SCALABLE MULTIPLE DESCRIPTION VIDEO CODING WITH FLEXIBLE NUMBER OF DESCRIPTIONS

Emrah Akyol\(^1\), A.Murat Tekalp\(^1,2\), and M.Reha Civanlar\(^1\)

\(^1\) College of Engineering, Koc University, Istanbul, Turkey
\(^2\) Department of Electrical and Computer Engineering, University of Rochester, Rochester, NY 14627

ABSTRACT

Multiple description video coding mitigates the effects of packet losses introduced by congestion and/or bit errors. In this paper, we propose a novel multiple description video coding technique, based on fully scalable wavelet video coding, which allows post encoding adaptation of the number of descriptions, the redundancy level of each description, and bit rate of each description by manipulation of the encoded bitstream. We demonstrate that the proposed method provides excellent coding efficiency, outperforming most other multiple description methods proposed so far. We also provide experimental results to show that varying the number of descriptions according to network conditions is superior to using a fixed number of descriptions, by means of NS-2 network simulation of a peer-to-peer video streaming system.

1. INTRODUCTION

Multiple description coding (MDC) addresses the problem of encoding source information using more than one independently decodable complementary bitstreams, which, when combined, can provide the highest level of quality and when used independently, can still provide an acceptable level of quality. This is made possible by introducing some redundancy in each description, which will be discarded if all streams are received. It is well known that MDC can provide robust video communication over unreliable networks, such as Internet or wireless networks, by utilizing path/server diversity at the cost of reduced compression efficiency.

There has been significant amount of early work on MDC employing nonscalable video codecs, including encoding even and odd frames independently [1], and utilizing motion estimation across descriptions, called motion compensated multiple description (MC-MDC) [2]. An excellent survey of multiple description techniques can be found in [3].

In order to adapt the rate of each description to varying network conditions, scalable video coders should be used to generate the multiple descriptions. One example of scalable MDC is based on motion compensated temporal filtering, where high frequency frames are grouped into two descriptions and missing frames are estimated using motion vectors in the two descriptions [4]. It is reported to outperform existing non-scalable MD video coders in compression performance while providing flexible rate allocation and redundancy control, although its performance is degraded under significant motion since estimating missing frames then becomes a difficult task. Furthermore, none of the available MDC methods addresses adaptation of the number of descriptions (post encoding) according to the network conditions.

Providing a variable (flexible) number of descriptions post encoding becomes an important concern in peer-to-peer (P2P) video streaming, where the number of available “good” source peers is not known a priori. Consider the scenario, where there are \(n\) source peers with scalable coded video available to start with, and each sends one out of \(n\) descriptions toward a common destination. In this setting the playback can start immediately and the performance of each source path can be measured at the destination for a given period. As a result of this measurement, the best \(k < n\) sources can be selected out of the initial \(n\), and they are requested to provide one of \(k\) descriptions which can be produced (post encoding) from the scalable descriptions available at each source.

To this effect, we propose a novel scalable multiple description video coding framework, which enables varying

\(i\) the number of descriptions,

\(ii\) the rate of each individual description, and

\(iii\) redundancy level of each description

on the fly (i.e., post encoding). These properties of the coder enable efficient adaptation to network conditions.

The main idea of our framework is to distribute encoded
spatio-temporal code-blocks to any given number of
descriptions. Since every code-block can be decoded at
any given rate, independent of any other code-block, we
can decode each code-block at high bitrate for only one
description. We decode the same code-block at low bitrate
for other descriptions. Different code-blocks, some of
which are coded at the high rate and others at the low rate
are mixed together to form a description. So if all
descriptions are received, decoder uses only code-blocks
which are coded at the high bitrates from all the
descriptions. On the other hand, if only one description is
received we still have an acceptable video quality with
code-blocks decoded at low-bitrates.

We discuss the proposed MD coding scheme in Section 2.
Comparative results are presented in Section 3.
Experimental results indicate that the proposed method
has high compression efficiency, outperforming MD
using motion compensated temporal filtering (MD-
MCTF) [4] up to 6-7db at different rate-redundancy
points. Conclusions are presented in Section 4.

2. PROPOSED METHOD

Proposed method is based on interframe wavelet based
scalable video coding. MCTF and 2-D spatial wavelet
filtering are used to perform motion compensated 3-D
wavelet transform, which creates a spatio-temporal
subband cube. All spatio-temporal subbands are divided
into non-overlapping regions called code-blocks. Similar
to most state-of-the art interframe wavelet coders, all
spatio-temporal code-blocks are encoded independently
by embedded bitplane coding.

