

AN OPTIMAL YET FAST PRUNING ALGORITHM TO REDUCE LATENCY IN MULTIVIEW PREDICTION STRUCTURES

Pablo Carballeira, Julián Cabrera, Fernando Jaureguizar, and Narciso García

ABSTRACT

We propose a new algorithm for the design of prediction structures with low delay and limited penalty in the rate-distortion performance for multiview video coding schemes. This algorithm constitutes one of the elements of a framework for the analysis and optimization of delay in multiview coding schemes that is based in graph theory. The objective of the algorithm is to find the best combination of prediction dependencies to prune from a multiview prediction structure, given a number of cuts. Taking into account the properties of the graph-based analysis of the encoding delay, the algorithm is able to find the best prediction dependencies to eliminate from an original prediction structure, while limiting the number of cut combinations to evaluate. We show that this algorithm obtains optimum results in the reduction of the encoding latency with a lower computational complexity than exhaustive search alternatives.

Index Terms — 3D Video, video-conference, multiview video coding, prediction structure, low latency, graph theory

1. INTRODUCTION

For several years, video technologies have targeted the development of systems that provide immersive viewing experiences. Today, the advances in three-dimensional (3D) display technologies have made 3D video an emerging and sustainable market in the near future. 3D Video (3DV) and Free Viewpoint Video (FVV) are new types of visual media that expand the user's experience beyond what is offered by 2D video [1], providing a 3D depth impression of the scene, and an interactive viewpoint selection. To support 3DV and FVV, several coding schemes for multiview video and additional data, such as depth, have been proposed. These have resulted in finished or ongoing standardization processes, such as MVC [2], 3D-AVC [3] or extensions of H.264 for 3D Video [4]. All of these coding schemes include multiview video as the fundamental element of the set of data that is coded and transmitted. The core of the coding technology for this multiview video has consistently been the extension of the hybrid-video coding scheme to the spatial dimension, exporting the temporal prediction concept to the inter-view component. Thus, complex prediction structures are constructed, with prediction relationships in two dimensions – temporal and spatial.

Regarding the design of multiview prediction structures, several options have been investigated to obtain efficient prediction structures in terms of rate-distortion (RD) performance. Merkle *et al* [5] proposed multiview prediction structures based on hierarchical B schemes in the temporal and spatial dimensions, that were adopted by the the Joint Video Team (JVT) as the non-normative structure for the MVC coding scheme [6]. These structures set the ground for the prediction structures of subsequent standards.

However, little effort has been made to optimize these multiview prediction structures. The work in [7] proposes to construct the interview prediction structure based upon the notion of minimum spanning tree. In [8], the authors locate the position of the base view using the global disparity information. The work in [9] proposes graph theory to determine the position of P frames and the coding order of B frames in multiview prediction structures using RD metrics.

In our previous work [10], we argued that the design of multiview prediction structures has been mostly focused on improving RD performance, ignoring important differences in the delay behavior of multiview encoders, which may be critical for delay constrained applications. In that work, we presented the problematic of delay analysis in multiview coding schemes and proposed a framework that uses graph theory to perform a systematic analysis and optimization of this delay. Using this framework, we were able to provide a methodology to reduce the encoding latency down to a given constraint with the minimum number of cuts in the prediction dependencies of the original prediction structure. This way, the penalty in RD performance is limited. In [11] we gave proof that, for multiple cuts, the optimum solution is found by an exhaustive search which is highly computationally intensive for usable prediction structures and a non-small number of cuts. Thus, we proposed a sub-optimal solution based on a tree-like search to limit this computational complexity.

