Tree-Structured Multiple Description Coding for Multiview Mobile TV and Camera-Phone Networks

Yongkai Huo and Lajos Hanzo
School of ECS, University of Southampton, SO17 1BJ, United Kingdom.
Tel: +44-23-8059 3125, Fax: +44-23-8059 4508
Email: {yh3g09,lh}@ecs.soton.ac.uk, http://www-mobile.ecs.soton.ac.uk

***Abstract*— Since multiview video communications facilitate the selection of several camera views of a given scene, it may be deemed to be a promising technique for mobile television and camera-phone networks. Similar to conventional single-view video, multiview video also suffers from packet loss events imposed by network congestion. Multiple description coding (MDC) constitutes an attractive candidate solution for mitigating the packet loss events in conventional video communications. Hence it may also be deemed to be an attractive solution for multiview video. In this paper, we propose a novel tree-structured MDC (T-MDC) scheme, which can be readily implemented with the aid of arbitrary video codecs, including multiview codecs. Moreover, the proposed philosophy enables the flexible creation of a variable number of unequal importance descriptions. Finally, we will conceive a joint multiple description coding and multiview video coding (MVC) scheme. We compare two specific tree-structured MDC schemes, namely a so-called height-one complete tree-structured MDC (HOCT-MDC) and a binary tree-structured MDC (BT-MDC) scheme, which we benchmarked against simulcasting aided single description coding in the context of both single and multiview video.**

I. Introduction

Multiple description coding (MDC) [1], [2] was proposed for overcoming the effects of channel impairments inflicting packet loss events in applications, where the employment of classic forward error correction (FEC) becomes inefficient. Hence MDC is particularly beneficial in the Internet, since retransmissions cannot be readily used in real-time interactive telephony. In MDC, the source is encoded into multiple descriptions, which may be transmitted to the receiver via multiple TCP-IP routes. When all the descriptions have been received, the receiver becomes capable of reconstructing a high quality replica of the source signal. When some of the descriptions are lost due to network congestion, the receiver still remains capable of reconstructing an acceptable quality of the source signal. Furthermore, MDC may also be potentially employed in wireless applications, such as video transmissions in cooperative networks [3] using relays for providing a diversity gain in order to combat channel errors.

Since multiview video communications facilitate the selection of several camera views of a given scene, it may be deemed to be a promising technique also for mobile multiview television. Multiview video coding has been developed for more than 20 years [4], [5] and a number of coding standards appeared, for example the multiview profile based on MPEG-2 [6]. Recently, the Joint Video Team (JVT) proposed multiview video coding as an amendment to H.264/AVC [7]. Similar to conventional single-view video, multiview video also suffers from packet loss events imposed by network congestion. Hence it is necessary to design techniques for multiview video communications in order to combat the packet loss events. Again, MDC [1], [2] has been introduced to overcome the deleterious effects of channel errors with the aid of source encoding diversity. More explicitly, the source may be encoded into several correlated representations/descriptions. Any subset of these descriptions may be independently decoded at the receiver. MDC may be deemed to be an attractive solution for multiview video streaming over unreliable networks, such as the Internet. However, there is a paucity of solutions on the topic. Before illustrating the proposed scheme, let us continue by reviewing the existing MDC solutions.

The financial support of the EPSRC under the auspices of the China-UK Science Bridge and that of the RC-UK under the India-UK Advanced Technology Centre (IU-ATC) is gratefully acknowledged.

The authors of [8] proposed a MDC system, where the input source signal is first decomposed into two subsources using a polyphase transform. Each of the subsources is quantized independently by a quantizer. Then each of the descriptions is multiplexed with the coarsely quantized version of the other subsource. Hence each description carries information about the other one, which may be used to combat packet loss events at the receiver. However, this method may become excessively complex, when more descriptions are required and it is not readily compatible with standardized video codecs, such as MPEG-4, H.264 etc [9].

Set Partitioning relying on Hierarchical Trees (SPIHT) [10] constitutes an image coding technique that is based on the embedded zerotree wavelet (EZW) algorithm, which organizes the wavelet coefficients in a so-called impartial orientation tree structure. In [11], Kim *et al.* proposed a 3-D SPIHT (3D-SPIHT) algorithm for multiple description coding. Then in [12], the authors propose a modified tree structure for 3D-SPIHT that is more efficient for employment in MD coding. They also designed a branch-pruning technique for generating multiple descriptions. However, this technique cannot be readily combined with multiview video codecs.

An odd-even frame separation based MDC codec designed for stereoscopic video communications was proposed in [13], where the odd and even indexed frames were encoded separately into two streams. At the receiver, each frame may be predicted by interpolation techniques applied in case of packet-loss events.

