Efficient Inter Mode Prediction Based on Model Selection and Rate Feedback for H.264/AVC

Kuan-I Lee, An-Chao Tsai, Jhing-Fa Wang, *Fellow, IEEE,* and Jar-Ferr Yang, *Fellow, IEEE*

***Abstract*—H.264/AVC is a standard developed for various low-complexity video applications and high-definition television. To improve coding performance, H.264/AVC may optionally adopt the rate-distortion optimization (RDO) method to find the best encoding mode among various inter and intra modes. However, the exhaustive RDO search among different modes increases the H.264/AVC encoder complexity and limits its application. In this paper, we propose an inter mode prediction algorithm for P slices based on spatial and temporal consistency analysis to reduce the complexity of the RDO computation. We apply the stochastic method to analyze the spatial consistency and use rate information for temporal consistency. The experimental results show a 0.03 peak signal-to-noise ratio loss, a 0.87% bit rate increase, and a 58.39% encoding time reduction on average.**

***Index Terms*—H.264/AVC, inter mode decision, spatial-temporal, stochastic analysis.**

I. Introduction

VIDEO compression reduces the required bandwidth for transmitting video signals. However, as video quality increases, higher bandwidth is required. In order to overcome this limitation, H.264/AVC, an emerging video coding standard, was jointly developed by ITU-T and ISO/IEC [1]–[3]. It includes many new video compression features, such as variable block size motion estimation (VBSME), multi-reference frames, and sub-pixel motion estimation (ME). In VBSME, a macroblock (MB) is divided into MB partitions (16 × 16, 16 × 8, 8 × 16, and P8 × 8) and sub-MB partitions (8×8, 8×4, 4×8, and 4×4) for ME. The sub-MB partitions are obtained by further dividing the $P8 \times 8$ block type, as shown in Fig. 1. Thus, there are seven inter modes ($Inter16 \times 16$, $Inter16 \times 8$, $Inter8 \times 16$, $Inter8 \times 8$, $Inter8 \times 4$, $Inter4 \times 8$, and $Inter4 \times 4$), two intra modes ($Intra16 \times 16$ and $Intra4 \times 4$), and one skip mode for an MB in inter frames. The rate distortion optimization (RDO) is optionally used to find the best coding mode [one with the minimum rate-distortion cost (RDcost)] to achieve the highest coding efficiency in H.264/AVC. Hence, the total number of RDO calculations to find the best mode is 768, which significantly increases the complexity of the H.264/AVC encoder [4].

The RDO procedure is as follows: RDO is used for the mode decision scheme to choose the best mode among seven inter modes, one SKIP mode, and two intra modes [5] as follows:

$$\begin{aligned} J(s, c, Mode|QP, \lambda_{Mode}) = SSD(s, c, Mode|QP) \\ + \lambda_{Mode} R(s, c, Mode|QP) \end{aligned} \quad (1)$$

where J denotes the RDcost function, which is used to find the mode with the minimum cost. SSD represents the sum of the squared difference between the original MB (s) and its reconstruction (c) with quantization parameter (QP). R represents the total number of coding bits for MB header information, motion vector (MV) information, and integer discrete cosine transform (DCT) quantization coefficients through context adaptive variable length coding (CAVLC) or context adaptive binary arithmetic coding. The Lagrangian multiplier λ_{Mode}, which depends on the QP value, is used for P slices as follows:

$$\lambda_{Mode} = 0.85 \times 2^{QP/3}. \quad (2)$$

Many fast algorithms have been proposed to decrease the inter mode computations for reducing the RDO complexity of H.264/AVC. The complexity reduction methods can be roughly divided into three categories: SKIP mode detection, mode prediction from spatial or temporal correlation, and hierarchical search of VBSME. Efficient SKIP mode prediction methods were proposed in [6]–[8]; however, they significantly degrade video quality. Methods in the second category use a variety of spatial features such as textures and edge information to predict the best mode [9], [10]. And there were few studies taking advantage of temporal correlation [11] or motion field [12], [13] to perform inter mode prediction recently. When the spatial and temporal features are considered together, more information can be used to select the most suitable mode. Finally, methods in [14], [15] performed hierarchical VBSME by first using the MB group (i.e., $Inter16 \times 16$, $Inter16 \times 8$, and $Inter8 \times 16$); the sub-MB group ($P8 \times 8$) is used when required.

In this paper, we focus on reducing redundant modes using the model selection criteria (MSC) and the proposed

Manuscript received April 19, 2009; revised September 17, 2009, December 30, 2009, April 18, 2010 and July 30, 2010; accepted October 24, 2010. Date of publication March 17, 2011; date of current version June 3, 2011. This work was supported in part by the National Science Council, Taiwan, under Grant NSC 99-2218-E-006-001. This paper was recommended by Associate Editor G. Wen.

K.-I. Lee is with Realtek Semiconductor Corporation, Hsinchu 300, Taiwan (e-mail: kuani@icwang.ee.ncku.edu.tw).

A.-C. Tsai and J.-F. Wang are with the Department of Electronic Engineering, National Cheng Kung University, Tainan 701, Taiwan (e-mail: anchao@icwang.ee.ncku.edu.tw; wangjf@csie.ncku.edu.tw).

J.-F. Yang is with the Technologies of Ubiquitous Computing and Humanity Center, National Cheng Kung University, Tainan 701, Taiwan (e-mail: jfyang@ee.ncku.edu.tw).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TCSVT.2011.2129730

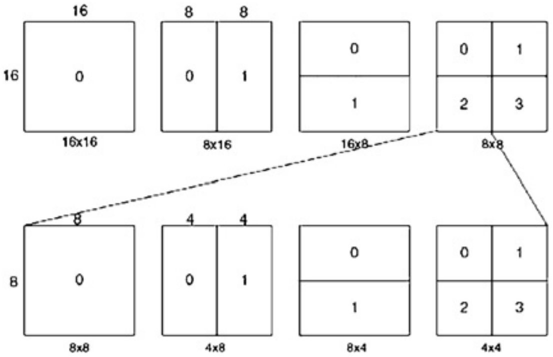


Fig. 1. Seven inter modes defined in H.264.

correctness scheme to reduce the computational complexity of RDO. MSC, which is based on stochastic analysis, is used to detect the spatial characteristic of the current MB before ME, and the correctness scheme uses rate information feedback to correct the predicted mode. Experimental results show that the proposed algorithm with spatial-temporal correlation can efficiently reduce the computation complexity with negligible coding loss.

The rest of this paper is organized as follows. The proposed spatial and temporal-based algorithms are explained in Sections II and III, respectively. The experiment results and discussion are presented in Section IV. Finally, the conclusions are given in Section V.

II. Proposed Algorithm: Spatial-Based

A. Overview of the Proposed Algorithm

According to (1), VBSME is based on spatial and temporal correlations for finding the minimum RDcost. However, most of the previous studies did not consider temporal correlation when performing inter mode coding. In this paper, we propose a new inter mode coding flow with feedback information to reduce computational complexity. The overall coding flow of the proposed algorithm includes two main steps, as shown in Fig. 2. The "PIM" stands for predicted inter mode of an MB. Details of each step are described below.

