10.4-5

Speedup Macroblock Mode Decision in H.264/SVC Encoding Using Cost-Sensitive Learning

Urvang Joshi, Rashad Jillani*, Chiranjib Bhattacharya, Hari Kalva*, and K. R. Ramakrishnan
Indian Institute of Science, Bangalore, India
*Florida Atlantic University, Boca Raton, FL 33431, USA

Abstract-- **In this paper we present a novel macroblock mode decision algorithm to speedup H.264/SVC Intra frame encoding. We replace the complex mode-decision calculations by a classifier which has been trained specifically to minimize the reduction in RD performance. This results in a significant speedup in encoding. The results show that machine learning has a great potential and can reduce the complexity substantially with negligible impact on quality. The results show that the proposed method reduces encoding time to about 70% in base layer and up to 50% in enhancement layer of the reference implementation with a negligible loss in quality.**

I. INTRODUCTION

H.264/SVC achieves efficient encoding, but at the same time the processing involved is quite complex and hence the codec is often unusable on recourse-constrained devices such as mobiles. The most time-consuming process in H.264/SVC encoding is MB mode decision. For Intra-frame prediction, H.264 has 4 prediction modes for 16x16 block-size and 9 modes each for 8x8 and 4x4 blocks-sizes [1]. This mode-selection is typically performed by trying all the modes and then choosing the one achieving best RD performance. This paper describes a novel method for mode-decision which uses a classifier to predict the optimal mode. In particular, the classifier is trained to minimize the loss in RD performance. This is a better approach than training the classifier to optimize zero-one error, because for our application, various wrong predictions may incur various penalties in terms of RD performance.

In this paper, our approach reduces the computationally expensive elements of encoding such as coding-mode evaluation to a classification problem with negligible complexity. The key contribution of this work is the exploration of Support Vector method in video encoding applications.

II. BACKGROUND AND MOTIVATION

SVC is based on a multilayer representation of the video with an AVC compliant base layer (BL). Enhancement layers (EL) can be added to increase the frame rate (temporal scalability), the spatial resolution (spatial scalability) and the quality (fidelity scalability) of the content. BL is encoded by using the standard AVC encoding tools while for ELs, additional inter-layer prediction (ILP) tools have been introduced which use BL for prediction in addition to encoding tools available within the same layer. We obtained classification data by using reference software JSVM with the configuration of two spatial layers of the format QCIF and CIF. Based on the cost-sensitive learning, we developed a classification algorithm and implemented in JSVM by replacing the traditional Intra prediction tools in AVC and ILP tools in SVC.

The proposed approach was developed based on the insights from our work on MPEG-2 to H.264 transcoding [2] and low complexity Intra MB encoding in H.264 [3], both of which exploit machine learning tools. The key idea behind this approach is to exploit the correlation between the structural information in a video frame and the corresponding mode decisions.

III. COST-SENSITIVE LEARNING IN H.264/SVC

We solve the mode-prediction problem using a *two-level approach*. At *first* level, we use a decision tree classifier (as used in [2], [3], [4] for H.264/AVC) to predict the block size to be 16x16 or 4x4. (We have not used 8x8 block size in our experiments). At *second* level, we use one classifier each for 16x16 and 4x4 block sizes, each of them having been trained using cost-sensitive learning described below.

[5] presents a general framework based on Support Vector method for building classifiers so as to *optimize* various *performance measures* as opposed to the *zero-one error*. In this approach, the classifier is trained by solving the following optimization problem:

$$\min_{w,\xi>0} \tfrac{1}{2} \| w \|^2 + C\xi \quad \text{s.t.}$$

$$\forall \bar{y}' \in \bar{Y} \setminus \bar{y} : w^T [\Psi(\bar{x}, \bar{y}) - \Psi(\bar{x}, \bar{y}')] \geq \Delta(\bar{y}', \bar{y}) - \xi$$

Where $\Psi(x, y)$ is a feature map describing the match between feature vector x & label y and Δ is the loss function. (See [5] for other notations). Note that optimal ξ upper bounds training loss $\Delta(y_{pred}, \bar{y})$ where $y_{pred} = \arg\max_y w^T \Psi(x, y)$ (See Theorem 1 in [5]).

The choices of Ψ and Δ are *application specific*. For our application, the classifiers for 16x16 & 4x4 block sizes will have 4 & 9 target classes respectively. Hence, we have chosen *feature map*

$$\Psi(x, y) = x \otimes \Lambda^c(y),$$

Which is similar to the one used for SVM-multiclass in [6]. In other words, vector $\Psi(x, y)$ will be an nk-dimensional vector (assuming x is n -dim feature vector and k is the number of modes) which is obtained by shifting feature vector x according to label y . In particular,

$$\Psi(x,y)[n*(y-1)+i] = x[i]; i = 0,...,n-1$$
$$= 0; otherwise$$

Due to this choice of Ψ , learned vector w will be a collection of vectors $v_1,...,v_k$ where v_i is the weight vector for the i^{th} mode.

