Abstract - Multiple description coding (MDC) schemes have become popular as robustness tools in packet switched networks where multiple paths with varying packet loss rates are available. In this paper we propose a new MDC scheme for video using 3D wavelet decomposition. The spatio-temporal coefficients are then grouped using either a random, uniform scheme or a hierarchical scheme. The descriptions are constructed by arranging a subset of these groups in two positions: primary and the secondary. The bit-allocation problem between the primary and secondary groups is formulated and solved using the generalized BFOS [4] algorithm. The bits allocated to the secondary position may be zero, in which case, the description is said to be constructed with no redundancy. Experiments using bi-orthogonal wavelet for the spatial and Haar wavelet for the temporal decomposition, along with embedded block coding with optimal truncation for the different groups have shown significant performance for test video sequences. We expect that, for finite packet loss probability, the MDC scheme with redundancy does better than the scheme that does not have redundancy. In this paper, we also present a discussion of the differences between our work with the most recent 3D based video MDC scheme [18].

1. INTRODUCTION

The main objective of traditional video compression algorithms is to maximize the coding gain, assuming error free channels. Even layered coding schemes assume that the base layer arrives at the decoder at all times. Wireless packet switched networks; however, do not guarantee the delivery or the time of delivery of every packet. In real time video applications, if a packet arrives too late or extremely corrupted, then it is considered lost. Packets may also be lost due to congestion at intermediate nodes in the network. Although traditional method like ARQ exists to mitigate the effects of packet loss to some extent, they could potentially involve unacceptable time delays for video applications; hence alternative techniques for error resilience are being aggressively sought after. Multiple description coding (MDC) is one such rapidly emerging technique that has attracted significant attention and seems well suited to address this issue in packet switched networks.

MDC uses the idea of diversity in transmission paths to achieve error resilience. In such a scheme, several representations of the source, called descriptions, are generated which are transmitted along different paths. The descriptions are designed in such a way that the quality of the received signal degrades gracefully with the increase in the number of lost descriptions and such that the redundancy between the description is kept as small as possible. The quality of the reconstructed video sequence depends on the number of descriptions received and not on which descriptions are actually received. Such a coding scheme is well suited to packet networks and fading wireless channels where there is inherent diversity in paths. In this paper we propose a new 3D multiple description video codec using 3D wavelet decomposition for the video sequence and by adapting the embedded block coding with optimal truncation (EBCOT) algorithm [7] developed for images (and used in JPEG 2000[7], MJPEG2000) to our video coding scheme.

3D video coding could be regarded as an extension of 2D wavelet coding by the addition of time as a third dimension and significant work has been done on single description 3D video coding schemes [10,11,12,17]. Significant work has also been done on multiple descriptions coding of images [1,2,5,8,9]. However, multiple description video coding is a relatively nascent area, offering many opportunities for further exploration. Very recently, a MDC scheme was developed for video [15] in which adjacent pixels are interleaved to form a multiple description code where each description uses a non-redundant error concealment-coding scheme. The reconstruction error due to lost descriptions is corrected by interpolation at the receiver. Other works on multiple description video coding include [14,16,18,19].

Recently Kim et al [18] proposed a 3D video MDC, where they used temporal partitioning and frequency partitioning to form groups. In the temporal partitioning scheme, two sub streams are formed from the odd and even frames of the video sequence. In the frequency domain partitioning, wavelet coefficients are grouped in a hierarchical fashion (similar to the one proposed by Cho et al [19]). In both partitioning schemes the groups are entropy coded using 3-D set partitioning in hierarchical trees (3-D SPIHT). This frequency domain grouping scheme is similar to our hierarchical selection scheme but our scheme also uses redundant groups in forming the descriptions. In our experiments the descriptions are coded using EBCOT. In this paper, we present a new 3-D multiple description video coder. Experimental results show that the proposed approach leads to more graceful degradation of video quality with increase in the loss of descriptions.

2. THE PROPOSED 3-D MULTIPLE DESCRIPTION VIDEO CODER

Figure 1: 3D wavelet transform of video sequence.

