# RATE-DISTORTION ESTIMATION FOR H.264/AVC CODERS

*Yu-Kuang Tu*, Jar-Ferr Yang*, and Ming-Ting Sun*[+]

*Institute of Computer and Communication Engineering, Department of Electrical Engineering,
National Cheng Kung University, Taiwan
[+]Department of Electrical Engineering, University of Washington, USA

## ABSTRACT

In a video coder, the optimal coding mode decision for each coding block could be achieved by exhaustively calculating the Lagrange cost (which includes the coding distortion plus the Lagrange parameter times the coding bit consumption) of all possible modes. The best mode can then be chosen as the one with the minimum Lagrange cost. To speed up the computationally intensive Lagrange cost computation, in this paper, we propose transform-domain bit-rate estimation and distortion measures for the inter-mode decision in H.264/AVC coders. With the proposed scheme, entropy coding, inverse DCT, and pixel-reconstructions are not required in the process. Simulation results show that the proposed estimation method is accurate for the inter-mode decision and about 46.42% time reduction can be achieved.

## 1. INTRODUCTION

International video coding standards such as MPEG-1, MPEG-2, MPEG-4, H.263, and H.264/AVC [1] only specify the decoding process; they give considerable flexibility for optimizing the encoder for coding performance improvement and complexity reduction. To compare the performance of different video encoders, the rate-distortion (R-D) performance is often used. The rate-distortion optimization for video encoding by using Lagrangian techniques is addressed in [2]. It controls the H.263 video encoder to optimally decide a mode such as INTRA, SKIP, INTER-$16 \times 16$, or INTER-$8 \times 8$ for the encoding. Although different coding options for a macroblock (MB) are supported in different standards, the Lagrangian optimization technique gives a general way to choose an optimal coding mode [3]. For instance, a video encoder can be optimized via performing the mode-decision by minimizing the Lagrangian cost $J$,

$$J = D + \lambda \cdot R, \qquad (1)$$

where $\lambda$ is the Lagrange multiplier, $D$ is the reconstruction distortion, and $R$ is the number of coded bits of each coding unit. In an H.264/AVC encoder, inter-modes include the partition of a macroblock into $16 \times 16$, $16 \times 8$, $8 \times 16$, and P$8 \times 8$ modes. In the P$8 \times 8$ mode, each $8 \times 8$ block could be further divided into 4 possible partition modes such as $8 \times 8$, $8 \times 4$, $4 \times 8$, and $4 \times 4$ subblock modes. The overall minimization of the Lagrange cost should cover all computation of the costs of 3 macroblock sizes and 16 sub-macroblock sizes for inter prediction. For intra prediction, Intra_$16 \times 16$ and Intra_$4 \times 4$ are used for predicting the content of a macroblock from the reconstructed pixels in the adjacent blocks coded previously. One of 4 modes should be selected for the Intra_$16 \times 16$ prediction, and one of 9 modes for each $4 \times 4$ block. A total of 144 ($9 \times 16$) costs have to be computed for the Intra_$4 \times 4$ macroblock mode decision. Therefore, the actual calculations of the distortion and the required bit-consumption for all candidate modes are very computationally intensive.

In this paper, we focus on the inter-mode decision and propose an efficient bit-rate estimation function and distortion measures, both directly obtained in the transform domain. Our goal is to reduce the computation load for rate-distortion optimization. From experimental results, the accuracy of proposed algorithm is justified.

## 2. MODE DECISION WITH RATE-DISTORTION OPTIMIZATION

The rate-distortion tradeoff can be optimized by properly selecting coding options for different coding blocks in the image. In a video picture, the content is usually divided into $16 \times 16$ macroblocks. The rate-distortion optimization for the macroblock mode decision is usually based on Lagrange techniques. For each macroblock, the optimal coding mode is the one with the minimum Lagrange cost. Suppose that a macroblock has $K$ possible modes. In the $i$th mode, a macroblock is first evenly divided into $n_i$ sub-blocks each with $N_1 \times N_2$ pixels. The cost of the $i$th mode is

$$J_i = D_i + \lambda \cdot R_i = \sum_{j=1}^{n_i} D_{ij} + \lambda \cdot R_{ij}, \; i = 1, 2, \ldots, K. \qquad (2)$$

The distortion measure $D_i$ is the Sum of Squared Differences (SSD) between the reconstructed and the original pixels in a macroblock, and $R_i$ is the required bits to encode the macroblock. In this paper, since we focus on the inter mode decision, the rate $R_i$ generated after entropy coding may include the bit consumption of

*0-7803-9332-5/05/$20.00 ©2005 IEEE*

quantized DCT coefficients, motion vector data, and the header.