Due to the time-lag amongst frames, the odd-even frame-index based methods may be viewed as employing temporal down-sampling. In contrast to the above-mentioned time-domain down-sampling, a spatial-domain down-sampling technique was proposed in [14]. The authors proposed two MDC schemes, namely a so-called drift-compensation based multiple description video codec (DC-MDVC) and independent flow-based multiple description video codec (IF-MDVC). The DC-MDVC was restricted to the employment of only two descriptions, while the IF-MDVC philosophy was more flexible.

Against this background, in this paper, we propose a novel tree-structured multiple description codec (T-MDC), which is flexible and may be combined with arbitrary video codecs. The technique advocated splits the original video signal into a pre-set number of correlated descriptions in time-domain, while retaining the correlation among the video frames within each description. Furthermore, our proposed scheme is also capable of splitting the video stream into multiple descriptions of unequal importance.

This rest of this paper is organized as follows. In Section II, we introduce a new tree-structured creation of multiple video descriptions. Section III presents the corresponding reconstruction algorithm designed for both lossless and lossy video compression. The framework of applying the proposed T-MDC to MVC is described in more detail in Section IV. The performance of two T-MDC structures is characterized in Section V in the context of both conventional and multiview video communications. Finally, we offer our conclusions in Section VI.

978-1-4244-8325-9/11/$26.00 ©2011 IEEE

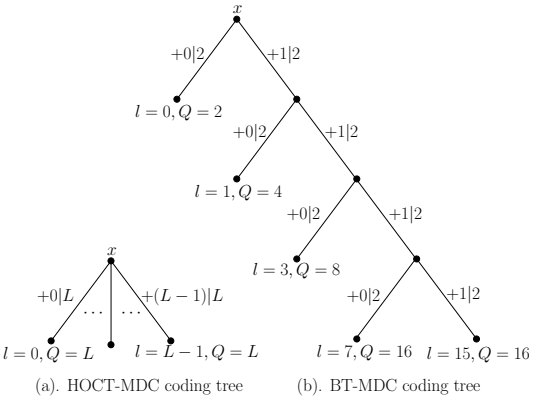
II. Tree-Structured Multiple Description Creation

It is widely recognized that the addition of a constant luminance level to a video signal does not affect the correlation among video frames. In this study, we consider gray-scale video sequences, but the proposed technique may be readily extended to the color *YUV* or *RGB* video formats. Let us commence our discourse by illustrating a simple method of expressing a single pixel value (positive integer) with the aid of multiple integers using an example. Let us assume that the pixel value equals x, which may be expressed as

$$\{x|L, (x+1)|L, (x+2)|L, \cdots, (x+L-1)|L\}, \tag{1}$$

where $|$ represents the aliquot part upon division by L, which may also be viewed as the action of quantization. For example, if we have $x = 50, L = 4$, x may be expressed as $\{12, 12, 13, 13\}$. Provided that all the integers $\{12, 12, 13, 13\}$ are known, the original value can be readily recovered.

Based on Eq. (1), let us now discuss the method of creating multiple video descriptions, which may be of practical importance in numerous applications, including both equal and unequal importance descriptions. For example, equal-importance descriptions may display different camera-angles in interactive TV [15] or the left-eye/right-eye views of stereoscopic video [16]. By contrast, less important descriptions-containing for example a low-resolution version of a scene may be dropped during instances of network-congestion [1]. For a $(c \times r)$-pixel luminance video frame, each pixel may be expressed as $x_{i,j}, 0 \leq i < c, 0 \leq j < r$. Let us assume that we want to describe the video clip using L descriptions. By applying Eq. (1) to each pixel of a frame, we can generate a specific description of the frame for each offset l $(0 \leq l < L)$. Each pixel of the l^{th} description corresponding to L in the temporal domain is formulated as $(x_{i,j} + l)|L$. Now we have L correlated descriptions of the video frame, each of which carries a quantized but equal-importance version of the original video, while also containing independent high frequency information. We may now treat the original video signal as the root of the corresponding coding tree, where the different but equal-importance descriptions are the leaves of the coding-tree. We may refer to the coding-tree as the height-one complete tree (HOCT), which is exemplified in Fig.1 (a).



(a). HOCT-MDC coding tree (b). BT-MDC coding tree

Fig. 1. Examples of proposed coding-tree, where $+l|Q$ means the addition of l to each video pixel and then its quantization with Q.

In the coding-tree generation procedure, only two operations, namely the addition of a DC component and the quantization are required. The higher the value of L, the more of the originally different pixels become identical after retaining the aliquot part, which results in a more correlated sequence than the original one, because the smaller pixel-differences representing the high-frequency components disappear after retaining the aliquot part. Hence the resultant aliquot part becomes more amenable to compression. More explicitly, this property facilitates the encoding of these correlated aliquot-based descriptions into bitstreams using any existing lossy or lossless compression schemes, including standardized video codecs.

