An Efficient Inter-Prediction Mode Decision Method for Temporal Scalability Coding With Hierarchical B-Picture Structure

Bumshik Lee, *Member, IEEE*, and Munchurl Kim, *Member, IEEE*

***Abstract*—Although the temporal scalability coding supported by both H.264/AVC and H.264/SVC provides a good means of adapting bitstream scalability in temporal dimension, it entails heavy computational complexity due to bidirectional motion estimation and compensation for B-pictures in temporal scalability levels. To make it more applicable, this paper introduces a fast inter-prediction mode decision method which is efficiently performed using statistical hypothesis testing. The hypothesis testing is performed on mean and variance of pixels in 16 × 16 and 8 × 8 blocks to decide early termination against further processing for their sub-blocks. Fallback check is employed to compromise the tradeoff between the compression efficiency and encoding time saving. The proposed method exhibits effective performance for early termination in the inter-prediction mode decision, thus leading to a significant reduction of the total encoding time up to average 64.8% with small increments in bit amounts and small degradation in visual quality.**

***Index Terms*—Fast mode decision, H.264/AVC, scalable video coding, statistical hypothesis testing, temporal scalability.**

I. Introduction

THE hierarchical B-picture structure in both H.264/AVC and H.264/SVC [1] entails all B-pictures within two consecutive key pictures to be decomposed into different temporal scalability levels in order to achieve various temporal resolutions. Also, the rate distortion (RD) optimized inter-prediction mode decision in the hierarchical B-picture structure causes large computational complexity during the encoding process. Therefore, this paper focuses on reduction of the computational complexity for temporal scalability by performing fast inter-prediction mode decision in the hierarchical B-picture coding. Although much research has been conducted to reduce encoding complexity for the conventional group of picture (GOP) structures such as IPPP or IBBP in H.264/AVC [2]–[4] and some effort has been made for H.264/SVC [5], [6], few efforts of complexity reduction have been made for temporal scalability coding in the hierarchical B-picture structure. Liu *et al.* [2] proposed a complexity reduction algorithm based on macroblock (MB) motion homogeneity. The MBs are classified into four motion activity classes according to the predefined threshold values for efficient mode decision. In [3], motion continuity is utilized for efficient mode decision, which is decided based on the edge map detected by a Sobel operator. Yu *et al.* [4] proposed an early termination method of mode selection using a hierarchical decision process using the pixel variance or SAD values. A few studies on complexity reduction of H.264/SVC encoders have been carried out. Li *et al.* [5] proposed a method of reducing the encoding complexity of H.264/SVC with full scalabilities. Lee *et al.* [6] proposed a low complexity encoding scheme for spatial scalability. Although these methods are applicable to the inter-prediction mode decision or other scalabilities of the SVC encoder, they are not efficient for temporal scalability coding since they do not utilize the characteristics of the hierarchical B-picture structure. Since the frames in the hierarchical B-picture structure have higher dependency each other and different characteristics inside a GOP, the coding efficiency might be readily degraded due to the error propagation into the higher temporal scalability layer or encoding time saving is not significant unless the property of the hierarchical B-picture structure is fully deployed. The conventional methods have been developed without the consideration of such characteristics in temporal scalability.

In this paper, an efficient early termination method for inter-prediction mode decision in hierarchical-B picture coding for temporal scalability is introduced by comparing the pixel values of the current block to be coded with those of a motion compensated reference block using statistical hypothesis testing. The confidence intervals (CI) for the hypothesis testing are adaptively applied in different temporal scalability levels of the hierarchical B-picture structure. In order to compromise the speed-up of encoding time and the resulting degradation of visual quality, fallback check is employed for the decision results by the hypothesis testing.

This paper is organized as follows: Section II introduces the hierarchical B-picture structure; Section III briefly describes statistical hypothesis testing (SHT) and the proposed fast inter-prediction mode decision method based on SHT for hierarchical B-picture coding; the experimental results are presented in Section IV; and Section V concludes this paper.

II. Hierarchical B-Picture Structure

Fig. 1 shows an example of a hierarchical B-picture structure with four temporal scalability levels $(N+1)$, which constitutes one single GOP in size $8\,(=2^N, N=3)$. In Fig. 1, the pictures I and P are defined as key-pictures and constitute temporal scalability level 0 (TSL_0). The temporal scalability level 1 (TSL_1) consists of the B-pictures B_1. Accordingly, the temporal scalability levels 2 and 3 (TSL_2 and TSL_3) consist of the B-pictures B_2's and B_3's, respectively.

In the hierarchical B-picture coding structure, quantization parameters (QP) are differently applied according to the temporal scalability levels according to different picture importance. Gradually increased QP values are assigned toward

Manuscript received October 22, 2010; revised August 15, 2011; accepted December 20, 2011. Date of publication February 24, 2012; date of current version May 18, 2012. This work was supported by the IT R&D program of MKE/KEIT [10039199, A Study on Core Technologies of Perceptual Quality based Scalable 3D Video Codecs].