B. Spatial-Based Mode Prediction Scheme

Spatial information can be used to reduce computation complexity. Image segmentation, a well-known technique used in spatial analysis, can be applied to VBSME [16], [17]. Many methods have been developed for image segmentation, such as Markov random field (MRF) [18] and random sample consensus [19]. In the proposed method, we focus on MRF with a model selection criterion for reducing the number of inter coding modes of H.264/AVC.

1) *Spatial Feature Extraction: Sample Variance:* We assume that the pixel data ($x_{i,\,j}$) in an MB region are independent and that each data point is normally distributed with mean (μ) and variance (σ^2). Applying the maximum likelihood estimate (MLE) theorem to estimate the mean and variance based on the observed sample data, the continuous joint probability density function of these pixel data can be described as follows:

TABLE I

Conditions of SKIP Mode

1) The best motion compensation block size for the MB is *Inter*16×16.
2) The best reference frame is the previous frame.
3) The best MV is the predicted MV.
4) The transform coefficients of the 16×16 block size are all quantized to zero.

$$f\left(x_{0,0}, x_{0,1}, \ldots, x_{15,15} | \mu, \sigma^2\right) = \frac{1}{\left(\sqrt{2\pi\sigma^2}\right)^{256}} \prod_{i=0,j=0}^{i=15,j=15} e^{-\frac{\left(x_{i,j}-\mu\right)^2}{2\sigma^2}}$$
$$0 \le x_{i,j} \le 255. \tag{3}$$

The maximum likelihood estimators, i.e., the sample mean ($\hat{\mu}$) and the sample variance ($\hat{\sigma}^2$), used for MLE are as follows:

$$\begin{aligned} \hat{\mu} &= \tfrac{1}{R} \textstyle\sum_{(i,j)\in R} x(i,\,j) \\ \hat{\sigma}^2 &= \tfrac{1}{R} \textstyle\sum_{(i,j)\in R} \left[x(i,\,j) - \hat{\mu}\right]^2. \end{aligned} \tag{4}$$

The sample variance ($\hat{\sigma}^2$) is a useful measurement criterion for smooth areas. When it is larger than a predefined threshold, we categorize the MB as a texture but homogenous with a smaller value. It performs image classification on a single raw frame effectively, but we are concerned with the relationship between two successive frames, i.e., the basic demand for inter frame coding. Fig. 3(a) shows the stochastic amounts of the homogeneous block (i.e., *SKIP* and *Inter*16 × 16) in the *Foreman* test sequence encoded by the JM reference software. The results show that these blocks have a wide distribution of the sample variance. Even when the sample variance is 3000, the MB still belongs to the *SKIP* mode. This might be due to some texture content (non-homogeneous) MBs being either stationary or having only one MV. Thus, the *SKIP* or *Inter*16 × 16 is the most suitable mode for this texture MB. This affects the selection of a suitable threshold for deciding the *SKIP* mode. This circumstance can be solved by replacing $x_{i,j}$ with (5), i.e., taking the residual frame instead of the raw frame, where the suffix t is the frame index in a video sequence. Comparing Fig. 3(a) and (b), we can see that the distribution of $\hat{\sigma}^2$ is improved and that the threshold setting is more explicit as follows:

$$x_{i,j} = \left|(x_{i,j})_t - (x_{i,j})_{t-1}\right|. \tag{5}$$

As a result, the coding MBs can be classified as homogenous or textures. The homogenous MBs can be encoded by *SKIP* or *Inter*16 × 16 mode by the SKIP mode conditions, as shown in Table I. The intra modes are omitted in the RDCost computation. Texture MBs are classified using the MSC method, as described below.

2) *Spatial Feature Extraction: MSC:* The MSC model is derived from the MRF theorem, which is a probabilistic model describing the spatial correlation of an image. Several methods can be utilized to estimate the MRF parameters, such as iterative condition model (ICM) [20], [21] or the high confidence first (HCF) model [22]. Some ICM models require iterative computation while HCF models use the given data to estimate parameters to describe the image model. Since

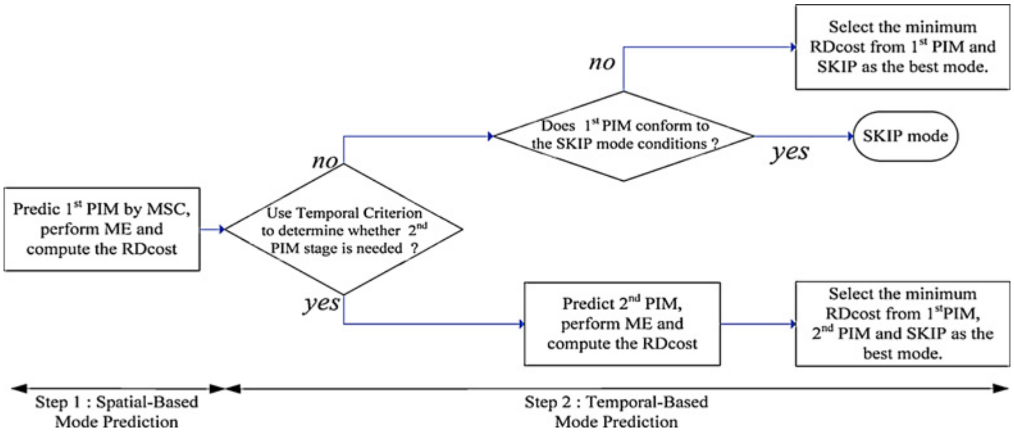

Predic 1st PIM by MSC, perform ME and compute the RDcost
Use Temporal Criterion to determine whether 2nd PIM stage is needed ?
no
yes
Does 1st PIM conform to the SKIP mode conditions ?
no
yes
Select the minimum RDcost from 1st PIM and SKIP as the best mode.
SKIP mode
Predict 2nd PIM, perform ME and compute the RDcost
Select the minimum RDcost from 1stPIM, 2nd PIM and SKIP as the best mode.
Step 1 : Spatial-Based Mode Prediction
Step 2 : Temporal-Based Mode Prediction

Fig. 2. Proposed flow of inter mode decision.