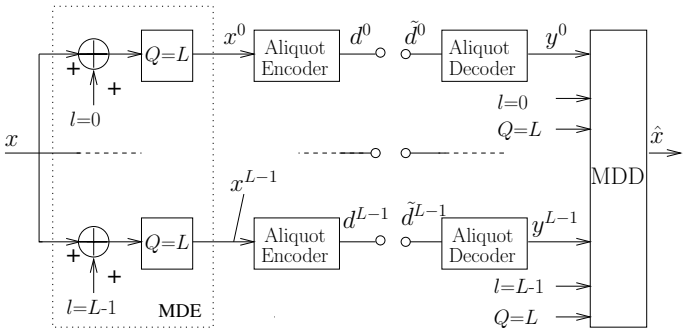


Fig. 2. The proposed HOCT-MDC architecture. The *Aliquot Encoder* and *Aliquot Decoder* represent the codecs employed to encode and reconstruct the aliquot part descriptions, $d^0 \cdots d^{L-1}$ represent the generated bitstreams, while *MDD* represents the multiple descriptions decoder for video reconstruction.

More specifically, L different video descriptions resulting in L compressed bit streams may be generated independently using for example lossy video compression schemes, such as the MPEG-2 and H.264 codecs [9]. The proposed codec architecture is displayed in Fig. 2, where the offsets $l = 0, \cdots, L-1$ and quantizers resolutions $Q = L, \cdots, Q = L$ may be associated with any specific coding-tree structure. After the aliquot encoding stage of Fig. 2, the L bitstreams having indices of $l = 0, \cdots, L-1$ may be transmitted to the receiver via multiple wireless channels or TCP-IP routes. Provided that no packets are lost during transmission, L descriptions will be received at the receiver. By contrast, less than L descriptions may be received in the presence of packet erasures imposed by route-congestion or channel-errors. To simplify our discourse, let us assume that all the L bitstreams are received at the receiver. Then the receiver has to recover the original video, namely x in Fig. 2, from the L bitstreams. Two reconstruction stages are involved in the recovery procedure of Fig. 2, namely the reconstruction of the L aliquot descriptions from the L received bitstreams, followed by the reconstruction of the original video x from the reconstructed L aliquot descriptions. The aliquot reconstruction may be accomplished by the decoder pair of the aliquot encoder. The second reconstruction stage, namely the reconstruction of the original video clip will be illustrated in great detail in Section III.

Again, the above method generates multiple descriptions of equal importance. However, in many practical multi-rate transceivers [17] or multi-route Internet based applications, it may be beneficial to transmit different-importance descriptions, so that the least important ones may be dropped in case of low instantaneous channel qualities [9] or network-congestion. Our scheme is also capable of generating descriptions of different importance by repeatedly and hierarchically splitting any specific subset of the L descriptions. Specifically, if we only employ the quantizer value of $L = 2$, we generate a binary-tree based MDC (BT-MDC), which is exemplified in Fig. 1 (b). Theoretically any specific tree structure of descriptions may be generated. Using the proposed method, we can then remember the description offsets l and the quantizer resolution Q along the path from the root to any specific leaf of the coding tree. Finally, we can combine all [offset, quantizer]=$[l, Q]$ pairs into a single parameter, as exemplified in Fig. 1 (b). This implies that generating a more complex coding-tree structure only modestly increases the encoding complexity. Furthermore, by simply assigning $[l, Q]$ pairs of Fig. 2 to the [offset, quantizer] parameter pair of the leaves seen in Fig. 1 (b), we can readily generate the corresponding architecture for the coding-tree of Fig. 1 (b).

III. Video Reconstruction

This section outlines the reconstruction of the original video clip. For simplicity, we consider the reconstruction algorithm of the Multiple Description Encoder (MDE) of Fig. 2. Other tree-structured MDC schemes may be readily reconstructed by simply

modifying the reconstruction parameters of the MDE considered. Here we will commence by illustrating a simple reconstruction technique conceived for lossless coding. However, in practical video systems typically "lossy" compression is used. Hence Section III-B illustrates further "lossy" reconstruction techniques required for practical applications. Before detailing the reconstruction methods, we stipulate the following assumptions:

- $x_{i,j}$: the pixel value at position (i, j) in a specific video frame;
- L: the quantizer resolution;
- $S_L = \{0, \cdots, L-1\}$: the set of descriptions received at the receiver, where each element corresponds to an offset invoked for creating a specific description;
- $X_m = \{0, \cdots, 2^m - 1\}$: the set of m-bit luminance pixel values;
- $x_{i,j}^l$: the aliquot part of the pixel $x_{i,j}$ upon division by L in the l^{th} description, $l \in S_L$;
- $d^l, \tilde{d}^l$: the l^{th} bitstream at the transmitter and receiver respectively, $l \in S_L$;
- $y_{i,j}^l$: the l^{th} aliquot part of the pixel $x_{i,j}$ reconstructed by the aliquot decoder at the receiver, $l \in S_L$;
- $\hat{x}_{i,j}$: the estimated pixel value at position (i, j);