The authors are with the KAIST, Daejeon 305-701, Korea (e-mail: lbs@kaist.ac.kr; mkim@ee.kaist.ac.kr).

Digital Object Identifier 10.1109/TBC.2012.2184154

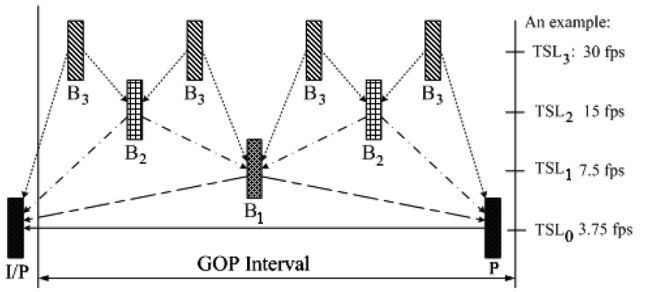

Fig. 1. Example of a hierarchical B-picture structure for temporal scalability (TSL: Temporal Scalability Level).

TABLE I
PROPORTIONS OF SELECTED BLOCK MODES IN DIFFERENT TEMPORAL SCALABILITY LEVELS (TSL'S): UNIT (%)

Seq.(CIF)	QP	TSL	DIRECT	16×16	16×8	8×16	P8×8
Mother& Daughter	28	1	52.9	22.4	8.1	7.2	9.3
		2	66.4	18.9	4.6	4.6	5.4
		3	85.7	8.9	1.7	2.1	1.5
Foreman	28	1	26.8	34.3	15.4	13.0	10.6
		2	40.6	34.7	10.7	8.8	5.2
		3	64.6	26.4	4.6	3.8	0.6
Harbour	28	1	24.8	25.5	6.0	16.6	27.1
		2	29.0	29.3	7.9	12.8	21.0
		3	41.1	31.7	7.0	12.3	7.9

higher TSLs [7]. The proportions of selected block modes are significantly different according to the TSLs. Table I shows the occurrence rates of the respective selected block modes according to temporal scalability levels.

As shown in Table I, the portions of the smaller block size, P8 × 8 become reduced going toward higher temporal scalability levels ($\mathrm{TSL}_2 \rightarrow \mathrm{TSL}_3$). This indicates that the frames in lower temporal scalability levels should maintain higher fidelity of image quality by selecting finer partitioned block modes, since they can be more frequently referred by B-pictures in higher temporal scalability levels. Motivated by this observation, a fast mode decision algorithm is proposed by taking into account the selected block mode distributions in different temporal scalability levels. Considering the proportions of the large block mode (16 × 16) in different TSL's, the CIs for hypothesis testing are adaptively set according to temporal scalability, which is described in the following section.

III. PROPOSED FAST INTER-PREDICTION MODE DECISION USING STATISTICAL HYPOTHESIS TESTING

In general, the means and variances of the pixels in the current block to be coded are often similar to those of the pixels in the motion-compensated reference block when the corresponding block partition is the final block mode. Therefore, we perform two statistical hypothesis tests on both mean and variance for two groups (16 × 16 or 8 × 8) of pixels from the current block and a motion-compensated reference block. If both tests are accepted, it is assumed that the two group pixels are very similar and thus the corresponding block partition (16 × 16 or 8 × 8) under test becomes the final block mode as either 16 × 16 or 8 × 8 mode, leading to early termination on the block mode under consideration without further processing on smaller partitioned block modes. We use Z-test statistic on variance and t-test statistic on mean for hypothesis testing, respectively. If we apply this scheme for 16 × 16 and 8 × 8 blocks in inter-prediction mode decision, then we can do early termination on the mode decision (16 × 16 or 8 × 8) when the two null hypotheses are accepted for both mean and variance tests. That is, no further examination for block mode decision is necessary for the block modes with smaller block sizes, thus resulting in the considerable computation savings. Let X_C and X_R be two independent normal distributions from the current and reference blocks such that

$$X_C \sim N\left(\mu_C, \sigma_C^2\right) \text{ and } X_R \sim N\left(\mu_R, \sigma_R^2\right) \tag{1}$$

where μ_C and μ_R indicate the true means of the pixels in the current and reference blocks, and are assumed to be known *a prior*. σ_C^2 and σ_R^2 are the true variances of the pixels in the current and reference blocks, respectively. We then define the following hypotheses to test the variances of pixel values between two blocks:

$$H_0 : \sigma_C^2 = \sigma_R^2 \text{ and } H_1 : \sigma_C^2 \neq \sigma_R^2 \tag{2}$$