In our scheme, the video frames are first decomposed over space and time dimension using a 3-D separable wavelet transform as shown in Figure 1. We use a reversible 5/3 wavelet for spatial decomposition and Haar wavelets for temporal decomposition. The
number of frames in each decomposition, which corresponds to the
depth in the time dimension, can be varied depending on the amount
of motion in the sequence with fewer frames being used if there is
less motion in the sequence. Note that in general, as the number of
frames in a GOF increases the complexity of the wavelet
decomposition increases. Hence the choice of number of frames in
the GOF will depend on the amount of motion and the limitations on
complexity. In our experiments we have found that using 4 frames
per GOF gives reasonably good results and increasing the number of
frames to more than 20 does not cause any significant performance
improvement. The numbers of levels of decomposition will
determine the number of groups in the formulation.

Each description consists of three components: the single coefficient
from the LL band, the primary position and the secondary position.
Figure 2 depicts one example of constructing descriptions from
groups.

Figure 2. Formation of descriptions from four groups

The groups in the primary positions are typically coded with higher
rate than the groups in the secondary positions. There may be one or
more groups occupying the primary position and zero or more
groups in the secondary position. The number of primary and
secondary groups and the rates at which each of these groups are
coded, is kept constant for all descriptions in order to ensure that
each additional description received causes roughly equal
improvement in the quality of the reconstructed signal – one of the
primary goals of designing multiple description codes. In the
following subsection, we discuss the formation of groups of subband
coefficients.

2.1. Grouping subband coefficients

The first step towards constructing the multiple descriptions is to
group the wavelet coefficients into different basic units called
groups. The coefficients forming each group are selected uniformly
or hierarchically from all bands and levels in the decomposition, to
ensure that there is sufficient representation both spatially and in
terms of frequency. We hypothesize that uniform, random selection
will work better than other structure schemes such as a hierarchical
scheme because the hierarchical scheme precludes grouping
spatially diverse coefficients into one group (since the coefficients
that are grouped will have a parent descendant relationship). We
will present experimental results to demonstrate that this is true.

The second design issue in our MDC scheme is the allocation of rate
between the different positions in a single description. The
following section examines the rate allocation problem in more
detail, assuming an underlying structure for the formation of the
multiple descriptions from the different groups.

2.2. Rate allocation between the groups

In our MDC formulation, the redundant information that will be
used to offset the effect of network loss, is embedded in the
secondary positions in a description. Since in general the amount of
redundancy needed to protect the source against network errors will
depend on the packet loss rate, it can be argued that the number of
bits allocated to the primary and the secondary groups would
depend on the packet loss statistics of the network. We therefore
optimize the redundancy level in the descriptions in a rate-distortion
sense with respect to the packet loss statistics. To do this, the
contribution of each position to the overall distortion of the
reconstructed video, needs to be calculated. For this, we need to
calculate the probability that any group will be used in the
reconstruction. This is illustrated using an example in the next
paragraph. Once these distortions and the corresponding rates are
known, we can cast our problem in a framework similar to the
problem of bit allocation between different sources in source
coding. Consequently, the problem can be solved using the
generalized BFOS [4] algorithm. Note that the rate allocation
procedure described here assumes some underlying structure for the
formation of descriptions from groups; we do not tackle the problem
of optimizing the structure itself, which might well be an intractable
problem.

To illustrate our bit allocation formulation, let us consider the
example where each description is made of exactly one primary and
one secondary position. Let us also assume that the coefficients are
grouped into four basic groups and let us further consider the set of
descriptions formed by taking all possible 2-group combinations
such that the same group does not occur in both the primary and
secondary positions of the same description. By this method twelve
descriptions can be formed. If the packet loss rate of the channel is
ε, then the probability that the i\textsuperscript{th} group is used for the
reconstruction of the final image from the primary position is given as:

\[
    P_{\text{pri}} = \frac{M-1}{M} \left(1-\varepsilon\right)^i \left(\varepsilon\right)^{M-1-i}
\]

Where M is the total number of groups. The probability that each
group is used for the reconstruction of the final image from the
secondary position is given by:

\[
    P_{\text{sec}} = \left(1 - \varepsilon^{M-1}\right) \left(\varepsilon^{M-1}\right)
\]

These probabilities are then used to weight the distortion associated
with using the groups in the primary and secondary positions generating
two tables of rate and distortion points – one for the
primary and one for the secondary position. These tables are then
fed to the generalized BFOS algorithm to pick the R-D optimal rates
for each of the groups in a description.

