A Novel Error Resilient Temporal Adjacency Based Adaptive Multiple State Video Coding over Error Prone Channels

Ragunathan M. and Mala C.

National Institute of Technology (Deemed University, Govt. of India),
Department of Computer Science and Engineering,
Tiruchirappalli, Tamil Nadu, India. Pin Code – 620 015
`mragunathan@gmail.com, mala@nitt.edu`

Abstract. Video streaming applications have been gaining interest rapidly in various perspectives from entertainment to e-learning. Practically, these applications suffer from inevitable loss in the transmission channels. Hence it is a challenging task to improve the quality of video streaming over the error prone channels. Multiple Description Coding (MDC) is a promising error resilient coding scheme which sends two or more descriptions of the source to the receiver to improve the quality of video streaming over error prone channels. Depending on the number of descriptions received, the reconstruction-distortion gets reduced at the receiver. Multiple State Video Coding (MSVC) is a MDC scheme based on frame-wise splitting of the video sequence into two or more sub-sequences. Each of these sub-sequences is encoded separately to generate descriptions, which can be decoded independently on reception. Basic MSVC is based on the separation of frames in a video into odd and even frames and sending each part over a different path. The drawbacks or certain subtleties of the basic MSVC such as lack of meaningful basis behind the frame wise splitting, inability to support adaptive streaming effectively, less error resiliency are brought out and discussed. Thus to overcome them and to improve the quality of video streaming, the design of a novel MSVC scheme based on the temporal adjacency between video frames is proposed in this paper. This temporal adjacency based splitting of the video stream into N sub-sequences also enables the proposed scheme to adapt to varying bandwidths in heterogeneous environments effectively. The simulation results show that the proposed scheme also outperforms Single State Video Coding (SSVC) scheme in terms of the sensitivity of perception of the reconstructed video sequence, under various loss scenarios.

1   Introduction

The demand for good quality Multimedia services over networks has been growing. Particularly Video streaming applications have been gaining interest rapidly in various perspectives from entertainment to e-learning. In these applications, since video data are voluminous it requires high compression before transmission using various encoding schemes. The transmission systems of the video streaming

applications rely on the reliability of the networks and impose stringent timing requirements as they deal with time based media. But all real networks are unreliable despite wired or wireless. Hence practically errors are inevitable and these applications suffer from inevitable loss in the transmission channels. This lossy transmission affects the quality of the reconstructed video. There exist conventional methods for resolving these packet losses and improving the quality of streaming such as Retransmission [11] and Forward Error Correction (FEC) [9]. However, if the transmission system was not able to afford an additional back-channel, or if the additional Round Trip Time (RTT) delay was not bearable, then the Retransmission cannot be employed. Also FEC approaches are designed to be effective only for losses less than a threshold. Obviously this cannot be guaranteed due to the highly dynamic nature of the networks. Since it is almost impossible to guess the threshold value, the scheme proves to be inefficient. Thus the demand for an Error resilient coding scheme naturally increases. MDC [6] is a promising error resilient coding scheme which sends two or more descriptions of the source to the receiver to improve the quality of video streaming over error prone channels. Depending on the number of descriptions received, the reconstruction-distortion gets reduced at the receiver. MSVC is a MDC scheme based on frame-wise splitting of the video sequence into two or more sub-sequences. Each one of these sub-sequences can be encoded separately to generate descriptions, which can then be decoded independently on reception. The advantages of MSVC are that the streams are independently decodable and they provide bi-directional data, that is, certain past and future frames are known appropriate to any instant of time. This helps in effective reconstruction as the received independent frames can recover the state of corrupted streams. The basic MSVC scheme [2] suggests the splitting of the given sequence into two sub-sequences. The original sequence is split frame-wise into two groups. One consisting of the odd frames and the other consisting of the even frames. Each of these two groups is separately encoded to form two descriptions or sub-sequences, one containing the odd and the other containing the even frames. Then these two sub-sequences are sent to the destination through diverse paths. Due to the diverse paths maintained, the reception would be experiencing only the average behavior of all the channels through which the descriptions are sent. Naturally this improves the performance of the system. The improvement in performance can be justified by the argument that the probability that all of the multiple paths will be congested simultaneously is less than the probability that the single path is congested.

