Multiview Coding Mode Decision With Hybrid Optimal Stopping Model

Tiesong Zhao, *Member, IEEE*, Sam Kwong, *Senior Member, IEEE*, Hanli Wang, *Senior Member, IEEE*, Zhou Wang, *Senior Member, IEEE*, Zhaoqing Pan, and C.-C. Jay Kuo, *Fellow, IEEE*

***Abstract*— In a generic decision process, optimal stopping theory aims to achieve a good tradeoff between decision performance and time consumed, with the advantages of theoretical decision-making and predictable decision performance. In this paper, optimal stopping theory is employed to develop an effective hybrid model for the mode decision problem, which aims to theoretically achieve a good tradeoff between the two interrelated measurements in mode decision, as computational complexity reduction and rate-distortion degradation. The proposed hybrid model is implemented and examined with a multiview encoder. To support the model and further promote coding performance, the multiview coding mode characteristics, including predicted mode probability and estimated coding time, are jointly investigated with inter-view correlations. Exhaustive experimental results with a wide range of video resolutions reveal the efficiency and robustness of our method, with high decision accuracy, negligible computational overhead, and almost intact rate-distortion performance compared to the original encoder.**

***Index Terms*— Inter-view prediction, mode characteristics, mode decision, multiview video coding, optimal stopping.**

I. Introduction

CONSEQUENTLY with H.264 Advanced Video Coding (AVC) [1] and Scalable Video Coding (SVC) [2], Multiview video coding (MVC) techniques [3] are developed to code Free-viewpoint TeleVision (FTV), Three-Dimensional TeleVision (3DTV) and many other multimedia applications. To fulfill the diversified requirements of stereoscopic and multiview storage, transmission and display, MVC is designed to be reliable, flexible, interactive, and with a good tradeoff between the quality of reconstructed views and the corresponding bit rates [4]. In MVC, besides Motion Estimation (ME) technique to remove temporal redundancy, Disparity Estimation (DE) is also adopted to further remove the inter-view redundancy [5], with a new prediction structure shown in Fig. 1. In each view, either IBBP or Hierarchical B Picture (HBP) [6] is supported. Among all views, the first view (*i.e.,* S0 in Fig. 1), namely, base view, is coded independently; while in the other views, parts or all pictures are predicted with both temporal ME and inter-view DE, which can remarkably improve the compression performance. Nevertheless, to obtain high compression efficiency, the computational complexity is also remarkably increased, which make the encoder optimization a necessity for real-time and mobile applications.

To address this issue, many researchers have focused on mode decision and ME/DE algorithms since in a video encoder, almost all coding time is consumed by mode decision and the related ME process [7]. These efforts have resulted in several efficient mode decision and fast ME/DE algorithms, including [8]–[26]. Among all these algorithms, neighboring prediction is usually employed, by investigating the neighboring and/or reference information, including but not limited to modes, textures, motions, Rate-Distortion (RD) costs, reference indexes, and so on. Besides that, several methods are also developed to select the candidate modes and determine the early termination conditions. In coding methods based on mode correlations [8], [10]–[12], [14], [15], [18]–[21], [24]–[26], parts of coding modes are checked first and then the remaining modes are decided accordingly, based on the statistical correlations between different modes. Specially, early Skip mode decision methods [10]–[12], [15], [18], [20], [21], [24]–[26] develop early termination condition after checking the Skip mode. If a pre-defined early termination condition is fulfilled, Skip mode is considered as the best and thus the checking of all other modes could be skipped. This method is with high efficiency for low bit rate coding due to the statistical fact that most of the best modes are Skip modes in this case. In RD cost based threshold estimation and ME/DE optimization methods [8], [10]–[15], [18]–[21], [23]–[26], the mode decision, multiple reference selection and ME/DE processes are early terminated with thresholds derived

Manuscript received April 30, 2012; revised December 9, 2012; accepted December 10, 2012. Date of publication December 20, 2012; date of current version February 12, 2013. This work was supported in part by the Hong Kong Research Grants Council General Research Fund, under Project 9041495 (CityU 115109) and City University of Hong Kong Grant 9610025, and the Program for Professor of Special Appointment (Eastern Scholar) at Shanghai Institutions of Higher Learning; the Program for New Century Excellent Talents in University of China under Grant NCET-10-0634; the Shanghai Pujiang Program under Grant 11PJ1409400; the National Natural Science Foundation of China under Grant 61102059 and Grant 61272289; the Fundamental Research Funds for the Central Universities under Grant 0800219158; the National Basic Research Program (973 Program) of China under Grant 2010CB328101; the National Sciences and Engineering Research Council of Canada; and the Ontario Ministry of Research and Innovation. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Joan Serra-Sagrista.

T. Zhao, S. Kwong, and Z. Pan are with the Department of Computer Science, City University of Hong Kong, Kowloon, Hong Kong (e-mail: ztiesong@uwaterloo.ca; cssamk@cityu.edu.hk; zqpan3@student.cityu.edu.hk).

H. Wang is with the Department of Computer Science & Technology and Key Laboratory of Embedded System and Service Computing, Ministry of Education, Tongji University, Shanghai 200092, China (e-mail: hanliwang@tongji.edu.cn).

Z. Wang is with the Department of Electrical and Computer Engineering, University of Waterloo, Waterloo, ON N2L 3G1, Canada (e-mail: zhouwang@ieee.org).

C.-C. J. Kuo is with the Ming Hsieh Department of Electrical Engineering and Signal and Image Processing Institute, University of Southern California, Los Angeles, CA 90089-2564 USA (e-mail: cckuo@sipi.usc.edu).

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/TIP.2012.2235451

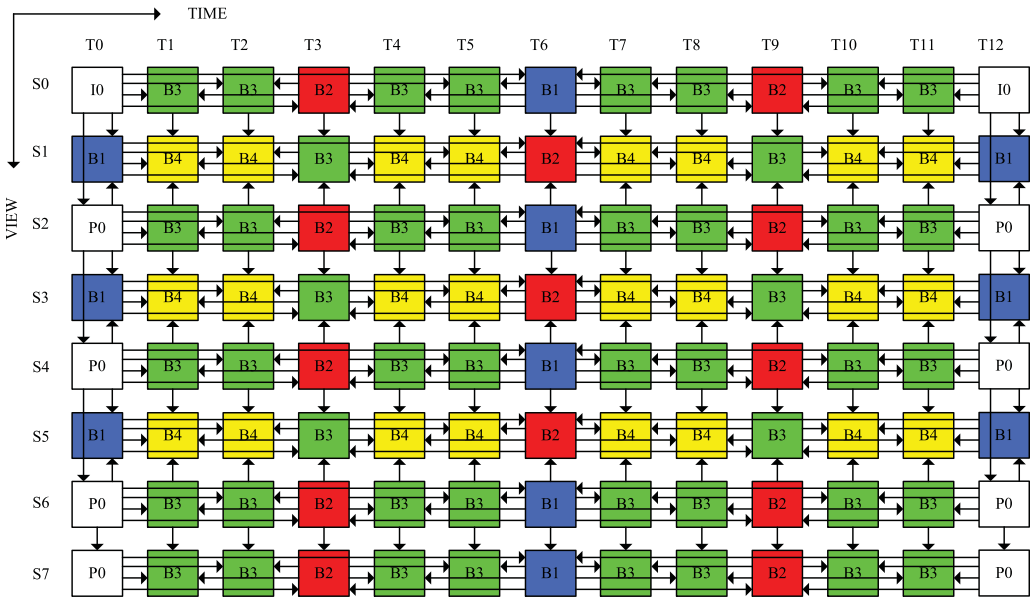


Fig. 1. Multiview video coding with 8 views.

from RD costs, Motion Vector (MV) and Disparity Vector (DV) correlations. In addition, the inter-view depth map and disparity characteristics could be also exploited to achieve more computational complexity reduction [16], [22].

Despite of all the above efforts, multiview coding mode decision could still be further improved, by theoretically developing efficient and robust model, to make a good tradeoff between computational complexity and decision accuracy. In our previous work [27], optimal stopping theory is employed in mode decision, by assuming all coding modes are with identical computational complexity. This method could achieve good performance; but, it could not well address the decision problem when coding complexities of different modes are entirely different. To address this issue and further improve the decision-making performance, in this paper, a new optimal stopping model is proposed to illustrate the relationship between decision accuracy and the actual coding time in a generic mode decision problem. After that, the new model and the model in [27] are compared and resulted in a hybrid model, with several model selection rules for diversified decision-making requirements. Besides, multiview coding mode characteristics are investigated and utilized to ulteriorly improve multiview coding performance, which is finally justified with both random trails and multiview sequences.

The rest of the paper is organized as follows. In Section II, the hybrid model is derived and justified, with investigation of mode characteristics in MVC encoder. Section III provides the overall algorithm with parameter estimations. The simulation results based on MVC codec are given in Section IV. Finally, Section V concludes the paper.