2.3. The MDC-video algorithm

The MDC encoding algorithm is then stated as follows:

At the transmitter:
1. The video sequence is 3D wavelet transformed using fixed
   number of frames in the GOFs.
2. For uniform grouping: Generate and apply the mask for
each level of decomposition to select coefficients for each
group.
3. For hierarchical grouping: All the offsprings of every LL
   band coefficients (like a tree) form a single group [3,19].
4. Fix the descriptions structure. Generate R-D tables for each
   of the positions (primary and secondary).
5. Use the BFOS algorithm to obtain the optimal rate for each
   of the component groups in the description.
6. Code the descriptions using the EBCOT algorithm at the
   specified optimal rate.
At the receiver, from all the descriptions that are received we extract the highest rate bit stream corresponding to each group. These are then decoded using the EBCOT algorithm and the resulting coefficients are inverse wavelet transformed to reconstruct the video sequence.

3. EXPERIMENTAL SETUP AND RESULTS

In our experiment we used the Kakadu software [7] for coding the wavelet coefficients. The wavelet coefficients selected using random mask to form individual groups are coded independently.

In the first experiment, we construct descriptions without any redundancy – no secondary groups. Figure 3 plots the PSNR of the reconstructed sequence for each of the 4 frames in a GOF as a function of the number of descriptions received in a five descriptions case, MDC system operating at an overall rate of 0.12 bpp. As can be seen from this figure, the quality of the reconstructed video sequence decreases gracefully with the loss in descriptions.

In the second experiment, we construct 49 descriptions from 49 groups (no redundancy MDC) for each GOF and compare the performance of uniform grouping MDC scheme with a hierarchical grouping MDC scheme like that used in the [18]. A similar hierarchical grouping scheme was used in MDC-SPIHT algorithm for images [3]. For this experiment the overall bit rate for both schemes was set at 0.7 bpp making the individual descriptions at 0.0143 bpp. Figure 4 shows the performance of uniform grouping scheme compared with that of the hierarchical scheme. The PSNR values are averaged for ten iterations of random packet (description) loss. The x-axis shows the average number of descriptions lost in a single GOF and the y-axis is the averaged PSNR of the entire sequence (100 frames). As can be seen from this figure, uniform grouping scheme performs better than the hierarchical grouping scheme for all the error rates, with the performance improvement being largest at smaller packet loss probabilities. Figure 5 shows the variance in the PSNR values in the ten realizations. As can be seen from this figure, uniform grouping scheme shows less variance in performance in comparison to the hierarchical grouping scheme showing that uniform grouping scheme is more reliable than the other. Finally Figure 6 gives a perceptual idea of the quality of the reconstructed frames for both uniform grouping and hierarchical grouping when two out of 49 descriptions are lost. The four images on the left column are reconstructed using uniform grouping scheme and the four images on the right column are reconstructed using hierarchical grouping scheme. The average PSNR for the left column images is 43.7 dB and for the right column images it is 29.1 dB.

In our third experiment we study the effect of the proposed bit allocation strategy. Twelve descriptions were formed for each GOF using 4 groups, with redundancy. This was compared with a system involving twelve descriptions without redundancy. The GBFOS algorithm along with the weights as derived in Section 2.2, gives 85% of the overall rate for the primary. There is not much variation in the rate allocation with variation in packet loss, since the value of $\varepsilon$ is raised to a high integer power and is small to begin with this. The average PSNR values over forty frames for twenty realizations with and without redundancy are shown in Figure 6. The MDC scheme with redundancy performs better by a maximum of about 6 dB ($\varepsilon = 0.35$). As expected there is no need for redundancy when there is no packet loss.

4. CONCLUSION

This paper proposed a new multiple description video coder using 3D wavelet transforms and the EBCOT compression algorithm.
A uniform grouping scheme was proposed to select coefficients into basic units called groups used to make the descriptions. It was shown that this scheme outperformed a hierarchical selection scheme by about 7 dBs at best and about 3 dBs at worst. An R-D optimization framework was proposed to allocate rates between the positions in a description in a way that is consistent with the packet loss statistics of the network. This optimization provides substantial quality improvements for the packet loss rates that were tested out.

Figure 6: Visual comparison of our algorithm with the hierarchical selection scheme when two out of 49 descriptions are lost. The left column shows the reconstructed four frames using uniform grouping scheme and the right column shows the reconstructed frames using hierarchical grouping scheme.

5. REFERENCES