In the proposed framework, we generate N descriptions,
each composed of code-blocks extracted at M different
bitrates from a single scalable coded video bitstream.
Every description includes motion vectors at full rate.
The case of N=2 descriptions derived from M=2 rates
(high rate and low rate) is illustrated in Figure 1.

One code-block when extracted at the high bitrate belongs
to one description, and when extracted at the low-rate
belongs to other descriptions. This is repeated for all
code-blocks following a regular pattern, except for the
lowest frequency code-blocks in both time and space.
Those code-blocks are decoded at high bit-rate in all
descriptions, since they affect the visual quality more than
other code-blocks.

Since the codeblocks at different rates are extracted from
a single scalable coded video, the number of rates M as
well as the number of descriptions N generated by various
combinations of them for each codeblock are totally
flexible and can be varied post encoding. The case N=3
descriptions derived from M=2 rates is illustrated in
Figure 2. The best M bitrates for each code-block can be
found by performing rate distortion optimization for each
block as in EBCOT [5].

The proposed scalable multiple description video coding
framework enables varying:

i) the number of descriptions: Code-blocks extracted
   at any number of bitrates can be distributed to any number
   of descriptions except for extracting the lowest spatio-
   temporal frequency band code-blocks at the high rate in
   all descriptions. Every description has some of code-
   blocks extracted at high rate and other code-blocks at low

Figure 1: Proposed MDC method for N=2 descriptions derived
from M=2 rates (high and low rates). The number of
descriptions N, and rates M can be adapted post encoding.

Figure 2: Proposed MDC method for N=3 descriptions coded at
M=2 rates.
rate where the number of the code-blocks extracted at high rate decreases as the number of descriptions increase.  

**ii) the rate of each individual description:** Since embedded bitplane coding is used for encoding every code-block, each description is rate scalable.  

**iii) redundancy in each description:** An individual description at any fixed rate can be generated with several high and low decoding rates. As the high rate increases; low rate should decrease to provide the same rate for the description. Hence, it is possible to control the redundancy level by changing high and low rates of code-blocks.

### 3. COMPARATIVE RESULTS

The proposed method is compared to other multiple description coders with redundancy-distortion curves for some fixed bitrates. We used a wavelet coder based on JPEG-2000 [5], which uses EBCOT for rate allocation, however, other wavelet based scalable coders can also be used. 3 level spatial and 4 level temporal decompositions are used in the coder.

Comparative results of the proposed coder and MD-MCTF are provided below for the case where \(N=2\) descriptions at three different bitrates and four redundancy levels when only one description is received.

**Table-1: Foreman QCIF 30fps; PSNR of the proposed method/MD-MCTF [4].**

<table>
<thead>
<tr>
<th>Redundancy</th>
<th>100kbps</th>
<th>200kbps</th>
<th>300kbps</th>
</tr>
</thead>
<tbody>
<tr>
<td>20%</td>
<td>NA/27.4</td>
<td>31.64/28.8</td>
<td>34.75/29.5</td>
</tr>
<tr>
<td>30%</td>
<td>29.44/27.7</td>
<td>34.03/29.4</td>
<td>37.02/30.8</td>
</tr>
<tr>
<td>40%</td>
<td>31.09/28.0</td>
<td>34.99/30.2</td>
<td>38.01/31.9</td>
</tr>
<tr>
<td>50%</td>
<td>32.35/28.1</td>
<td>35.76/30.2</td>
<td>38.59/32.0</td>
</tr>
</tbody>
</table>

**Table-2: Akiyo QCIF 30fps; PSNR of the proposed method/MD-MCTF [4].**

<table>
<thead>
<tr>
<th>Redundancy</th>
<th>50kbps</th>
<th>100kbps</th>
<th>200kbps</th>
</tr>
</thead>
<tbody>
<tr>
<td>20%</td>
<td>29.21/31.9</td>
<td>36.36/36.0</td>
<td>46.21/39.0</td>
</tr>
<tr>
<td>30%</td>
<td>31.03/32.0</td>
<td>37.76/36.2</td>
<td>46.82/40.1</td>
</tr>
<tr>
<td>40%</td>
<td>32.15/32.0</td>
<td>38.67/36.3</td>
<td>46.95/41.0</td>
</tr>
<tr>
<td>50%</td>
<td>35.11/32.0</td>
<td>39.33/36.3</td>
<td>47.02/41.0</td>
</tr>
</tbody>
</table>

