of each descriptor. The sorting operation orders the coefficients by decreasing absolute value, and is intended to ease the interpolation of the missing coefficients at the decoder when a descriptor is lost. The idea is to make the decoder in the condition of reconstructing the sorted sequence of coefficients, part of which is available and part missing. Being the sequence ordered, the missing coefficients can be successfully recovered at the decoder by a simple monotonic decreasing interpolator. The sorting information is clearly responsible for an overhead. In fact, besides the selected coefficients,

the encoder has to transmit in each elementary descriptor the position of the cancelled coefficients in the sorted array. It is therefore important that this information is efficiently represented. This leads to point (ii): two streams have to be transmitted: the complete information on the preserved coefficients and the positions and sign of the cancelled ones (the sign is needed for the ordering is performed in absolute value). The first stream is perfectly compatible with the standard case, and can be handled as in a standard encoder, while the second stream should be managed by a specific syntax and an appropriate entropy coding.
The overall scheme of the encoder is shown in Fig. 2. The details about the encoder and decoder are given in sections 3.1 and 3.2, respectively.

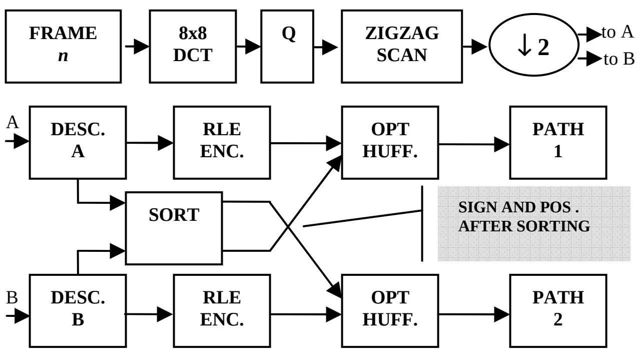

Fig. 2 the MDC scheme: the upper part is in common, the lower part represents the generation of the two descriptors.

3.1 Encoder

This section describes the exact procedure implemented to achieve the two descriptors. The quantized DCT coefficients are zig-zag scanned (Fig. 3), producing a string of 64 coefficients. Due to its importance for the reconstruction process, the DC component is embedded in both descriptors coded in DPCM. The AC coefficients are instead split in an even-odd manner (A, B coefficients in Fig. 3) and associated to the respective descriptors.

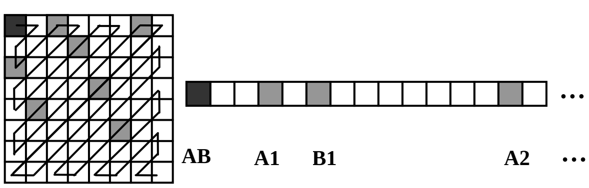

Fig. 3 Zig-zag scan and odd/even split of the DCT coefficients. The black position is the DC, while the grey ones represent the non-null AC coefficients.

According to a JPEG-like scheme, each preserved AC coefficient is then coded with a NNNN|SSSS string that represents the relevant run-level, 2D-VLC coded (Huffman), and associated to an adequate number of precision bits. Referring (without loss of generality) to descriptor A, at this point only one half of the non-null coefficients are transmitted (positions 'A'), while the remaining information should be reconstructed at the decoder. To allow interpolating the coefficients 'B', first of all we need to know their positions. Furthermore, being the DCT coefficients highly uncorrelated, some additional information is needed to achieve a meaningful interpolation. In the proposed scheme, such additional information refers to the position of the skipped AC coefficients in a list sorted in descending order. The above information can be again represented by an 8-bit string of the type: NNNN|S|PPP, where NNNN represents the RLE position of the skipped coefficient, S its sign, and PPP its position. The sequence of such 8-bit codewords is then 3D-VLC encoded with Huffman and represents the overhead of descriptor A. Since the three bits PPP allow encoding up to 8 positions, the ordering and encoding of skipped coefficients is performed only on the first 8 non-null AC coefficients in zig-zag order, completely discarding B coefficients from the 10-th position on. Due to the lowpass nature of image blocks, this is usually sufficient to preserve all the significant information in the block. With reference to the example of Fig. 3, Fig. 4 shows an example of the encoding of descriptor A.

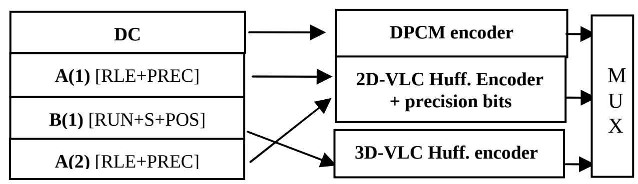

Fig. 4 Example of encoding of descriptor A

3.2 Decoder

If both descriptors A and B are correctly received, the decoder can discard the overhead information and reconstruct the original DCT block by merging the odd and even coefficients. After the merge, the block is completely JPEG-equivalent and then a standard decoder can be applied. On the other side, if one description is lost during the transmission, the decoder should exploit the full information of the single descriptor received. This process is sketched in Fig. 5 for descriptor A.