II. Proposed Hybrid Optimal Stopping Model

A typical optimal stopping problem is defined by a sequence of random variables with known joint distribution and a sequence of real-valued reward functions. The decision-maker checks these variables sequentially to get the observed value and finds a time to stop, aiming to maximize the expected reward [28], [29]. By regarding the coding modes as random variables and exploiting joint distribution of mode parameters, we could also address the mode decision problem with optimal stopping theory.

A. Related Work

In [30], Ferguson *et al.* proposed an optimal stopping problem named duration problem. During a decision process, if a variable is with better observed value than any variable before it, then it is called a Relatively Best Object (RBO). The objective of duration problem is to find a time to stop with the maximum expected duration until the next RBO. In this problem, a longer expected duration indicates both a higher probability of no RBO after the stop, and a larger time saving without unnecessary variable examination. In other words, the duration problem could achieve a good tradeoff between decision accuracy and time reduction.

By introducing prior probabilities for all variables to be the best, the authors developed an optimal stopping model for mode decision [27], which could be denoted as Probability-based-Model, or P-Model in the following text. In this model, we assume there are N candidates modes (*i.e.*, random variables), denoted as $X_k, k = 1, 2, \ldots, N$; and the corresponding probabilities to be the best mode are predicted as $\hat{p}_k, k = 1, 2, \ldots, N$. Besides, to keep almost intact RD performance in mode decision, we define a constraint threshold of decision performance as $\tau \in [N, N + 1)$; a larger τ indicates a better decision performance. Hence, P-Model aims to maximize the expected duration with a decision performance constraint. The solution to this problem consists of two steps. We first rank these modes with a descending order of probabilities

$$\hat{p}_i^p \geq \hat{p}_j^p, \forall i, j \in [1, N], i < j \tag{1}$$

where the superscript p denotes P-Model, and $p_k^p, k = 1, 2, \ldots, N$ represent the sorted probabilities; then, we check

TABLE I
PERCENTAGES NEEDED TO BE THE SAME MODE IN INTER-VIEW PREDICTION

Sequence	Qp	Upper	Left	Upper || Left	Forward	Backward	Forward || Backward
Ballet (1024 × 768)	20	69.39	69.12	81.05	69.45	68.90	81.30
	36	85.82	85.18	91.23	90.91	90.79	94.77
Breakdancer (1024 × 768)	20	44.04	44.80	61.10	38.68	39.35	55.19
	36	74.43	73.37	84.17	78.23	76.73	86.66
Champagnetower (1280 × 960)	20	76.89	76.66	85.78	81.06	81.79	89.66
	36	93.99	93.98	96.50	98.03	97.70	99.04
Dog (1280 × 960)	20	68.35	65.97	80.04	68.19	70.61	81.31
	36	88.95	86.60	92.70	94.81	94.52	96.73

all the N modes sequentially, with the optimal stop mode at

$$K_*^p = \max\left\{K_\alpha^p, K_\beta^p\right\} \tag{2}$$

where

$$K_\alpha^p = \min\left\{k \geq 1 : \sum_{i=1}^{k} \hat{p}_i^p \sum_{j=k}^{N} \frac{1}{\sum_{r=1}^{j} \hat{p}_r^p} > \tau - k\right\} \tag{3}$$

$$K_\beta^p = \min\left\{k \geq 1 : \hat{p}_{k+1}^p \sum_{j=k+1}^{N} \frac{1}{\sum_{r=1}^{j} \hat{p}_r^p} \leq 1\right\}. \tag{4}$$

In P-Model, all modes are assumed to be with identical computational complexity and thus, the duration is measured by the number of modes to be tested until the next RBO. Hence, it could not fully investigate the relationship between the two major measurements in mode decision problem, as computational complexity reduction and final decision performance. In addition, due to the engineering fact that different coding modes are usually with different computational complexities, P-Model cannot well address the decision problem when two coding modes are with similar probabilities but totally different complexities. Therefore, the optimal stopping model in mode decision should be further exploited with computational complexities.

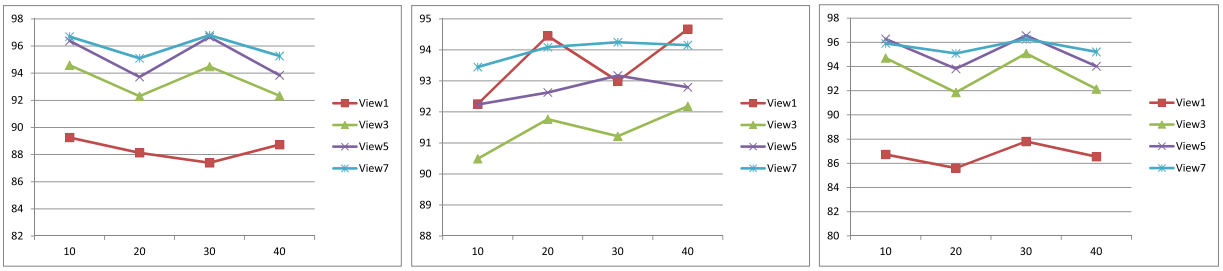
B. Inter-View Mode Characteristics

An intuitive and effective method to measure the computational complexity of a coding mode is the time proportions consumed among all coding modes. To develop a new optimal stopping model with both prior probabilities and time consuming, there are three problems yet to be addressed, as how to rank all these modes with probabilities and time proportions, where to stop in the sequential list, and how to estimate the decision performance. The discussion and derivation of this problem result in a Probability-and-Time-based-Model, or PT-Model in this work.

In multiview coding, the probability and coding time of each mode could be predicted based on statistical history information and spatial/inter-view correlations. Despite of disparity, there still exists high probability when co-located MacroBlocks (MBs) in different coding views have exactly the same coding mode. To justify this, for each MB with inter-view predictions, the upper, left, forward inter-view co-located and backward inter-view co-located MBs are separately evaluated with the probability to have the same mode. These probabilities are summarized in Table I, where four benchmark sequences (*Ballet*, *Breakdancer*, *Champagnetower* and *Dog*) are tested with 8 views and different Quantization parameters (Qps); two Group-Of-Pictures (GOPs) are tested with GOP size 12; fast search is enabled with search range 96; and || denotes the "or" condition. From the table, there exist high probabilities when these MBs are with the same mode to the coding MB; and these probabilities are even higher as Qp gets larger. As a result, the inter-view correlations could be utilized in prediction of mode characteristics.

To further investigate how much time reduction we could achieve with mode decision algorithm, we store the coding modes of several benchmark sequences first, and then reload these modes in the same coding environments. The time consuming of store/reload operations is negligible compared with the entire coding process, and thus it could be ignored. In such a case, the original and reloaded schemes achieve exactly the same coding performance, including bit rates and video quality. However, during the store-and-reload process, the computational complexity is significantly reduced and consequently, the coding time is saved. Intuitively, the time reduction between the two schemes shows ideal mode decision curves in Fig. 2, which indicates the maximum time reductions without any compression efficiency loss. In this figure, three benchmark sequences (*Ballet*, *Breakdancer* and *Dog*) are tested, with Qp from 10 to 40 and the other parameters same to Table I.

From Fig. 2, the overall coding time could be significantly saved in all cases, which is because large mode partitions are more probable to be the best mode while the time proportions of these modes are relatively small. Therefore, there exists an enormous potential to reduce computational complexity by skipping unnecessary coding modes, or even predicting the best coding mode before the whole mode decision process. Theoretically, mode decision algorithm can achieve the same time reduction to Fig. 2 with intact coding performance to the original encoder. However, due to large amount of video data and exhaustive computation in RD Optimization (RDO), the best coding mode can only be exactly obtained after the comprehensive mode decision process. Hence, by skipping unnecessary coding modes, the mode decision algorithms would have a little loss in coding efficiency including



(a) (b) (c)

Fig. 2. Ideal inter-view mode decision curve examples in percentage. 8 views are tested with view order 0-2-1-4-3-6-5-7. Vertical axis: time reduction (%). Horizontal axis: Qp. (a) *Ballet* (1024 × 768). (b) *Breakdancer* (1024 × 768). (c) *Dog* (1280 × 960).

Peak-Signal-to-Noise-Ratio (PSNR) and bit rates; on the other hand, the time reduction can be improved to be closer to the ideal mode decision curves.

C. Optimal Stopping Model With Time Proportions

In P-Model, the duration is measured with the number of random variables to be examined during the interval. In PT-Model, we let the estimated examination time of variable X_k be $\hat{t}_k$ with $\sum_{k=1}^{N} \hat{t}_k = 1$, then the duration could be measured with "real" time. Assume we early terminate at k with the next RBO at T_k, then the duration time is $\sum_{j=k+1}^{T_k} \hat{t}_j$. Especially, a virtual variable $N+1$ is defined with the examination time $\hat{t}_{N+1}$. If we early terminate at k, then $T_k = N$ indicates that the next RBO is the last variable; while $T_k = N+1$ indicates that there is no RBO after k, in other words, the best observation value exists in the first k variables.

