An analysis for the paper

Interactive Streaming of Stored Multiview Video Using Redundant Frame Structures [1]

Richard A. Kramer, Oregon State University, Member IEEE

Abstract

This paper's objective it two-fold:

First, this paper provides an analysis of the paper "Interactive Streaming of Stored Multiview Video Using Redundant Frame Structures" [1] and the related subject matter.

Second, in the sub-section entitled "Weaknesses of the Proposed Solutions / Additional Considerations", this paper presents a future vision for research to take the subject matter of "interactive multiview video" to the next level. Specifically I introduce the concept of interactive *free viewpoint live* multiview video streaming *using network coding*.

Specific to "Interactive Streaming of Stored Multiview Video Using Redundant Frame Structures" [1], the paper teaches:

1) The use of an interactive network,
2) for a client to send a request to a server to switch frame views for multiview video playback,
3) while optimizing storage and transmission costs based on frame structures.

> *"Thus, as a client is playing back successive frames (in time) for a given view, it can send a request to the server to switch to a different view while continuing uninterrupted temporal playback. Noting that standard tools for random access (i.e., I-frame insertion) can be bandwidth-inefficient for this application, we propose a redundant representation of I-, P-, and "merge" frames, where each original picture can be encoded into multiple versions, appropriately trading off expected transmission rate with storage, to facilitate view switching.* [1] at Abstract.

Thus, the objective of "Interactive Streaming of Stored Multiview Video Using Redundant Frame Structures" is to optimize and evaluate the trade-offs between server storage (which the paper refers to as "storage cost") and transmission bandwidth requirements (which the paper refers to as "transmission cost") using frame structures.

In order to evaluate this subject matter, a thorough understanding of multiview coding is (was) required. Recommended reading includes: (1) "Overview of the Stereo and Multiview Video Coding Extensions of the H.264/MPEG-4 AVC Standard" [2], and (2) the actual H.264/MPEG-4 AVC (Advanced Video Coding) standard [3]. This material is further summarized in my presentation entitled "An Introduction to the Problem: Interactive Free Viewpoint Live Multiview Video Streaming Using Network Coding" [7].

"Interactive Streaming of Stored Multiview Video Using Redundant Frame Structures" teaches optimization and evaluates the impact on storage cost versus transmission cost using various video frame structures including:

1) I-Frames (Intra-frames)
2) P-Frames (Prediction-frames)
3) M-Frames (Merge-frames–explained below)

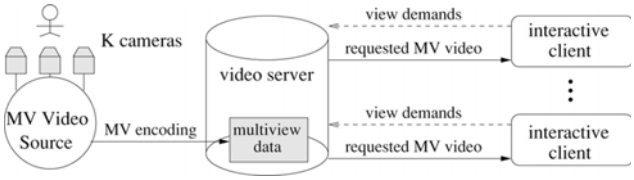
Thus, in this paper, I provide a detailed analysis of "Interactive Streaming of Stored Multiview Video Using Redundant Frame Structures" organized into the following sections:

1. **The Problems to be Solved**
2. **The Proposed Solutions**
3. **Strengths of the Proposed Solutions / Effectiveness**
4. **Weaknesses of the Proposed Solutions**
5. **Additional Considerations**
6. **Conclusion**

Overall, "Interactive Streaming of Stored Multiview Video Using Redundant Frame Structures" is truly a tradeoff analysis, and offers frame structures that "reduc[e] expected transmission rate by up to 45% compared to I-frame insertion approach, at twice the storage costs" (*see* [1] at pg. 746). Thus while optimization in transmission cost is obtained, it is at the expense of storage cost and vice-versa based on the frame structures used.

1. The Problem to be Solved

To expand, the paper "Interactive Streaming of Stored Multiview Video Using Redundant Frame Structures" teaches an interactive multiview video streaming system. As shown immediately below, the interactive multiview video streaming system entails "K" cameras, a video server for the storage of the multiview video content and interactive clients[1].



([1] at Fig. 3)

From above, the objective of "Interactive Streaming of Stored Multiview Video Using Redundant Frame Structures" is to evaluate the trade-offs between server storage and transmission bandwidth requirements by introducing and evaluating alternative frame structures.