Our goal in this paper is to propose a novel Multiple State Video Coding scheme. The novelty of the proposed scheme is the way by which the given video sequence is split into sub-sequences. In the basic MSVC scheme suggested in [2], [4] the basis of splitting the video frames is thze frame number i.e. whether they are odd or even. Obviously these subsequences alone or together can provide an acceptable quality of reconstruction despite losses. However the notion behind the scheme is purely mathematical. Almost the same effect can be got by splitting the video into three sub-sequences, where one containing the frames which are divisible by 3, the other containing the frames divisible by 2, and the third sub-sequences that are not divisible by both 2 and 3. Thus the absence of a meaningful notion behind the splitting of the video frames place no boundary on, the basis of splitting and the number of sub-sequences to be generated. The proposed scheme overcomes these subtleties with a

meaningful notion as its basis. The notion behind the proposed MSVC scheme is the *summary* and *non-summary* frames. The summary frames are the snapshots or relatively a small number of frames of the original complete video sequence which provide the summary of the video content. Naturally the frames that are not summary frames are non-summary frames. In short the summary frames are those frames which are temporally adjacent, that is they are visually dissimilar or their content difference is large. The interesting aspect of this notion is that the summary and non-summary frames alone / together can provide a comprehendible reconstructed video. This aspect is interesting since it holds despite the fact that the summary frames contain a relatively small number of frames of the original video, whereas the non-summary frames contain relatively large number of frames. This is due to the characteristics of the summary frames, that they can summarize the whole content with relatively small number of frames but they are substitutable by the non-summary frames. Through the reception of summary frames we can ensure a comprehendible video, however their loss are substitutable by the non-summary frames, thereby increasing the error resiliency. Also the transmission of summary frames requires less bandwidth compared to that of non-summary frames / original sequence, thereby increasing the adaptive nature when streaming to heterogeneous clients. Thus a novel error resilient temporal adjacency based adaptive MSVC scheme is proposed. Then the performance of the proposed scheme is compared with the SSVC scheme, in terms of the sensitivity of perception of the reconstructed video sequence. The performance analysis is done *shot wise* under various loss scenarios such as single and burst error, when the channel losses are independent and dependent. Finally a discussion on the ability of the scheme to adapt to the varying bandwidths is presented. Further the discussion investigates the application of the proposed scheme in the context of adaptive streaming to heterogeneous clients.

The remainder of the paper is organized as follows. Section 2 discusses the existing schemes, MDC and basic MSVC. In Section 3, the design of the proposed scheme is presented. The performance analysis over SSVC is presented in Section 4. This is followed by Section 5, where the effectiveness of the proposed scheme over the basic MSVC scheme [2] is discussed. Section 6 concludes the paper.

2   Existing Schemes

Multiple Description Coding (MDC) has emerged as a power framework for robust coding and transmission of the video data over lossy networks. The two main problems of the real networks [5] are the limited bandwidth and the packet losses. Limitation on bandwidth is natural and can be resolved by duly considering it. The conventional methods of resolving packet losses such as retransmission and FEC cannot be employed effectively as mention in Section 1. MDC is a promising error resilient coding scheme. MDC was invented in Bell Labs primarily for telephone conversations. But it is used for image, video/audio transmission. Multiple Description Coding *"represents a single information source with several chunks of data (i.e. descriptions) so that the source can be approximated from any subset of chunks"* [3]. MDC is applied to some major coding techniques such as scalar quantization, vector quantization of motion vectors [10], correlating transforms, or

quantized frame expansions. A summary of the state-of-the-art system designs can be found in [3]. High rates, low latency requirements, and error drift are however the main problems encountered in MDC schemes for video streaming due to possible desynchronization of encoders and decoders [4].