With the predicted probabilities $\hat{p}_k, k = 1, 2, \ldots, N$, when stopping at k, the probability of the next RBO could be derived as [27]

$$P(T_k = j) = \begin{cases} \sum_{i=1}^{k} \hat{p}_i \left\{ \frac{1}{\sum_{r=1}^{j-1} \hat{p}_r} - \frac{1}{\sum_{r=1}^{j} \hat{p}_r} \right\}, & j \in (k, N], \\ \sum_{i=1}^{k} \hat{p}_i, & j = N+1. \end{cases} \tag{5}$$

Ulteriorly, the expected next RBO could be predicted as

$$\begin{aligned} E(T_k) &= (N+1) \cdot P(T_k = N+1) \\ &\quad + \sum_{j=k+1}^{N} \left[j \cdot P(T_k = j) \right] \\ &= k + \sum_{i=1}^{k} \hat{p}_i \sum_{j=k}^{N} \frac{1}{\sum_{r=1}^{j} \hat{p}_r} \end{aligned} \tag{6}$$

and $\forall k \in [1, N]$, the expected duration time is

$$\begin{aligned} y_k &= \sum_{j=k+1}^{N} \left[\sum_{r=k+1}^{j} \hat{t}_r \cdot P(T_k = j) \right] \\ &\quad + \sum_{r=k+1}^{N+1} \hat{t}_r \cdot P\left(T_k = N+1\right) \\ &= \sum_{i=1}^{k} \hat{p}_i \sum_{j=k}^{N} \frac{\hat{t}_{j+1}}{\sum_{r=1}^{j} \hat{p}_r}. \end{aligned} \tag{7}$$

To find a k to maximize the expected duration, we first make y_k a unimodal function [30], with the necessary and sufficient condition:

1) $y_1 - y_0 > 0$;
2) $\forall m, n \in [1, N-1], m > n, y_{n+1} - y_n \leq 0 \Rightarrow y_{m+1} - y_m < 0$;
3) $y_N - y_{N-1} < 0$.

In this condition, $y_1 - y_0 > 0$ is always true because $y_0 = 0$ in the sense of no duration obtained if there is no variable examined. To fulfill the second item of the above condition, a sufficient condition could be derived to rank all modes in PT-Model, as

$$\frac{\hat{p}_i^{pt}}{\hat{t}_i^{pt}} \geq \frac{\hat{p}_j^{pt}}{\hat{t}_j^{pt}}, \quad \forall i, j \in [1, N], \quad i < j \tag{8}$$

where the superscript pt denotes PT-Model, p_k^{pt}, $k = 1, 2, \ldots, N$ and t_k^{pt}, $k = 1, 2, \ldots, N$ represent the sorted probabilities and time proportions in PT-Model. With $y_N - y_{N-1} < 0$ and the rank condition in Eq. (8), it could be obtained that

$$0 < \hat{t}_{N+1} < \frac{\hat{t}_N^{pt}}{\hat{p}_N^{pt}} \leq 1. \tag{9}$$

In this work, we set $\hat{t}_{N+1}$ as the average time proportion of examining each random variables or modes

$$\hat{t}_{N+1} = \frac{1}{N}. \tag{10}$$

Correspondingly, after sorting, the expected next RBO and the expected duration could be updated as

$$E(T_k^{pt}) = k + \sum_{i=1}^{k} \hat{p}_i^{pt} \sum_{j=k}^{N} \frac{1}{\sum_{r=1}^{j} \hat{p}_r^{pt}} \tag{11}$$

$$y_k^{pt} = \sum_{i=1}^{k} \hat{p}_i^{pt} \left\{ \sum_{j=k}^{N-1} \frac{\hat{t}_{j+1}^{pt}}{\sum_{r=1}^{j} \hat{p}_r^{pt}} + \frac{1}{N} \right\}. \tag{12}$$

In many engineering applications, the duration problem itself could achieve a good balance between decision accuracy and time reduction, and thus we could just utilize the maximum value of the unimodal function y_k^{pt}. However, in mode decision and some related problems, the objective is

TABLE II
COMPARISON BETWEEN PT-MODEL AND P-MODEL WITH RANDOM DATA

A **versus** B	A > B	A = B	A < B	**AVE** (A − B)
$E(T_k^{pt})$ versus $E(T_k^p)$	49.00%	20.08%	30.92%	0.0285
y_k^{pt} versus $y_k^p t$	48.69%	20.08%	31.23%	0.0984
TS^{pt} versus TS^p	39.43%	37.40%	23.17%	10.48

TABLE III
$E(T_k^{pt}) - E(T_k^p)$ WITH AVERAGE MODE CHARACTERISTICS

Qp	*Ballet*	*Breakdancer*	*Champagnetower*	*Dog*
20	0.0000	0.0000	0.2631	0.0000
26	0.0011	0.0000	−0.0954	0.2740
32	0.3027	0.3206	0.0284	−0.1727
38	−0.1456	0.1845	0.0203	−0.1107

TABLE IV
$y_k^{pt} - y_k^p$ WITH AVERAGE MODE CHARACTERISTICS

Qp	*Ballet*	*Breakdancer*	*Champagnetower*	*Dog*
20	0.0000	0.0000	0.0478	0.0000
26	−0.0003	0.0000	0.1118	0.0546
32	0.0638	0.0685	−0.0039	0.0912
38	0.1423	0.0325	−0.0024	0.1249

TABLE V
$TS^{pt} - TS^p$ (%) WITH AVERAGE MODE CHARACTERISTICS

Qp	*Ballet*	*Breakdancer*	*Champagnetower*	*Dog*
20	0.00	0.00	−1.70	0.00
26	0.00	0.00	13.56	−1.48
32	−1.05	−0.95	0.00	13.41
38	17.75	−1.17	0.00	15.19

to maximize the time reduction with a negligible loss in RD performance. Hence, we employ a constraint in decision performance, and design the objective as

$$\max\left\{y_k^{pt}\right\}, \quad s.t. \quad E(T_k^{pt}) > \tau, k \in [1, N] \tag{13}$$

where $\tau \in [N, N+1)$ in the sense of the expected next RBO is after the last variable to be examined.

Since that y_k^{pt} is a unimodal function, and $E(T_k^{pt})$ could be proved as a monotonically increasing function, the solution to Eq. (13) could be finally derived as

$$K_*^{pt} = \max\left\{K_\alpha^{pt}, K_\beta^{pt}\right\} \tag{14}$$

where

$$K_\alpha^{pt} = \min\left\{k \geq 1 : E\left(T_k^{pt}\right) > \tau\right\} = \min\left\{k \geq 1 : \sum_{i=1}^{k} \hat{p}_i^{pt} \sum_{j=k}^{N} \frac{1}{\sum_{r=1}^{j} \hat{p}_r^{pt}} > \tau - k\right\} \tag{15}$$

$$K_\beta^{pt} = \min\left\{k \geq 1 : y_{k+1}^{pt} - y_k^{pt} \leq 0\right\} = \min\left\{k \geq 1 : \frac{\hat{t}_{k+1}^{pt}}{\hat{p}_{k+1}^{pt}} - \sum_{j=k+1}^{N-1} \frac{\hat{t}_{j+1}^{pt}}{\sum_{r=1}^{j} \hat{p}_r^{pt}} \geq \frac{1}{N}\right\}. \tag{16}$$

D. Hybrid Optimal Stopping Model

In our previous work [27], P-Model is developed with the candidate rank condition as in Eq. (1) and the optimal stopping condition as in Eqs. (2–4); while in PT-Model, the candidate rank condition and optimal stop are decided by Eq. (8) and Eq. (14–16), respectively. To compare the two models, we employ both random trails and average mode characteristics, with the same parameters in [27]. Three criteria are employed, as the expected next RBO $E(T_k^{pt})$ vs. $E(T_k^p)$, the expected duration y_k^{pt} vs. y_k^p, and the expected time reduction TS^{pt} vs. TS^p (%) after optimal stop.

First, both PT-Model and P-Model are examined and compared with 1 000 000 random trails. In each trail, there are $N = 6$ random $(\hat{p}_k, \hat{t}_k)$ variables. The comparison results are summarized in Table II, with the probabilities of PT-Model to be better, equal and worse than P-Model shown in terms of percentage. In addition, the average differences are also given in the last column. We could notice that, in most of cases, as well as the average values, PT-Model is superior; while still in some cases, P-Model has better performance due to different rank conditions. For example, in a random $(\hat{p}_k, \hat{t}_k)$ trail: X_1(0.2304, 0.0134), X_2(0.0576, 0.0673), X_3(0.2762, 0.3493), X_4(0.1909, 0.0486), X_5(0.0614, 0.2333) and X_6(0.1835, 0.2881), after sorting with different conditions, we could obtain that $E(T_k^{pt}) > E(T_k^p)$, $y_k^{pt} < y_k^p$ and $TS^{pt} < TS^p$.

To compare the performances of P-Model and PT-Model in mode decision, the average probability and time proportion of each mode are utilized. Four benchmark sequences (*Ballet*, *Breakdancer*, *Champagnetower* and *Dog*) are evaluated Qps from 20 to 38, and the other configuration parameters are the same to Table I. The average differences between the expected next RBO $E(T_k^{pt})$ vs. $E(T_k^p)$, the expected duration y_k^{pt} vs. y_k^p, and the expected time reduction TS^{pt} vs. TS^p (%) are given in Tables III–V, respectively. From the table, PT-Model has better decision performance for smaller Qp, or high bit rate coding, and better time reduction for larger Qp, or low bit rate coding. The reason is for all coding modes, the mode probabilities are remarkably changed as Qp changes, while the time proportions are almost the same. Hence, the model preference is also changed due to different rank conditions in different models.