In this paper, we propose a new approach to the algorithm of multiple edge pruning that obtains optimum results with important savings in computational complexity with respect to the exhaustive search. This new algorithm makes use of the properties of the delay analysis in the directed acyclic graph encoding latency (DAGEL) model [10] to limit the number of combinations of cuts in prediction links that need to be evaluated, discarding those combinations that are not amenable to be the optimum. We prove that the optimum solution, in terms of encoding latency reduction, is necessarily in the set of link combinations generated by the algorithm. We name this algorithm Fast Optimum Search (FOS). The results will show that the FOS saves between one and six orders of magnitude in computational complexity with respect to an exhaustive search, while obtaining optimum results, for commonly used prediction structures.

This paper is organized as follows: in Section 2 we give a brief reminder of the concepts of the DAGEL model. In Section 3 we present the FOS for multiple edge pruning. In Section 4 we give results of computational complexity savings achieved by the FOS and in Section 5 we present the conclusions.

2. DAGEL MODEL AND EDGE PRUNING

In [10], we proposed a the DAGEL model to perform a systematic analysis of the encoding delay in multiview video encoders. From any arbitrary multiview prediction structure a direct acyclic graph

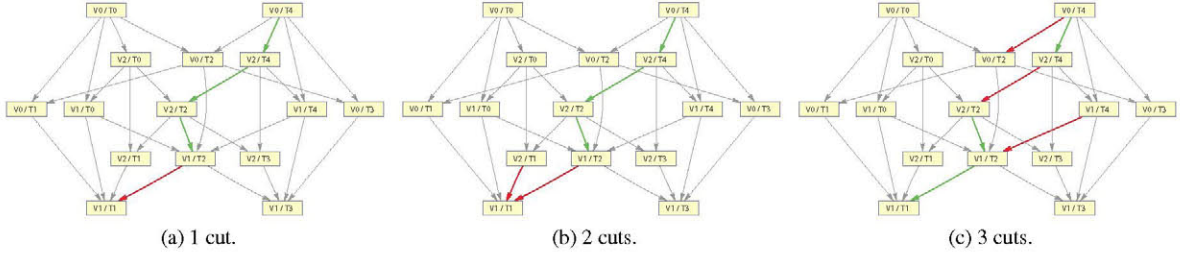


Figure 2: Optimum reduction of the encoding latency for (a) one, (b) two and (c) three cuts of edges in the DAG. The edges in red are the ones selected to cut. The edges that form the critical path of the initial prediction structure are marked in green. The initial prediction structure is the one depicted in Figure 1.

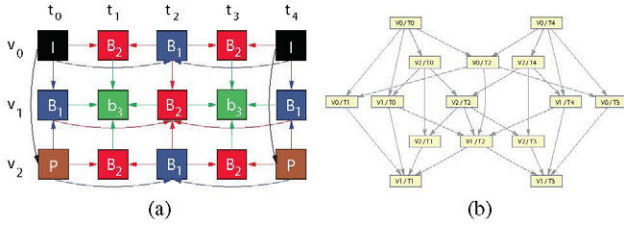


Figure 1: (a) Example of a JMVM prediction structure with three views and a GOP size of four frames and (b) the DAG extracted from it. v_i/t_j represents frame j of view i as signaled on the GOP structure.

(DAG) is extracted. As explained in [10], each node in the DAG represents a frame and each edge a prediction relationship. Figure 1 shows an example of a multiview prediction structure and the associated DAG.

In the DAGEL model, each edge of this DAG has an associated cost value that indicates the delay added by a parent frame to the encoding process of its child frame, i.e. the edge that links the parent frame x_j^i (j -th frame of view i) with frame x_l^k , has a cost $\omega_{j,l}^{i,k}$ that represents the delay added by the processing time of x_j^i to the coding process of x_l^k . Similarly, the paths in the graph have a cost that is computed by adding the costs $\omega_{j,l}^{i,k}$ of the edges in the graph. Then, the encoding latency (Lat_{cod}) of the prediction structure is computed by solving:

$$Lat_{cod} = \max_{x_j^i} \left(p_{del_j^i} + \Delta t_{proc_j^i} \right), \quad (1)$$

where $p_{del_j^i}$ is the path with the highest cost among those ending in node x_j^i , and $\Delta t_{proc_j^i}$ is the processing time of x_j^i . $p_{del_j^i}$ can be obtained by means graph theory algorithms [12]. The path $p_{del_j^i}$ that maximizes (1) is the *critical path* of the DAGEL model.

