Using Machine Learning for Fast Intra MB Coding in H.264

Hari Kalva and Lakis Christodoulou
Dept. of Computer Science and Engineering
Florida Atlantic University
Boca Raton, FL 33431

ABSTRACT

H.264 is a highly efficient and complex video codec. The complexity of the codec makes it difficult to use all its features in resource constrained mobile devices. This paper presents a machine learning approach to reducing the complexity of Intra encoding in H.264. Determining the macro block coding mode requires substantial computational resources in H.264 video encoding. The goal of this work to reduce MB mode computation from a search operation, as is done in the encoders today, to a computation. We have developed a methodology based on machine learning that computes the MB coding mode instead of searching for the best match thus reducing the complexity of Intra 16x16 coding by 17 times and Intra 4x4 MB coding by 12.5 times. The proposed approach uses simple mean value metrics at the block level to characterize the coding complexity of a macro block. A generic J4.8 classifier is used to build the decision trees to quickly determine the mode. We present a methodology for Intra MB coding. The results show that intra MB mode can be determined with over 90% accuracy. The proposed can also be used for determining MB prediction modes with an accuracy varying between 70% and 80%.

Keywords: H.264, fast encoding, machine learning, MB mode

1. INTRODUCTION

Recent developments in video encoding such as the H.264 and VC1 have resulted in highly efficient compression. These new generation codecs are highly efficient and result in equivalent quality video at 1/3 to ½ of MPEG-2 video bitrates. The complexity of this new encoder, however, is 10 times as complex [1]. The compression efficiency has a high computational cost associated with it. The high computational cost is the key reason why these increased compression efficiencies cannot be exploited across all application domains. Resource constrained devices such as cell phones, embedded cameras, and video sensors use simpler encoders or simpler profiles of new codecs to tradeoff compression efficiency and quality for reduced complexity. The new video codecs from Microsoft and Real are also based on hybrid coding techniques similar to H.264 and are comparable in complexity and quality.

The compressions efficiency of these new codecs has increased mainly because of the large number of coding options available. For example, the H.264 video supports Intra prediction with 3 different block sizes and Inter prediction with 8 different block sizes. The encoding of a macro block (MB) involves evaluating all the possible block sizes. As the number of reference frames is increased, the complexity increases proportionally. New approaches are necessary to drastically reduce the encoding complexity without sacrificing quality. We are developing low complexity encoding tools based on machine learning techniques. The goal is to reduce MB mode and motion vector search to a decision tree with a motion search in a small window. The proposed approach is based on the hypothesis that video frames can be characterized for the purpose of encoding and the encoding complexity can be drastically reduced. Our work on MPEG-2 to H.264 transcoding has shown that machine learning tools can be effectively used to classify a MPEG-2 residual MB into one of the possible H.264 coding modes [2,3]. This paper reports our work on Intra MB coding based on this methodology. Inter MB coding is not considered in this paper. The key contribution of this paper is the machine learning based methodology for reducing the complexity of Intra MB coding in H.264 encoding. Machine learning has been used in image and video processing primarily for content analysis and understanding. We have not come across any work on the use of machine learning for reducing the complexity of the encoding process.

2. RELATED RESEARCH

While there is existing work on reducing the complexity of Intra MB coding in H.264, none of these approaches use machine learning and are still relatively complex. Most of these approaches are based on limiting the number of prediction modes evaluated. One approach to reducing the complexity of the Intra 4x4 prediction modes is reported in [4]. In this approach only the most probable prediction modes are evaluated resulting in reduced complexity. The Intra

