Video Coding Mode Decision As a Classification Problem

Rashad Jilani[a], Urvang Joshi[b], Chiranjib Bhattacharyya[b], Hari Kalva[a], K R Ramakrishnan[b]

[a]Computer Science and Engineering, Florida Atlantic University, Boca Raton, FL, USA 34341
[b]Computer Science and Automation, Indian Institute of Science, Bangalore, India 560012

ABSTRACT

In this paper, we show that it is possible to reduce the complexity of Intra MB coding in H.264/AVC based on a novel chance constrained classifier. Using the pairs of simple mean-variances values, our technique is able to reduce the complexity of Intra MB coding process with a negligible loss in PSNR. We present an alternate approach to address the classification problem which is equivalent to machine learning. Implementation results show that the proposed method reduces encoding time to about 20% of the reference implementation with average loss of 0.05 dB in PSNR.

Keywords: H .264/AVC, intra MB, complexity, chance constrained, mode decision, classifier, machine learning

1. INTRODUCTION

Resource constrained devices typically manage the complexity by using a subset of possible coding modes thereby sacrificing video quality. This quality and complexity relationship is evident in most video codecs used today. Most H.264/AVC encoder implementations on mobile devices currently do not implement the standard profiles fully due to high complexity. For example Intra 4x4 prediction is usually not implemented due to complexity. Most of the traditional complexity reduction approaches in video coding are based on eliminating a subset of allowed coding modes and sacrifice quality for reduced complexity. The traditional approaches have not been able to reduce the encoding complexity enough to enable the use of advanced video coding features on resource constrained devices. We developed a machine learning based approach to reduce the computationally expensive elements of encoding such as coding-mode evaluation to a classification problem with much less complexity.[1] We successfully adopted this technique for transcoding and produced much better results.[2] We continued to improve our methodology by employing various techniques to improve our results in the classification domain. The key contribution of this work is the exploration of machine learning and mathematical formulation of video coding mode decision as a classification problem.

2. BACKGROUND

While there are several techniques investigated for fast intra mode decision in H.264/AVC encoding[3],[4] there is no known published results on the use of coding mode classification in video encoding. Traditional approaches reduce the computational cost compared to H.264/AVC reference implementation. The complexity reduction, however, is not sufficient to enable the use of complex features on resource constrained devices. The proposed approach was developed based on the insights from our work on MPEG-2 to H.264 transcoding that exploited machine learning tools[2].[5] The key idea behind this approach is to exploit the correlation between the structural information in a video frame and the corresponding H.264/AVC MB mode decisions. Our proposed approach is summarized in Fig. 1. The key idea behind this approach is to determine encoder decisions such as MB coding mode decisions that are computationally expensive using the easily computable features derived from uncompressed video. Chance constrained technique is used to deduce the classifier based on such features. Once the classifier is obtained, the encoder coding mode decisions that are normally done using cost-based models that evaluate all possible coding options are replaced with a binary decision tree. We believe this simple approach has the potential to significantly reduce encoding complexity and affect the way encoders are used in mobile devices. The results of this technique may not be necessarily interpreted as an improvement to the results of our previous work based on machine learning but we have successfully explored alternate approach by keeping the problem in the classification domain.