As proven in [11], if a unique cut is to be performed in the set of prediction links, the best cut in terms of encoding latency reduction is necessarily on its critical path. In the case of multiple edge pruning, a greedy solution that iteratively cuts an edge in the critical paths is generally sub-optimal. To obtain the optimum solution for a given number of cuts, an exhaustive search of all the possible cut combinations is needed. Figure 2 depicts an example of this phenomenon by showing optimum edge selection for an increasing number of cuts in the initial prediction structure shown in Figure 1. Figure 2 depicts the DAGs for the cases of one, two, and three cuts respectively. The results show that the selected edge in the case of one cut is maintained for the case of two cuts, while in the case three cuts three different edges of the DAG are selected.

		3 views GOP 4	3 views GOP 8	3 views GOP 16
n_C	2	435	1.89×10^3	7.88×10^3
	3	4060	3.78×10^4	3.26×10^5
	4	2.74×10^4	5.58×10^5	1.00×10^7
	5	1.43×10^5	6.47×10^6	2.44×10^8
	6	5.94×10^5	6.15×10^7	4.93×10^9
		5 views GOP 4	5 views GOP 8	5 views GOP 16
n_C	2	1.43×10^3	5.99×10^3	2.45×10^4
	3	2.48×10^4	2.16×10^5	1.80×10^6
	4	3.16×10^5	5.77×10^6	9.85×10^7
	5	3.16×10^6	1.22×10^8	4.29×10^9
	6	2.58×10^7	2.14×10^9	1.55×10^{11}

Table 1: Computational complexity of the Exhaustive Search measured by the number of prediction structures evaluated with the DAGEL model.

3. FAST OPTIMUM SEARCH FOR MULTIPLE EDGE PRUNING

The example in Figure 2 provides the intuition of the need of an exhaustive search of all the possible cut combinations. However, for most useful multiview prediction structures, the exhaustive search of all cut combinations is a computationally intensive algorithm: in the case of n_C cuts over the n_L links, it implies the evaluation of $C_{n_L}^{n_C}$ prediction structures, where $C_{n_L}^{n_C}$ is:

$$C_{n_L}^{n_C} = \frac{n_L!}{n_C! (n_L - n_C)!}. \quad (2)$$

Since in multiview prediction structures, n_L is usually high, if the number of cuts is high, the optimum latency reduction algorithm would require a high computational load for all the possible combinations. Table 1 shows the complexity of the exhaustive search for different JMVM [6] prediction structures, and different values of n_C , measured by the number of prediction structures that need to be analyzed using the DAGEL model.

In [10], we proposed a sub-optimal algorithm for multiple edge pruning, the Tree-Search. This algorithm reduced the number of prediction structures to analyze by selecting a series of prediction structure candidates in a tree-decision manner at the cost of obtaining a sub-optimal solution to the problem of multiple edge pruning. To avoid this sub-optimal approach, we propose here a new approach to multiple edge pruning that obtains the same result as the exhaustive search with a reduced computational complexity. This algorithm takes advantage of the concept that for a