MSVC is a MDC scheme which is inspired by the frame-wise splitting of the video stream into sub-sequences. Each of these sub-sequences can be independently decodable. Thus MSVC is a kind of MDC with the novelty of the additional state recovery capability [1]. The MSVC scheme is less complex than MDC. The other advantages of MSVC include low delay property, nil error propagation, and the bi-directional information the streams provide. The goal of MSVC is to combine the high compression and the high error resilience. The advantage of MSVC over SSVC schemes is that in MSVC the streams received provide bi-directional information, that is the information about the future and the past frames received through multiple descriptions, which help in increasing the state recovery property during reconstruction. Whereas in SSVC when a previously decoded frame is lost the quality degrades until the state is refreshed by the next I-frame. Basic MSVC is based on the separation of video frames into odd and even frames in a stream and sending each part over a different path. Here the path diversity is maintained in transmitting the two descriptions. This is to achieve a better reconstruction quality since the independent loss patterns in the two channels results in only the average path behavior. However the basic MSVC have certain drawbacks such as lack of meaningful basis behind the frame wise splitting, inability to support adaptive streaming effectively and less error resiliency. To overcome them and to improve the quality of video streaming, the design of a novel MSVC scheme based on the temporal adjacency between video frames is proposed in the next section. Complete information about the basic MSVC scheme can be found in [1].

3   The Proposed MSVC Scheme

This section first introduces the basic idea of summary and non-summary frames behind the proposed MSVC scheme. The aspects of the proposed scheme's basic notion which increases the error resilience capability and adaptive nature of the coding scheme are then discussed. The algorithm for selecting the summary frames from a given video stream is presented. This is followed by the design of a two-state MSVC following the proposed scheme. Then the necessary details about extending this novel MSVC scheme to N states are presented.

The inspiration for the notion of summary and non-summary frames is got from [8], where the summary and non-summary frames were introduced and used for developing a disruption tolerant content aware video coding. The proposed scheme incorporates the notion of summary and non-summary frames with subtle changes into the context of MSVC.

The 'summary frames' in short are the snapshots or relatively a small number of frames of the original complete video sequence which provide a summary of the video content. As the name suggests, by viewing the summary frames of a video the user can comprehend the content of the video. Naturally, if a set of frames are summary frames of a video, then the frames that are not summary frames are the set

of 'non-summary frames'. It can be well perceived that the non-summary frames can provide a video of quality very close to the original complete sequence. Since only a relatively small number of frames of the original sequence form the summary frames, the video is not much altered without them.

The interesting aspect of this notion is that the summary and non-summary frames alone/together can provide a comprehendible reconstructed video. This aspect is interesting since it holds despite the fact that the summary frames contain relatively small number of frames of the original video, whereas the non-summary frames contain relatively large number of frames that are not summary frames. This is due to the characteristics of the summary frames, that they can summarize the whole content with relatively small number of frames but they are substitutable by the non-summary frames. This follows that the reception of summary frames ensures the comprehension of the video, however their loss are substitutable by the non-summary frames, thereby increasing the error resiliency of the scheme. Also the transmission of summary frames requires less bandwidth compared to that of non-summary frames/ original sequence, thereby increasing the adaptive nature of the scheme when streaming to heterogeneous clients.

3.1   Algorithm - For Selecting the Summary Frames of a Given Video Sequence

The algorithm for generating two sets of frames i.e. the summary frames and non-summary frames from a given video sequence is given below.

Let F be the set containing all the frames in the given sequence. Let Fs denote the summary frames set and Fns denote the non-summary frame set. Let L be a list built of all the frames in the set F and f_i denote the frame i in the list L. Let D be a two dimensional array containing the content difference of the frames in the list L. For example D(i,j) contains the content difference of the frames L(i) and L(j).

```
Initially the set Fs contains all the frames of the
video and the set Fns has no frames i.e. empty.

1. Find the minimum content difference D(i,j)in the
array D and delete the frame fi if the frames fi and its
predecessor are less temporally adjacent than that of
fj and its successor. Else delete the frame fj from the
set Fns.

2. Include fi or fj to the set Fs depending on whether
the frame fi or fj of is deleted, respectively.

3. Update the D array with the content difference
between the frame deleted and its adjacent frame.

4. Repeat steps 1 and 2 until there is no frames in Fs
or when the number of frames in Fns is equal to R.
```

The variable R can be defined as the ratio between the total number of frames present in the original video to the number of summary frames required. Naturally with this variable the number of summary frames and how densely are they required to be spaced can be controlled. A basic interpretation may be - if R is large the summary frames will be too sparsely spaced and vice versa.