In the H.264/AVC reference software [4], the Lagrange parameter for mode decision $\lambda_{\text{MODE}}$ is a function of the quantization parameter (QP):

$$\lambda_{\text{MODE}}(\text{QP}) = 0.85 \cdot 2^{(\text{QP}-12)/3}. \quad (3)$$

In the motion search stage, $\lambda_{\text{MOTION}} = \sqrt{\lambda_{\text{MODE}}}$ is used for the rate-constrained motion estimation when Sum of Absolute Differences (SAD) is considered, and $\lambda_{\text{MOTION}} = \lambda_{\text{MODE}}$ when SSD is used.

## 3. EFFICENT BIT-RATE ESTIMATION

In DCT-based block video coding, an $N \times N$ residual block is obtained after prediction, and then processed by DCT, quantization, and entropy encoding. The quantization process can be represented as

$$\hat{Y}(i,j) = round\left(Y(i,j)/qs\right), \quad (4)$$

where $Y(i,j)$ and $\hat{Y}(i,j)$, $i, j = 0, 1, \ldots, N-1$, are the original and the quantized $N \times N$ DCT coefficients, $qs$ is the quantization step-size which is referred by QP specified by the standard. In (4), round( ) is the rounding operation. The magnitudes and the number of nonzero quantized DCT coefficients $\hat{Y}(i,j)$ are determined by $qs$. Recently, the accuracy of $\rho$ (number of zero DCT coefficients) -domain rate-model has been justified in rate-control and bit-allocation schemes [5]. Here, we model the bit consumption of the quantized DCT coefficients, $\hat{B}_{\text{COF}}$ as,

$$\hat{B}_{\text{COF}} = \alpha \cdot N_{nz} + \beta \cdot E_{QTC}, \quad (5)$$

where $N_{nz}$ denotes the number of nonzero quantized DCT coefficients, and $E_{QTC}$ represents the $l_p$-norm of the quantized DCT coefficients defined by

$$E_{QTC} = \left( \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} \left| \hat{Y}(i,j) \right|^p \right)^{1/p}. \quad (6)$$

For computation reduction, we could choose $p = 1$. In [6], it has been found that the above model is accurate in the MPEG-4 rate-control scheme for small $qs$. For H.264/AVC, we should modify the above bit-estimation function for better mode decision.

In the minimization problem addressed in Section 2, $\lambda_{\text{MODE}}(\text{QP})$ will affect the search results, motion-compensated errors, properties of quantized DCT coefficients, mode-decision, and the coding bits. As a result, Eq. (5) can be rewritten in a general form as

$$\begin{aligned} \hat{B}_{\text{COF}}\left(\lambda(\text{QP}_\text{M}), \text{QP}_Q, \text{MODE}\right) \\ = \alpha\left(\lambda(\text{QP}_\text{M}), \text{QP}_Q, \text{MODE}\right) \cdot N_{nz}\left(\text{QP}_Q, \text{MODE}\right) \\ + \beta\left(\lambda(\text{QP}_\text{M}), \text{QP}_Q, \text{MODE}\right) \cdot E_{QTC}\left(\text{QP}_Q, \text{MODE}\right), \end{aligned} \quad (7)$$

where $\lambda(\text{QP}_\text{M})$ as in (3) is used for mode decision and motion estimation, and $\text{QP}_Q$ is used for quantization.

Since for H.264/AVC without rate-control, $\text{QP}_Q = \text{QP}_\text{M} = \text{QP}$, ($\lambda(\text{QP}_\text{M})$, $\text{QP}_Q$, MODE) can be denoted as (QP, MODE) for clarity. When computing the total coding bits $B_{\text{TOT}}$(QP, MODE) for one mode of a coding unit, it consists of $B_{\text{COF}}$(QP, MODE), the bits of motion vector data (MVD) $B_{\text{MVD}}$(QP, MODE), and the bits of header $B_{\text{HEADER}}$(QP, MODE) including coding mode, reference frame, and CBP, etc.:

$$\begin{aligned} B_{\text{TOT}}(\text{QP}, \text{MODE}) = B_{\text{COF}}(\text{QP}, \text{MODE}) \\ + B_{\text{HEADER}}(\text{QP}, \text{MODE}) + B_{\text{MVD}}(\text{QP}, \text{MODE}). \end{aligned} \quad (8)$$