As a conclusion, both of the two models, P-Model and PT-Model, have advantages in different cases. Hence, a hybrid model could be developed by incorporating the advantages of the two models. Considering in the above comparisons, there exists a low probability when PT-Model is superior to P-Model in all the three criteria (4.26% with random trails), and a negligible probability when P-Model is superior to PT-Model in all criteria (almost 0.00% with random trails), we can use one of the three criteria to decide which model

would be selected:

1) aim to achieve higher expected decision accuracy. The expected next relatively best candidate, $E(T_k^{pt})$ and $E(T_k^p)$ are predicted and compared; then, the model with better expected T_k value is selected;
2) aim to achieve more duration time after optimal stop. The two models are compared with the duration time y_k^{pt} and y_k^p; then the model with larger duration is selected;
3) Aim to achieve more time reduction. The two models are separately utilized to predict TS^{pt} and TS^p; then, the model with more time reduction is selected in the decision process.

The above three selection rules could be employed in different problems, depending on the tradeoff between decision accuracy and time saving. In mode decision, to achieve almost intact RD performance compared with the original encoder, the first rule is finally selected.

III. Proposed Overall Algorithm

In inter-view prediction, there are $N = 6$ modes to be decided, as Skip/Direct, 16 × 16, 16 × 8, 8 × 16, 8 × 8 (in this mode, each of the four sub-blocks can be coded with sub-8×8, sub-8×4, sub-4×8, sub-4×4 or Skip-8×8) and intra modes, where intra modes would be checked together due to high correlations between each other. To employ the hybrid model, the parameters including mode probabilities and time proportions should be estimated before the coding process. In this work, the statistical method is employed, with neighboring prediction and inter-view correlations.

A. Estimation of Parameters

With a small number of variables, the optimal stop point might be not changed if there are small prediction errors in parameters. Take the aforementioned trail X_1(0.2304, 0.0134), X_2(0.0576, 0.0673), X_3(0.2762, 0.3493), X_4(0.1909, 0.0486), X_5(0.0614, 0.2333) and X_6(0.1835, 0.2881) for example, in which PT-Model is selected because $E(T_k^{pt}) > E(T_k^p)$, and the early termination is after checking X_1, X_4, X_2, X_3 and X_6. If the probabilities of X_1 and X_4 are estimated as 0.3304 (increased by 0.1) and 0.0909 (decreased by 0.1), the optimal stopping conditions in P-Model and PT-Model are kept unchanged; also, PT-Model would be selected with early termination after checking X_1, X_4, X_2, X_3 and X_6. The reason is, for a smaller N, the parameters should be greatly changed to skip one more variable, with a strict constraint in decision performance. Based on this observation, the parameters $\hat{p}_k, \hat{t}_k, k = 1, 2, \ldots, N$ could be approximately estimated with statistics and neighboring prediction.

In statistical methods, both adaptive and fixed methods could be adopted, depending on the statistic characteristics of the parameters. In general, adaptive method has better performance when the parameters usually change, such as the mode probabilities $\hat{p}_k, k = 1, 2, \ldots, N$, which would change due to different textures, motions and coding parameters. Hence, we employ the adaptive mode probability prediction method in [27], with four reference modes in Table I, as the upper mode, the left mode, the forward inter-view mode and the backward inter-view mode. Four adaptively updated prediction matrices are defined as:

1) upper prediction matrix $\mathbf{Tu}_{M,k}, k = 1, 2, \ldots, N, M = 1, 2, \ldots, N$, which indicates the percentage of mode k when the upper mode is M;
2) left prediction matrix $\mathbf{Tl}_{M,k}, k = 1, 2, \ldots, N, M = 1, 2, \ldots, N$, which indicates the percentage of mode k when the left mode is M;
3) forward inter-view prediction matrix $\mathbf{Tf}_{M,k}, k = 1, 2, \ldots, N, M = 1, 2, \ldots, N$, which indicates the percentage of mode k when the co-located mode in forward view is M;
4) backward inter-view prediction matrix $\mathbf{Tb}_{M,k}, k = 1, 2, \ldots, N, M = 1, 2, \ldots, N$, which indicates the percentage of mode k when the co-located mode in backward view is M.

With the above prediction matrices, for an MB to be coded, the probability $\hat{p}_k$ could be estimated as

$$\hat{p}_k \approx \frac{\mathbf{Tu}(M_u, k) + \mathbf{Tl}(M_l, k) + \mathbf{Tf}(M_f, k) + \mathbf{Tb}(M_b, k)}{\sum_{r=1}^{N} \left(\mathbf{Tu}(M_u, r) + \mathbf{Tl}(M_l, r) + \mathbf{Tf}(M_f, r) + \mathbf{Tb}(M_b, r)\right)} \tag{17}$$

where M_u, M_l, M_f and M_b represent the upper, left, forward inter-view and backward inter-view modes, respectively. If either of the upper, left, forward inter-view and backward inter-view MB is not available, the corresponding prediction matrix can be set as 0 and Eq. (17) is still applicable in probability estimation.

Fixed method is preferred when the parameters are almost the same with different coding parameters, such as the time proportions $\hat{t}_k, k = 1, 2, \ldots, N$, which are quite similar with the same Qp. Since small prediction error in parameters is acceptable with a small value of N, we employ fixed time proportions for different Qps. In this work, the average time proportions for all coding modes are obtained by coding 2 GOPs of 9 benchmark sequences (*Ballroom*, *Exit*, *Race1*, *Vassar*, *Ballet*, *Breakdancer*, *Doorflowers*, *Champagnetower* and *Dog*), with Qps from 8 to 44. Based on the statistics, the average time proportions for mode k is estimated by

$$\hat{t}_k(Qp, L) = \sum_{r=0}^{M} \alpha_r Qp^r \tag{18}$$

where α_r is a weighting parameter; L indicates the temporal level in a GOP; $M = 3$ (*i.e.,* quadratic fit) for Skip mode and $M = 2$ (*i.e.,* linear fit) for the other inter modes. In Table VI, $\alpha_r, r = 1 \ldots M$ are given for different modes, temporal levels and multiview prediction structures, where L_{MAX} represents the highest temporal level.

After optimal stop, the prediction matrices are updated with the coding results. Take PT-Model for example and assume the best mode is obtained as $j \le K_*^{pt}$, the posterior probability for mode k to be the best could be derived as [27]

$$\widehat{p}_k = \begin{cases} \sum_{r=1}^{K_*^{pt}} \hat{p}_r^{pt}, & \text{if } k = j, \\ 0, & \text{if } k \le K_*^{pt}, k \ne j, \\ \hat{p}_k^{pt}, & \text{otherwise.} \end{cases} \tag{19}$$

TABLE VI
PARAMETERS OF TIME PROPORTIONS FOR ALL MODES

Mode	$L \le L_{MAX}-4$			$L = L_{MAX}-3$			$L = L_{MAX}-2$			$L = L_{MAX}-1$			$L = L_{MAX}$		
	α_2	α_1	α_0	α_2	α_1	α_0	α_2	α_1	α_0	α_2	α_1	α_0	α_2	α_1	α_0
MBs With Double Inter-View Predictions															
Skip	1.0e − 4	−0.001	0.245	1.2e−4	−0.006	0.171	1.3e−4	−0.006	0.176	1.4e−4	−0.007	0.184	1.4e−4	−0.007	0.182
16 × 16		0.244	8.105		0.223	7.815		0.219	7.882		0.216	7.901		0.213	7.969
16 × 8		0.145	9.438		0.117	10.09		0.112	10.27		0.108	10.40		0.104	10.54
8 × 16		0.148	10.82		0.120	11.42		0.114	11.59		0.109	11.76		0.104	11.92
8 × 8		−0.555	70.14		−0.465	69.99		−0.453	69.59		−0.443	69.30		−0.432	68.99
MBs With Single Inter-View Prediction															
Skip	−1.2e − 5	0.004	−0.000	1.0e−4	−0.004	0.405	9.0e−5	−0.004	0.302	7.3e−5	−0.003	0.288	6.9e−5	−0.003	0.284
16 × 16		0.306	7.691		0.205	8.019		0.174	8.690		0.153	9.189		0.139	9.590
16 × 8		0.112	9.616		0.093	10.27		0.083	10.78		0.073	11.16		0.063	11.55
8 × 16		0.134	10.89		0.098	11.45		0.085	12.00		0.073	12.37		0.063	12.71
8 × 8		−0.624	69.29		−0.403	67.48		−0.347	66.54		−0.302	65.30		−0.268	64.13

And the prediction matrices could be linearly updated, such as

$$\mathbf{Tu}(M_u, k) = \mathbf{Tu}(M_u, k) \cdot (1 - \gamma) + \widehat{p}_k \cdot \gamma \qquad (20)$$

where γ is a regulation parameter with a typical value 0.08 based on exhaustive experiments.