where the null hypothesis H_0 indicates that two pixel groups originate from the same source. That is, if H_0 is accepted, then the two pixel groups in the current and reference blocks are considered to be from the same source and hence the block mode (16 × 16 or 8 × 8 mode) under test is determined as the final selected mode. The Z-test statistic for variance test is given by

$$Z_0 = \frac{S_C - S_R - (\sigma_C - \sigma_R)}{S_P \sqrt{\frac{1}{2n_C} + \frac{1}{2n_R}}} \tag{3}$$

where n_C and n_R indicate the numbers of the original and reference block pixels under test, respectively, which can be the size of a block, 16 × 16 or 8 × 8. S_C and S_R are the standard deviation estimates of pixels in the original and reference blocks, respectively [8]. S_p^2 is a pooled variance estimate of the common standard deviation and is defined as

$$S_P^2 = \frac{(n_C - 1)S_C^2 + (n_R - 1)S_R^2}{n_C + n_R - 2} \tag{4}$$

If the confidence interval is determined for a given level of significance α, we accept the null hypothesis H_0 if $-Z_{\alpha/2} \leq Z_0 \leq Z_{\alpha/2}$. Otherwise we reject H_0. Here $Z_{\alpha/2}$ indicates the critical point for acceptance region under null hypothesis for a given level of significance α.

For the variance test in the proposed fast inter-prediction mode decision method, we compensate the variance of the reference block pixels by adding the average of squared residual pixel values which is equal to the quantization error. Thus, more homogeneous blocks can be detected as the final block modes even in the low bitrate region where the blocks are encoded with higher QP values. The variance under test for the reference block is then compensated with a squared sum of residuals as

$$\tilde{S}_R^2 = S_R^2 + \frac{1}{n_R} \sum_{k=0}^{n_R - 1} |x(k) - \hat{x}_r(k)|^2 \tag{5}$$

where $x(k)$ and $\hat{x}_r(k)$ are the k-th pixel values of the original block and the reconstructed reference block, respectively. Therefore, S_R^2 is substituted with compensated variance estimate $\tilde{S}_R^2$ in (3). By doing so, the quantization effect on variance estimate can be compensated in the test statistic in (3).

Another hypothesis test for the similarity between two groups of pixels is the mean test. The mean of the original block pixels is often similar to that of the reconstructed block pixels for little or small block motion. In this case, large block modes are more likely to be selected. However, the two means may be different each other for the blocks with large or local non-rigid motion, which usually leads to smaller block partitions for block mode decision. Hence, in the proposed fast inter-prediction mode decision method, the homogeneous regions with no or small motion can be detected in an early stage by accepting the null hypothesis on mean test. In case of the mean test, we assume that the unknown variances σ_C^2 and σ_R^2 in the original and reference blocks may not be the same due to the quantization effect. We therefore use t-test statistic on mean test for two groups of block pixels and establish the following hypotheses:

$$H_0 : \mu_C = \mu_R \text{ and } H_1 : \mu_C \neq \mu_R \tag{6}$$

where μ_C and μ_R are the true means of the original block pixels and motion-compensated reference block pixels, respectively. The null hypothesis H_0 indicates that the two pixel groups originated from the same source and the alternative hypothesis H_1 assumes that they are from different sources. That is, if H_0 under mean test is accepted, the two pixel groups in the current and reference blocks are considered to be from the same source. Thus, for the both accepted mean and variance tests under null hypotheses, the block mode (16×16 or 8×8 mode) under consideration is determined as the final selected mode, thus leading to early termination without further processing for smaller partitioned block modes. To test the hypothesis on the two means in (6), we use the following t-test statistic, which approximately follows a t-distribution [9] as

$$t = \frac{\bar{X}_C - \bar{X}_R - (\mu_C - \mu_R)}{\sqrt{\frac{S_C^2}{n_C} + \frac{S_R^2}{n_R}}} \tag{7}$$

with degrees of freedom given by

$$\nu = \frac{\left(S_C^2/n_C + S_R^2/n_R\right)^2}{\left(S_C^2/n_C\right)^2/(n_C+1) + \left(S_R^2/n_R\right)^2/(n_R+1)} - 2 \tag{8}$$

where $\bar{X}_C$, $\bar{X}_R$, S_C^2 and S_R^2 are the sample means and variances of pixel intensities within the current and the reference blocks, respectively. Notice that S_R^2 in (7) and (8) is substituted with $\tilde{S}_R^2$ in (5) to take into account the quantization effect.

Fig. 2 shows the distributions of t- and Z-test statistics under the null hypotheses for *Forman* sequences with $\mathrm{QP} = 28$ and $\mathrm{GOP\ size} = 16$. The both distributions can be well approximated as normal distributions in each temporal scalability level.