In the motion estimation step, we can obtain $B_{\text{MVD}}$(QP, MODE) during search. If the coding scheme is UVLC in H.264/AVC, $B_{\text{HEADER}}$(QP, MODE) is also easy to obtain by Look-Up Table (LUT). After we add the bit-consumption of DCT coefficients and CBP as whole $B_{\text{COF}}$(QP, MODE), the function in (7) will be accurate. By adopting this bit-rate estimation function, the actual entropy coding process (CAVLC) can be skipped when performing the rate-distortion optimized mode decision, and thus computational load is alleviated.

Parameters $\alpha$(QP, MODE) and $\beta$(QP, MODE) can be obtained from empirical results in the H.264/AVC encoding of several video sequences. In order to be adaptive to the characteristics of different video sequences or different frames in a video sequence, parameters $\alpha$(QP, MODE) and $\beta$(QP, MODE) can be updated by linear regression. In our simulations, we use the same $\alpha$(QP) and $\beta$(QP) for each MODE. Under a certain QP, $\alpha$(QP) and $\beta$(QP) are updated as

$$\alpha = \frac{\left(\sum_{k=1}^{n} E_k^2\right)\left(\sum_{k=1}^{n} N_k B_k\right) - \left(\sum_{k=1}^{n} N_k E_k\right)\left(\sum_{k=1}^{n} B_k E_k\right)}{\left(\sum_{k=1}^{n} E_k^2\right)\left(\sum_{k=1}^{n} N_k^2\right) - \left(\sum_{k=1}^{n} N_k E_k\right)^2} \quad (9)$$

and

$$\beta = \frac{\left(\sum_{k=1}^{n} N_k^2\right)\left(\sum_{k=1}^{n} B_k E_k\right) - \left(\sum_{k=1}^{n} N_k E_k\right)\left(\sum_{k=1}^{n} N_k B_k\right)}{\left(\sum_{k=1}^{n} E_k^2\right)\left(\sum_{k=1}^{n} N_k^2\right) - \left(\sum_{k=1}^{n} N_k E_k\right)^2}, \quad (10)$$

where $n$ is the number of observed macroblocks in the past e.g., the previous frame. For simplicity, we abbreviate $N_{nz,k}$(QP), $E_{QTC,k}$(QP) and $B_{\text{COF},k}$(QP) as $N_k$, $E_k$, and $B_k$, respectively.

## 4. COMPLEXITY-REDUCED DISTORTION MEASURE

In H.264/AVC, for a $4 \times 4$ luma residual block $\boldsymbol{X}$, its transform coefficients matrix $\boldsymbol{Y}$ can be expressed as [7]

$$\boldsymbol{Y} = T(\boldsymbol{X}) = \left(\boldsymbol{C}_f \boldsymbol{X} \boldsymbol{C}_f^T\right) \otimes \boldsymbol{E}_f = \boldsymbol{W} \otimes \boldsymbol{E}_f, \quad (11)$$

where $\mathbf{C}_f$ is the forward integer transform in H.264/AVC and $\boldsymbol{E}_f$ is the post-scaling factor matrix which is absorbed

in the quantization process. The reconstructed residual block $\hat{\boldsymbol{X}}$ is obtained from the inverse transform of the inverse quantized coefficients

$$\hat{\boldsymbol{X}} = \boldsymbol{C}_i^T \left(\hat{\boldsymbol{Y}} \otimes \boldsymbol{E}_i\right) \boldsymbol{C}_i = \tfrac{1}{64} \boldsymbol{C}_i^T \boldsymbol{W}' \boldsymbol{C}_i , \qquad (12)$$

where $\mathbf{C}_i$ is the inverse integer transform in H.264/AVC and $\boldsymbol{E}_i$ is the pre-scaling factor matrix, which could be absorbed in the inverse quantization process. Therefore, the quantization error matrix, which is denoted as the Spatial Domain Distortion (SDD) matrix between the reconstructed block and the original block, is expressed by