B. Overall Algorithm

Finally, the overall algorithm consists of five steps as follows. In this work, to achieve high coding performance, we set $\tau = N + 4/5$; and for anchor pictures (*e.g.,* T0, T12 in Fig. 1), all modes are checked in I/P frames.

Step 1: *Parameter Estimation:* estimate the probabilities $\hat{p}_k, k = 1, 2, \ldots, N$ with spatial/inter-view neighboring MBs; estimate the time proportions $\hat{t}_k, k = 1, 2, \ldots, N$ based on the average time statistics.

Step 2: *Model Prediction:* predict the rank conditions and early terminations for both P-Model and PT-Model; predict $E(T_k^{pt})$ and $E(T_k^{p})$ with early termination.

Step 3: *Model Selection:* if $E(T_k^{pt}) >= E(T_k^{p})$, then PT-Model would be selected; otherwise P-Model would be selected.

Step 4: *Mode Decision:* initialize the candidate mode list, check all the candidate modes with optimal stop based on the model selected in Step 3, and decide the best coding mode among all examined modes. Parts of sub-inter modes could also be skipped based on the most probable mode (*i.e.,* the first mode in the candidate mode list): sub-8 × 4 could be skipped if the most probable mode is not 16 × 8 or 8 × 8; sub-4 × 8 could be skipped if the most probable mode is not 8 × 16 or 8 × 8; sub-4 × 4 could be skipped if the most probable mode is not 8 × 8.

Step 5: *Parameter Update:* update the posterior probabilities and the related prediction matrices.

IV. EXPERIMENTAL RESULTS

To evaluate the coding performance of our method, it is implemented on MVC reference software JMVC [31] and compared with Shen's low-complexity mode decision algorithm [24], in which Global Disparity Vector (GDV) based neighbor mode prediction, RD cost based early termination and All Zero Block (AZB) detection are jointly employed in inter-view mode decision, and Zeng's fast mode decision algorithm [25], which consists of weighted neighor prediction, early Skip mode decision, and Predicted MV (PMV) based mode classification, ME/DE seleciton, and so on. To examine the coding performances with various scenes and bit rates, two 320 × 240 sequences (*Flamenco1* and *Golf1*), four 640 × 480 sequences (*Ballroom*, *Exit*, *Race1*, *Vassar*), six 1024 × 768 sequences (*Ballet*, *Breakdancer*, *Doorflowers*, *Jungle*, *Lovebird1*, *Uli*) and two 1280 × 960 sequences (*Champagnetower* and *Dog*) are tested, with configuration parameters given in Table VII.

TABLE VII
SIMULATION ENVIRONMENT FOR INTER-VIEW MVC

Encoder	JMVC 8.3.1 [31]
Qps	20, 26, 32, 38
Resolutions	320 × 240, 640 × 480, 1024 × 768, 1280 × 960
Configurations	GOP size: 12 Frames coded: 2 GOPs × 8 views ViewOrder: 0-2-1-4-3-6-5-7 Number of reference frames: 2 RDO: enabled BiPredIter: 4 IterSearchRange: 8 ME: fast search with 1/4 pixel Search range: ±96 Entropy coding: CABAC

A. Comparison of Mode Decision Algorithms

In the comparison, 8 views are coded, with view order 0-2-1-4-3-6-5-7, as shown in Table VII. In even views (*i.e.,* view 0, 2, 4, 6), inter-view predictions only exist in anchor pictures, and thus these views could be optimized with single-view mode decision methods. For the odd views, three algorithms, as Shen's [24], Zeng's [25] and the proposed hybrid model, are implemented with comprehensive experimental results summarized in Tables VIII (for the sequences to train

TABLE VIII
SIMULATION RESULTS OF ODD VIEWS FOR THE SEQUENCES USED TO TRAIN TIME PARAMETERS

Sequence	Qp	Shen's [24]			Hybrid Model*			Zeng's [25]			Hybrid Model		
		TS	ΔPSNR	ΔBR	TS	ΔPSNR	ΔBR	TS	ΔPSNR	ΔBR	TS	ΔPSNR	ΔBR
Ballroom	20	55.56	−0.022	0.17	68.02	−0.035	1.10	65.76	−0.101	2.89	67.41	−0.040	1.32
	26	70.18	−0.026	0.32	73.05	−0.035	1.58	66.11	−0.090	4.54	72.65	−0.048	1.84
	32	77.49	−0.036	0.28	76.40	−0.040	2.41	65.55	−0.111	6.11	75.64	−0.057	2.34
	38	80.45	−0.083	1.70	79.22	−0.07	2.16	65.07	−0.146	7.87	78.10	−0.078	2.10
	Average	70.92	−0.053	1.51	74.17	−0.104	3.19	65.62	−0.297	8.63	73.45	−0.123	3.57
Exit	20	54.37	−0.021	0.64	71.15	−0.036	1.07	64.70	−0.126	3.18	70.93	−0.040	1.01
	26	76.54	−0.067	4.67	78.95	−0.033	2.40	67.13	−0.111	6.41	78.07	−0.042	2.41
	32	84.46	−0.174	8.55	82.02	−0.063	2.67	68.33	−0.174	9.14	81.64	−0.078	3.03
	38	87.80	−0.347	9.28	84.87	−0.125	3.71	71.38	−0.276	8.66	84.05	−0.130	3.15
	Average	75.79	−0.276	12.01	79.25	−0.116	4.79	67.89	−0.334	14.19	78.67	−0.126	5.18
Race1	20	62.07	−0.053	0.19	74.41	−0.049	0.60	49.90	−0.177	5.45	74.88	−0.052	0.96
	26	71.18	−0.098	0.83	76.86	−0.047	0.70	50.30	−0.189	4.82	77.05	−0.054	1.18
	32	77.55	−0.331	8.45	78.10	−0.049	0.35	49.18	−0.176	3.84	78.30	−0.057	1.18
	38	81.41	−0.904	28.65	78.88	−0.071	0.78	48.61	−0.186	3.03	78.94	−0.078	1.68
	Average	73.05	−0.749	14.21	77.06	−0.074	1.85	49.49	−0.346	9.27	77.29	−0.106	2.70
Vassar	20	58.53	−0.013	−0.20	75.25	−0.025	0.02	79.60	−0.112	−0.04	74.82	−0.028	−0.03
	26	78.11	−0.026	−0.22	86.25	−0.022	0.60	79.79	−0.053	0.52	85.69	−0.025	0.40
	32	85.71	−0.012	−0.28	90.52	−0.031	0.54	79.20	−0.016	1.28	89.88	−0.031	0.24
	38	89.34	−0.043	−1.06	91.62	−0.053	0.11	78.79	−0.024	2.26	91.31	−0.051	0.35
	Average	77.92	−0.014	0.65	85.91	−0.037	1.93	79.35	−0.067	3.29	85.42	−0.036	1.76
Ballet	20	67.72	−0.035	1.77	78.16	−0.024	1.42	68.17	−0.058	2.18	77.91	−0.028	1.27
	26	83.08	−0.110	5.89	81.32	−0.038	1.91	69.21	−0.052	4.59	81.17	−0.048	2.23
	32	87.04	−0.334	8.13	84.17	−0.057	1.31	69.96	−0.080	6.53	83.95	−0.082	2.34
	38	89.07	−0.593	2.04	85.90	−0.111	0.39	70.55	−0.120	8.33	85.76	−0.146	2.36
	Average	81.73	−0.317	16.60	82.39	−0.079	3.38	69.47	−0.213	9.26	82.20	−0.117	5.04
Breakdancer	20	45.98	−0.013	0.02	72.40	−0.027	0.95	48.90	−0.118	4.24	72.58	−0.028	0.83
	26	62.59	−0.034	0.22	73.22	−0.044	1.19	53.17	−0.14	7.19	74.13	−0.047	1.11
	32	73.33	−0.134	−0.21	76.49	−0.048	0.89	57.04	−0.221	9.72	77.03	−0.052	0.71
	38	79.88	−0.320	−0.33	79.44	−0.088	0.79	61.88	−0.312	12.08	79.65	−0.100	1.45
	Average	65.44	−0.082	4.16	75.39	−0.067	3.22	55.25	−0.349	18.59	75.85	−0.073	3.35
Doorflowers	20	74.31	−0.050	1.80	83.56	−0.022	0.86	76.87	−0.072	1.18	83.41	−0.024	0.85
	26	88.60	−0.104	4.79	89.27	−0.037	3.07	78.52	−0.035	1.33	89.03	−0.040	2.52
	32	92.85	−0.220	11.20	91.85	−0.050	4.10	78.57	−0.033	1.58	91.56	−0.053	3.58
	38	93.40	−0.354	10.80	93.85	−0.096	3.68	77.44	−0.042	2.68	93.48	−0.092	2.63
	Average	87.29	−0.28	17.64	89.63	−0.105	5.85	77.85	−0.077	4.13	89.37	−0.098	5.27
Champagnetower	20	78.64	−0.052	0.00	84.43	−0.044	0.60	78.45	−0.081	0.59	84.34	−0.046	0.43
	26	88.32	−0.111	1.41	89.69	−0.056	1.10	80.44	−0.069	1.05	89.56	−0.071	1.50
	32	92.35	−0.225	0.55	93.61	−0.124	1.23	80.95	−0.055	1.58	93.43	−0.143	1.64
	38	93.94	−0.189	2.80	96.37	−0.129	0.10	81.41	−0.068	1.81	96.13	−0.169	−0.04
	Average	88.31	−0.186	6.39	91.03	−0.114	4.09	80.31	−0.108	3.56	90.87	−0.142	4.73
Dog	20	68.33	−0.034	−0.56	81.15	−0.019	−0.33	70.31	−0.063	1.17	80.24	−0.019	−0.38
	26	78.29	−0.022	−0.45	84.42	−0.018	−0.27	71.50	−0.054	3.24	83.25	−0.021	−0.11
	32	82.43	−0.051	−0.91	84.79	−0.034	−0.31	72.03	−0.077	5.06	83.84	−0.033	−0.09
	38	83.93	−0.193	0.26	85.88	−0.049	−0.14	71.98	−0.097	7.98	85.14	−0.049	0.16
	Average	78.24	−0.041	1.67	84.06	−0.019	0.83	71.46	−0.191	7.51	83.12	−0.026	1.00
Average		**77.63**	**−0.222**	**8.32**	**82.10**	**−0.079**	**3.24**	**68.52**	**−0.220**	**8.71**	**81.80**	**−0.094**	**3.62**

