10.4-1

Low Complexity H.264 Video Encoder Design Using Machine Learning Techniques

Paula Carrillo, Tao Pin[†], Hari Kalva.
Department of Computer Science and Engineering, Florida Atlantic University, Boca Raton, FL 33431. [†]Department of Computer Science and Technology, Tsinghua University, Beijing, China

Abstract-- **H.264/AVC encoder complexity is mainly due to variable size in Intra and Inter frames. This makes H.264/AVC very difficult to implement, especially for real time applications and mobile devices. The current technological challenge is to conserve the compression capacity and quality that H.264 offers but reduce the encoding time and, therefore, the processing complexity. This paper applies machine learning technique for video encoding mode decisions and investigates ways to improve the process of generating more general low complexity H.264/AVC video encoders. The proposed H.264 encoding method decreases the complexity in the mode decision inside the Inter frames. Results show, in a 67.81% average reduction of complexity and 0.2 average decreases in PSNR and an average bit rate increase of 0.04% for different kinds of videos and formats.**

I. INTRODUCTION

H.264 video coding standard is the latest block-oriented motion-compensation-based codec standard developed by the ITU-T Video Coding Experts Group (VCEG) together with the ISO/IEC Moving Picture Experts Group (MPEG). H.264 can achieve considerably higher coding efficiency than previous standards. Unfortunately, this comes at a cost in considerably increased complexity at the encoder mainly due to motion estimation and mode decision. The high-computational complexity of H.264 and real-time requirements of video systems represent the main challenge to overcome on the development of efficient encoder solutions.

Different techniques to reduce complexity in H.264 have been proposed in the literature, but few approaches have been done using machine learning. This paper is focused on reducing the complexity of the H.264 Intel IPP encoder using machine learning techniques. The idea behind using machine learning is to exploit structural similarities in video in order to make optimal prediction modes through the use of fast if-else statements instead of the usual cumbersome Sum of Absolute Differences (SAD) and cost evaluations.

The rest of the paper is organized as follows. Section 2 presents a background of the applied technique. Section 3 shows experiments and results obtained. Section 4 draws our Conclusions.

This work was supported in part by the National Science under OISE-0730065. Any opinions, findings and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect those of the National Science Foundation.

II. BACKGROUND AND MOTIVATION

In the H.264 standard, the MB mode decision in Inter frames is the most computationally expensive process due to the use of the variable block-size (16x16, 16x8, 8x16, 8x8, 8x4, 4x8, 4x4) motion estimation. Different approaches for fast inter mode decisions in H.264 encoding have been proposed [1 to 6]. All these approaches try to reduce the computational cost of making MB mode predictions by skipping all/some of the mode calculations and comparisons that the standard does. Previous works made fast MB mode type classification by experimentally setting some thresholds of a metric. The metric used is based on observation of the structural similarities presented in videos. Some of the used metrics are: frame residual information [1], amplitude of edge vectors [2], correlation of collocated MB mode type [3], MB index complexity classification through energy variance [4], between others. Our approach is summarized in Figure 1. The key idea behind this approach is to determine encoder decisions such as MB coding mode decisions that are computationally expensive using easily computable features derived from uncompressed video. A machine learning algorithm is used to deduce the classifier /decision tree based on such features. Once a tree is trained, the encoder coding mode decisions that are normally done using cost-based models that evaluate all possible coding options are replaced with a decision tree. Decision trees are in effect if-else statements in software and require negligible computing resources. As in the previous works our proposed approach takes also MB metrics in order to exploit the video structure, but with the difference that we train classifiers with a set of metrics instead of trying to manually find the correct threshold for each one at the time. This difference permits us to generate more general and accurate MB mode type classifications

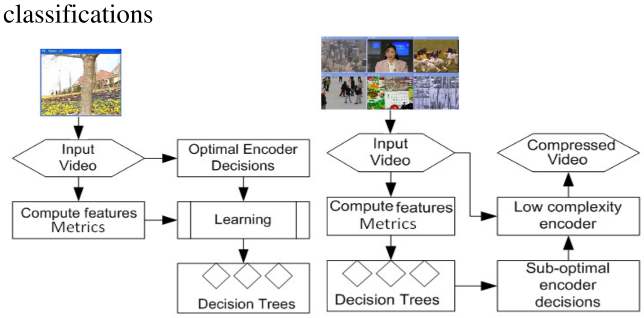

Figure 1. Applying machine learning in video coding

III. EXPERIMENTS AND RESULTS

This paper proposes a three level topology Tree for Inter mode decision. In the first level, an improvement in speed up is achieved with a Skip early decision. Also, after observation of the residual and metric's information, a logic mode sub-classification for the first level was chosen as {Skip, Intra 16x16 and the rest of the modes}. Skip shows reduced residual information, while Intra 16x16 shows uniform residual information and the rest of the modes show a more diverse residual and metrics values. In the second level, there is a consequent division between Inter 8x8 and sub modes against Inter 16x16 and sub modes. The implementation is not going inside the 8x8 modes. On the other hand, if the other leaf is selected, a third classification between Inter 16x16 sub modes {16x16, 16x8, 8x16} and the Intra 4x4 is evaluated. Observations of the metrics detected a high similitude between 16x8, 8x16 and Intra 4x4, and for that reason, these modes were at the same level decision, in an attempt to reduce a negative impact of a bad mode selection.