Based on the above notations, in an entire $(c \times r)$-pixel video frame of the l^{th} $(l \in S_L)$ description, each pixel at position (i, j) may be expressed in aliquot part form as $x_{i,j}^l$. Recall from Fig. 2, that the l^{th} description consisting of a sequence of video frames with $c \times r$ aliquot pixels, will be encoded using either lossy or lossless aliquot encoders and the corresponding bitstream $d^l, l = 0, \cdots, L-1$ will be generated. Then, the L bitstreams $d^0, \cdots, d^{L-1}$ may be transmitted to the receiver through a number of wireless channels or TCP-IP routes. At the receiver, the L received bitstreams $\tilde{d}^0, \cdots, \tilde{d}^{L-1}$ must be decoded using the aliquot decoder in order to reconstruct the L aliquot descriptions. In an entire frame of the l^{th} $(l \in S_L)$ aliquot description reconstructed by the aliquot decoder, each reconstructed pixel at position (i, j) may be expressed as $y_{i,j}^l$. Then, for each original pixel we obtain L aliquot parts $y_{i,j}^l, l \in S_L$ at the receiver, which we will decode for recovering the original pixel $x_{i,j}$ as detailed in the following two sections.

Note that here we continue our discourse based on the assumption that all descriptions $S_L = \{0, \cdots, L-1\}$ are received at the receiver, albeit the reconstruction procedure may rely on any subset of the full set S_L.

A. Reconstruction for Lossless Aliquot Encoding

Naturally, the original video sequence may be recovered from any of its L descriptions, but an improved video quality may be expected upon beneficially combining several descriptions. This will be further detailed below. Upon Eq. (1), for each received description $l \in S_L$, we arrive at:

$$y_{i,j}^l \cdot L - l \leq \hat{x}_{i,j} < (y_{i,j}^l + 1) \cdot L - l. \tag{2}$$

Provided that the transmitted aliquot part is perfectly received, namely $y_{i,j}^l = x_{i,j}^l$, then we have

$$\hat{x}_{i,j} \begin{cases} < \min\left\{(y_{i,j}^l + 1)L - l | l \in S_L\right\} \\ \geq \max\left\{y_{i,j}^l L - l | l \in S_L\right\}. \end{cases} \tag{3}$$

When some of the descriptions become unavailable owing to channel-induced packet loss events, we may simply choose the average of the available pixel values for the estimated value of $\hat{x}_{i,j}$.

B. Reconstruction for Lossy Aliquot Encoding

In case of lossy compression and error-infested channel decoded scenarios, the aliquot part obtained at the receiver after channel decoding and aliquot reconstruction may be expressed as $y_{i,j}^l = x_{i,j}^l + \delta$, where the reconstruction error δ is introduced by the aliquot decoder and channel decoder. Here we ignore the effects of transmission errors for simplicity. Then δ is solely the lossy aliquot decoder's reconstruction error. We now have to recover the original video pixel $x_{i,j}$ based on the received aliquot parts $y_{i,j}^0, \cdots, y_{i,j}^{L-1}$. Below, we will now introduce the direct mathematical formulation of reconstructing the original pixel $x_{i,j}$, which depends on the probability of $x \in X_m$ conditioned on the aliquot parts $y_{i,j}^0, \cdots, y_{i,j}^{L-1}$ reconstructed by the aliquot decoder of Fig. 2. Since we will focus our attention on a single pixel here, we simplify our notations by treating a pixel value without its position index, i.e. we use x^l instead of $x_{i,j}^l$. Furthermore, the notation y_0^{L-1} represents $y_{i,j}^0, \cdots, y_{i,j}^{L-1}$.

Let us now assume that we have received all the aliquot values y_0^{L-1} from L different reconstructed aliquot descriptions at a given pixel position, although as noted above, even a single description is sufficient for adequately reconstructing the original video sequence. Naturally, having multiple descriptions is expected to improve the reconstructed video quality. Based on these received aliquot values, the original pixel value x can be recovered using for example either

- the MMSE estimation rule of

$$\hat{x} = \sum_{x \in X_m} x \cdot p\left(x | y_0^{L-1}\right), \tag{4}$$

- or the MAP estimation rule of

$$\hat{x} = \arg\max_{\forall x \in X_m} p\left(x | y_0^{L-1}\right). \tag{5}$$

Based on Bayes' theorem and on the chain rule of probability, the *a-posteriori* probability of occurrence $p\left(x | y_0^{L-1}\right)$ in Eq. (4) and Eq. (5) may be formulated as follows

$$p\left(x | y_0^{L-1}\right) = \frac{p\left(y_0^{L-1} | x\right) \cdot p\left(x\right)}{\sum\limits_{u \in X_m} p\left(y_0^{L-1} | u\right) \cdot p\left(u\right)}. \tag{6}$$