*Only view 1, 3, 5 are tested among all odd views.

time parameters) and IX (for the other sequences). Three evaluation criteria are used, as TS (%) for time reduction, ΔPSNR (dB) for PSNR increase and ΔBR (%) for bit rate increase. For each sequence and each Qp, the average results of the odd views are given; To jointly investigate the RD performance, in this table, the average ΔPSNR and average ΔBR are Bjontegaard's average PSNR increase (BDPSNR) and Bjontegaard's average bit rate increase (BDBR) [32], separately.

In Shen's algorithm, GDV prediction is used to predict coding information, which is not applicable for views with only forward inter-view prediction (*i.e.,* view 7 in Table VII). Hence, we compare this algorithm and our hybrid model with view 1, 3, 5 only (marked with * in Tables VIII and IX). From the two tables, Shen's algorithm could achieve high computational complexity reduction in average, with acceptable RD performance. Nevertheless, the time reduction is not so robust compared with that of our method; also, there exists relatively

TABLE IX
SIMULATION RESULTS OF ODD VIEWS FOR THE TESTING SEQUENCES

Sequence	Qp	Shen's [24]			Hybrid Model*			Zeng's [25]			Hybrid Model		
		TS	ΔPSNR	ΔBR	TS	ΔPSNR	ΔBR	TS	ΔPSNR	ΔBR	TS	ΔPSNR	ΔBR
Flamenco1	20	66.13	−0.092	2.23	61.75	−0.084	1.49	77.30	−0.300	8.02	61.64	−0.100	1.67
	26	74.48	−0.118	2.03	67.54	−0.073	1.80	78.23	−0.292	6.85	67.52	−0.099	1.44
	32	81.10	−0.157	0.70	74.75	−0.100	0.40	78.47	−0.235	5.94	74.80	−0.103	0.08
	38	87.20	−0.295	−0.92	81.41	−0.134	−0.34	78.19	−0.212	5.82	81.39	−0.133	−0.23
	Average	77.23	−0.195	4.66	71.36	−0.130	3.12	78.05	−0.524	13.71	71.34	−0.135	3.31
Golf1	20	76.15	−0.085	0.44	90.85	−0.117	0.46	85.07	−0.085	0.33	90.49	−0.121	0.96
	26	79.13	−0.077	0.29	90.35	−0.015	1.06	84.74	−0.053	0.06	90.12	−0.039	1.47
	32	82.86	−0.028	−0.14	93.78	−0.025	0.87	83.45	−0.012	−0.22	93.42	−0.047	0.65
	38	88.64	−0.007	0.71	93.96	−0.019	2.26	81.29	−0.004	−0.13	93.65	−0.053	1.22
	Average	81.70	−0.056	1.30	92.24	−0.077	1.67	83.64	−0.032	0.69	91.92	−0.104	2.22
Jungle	20	51.88	−0.038	1.35	67.11	−0.038	1.80	60.67	−0.131	4.64	66.48	−0.045	1.91
	26	65.25	−0.097	2.64	71.94	−0.064	2.15	64.55	−0.185	5.43	71.39	−0.075	2.21
	32	73.57	−0.211	3.57	76.71	−0.090	1.85	67.27	−0.235	5.71	76.28	−0.103	1.91
	38	79.25	−0.352	3.75	81.21	−0.111	1.60	69.89	−0.264	5.99	80.84	−0.133	1.52
	Average	67.49	−0.267	7.87	74.24	−0.141	4.08	65.59	−0.398	11.70	73.75	−0.156	4.45
Lovebird1	20	80.42	−0.066	0.25	84.72	−0.030	0.33	84.80	−0.128	2.31	84.42	−0.029	0.39
	26	85.67	−0.036	−0.28	86.95	−0.022	0.37	83.43	−0.063	1.86	86.69	−0.026	0.34
	32	88.43	−0.032	−0.36	90.28	−0.024	0.45	82.07	−0.053	2.75	90.10	−0.028	0.46
	38	89.45	−0.048	−0.61	92.26	−0.042	−0.14	81.35	−0.061	4.39	92.05	−0.047	0.02
	Average	85.99	−0.031	0.89	88.55	−0.038	1.10	82.91	−0.150	4.63	88.32	−0.041	1.26
Uli	20	55.13	−0.037	1.06	68.48	−0.040	1.94	64.84	−0.111	3.19	67.89	−0.041	1.85
	26	69.35	−0.077	1.58	71.92	−0.055	1.94	67.04	−0.146	3.61	71.47	−0.059	1.87
	32	76.65	−0.172	2.72	76.77	−0.072	1.80	69.01	−0.192	3.26	76.35	−0.082	1.71
	38	81.14	−0.289	2.64	81.20	−0.089	1.42	71.00	−0.196	2.72	80.93	−0.113	1.35
	Average	70.56	−0.206	6.13	74.59	−0.126	3.65	67.97	−0.278	8.14	74.16	−0.132	3.79
Average		**76.59**	**−0.151**	**4.17**	**80.20**	**−0.102**	**2.72**	**75.63**	**−0.276**	**7.77**	**79.90**	**−0.114**	**3.01**

*Only view 1, 3, 5 are tested among all odd views.

high RD performance loss in most cases as well as the average performance. The reasons why this algorithm could not always achieve good performance are as follows. First of all, in this algorithm, the early Skip mode termination method is developed based on JMVM platform (the multiview reference software before JMVC), in which motion skip mode [33] and some other techniques are enabled to highly improve the overall coding performance. Thus, the early termination after Skip mode would not cause large RD loss due to the existence of motion skip mode. In JMVC, the aforementioned techniques are excluded, and in such a case, the early termination scheme designed for JMVM would cause relatively large quality degradation. Secondly, as a frequently-used mode decision method, the early Skip mode termination would result in unstable time reduction performance for various sequences with different motions, textures and coding parameters, such as *Breakdancer* in Table VIII and *Jungle* in Table IX. Thirdly, this method could not work for view 7 due to limitation of GDV prediction.

Similar conclusions could also be drawn with Zeng's algorithm, in which ME/DE selection is not included for fair comparison in mode level. In this algorithm, RD cost based early Skip mode termination is also developed on JMVM platform, which could not always achieve robust and efficient time reduction with good RD performance in JMVC encoder, as discussed above. In addition, the algorithm is designed for the views with both forward and backward inter-view predictions, and the thresholds are also derived based on this fact, which may not work well for views with only forward prediction (*i.e.,* view 7 in Table VII). Finally, the overall performance could still be further improved because mode classification in this algorithm is based on neighbor prediction and without feedback and adaptation, which may results in error propagation when prediction error exists.