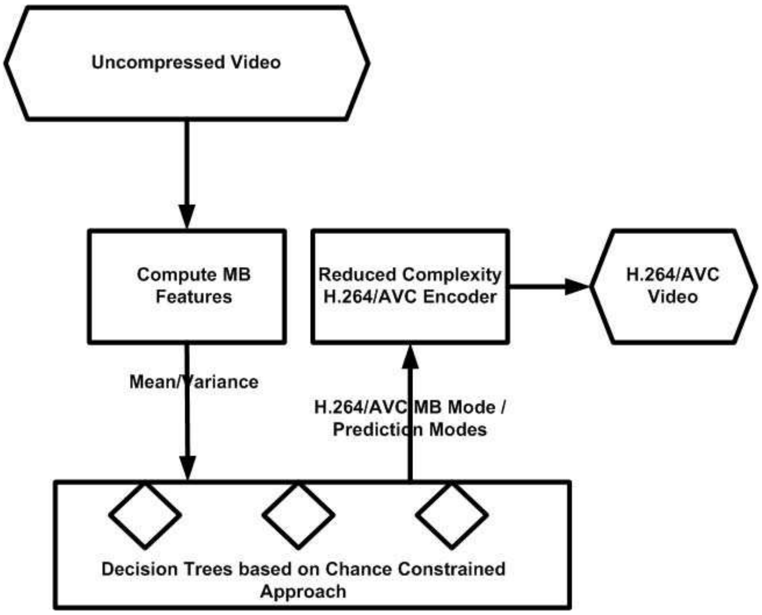


Figure 1. Applying chance constrained classifier in video coding

3. CHANCE CONSTRAINED APPROACH

The features that are of significant value when taking Intra-coding mode decision in H.264 are those which capture the similarity between current Macroblock (MB) and neighboring blocks as well as the ones which capture the self-similarity of the MB. The importance of self-similarity comes from the fact that coding mode depends upon whether the MB is homogeneous or contains highly detailed information. For Instance, to minimize Rate-Distortion, optimum coding mode size of a *homogeneous* MB would be *large* and vice versa. Metrics such as mean and variance of intensity values of an MB provide a good measure of self-similarity. We can also partition each 16x16 MB into 16 sub-macroblocks (SMBs) of size 4x4 and then we can take the mean and variance of intensity values within each SMB; which will signify how different are the intensity values and homogeneity in various parts of the MB. The similarity with neighboring MBs can be captured using the difference in intensity values between current and neighboring MBs. These differences can again be in terms of mean and variance of pixel wise differences of intensities.

3.1 Mathematical formulation

It is clear that pairs of means and variance values in our problem play a significant role in MB coding-mode decision. From this viewpoint, we can consider an approach similar to the chance constrained approach introduced in.[6] This approach considers each pair of mean-variance values being the mean & variance of a random variable. We use this approach to derive an equivalent formulation for our problem. Let the data of each class be specified by the first two moments, i.e., mean μ and covariance Σ. Let X_1 and X_2 represent the random vectors that generate the data points of the positive and negative classes respectively. Assume that distributions of X_1 and X_2 can be modeled using mixture models, with component distributions having *diagonal covariance matrices*; i.e. the component distributions are *decorrelated.* Let k_1 and k_2 be the number of components in the mixture model of positive class and negative class respectively and let $k = k_1 + k_2$.

$$X_1 \sim f_{X_1} = \sum_{j=1}^{k_1} \rho^j f_{X^j} \qquad X_2 \sim f_{X_2} = \sum_{j=k_1+1}^{k} \rho^j f_{X^j}$$

$$\begin{aligned} \text{s.t.} \sum_{j=1}^{k_1} \rho^j = & \quad 1 \quad = \sum_{j=k_1+1}^{k} \rho^j \\ \text{and} \qquad \Sigma^j \quad = & \quad diag(\sigma_{j_1}^2, \sigma_{j_2}^2, \ldots, \sigma_{j_n}^2) \qquad \forall j \end{aligned}$$

where $\sigma_{j_i}^2$ is the variance of i^{th} component of X^j.

We want to learn a classifier $< w, b >$ which separates the two classes with high probability, i.e.

$$
\begin{aligned}
P(w^T X^j + b \geq 1) &\geq \eta, \qquad j = 1 \rightarrow k_1 \\
P(w^T X^j + b \leq -1) &\geq \eta, \qquad j = k_1 + 1 \rightarrow k \\
X^j \sim f_{X^j}, & \qquad j = 1 \rightarrow k
\end{aligned} \tag{1}
$$

where η is a user-defined parameter, which lower-bounds the classification accuracy.

To handle the case of outliers and almost linearly separable cases, we can introduce slack variables ξ^j and obtain the following formulation with relaxed constraints:

$$
\begin{aligned}
\min_{w,b,\xi} \sum_{j=1}^{k} & \quad \xi^j \qquad \text{s.t.} \\
P(w^T X^j + b \geq 1 - \xi^j) &\geq \eta, \qquad j = 1 \rightarrow k_1 \\
P(w^T X^j + b \leq -1 + \xi^j) &\geq \eta, \qquad j = k_1 + 1 \rightarrow k \\
\xi^j &\geq 0, \qquad j = 1 \rightarrow k \\
||w||_2 &\leq W \\
X^j &\sim f_{X^j}, \quad j = 1 \rightarrow k
\end{aligned} \tag{2}
$$