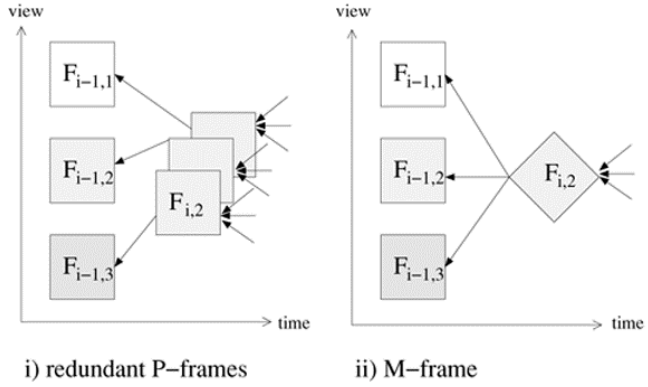
> *"In this paper, we develop heuristics and optimization algorithms that construct good frame structures for IMVS* [Interactive Multiview Video Streaming] *using available I-, P, and implementations of M-frames as building blocks.* [1] at pg. 746.

[1] Oddly, while "Interactive Streaming of Stored Multiview Video Using Redundant Frame Structures" shows multiple interactive clients, the focus is for optimization *for only a single client* which I find to be a major deficiency.

2. The Proposed Solutions

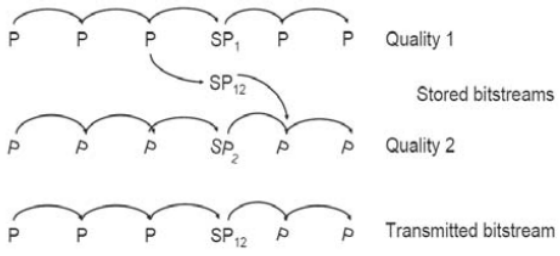
In order to optimize storage cost and transmission cost, "Interactive Streaming of Stored Multiview Video Using Redundant Frame Structures" introduces a number of alternative frame structures, called M-Frames (Merge Frames) and further considers the impact of probability that a user will change the view versus the user continuing to view the present view.

A. The M-Frame is the combination of multiple frames (views) into a single frame structure, thus rather than storing and/or transmitting multiple I-Frames or P-Frames in order to render a new view in the future, instead, only a single M-Frame is required as shown below, whereas a square "F" represents a generic frame (I- or P-frame) of a given view and the diamond "F" represents an M-frame for multiple views.



([1] at Fig. 1(a))

In all, three (3) new/different M-Frame structures are introduced and are further evaluated. Each of the proposed M-Frame types are based on *Distributed Source Coding* (DSC – see glossary at end) and *SP-Frames*, whereas SP-Frames (Super P-Frames) are part of the ITU-T H.264 standard [3] and allow switching between two video streams at a random moment in time [4].



([4] at Fig. 1)

The three (3) types of M-Frames considered are:

1) ***DSC0-Implementation of M-Frame*** which entails encoding a single DSC component of all possible transitions into a future new view.
2) ***DSC1-Implementation of M-Frame*** which entails encoding the differentials for all possible transitions into a future new view, using "P" (Predictor) components.
3) ***DSC0+1-Implementation of M-Frame*** which entails taking the output from the *DSC1-Implementation of M-Frame* process described above to generate a single DSC component as described in the above *DSC0-Implementation of M-Frame* process.

> *"In a nutshell, good frame structures should contain the "right" mixture of redundant P-frames (for bandwidth efficiency) and M-frames (for storage efficiency) for a given Lagrangian multiplier.* [1] at pg. 752.

B. The impact of probability (α) to change views: As an additional consideration, "Interactive Streaming of Stored Multiview Video Using Redundant Frame Structures" considers the probability that a user will actually change views versus maintain the current view. From this, the probability that the initial I-Frame view from camera K is shown to be $q(I_{0,K}^{0})=1$, and the probability that an alternative view is selected is determined by the sum of the probabilities (number of) alternative views $q(F_{i,j})$ that the user can select a new view, scaled by the probability (e.g. likelihood) that the user will make a transition (α) to a new view as follows:

$$q(I_{0,K^o}) = 1$$
$$q(F_{i+1,k}) = \sum_{F_{i,j}|F_{i,j}\leftarrow F_{i+1,k}} q(F_{i,j})\alpha_{i,j}(k).$$

Where the variable "I" represents an I-Frame, "F" represents a frame at a point in time, "K" refers to a camera view, "i" is for the frame in time and "j" denotes a specific view.