TABLE X
COMPARISON OF DECISION ACCURACY (%) IN MODE DECISION

Algorithms		Shen's [24]	Zeng's [25]	Hybrid Model
Qp	20	95.63	86.33	95.40
	26	96.88	92.05	96.79
	32	97.51	95.06	97.57
	38	98.00	96.98	98.23
view	1	96.98	93.55	97.19
	3	96.82	92.97	96.98
	5	97.22	92.27	96.88
	7	–	91.62	96.94
Average		97.00	92.60	97.00

To avoid the limitation of reference software, our method is designed with a theoretical and adaptive model, which does not depend on coding techniques and prediction structures. Compared with the above two algorithms, the time reduction of our method is more robust, and it is closer to the aforementioned ideal curves in Fig. 2. This is mainly due to investigation of

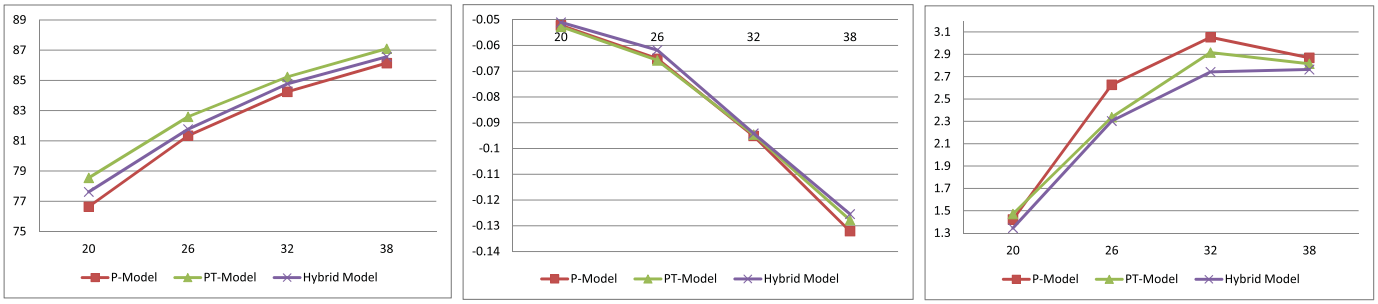


Fig. 3. Comparison of P-Model, PT-Model and hybrid model in terms of average TS, ΔPSNR and ΔBR. Horizontal axis: Qp. Vertical axis. (a) TS (%). (b) ΔPSNR (dB). (c) ΔBR (%).

mode characteristics, including both mode probabilities and time proportions, with optimal stopping theory. In addition, the scheme selection in hybrid model provides a better tradeoff between time reduction and coding efficiency, and thus the RD performance is kept almost intact for all sequences and Qp settings. In average, our method could achieve 81.12% time reduction, with −0101 dB BDPSNR and 3.40% BDBR, which could justify the efficiency and robustness of the proposed hybrid model.

B. Decision Accuracy and Computational Overhead

In mode decision, the decision accuracy is not directly related to RD performance. Nevertheless, it can be another measurement in a single mode decision process, especially for our theoretical model, which aims to achieve a good tradeoff between computational complexity reduction and final decision performance. In Table X, the average decision accuracies of Shen's [24], Zeng's [25] and our method are evaluated with four sequences, as *Flamenco1* (320 × 640), *Ballroom* (640 × 480), *Ballet* (1024 × 768) and *Champagnetower* (1280 × 960). From the table, our method could achieve an average decision accuracy of 97.00%, which is similar to Shen's algorithm[1] (with the decision accuracy of view 7 not available) and better than Zeng's algorithm. Another fact could be noticed that for all these algorithms, the decision accuracy increases when Qp gets larger. This is mainly because as Qp increases, more MBs would be coded with large partitions, and the early termination after Skip (in Shen's and Zeng's algorithms) are more probable; also, the mode probabilities are more centralized, which would results in smaller K_*^p/K_*^{pt} in hybrid model.

To further justify the efficiency of our method, the computational overhead is also evaluated. As shown in Section III-B, our method consists of five major steps, as parameter estimation, model prediction, model selection, mode decision and parameter update. Among all these steps, only Step 3 (mode decision) is for complexity reduction and the other steps may bring computational overhead. To evaluate this, the computational time of these steps are examined and shown in in Table XI (since the even views are not optimized, the overhead of hybrid model is not available for these views), with the same sequences to Table X. Compared with the overall algorithm, the average computational overhead is 0.150% for parameter estimation and update (Step 1, 5) and 0.017% for hybrid model prediction and selection (Step 2, 3). These figures are even smaller compared with the original algorithm, considering more than 80% encoding time could be saved in our method. Hence, the hybrid model could achieve a good tradeoff between computational complexity and decision performance with a negligible computational overhead.

TABLE XI
COMPUTATIONAL OVERHEAD (%) IN HYBRID MODEL

Module		Paramter: Step 1, 5	Model: Step 2, 3
Qp	20	0.146	0.012
	26	0.116	0.014
	32	0.148	0.018
	38	0.191	0.023
view	0	0.080	–
	1	0.181	0.025
	2	0.070	–
	3	0.176	0.026
	4	0.083	–
	5	0.172	0.025
	6	0.152	–
	7	0.288	0.056
Average		0.150	0.017

C. Comparison of Optimal Stopping Models

Finally, to analyze the improvement of hybrid model compared with P-Model [27] and PT-Model, these algorithms are separately evaluated with JMVC encoder [31] and configuration parameters in Table VII. The average results of TS, ΔPSNR and ΔBR are intuitively shown in Fig. 3. From this figure, the proposed hybrid model is remarkably superior to P-Model in all the three criteria, due to adoption of coding time proportions for more time reduction, and hybrid model

[1]From Tables VIII, IX and X, Shen's algorithm is with a similar decision accuracy to our method, but the RD loss of this method is much higher. The reason is, without motion skip mode, early termination after Skip mode would cause large RD loss, as discussed in Section IV-A. Among all incorrect decisions of this method, the probability of false acceptance of Skip mode is 68.74%; and due to high RD cost penalty when false accepting Skip mode [34], the average RD cost is increased by 27.44% when incorrect decision exists. While in our method, the aforementioned two figures are 31.19% and 8.45%, separately.

selection rule for higher expected decision accuracy. Besides, the hybrid model could also achieve obviously better video quality and less bit rates than PT-Model, with similar time reduction (less than 1%), which is acceptable considering the decision accuracy required to keep almost intact RD performance.

V. Conclusion

In this work, to achieve a good tradeoff between time reduction and decision performance in mode decision, a hybrid optimal stopping model is developed and implemented in multiview encoder, with investigation of the mode probabilities and time proportions in inter-view coding mode decision. Exhaustive experimental results demonstrate the efficiency and robustness of our model, which is also superior to P-Model [27] and several other recent algorithms. Besides, this model could also be widely employed including but not limited to mode decision and ME process in video coding.

References

[1] *Advanced Video Coding for Generic Audiovisual Services*, ISO/IEC Standard 14496-10:2005(E), Mar. 2005.

[2] T. Wiegand, G. Sullivan, J. Reichel, H. Schwarz, and M. Wien, "Joint draft 11: Scalable video coding," JVET, Geneva, Switzerland, Tech. Rep. Doc.JVT-X201, Jul. 2007.

[3] K. Mueller, P. Merkel, A. Smolic, and T. Wiegand, "Multiview coding using AVC," ISO/IEC, Bangkok, Thailand, Tech. Rep. Doc.M12945, 2006.

[4] M. Flierl and B. Girod, "Multiview video compression," *IEEE Signal Process. Mag.*, vol. 24, no. 6, pp. 66–76, Nov. 2007.

[5] M. Flierl, A. Mavlankar, and B. Girod, "Motion and disparity compensated coding for multiview video," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 17, no. 11, pp. 1474–1484, Nov. 2007.

[6] M. Merkle, A. Smolic, K. Muller, and T. Wiegand, "Efficient prediction structures for multiview video coding," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 17, no. 11, pp. 1461–1473, Nov. 2007.

[7] Y. Huang, B. Hsieh, S. Chien, S. Ma, and L. Chen, "Analysis and complexity reduction of multiple reference frames motion estimation in H.264/AVC," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 16, no. 4, pp. 507–522, Apr. 2006.

[8] L.-F. Ding, P.-K. Tsung, S.-Y. Chien, W.-Y. Chen, and L.-G. Chen, "Computation-free motion estimation with inter-view mode decision for multiview video coding," in *Proc. 3DTV-Conf.*, May. 2007, pp. 1–4.

[9] L. Shen, T. Yan, Z. Liu, Z. Zhang, P. An, and L. Yang, "Fast mode decision for multiview video coding," in *Proc. Int. Conf. Image Process.*, Nov. 2009, pp. 2953–2956.

[10] W. Zhu, W. Jiang, and Y. Chen, "A Fast inter mode decision for multiview video coding," in *Proc. Inf. Eng. Comput. Sci.*, Dec. 2009, pp. 1–4.

[11] L. Shen, Z. Liu, T. Yan, Z. Zhang, and P. An, "Early SKIP mode decision for MVC using inter-view correlation," *Signal Process., Image Commun.*, vol. 25, no. 2, pp. 88–93, Feb. 2010.

[12] T.-Y. Kuo, Y.-Y. Lai, and Y.-C. Lo, "Fast mode decision for non-anchor picture in multiview video coding," in *Proc. IEEE INt. Symp. Broadband Multimedia Syst. Broadcast.*, Mar. 2010, pp. 1–5.

[13] W. Zhu, X. Tian, F. Zhou, and Y. Chen, "Fast disparity estimation using spatio-temporal correlation of disparity field for multiview video coding," *IEEE Trans. Consumer Electron.*, vol. 56, no. 2, pp. 957–964, May 2010.

[14] L. Shen, Z. Liu, T. Yan, Z. Zhang, and P. An, "View-adaptive motion estimation and disparity estimation for low complexity multiview video coding," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 20, no. 6, pp. 925–930, Jun. 2010.

[15] H. Zeng, K.-K. Ma, and C. Cai, "Mode-correlation-based early termination for multi-view video coding," in *Proc. Int. Conf. Image Process.*, Sep. 2010, pp. 3405–3408.