Visual Communications and Image Processing 2007, edited by Chang Wen Chen, Dan Schonfeld, Jiebo Luo,
Proc. of SPIE-IS&T Electronic Imaging, SPIE Vol. 6508, 65082U, © 2007 SPIE-IS&T · 0277-786X/07/$18

prediction complexity reduction approach proposed in [5] computes an edge histogram to locate a spatial edge and selects the prediction modes based on the edge angle. The method proposed in [6] computes the difference between the current block and the predicted block. The sum of absolute differences (SAD) and the sum of absolute coefficients of the Hadamard transform (SATD) of the residual are computed and early termination of the prediction modes is determined based on the thresholds of SAD and SATD. The complexity reduction approach proposed in [7] based on the correlations that exist on the prediction modes of neighboring blocks. All these approaches reduce the computational cost compared to the H.264 reference software. The key difference with our approach is that in our approach we separate the Intra MB mode and Intra prediction mode decisions. Intra MB mode is determined as Intra 16x16 or Intra 4x4 without computing any prediction modes. The appropriate prediction modes for the MB mode are then determined. Since the MB mode is determined first, our approach right away eliminates the computation of any prediction modes for the MB mode that is not selected. If the MB mode is determined to be Intra 16x16, there is no need to evaluate any prediction modes for the 4x4 sub blocks. Furthermore, since MB mode is determined independently, any other approach can be used to determine prediction modes giving more flexibility in the quality vs. computation tradeoff.

3. MACHINE LEARNING BASED APPROACH TO VIDEO ENCODING

Machine learning has been widely used in image and video processing for applications such as content based image and video retrieval (CBIR), content understanding, and more recently video mining. Video encoding was not considered complex enough to use machine learning approaches. Furthermore, classifying macroblocks (MB) in natural images and video is extremely difficult given the large problem space. The complexity of H.264 video encoding and the expected increase in complexity in next generation video encoding such as H.265 is enough of a motivation to consider new approaches. Our approach is based on using simple mean and variance operations and classifying the MBs based on the relative metrics; for example, how close are the mean values of the neighboring pixel blocks. These seemingly simple metrics give very good performance in determining MB mode and prediction mode of MBs. We developed a hierarchy of decision trees based on the relative mean metrics to compute Intra MB modes quickly.

The Weka data mining tool was used in training and evaluating the decision trees [8]. Selecting the right learning algorithm is a difficult problem. We use the widely studied and used C4.5 algorithm [9]. The C4.5 learning algorithm is considered a generic learning algorithm with broad applicability [10]. The Java implementation of this algorithm in Weka is referred to as J4.8. The Weka tool input is an attribute relation file format (ARFF). The file contains the attributes (e.g., mean of 4x4 sub blocks) that are used to classify a target class (e.g, Intra MB mode). The output of Weka is a decision tree built with the J4.8 algorithm. Figure 1 shows the high level process for the proposed fast Intra coding algorithm. The uncompressed video is encoded with H.264 and at the same time, the means of the 4x4 sub blocks of a MB and the variance of the means of the 16 4x4 sub-blocks are computed. These values together with the MB mode as determined by a H.264 encoder are supplied to Weka/J4.8 to obtain a decision tree. The decision is then used in the encoder instead of the actual mode search code used in JM10.2 H.264 encoder. The decision trees are just *if-else* statements and have negligible computational complexity. Depending on the decision tree, the mean values used are different as discussed in the following sections. The set of decision trees used in the H.264 Intra MB coding are used in a hierarchy to arrive at the Intra MB mode and Intra prediction mode quickly. The trees are trained using 396 MBs from one Intra coded CIF sized (352x288) video frame.