I consider both the use of M-Frames and the use of probability α of view selection to be extremely useful in order to optimize storage and transmission costs because both factors have a direct and dramatic impact on both server storage and transmission costs.

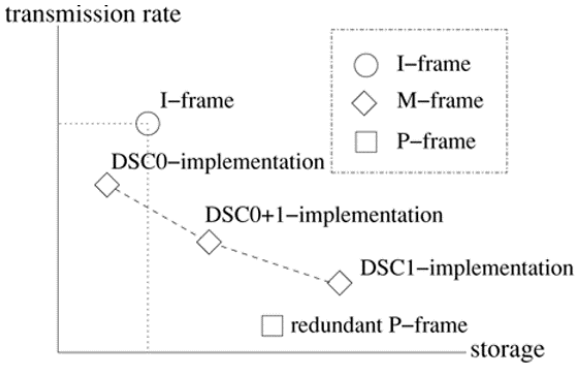
3. Strengths of the Proposed Solutions / Effectiveness

In my opinion, "Interactive Streaming of Stored Multiview Video Using Redundant Frame Structures" provides an effective solution for optimizing frame structures for the interactive viewing of stored multiview video content and provides other results as follows:

A. Frame Structure optimization: "Interactive Streaming of Stored Multiview Video Using Redundant Frame Structures" provides a trade-off analysis of the abovementioned "M-Frame" structures as compared to traditional "I-Frame" and "P-Frame" structures[2]. The analysis of each frame structure is as follows:

1) **"I-frame":** In this scenario, only native camera I-Frames are stored and the decoder at the client generates the redundant P-frames, thus minimizing the amount of video stored at the server. As would be expected, the scenario of only I-Frames results in the highest transmission rate (worst cost) and near lowest storage cost. This matches the logical expectation that spatially compressed video alone, (with no additional "P-frames" being stored coincident with the I-Frames) would require the least storage at the server and required the highest transmission rate.

[2] I note that an analysis of the standard "B-Frame" (Bi-directional Frame) structure is absent.



([1] at Fig. 1(b) showing trade-off of the various frame structure types)

2) **"Redundant P-frame":** In this scenario, redundant P-frames are generated and stored at the server, along with the I-frames. Because primarily only P-frames are transmitted to render a future view, this scenario requires the lowest transmission rate, and near highest storage cost.

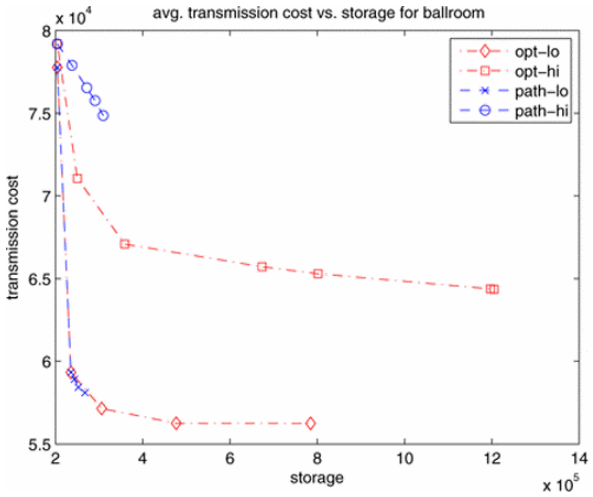
3) **"DSC0-implementation":** This frame structure (see above) offers improved storage and transmission costs as compared to the "I-Frame" scenario.

4) **"DSC0+1-implementation" and "DSC1–implementation"**: These frame structures (described above) offer additional trade-offs between storage and transmission costs.

B. Probability that the user will change views: "Interactive Streaming of Stored Multiview Video Using Redundant Frame Structures" provides an analysis of the probability α (see discussion above) for two *sets* of frame structures, whereas each frame structure *set* represents a collection of views over time. The frame structure sets are: (1) a "path" frame structure set, and (2) a "tree" frame structure set.