Probabilistic constraints in (2) can be simplified using Chebyshev-Cantelli Inequality as in[6] to get the following equivalent formulation (3):

$$
\begin{aligned}
\min_{w,b,\xi} \sum_{j=1}^{k} & \quad \xi^j \qquad \text{s.t.} \\
y_j(w^T \mu^j - b) &\geq 1 - \xi^j + \kappa ||\Sigma^{1/2} w||, \quad j = 1 \rightarrow k \\
\xi^j &\geq 0, \qquad j = 1 \rightarrow k \\
W &\geq ||w||_2
\end{aligned} \tag{3}
$$

where $\kappa = \sqrt{\frac{\eta}{1-\eta}}$

Formulation in (3) can be further simplified as follows: Let Z^j be a random vector such that $Z^j = \Sigma^{-1/2} X^j$. Hence $\mathbb{E}(Z^j) = \Sigma^{-1/2} \mu^j$ and $cov(Z^j) = I$

Now using Z^j in place of X^j in (3), we can obtain following equivalent formulation:

$$
\begin{aligned}
\min_{w,b,\xi} \sum_{j=1}^{k} & \quad \xi^j \qquad \text{s.t.} \\
y_j(w^T \mu_Z^j - b) &\geq 1 - \xi^j + \kappa ||w||, \quad j = 1 \rightarrow k \\
\xi^j &\geq 0, \qquad j = 1 \rightarrow k \\
W &\geq ||w||_2
\end{aligned} \tag{4}
$$

where $\mu_Z^j = \mathbb{E}(Z^j) = \Sigma^{-1/2} \mu^j$

This classification problem is an instance of a Second Order Cone Program (SOCP), which can be efficiently solved by solvers like *sedumi*.

3.2 Geometrical interpretation

Geometric interpretation of (4) turns out to be classifying most of the *spheres* - with center being the mean value and radius κ - correctly, as opposed to classifying *points* in the usual case. This can be seen as follows: Let the set of points lying in sphere B with center c and radius r be denoted by $B(c,r) = \{x|(x-c)^T(x-c) \leq r^2\}$. Now, the problem of classifying points in $B(\mu,\kappa)$ correctly (including the case of outliers) is:

$$w^T x - b \geq 1 - \xi, \quad \forall x \in B(\mu,\kappa) \tag{5}$$

The constraints in (5), which imply that the whole sphere should be on the positive halfspace of the hyperplane $w^T x - b = 1 - \xi$, can be replaced by a single constraint:

$$w^T x_0 - b \geq 1 - \xi, \quad \text{where } x_0 = \arg\min_{x \in B(\mu,\kappa)} (w^T x - b) \tag{6}$$

which specifies that the point nearest to the hyperplane $w^T x - b = 1 - \xi$ should be on the positive halfspace.

Now, x_0 can be found as follows: Drop a perpendicular to the hyperplane from centre of the sphere. The point at which the perpendicular intersects with the sphere is x_0. Using this geometry and noting that the sphere has center μ and radius κ, we get $x_0 = \mu - \kappa \frac{w}{||w||}$

Now, $x_0 \in B(\mu,\kappa)$. Hence, from (5),

$$w^T x_0 - b \geq 1 - \xi \tag{7}$$

Putting value of x_0 in (7), we get

$$w^T \mu - b \quad \geq \quad 1 - \xi + \kappa ||w||, \quad j = 1 \rightarrow k \tag{8}$$

which is similar in form to the Second Order Cone (SOC) constraint in (3) considering that we have considered positive half-space here ($y = 1$).

This is geometrically illustrated in Fig. 2. All blue spheres have positive, while red sphere have negative labels. Note that except the red sphere intersecting the hyperplane, all spheres are classified correctly.

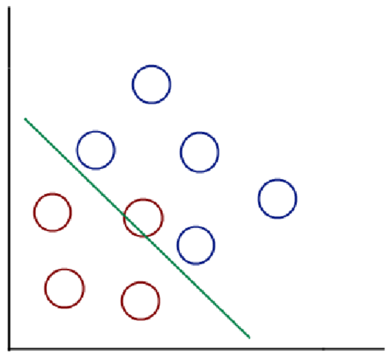


Figure 2. Geometrical Interpretation of SOC Constraint

3.3 Implementation

Our mode-decision problem is a multiclass problem consisting of 13 possible class labels: 4 prediction modes for block size of 16x16 and 9 prediction modes for block size of 4x4. As in the approach used in,[2] we solve this problem using a set of 12 binary sub-classifiers arranged as a hierarchical tree.