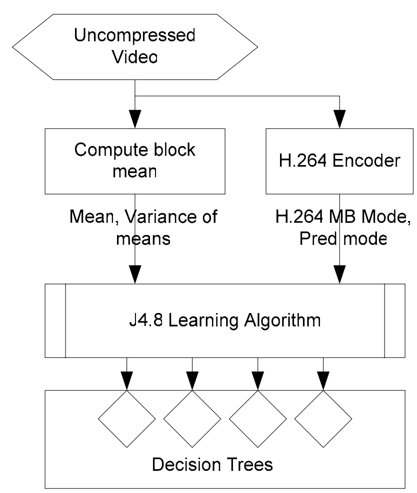

Fig. 1. Building decision trees

Figure 3 shows the hierarchical decision tree used in the proposed Intra MB encoder. The nodes of the tree (circles) are the decision trees and the leaves of the tree (rectangles) are the final decisions. Each node makes a binary decision and additional nodes down in the hierarchy are used to make further classification if necessary. As shown in the figure, the MB modes are classified into Intra16x16 and Intra 4x4 targeting mobile applications. Intra 8x8 mode is not considered. The prediction mode decisions do not support mode 3 in Intra 16x16 and modes 5, 6, 7, and 8 in Intra 4x4. Reducing the prediction modes is necessary to simplify the decision tree. This use of the reduced set of prediction modes is expected to have negligible impact on the PSNR as shown in Figure 3. The hierarchical decision tree uses 7 binary decisions; a maximum of 3 decisions are necessary for Intra 16x16 and 4 are necessary for Intra 4x4.

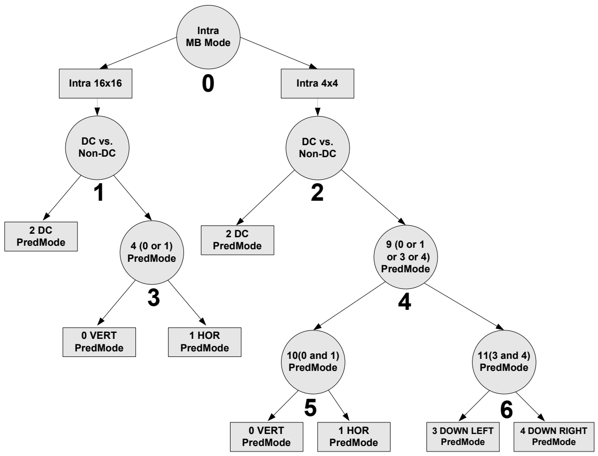

Fig. 2. Hierarchical Decision Tree

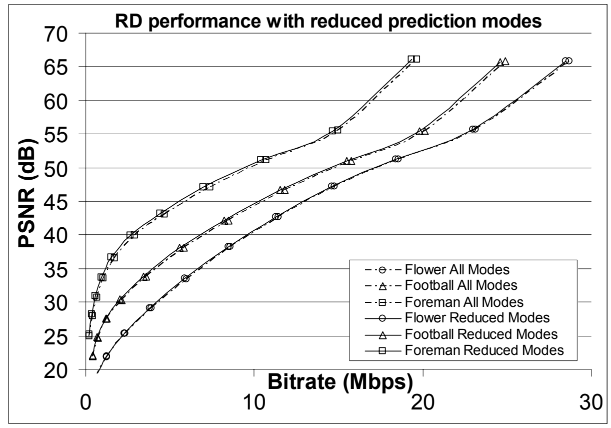

Fig. 3. RD Performance with reduced prediction modes

3.1. Intra MB Mode Decision (Node 0)

An Intra MB is coded as Intra 16x16 or Intra 4x4. Intra 16x16 is used for areas that are relatively uniform and Intra 4x4 is used for areas that are non-uniform and have more detail. The inputs to this classification are the means of the 16 4x4 sub-blocks of a MB and the variance of these means. Intuitively, the variance would be small for Intra 16x16 and large for Intra 4x4 coded MBs. The Intra MB mode is determined without evaluating any prediction modes. This method right away eliminates the evaluation of the prediction modes of the MB mode that is not selected. The sub-block mean computation takes 256 simple operations (240 additions and 16 shifts) and variance computation takes 32 additions and 16 multiplications – a total of 304 operations.