The chance of large block modes to be selected may not increase towards higher TSL's, rather remains little changed in all TSL's with the same and fixed critical points $Z_{\alpha/2}$ and $t_{\alpha/2}$. This will cause the decrease in encoding time saving. In order

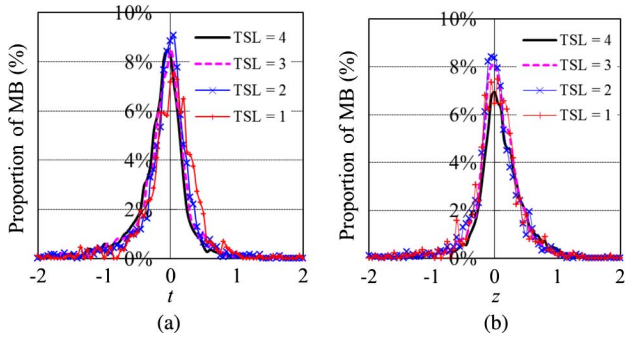

Fig. 2. t and Z distributions under the null hypothesis for the *Foreman* sequences: $\mathrm{QP} = 28$. (a) t distribution. (b) Z distribution.

TABLE II
LEVEL OF SIGNIFICANCE ACCORDING TO THE PROPORTIONS $(P_{LMB}, 0)$ OF THE LARGE BLOCK MODES IN TSL_0

$P_{LBM,0}$	[0.0 0.3)	[0.3 0.6)	[0.6 1.0]
$\alpha_0/2$	0.35	0.40	0.45

to take into account the increase in the proportion of large block modes in higher TSL's, the level of significance (α_l) for the critical points $Z_{\alpha/2}^{(l)}$ and $t_{\alpha/2}^{(l)}$ in TSL_l is set based on the level of significance (α_{l-1}) in TSL_{l-1} in addition with the proportional difference between the two portions of large block modes in TSL_{l-1} and TSL_0. Therefore, we let the level of significance α be adjusted by considering the proportions of large block modes (SKIP and 16×16 modes) selected in different TSL's and the QP values. In order to determine the confidence interval in TSL_l, the corresponding level of significance (α_l) is given by

$$\alpha_l = \alpha_{l-1} + w_{l,QP} \cdot (P_{LBM,0} - P_{LBM,l-1}), 0 \leq \alpha \leq 1, l \geq 2 \tag{9}$$

where $w_{l,QP}$ is a weight in TSL_l for a given QP value. $P_{LBM,l-1}$ is the proportion of the large block modes (SKIP and 16×16 modes) in TSL_{l-1}, and $P_{LBM,0}$ is the proportion of the large block modes in TSL_0 where all frames are key-pictures. Notice that the proposed fast inter-prediction mode decision method is not applied in TSL_0. As the proportion of the large block modes increases towards higher TSL's, the confidence intervals for Z- and t-test statistics also increase so that the null hypotheses on mean and variance are more frequently accepted. In (9), the level of significance α_1 in TSL_1 should be carefully determined to prevent from aggressive acceptance on null hypotheses which may otherwise cause large false alarms [9] in higher TSL's, thus resulting in more visual quality degradation in the reconstructed frames. We therefore set α_0 in a conservative manner with respect to the proportion of the large block modes in TSL_0 as shown in Table II. The weight value increases as the temporal scalability levels get higher so that larger block modes are favorably selected in the higher TSL's. Table III shows the weight values in different TSL's with their respective QP ranges. The weight values are set empirically but fixed in the experiments. Fig. 3 shows the average values of the critical points for Z- and t-test statistics in four TSL's for QP values according to (9), Tables II and III for *Mobile&Calendar* and *Foreman* sequences in CIF format.

TABLE III
WEIGHT VALUES FOR DIFFERENT TSL'S AND QP VALUES

QP ranges	TSL_1	TSL_2	TSL_3	above TSL_4
QP ≤ 24	0.10	0.10	0.10	0.15
24< QP ≤ 28	0.10	0.10	0.10	0.20
28< QP ≤ 32	0.10	0.10	0.20	0.30
32< QP ≤ 36	0.10	0.10	0.30	0.60
36< QP ≤ 40	0.10	0.20	0.40	0.80
40< QP	0.10	0.30	0.50	1.00

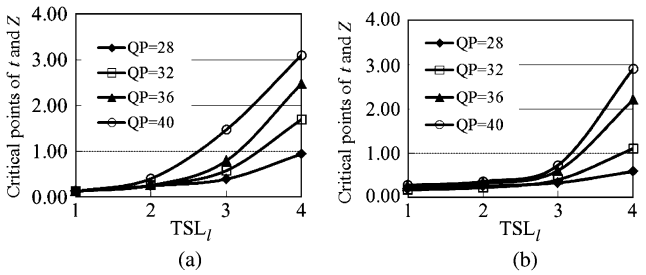

Fig. 3. Average values of critical points versus TSL's for different QP values. (a) *Mobile&Calendar*. (b) *Foreman*.