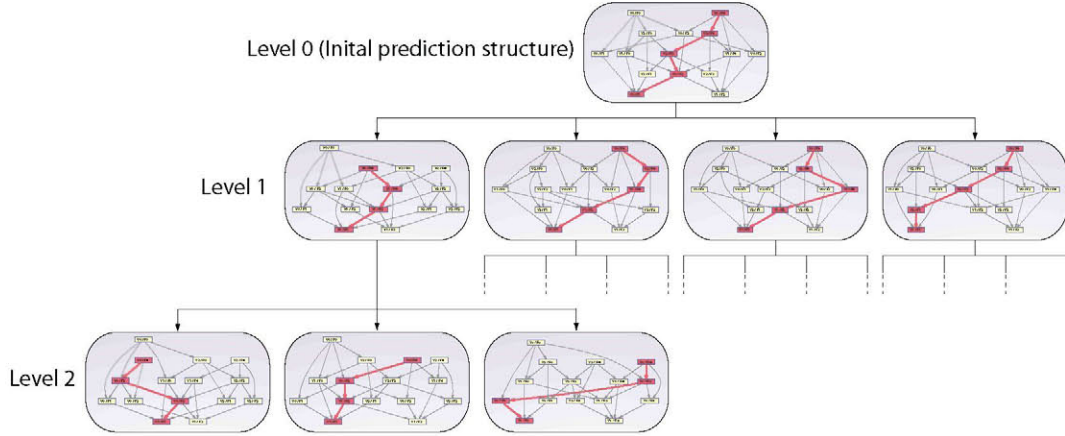


Figure 3: Example of the FOS for $n_C = 2$. Each node of the tree depicts a DAG with its critical path marked in red. For clarity only the branches resulting from one of the nodes in level 1 are depicted in level 2.

given multiview prediction structure the optimum cut is necessarily in the set of links that constitute the critical path. We name this algorithm the Fast Optimum Search (FOS). In the following, we explain the algorithm of the FOS:

Let S_o be the initial multiview prediction structure and n_C a given number of cuts. The following operations are performed:

1. The critical path of S_o is computed by means of the DAGEL model. Let $n_{CP,0}$ be the size of the critical path of S_o .
2. $n_{CP,0}$ prediction structures are obtained as children of structure S_o in the immediate lower level by cutting each of the $n_{CP,0}$ edges that form the critical path of S_o .
3. Steps 1 and 2 are repeated recursively for each children prediction structure until a tree of $n_C + 1$ levels is obtained.
4. Lat_{cod} is computed for each of the prediction structures of the lowest level and the prediction structure with the minimum value is selected.

Figure 3 shows an example of the FOS for the prediction structure in Figure 1 and $n_C = 2$. It depicts the DAG of an initial prediction structure (level 0) and the children DAGs that are obtained from it by cutting each of the links in the critical path (level 1). Level 2 of the tree is obtained analogously but only the prediction structures derived from the first structure in level 1 are shown.

3.1. Optimality of the Fast Optimum Search

The optimality of the FOS algorithm can be proven by structural induction: we know from [11] that if one cut ($n_C = 1$) is performed, it must be one of the edges of the critical path of the original prediction structure (S_o).

In the case of $n_C = 2$, one of the cuts is necessarily in the critical path of S_o (if not, the critical path of S_o exists and Lat_{cod} is that of S_o). Given that level 1 of the tree is generated by all possible cuts in the critical path of S_o , and that for each of these structures the optimum cut is found in one of the edges of their critical paths, the optimum structure is necessarily one of those in level 2 of the tree.

This reasoning can be extended for any arbitrary number of cuts n_C , by proving recursively that the optimum solution exists for all levels from 1 to n_C . Thus, for any number of cuts, the optimum prediction structure in terms of encoding latency reduction is found by using this FOS algorithm.

4. COMPUTATIONAL COMPLEXITY

This section provides a comparison of the computational complexity between the FOS algorithm and the exhaustive search. To illustrate the reduction of computational complexity that is achieved using the FOS algorithm in comparison with the exhaustive search, we have compared both algorithms in terms of the number of prediction structures that have to be analyzed using the DAGEL model.

For a given initial prediction structure, and a given value of n_C , the computational complexity is measured by:

$$\log_{10}(\text{Comp}(X)), \quad (3)$$

where $\text{Comp}(X)$ is the number of prediction structures computed using the DAGEL model for a given algorithm (Exhaustive Search or FOS). With this measure, we obtain a clear insight of the difference in order of magnitude of the number of operations that have to be computed for both algorithms.