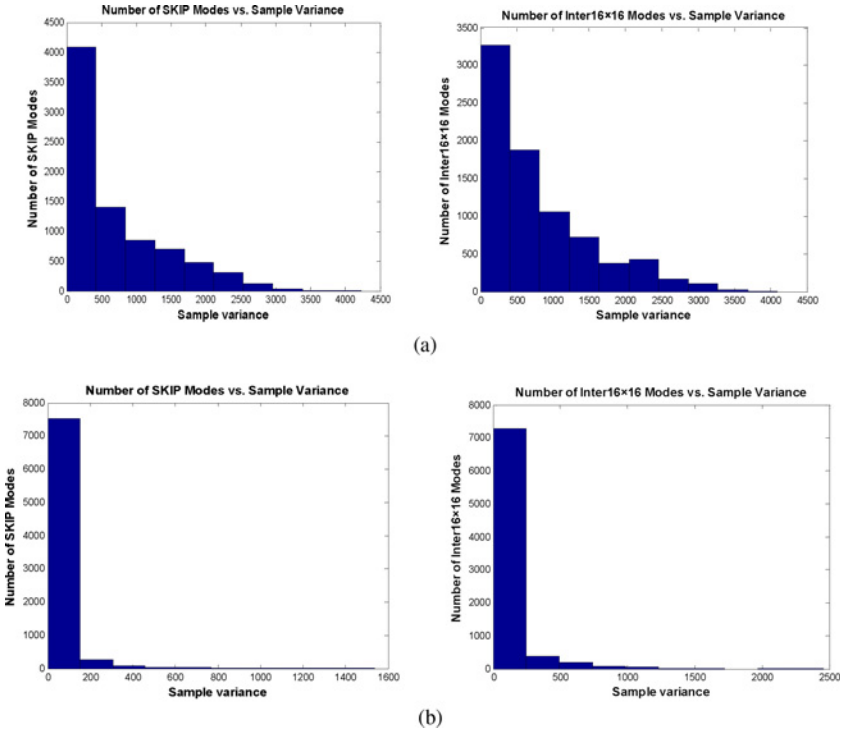


Fig. 3. Sample variance distribution of homogenous modes (SKIP, 16 × 16). (a) Extracted from a single frame. (b) Extracted from the residual frame (*Foreman*.QCIF).

iterative methods require a lot of computation, they are not suitable for real-time applications. Hence, the stochastic model selection criterion is selected in this paper.

The general form of the MSC [23] is expressed as follows:

$$G(k) = \ln[f(k)] - a(N)M(k) - b(k,n) \tag{6}$$

where k represents distinct textures in the region, *f(k)* is the joint probability density function (PDF), and $M(k)$ is the number of constants in the specified texture. Both a (N) and $b(k, n)$ are specified subject to the various MSC, which greatly affects the feature extraction efficacy. Refer to [24] for more details about (6). In this paper, (3) can be substituted for the PDF in (6). Then, the Gaussian log-likelihood function, $\ln[f(k)]$, has the mathematical form as follows:

$$\begin{aligned}
&\ln\left[f(x|\mu,\sigma^2)\right] = \textstyle\sum_{(i,j)\in R}\left(\ln\left[\frac{1}{\sqrt{2\pi\sigma^2}}\right] - \frac{\left(x_{i,j}-\mu\right)^2}{2\sigma^2}\right)\\
&0 \le x_{i,j} \le 255\\
&\quad = \textstyle\sum_{(i,j)\in R}\left(\ln 1 - \frac{1}{2}\ln 2\pi\sigma^2\right) - \frac{1}{2\sigma^2}\sum_{(i,j)\in R}\left(x_{i,j}-\mu\right)^2\\
&\quad = -\tfrac{R}{2}\ln\sigma^2 - \tfrac{R}{2}\ln 2\pi - \tfrac{R}{2}\\
&\quad = -\tfrac{R}{2}\ln\sigma^2 + const. \quad \text{where } R = 16\times 16.
\end{aligned} \tag{7}$$

$a(N)M(k)$ and $b(k, n)$ in (6) decide which selection criteria we are going to use. In this paper, we integrate the two parameters, shown in (8), into (6), then the selection criteria becomes to Won's modified Akaike information criterion [17] as follows:

$$\begin{cases} a(N)M(k) & = N^c \times M(k)\\ b(k,n) & = 0 \end{cases} \tag{8}$$

TABLE II
PERCENTAGE OF P8 × 8 MODE IN VARIOUS SEQUENCES AND QPS

Sequence	QP	SKIP (%)	16 × 16 (%)	16 × 8 (%)	8 × 16 (%)	P8 × 8 (%)	Intra (%)
Miss-America (QCIF)	20	39.43	28.46	8.95	8.62	**9.58**	4.92
	24	60.30	18.66	6.70	6.40	**5.55**	2.35
	28	70.80	15.89	4.77	4.84	**2.51**	1.17
	32	79.04	13.06	3.19	3.36	**0.86**	0.46
Foreman (CIF)	20	10.48	28.81	12.08	12.25	**31.59**	4.77
	24	21.0	29.09	12.11	12.70	**20.36**	4.66
	28	19.55	36.56	11.41	12.43	**10.54**	9.50
	32	32.99	33.68	8.99	9.88	**4.67**	9.79
News (CIF)	20	64.07	13.51	4.58	5.15	**11.48**	1.19
	24	70.28	10.91	4.23	4.97	**8.39**	1.18
	28	75.50	9.27	3.79	4.44	**5.75**	1.21
	32	79.75	8.59	3.127	3.71	**3.50**	1.30
Mother & Daughter (CIF)	20	49.82	18.59	8.37	9.12	**12.92**	1.15
	24	59.75	17.26	7.69	8.16	**6.46**	0.66
	28	67.81	17.24	5.62	6.14	**2.63**	0.55
	32	75.85	15.25	3.45	3.99	**0.85**	0.61

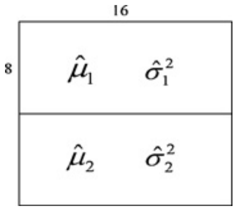


Fig. 4. Example for $M(k)$, k = Inter16 × 8 calculation.

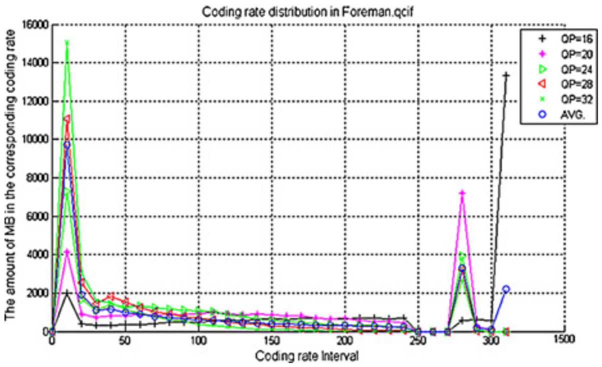
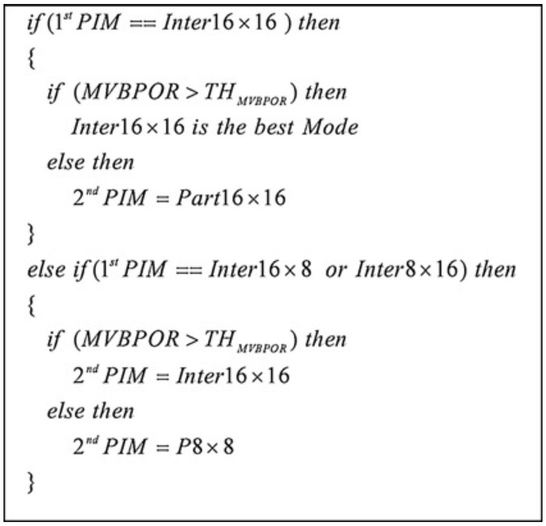


Fig. 5. BR versus MB order for *Foreman*.QCIF.



```
if (1st PIM == Inter16×16 ) then
{
  if (MVBPOR > TH_MVBPOR) then
    Inter16×16 is the best Mode
  else then
    2nd PIM = Part16×16
}
else if (1st PIM == Inter16×8 or Inter8×16) then
{
  if (MVBPOR > TH_MVBPOR) then
    2nd PIM = Inter16×16
  else then
    2nd PIM = P8×8
}
```

Fig. 6. Pseudocode for the second prediction stage.