Furthermore, let us define the aliquot reconstruction error $\delta^l = y^l - x^l$ of the l^{th} description, which is solely introduced by the aliquot codec, since the channel effects are ignored. Then the PDF $p\left(y_0^{L-1} | x\right)$ in Eq. (6) may be formulated as follows

$$\begin{aligned} p\left(y_0^{L-1} | x\right) &= p\left(y^0, \cdots, y^{L-1} | x\right) \\ &= p\left(\delta^0, \cdots, \delta^{L-1} | x\right) = p\left(\delta_0^{L-1} | x\right). \end{aligned} \tag{7}$$

Let us now discuss the calculation of the joint probability $p\left(\delta_0^{L-1} | x\right)$ in Eq. (7). Let us consider a simple video codec comprised of *a quantizer* and a *bit mapper* as the aliquot codec. Then the aliquot reconstruction error $\delta^l, l \in S_L$ arises solely due to the *quantizer*. The aliquot reconstruction errors $\delta^0, \cdots, \delta^{L-1}$ are independent of each other, when the original pixel x is given. When a more complex aliquot codec is employed, such as H.264, the aliquot reconstruction errors δ_0^{L-1} introduced by the aliquot codec may be deemed to be independent of each other. Hence their joint probability is given by the product of the individual probabilities, usually by

$$p\left(\delta_0^{L-1} | x\right) = p\left(\delta^{L-1} | \delta_0^{L-2}, x\right) \cdots p\left(\delta^0 | x\right) = \prod_{l \in S_L} p\left(\delta^l | x\right). \tag{8}$$

Upon combining Eq. (6) , 7 and (8), the *a-posteriori* probability of pixel x conditioned on all the L reconstructed aliquot parts y_0^{L-1} may be expressed as

$$p(x | y_0^{L-1}) = \frac{\prod\limits_{l \in S_L} p(\delta^l = y^l - x^l | x) \cdot p(x)}{\sum\limits_{u \in X_m} p(u) \cdot \prod\limits_{l \in S_L} p(\delta^l = u^l - x^l | u)}, \tag{9}$$

where $p(\delta^l | x)$ is the distribution of the reconstructed aliquot part error conditioned upon the pixel value x, while $p(x)$ is the distribution of the original pixels. Eq. (9) can then be used for video reconstruction from the L descriptions, provided that the two PDFs $p(\delta^l | x)$ and $p(x)$ are known at the receiver. In practice, these PDFs have to be evaluated for a representative video training sequence and stored at the receiver.

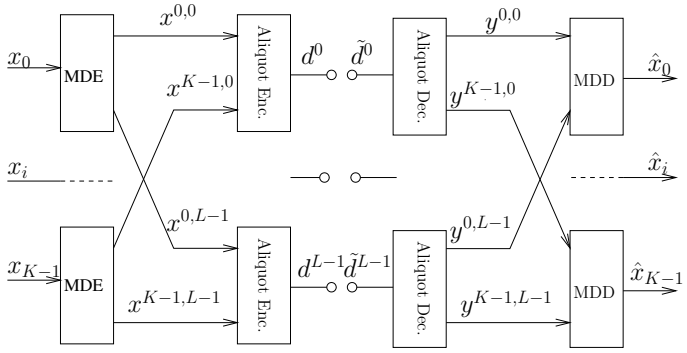


Fig. 3. Framework of multiview coding with tree-structured-MDC. *MDE* represents a multiple description encoder, while *MDD* represents its multiple description decoder pair.

IV. Multiple Description Coding of Multiview Video

In this section, we employ the proposed HOCT-MDC architecture of Fig. 2 for multiview video transmission. As shown in Fig. 3, K different camera-views are input to the system and L description streams are created. In order to encode each of the K video input streams considered using the MDE of Fig. 2, K HOCT-MDC encoders are employed. Each of the K input views generates L aliquot descriptions, as seen in Fig. 3. After the HOCT-MDC encoder stage, a total of $(K \times L)$ aliquot part descriptions are created, which may be grouped into L number of K-aliquot camera views based on their *offset* l and *quantizer* resolution L. The grouping seen in Fig. 3 encodes the K correlated camera-view jointly and hence it is expected to achieve a certain compression. These L K-aliquot camera views are then input to L multiview video encoders, namely to the *Aliquot Enc.* of Fig. 3, each of which may have different offsets and quantization parameters. Each of the L K-aliquot encoders will generate a bitstream independently, each of which may be transmitted via different wireless-channels or TCP-IP routes to the receiver.