$$\begin{aligned} \boldsymbol{D} &= ((\boldsymbol{X} - \hat{\boldsymbol{X}}))^2 \\ &= \left(\left((\boldsymbol{C}_f)^{-1} \boldsymbol{W} \left(\boldsymbol{C}_f^T\right)^{-1} - \tfrac{1}{64} \boldsymbol{C}_i^T \boldsymbol{W}' \boldsymbol{C}_i\right)\right)^2 \\ &= \left(\left((\boldsymbol{C}_f)^{-1} \boldsymbol{W} \left(\boldsymbol{C}_f^T\right)^{-1} - \tfrac{1}{64} (\boldsymbol{C}_f)^{-1} \boldsymbol{P} \boldsymbol{W}' \boldsymbol{P} \left(\boldsymbol{C}_f^T\right)^{-1}\right)\right)^2 \\ &= \left(\left((\boldsymbol{C}_f)^{-1} \left(\boldsymbol{W} - \tfrac{1}{64} \boldsymbol{P} \boldsymbol{W}' \boldsymbol{P}\right) \left(\boldsymbol{C}_f^T\right)^{-1}\right)\right)^2 , \end{aligned} \qquad (13)$$

where $((\boldsymbol{B}))^2$ denotes the matrix whose elements are the squared values of the elements of $\boldsymbol{B}$ in the corresponding positions and $\boldsymbol{P}$ = diag(4, 5, 4, 5). The Transform Domain Distortion (TDD) matrix, $T(\boldsymbol{D})$ obtained from the transformed quantization error matrix is given by

$$\begin{aligned} T(\boldsymbol{D}) &= ((\boldsymbol{Y} - \hat{\boldsymbol{Y}}))^2 = \left(\left(T(\boldsymbol{X} - \hat{\boldsymbol{X}})\right)\right)^2 \\ &= \left(\left(\boldsymbol{C}_f (\boldsymbol{C}_f)^{-1} \left(\boldsymbol{W} - \tfrac{1}{64} \boldsymbol{P} \boldsymbol{W}' \boldsymbol{P}\right) \left(\boldsymbol{C}_f^T\right)^{-1} \boldsymbol{C}_f^T\right) \otimes \boldsymbol{E}_f\right)^2 \\ &= \left(\left(\boldsymbol{W} - \tfrac{1}{64} \boldsymbol{P} \boldsymbol{W}' \boldsymbol{P}\right)\right)^2 \otimes \left((\boldsymbol{E}_f)\right)^2 . \end{aligned} \qquad (14)$$

For an $8 \times 8$ chroma residual block, we need to perform another $2 \times 2$ transform of $\boldsymbol{Z}$, which consists of four DC terms $\boldsymbol{W}_{(0,\,0)}$. We also consider the transform-domain quantization error:

$$\begin{aligned} &\left(\left(T(\boldsymbol{W}_{(0,0)} - \tfrac{1}{64} P_{(0,0)}^2 \boldsymbol{W}'_{(0,0)})\right)\right)^2 \\ &= \left(\left(\boldsymbol{H}(\boldsymbol{H})^{-1} \left(\boldsymbol{Z} - \tfrac{1}{64} \boldsymbol{M} \boldsymbol{Z}' \boldsymbol{M}\right) \left(\boldsymbol{H}^T\right)^{-1} \boldsymbol{H}^T\right)\right)^2 \otimes \left((\boldsymbol{E}_H)\right)^2 \\ &= \left(\left(\boldsymbol{Z} - \tfrac{1}{64} \boldsymbol{M} \boldsymbol{Z}' \boldsymbol{M}\right)\right)^2 \otimes \left((\boldsymbol{E}_H)\right)^2 , \end{aligned} \qquad (15)$$

where $\boldsymbol{H}$ is the $2 \times 2$ non-normalized Hadamard transform, $\boldsymbol{E}_H$ is the normalization matrix and $\boldsymbol{M}$ = $P_{(0,\,0)}$diag(2, 2). Since the post-scaling factor of the DC term of each $4 \times 4$ transform is absorbed in the quantization process after repeated transform is applied, the transform-domain quantization error of the gathered $2 \times 2$ chroma DC block is

$$T(\boldsymbol{D}_{(0,0)}) = \tfrac{1}{64} \left(\left(\boldsymbol{Z} - \tfrac{1}{64} \boldsymbol{M} \boldsymbol{Z}' \boldsymbol{M}\right)\right)^2 . \qquad (16)$$

In a $4 \times 4$ block, the total TDD is the sum of each element in $T(\boldsymbol{D})$. For a luma transform coefficient:

$$T(\boldsymbol{D})_{(u,v)} = \left(\left(W_{(u,v)} - \tfrac{1}{64} W'_{(u,v)} P_{(u,u)} P_{(v,v)}\right) E_{f\,(u,v)}\right)^2 . \qquad (17)$$