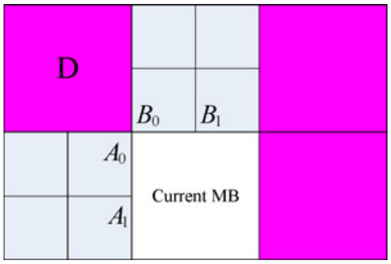


Fig. 7. Neighborhood information of the current MB.

where N is the sample number, c is a prespecified constant that is set to 0.6, and $M(k)$ represents the number of free parameters (i.e., mean and variance) in the specified criteria. In our application, $M(k)$ has only two possible values, 2 or 4, depending on the selected feature k. We illustrated the value $M(k = Inter16 \times 8)$ case in Fig. 4. There are four free parameters, $\hat{\mu}_1$, $\hat{\sigma}_1^2$, $\hat{\mu}_2$, and $\hat{\sigma}_2^2$, as a result $M(k = Inter16 \times 8) = 4$.

Summarize (6)–(8). The likelihood function is

$$G(k) = -\frac{R}{2} \ln \sigma^2 - N^{0.6} \times M(k) + const. \tag{9}$$

To obtain the appropriate mode for the current MB, the texture (k) that can maximize the likelihood equation, G(k), is estimated. The sample mean and variance, shown in (4), can be reused to solve it as follows:

$$\hat{k} = \underset{k \in 16\times8, 8\times16}{\arg\max} \; G(k). \tag{10}$$

However, MSC is not precise enough since there are only 64 sample data points in the sub-MB region (8 × 8) [25]. Table II shows the percentages of P8 × 8 in *Miss-America* (QCIF), *Foreman* (CIF), *News* (CIF), and *Mother & Daughter* (CIF).

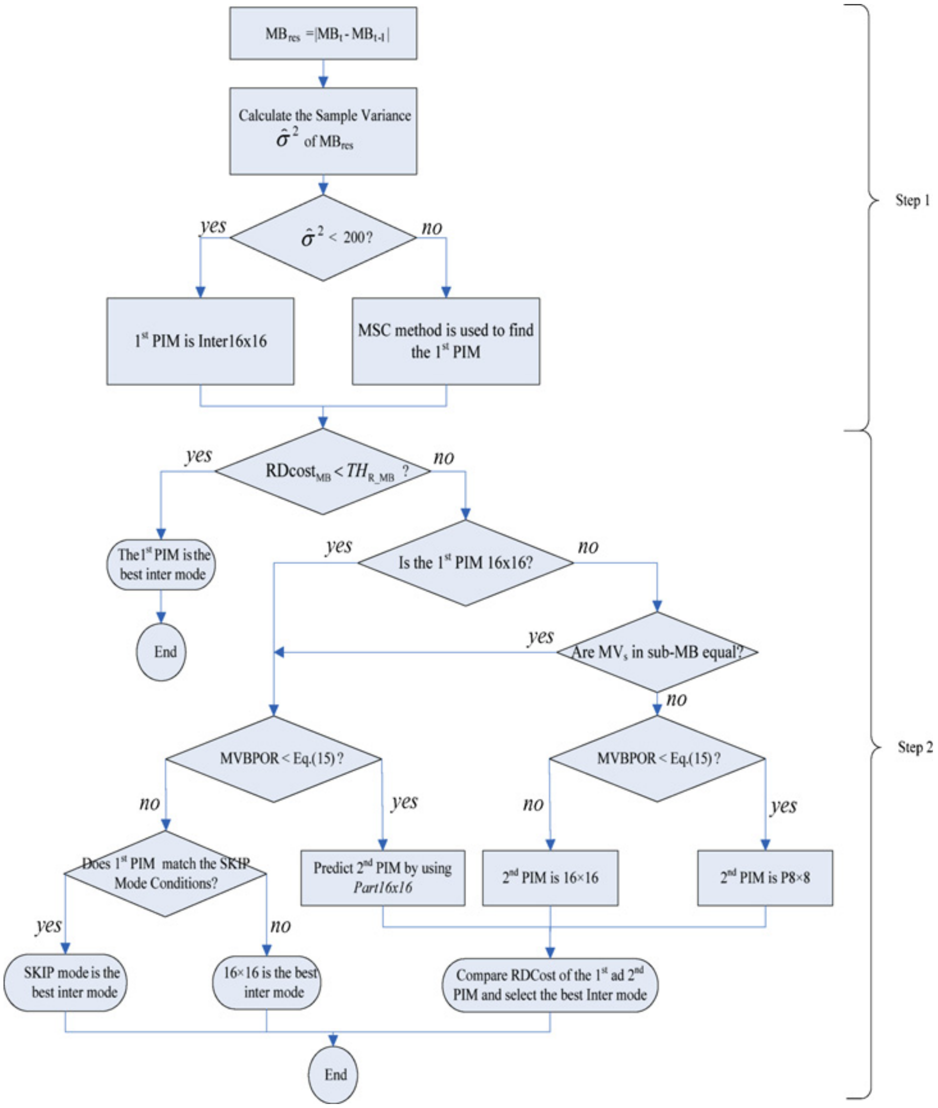

MBres =|MBt - MBt-1|
Calculate the Sample Variance σ̂² of MBres
σ̂² < 200?
yes
no
1st PIM is Inter16x16
MSC method is used to find the 1st PIM
RDcostMB < THR_MB ?
yes
no
The 1st PIM is the best inter mode
End
Is the 1st PIM 16x16?
yes
no
Are MVs in sub-MB equal?
yes
no
MVBPOR < Eq.(15)?
no
yes
MVBPOR < Eq.(15)?
no
yes
Does 1st PIM match the SKIP Mode Conditions?
Predict 2nd PIM by using Part16x16
2nd PIM is 16×16
2nd PIM is P8×8
yes
no
SKIP mode is the best inter mode
16×16 is the best inter mode
Compare RDCost of the 1st ad 2nd PIM and select the best Inter mode
End
Step 1
Step 2

Fig. 8. Overall scheme of the proposed algorithm.

We observed that the P8 × 8 mode makes up a small portion all the modes; therefore, we adopt the brute force search method instead of MSC to maintain video quality, if necessary.

III. Proposed Algorithm: Temporal-Based

A. *Temporal-Based Mode Prediction Scheme*

In Section II, we predicted an inter mode based on spatial feature extraction, i.e., the Step 1 (shown in Fig. 2), with negligible peak signal-to-noise ratio (PSNR) degradation. In [26], using the Step 1 alone can only keep the performance on PSNR but result in a tremendous increase in the coding bit rate (BR) compared to that of JM reference code. This indicates that performing inter mode selection using spatial information can be further improved. As a result, temporal features are considered as feedback after Step 1. We introduce the rate feedback mechanism to determine whether Step 1 results in unexpectedly high RDcost. If it does, then the following section is taken into account.