Our coder outperforms MD-MCTF significantly at medium motion sequence (Foreman), however, for sequences with low motion (Akiyo) MD-MCTF performs comparable (at some low rates better) to our coder because MD-MCTF can properly estimate missing frames at sequences with low motion. Since MD-MCTF is reported to outperform MC-MDC [2] and domain based MD coder [6], and we observed that our coder performs better than MD-MCTF in most rate-redundancy levels, we did not compare our coder to MC-MDC or domain based MD coder.

To simulate the packet losses, we set up a simulation with NS-2 network simulator [7]. The luminance component of the Foreman sequence in QCIF format is coded with wavelet coder with 3 spatial and 3 temporal decomposition levels for 296 frames at 30fps. Other than the lowest frequency frame in the temporal decomposition, every frame is packetized into one packet with maximum size of 1000 bytes. Every spatial resolution in the lowest frequency frame is put into one packet. All motion vectors for a GOP length 8 are put into a total of 2 packets. Traffic trace files generated in the coder are used in the ns-2 simulation to specify the timing and size of each packet. The packet sending interval is determined in accordance with the playing rate to avoid buffer overflows/underflows at the receiver side.

To show the use of adapting the number of descriptions to the number of available channels; we describe this example scenario: There are 8 senders who have the encoded video and description generator to generate any number of descriptions. The last hop link is set as the bottleneck link with 100kbps bandwidth and high error rate. All senders send multiple descriptions over disjoint paths. Every path from senders to receiver shares one link with 200kbps bandwidth with external traffic. External cross traffic is randomly specified as %50 of the link capacity with exponentially distributed packet sizes and sending intervals.

The shared link can sometimes be heavily loaded with external 1Mbps constant bitrate traffic, so the path enters loaded state. Two of the eight available paths are in loaded state initially. At 9th sec, randomly four of them also become loaded with external traffic which runs till the end of the simulation. The simulation time is the two full play time of the video, \(T=20\) sec. The time period between the time at path entering loaded state and sender sides become aware of that event, the recognition time, is assumed to be \(t_{rec}=1\) sec. The rate of each description is set to the bottleneck bandwidth \(R=100\)kbps.

The proposed system starts with \(N=8\) descriptions. After a recognition time \(t_{rec}=1\) sec, it adapts the number of generated descriptions to \(N=6\). At time \(t=10\) sec, after a \(t_{rec}=1\) sec time from congestion beginning time, it changes the number of descriptions to \(N=2\).

For comparison purposes, the performance of a test system with fixed number of descriptions is also simulated. Since the number of available good channels alternates between two and six during the simulation, the
number of descriptions is fixed to $N=4$. The test system begins with sending four descriptions over four channels, and uses remaining four paths to send four identical back-up descriptions. After a time of $t_{rec}=1\text{sec.}$, it stops sending two of the back-up descriptions which are on loaded paths. After time $t=9\text{ sec}$, when congestion begins in four of the channels, it stills continues to send four descriptions over four paths although two of them are heavily loaded.

We compared the proposed system with the test one with two different packet loss rates as %5 and %20. Packet loss rates are found by analyzing the ns-2 output trace files of the paths which do not experience congestion. The high and low bitrates which determine the amount of redundancy is found according to the packet loss rate estimated in the first $t_{rec}=1\text{sec}$ time period. [8] The results are found by averaging 15 realizations of the simulation.

The proposed system outperforms the fixed number of descriptions by 2.55 dB at %5 loss rate, and 0.77dB at %20 loss rate. The PSNR gap between the adaptive and fixed number of descriptions decreases as loss rate increases. The reason behind this observation is that resending the identical descriptions performs quite close to sending multiple descriptions at very high loss rates.

4. CONCLUSIONS

A flexible multiple description video coding framework with high compression efficiency based on fully scalable wavelet video coding is proposed. The main novelty of the proposed framework is the adaptation of the number of descriptions post encoding, in addition to properties of other existing scalable MDC coders such as rate and redundancy level adaptation. We note that the proposed method can be applied to any type of scalable video coder which utilizes independent code-block coding as the sub-band entropy coding technique.

5. REFERENCES


