Multiple Description Video Transcoding

Ali El Essaili*, Shoaib Khan*, Wolfgang Kellerer†, and Eckehard Steinbach*
*Munich University of Technology, Munich, Germany
†DoCoMo Communications Laboratories Europe GmbH, Munich, Germany

***Abstract*— In this paper we introduce the concept of multiple description video transcoding (MDVT). MDVT converts a single description encoded video into two or more descriptions at an intermediate node in the network. The objective of our MDVT approach is to adapt the video transmission to a multi-radio environment where two or more independent transmission paths exist between the intermediate node and the receiver. The sender does not have to be aware of the transcoding process and the multi-path transmission. MDVT can for instance be applied for multi-mode terminals that are simultaneously connected to two wireless access technologies, e.g., UMTS and WLAN. We compare MDVT with multiple description coding at the sender (MDC-S) as well as with MDC at the intermediate node (MDC-I) where the incoming single description video is decoded and re-encoded into multiple descriptions and the transmission is optimized separately for each path. We present a fast greedy method that can be used to perform multiple description video transcoding in real-time at low complexity. Our experimental results show that we can achieve performance similar to MDC-S where the sender has to be aware of the availability of multiple paths. Compared to MDC-I we observe more than 2 dB gain in reconstruction quality.**

***Index Terms*—Video transcoding, multiple description video coding, forward error correction.**

I. Introduction

VIDEO transcoding, the process of converting a video from a format to another, has been thoroughly addressed to cope with the heterogeneity of the links between a transmitter and a receiver [1]. Due to the diversity of multimedia applications and the different bandwidth characteristics of end users, intermediate nodes are often deployed in the network to adapt the transmitted video stream to the distribution channel.

On the other hand, multiple description (MD) video coding at the sender can provide improved error resilience when combined with path diversity [2]. The video source is decomposed into multiple descriptions which are transmitted over independent paths. Each description can be decoded independently at the receiver side, and the reception quality improves with the number of received descriptions. For a summary of MD coding techniques, refer to ([3], [4]).

Meanwhile, there has been very little relevant work on Multiple Description Video Transcoding (MDVT). By MDVT we mean converting a single description encoded video into two or more descriptions without going through a decoder-encoder implementation at some intermediate node in the network. [5] proposed an algorithm that could be used to split a one-layer video stream into two descriptions by partitioning of the DCT coefficients. However, there is no adaptation to the characteristics of the transmission paths. Besides, this approach depends on the video properties of the input stream and is limited to non-layered video streams. In this paper we propose a MDVT method based on forward error correction (FEC). While the rest of the paper assumes a video bit stream as an input to the intermediate node, the proposed approach can be considered for different media applications. In a typical scenario, the intermediate network node, e.g., a multi-system radio network controller, is connected to the receiver through two paths (e.g., a UMTS interface and a WLAN interface). As the transmitter is unaware of the two paths' profiles, transcoding is required at the intermediate node. The network node generates two descriptions and protects them with an appropriate amount of FEC.

FEC has already been studied for multiple description coding (MDC) ([6], [7]). In [6], a scalable video stream is marked at N different layers. Each layer $i, i = 1 \ldots N$ is further decomposed into i sections and protected with (N, i) Reed Solomon (RS) code. This results in N descriptions, and the i^{th} layer is recovered when i out of N descriptions are received. The rate boundaries of the scalable video stream are being altered to determine the optimal FEC allocation. When the optimization of the FEC allocation is carried out at an intermediate network node, the rate boundaries are already fixed by the source and the optimization will result in a suboptimal solution. In [7], FEC-based MDC has been considered for multicast overlay streaming. The source applies FEC and determines the optimal (N, i) parameters for each connected node. Intermediate nodes are deployed to truncate and repack the encoded stream with an appropriate FEC proportion so that it matches with the end users' rate and loss characteristics. While this method can serve a large number of users, it leads to degradation in performance compared to traditional FEC-based MDC where the end user's peak signal to noise ratio (PSNR) is directly maximized by the source.

In this paper we present a new framework to apply MDVT at an intermediate network node. The transcoding is independent of the video source, i.e., the source does not have to know about the transcoding at the intermediate node. The intermediate node jointly optimizes a layered video stream transmission over two available paths to maximize the reconstruction quality at the receiver. It determines the optimal layer partitioning and protection depending on the loss and rate characteristics of both paths. We propose a fast greedy algorithm to transcode the video stream at low time complexity. To the best of our knowledge this is the first practical approach for multiple description video transcoding at intermediate network nodes.