1) *Rate Feedback Scheme for First PIM:* After performing ME and RDcost computation of the first PIM of an MB from Section II-B, we apply the criterion to test whether the first PIM is appropriate for the current MB as follows:

$$\text{The Best Mode} = \text{1stPIM} \quad \in \quad \text{RDcost}_{\text{MB}} < TH_{\text{R\_MB}}. \tag{11}$$

If (11) is satisfied, the first PIM becomes the final coding mode; otherwise, the second prediction mechanism for the current MB is required.

It is apparent to set the condition (11), because when the first PIM is the best mode, the coding BR should be reasonable, or it will result an irrational effect on BR. Therefore, a proper threshold must be obtained to verify the correctness of the coding rate. From the empirical analysis, we observed that when the smaller threshold (TH_{R_MB}) is used, the less BR increasing of the encoding performance. Accordingly, we set $TH_{R\_MB} = 11$ to decide the first possible coding mode. Fig. 5 shows the statistics results of the coding BRs for the best coding mode versus various QP values, where the *x*-

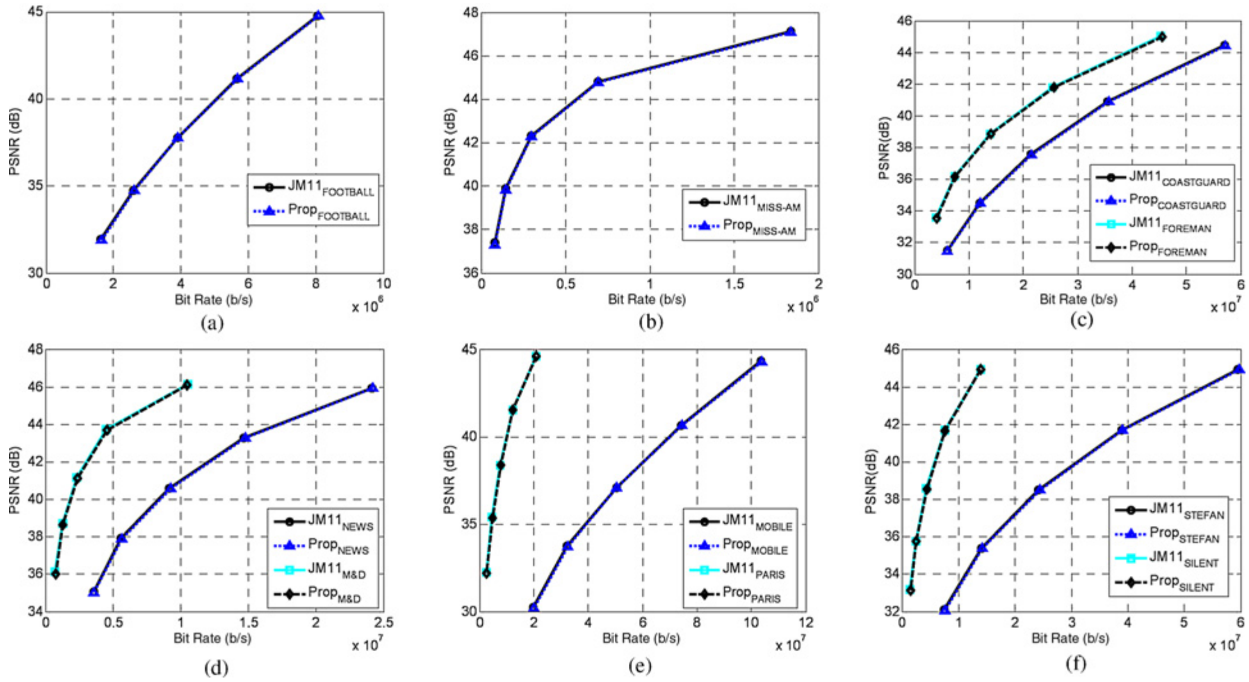


Fig. 9. R-D curves for JM11 and the proposed method. (a) QCIF Sequence: *Foot*. (b) QCIF sequence: *Miss-America*. (c) CIF sequence: *Coastguard* and *Foreman*. (d) CIF sequence: *News* and *Mother & Daughter*. (e) CIF sequence: *Mobile* and *Paris*. (f) CIF sequence: *Stefan* and *Silent*.

axis stands for the best mode coding rate interval from 0 to 1500, and the y-axis stands for the amount of MB falling in the corresponding rate interval. Although the rate falling in 270–280 takes the second place, it does not affect the whole encoding performance when we set $TH_{R\_MB} = 11$. It is because that our purpose is aim to save the encoding time in this step, and the TH_{R_MB} can help us to filter most part of MB that do not need the second step process.

2) *Temporal Feature Extraction:* In H.264/AVC, the total number of coding bits for the current MB is determined by variable length coding (VLC). The motion characteristic of the current MB can be obtained from the coded bits. In order to choose a candidate mode, we analyze the rate information (R_{MB}). The R_{MB} comprises parameters [27] as follows:

$$R_{MB} = R_{header} + R_{res} + R_{mv} \tag{12}$$

where R_{header} represents the coding bits for the header information, R_{res} represents the required bits for the predictive residue which is derived from integer DCT, quantization, VLC (the value depend on the QP value), and R_{mv} reveals the MV information for the current MB that is estimated by look-up table shown in Table III, where ΔMV [28] is given as follows:

$$\Delta MV = \left| MV_{predicted} - MV_{current} \right| . \tag{13}$$

Consequently, we analyze the motion density of the MB from R_{MB} using the proposed method, which is the MV bit percentage of the rate term (*MVBPOR*). We can use *MVBPOR* to determine whether the PIM selected by MSC is suitable. MVBPOR is defined as follows:

$$MVBPOR = \frac{R_{mv}}{R_{MB}}. \tag{14}$$

TABLE III
CODING BITS OF RMV BY LOOK-UP TABLE

|ΔMV|	RMV (bits)
0	1
1	3
2, 3	5
4, 5, 6, 7	7
8, 9, …, 15	9
16, 17, …, 31	11
32, 33, …, 63	13
64, 65, …, 127 128, 129, …, 255	15

MVBPOR can be viewed as an encoding BR weight. It indicates the BR percentage between the current encoding MV and the total encoding rate. When *MVBPOR* is larger than the specified threshold (TH_{MVBPOR}), the MV information takes up the major portion of the total coding bits. This means that the first PIM takes a significant number of bits on the current MB. In order to reduce the BR, the second PIM should be *Inter16×16*, *Inter*16 × 8, or *Inter*8 × 16 depending on the first PIM. It is because the first PIM is expected to induce a small RDCost performance if it is well predicted. Accordingly, when *MVBPOR* is larger than TH_{MVBPOR}, the encoded MB may belong to a moving block, since the first PIM takes lots portion of rate on MV bits. We then select the mode that needs less MV bits, such as 16 × 16, 16 × 8, or 8 × 16 but not P8 × 8, as the second mode to get another RDCost, which may be smaller in contrast to the first PIM.