When a larger block consisting of several $4 \times 4$ blocks is considered, we can group the integer DCT coefficients with the same frequency ($u$, $v$) together. Furthermore, since different frequency components may share the same post-scaling factor in $\boldsymbol{E}_f$, we can also collect those components and classify them into three groups: 1) those with $u$ and $v$ are even, 2) those with $u$ and $v$ are odd and 3) otherwise. Note that the quantization error discussed above does not take into account the clipping function applied when the reconstructed prediction error is added to the compensated signal. However, from our experiments shown in Section 5, the total TDD of a block is close to the actual SDD. If TDD is used as the distortion measure, the computation of IDCT and block reconstruction can be saved.

## 5. EXPERIMENTAL RESULTS

To verify the proposed rate-distortion estimation algorithm, several test video sequences are used in simulations under different QPs. By using the proposed estimation, the estimated curves are plotted to compare to the real $B_{\mathrm{COF}}$-$qs$ curves. In Fig. 1, estimation results of some macroblocks are given. As it can be seen, the proposed estimation achieves precise prediction of the coding bits. We also compare the total transform domain distortion of each marcoblock to the total spatial domain reconstructed distortion. As shown in Fig. 2, the transform domain estimation is very close to that of spatial domain.

Finally, we use the bit-rate estimation function and transform-domain distortion measure for the rate-distortion optimization in the H.264/AVC reference software JM version 8.2 [4]. In the simulations, $16 \times 16$, $16 \times 8$, $8 \times 16$, and $8 \times 8$ block-sizes are utilized for the inter prediction. Some important parameters are set as follows: (a) Sequence type is IPPP…, (b) Search range is 33 × 33, (c) Number of reference frames is 1, (d) Hadamard transform is used, and (e) Entropy coding method is CAVLC. The proposed algorithm achieves 46.42% time reduction in average as list in Table 1 while only slight performance degradation, as shown in Fig. 3, is introduced compared to the original H.264/AVC encoder with the rate-distortion optimization.

## 6. CONCLUSION

Rate-distortion optimization plays an important role in optimizing a coding scheme like H.264/AVC which possesses many coding options for each coding unit. However, the process is very computation demanding. We first propose an efficient bit-rate estimation function via analyzing the number and levels of nonzero quantized DCT coefficients with coding bits consumption. Then, we measure the distortion in the transform domain by using the integer (Q/IQ) DCT coefficients and mathematical manipulations. With the proposed scheme, entropy coding, IDCT, and block reconstruction can be skipped during the mode decision.

## REFERENCES

[1] Joint Video Team, "Draft ITU-T Recommendation and Final Draft International Standard of Joint Video Specification (ITU-T Rec.H.264 | ISO/IEC 14496-10

AVC),” Joint Video Team (JVT) of ISO/IEC MPEG and ITU-T VCEG, JVT-G050, March 2003.

[2] T. Wiegand, M. Lightstone, D. Mukherjee, T. G. Campbell, and S. K. Mitra, “Rate-distortion optimized mode selection for very low bit rate video coding and the emerging H.263 standard,” *IEEE Trans. Circuits Syst. Video Technol.*, vol. 6, no. 2, pp. 182-190, Apr. 1996.

[3] T. Wiegand, H. Schwarz, A. Joch, F. Kossentini, and G. J. Sullivan, “Rate-constrained coder control and comparison of video coding standards,” *IEEE Trans. Circuits Syst. Video Technol.*, vol. 13, no. 7, pp. 688-703, July 2003.

[4] Joint Video Team (JVT) Reference Software [Online]. http://bs.hhi.de/~suehring/tml/download/

[5] Z. He and S. K. Mitra, “Optimum bit allocation and accurate rate control for video coding via $\rho$-domain source modeling,” *IEEE Trans. Circuits Syst. Video Technol.*, vol. 12, no. 10, pp.840-849, Oct. 2002.

[6] S.-C. Chang, J.-F. Yang, C.-F. Lee and J.-N. Hwang, “A novel rate predictor based on quantized DCT indices and its rate control mechanism,” *Signal Processing: Image Commun.*, vol. 18, Issue 6, pp. 427-441, July 2003.

[7] A. Hallapuro and M. Karczewicz, “Low complexity transform and quantization,” in Joint Video Team (JVT), Jan. 2002 Docs. JVT-B038 and JVT-B039.