3.2. Intra 16x16 Prediction Mode Decision (Nodes 1, 3)

When the Intra 16x16 MB decision is made, the next step is to determine the prediction modes. Prediction modes 0, 1, and 2 are supported. The Intra 16x16 prediction modes in H.264 depend on the edge pixel values in the neighboring MBs. The prediction direction is determined based on how close the mean of the current MB (μ_C) pixels are to the mean of the bottom row of the above MB (μ_{BR}) and right column of the MB to the left (μ_{RC}). The decision tree is thus made using relative means: $|\mu_C - \mu_{BR}|, |\mu_C - \mu_{RC}|$ and $|\mu_C - (\mu_{BR} + \mu_{BR})/2|$. The decision tree first uses a binary decision to classify DC vs. non-DC modes (node 1) and then uses a separate tree (node 3) for classifying non-DC modes into horizontal and vertical predictions. The computations required are 16 operations to compute the mean of the mean of the current MB using the means of the 4x4 sub-blocks computed in the first step, 33 operations to calculate the relative means – a total of 50 simple operations (add/subtract/shift/absolute).

3.3. Intra 4x4 Prediction Mode Decision (Nodes 2, 4, 5, 6)

For Intra 4x4 MBs, the next step is to determine the prediction direction for the sub-blocks. Prediction modes 0-4 are supported. Similar to Intra 16x16 prediction modes, the Intra 4x4 prediction modes depend on the pixel values on the neighboring 4x4 sub-blocks. The classification is done using: $|\mu_C - \mu_{BR}|, |\mu_C - \mu_{RC}|$, and $|\mu_{BR} - \mu_{RC}|$ where the mean values refer to the 4x4 sub-block, top-row of the sub-block, and the right-column of the sub-block. Node 2 performs a DC vs. non-DC mode classification, node 4 performs diagonal vs. non-diagonal classification, and nodes 5 and 6 further classify modes 0,1 and 3,4 respectively. The computations required per sub-block are 8 simple operations for the mean of neighboring pixels and three absolute value computations – a total of 11 operations. For an Intra 4x4 MB we have 16 sub-blocks that require a total of 176 simple operations.

4. Performance Evaluation

A 4x4 sub-block requires 322 operations to evaluate all the five prediction modes, modes 0-4, which are used in the proposed approach. This is a total of 5152 operations for the 16 sub-blocks of the MB (luma component). For Intra 16x16 prediction modes, evaluating the prediction modes 0, 1, and 2 requires 874 operations per MB. Using the reference implementation such as JM10.2 requires 6026 operations per MB. With the proposed approach, the Intra 16x16 mode requires 304 operations for MB mode computations and 50 operations for prediction mode computations – a total of 354 operations per MB. For Intra 4x4 MB, the proposed method requires 304 operations for MB mode computations and 176 operations for prediction mode computations – a total of 480 operations. With the proposed approach Intra 16x16 MB mode computation is *17 times faster* than the standard and for Intra 4x4 MBs this is *12.5 times faster*. The decision trees are if-else statements that are computationally inexpensive to implement. The effectiveness of the proposed approach is summarized in the table below. Intra MB mode can be determined with 90% accuracy. While Intra prediction mode accuracy is about 75%; this loss in accuracy is not significant in terms of RD performance. The proposed approach shows good promise and can be used to reduce the complexity of encoding in mobile devices.

Table1: Mode Matching Performance of the Seven Decision Trees (See Fig. 2 for the decision tree nodes)

Node:Descr.	Football	Forman	Akiyo	Mobile	Vectra	Mother	Crew	Table	Stefan	Dancer	Container	Average
0:MBMode	97%	93%	79%	99%	90%	90%	82%	88%	96%	89%	84%	89.87%
1:Intra16x16	45.38%	64%	74.67%	75.00%	77.14%	80%	69.67%	45.15%	81.08%	83.46%	72.38%	69.79%
3:Intra16x16	51.85%	60.67%	74.43%	40%	76.22%	82.91%	84.58%	73.81%	84.16%	75.58%	72.02%	70.57%
2:Intra4x4	78.87%	80.65%	81.95%	84.09%	90.19%	79.61%	87.78%	77.96%	85.90%	79.68%	84.81%	82.86%
4:Intra4x4	63.98%	66.37%	78.95%	72.45%	87.99%	66.99%	83.19%	69.93%	78.82%	81.40%	80.19%	75.48%
5:Intra4x4	81.44%	75.75%	84.06%	83.52%	77.05%	79.33%	86.89%	78.01%	76.69%	85.85%	75.30%	80.35%
6:Intra4x4	65.86%	72.54%	65.01%	64.13%	68.08%	61.67%	62.45%	62.81%	67.79%	55.59%	62.45%	64.40%