The rest of the paper is organized as follows. In Section 2, we describe our MDVT approach. In Section 3, we propose a fast and dynamic greedy method to solve the presented approach. In Section 4, we provide simulation results. In Section 5, we conclude the paper.

1-4244-1437-7/07/$20.00 ©2007 IEEE

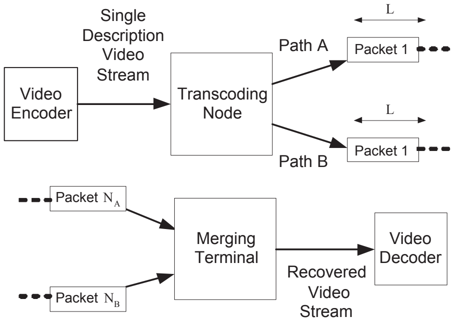


Fig. 1. Complete path between the transmitter and receiver including MDVT at an intermediate network node.

II. MDVT-APPROACH

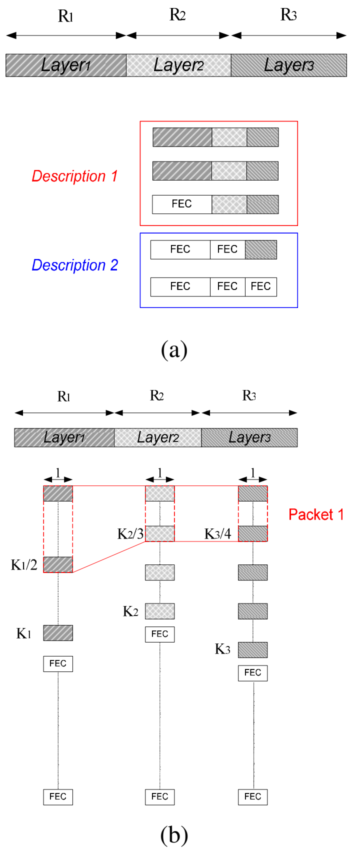
A. Overview

The single description video stream is generated from a standard hybrid or scalable video encoder. The transcoding node processes each received Group of Pictures (GOP) independently and transcodes it into two descriptions. The descriptions are formed of N_A and N_B packets of fixed size L respectively. These descriptions are transmitted over two independent paths to the receiver. At the receiver side, the descriptions are merged and the recovered video stream is input to the video decoder. The complete transmission chain is explained in Figure 1.

B. Approach

We consider a layered video stream (e.g., M layers) as an input to the transcoding node. $Layer_1$ could correspond to the base layer and $Layer_2$ to $Layer_M$ to the enhancement layers of a scalable video stream. Similarly, $Layer_1$ could stand for an I frame and $Layer_2$ to $Layer_M$ for subsequent P and B frames of a hybrid codec video stream. The layer boundaries are already specified by the source; $Layer_i$ has a fixed length equal to $R_i, i = 1, 2, \ldots, M$. We partition each layer into small segments of predefined size l. This will result in K_1 segments for $Layer_1$ where the last segment is zero-padded if R_1 is not a multiple of l. In the same way, $Layer_2$ to $Layer_M$ are partitioned into small segments of the same size l resulting in K_2 to K_M segments, respectively. We define the packet length L to be an integer multiple of the segment size l. We further define N_i as the number of packets that should be received in order to recover $Layer_i$. The contribution of $Layer_i$, $cont(i)$, specifies the number of segments of $Layer_i$ that should be allocated to each transmitted packet. As $Layer_i$ should be recovered from any N_i received packets, then $cont(i)$ equals to $\lceil \frac{K_i}{N_i} \rceil$.