It can be shown in Fig. 3 that the critical points are set for the acceptance regions to be wider with the larger QP values, which makes the large block modes more likely be accepted.

Fallback check is performed only on 16 × 16 blocks when both hypothesis tests are accepted. This is because early termination on 16 × 16 block mode can cause more performance degradation than that on 8 × 8 block mode when false decision is made. The fallback check is to see whether a normalized difference between SAD value of the current MB (n-th MB) and averaged SAD value of MBs from the 1-st to the $(n-1)$-th MB exceeds a predefined threshold with a quantization step size. The fallback check on 16 × 16 block mode is performed according to

$$\frac{1}{16\times 16}\left\{\frac{1}{n-1}\sum_{i=0}^{n-1}|SAD_i| - SAD_c\right\} > c\cdot q_s \qquad (10)$$

where SAD_i and SAD_c are the SAD values of the best modes in the i-th MB and the current MB (n-th MB), respectively. q_s is a quantization step size and c is a constant value. The fallback check has trade-off between encoding speed and coding performance. When the inequality in (10) holds true after fallback check is performed, we call it "true" fallback otherwise "false" fallback. Fig. 4 shows the performance of fallback check in terms of probability of "true" fall back, encoding time saving, PSNR degradation and bitrate increase versus c for *Foreman* CIF sequences with QP = 28 and GOP size = 16. In Fig. 4(a), as c gets larger, the proportion of "true" fallback decreases so that encoding speed increases. In Fig. 4(b), relative encoding time is plotted as the ratio of the encoding time of the proposed method over the original JSVM 9.16. Fig. 4(c) and (d) are the BDPSNR and BDBR [11] plots versus c. It can be readily seen that RD performance is degraded as c increases, while encoding time saving is increased up to 48% for $c = 50$. For $c \geq 45$ in Fig. 4(a), there is always "false" fallback, and encoding time saving reaches about 48% but bitrate increases up

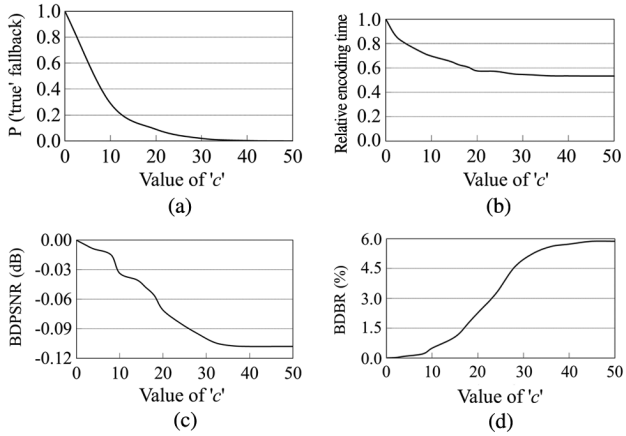

Fig. 4. Performance of fallback check with QP = 28 for *Foreman* CIF sequence. (a) Probability of "true" fallback (b) Relative encoding time. (c) BDPSNR(dB) vs. c. (d) BDBR(%) vs. c.

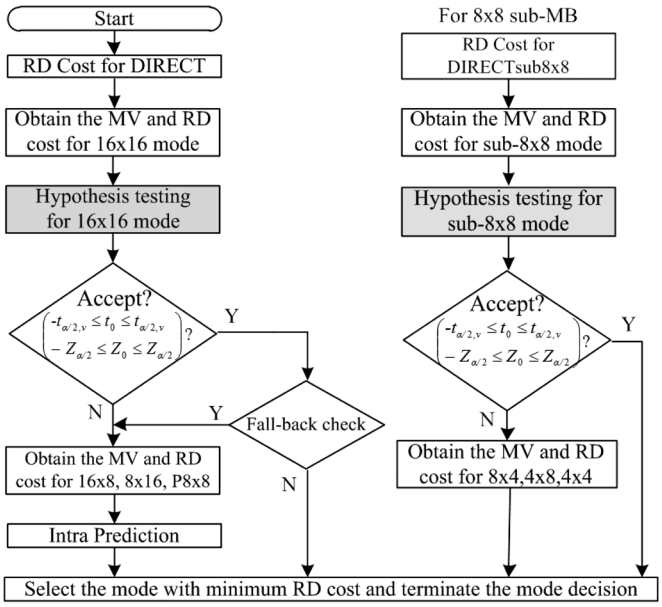

Fig. 5. Block diagram of the proposed fast inter-prediction mode decision method.

to 6% in BDBR. As shown in Fig. 4(a)–(c), a trade-off is necessary between encoding speed and coding efficiency. Therefore c should be carefully chosen to compromise the encoding speed and coding efficiency. c can be reasonably selected to be 20 for our experiments with all sequences and various quantization parameter values.