Table 1. The percentage of time reduction

| Sequence | Quantization Parameter, QP | | | | Average |
|---|---|---|---|---|---|
| | 22 | 28 | 34 | 40 | |
| Foreman | 47.78% | 39.39% | 27.60% | 25.15% | 46.42% |
| Stefan | 71.94% | 64.59% | 50.82% | 44.10% | |

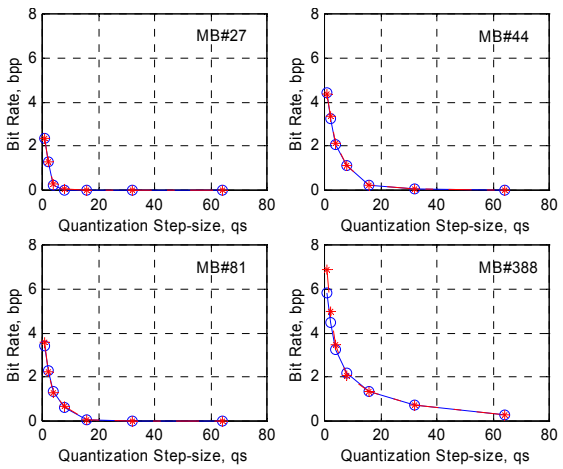
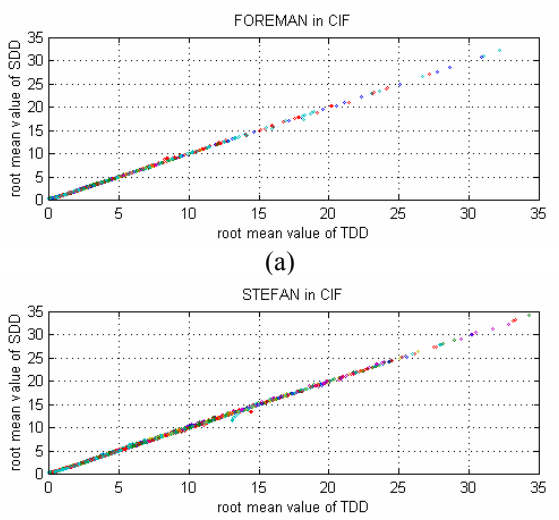


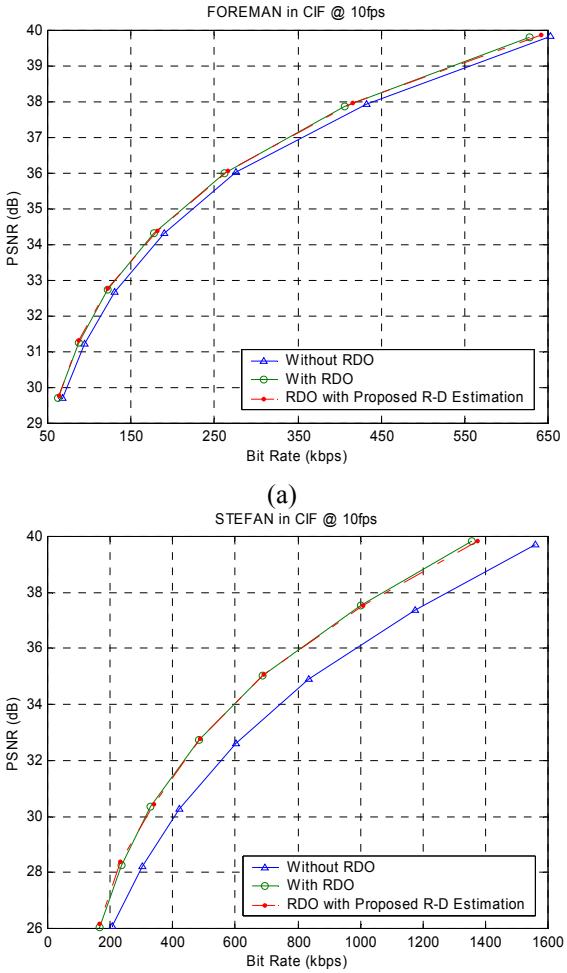

Fig. 1. Comparison between estimated (dash-dot line with star) and actual *B-qs* (solid line with hollow circle) curves.




(a)

(b)

Fig. 2. Macroblock-based TDD versus SDD.




(a)

(b)

Fig. 3. Comparison of *R-D* curves among the original H.264/AVC and the proposed algorithm.