For each sub-classifier we use a set of features in the form of mean-variance pairs of various intensity values relevant to that particular sub-classification. Each classifier is then trained using the formulation in (4) using *yalmip*, which is an interface to SOCP solver *sedumi*, to obtain model parameters w and b.

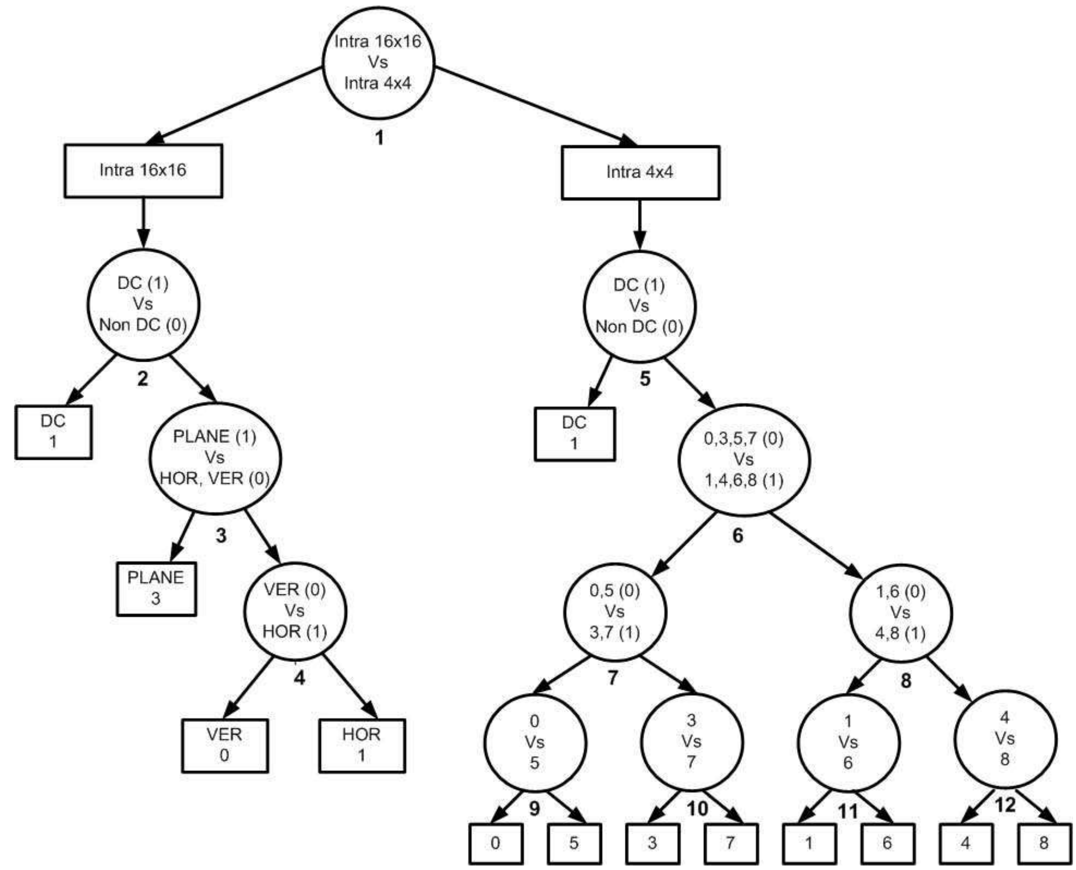


Figure 3. Decision tree used in Intra MB coding

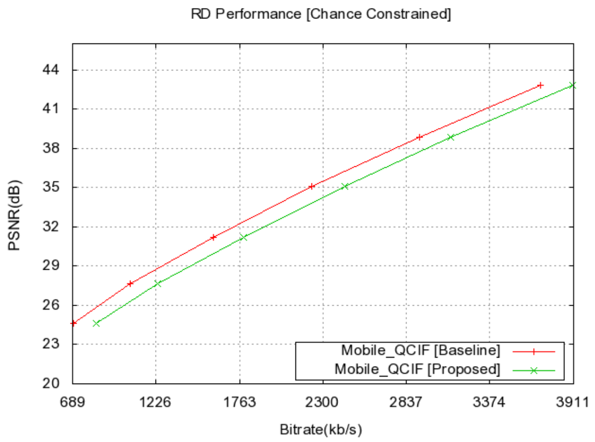
For testing, firstly the required features of the incoming macroblock are computed. Then, the mean values are scaled using the equation $\mu^j_{scaled} = \Sigma^{-1/2}\mu^j$. Note that this computation is done only once per macroblock and then reused at all levels of the tree. Note also that, as Σ is a diagonal matrix, computation of $\Sigma^{-1/2}$ as well as matrix multiplication are very low cost operations. After this initial computation, each sub-classifier takes a decision based on $sign(w^T\mu'^j_{scaled} - b)$ where μ'^j_{scaled} consists of only those mean values which are relevant to sub-classification at this level of tree.

Note that one major task in our problem is the speed-up achieved in encoding as compared to standard reference encoder (JM 14.2). This is clearly achieved here, as all operations required during testing are low-cost operations as can be seen in the last paragraph. The experimental results support this claim as well.