3.2   Design of a Two-State MSVC Based on the Proposed Scheme

For the video sequence to be transmitted two subsequences are generated – one containing 'the summary frames' and the other containing 'the non-summary frames'. This can be done by processing the given video sequence with the above algorithm to generate two sets of frames F_s and F_{ns} .These two sets of frames are encoded separately to form two sub-sequences or descriptions that are independently decodable. These two subsequences are transmitted to the destination following two different paths. When at the reception the summary frame sequence alone is received, it can provide the summary of the video. If the non-summary frames sequence alone is received it can also provide the video sequence of quality nearly equal to the original sequence as said before. If both are received with some losses then they can be decoded and combined to get an optimum quality video. Thus we have two descriptions of the video and thereby a two state MSVC. Since the proposed notion of summary and non-summary frames is used, this forms the design of a two-state MSVC based on the proposed scheme.

3.3   Extending the Design to N Descriptions or State

In the above algorithm for a video sequence two sets of frames F_s and F_{ns} are got as output. Each of these two sets of frames is encoded to generate a description. Thus we have two descriptions and hence the two state MSVC. To extend it to more than two states say 'n', 'n' descriptions have to be generated. To generate 'n' descriptions, 'n' set of frames are necessary as it was two sets F_s and F_{ns} for n=2 i.e. in two state MSVC. This can be done by applying the above discussed algorithm again with F_{ns1} (let Fns1 be the non-summary frames got by applying the algorithm for the first time on the original sequence) as F. By this two sets of frames F_{s2} and F_{ns2} can again be got. This is continued (n-1) times to get 'n' sets: $F_{n(1)}$, $F_{n(2)}$,$F_{n(3)}$…$F_{n(n-1)}$ and $F_{ns(n-1)}$. Thus we have n-1 summary frame sets and 1 non-summary frame set, totally 'n' sets. From these 'n' sets 'n' descriptions can be generated and thereby this leads to an 'n' state MSVC technique.

4   Performance Analysis over SSVC

This section provides the results of performance analysis of the proposed MSVC scheme over the SSVC scheme. A 3-state MSVC codec based on the proposed scheme is developed using C-Language. The encoder encodes the given video sequence into three sub-sequences namely ns, s1 and s2. Where s1 is the sub-sequence got by summarizing the given video with the summarizing ratio R1 and let ns' be the resulting non-summary frames. s2 and ns are the summary frames and non-summary frames got by further summarizing the non-summary frames ns' with the summarizing ratio R2. The error patterns while transmitting the sub-sequences to the destination through diverse paths are simulated. The error pattern simulation is done for various scenarios.

   The comparison is done in terms of Sensitivity Of Perception (SOP), which is a number signifying the sensitivity perceived by the user when a transition from one frame to the other takes place while streaming a video. The decoder accepts the

frames transmitted after error simulation is done manually and it decodes them independently. The lost frames are reconstructed by inserting *black frames* in order to improve the sensitivity of perception and thereby making the performance analysis clearer.

When the sub-sequences are transmitted via three diverse paths, the losses in the three channels may be independent or dependent. The loss patterns of the three channels will be independent if they do not have any partial common path. But the losses will be dependent if two or three of them have partial common paths. Now the performance is analyzed for both the cases of independent and dependent channel losses. And in each of these two cases both the single frame loss and burst frame losses are considered. Thus an analysis of SSVC over proposed MSVC is done for four scenarios.

For this a reference video sequence that has 110 frames and 4 shots is considered. However our analysis is confined to the first shot which consists of frames 0-38. The summary ratios R1 and R2 used are 4 and 2 respectively.