On the contrary, when *MVBPOR* is smaller than TH_{MVBPOR}, it means that the first PIM spends fewer portions of rate on MV bits but more on residual frame. Therefore, we select the

TABLE IV

CONFIGURATION SETTINGS OF THE SIMULATION ENVIRONMENT

Input Benchmarks Sequences	QCIF: *Football*, *Miss-America*
	CIF: *Coastguard*, *Foreman*, *Mother & Daughter*, *Mobile*, *News*, *Paris*, *Silent*, *Stefan*
Profile IDC	Baseline
Frame rate	30 Hz
Intra period	0 (IPPP)
No. of B frames	0
Frames to be encoded	300
QP	16, 20, 24, 28, 32
Hadamard transform	Enabled
RDO	Enabled
Motion search range	16
ME	Full search
Symbol mode	CAVLC

mode that needs to be encoded with more MV bits, such as $P8 \times 8$ or modes from the $Part16 \times 16$ method, as the second mode to check if we can get a smaller RDCost from the second PIM with more detail motion information.

In general, TH_{MVBPOR} can be defined as one third, since R_{mv} only takes one third of R_{MB}. Moreover, in order to accelerate the performance in high QPs, we suggest another threshold. In some general cases, when QP is 16, 20, 24, 28, and 32, we found that a smaller QP is always biased toward higher video quality and results in a larger R_{res} term than R_{mv}. To avoid the influence of QP on *MVBPOR*, we add a QP factor to the TH_{MVBPOR} threshold design. Then, we can adaptively renew TH_{MVBPOR} to QP using the equation as follows:

$$TH_{MVBPOR} = \frac{0.33}{\lambda_{mode}}. \tag{15}$$

3) *Second Inter Mode Prediction Stage:* Three processes are used for the second prediction stage (pseudocode in Fig. 6), which are sub-MB ME ($P8 \times 8$), $Part16 \times 16$, and $Inter16 \times 16$, respectively. $P8 \times 8$ and $Inter16 \times 16$ are the original inter modes in H.264/AVC. Therefore, we only explain the procedure of $Part16 \times 16$. We generate the second PIM by using the neighboring MBs motion information.

Fig. 7 shows all the possible neighborhood information of the current MB. The possible cases are

$$\begin{cases} \text{case A-1} : MV_{\mathrm{A0}} = MV_{A1} \& MV_{\mathrm{B0}}! = MV_{B1} \\ \text{case A-2} : MV_{\mathrm{A0}}! = MV_{A1} \& MV_{\mathrm{B0}} = MV_{B1} \\ \text{case A-3} : MV_{\mathrm{A0}}! = MV_{A1} \& MV_{\mathrm{B0}}! = MV_{B1}. \end{cases} \tag{16}$$

In case A-1, block A can be $Inter8 \times 16$ or $Inter16 \times 16$ and block B can be $Inter16 \times 8$ or $P8 \times 8$. Thus, the current MB is partitioned into Inter8 × 16. In case A-2, block A can be $Inter16 \times 8$ or $P8 \times 8$ and block B can be Inter16 × 8 or Inter16 × 16. We can partition the current MB into Inter16×8. In case A-3, block A can be $Inter16 \times 8$ or $P8 \times 8$ and block B can be $Inter8 \times 16$ or $Inter16 \times 16$. Thus, we need to use $P8 \times 8$ mode to encode the current MB.

B. Overall Proposed Algorithm

The overall proposed algorithm flow is shown as Fig. 8.

TABLE V

COMPARISON RESULTS OF PROPOSED AND PREVIOUS WORKS

Sequence	Method	ΔPSNR$_\mathrm{Y}$ (dB)	ΔBR (%)	ΔT (%)
Miss-American (QCIF)	**Proposed**	**−0.06**	**1.00**	**−61.13**
	Grecos	−0.10	0.14	−58.95
	Wu	−0.13	0.34	−47.94
Football (QCIF)	**Proposed**	**−0.00**	**0.50**	**−51.52**
	Grecos	−0.14	3.65	−37.91
	Wu	0	0.61	−36.81
Coastguard (CIF)	**Proposed**	**−0.03**	**0.64**	**−51.82**
	Grecos	−0.13	0.90	−37.02
	Ri	−0.1	1.8	−39.4
	Chen	−0.14	2.3	−19.3
Foreman (CIF)	**Proposed**	**−0.03**	**1.00**	**−63.01**
	Grecos	−0.13	1.28	−39.85
	Wu	−0.02	0.1	−24.08
	Ri	−0.05	2.1	−40.9
	Chen	−0.13	3	−22
Mother & Daughter (CIF)	**Proposed**	**−0.06**	**1.27**	**−58.37**
	Grecos	−0.09	1.08	−56.30
	Ri	−0.03	1.6	−53.2
	Chen	−0.12	4.5	−23.6
Mobile (CIF)	**Proposed**	**−0.02**	**0.44**	**−53.96**
	Grecos	−0.15	1.04	−33.70
	Wu	−0.01	0.13	−9.97
	Ri	−0.01	2.9	−38.6
News (CIF)	**Proposed**	**−0.03**	**0.55**	**−47.69**
	Wu	−0.06	1.18	−42.62
Paris (CIF)	**Proposed**	**−0.02**	**0.92**	**−66.96**
	Grecos	−0.15	5.01	−62.71
	Wu	−0.06	1.28	−25.18
Silent (CIF)	**Proposed**	**−0.02**	**1.45**	**−66.30**
	Grecos	−0.08	3.77	−65.91
	Wu	−0.03	0.95	−42.28
	Chen	−0.12	2.8	−24
Stefan (CIF)	**Proposed**	**−0.03**	**0.94**	**−63.16**
	Grecos	−0.16	1.36	−35.67
	Ri	−0.02	2.5	−41.0
	Chen	−0.07	2.3	−19.2
Average	**Proposed**	**−0.03**	**0.87**	**−58.39**
	Grecos	−0.12	1.95	−47.55
	Wu	−0.04	0.65	−32.69
	Ri	−0.04	2.18	−42.62
	Chen	−0.11	2.98	−21.62

IV. EXPERIMENTAL RESULTS

The proposed fast inter mode decision algorithm was implemented and integrated with the JVT reference software JM11.0 [29]. The simulations were run on an Intel 3.40 GHz Pentium D processor with 1 GB DDR random access memory. The operating system was Windows XP SP2. The configuration settings of the simulation environment are shown in Table IV.

The benchmarks sequences have various motion characteristics. For example, *Silent*, *Mother & Daughter*, *Miss-American*, and *Coastguard* have a lot of static scenes and few moving objects. *Foreman*, *Football*, and *Stefan* have high-motion objects and many scene changes. The remaining sequences, *Mobile*, *News*, and *Paris*, contain either low-motion or high-motion in many short scenes, which increase the coding challenge.