Fig. 5 shows the overall scheme for the proposed fast inter-prediction mode decision method. If either hypothesis on mean or variance test for 16 × 16 MB is not accepted, the mode decision process is performed for its smaller blocks, such as 16 × 8, 8 × 16, and 8 × 8. If the 8 × 8 block mode results in the least RD cost, the two hypothesis tests are then again performed in the same manner as the case of 16 × 16 MB. Accordingly, if the two hypothesis tests for the 8 × 8 block are accepted, then the mode decision process for smaller blocks of 8 × 4, 4 × 8, and 4 × 4 is skipped and the 8 × 8 block is determined to be the best block mode. In order to maintain good coding efficiency from

TABLE IV
EXPERIMENTAL CONDITIONS FOR THE PROPOSED MODE DECISION METHOD

Test conditions		Conditions	
		$-t_{a/2}$ and $-Z_{a/2}$	Fallback check
Proposed	*Case* 1	(9) not applied but fixed	On
	Case 2	(9) applied	Off
	Case 3	(9) applied	On

early termination on the block mode decision, the proposed fast inter-prediction mode decision method is conducted only on the B-pictures, not key-pictures in TSL_0 inside each GOP.

IV. EXPERIMENTAL RESULTS

The experiments for the proposed inter-prediction mode decision method were performed using H.264/SVC reference software, Joint Scalable Video Model (JSVM) 9.16 [10]. Four QPs of 28, 32, 36 and 40 are used. The search range for motion estimation is set to 32 pixels. The number of reference frames is set to 2 and GOP sizes 8 and 16 are used. The computer used for simulation is a PC with a 3.2GHz Pentium IV CPU and 1 GB RAM.

We set up the experimental conditions with 3 cases to test the performance of the proposed fast inter-prediction mode decision method. The three cases are summarized in Table IV. Note that *Case* 3 is our proposed method with confidence interval adjustment in TSL's and fallback check for fast inter-prediction mode decision in the hierarchical B-pictures. In addition, Liu's [2] and Li's [5] methods are compared with the proposed method with three cases. Performance of the proposed fast inter-prediction mode decision method is measured in terms of BDBR (%), BDPSNR (dB) [11] and average encoding time saving. Table V shows the performance of the proposed fast inter-prediction mode decision method with three cases and the proposed method is also compared with Liu's [2] and Li's [5] method in terms of coding efficiency and encoding time saving. As shown in Table V, the proposed method with *Case* 1 yields better performance of BDBR and BDPSNR but results in less encoding time saving compared to *Case* 2 and *Case* 3, because it does not adopt the TSL-dependent levels of significance in (9), rather uses a fixed level of significance in mean and variance tests, leading to conservative mode decision for early termination. Since *Case* 2 does not perform fallback check after the two hypothesis tests are accepted, it results in the largest encoding time saving for all test sequences compared with the others. But, it produces larger BDBR increases and larger BDPSNR drops compared to *Case* 3. Although *Case* 3 has smaller encoding speed-up due to fallback check compared with *Case* 2, it produces a lot less performance degradation in BDBR and BDPSNR.

The Li's method shows better performance than the proposed method with all three cases in perspective of encoding time saving for most of sequences except *Mobile&Calendar* and *Harbor* sequences, but severely suffers from performance degradation in BDPSNR and BDBR which cannot be acceptable. On the other hand, although the Liu's method maintains the coding efficiency, encoding time savings are not significant due to the fixed threshold values.

TABLE V
PERFORMANCE COMPARISONS OF THE PROPOSED METHOD WITH THREE CASES AND LIU'S [2] AND LI'S METHOD [5]