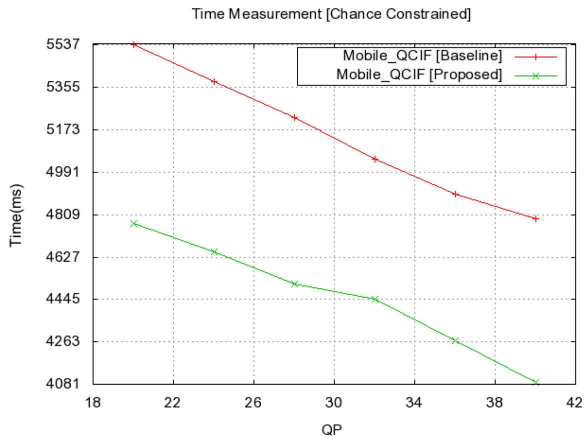
Another major task is the performance in terms of Bitrate penalty (/Rate Distortion / PSNR). This highly depends upon the performance of the classifier. Experimental results suggest that bitrate performance of the approach is good too, that is, the bitrate penalty of the video encoded using this technique is very low.

4. EXPERIMENTS, RESULTS, AND DISCUSSION

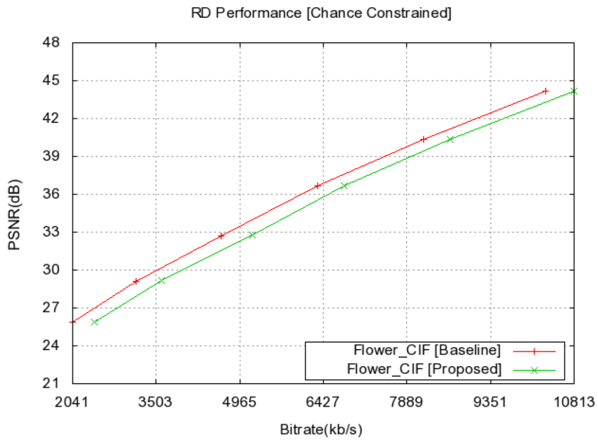
Intra MBs in H.264 are coded as Intra 16x16, Intra 4x4, or Intra 8x8. The baseline profile used in mobile devices does not support Intra 8x8 mode and this mode will not be discussed further in this paper. Intra modes also have associated prediction modes; Intra 16x16 has 4 prediction modes and Intra 4x4 has 9 prediction modes. Baseline profile encoders typically evaluate both Intra 16x16 and Intra 4x4 modes and the associated prediction modes before making MB mode decisions. In the proposed chance constrained based approach we separate the Intra MB mode and Intra prediction mode decisions. Intra MB mode is determined as Intra 16x16 or Intra 4x4 without evaluating 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