The graphs got by simulation are given in Fig.1 and Fig.2. Fig 1 corresponds to the Scenario 1- Single frame loss in two channels (independent) and Fig 2 corresponds to the Scenario 2- Burst frame loss in two channels (dependent).
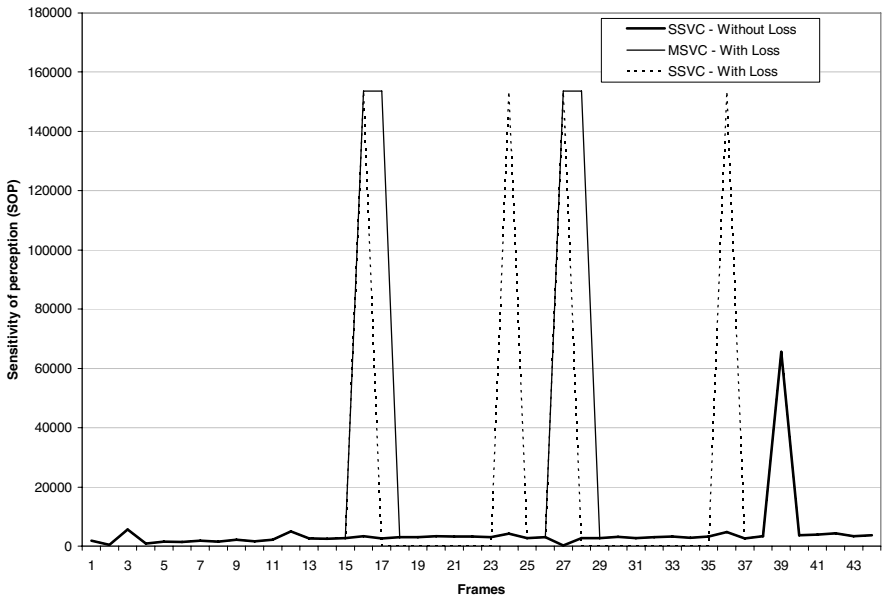

Fig. 1. Scenario-Single frame loss in two channels (independent)

Scenario 1: In this scenario frame 27 is lost in channel s1. And frame 16 is lost in channel s2. The dark thick line corresponds to a streaming of the normal lossless video. The dotted line corresponds to the streaming of the reconstructed SSVC – with loss. The normal dark line corresponds to that of MSVC. It can be seen that both the SSVC (dotted) and MSVC (normal) are getting deviated at the frame number 16,

since it is lost (to improve the SOP this frame is replaced by black frame for analysis sake). The MSVC returns to normal streaming at frame 17 but the SSVC returns to the normal streaming only after the state is refreshed by the next I-frame (frame 24) received. Thus the viewer is subjected to a bad SOP for long time in SSVC, whereas it is not so in MSVC. The same argument holds for the loss of frame 27 in channel s2.
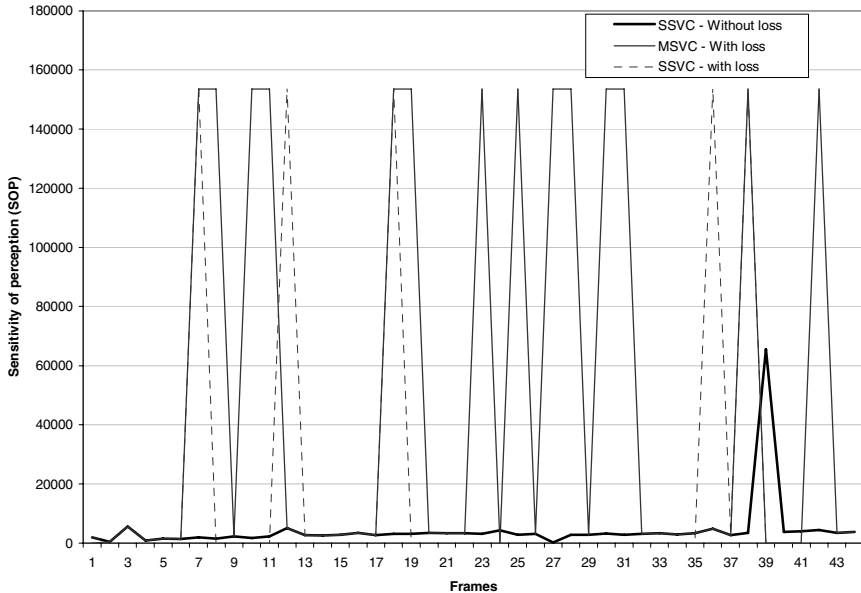

Fig. 2. Scenario-Burst frame loss in two channels (dependent)

Scenario 2: In this scenario frames 7, 10, 27 are lost in channel s1 and frames 18 23 24 30 38 are lost in channel ns. Here too the viewer is subjected to a bad SOP for long time in SSVC, whereas it is not so in MSVC. This can be seen from the graph of Fig.2.

5   Effectiveness of the Proposed Scheme over the Basic MSVC Scheme

The effectiveness issues apart from those mentioned in the introduction part of this paper is presented in this section.