Test sequences (CIF)		GOP size = 8			GOP size = 16		
		BDBR	BDPSNR	ΔT_{avg}	BDBR	BDPSNR	ΔT_{avg}
Paris	*Case* 1	2.31	0.13	33.1	2.49	-0.13	36.6
	Case 2	9.39	-0.48	61.2	10.93	-0.54	61.5
	***Case* 3**	**3.41**	**-0.18**	**56.0**	**2.86**	**-0.15**	**52.5**
	Liu[2	1.83	-0.14	42.1	0.25	-0.01	38.6
	Li[5]	4.20	-0.23	58.6	6.19	-0.33	60.7
Tempete	*Case* 1	0.58	-0.02	20.8	0.87	-0.03	30.5
	Case 2	2.79	-0.11	64.7	3.16	-0.12	63.1
	***Case* 3**	**0.60**	**-0.02**	**53.9**	**0.90**	**-0.03**	**55.1**
	Liu[2]	2.87	-0.11	43.2	2.40	-0.09	34.7
	Li[5]	6.21	-0.24	52.2	9.34	-0.35	56.9
Foreman	*Case* 1	2.68	-0.11	21.5	2.95	-0.13	35.1
	Case 2	6.93	-0.30	53.0	7.21	-0.31	55.0
	***Case* 3**	**3.81**	**-0.17**	**49.8**	**3.06**	**-0.13**	**52.1**
	Liu[2]	1.69	-0.07	20.4	1.23	-0.05	25.6
	Li[5]	11.07	-0.50	54.9	16.33	-0.71	65.6
City	*Case* 1	0.95	-0.05	15.9	1.12	-0.06	17.8
	Case 2	3.52	-0.17	57.7	3.44	-0.17	60.7
	***Case* 3**	**1.91**	**-0.09**	**55.8**	**1.54**	**-0.08**	**59.6**
	Liu[2]	2.85	-0.03	29.0	3.00	-0.15	28.6
	Li[5]	8.02	-0.40	61.2	3.84	-0.33	63.9
Harbour	*Case* 1	1.71	-0.07	31.1	2.24	-0.08	31.0
	Case 2	4.83	-0.18	68.6	4.59	-0.17	62.6
	***Case* 3**	**2.54**	**-0.09**	**64.8**	**2.50**	**-0.09**	**60.0**
	Liu[2]	0.70	-0.03	30.4	0.97	-0.04	23.3
	Li[5]	5.41	-0.21	48.5	7.26	-0.27	48.0
Mobile& Calendar	*Case* 1	2.88	-0.11	34.8	4.13	-0.16	40.1
	Case 2	4.98	-0.20	69.9	6.35	-0.24	68.9
	***Case* 3**	**3.30**	**-0.13**	**64.2**	**3.95**	**-0.15**	**64.2**
	Liu[2]	2.66	-0.11	23.8	2.47	-0.10	20.9
	Li[5]	6.66	-0.27	48.6	9.70	-0.38	50.4
Average	*Case* 1	1.85	-0.04	26.2	2.30	-0.10	31.8
	Case 2	5.41	-0.24	62.5	5.95	-0.26	61.9
	***Case* 3**	**2.60**	**-0.11**	**57.4**	**2.47**	**-0.11**	**57.3**
	Liu[2]	2.10	-0.08	31.5	1.72	-0.07	29.4
	Li[5]	6.93	-0.31	54.0	8.78	-0.40	57.6

Fig. 6 illustrates relative encoding time versus CI lengths in mean and variance tests for *Case* 1. It can be observed in Fig. 6 that increasing the CI results in the reduction of computational complexity but entails RD performance degradation. Therefore, it is important to set an appropriate CI. As shown in Fig. 6(a), encoding time savings are reduced as the CI increases. The RD performance is degraded up to 7.43% increment in bitrate for $\mathrm{CI} = 60\%$ as shown in Fig. 6(b). As shown in Table V, *Case* 3 outperforms *Case* 1 in encoding time savings for all test sequences. The CI for *Case* 1 is set to 30% in order to have comparable RD performance as *Case* 3. While *Case* 1 does not result in significant encoding time savings for sequences in the higher QP range, *Case* 3 achieves significant encoding time saving over the entire range of bitrates with negligible amounts of RD degradation. This is due to the fact that the null hypotheses in *Case* 1 are more likely to be rejected in both mean and variance tests for the range of low bitrates with a fixed confidence interval, which corresponds to the case of lower pixel similarity caused by coarser quantization between the current blocks and reference blocks. Secondly, the fallback check following the both hypothesis tests further enhances the performance of the proposed

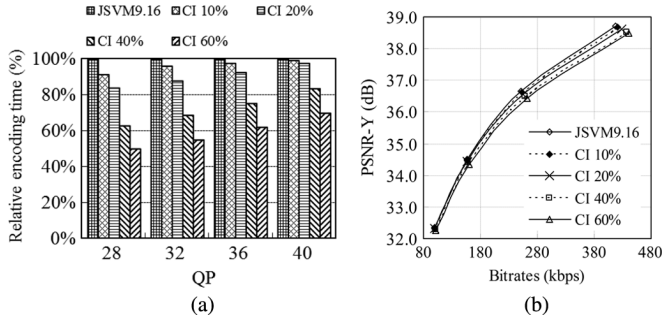

Fig. 6. RD performances and encoding time savings for *Case* 1 (*Foreman* CIF): (a) Relative encoding time; (b) RD curves for various confidence intervals.