The performance of the proposed method was measured using the average PSNR difference of luma (ΔPSNR$_\mathrm{Y}$), the increase of BR (ΔBR), and the decrease of total encoding time (ΔT). The results are shown in (17)–(19), respectively, as follows:

$$\Delta\mathrm{PSNR_Y(dB)} = PSNR_{Y\_prop} - PSNR_{Y\_JM} \tag{17}$$

$$\Delta\mathrm{BR}(\%) = \frac{BR_{prop} - BR_{JM}}{BR_{JM}} \times 100\% \tag{18}$$

$$\Delta\mathrm{T}(\%) = \frac{T_{prop} - T_{JM}}{T_{JM}} \times 100\%. \tag{19}$$

Table V shows the results compared with several existing algorithms, such as those of Grecos [7], Wu [10], Ri [30], and Chen [31]. Each method has its own limitations. For example, Wu's method performs good R-D on static sequences but its computation reduction is poor. Grecos's and Ri's methods have good computation reduction at the cost of R-D performance. Chen's method performs well only on some sequences. The proposed algorithm achieves a 58% total encoding time reduction as well as appropriate R-D performance with only a 0.03 PSNR decrease and a 0.87% BR increase on average.

The R-D performance of the proposed method is shown in the form of R-D curves in Fig. 9. The proposed approach is close to the JM standard, with less PSNR loss and BR increase.

The above experimental results indicate an efficient algorithm design should consider both computation complexity reduction and coding performance degradation. The proposed scheme based on spatial and temporal analysis was efficient for all video sequences; it adaptively adjusts the computation complexity with QP.

V. Conclusion

In this paper, we proposed an efficient inter mode decision method based on spatial-temporal analysis to reduce R-D computation. The proposed method reduced computation complexity about to 58.39% while maintaining coding performance with negligible PSNR 0.03 loss and a 0.87% increase in BR on average. The proposed algorithm can adaptively adjust the computation complexity according to QP and maintains R-D performance. In addition, to implement a ME processor in very large scale integrated (VLSI) hardware design, our proposed method has the extreme regularity on the VLSI design to be an accelerator for the processor.

References

[1] T. Wiegand, G. J. Sullivan, G. Bjontegaard, and A. Luthra, "Overview of the H.264/AVC video coding standard," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 13, no. 7, pp. 560–576, Jul. 2003.

[2] *Draft ITU-T Recommendation and Final Draft International Standard of Joint Video Specification*, document ITU-T Rec. H.264/ISO/IEC 14496-10 AVC, 2003.

[3] J. Ostermann, J. Bormans, P. List, D. Marpe, M. Narroschke, F. Pereiera, T. Stockhammer, and T. Wedi, "Video coding with H.264/AVC: Tools, performance, and complexity," *IEEE Circuits Syst. Mag.*, vol. 4, no. 1, pp. 7–28, Apr. 2004.

[4] B. Jeon and J. Lee, *Fast Mode Decision for H.264*, document JVT-J033, ITU-T Q.6/16, 2003.

[5] T. Wiegand, *Joint Model Number 1 (JM-1)*, document JVT-A003, Joint Video Team (JVT) Meeting, Pattaya, Thailand, Dec. 4–6, 2001.

[6] I. Choi, J. Lee, and B. Jeon, "Fast coding mode selection with rate-distortion optimization for MPEG-4 part-10 AVC/H.264," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 16, no. 12, pp. 1557–1561, Dec. 2006.

[7] C. Grecos and M. Y. Yang, "Fast mode prediction for the baseline and main profiles in the H.264 video coding standard," *IEEE Trans. Multimedia*, vol. 8, no. 6, pp. 1125–1134, Dec. 2006.

[8] A. Saha, K. Mallick, J. Mukherjee, and S. Sural, "SKIP prediction for fast rate distortion optimization in H.264," *IEEE Trans. Consumer Electron.*, vol. 53, no. 3, pp. 1153–1160, Aug. 2007.

[9] B. Feng, G. Zhu, and W. Liu, "Complexity scalable inter modes decision algorithm for H.264 based on spatial correlation," in *Proc. Adv. Commun. Technol., ICACT*, vol. 2. 2006, pp. 963–966.

[10] D. Wu, F. Pan, K. P. Lim, S. Wu, Z. G. Li, X. Lin, S. Rahardja, and C. C. Ko, "Fast intermode decision in H.264/AVC video coding," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 15, no. 6, pp. 953–958, Jul. 2005.

[11] B.-G. Kim, "Novel inter-mode decision algorithm based on macroblock (MB) tracking for the P-slice in H.264/AVC video coding," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 18, no. 2, pp. 273–279, Feb. 2008.

[12] T. Y. Kuo and C. H. Chan, "Fast variable block size motion estimation for H.264 using likelihood and correlation of motion field," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 16, no. 10, pp. 1185–1195, Oct. 2006.

[13] L. Shen, Z. Liu, Z. Zhang, and X. Shi, "Fast inter mode decision using spatial property of motion field," *IEEE Trans. Multimedia*, vol. 10, no. 6, pp. 1208–1214, Oct. 2008.

[14] P. Yin, H. Y. Tourapis, A. M. Tourapis, and J. Boyce, "Fast mode decision and motion estimation for JVT/H.264," in *Proc. Image Process., ICIP*, vol. 3. 2003, pp. 853–856.

[15] A. Chia, W. Yu, G. R. Martin, and H. Park, "Fast inter-mode selection in the H.264/AVC standard using a hierarchical decision process," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 18, no. 2, pp. 186–195, Feb. 2008.

[16] C. S. Won, "Variable block size segmentation for image compression using stochastic models," in *Proc. Image Process.* vol. 3. 1996, pp. 975–978.

[17] C. S. Won and H. Derin, "Unsupervised segmentation of noisy and textured images using Markov random fields," in *Proc. CVGIP: Graphic. Models Image Process.*, vol. 54. 1992, pp. 303–328.

[18] R. Szeliski, R. Zabih, D. Scharstein, O. Veksler, V. Kolmogorov, A. Agarwala, M. Tappen, and C. Rother, "A comparative study of energy minimization methods for Markov random fields with smoothness-based priors," *IEEE Trans. Patt. Anal. Mach. Intell.*, vol. 30, no. 6, pp. 1068–1080, Jun. 2008.

[19] O. Chum and J. Matas, "Optimal randomized RANSAC," *IEEE Trans. Patt. Anal. Mach. Intell.*, vol. 30, no. 5, pp. 1472–1482, Aug. 2008.

[20] J. Park and L. Kurz, "Image enhancement using the modified ICM method," *IEEE Trans. Image Process.*, vol. 5, no. 5, pp. 765–771, May 1996.

[21] A. Dogandzic and B. Zhang, "Complex amplitude estimation and adaptive matched filter detection in low-rank interference," *IEEE Trans. Signal Process.*, vol. 55, no. 3, pp. 1176–1182, Mar. 2007.

[22] Y. Tsaig and A. Averbuch, "Automatic segmentation of moving objects in video sequences: A region labeling approach," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 12, no. 7, pp. 597–612, Jul. 2002.

[23] S. Z. Li, *Markov Random Field Modeling in Image Analysis*. New York: Springer-Verlag, 2001.