As an example we consider in Figure 2(a) a 3-layer video stream which is transcoded into two descriptions of $N_A = 3$ and $N_B = 2$ packets. The transcoder determines the optimal (N_1, N_2, N_3) allocation set for $(Layer_1, Layer_2, Layer_3)$. For instance, in Figure 2(a), the optimization result for the allocation set is equal to $(2, 3, 4)$. This means that a minimum of 2, 3, and 4 packets should be received to recover $Layer_1$, $Layer_2$, and $Layer_3$, respectively. To satisfy this condition, each $Layer_i, i = 1 \ldots 3$ is partitioned into N_i equal parts of length $\lceil \frac{K_i}{N_i} \rceil$. The first N_i transmitted packets are filled with the data segments of $Layer_i$. A Reed-Solomon code is further applied to each layer to generate the corresponding FEC parts which are allocated to the rest of the transmitted packets. The result of the optimization is illustrated in Figure 2(b) where $(\lceil \frac{K_1}{2} \rceil, \lceil \frac{K_2}{3} \rceil, \lceil \frac{K_3}{4} \rceil)$ segements of $(Layer_1, Layer_2, Layer_3)$ are transmitted in each packet, respectively.



(a)

(b)

Fig. 2. Schematic description of the MDVT approach. (a) The video stream is transcoded into two descriptions. The descriptions are formed of 2 and 3 packets of fixed packet length. (b) Allocation of the sections of the video stream across the transmitted packets.

C. Problem formulation

Given (N_A, P_A) and (N_B, P_B), the packet budget and packet loss rates (PLR) of paths A and B respectively, the expected distortion function for a GOP is defined by:

$$E\{D(N_A, N_B, P_A, P_B, N_1, \ldots, N_M)\} = \sum_{i=N_A+N_B-N_1+1}^{N_A+N_B} P(N_A+N_B, i) D(Layer_1) + \sum_{i=N_A+N_B-N_2+1}^{N_A+N_B-N_1} P(N_A+N_B, i) D(Layer_2) \ldots + \sum_{i=N_A+N_B-N_M+1}^{N_A+N_B-N_{M-1}} P(N_A+N_B, i) D(Layer_M)$$

Each term in the expected distortion function determines the expected distortion at the receiver for different packet loss combinations. The two paths are disjoint and $P(a,b)$ is the binomial probability that b out of a transmitted packets on paths A and B are lost. $D(Layer_i), i = 1 \ldots M$, represents the GOP distortion when the video stream is truncated at $Layer_i$. The distortion values are needed to determine the optimal data and FEC allocation at the intermediate node. They are calculated at the sender and sent along with the video stream (e.g., [8]).

We search for the optimal allocation of the segments of $Layer_1$ to $Layer_M$ across the transmitted packets such that the expected distortion at the receiver is minimized. The problem is formulated as follows:

$$\{N_1, \ldots, N_M\} = argmin\, E\{D(\ldots)\}$$

$$Subject\ to:$$

$$\sum_{i=1}^{M} cont(i) = \sum_{i=1}^{M} \lceil \frac{K_i}{N_i} \rceil \leq \frac{L}{l} \quad (1)$$

$$N_1 \leq N_2 \leq \cdots \leq N_M \quad (2)$$

(1) stands for the packet budget constraint, i.e., the sum of the contributions of the different layers to each of the transmitted packets should not exceed the packet length. This constraint also solves the packet filling problem. Once the $(N_1, \ldots, N_M)$ values are determined, the contribution of each layer to each packet is known and the data and FEC segments of each layer can be allocated to the transmitted packets. This constraint also guarantees a fixed packet size across the GOP. (2) is a direct consequence of the dependency between the different layers of the GOP.

III. MDVT Algorithm

The optimal solution to our non-convex allocation problem could be obtained by dynamic programming methods [9]. However, complexity is a major limitation of any transcoding approach. The high complexity of dynamic programming methods makes them unsuitable to our application. Therefore, we follow a greedy allocation technique to determine the $\underline{N} = (N_1, \ldots, N_M)$ values. From (1), we notice that incrementing any N_i value while keeping the rest of N_j values constant $(i.e, j = 1 \ldots M, j \neq i, \forall i)$ will decrease the total rate. This also means that more packets should be correctly received to recover $Layer_i$ and will thus increase the expected distortion at the receiver. We respectively denote by $L(\underline{N})$ and $D(\underline{N})$ the total rate and the total distortion of the GOP for a given $\underline{N}$.