For the experiments we have used JMVM prediction structures [6] with IBP prediction in the spatial dimension. To cover a broad range of multiview prediction structures, we have used different combinations of 3 and 5 views and GOP sizes of 4, 8 and 16 frames. The test cases cover different values of the n_C (ranging from 2 to 6).

Figure 4 shows the computational complexity results, measured as previously described, for different original prediction structures, and different values of n_C . The results evidence a relevant reduction of the computational complexity for all evaluated cases for the FOS algorithm, which is even more evident for larger prediction structures and number of cuts. For example, for a prediction structure of 5 views and a GOP size of 16 frames, the number of times that the DAGEL model needs to be computed for $n_C = 6$, in the exhaustive method, is 1.55×10^{11} (see Table 1). Instead, this number of computations can be reduced 6 orders of magnitude using the FOS. This great saving in the computational cost of the search and the optimality of the solution proposed by the FOS make this new algorithm prevail over previous approaches for multiple edge pruning in multiview prediction structures.

5. CONCLUSIONS

We have presented the Fast Optimum Search for multiple pruning of prediction links in multiview prediction structures. By consid-

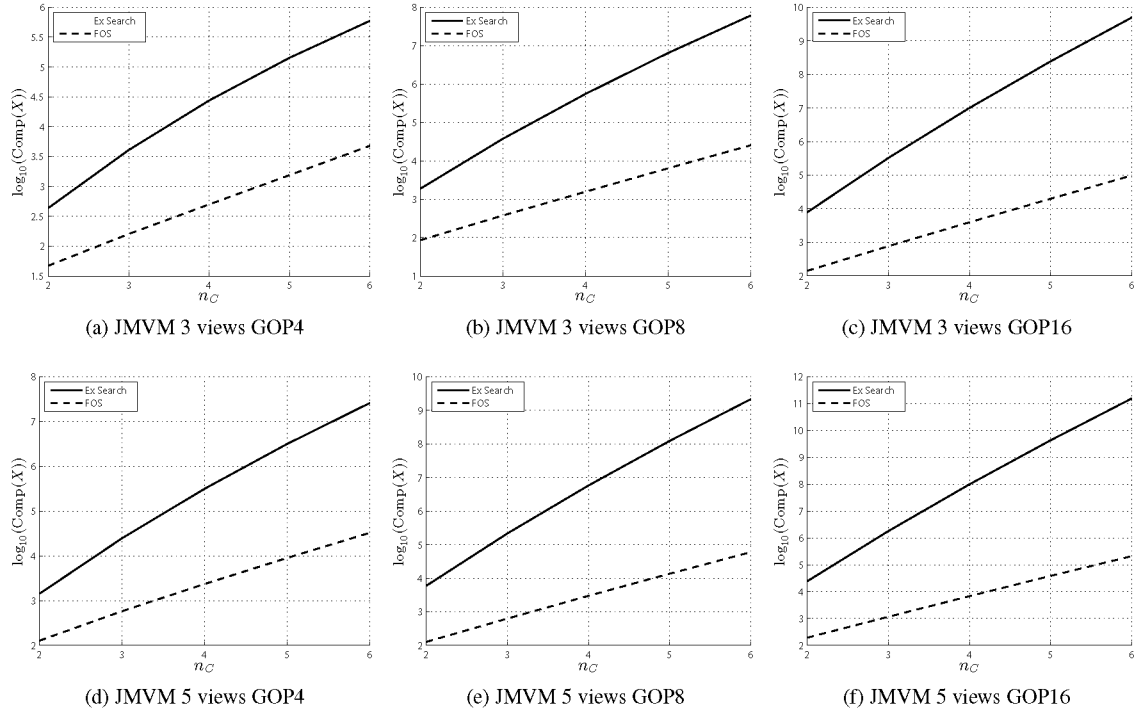


Figure 4: Computational complexity of the exhaustive search and FOS. Results in a logarithmic scale for different initial prediction structures and values n_C .