The performance improvement of the basic MSVC over SSVC is justified by the argument that the probability that all of the multiple paths will be congested simultaneously is less than the probability that the single path is congested [7]. Following the same argument the proposed scheme extends the two states to N states to improve the performance further since the probability that all of the N paths will be congested simultaneously is less than the probability that the two paths are simultaneously congested. Also the loss of frames in either of the channel where the

odd / even frames sub-sequences are sent will just mean loss of frames. Nothing can be inferred about the reconstruction quality from the knowledge of losses in these channels before transmission by any of the channel look up techniques. But the knowledge about the loss of frames in any of the channels through which the summary or non-summary sub-sequences are sent convey meaning. Suppose the loss is only in the non-summary frames implies that we can have a quality of video that is comprehendible. Similarly the loss of frames only in the channel through which the summary frames sub-sequence is sent implies that they can be substituted by the non-summary frames that are receive, thereby implying an acceptable quality of reconstruction.

6   Conclusions

The work presented in this paper explored the subtleties of the MSVC scheme proposed in [2], [1]. The exploration resulted in the conclusion that the basis of frame-splitting is purely mathematical. Thus the absence of a meaningful notion behind the splitting of the video frames in the basic MSVC [1], [4] placed no boundary on the basis of splitting and the number of sub-sequences to be generated. Also a drawback of the basic MSVC, that it does not possess effective adaptive streaming capability is opened up. To overcome the drawbacks a novel MSVC scheme is proposed. The scheme is based on the notion of summary and non-summary frames. The justification for the better performance of the proposed MSVC scheme over the basic MSVC scheme, due to its meaningful notion behind splitting the video frames is presented. The performance is better in terms of quality of the video streaming over error prone channels and the effectiveness of adaptation when streaming to heterogeneous clients. The codec for the proposed MSVC is developed and the various loss scenarios when the encoded video stream is transmitted are simulated. The simulation results concluded that the proposed MSVC scheme outperformed the SSVC, in terms of the Sensitivity Of Perception (SOP).

References

1. Vorgelegt von, Diplom Ingenieurin, Siila Ekmekciaus, Frankfurt am Main. "Multi-State Video Coding over Error Prone Channels", Thesis,Technical University Berlin, December 2004.
2. J. G. Apostolopoulos. "Reliable video communication over lossy packet networks using multiple state encoding and path diversity", Visual Communications and Image Processing, Proceedings of SPIE, vol. 4310, pp.392–409, 2001.
3. V. K. Goyal. "Multiple description coding: compression meets the network", IEEE Signal Processing Magazine, vol. 18, no. 5, pp. 74–93, 2001.
4. Sila Ekmekci, Thomas Sikora. "Unbalanced Quantized Multiple State Video Coding". EURASIP Journal on Applied Signal Processing, Article ID 14694, pp.1–10, 2006.
5. Fisichee Naturali, Notes on "Networked Multimedia Systems - error detection and control schemes", Università degli Studi di Verona, 2001.
6. Raman Venkataramani. "Multiple Description Coding with Many Channels". DIMACS workshop on network information theory, Harvard University, Cambridge, MA, 2003.

7. Sila Ekmekci, Thomas Sikora. "Multi-State vs. Single-State Video Coding over Error-Prone Channels", in the Thirty-Seventh Asilomar Conference on Signals, Systems and Computers, vol.2, pp.1544-47, Nov. 2003.

8. Tiecheng Liu and Srihari Nelakuditi. "Disruption-Tolerant Content Aware Video Streaming", Proceedings of the 12th annual ACM International Multimedia Conference, pp. 420 – 423, 2004.

9. R. Puri, K. Ramchandran, K.W. Lee and V. Bharghavan. "Forward error correction (FEC) codes based multiple description coding for Internet video streaming and multicast", Signal Processing: Image Communication, vol. 16, no. 8, pp. 745–762, 2001.

10. C.S. Kim and S.U. Lee. "Multiple description motion coding algorithm for robust video transmission", Proceedings of IEEE International Symposium on Circuits and Systems, vol. 4, pp. 717–720, Geneva, Switzerland, May 2000.

11. Injong Rhee and Srinath R. Joshi. "Error Recovery for Interactive Video Transmission over the Internet", IEEE Journal on Selected Areas in Communication, vol.18, no.6, pp. 1033-1049, June 2000.