We define the slope that results from incrementing N_i by an integer step size s as:

$$slope(i) = \frac{D(\underline{N} + se_i) - D(\underline{N})}{L(\underline{N}) - L(\underline{N} + se_i)}, e_i = \begin{cases} 1 & i^{th} \text{ position} \\ 0 & \text{elsewhere} \end{cases}$$

The basic greedy allocation algorithm is described as follows:

1: As a starting point, we determine the minimum N_i value for each layer. This value corresponds to the best possible solution for that individual layer assuming the worst case scenario for the rest of the layers of the GOP. This is done, for instance, by saying that we want to transmit $Layer_i$ only; determine the minimum N_i value that satisfies (1) and (2). As a result, we have an initial $\underline{N}$ with a total rate $L(\underline{N}) > \frac{L}{l}$.

2: Calculate the $slope(i)$ that results from every possible N_i increment, $i = 1 \ldots M$. Each N_i is incremented by a variable integer step size that allows a valid slope.

3: Determine the minimum $slope(i)$ value and increment its corresponding N_i.

4: Repeat steps 2 and 3. Stop when $L(\underline{N}) = \frac{L}{l}$.

For low transmission rates or high packet loss rates, dropping lower priority layers will allow more important layers to enjoy more protection. This can be achieved by allowing the algorithm to truncate the layered stream. The MDVT algorithm is thus executed in two phases:

A: Determine the truncating point: Apply the basic algorithm; truncate the stream at $Layer_j$ if $N_{j+1} > N_A + N_B$. Due to the dependency between layers, all $Layer_k, j+1 \leq k \leq M$ will be dropped.

B: Apply the basic algorithm to determine the $(N_1, \ldots, N_j)$ values.

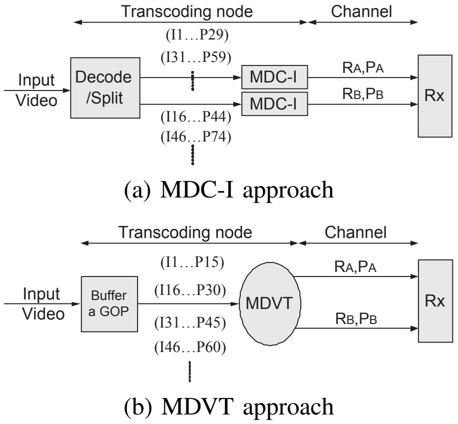
IV. Simulation Results

The experiments performed to test our MDVT algorithm use the H.264/AVC codec [10]. An independent Bernoulli packet loss channel is considered, the packet length L is fixed to 512 bytes, and the segment size l is fixed to 1 byte.

A. MDVT vs. MDC-I

An alternative method of transcoding a video stream over two paths is to decode the video stream and split it into two subsets of even and odd frames. In MDC-I each subset is encoded and error protected separately and transmitted over a different path. One could consider this as a form of joint source channel coding applied separately to each description. On the other hand, in our MDVT approach, we consider the two paths as one virtual path and we optimize jointly over both paths. We use the MDVT algorithm to transcode the video stream into two descriptions. To compare both approaches, we consider a *foreman* test sequence (QCIF size) of $IPP\ldots$ structure. A GOP length of 15 frames is used and the results are averaged over 24 GOP. (R_A, P_A) and (R_B, P_B) define the bit rate and PLR on both paths. The above comparison is illustrated in Figure 3.

The mean bit rate of the input video stream is equal to 156 kbps and the generated odd and even streams have a mean bit rate of 90 kbps. We consider a two balanced paths scenario with a PLR of 10%. We compare the MDVT approach and the MDC-I approach by varying the total transmission rate. Figure 4 shows the gain achieved by jointly optimizing over both paths. As the redundancy in the transmitted descriptions increases, the MDVT approach can efficiently distribute the FEC segments across the different frames so that the composite



(a) MDC-I approach

(b) MDVT approach

Fig. 3. Schematic description of MDVT and MDC-I transcoding approaches.

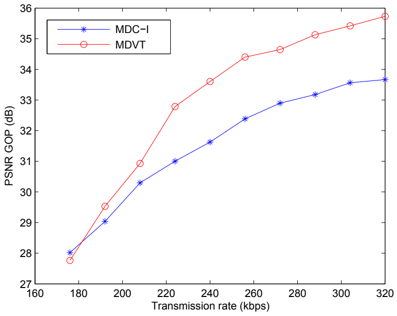


Fig. 4. Comparing between MDVT and MDC-I transcoding approaches.

channel's reconstruction quality is maximized. At the source rate limit, both methods converge to the same point.