Next, to minimize reduction in RD performance, we have chosen the *loss function*

$$\Delta(\overline{y},y) = Cost(y) - Cost(\overline{y})$$

Where $Cost(y)$ = RD cost incurred when mode y is chosen. Thus $\Delta(\overline{y},y)$ in essence denotes the *extra* RD cost incurred when we chose a non-optimal mode y instead of optimal one $\overline{y}$.

With this choice of Ψ and Δ , we *train* one classifier each for block sizes 16x16 and 4x4 using the framework above by tuning parameter C to learn vector w .

During encoding, the *mode-prediction* for a given MB with feature vector x is performed as follows: for each mode $\overline{y}$, we calculate its score using

$$score(\overline{y}) = w^T\Psi(x,\overline{y})$$

Then the mode with the lowest score is predicted by the classifier.

IV.EXPERIMENTS AND RESULTS

We implemented the classification algorithm in order to evaluate the performance of our algorithm. The classification algorithm was implemented in SVC reference software JSVM 9.17. The Intra MB decisions in base layer and enhancement layers were replaced by our algorithm. Test sequences are encoded with all Intra frames and performance evaluated by comparing the RD performance and speedup of the modified encoder with that of standard SVC encoder. The results show that the MB mode decisions can be made with very high accuracy; as shown in Fig 1. Prediction mode decisions are complex and introduce a small loss in PSNR. The maximum PSNR loss suffered in this case is about 0.5 dB.

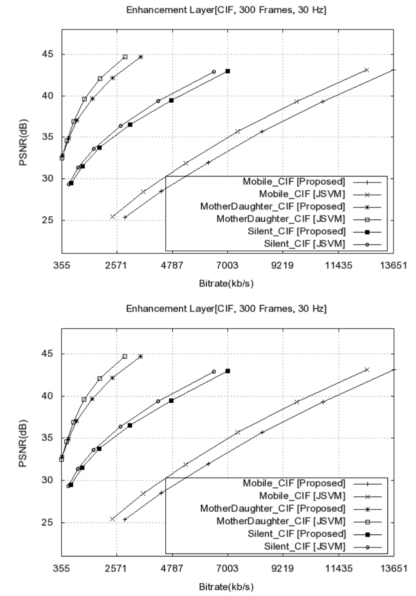
Table I shows the classification performance for various QCIF and CIF sequences. For base layer, we obtain a speedup of about 50% while for enhancement layer; a speedup of about 70% is obtained. Fig. 1 shows the RD performance for base and enhancement layers.

V.CONCLUSIONS

This paper presents a novel approach to H.264/SVC Intra MB mode computation based on machine learning. The proposed approach has great potential to reduce the computational complexity. The results of the implementation in JSVM 9.17 show that the RD performance is very close to the reference encoder. We believe the proposed approach is also applicable to Inter mode prediction and is expected to substantially reduce the encoding complexity.

TABLE I
CLASSIFICATION RESULTS BY COST-SENSITIVE LEARNING

	QCIF			CIF		
		ΔBR	ΔT		ΔBR	ΔT
Sequence	ΔPSNR	%	%	ΔPSNR	%	%
Akiyo	-0.292	2.38	-50.5	-0.31	3.08	-70.91
Flower	-0.465	1.86	-57.16	-0.474	3.01	-77.4
Foreman	-0.223	1.85	-52.25	-0.26	2.28	-67.27
Mobile	-0.404	2.75	-57.42	-0.377	4.76	-68.05
Mot Dau	-0.286	3.81	-51.26	-0.265	4.71	-79.02
Silent	-0.303	1.19	-52.76	-0.314	2.88	-73.42
Average	-0.33	2.31	-53.56	-0.33	3.45	-72.68



(a) EL RD Performance (b) BL RD Performance

Fig. 1. RD performance for BL and EL using cost-sensitive learning.

REFERENCES

[1] Schwarz, H. and Marpe, D. and Wiegand, T., "Overview of the Scalable Video Coding Extension of the H.264/AVC Standard," *Circuits and Systems for Video Technology, IEEE Transactions on*, vol. 17, 2007.

[2] G. Fernandez-Escribano, H. Kalva, P. Cuenca, and L. Orozco-Barbosa, “Very Low Complexity MPEG-2 to H.264 Transcoding Using Machine Learning”, Procs of the ACM Multimedia 2007, Santa Barbara, CA.

[3] H. Kalva, and L. Christodoulou, “Using Machine Learning for Fast Intra MB Coding in H.264”, Procs of VCIP 2007, January 2007.

[4] R. Jillani, and H. Kalva., “Low Complexity Intra MB Encoding in H.264/AVC”, Proceedings of the IEEE ICCE, Las Vegas, USA, 2008.

[5] T. Joachims, "A support vector method for multivariate performance measures", Procs of the 22nd ICML, New York, USA..

[6] I. Tsochantaridis, T. Hofmann, T. Joachims, and, Y. Altun, "Support vector machine learning for interdependent and structured output spaces" in ICML ’04, New York, NY, USA 2004.