5. CONCLUSIONS

This paper presents an innovative approach based on machine learning to reduce the complexity of Intra MB coding. The proposed method relies on low complexity features such as mean of dependent pixels to build a decision tree for determining the Intra MB mode and MB prediction mode. The proposed approach uses a generic classification algorithm and reduces the Intra MB mode coding to a classification problem. The paper shows that when the right features are used, machine learning can be effectively used even for problems such as video encoding. The results show that the proposed method reduces the Intra 16x16 MB computations by 17 times and Intra 4x4 MB mode complexity is reduced by 12.5 times. The decision trees are able to classify the Intra MB coding mode with over 90% accuracy and the prediction mode computations by about 75% accuracy. We are currently implementing this decision tree to evaluate the RD performance. Our experience with MB mode computations in video transcoding show that around 75% mode match is sufficiently high to give RD performance close to the reference implementation.

REFERENCES

1. Implementation Studies Group, “Main Results of the AVC Complexity Analysis”. MPEG Document N4964, ISO/IEC JTC11/SC29/WG11, July 2002.
2. G. Fernandez-Escribano, H. Kalva, P. Cuenca, and L. Orozco-Barbosa, “Speeding-up the Macro block Partition Mode Decision for MPEG-2 to H.264 transcoding,” Accepted for Publication, ICIP 2006, Atlanta, September 2006.
3. G. Fernandez-Escribano, H. Kalva, P. Cuenca, and L. Orozco-Barbosa, “Very Low Complexity MPEG-2 to H.264 Transcoding Using Machine Learning,” Accepted for Publication, ACM Multimedia 2007, Santa Barbara, CA, October 2006.
4. B. Meng, O. Au, C.-W. Wong and H.-K. Lam, “Efficient Intra-Prediction Mode Selection for 4x4 Blocks in H.264,” International Conference on Multimedia and Expo, ICME 2003, pp. 521-524.
5. Pan, F.; Rahardja, L.S.; Lim, K.P.; Wu, L.D.; Wu, W.S.; Zhu, C.; Ye, W.; Liang, Z., "Fast intra mode decision algorithm for H.264-AVC video coding," International Conference on Image Processing, ICIP 2004, vol.2, pp. 781-784.
6. Kim, C. Shih, H.-H. Kuo, C.-C.J., "Feature-based intra-prediction mode decision for H.264," International Conference on Image Processing, ICIP 2004, Vol. 2, pp. 769- 772.
7. J. E. Fritts, H. Zhang, and F. W. Steiling, "Fast intra-frame mode selection for H.264," SPIE Electronic Imaging - Image and Video Communications 2005, San Jose, Jan. 2005.
8. Ian H. Witten and Eibe Frank (2005) "Data Mining: Practical machine learning tools and techniques", 2nd Edition, Morgan Kaufmann, San Francisco, 2005.
9. J.R. Quinlan, “C4.5: Programs for Machine Learning”, Morgan Kaufmann, 1993.
10. Baskiotis, N. and Sebag, M. 2004. C4.5 competence map: a phase transition-inspired approach. In Proceedings of the Twenty-First international Conference on Machine Learning (Banff, Alberta, Canada, July 04 - 08, 2004). ICML '04, vol. 69.
11. H. Kalva and B. Petljanski, "Exploiting Directional Features in MPEG-2 for H.264 Transcoding," IEEE Transactions on Consumer Electronics, Vol. 52, No. 2, May 2006, pp. 706-711.