Besides the gain in reception quality, one major benefit from using the MDVT approach is its low complexity. The transcoding is done on the fly and the execution time of the MDVT algorithm does not exceed a couple of ms per GOP.

B. MDC-S

In this subsection we study the case where the sender is aware of the existence of the two paths. The transmission rate is the aggregate rate over both paths and is equal to the sum of the source coding rate, Rs, and the channel coding rate, Rc. In our transcoding scenario, Rs is already determined by the video encoder. On the other hand, when the two descriptions are generated at the transmitter, the encoder can choose between different Rs operational modes for a given transmission rate. In Figure 5, we consider the flexibility of encoding the video at different source coding rates and source distortions. We consider a *foreman* test sequence of $IBP\ldots$ structure and a GOP size of 16 frames. Figure 5 shows the average reconstruction quality (PSNR) at the receiver as a function of the transmission rate and for a PLR of 15%. Each curve corresponds to the MDVT algorithm applied for a given source coding rate. For a particular transmission rate, each curve represents a different source and channel rate combination. By having the flexibility to choose between different operational modes, the optimal trade off between source and channel coding can be determined for a given rate budget. This eventually leads to a gain at the receiver compared to a single mode operation.

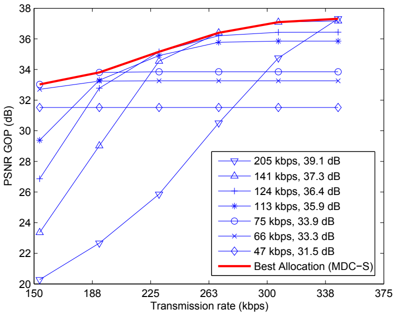


Fig. 5. PSNR at the receiver as a function of the transmission rate for different source coding rates.

V. Conclusion

In this paper we introduce multiple description video transcoding as a novel approach for multi-path communication. In MDVT a single description stream is transcoded into two or more descriptions at an intermediate network node. We present a practical framework for MDVT and propose a fast transcoding algorithm. Our simulation results show that our proposed approach can provide a similar performance to MDC at the sender (MDC-S) and outperforms MDC at the intermediate node (MDC-I) where the transmission is optimized separately for each path.

References

[1] J. Xin, C-W. Lin, and M-T. Sun, *Digital video transcoding*, Proceedings of the IEEE, vol. 93, no.1, pp. 84-97, Jan. 2005.

[2] J. G. Apostolopoulos, *Reliable video communication over lossy packet networks using multiple state encoding and path diversity*, Proc. Visual Communications and Image Processing, pp. 392-409, Jan. 2001.

[3] V. K. Goyal, *Multiple Description Coding: Compression meets the network*, IEEE Signal Processing Mag., vol. 18, no.5 pp. 74-93, Sept. 2001.

[4] Y. Wang, A. R. Reibman, and S. Lin, *Multiple description coding for video delivery*, Proceedings of the IEEE, vol. 93, no. 1, pp. 57-70, Jan. 2005.

[5] A. Reibman, H. Jafarkhani, Y. Wang, and M. Orchard, *Multiple description video using rate-distortion splitting*, ICIP 2001 Proceedings, 2001, pp. 978-981.

[6] R. Puri and K. Ramchandran, *Multiple description source coding using forward error correction codes*, in Proc. 33rd Asilomar Conf. Signals, System Comp., vol. 1, pp. 342-346, Pacific Grove, CA, Oct. 1999.

[7] G. Wang, S. Futemma, and E. Itakura, *FEC-based scalable multiple description coding for overlay network streaming*, Consumer Communications and Networking Conference (CCNC), Las Vegas, Nevada, Jan. 2005.

[8] W. Tu, W. Kellerer, and E. Steinbach, *Rate-Distortion optimized video frame dropping on active network nodes*, in Packet Video Workshop 2004, Irvine, California, Dec. 2004.

[9] A. V. Trushkin, *Bit number distribution upon quantization of a multivariate random variable*, Problems of Information Transmission, vol. 16, pp. 76-79, 1980.

[10] JVT H.264/MPEG-4 AVC reference software at http://iphome.hhi.de/suehring/tml/download/.