[16] G. Cernigliaro, F. Jaureguizar, J. Cabrera, and N. Garcia, "Fast mode decision for multiview video coding based on scene geometry," in *Proc. Int. Conf. Image Process.*, Sep. 2010, pp. 3429–3432.

[17] M. Ai and J. Wang, "A fast mode decision algorithm for multiview video coding," in *Proc. Image Signal Process. Int. Congr.*, vol. 7. Oct. 2010, pp. 3252–3257.

[18] B. Zatt, M. Shafique, S. Bampi, and J. Henkel, "An adaptive early skip mode decision scheme for multiview video coding," in *Proc. Picture Coding Symp.*, Dec. 2010, pp. 42–45.

[19] G. C.-C. Chan, J.-P. Lin, and A. C.-W. Tang, "Online statistical analysis based fast mode decision for multi-view video coding," in *Proc. Picture Coding Symp.*, Dec. 2010, pp. 478–481.

[20] X. Liu, K. Sohn, L. T. Yang, and W. Zhu, "Intelligent mode decision procedure for MVC inter-view frame," in *Proc. 13th Int. Conf. Comput. Sci. Eng.*, Dec. 2010, pp. 184–189.

[21] P.-J. Lee, H.-J. Lin, S.-H. Huang, and W.-J. Wang, "A fast mode determination algorithm for multi-view video coding," in *Proc. Consumer Electron. Int. Conf.*, Jan. 2011, pp. 689–690.

[22] Y.-H. Lin and J.-L. Wu, "A depth information based fast mode decision algorithm for color plus depth-map 3D videos," *IEEE Trans. Broadcast.*, vol. 57, no. 2, pp. 542–550, Apr. 2011.

[23] J.-C. Chiang, W.-C. Chen, L.-M. Liu, K.-F. Hsu, and W.-N. Lie, "A fast H.264/AVC-based stereo video encoding algorithm based on hierarchical two-stage neural classification," *IEEE J. Sel. Topics Signal Process.*, vol. 5, no. 2, pp. 309–320, Apr. 2011.

[24] L. Shen, Z. Liu, P. An, R. Ma, and Z. Zhang, "Low-complexity mode decision for MVC," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 21, no. 6, pp. 837–843, Jun. 2011.

[25] H. Zeng, K.-K. Ma, and C. Cai, "Fast mode decision for multiview coding using mode correlation," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 21, no. 11, pp. 1659–1666, Nov. 2011.

[26] Y. Zhang, S. Kwong, G. Jiang, X. Wang, and M. Yu, "Statistical early termination model for fast mode decision and reference frame selection in multiview video coding," *IEEE Trans. Broadcast.*, vol. 58, no. 1, pp. 10–23, Mar. 2012.

[27] T. Zhao, S. Kwong, H. Wang, and C.-C. J. Kuo, "H.264/SVC mode decision based on optimal stopping theory," *IEEE Trans. Image Process.*, vol. 21, no. 5, pp. 2607–2618, May 2012.

[28] J. Gilbert and F. Mosteller, "Recognizing the maximum of a sequence," *J. Amer. Statist. Assoc.*, vol. 61, no. 313, pp. 35–73, 1966.

[29] T. S. Ferguson. (2010). *Optimal Stopping and Applications* [Online]. Available: http://www.math.ucla.edu/∼tom/Stopping/Contents.html

[30] T. S. Ferguson, J. P. Hardwick, and M. Tamaki, "Maximizing the duration of owning a relatively best object," in "Strategies for sequential search and selection in real time," *Contemp. Math.*, vol. 125, pp. 37–57, Mar. 1992.

[31] *Joint Draft 8.0 on Multiview Video Coding*, ISO/IEC Standard ITU-T VCEG/JVT-AB204, 2008.

[32] G. Bjontegaard, "Calculation of average PSNR differences between RD-curves," JVET, Austin, TX, Tech. Rep. VCEG-M33, Apr. 2001.

[33] H.-S. Koo, Y.-J. Jeon, and B.-M. Jeon, "MVC motion skip mode," JVT, San Jose, CA, Tech. Rep. JVT-W081, Apr. 2007.

[34] Y.-C. Lin, T. Fink, and E. Bellers, "Fast mode decision for H.264 based on rate-distortion cost estimation," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, Apr. 2007, pp. 1137–1140.

Tiesong Zhao (M'12) received the B.S. degree in electrical engineering from the University of Science and Technology of China, Hefei, China, and the Ph.D. degree in computer science from the City University of Hong Kong, Kowloon, Hong Kong, in 2006 and 2011, respectively.

He is currently a Post-Doctoral Fellow with the Department of Electrical and Computer Engineering, University of Waterloo, ON, Canada. His current research interests include video coding and perceptual video coding.

Sam Kwong (SM'04) received the B.Sc. and M.Sc. degrees in electrical engineering from the State University of New York at Buffalo, in 1983, and the University of Waterloo, Canada, in 1985, respectively, and the Ph.D. degree from the University of Hagen, Hagen, Germany, in 1996.

He was a Diagnostic Engineer with the Control Data Canada, from 1985 to 1987. He joined the Bell Northern Research Canada, Ottawa, Canada, as a Member of Scientific Staff. He joined the City University of Hong Kong, Kowloon, Hong Kong, as a Lecturer at the Department of Electronic Engineering in 1990, where he is currently a Professor with the Department of Computer Science. His current research interests include video coding, evolutionary algorithms, speech processing and recognition, and digital watermarking.

Dr. Kwong was an Associate Editor of the IEEE TRANSACTIONS ON INDUSTRIAL ELECTRONICS and the IEEE TRANSACTIONS ON INDUSTRIAL INFORMATICS.

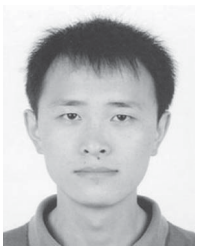
Hanli Wang (M'08–SM'12) received the B.S. and M.S. degrees in electrical engineering from Zhejiang University, Hangzhou, China, in 2001 and 2004, respectively, and the Ph.D. degree in computer science from the City University of Hong Kong, Kowloon, Hong Kong, in 2007.

He is currently a Professor with the Department of Computer Science and Technology, Tongji University, Shanghai, China. He was a Research Fellow with the Department of Computer Science, City University of Hong Kong, from 2007 to 2008. From 2007 to 2008, he was a Visiting Scholar with Stanford University, Palo Alto, CA. From 2008 to 2009, he was a Research Engineer with Precoad, Inc., Menlo Park, CA. From 2009 to 2010, he was an Alexander von Humboldt Research Fellow with the University of Hagen, Hagen, Germany. His current research interests include digital video coding, image processing, pattern recognition, and video content analysis.

Zhou Wang (S'99–A'01–M'02–SM'12) received the Ph.D. degree in electrical and computer engineering from the University of Texas at Austin, Austin, in 2001.

He is currently an Associate Professor with the Department of Electrical and Computer Engineering, University of Waterloo, Waterloo, ON, Canada. He has authored or co-authored more than 100 papers in journals and conferences with over 10,000 citations (Google Scholar). His current research interests include image processing, coding, and quality assessment, computational vision and pattern analysis, multimedia communications, and biomedical signal processing.

Dr. Wang was a recipient of the 2009 IEEE Signal Processing Society Best Paper Award, the ICIP 2008 IBM Best Student Paper Award as a senior author, and the 2009 Ontario Early Researcher Award. He was an Associate Editor of the *IEEE Signal Processing Letters* from 2006 to 2010. He has been a member of the IEEE Multimedia Signal Processing Technical Committee since 2013 and an Associate Editor of the IEEE TRANSACTIONS ON IMAGE PROCESSING since 2009 and *Pattern Recognition* since 2006. He was a Guest Editor of the IEEE JOURNAL OF SELECTED TOPICS IN SIGNAL PROCESSING from 2007 to 2009 and the *EURASIP Journal of Image and Video Processing* from 2009 to 2010. He has been a Guest Editor of *Signal, Image, and Video Processing* since 2011.



Zhaoqing Pan received the B.S. degree (Hons.) in computer science and technology from Yancheng Normal University, Yancheng, China, in 2009. He is currently pursuing the Ph.D. degree with the Department of Computer Science, City University of Hong Kong, Kowloon, Hong Kong SAR.

His current research interests include motion estimation and mode decision in video coding.

C.-C. Jay Kuo (F'99) received the B.S. degree from National Taiwan University, Taipei, Taiwan, in 1980, and the M.S. and Ph.D. degrees from the Massachusetts Institute of Technology, Cambridge, in 1985 and 1987, respectively, all in electrical engineering.

He is currently a Professor of electrical engineering, computer science, and mathematics with the Department of Electrical Engineering and Integrated Media Systems Center, University of Southern California, Los Angeles. His current research interests include digital image and video analysis and modeling, multimedia data compression, communication and networking, and biological signal and image processing. He has authored or co-authored about 200 journal papers, 850 conference papers, and 10 books.

Dr. Kuo is a fellow of the American Association for the Advancement of Science and the International Society for Optical Engineers.