ering the properties of the analysis of the encoding delay using the DAGEL model, we prove that this algorithm limits the number of evaluated cut combinations while it guarantees an optimum result in terms of encoding latency reduction. The experimental results report computational complexity savings that go from one to six orders of magnitude with respect to the exhaustive search approach. Thus, this algorithm overcomes previous approaches by obtaining optimum results with an important reduction of the computational complexity. Therefore, it constitutes the best solution to date to the problem of design of low latency prediction structures by prediction link pruning.

6. REFERENCES

- [1] A. Smolic, K. Müller, P. Merkle, C. Fehn, P. Kauff, P. Eisert, and T. Wiegand, "3D video and free viewpoint video - technologies, applications and MPEG standards," in *Proc. of IEEE International Conference on Multimedia and Expo (ICME'06)*, Jul. 2006, pp. 2161–2164.
- [2] A. Vetro, P. Pandit, H. Kimata, A. Smolic, and Y. Wang, "Joint Draft 8.0 on Multiview Video Coding," *Doc. JVT-AB204, 28th JVT meeting, Hannover, Germany*, Jul. 2008.
- [3] ISO/IEC JTC1/SC29/WG11, "3D-AVC Test Model 7," *Output doc. N13753, 105th MPEG meeting, Vienna, Austria*, Aug. 2013.
- [4] G. Sullivan, J. Boyce, Y. Chen, J.-R. Ohm, C. Segall, and A. Vetro, "Standardized extensions of high efficiency video coding (hevc)," *IEEE Journal of Selected Topics in Signal Processing*, vol. 7, no. 6, pp. 1001–1016, Dec. 2013.
- [5] P. Merkle, A. Smolic, K. Müller, and T. Wiegand, "Efficient Prediction Structures for Multiview Video Coding," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 17, no. 11, pp. 1461–1473, Nov. 2007.
- [6] A. Vetro, P. Pandit, H. Kimata, A. Smolic, and Y. Wang, "Joint Multiview Video Model (JMVM) 8.0," *Doc. JVT-AA207, 27th JVT meeting, Geneva, Switzerland*, Apr. 2008.
- [7] D.-X. Li, W. Zheng, X.-H. Xie, and M. Zhang, "Optimising inter-view prediction structure for multiview video coding with minimum spanning tree," *IET Electronics Letters*, vol. 43, pp. 1269–1271(2), Nov. 2007.
- [8] P.-K. Park, K.-J. Oh, and Y.-S. Ho, "Efficient view-temporal prediction structures for multi-view video coding," *IET Electronics Letters*, vol. 44, pp. 102–104(2), Jan. 2008.
- [9] J.-W. Kang, Y.-Y. Lee, C.-S. Kim, and S.-U. Lee, "Coding order decision of b frames for rate-distortion performance improvement in single-view video and multiview video coding," *IEEE Transactions on Image Processing*, vol. 19, no. 8, pp. 2029–2041, Aug. 2010.
- [10] P. Carballeira, J. Cabrera, A. Ortega, F. Jaureguizar, and N. Garcia, "A framework for the analysis and optimization of encoding latency for multiview video," *IEEE Journal of Selected Topics in Signal Processing*, vol. 6, no. 5, pp. 583–596, Sep. 2012.
- [11] P. Carballeira, J. Cabrera, A. Ortega, F. Jaureguizar, and N. Garcia, "A graph-based approach for latency modeling and optimization in multiview video encoding," in *Proc. of IEEE 3DTV Conference (3DTV-CON'11)*, May 2011, pp. 1–4.
- [12] K. Thulasiraman and M. N. S. Swamy, *Graphs: Theory and Algorithms*. Wiley, 1992.