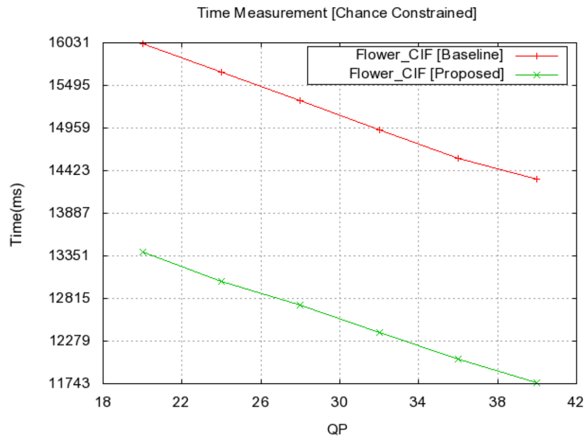
(a)



(b)



(c)



(d)

Figure 4. (a)(c) RD performance of the proposed H.264 Intra frame encoder with Chance Constrained classifier; (b)(d)Time complexity reduction.

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. Fig. 3 shows the hierarchical decision tree used in making H.264 intra MB mode and prediction mode decisions. The attributes used in the decision trees are the mean and variance of the means of the 4x4 sub blocks for MB mode decisions. For prediction mode decisions, the mean and variance of the pixels used for prediction are the attributes. Analytical results based on this approach show that the number of operations required for Intra mode determination is reduced by about 15 times.[7]

We implemented the decision tree in order to evaluate the performance of the chance constrained based decisions. The decision trees were implemented in H.264 reference software JM 14.2. The Intra MB decisions in JM 14.2 were replaced by a decision tree, a set of if-else statements. Test sequences are encoded with all Intra frames and performance evaluated by comparing the RD performance of an encoder with the proposed methods for Intra mode decisions and a standard encoder. The results show that MB mode decisions can be made with very high accuracy; as shown in Fig. 4.a-c, with MB mode decisions made by the proposed approach and the prediction modes determined by the reference software there is very minor difference compared with the reference encoder. Prediction mode decisions are complex and introduce a small loss in PSNR as shown in Fig. 4.b-d. The maximum PSNR loss suffered has been less than 1 dB. The encoding time of the proposed method is about 1/5 of the reference encoding time. Fig. 5 shows the RD performance and time complexity measurement