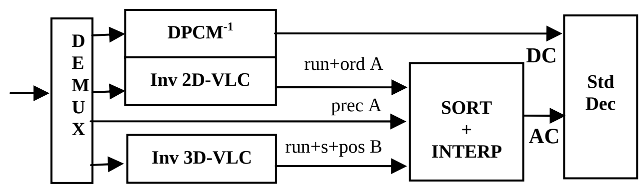

Fig. 5 Scheme of the decoder for descriptor A

The DC information is directly available at the decoder, as well as the position and value of the AC coefficients in positions 'A'. The position and sign of up to 4 AC coefficients in position 'B' can be recovered from the overhead information, together with their ordered position in the array of first (up to) 8 coefficients sorted by descending magnitude.
This allows constructing an ordered sequence, where the B coefficients are represented by missing points at known

positions. According to the typical rapid decay of the coefficients magnitude, a cubic spline curve can be used to interpolate them. Nevertheless, it has been experimentally verified that the use of a simple linear interpolation achieves comparable results with lower computation. Fig. 6 shows an example of linear interpolation of the missing coefficients.

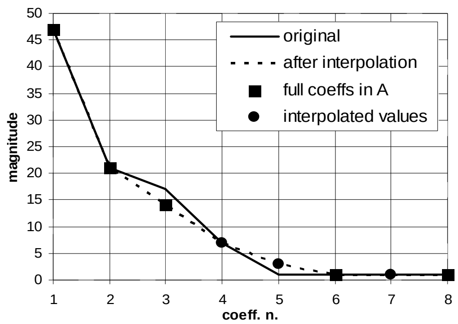

Fig. 6: Interpolation of the missing coefficients

4. EXPERIMENTAL RESULTS

In order to validate the proposed technique, a large number of standard sequences have been tested. The proposed MDC scheme has been inserted in a standard M-JPEG encoder in order to assess the average performance on large sequences of I-frames. The results have been then evaluated in terms of average PSNR and overhead, comparing the standard M-JPEG with the reconstruction achieved from a single descriptor (the sum of the two descriptors being equal to M-JPEG). Table 1 reports the results achieved on three of the adopted test sequences: "*Claire*", "*Mobile & Calendar*", and "*Paris*". It can be observed that the average quality loss for a single descriptor is around 3.5 dB, with an overhead lower than 30%.

Table 1. Simulation results for three test sequences

	M-JPG	proposed: 1 descr.	
	PSNR	PSNR	ρ
Claire	42.8 dB	38.8 dB	29.2%
Mobile	38.6 dB	35.1 dB	26.6%
Paris	39.5 dB	36.1 dB	27.2%

In Fig. 7, a reconstructed sample frame of the sequence "Paris" is shown, to allow a visual evaluation of the degradation. The artifacts are mainly related to a slight increase of the ringing effect near object contours.

A comparison has also be made with the very efficient MDC approach proposed in [5,6]. To this purpose, the MDC scheme has been embedded in a H.263 codec and tested on the well known sequence "*carphone*" at 200kbps. The results (Table 2) demonstrate that the two methods have a similar behavior in terms of quality/overhead, the PCT having the disadvantage of a higher computational burden and implementation complexity.

Table 2. Comparison with the PCT approach

	PCT		proposed method	
	ρ	PSNR	ρ	PSNR
Carphone	47%	27.17	48%	28.1

5. CONCLUSIONS

A novel approach to MDC coding has been proposed suitable for video streaming systems using path or server diversity. The main goal is to achieve a real-time, computationally effective procedure able to guarantee a good trade-off between quality and overhead, using a nearly-standard syntax. Experimental results highlight that the proposed approach succeeds in limiting the increase of the distortion, with a redundancy always below 30%. Further studies are being carried on to demonstrate the efficiency of the method for three or more descriptions.

Fig. 7: Detail from the test sequence *Paris*. Left: M-JPG, right: single description reconstruction.

6. REFERENCES

[1] J.K. Wolf, A.D. Wyner, and J. Ziv, "Source coding for multiple descriptions," Bell Syst. Tech. J., vol. 59, no. 8, pp. 1417-1426, Oct. 1980.

[2] J. G. Apostolopoulos, "Reliable video communication over lossy packet networks using multiple state encoding and path diversity," Proc. VCIP, San Jose, USA, pp. 392–409, Jan. 2001

[3] Xuguang Yang, K. Ramchandran, "Optimal subband filter banks for multiple description coding," IEEE Trans. On Information Theory, Volume: 46 , Issue: 7 , Nov. 2000, 2477 – 2490

[4] R. Neff and A. Zakhor, "Very low bit-rate video coding based on matching pursuits," IEEE Trans. Circuits Syst. Video Technol., vol. 7, pp. 158–171, Feb. 1997.

[5] M. T. Orchard, Yao Wang, V. Vaishampayan, A. R. Reibman, "Multiple Description Coding Using Pairwise Correlating Transforms," IEEE Trans. On Image Processing, Vol. 10, No. 3, March 2001

[6] Y. Wang, M. Orchard, and A. Reibman, "Optimal pairwise correlating transforms for multiple description coding," in Proc. IEEE Int. Conf. Image Processing (ICIP'98), Chicago, IL, Oct. 1998.

[7] A. Reibman, H. Jafarkhani, Y. Wang, M. Orchard, and Rohit Puri, "Multiple-Description Video Coding Using Motion-Compensated Temporal Prediction," IEEE Trans. Circuits Syst. Video Technol., vol. 12, no. 3, March 2002.

[8] Xiaoyi Tang, Avideh Zakhor, "Matching Pursuits Multiple Description Coding for Wireless Video," IEEE Trans. Circuits Syst. Video Technol., vol. 12, no. 6, June 2002.