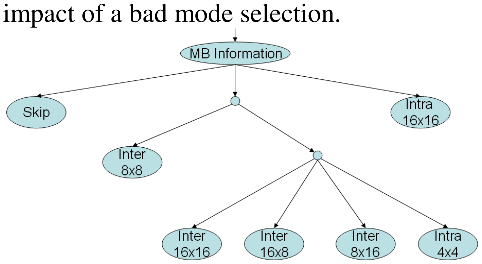

Figure 2. Decision tree used in Intra MB encoding

Figure 2 shows the hierarchical decision tree used in making H.264 inter MB mode and prediction mode decisions. The metrics used in the decision trees are the mean, variance, variance of means, edge detection, residual absolute sum, residual mean, residual variance, residual variance of means and means of variance. These metrics were calculated for the main MB shapes 16x16, 8x8 and 4x4.

We implemented the decision tree inside of in the Intel® H.264 encoder in order to evaluate the performance of the machine learning based decisions. The original Inter MB decisions were replaced by a decision tree, a set of if-else statements. Table 3 shows the test environment used. The results show that MB mode decisions can reduce the complexity of the encoder more than 65% with a slight PSNR penalty as shown in Tables 1 and Table 2.

HD Format (1280x720)	% Average Increase BitRate	Average decrease PSNR	% Average speed up Encoder time	% Average speed up MB Decision
QPs Fixed Experiments HD				
ParkrunHD.yuv	0,60%	0,04	60,82%	67,30%
ShieldsHD.yuv	1,75%	0,04	66,88%	72,35%
StockholmHD.yuv	-0,01%	0,02	68,35%	73,87%
SuperBowlHD.yuv	0,11%	0,03	72,33%	77,91%
Average	**0,61%**	**0,03**	**67,10%**	**72,86%**
BRs Fixed Experiments HD				
ShieldsHD.yuv	0,15%	0,52	65,40%	70,84%
ParkrunHD.yuv	-0,06%	0,20	67,38%	72,75%
StockholmHD.yuv	-0,01%	0,02	68,35%	73,87%
SuperBowlHD.yuv	0,11%	0,03	72,33%	77,91%
Average	**0,05%**	**0,19**	**68,37%**	**73,84%**

Table 1. Average results HD format

TV Format (720x480)	% Average Increase BitRate	Average decrease PSNR	% Average speed up Encoder time	% Average speed up MB Decision
QPs Fixed Experiments TV				
CrewTV.yuv	3,93%	0,02	64,24%	70,15%
HarbourTV.yuv	0,99%	0,07	62,84%	68,57%
IceTV.yuv	8,14%	0,09	71,57%	77,62%
SoccerTV.yuv	5,02%	0,07	63,25%	69,02%
Average	**4,52%**	**0,06**	**65,48%**	**71,34%**
BRs Fixed Experiments TV				
CrewTV.yuv	0,13%	0,07	66,76%	72,73%
HarbourTV.yuv	0,13%	-0,04	66,22%	71,51%
IceTV.yuv	0,01%	0,39	71,84%	77,75%
SoccerTV.yuv	-0,16%	0,43	64,19%	70,11%
Average	**0,03%**	**0,21**	**67,25%**	**73,02%**

Table2. Average results TV format

Processor: Intel® Core™ 2 CPU 6600 @ 2.4 GHz	
Number of frames: 240, IPPPP, GOP (Group of Pictures) =100	
Search algorithm: FS, range of search = 15.	
Entropy coding: CAVLC	**Frame Rate:** 30 Hz

Table 3 Experimental environment

IV. Conclusions

The results of using data mining and machine learning techniques prove the efficiency of these concepts in video compression. Even when using a highly optimized H.64 encoder from Intel, the proposed approach reduced the encoding time by over 65%. Another key contribution of this paper is the evaluation of the algorithms in a product ready and optimized implementation. Most of the published results are based on the reference JM software and do not reflect the true complexity of encoding. It is also important to clarify that the implementation tested over different video formats and context showed a reliable quality performance throughout the entire bit rate range.

Reference

[1] Xuan Jing and Lap-Pui Chua, "An Efficient Inter Mode Decision Approach for H.264 Video Coding" International Conference on Multimedia and Expo (ICME) 2004.
[2] D.Wu, S.Wu, K.P Lim, F. Pan, Z.G. Li, C. C. Ko "Fast Inter Mode Decision for H.264 Encoding". Consumer Electronics, 2004 IEEE,
[3] Cheng-Chang Lien, Chung-Ping Yu, "A Fast Mode Decision Method for H.264/AVC Using the Spatial-Temporal Prediction Scheme", ICPR 2006.
[4] Andy C. Yu, "Efficient block-size selection algorithm for inter-frame coding in H.264/Mpeg-4 AVC", IEEE International Conference on Acoustics, Speech and Signal Processing. 2004.
[5] S. Saponara, M. Casula, F. Rovati, D. Alfonso, L. Fanucci, "Dynamic Control of Motion Estimation Search Parameters for Low Complex H.264 Video Coding", IEEE Transactions on Consumer Electronics.
[6] S. Saponara, M. Melani, L. Fanucci, P. Terreni, "Adaptive algorithm for fast motion estimation in H.264/MPEG-4 AVC", Proc. Eusipco2004.
[7] Fernández, Kalva, Cuenca, Orozco, "A first approach to speeding-up the inter mode selection in MPEG-2/H.264 transcoders using machine learning", Multimed Tools Appl (2007) 35:225–240.
[8] Intel Integrated Performance Primitives Reference Manual: Volume 2