The receiver employs L multiview video decoders, namely the *Aliquot Dec.* blocks of Fig. 3, each of which reconstructs the K-aliquot descriptions from the L received bitstreams $\tilde{d}_0, \cdots, \tilde{d}_{L-1}$, as seen in Fig. 3. After the aliquot reconstruction stage, a total of $(K \times L)$ aliquot part descriptions are reconstructed. These $(K \times L)$ descriptions may be grouped into K L-aliquot part groups, which represent the K original camera views, that are then input to K multiple description decoders (MDD), each of which will reconstruct one of the original camera views $\hat{x}_l, 0 \leq x < K$.

V. Performance Results

This section evaluates the performance of both the proposed T-MDC (HOCT-MDC, BT-MDC) scheme as well as that of the entire system using MVC invoking the T-MDC. Simulcasting of several single descriptions (SD)[1] [18], [19] provides the lower bound of the achievable MDC performance and may always be employed for practical video streaming applications to transmit the same bitstream via different channels and routes in order to combat the packet loss events imposed by network congestion. In this section, we compare our system's performance to simulcast-SDC. Note that if a quantizer resolution of $Q = L = 1$ is used, then multiple description coding degenerates to conventional single description coding (SDC). We will employ the H.264 video codec for aliquot compression and the quantizer parameter (*QP*) values used in this section are those of the H.264 standard [7]. Furthermore, the peak signal-to-noise ratio (PSNR) of the luminance computed is used to quantify the video quality. In order to simplify our discourse, let us now introduce the notation G, denoting the total number of descriptions generated, as well as G_a denoting the number the descriptions available at the output of the channel.

[1]Here in simulcast-SDC, the original video is encoded into a conventional single bitstream. Then multiple duplicated copies of the single description bitstream are transmitted simultaneously to the receiver via multiple routes.

A. Multiple Description Codec Performance

In this section, we characterize the performance of our proposed MDC schemes. In all the simulations the 45-frame *Akiyo* video sequence in (176×144)-pixel quarter common intermediate format (QCIF) was input to the HOCT-MDC codec. The JM/AVC 17.2 H.264 scheme was used for encoding the aliquot part descriptions into bitstreams. Furthermore, the Intra-frame (*I*) refresh period was set to 15 and both the predicted (*P*) and bidirectional (*B*) frames were enabled. The scanning rate expressed in frame per second (FPS) was set to 15. These parameters jointly determine the bitrate.

Firstly, the comparison among the rate-distortion performances of the proposed HOCT-MDC, BT-MDC and the simulcast-SDC-H.264 lower bound is portrayed in Fig. 4. Upon increasing the bitrate, HOCT-MDC increasingly outperforms simulcast-SDC-H.264, albeit naturally, its performance saturates at the Y-PSNR upper bound. Observe in Fig. 4 that the BT-MDC scheme outperforms both simulcast-SDC-H.264 using $G = 5$ by about 4 dB at a bitrate of 200 kbps and HOCT-MDC using $L = 4$ by 2.5 dB at a bitrate of 125 kbps. The reason for the superiority of BT-MDC is that the BT-MDC scheme generates aliquot part descriptions of unequal importance, where the less important descriptions carry less high-frequency information. This property statistically decreases the average correlation amongst all the BT-MDC aliquot descriptions.

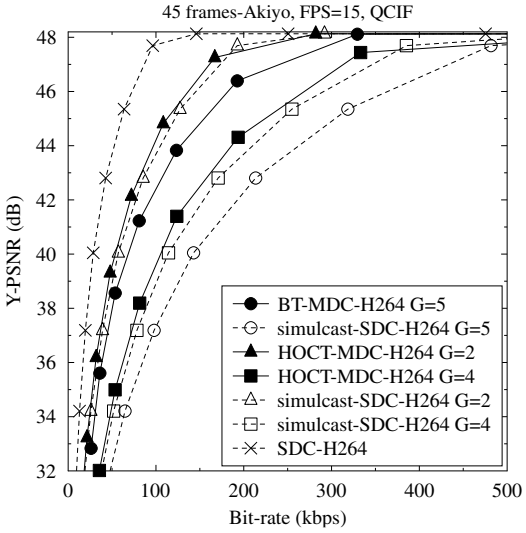


Fig. 4. Comparison of BT-MDC-H.264 using $G = 5$, HOCT-MDC-H.264 using $G = 2, 4$ and simulcast-SDC-H.264 using $G = 2, 4, 5$

Fig. 5 characterizes the rate-distortion performance when different number of descriptions are available at the receiver. An important point to mention is that the aliquot part descriptions were chosen by maximizing the distance of description offsets. For example, we chose aliquot descriptions associated with the offsets of $l = 0, 2$ for $G = 4, G_a = 2$. Observe in Fig. 5 that Y-PSNR increases gradually with the number of available descriptions increasing. Furthermore, except for the curves recorded when all the G aliquot descriptions were available, there is a PSNR upper bound for the curves, which is jointly determined by the number of available descriptions G_a and the quantizer $Q = 4$. For example, when $G_a = 1$ aliquot description is available, Y-PSNR increases slower upon increasing the bitrate, because some high frequency information is removed by the quantizer.