The path frame structure set assumes a low probability α that the user will switch views and subsequently, the frame structure set is optimized for a direct *path* (e.g., little to no adjacent view information is transmitted to the client).

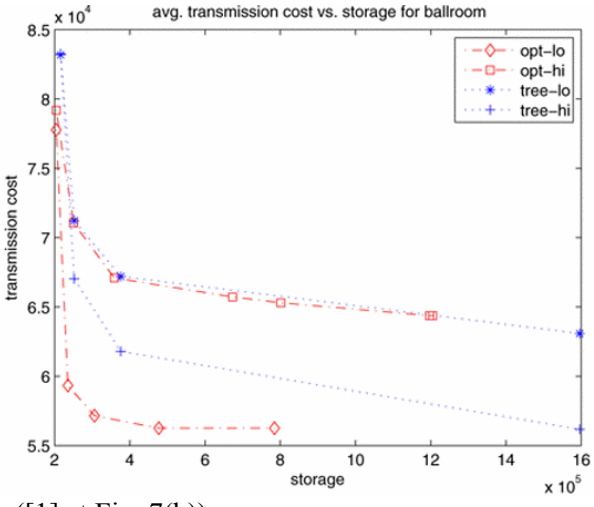
The results are shown below; the plot illustrates the transmission and storage costs *for a path frame structure set for infrequent (path-lo) versus frequent (path-hi) view changes.* From this analysis, near optimal performance is obtained when view changes are infrequent for a path frame structure set and conversely there is a heavy penalty for frequent view changes using a path frame structure set.



([1] at Fig. 7(a))

The tree frame structure set assumes a high probability that the user will change the view, thus a tree frame structure is provided, whereas the branches of the frame structure represents transmitted information to generate adjacent views.

The results are shown below; the plot illustrates the transmission and storage cost for *a tree frame structure set for infrequent (tree-lo) versus frequent (tree-hi) view changes.* As shown, near optimal performance is obtained when view changes are frequent for a tree frame structure set and there is a heavy penalty for infrequent view changes using a tree frame structure set.



([1] at Fig. 7(b))

I find that both (1) frame structure and (2) the probability α of a user switching a view are important because both factors have a direct and significant impact on storage and transmission costs. Nonetheless, while these factors are important, the effects of these factors on storage and transmission costs seem obvious.

4. Weaknesses of the Proposed Solutions

While I found "Interactive Streaming of Stored Multiview Video Using Redundant Frame Structures" to be insightful regarding the use of M-frames and the subsequent analysis of storage costs versus transmission costs, I found the that overall the paper was lacking in a number of areas as follows:

A. Data sharing between multiple clients was not considered: While the premise of "Interactive Streaming of Stored Multiview Video Using Redundant Frame Structures" relates to interactive clients (plural), I found the paper is deficient in accommodating *multiple* interactive clients. Such a deficiency fails to exploit the highly redundant nature of multiview video streaming data between clients.

B. Network characteristics where not considered: In the evaluation of the various methods taught by "Interactive Streaming of Stored Multiview Video Using Redundant Frame Structures" network latency, network errors, and other real-world network characteristics were not considered.

C. Layer separation and/or reuse/redundancy of non-VCL (Video Coding Layer) data as a means to reduce bandwidth was not disclosed: As explained in my accompanying presentation [8] entitled: "An Introduction to the Problem: Interactive Free Viewpoint Live Multiview Video Streaming Using Network Coding", significant data redundancies exist within each frame (M-Frame or otherwise); these data redundancies are separated into layers and can be sent separately (and only once).

Additional considerations are discussed immediately below in Section **"Additional Considerations"**.

5. Additional Considerations

I reviewed the paper "Interactive Streaming of Stored Multiview Video Using Redundant Frame Structures" because of its relevancy to the state-of-the art in the reduction of bandwidth over networks for interactive multiview video streaming content. With that in mind, I found many additional areas of research that warrant attention. Below, I outline a number of research areas and the related constraints that should be considered for a combined research area which I call "*interactive free viewpoint live multiview video streaming using network coding*":

A. Optimization of network bandwidth using live (versus stored) multiview video content: I propose research in the area of live multiview video streaming *as opposed to stored multiview video content.* With the additional complexity of *live m*ultiview video streaming, *network latency becomes an additional constraint in addition to the bandwidth constraint of the network.* Further, to obtain superior compression, B-Frame generation requires both a past "reference frame" as well as a "future" frame, thus adding additional complexity as compared to the methods taught in "Interactive Streaming of Stored Multiview Video Using Redundant Frame Structures".