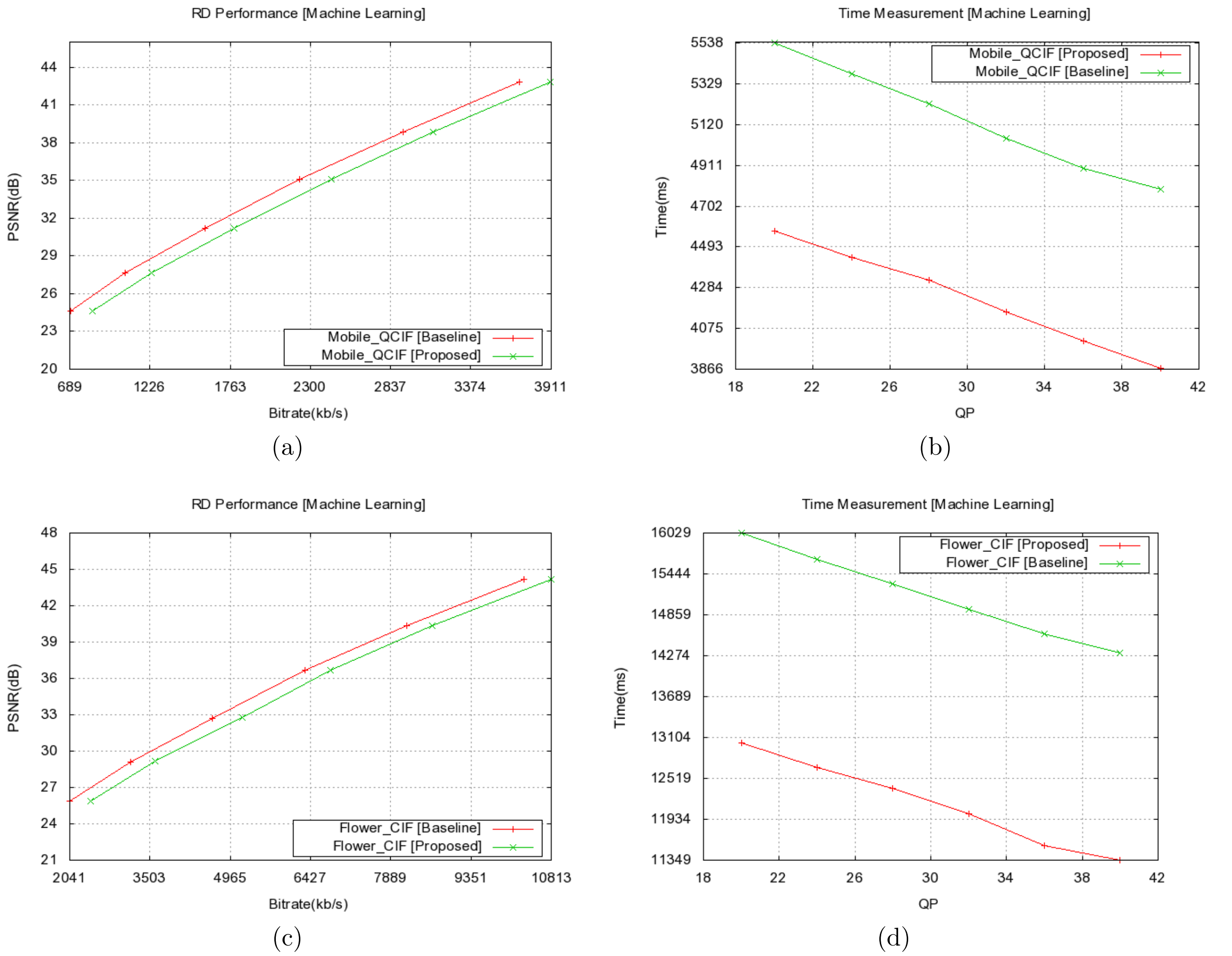


Figure 5. (a)(c) RD performance of the H.264 Intra frame encoder with Machine Learning classifier; (b)(d)Time complexity reduction.

of the H.264 Intra MB coder developed by machine learning approach in[1] for the same sequences used in Fig. 4. From both of these figures, it is clear that the performance of both of these classifiers is almost similar.

5. CONCLUSIONS

This paper presents a novel approach to H.264 Intra MB mode computation based on chance constrained classifier. The proposed approach has great potential to reduce the computational complexity and is a new type of classifer explored. The results of the implementation in JM 14.2 show that the RD performance is very close to the reference encoder and with each other. The encoding time for chance constrained classifier is reduced by about 1/5 compared 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.

REFERENCES

[1] Kalva, H., Kunzelmann, P., Jillani, R., and Pandya, A., "Low complexity h.264 intra mb coding," *Consumer Electronics, 2008. ICCE 2008. Digest of Technical Papers. International Conference on* , 1–2 (Jan. 2008).

[2] Fernández, G., Cuenca, P., Barbosa, L. O., and Kalva, H., "Very low complexity mpeg-2 to h.264 transcoding using machine learning," in [*MULTIMEDIA '06: Proceedings of the 14th annual ACM international conference on Multimedia*], 931–940, ACM, New York, NY, USA (2006).

[3] Pan, F., Rahardja, L., Lim, K., Wu, L., Wu, W., Zhu, C., Ye, W., and Liang, Z., "Fast intra mode decision algorithm for h.264-avc video coding," *Image Processing, 2004. ICIP '04. 2004 International Conference on* **2**, 781–784 Vol.2 (Oct. 2004).

[4] Kim, C., Shih, H.-H., and Kuo, C.-C., "Feature-based intra-prediction mode decision for h.264," *Image Processing, 2004. ICIP '04. 2004 International Conference on* **2**, 769–772 Vol.2 (Oct. 2004).

[5] Fernandez-Escribano, G., Kalva, H., Cuenca, P., and Orozco-Barbosa, L., "Speeding-up the macroblock partition mode decision in mpeg-2/h.264 transcoding," *Image Processing, 2006 IEEE International Conference on* , 869–872 (Oct. 2006).

[6] Nath, J. S., Bhattacharyya, C., and Murty, M. N., "Clustering based large margin classification: a scalable approach using socp formulation," in [*KDD '06: Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining*], 674–679, ACM Press, New York, NY, USA (2006).

[7] Kalva, H. and Christodoulou, L., "Using machine learning for fast intra mb coding in h.264," *Visual Communications and Image Processing 2007* **6508**(1), 65082U, SPIE (2007).