B. Performance of MVC with T-MDC

In this section, we characterize the performance of our proposed MDC-MVC scheme, employing the left 8 of 16 views of the 100-frame *Leaving-Laptop* sequence in (1024×768)-pixel resolution. The H.264 JMVC scheme is employed as the aliquot codec in Fig. 3 for encoding $K = 8$-aliquot camera views into $(K \times L)$ bitstreams. Moreover, the Intra-frame (*I*) refresh period was set to 15 and the

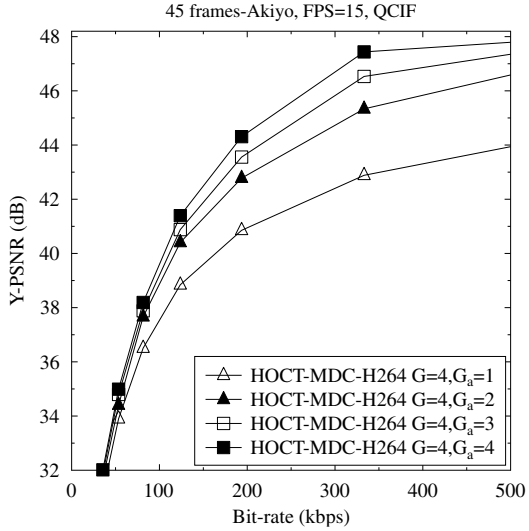


Fig. 5. Rate-distortion performance of HOCT-MDC-H.264 for $L = 4, G = 4, G_a = 1, 2, 3, 4$

frame scanning rate per second (FPS) was set to 16.67. This facilitates the evaluation of the bitrate, which can be adjusted by modifying the quantization parameters (*QP*).

The comparison among the rate-distortion performances of the BT-MDC-H.264-MVC scheme of Fig. 3, HOCT-MDC-H.264-MVC and simulcast-SDC-H.264-MVC recorded for $L = 2, L = 4$ is displayed in Fig. 6, where we observe that upon increasing the bitrate, HOCT-MDC-H.264-MVC increasingly outperforms simulcast-SDC-H.264-MVC. Quantitatively, for $G = 4$ HOCT-MDC-H.264-MVC outperforms simulcast-SDC-H.264-MVC by 0.7 dB at 2×10^5 kbps, because the reconstruction error of the aliquot part descriptions becomes lower upon increasing the accuracy of the high frequency information. The BT-MDC-H.264-MVC scheme outperforms simulcast-SDC-H.264-MVC using $G = 5$ by about 1.5 dB at a bitrate of 10^5 kbps. Alternatively, based on Fig. 6 we may argue that BT-MDC-H.264-MVC roughly halves the bitrate required for achieving 42 dB Y-PSNR.

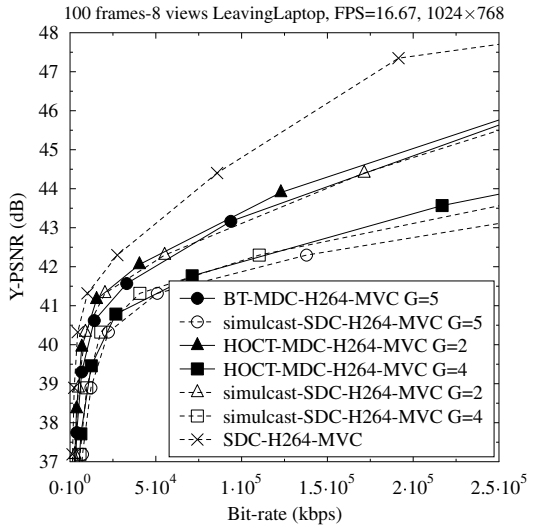


Fig. 6. Comparison among BT-MDC-H.264-MVC using $G = 5$, HOCT-MDC-H.264-MVC using $G = 2, 4$ and simulcast-SDC-H.264-MVC using $G = 2, 4, 5$

It may be concluded from Fig. 4 and Fig. 6 that HOCT-MDC outperforms simulcast-SDC in all the scenarios considered, while BT-MDC outperforms HOCT-MDC. This trend is not unexpected, since each of the simulcast-SDC descriptions carries the entire original video sequence, while each description in HOCT-MDC carries a coarse version of the original video along with unique high frequency information. It may also be concluded that BT-MDC creates descriptions containing the lowest correlations amongst the three schemes.