B. Multiple interactive clients: Based on the fact that differing views between disparate clients within a multiview video system have highly redundant data, an extension of the abovementioned sub-section "**A. Optimization of network bandwidth using live (versus stored) multiview video content"** should include *optimization to exploit the interdependencies of data between multiple clients*.

C. Network optimization based on the separation of non-VCL data from VCL data, recognizing that non-VCL data is often times highly redundant. Additional bandwidth efficiency can be obtained by separating multiview video streaming information *into layers rather than by frame structures alone.*

D. Network coding in P2P networks: Because multiview video content provides redundant information in layers as discussed immediately above, *opportunities to apply network coding*

such as Hierarchal Network Coding (HNC) [8] exists. To expand, HNC provides improved performance and low latency in environments where data redundancy is less than zero; thus offering improved server storage and server synchronization which can further be contrasted with WNLC (Within Layer Network Coding).

E. Prediction frames generated based on a given node having a "reference frame" rather than transmitting the information from the source. It would seem that additional network efficiency can be obtained by utilizing "reference frames" from other nodes and/or end-points in a Markov network as opposed to transmitting the "reference frame" from the source to each client *and from that, only generating prediction information to render a unique view for each user*.

Conclusion

I found the paper "Interactive Streaming of Stored Multiview Video Using Redundant Frame Structures" insightful and effective in providing insight into frame structure trade-offs related to storage and transmission costs. The paper was also effective in teaching what is near state-of-the art for interactive multiview video streaming. Based on this foundation (and deficiencies), additional research related to *interactive free viewpoint live multiview video streaming using network coding* invites significant additional research.

References

[1] Cheung, et al., "Interactive Streaming of Stored Multiview Video Using Redundant Frame Structures", IEEE Transactions on Image Processing, Vol. 20, No. 3, March 2011.

[2] Vertro, et al., "Overview of the Stereo and Multiview Video Coding Extensions of the H.264/MPEG-4 AVC Standard", Proceedings of the IEEE, 2011.

[3] ITU-T and ISO/IEC JTC 1, "Advanced Video Coding for Generic Audiovisual services, ITU-T Recommendation H.264", ISO/IEC 14496-10 (AVC), 2003.

[4] Setton, et al., "Video Streaming with SP and SI Frames", Stanford University, undated.

[5] Cheung, et al., "Distributed source coding application to low-delay free viewpoint switching in multiview video compression," Picture Coding Symp., PCS'07, Lisbon, Nov. 2007.

[6] Cheung, et al., "Distributed source coding techniques for interactive multiview video streaming", 27th Picture Coding Symp., May 2009.

[7] Kramer, "An Introduction to the Problem: Interactive Free Viewpoint Live Multiview Video Streaming Using Network Coding", Oregon State University, 2016.

[8] Kien Nguyen, Thinh Nguyen, and Sen-Ching Cheung, "Video Streaming with Network Coding", undated.

Glossary of Acronyms and Terms Used

DSC: Distributed source coding is an important problem in information theory and communications. DSC problems regard the compression of multiple correlated information sources that do not communicate with each other. By modeling the correlation between multiple sources at the decoder side together with channel codes, DSC is able to shift the computational complexity from encoder side to decoder side, thereby providing appropriate frameworks for applications with a complexity-constrained sender, such video/multimedia compression. One of the main properties of distributed source coding is that the computational burden in encoders is shifted to the joint decoder.

SP-Frame: Super P (Prediction) Frame; the SP-Frame is used in streaming media for switching from one video stream to another based on P-Frames. [4]

SI-Frame: Super I (Intra) Frame; the SI-Frame is used in streaming media for switching from one video stream to another based on P-Frames, whereas an error occurs and a reference frame is needed in the switched stream. [4]

WNLC: Within Layer Network Coding; *see* "Video Streaming with Network Coding", pgs. 13, 18. [8]

HNC: Hierarchical Network Coding; *see* "Video Streaming with Network Coding", pgs. 12, 18. [8]