method which corresponds to *Case* 3. It mitigates the degradation of RD performance by reducing the probability of type I error (also called false-alarm rate) for which, when H_0 is rejected, it is turned out to be true. The probability of type I error for 16 × 16 block mode is calculated by considering all the cases that when H_0 is rejected for the 16 × 16 block mode, the actual selected block modes turn out to be SKIP or 16 × 16 mode by the original JSVM 9.16. The probability of detection for 16 × 16 block mode are defined as

$$P_D = P(\text{true } H_0|\text{accept } H_0) = \frac{N_{H_0^{true},H_0^{accepted}}}{N_{H_0^{accepted}}} \quad (11)$$

where $N_{H_0^{true},H_0^{accepted}}$, is the number of MB's for which 16 × 16 or DIRECT modes become the final block modes determined by the original JSVM 9.16 (null hypothesis is true case) and two H_0's are accepted. $N_{H_0^{accepted}}$ is the number of MB's for which two H_0's are accepted, and N_T is the total number of MB's. The probability of false alarm for 16 × 16 block mode are defined as

$$P_{FA} = P(\text{true } H_0|\text{reject } H_0) = \frac{N_{H_0^{true},H_0^{rejected}}}{N_{H_0^{rejected}}} \quad (12)$$

where $N_{H_0^{true},H_0^{rejected}}$, is the number of MB's for which 16 × 16 or DIRECT modes become the final block modes by the original JSVM 9.16 and either of two H_0's is at least rejected. $N_{H_0^{rejected}}$ is the number of MB's for which either of two H_0's is at least rejected. In Table VI, the performance of the proposed fast inter-prediction mode decision method (*Case* 2, *Case* 3) is tabulated in terms of P_D and P_{FA} for three QP values. As shown in Table VI, P_D is greater than 0.75 for *Case* 2 and *Case* 3, and it increases with QP values when the fallback check is on. For the test sequences of small motion, P_D is larger for higher QP values. If the fallback check is turned off, P_D is reduced and P_{FA} is increased as shown in Table VI. Instead, encoding time saving is significantly increased at the expense of the coding efficiency degradation. The fallback check results in the reduction of false-alarm rates and increase of the probability of detection.

TABLE VI
PROBABILITIES OF DETECTION AND FALSE ALARM

Sequences (CIF)		QP = 28		QP = 32		QP = 36	
		P_D	P_{FA}	P_D	P_{FA}	P_D	P_{FA}
Foreman	Fallback check	0.77	0.71	0.86	0.75	0.91	0.83
	No fallback check	0.72	0.75	0.82	0.80	0.90	0.87
City	Fallback check	0.91	0.74	0.92	0.76	0.95	0.83
	No fallback check	0.81	0.61	0.91	0.80	0.94	0.87

V. CONCLUSION

Statistical hypothesis testing is used for early termination in the inter-prediction mode decision on hierarchical B-pictures for temporal scalability. In order to make it efficient the early termination in inter-prediction mode decision, two sets of hypotheses on mean and variance are effectively tested to check the similarity between the pixel values of the current block and those of a reference block for 16 × 16 and 8 × 8 block sizes. The confidence intervals are adjusted in temporal scalability levels by taking into account the proportion of large block mode selection in hierarchical B-picture GOP structure. In addition, a fallback check to reduce false-alarm rates is incorporated with hypothesis testing. It improves the detection power of the homogenous regions in 16 × 16 blocks and provides an effective compromise between encoding time saving and coding efficiency. The proposed fast inter mode decision method exhibited up to 64.8% of encoding time reduction with quality degradation of 0.11dB and average 2.54% bit increase.

REFERENCES

[1] H. Schwarz, D. Marpe, and T. Wiegand, "Overview of the scalable video coding extension of the H.264/AVC standard," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 17, no. 9, pp. 1103–1120, Sep. 2007.
[2] Z. Liu, L. Shen, and Z. Zhang, "An efficient intermode decision algorithm based on motion homogeneity for H.264/AVC," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 19, no. 1, pp. 128–132, Jan. 2009.
[3] 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.
[4] A. C. 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. 19, no. 2, pp. 186–195, Feb. 2008.
[5] H. Li, G. Li, and C. Wen, "Fast mode decision algorithm for inter-frame coding in fully scalable video coding," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 16, no. 7, pp. 889–895, Jul. 2006.
[6] B. Lee and M. Kim, "A low complexity mode decision method for spatial scalability coding," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 21, no. 1, pp. 88–95, Jan. 2011.
[7] M. Wien, H. Schwarz, and T. Oelbaum, "Performance analysis of SVC," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 17, no. 9, pp. 1194–1203, Sep. 2007.
[8] W. W. Hines, D. C. Montgomery, D. M. Goldsman, and C. M. Borror, *Probability and Statistics in Engineering*, 4th ed. Hoboken, NJ: Wiley, 2003.
[9] S. M. Kay, *Fundamentals of Statistical Signal Processing: Detection Theory*. Upper Saddle River, NJ: Prentice Hall, 1998.
[10] "Draft reference software for SVC," Joint Video Team (JVT) of ISO/IEC MPEG & ITU-T VCEG, JVT-AC203, Oct. 2008, 29th meeting.
[11] G. Bjontegaard, "Calculation of average PSNR differences between RD-curves," ITU-T Q6/SG16, Doc. VCEG-M33 Apr. 2001.