[24] S. L. Scolve, "Application of model-selection criteria to some problems in multivariate analysis," *Psychometrika*, vol. 52, no. 3, pp. 333–343, Sep. 1987.

[25] F. D. Ridder, R. Pintelon, J. Schoukens, and D. P. Gillikin, "Modified AIC and MDL model selection criteria for short data records," *IEEE Trans. Instrument. Measurement*, vol. 54, no. 1, pp. 144–150, Feb. 2005.

[26] K. I. Lee, A. C. Tsai, and J. F. Wang, "Fast inter mode decision strategy based on stochastic models selection for H.264/AVC," in *Proc. IEEE TENCON*, Oct.–Nov. 2007, pp. 1–4.

[27] H. Sun, X. Chen, and T. Chiang, *Digital Video Transcoding for Transmission and Storage*. Boca Raton, FL: CRC Press, 2005.

[28] I. E. G. Richardso, *H.264 and MPEG-4 Video Compression: Video Coding for Next-Generation Multimedia*. New York: Wiley, 2003, pp. 67–214.

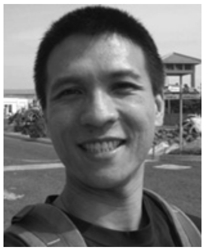
[29] Joint Video Term (JVT). *Reference Software JM 11.0* [Online]. Available: http://iphome.hhi.de/suehring/tml

[30] S.-H. Ri, Y. Vatis, and J. Ostermann, "Fast inter-mode decision in an H.264/AVC encoder using mode and Lagrangian cost correlation," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 19, no. 2, pp. 302–306, Feb. 2009.

[31] G. Chen, Y.-D. Zhang, S.-X. Lin, and F. Dai, "Efficient block size selection for MPEG-2 to H.264 transcoding," in *Proc. ACM Multimedia*, Oct. 2004, pp. 300–303.

Kuan-I Lee was born in Tainan, Taiwan, in 1983. He received the B.S. degree in mechanical engineering from National Central University, Taoyuan, Taiwan, in 2004, and the M.S. degree in electrical engineering from National Cheng Kung University, Tainan, in 2006.

He is currently a Staff Engineer with Realtek Semiconductor Corporation, Hsinchu, Taiwan. His current research interests include multimedia processing, wireless communication, and digital and analog circuit design.



An-Chao Tsai was born in Changhua, Taiwan, in 1980. He received the M.S. degree in electronic engineering from Da-Yeh University, Changhua, in 2004. He is currently pursuing the Ph.D. degree in electronic engineering from National Cheng Kung University, Tainan, Taiwan.

His current research interests include video coding technology, intra prediction, inter prediction, scalable video coding, interpolation technique, and very large scale integration architectural design.

Jhing-Fa Wang (S'82–M'83–SM'88–F'99) received the B.S. and M.S. degrees from the Department of Electrical Engineering, National Cheng Kung University (NCKU), Tainan, Taiwan, in 1979 and 1973, respectively, and the Ph.D. degree from the Department of Computer Science and Electrical Engineering, Stevens Institute of Technology, Hoboken, NJ, in 1983.

He is currently a Chair Professor with NCKU. He developed a Mandarin speech recognition system called Venus-Dictate, known as a pioneering system in Taiwan. He is currently leading a research group of different disciplines for the development of "advanced ubiquitous media for created cyberspace." He has published about 91 journal papers and 217 conference papers, and has obtained five patents since 1983. His current research interests include wireless content-based media processing, image processing, speech recognition, and natural language understanding.

Dr. Wang was the recipient of outstanding awards from the Institute of Information Industry in 1991 and the National Science Council of Taiwan in 1990, 1995, and 1997, respectively. He was an Associate Editor for the IEEE TRANSACTIONS ON NEURAL NETWORKS and the IEEE TRANSACTIONS ON VERY LARGE SCALE INTEGRATION SYSTEMS. He was invited to give the keynote speech at the 12th Pacific Asia Conference on Language, Information and Computation, Singapore. He has served as the General Chairman of the International Symposium on Communication, Taiwan, in 2001.

Jar-Ferr Yang (S'84–M'88–SM'98–F'07) was born in Keelung, Taiwan, on September 15, 1954. He received the B.S. degree from Chung Yuan Christian University, Taoyuan, Taiwan, in 1977, the M.S. degree from National Taiwan University, Taipei, Taiwan, in 1979, and the Ph.D. degree from the University of Minnesota, Minneapolis, in 1988, all in electrical engineering.

From 1979 to 1980, he was an Instructor with the Chinese Naval Engineering School, Kaoshiung, Taiwan, for his Navy Reserve Officers Training Corps Service. From 1981 to 1984, he was an Assistant Researcher with the Data Transmission and Network Design Research Group, Telecommunication Laboratories, Chung-Li, Taiwan. From 1982 to 1984, he was also an Adjunct Lecturer with Chung Yuan Christian University. In 1988, he joined the National Cheng Kung University, Tainan, Taiwan, as an Associate Professor and was promoted to Full Professor and Distinguished Professor in 1994 and 2004, respectively. He was the Chairman of the Center for Computer and Communication Research, National Cheng Kung University, from 1997 to 2000. In 2002, he was a Visiting Scholar with the Department of Electrical Engineering, University of Washington in Seattle, Seattle. From 2004 to 2008, he was the Chairperson of Graduate Institute of Computer and Communication Engineering and the Director of the Electrical and Information Technology Center. Currently, he is the Director of Technologies of Ubiquitous Computing and Humanity Center, National Cheng Kung University, supported by National Science Council (NSC), Taiwan. He has published over 92 journals and 141 conference papers. His current research interests include video, audio, and speech signal processing and coding, and living technology system designs and integrations. He also contributed to fast algorithms and efficient realization of video and audio coding.

Dr. Yang received the Government Study Abroad Scholarship from 1984 to 1988, which supported his advanced study at the University of Minnesota. In 2008, he received the NSC Excellent Research Award. From 2004 to 2005, he was a speaker in the Distinguished Lecturer Program by the IEEE Circuits and Systems Society. He was the Technical Program Co-Chair of the IEEE Asia Pacific Conference on Circuits and Systems in 2004 and the ninth IEEE International Workshop on Cellular Neural Networks and Their Applications in 2005. From 2004 to 2006, he was the Chair of the IEEE Signal Processing Society, Tainan Chapter, and Treasurer of the IEEE Tainan Section. From 2006 to 2007, he joined the Chair Committee of the IEEE Signal Processing Society. From 2006 to 2007, he was the Secretary of the IEEE Multimedia Systems and Applications Technical Committee (MSA TC), IEEE Circuits and Systems Society. From 2008 to 2009, he became the Chair of the MSA TC. Currently, he is an Associate Editor of the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY and the IEEE CIRCUITS AND DEVICES MAGAZINE. He is an Associate Editor of the *EURASIP Journal of Advances in Signal Processing* and was a Guest Editor of the Special Issue on "Advanced Video Technologies and Applications for H.264/AVC and Beyond." From 2007 to 2010, he was an Editorial Board Member of IET Signal Processing.