VI. Conclusion

A new tree-structure based MDC scheme was proposed for a flexible generation of multiple video descriptions, which may be compressed by arbitrary video codecs. Two specific structures were analyzed in detail, namely HOCT-MDC and BT-MDC. Since BT-MDC is capable of generating descriptions of unequal importance, which correspondingly reduces the correlation amongst the descriptions, it outperforms HOCT-MDC in the absence of packet loss events. Furthermore, diverse multiview schemes were detailed.

In our future work we will design appropriate rateless coding schemes for mitigating the transmission overhead of the proposed schemes in the context of multiview camera-phone networks.

References

[1] V. Goyal, "Multiple description coding: Compression meets the network," *IEEE Signal Processing Magazine*, vol. 18, pp. 74–93, September 2001.

[2] Y. Wang, A. Reibman, and S. Lin, "Multiple description coding for video delivery," *Proceedings of the IEEE*, vol. 93, no. 1, pp. 57–69, 2005.

[3] J. Laneman, D. Tse, and G. Wornell, "Cooperative diversity in wireless networks: Efficient protocols and outage behavior," *IEEE Transactions on Information Theory*, vol. 50, pp. 3062–3080, December 2004.

[4] M. Lukacs, "Predictive coding of multi-viewpoint image sets," *in Proc. IEEE Int. Conf. Acoust. Speech Signal Process.*, vol. 11, pp. 521–524, April 1986.

[5] A. Vetro, T. Thomas, and G. Sullivan, "Overview of the stereo and multiview video coding extensions of the H.264/MPEG-4 AVC standard," *Proceedings of the IEEE*, vol. 99, pp. 626–642, March 2011.

[6] H. Imaizumi and A. Luthra, *Three-Dimensional Television, Video, and Display Technologies*, ch. MPEG-2 Multiview Profile, pp. 169–181. Berlin, Heidelberg, and New York: Springer Verlag, 2002.

[7] Joint Video Team (JVT) of ISO/IEC MPEG and ITU-T VCEG, *ITU-T Rec. H.264/ISO/IEC 14496-10 AVC: Advanced Video Coding for Generic Audiovisual Services*, March 2010.

[8] W. Jiang and A. Ortega, "Multiple description coding via polyphase transform and selective quantization," *Proc. SPIE Conf. Visual Commun. and Image Processing*, vol. 3653, pp. 998–1008, February 1999.

[9] L. Hanzo, P. Cherriman, and J. Streit, *Video Compression and Communications: From Basics to H.261, H.263, H.264, MPEG2, MPEG4 for DVB and HSDPA-Style Adaptive Turbo-Transceivers*. New York: John Wiley, 2007.

[10] J. Shapiro, "Embedded image coding using zerotrees of wavelet coefficients," *IEEE Transactions on Signal Processing*, vol. 41, pp. 3445–3462, December 1993.

[11] B. Kim, Z. Xiong, and W. Pearlman, "Low bit-rate scalable video coding with 3-D set partitioning in hierarchical trees (3-D SPIHT)," *IEEE Transactions on Circuits Syst. Video Technol.*, vol. 10, pp. 1374–1387, December 2000.

[12] M. Biswas, M. Frater, and J. Arnold, "Multiple description wavelet video coding employing a new tree structure," *IEEE Transactions on Circuits Syst. Video Technol.*, vol. 18, pp. 1361–1368, October 2008.

[13] H. Karim, A. Sali, S. Worrall, A. Sadka, and A.Kondoz, "Multiple description video coding for stereoscopic 3D," *IEEE Transactions on Consumer Electronics*, vol. 55, pp. 2048–2056, November 2009.

[14] N. Franchi, M. Fumagalli, R. Lancini, and S. Tubaro, "Multiple description video coding for scalable and robust transmission over IP," *IEEE Transactions on Circuits Syst. Video Technol.*, vol. 15, pp. 321–334, March 2005.

[15] T. Kunert, *User-Centered Interaction Design Patterns for Interactive Digital Television Applications*. Dordrecht, Heidelberg, London and New York: Springer Verlag, 2009.

[16] W. Matusik and H. Pfister, "3D TV: A scalable system for real-time acquisition, transmission, and autostereoscopic display of dynamic scenes," *ACM Transactions on Graphics (TOG)*, vol. 23, pp. 811–821, August 2004.

[17] L. Hanzo, J. Blogh, and S. Ni, *3G, HSPA and FDD versus TDD Networking: Smart Antennas and Adaptive Modulation*. New York: John Wiley, 2008.

[18] F. Verdicchio, A. Munteanu, A. Gavrilescu, J. Cornelis, and P. Schelkens, "Embedded multiple description coding of video," *IEEE Transactions on Image Processing*, vol. 15, pp. 3114–3130, October 2006.

[19] O. Crave, B. Popescu, and C. Guillemot, "Robust video coding based on multiple description scalar quantization with side information," *IEEE Transactions on Circuits Syst. Video Technol.*, vol. 20, pp. 769–779, June 2010.