www.ietdl.org

Published in IET Image Processing
Received on 16th October 2009
Revised on 13th July 2010
doi: 10.1049/iet-ipr.2010.0149

ISSN 1751-9659

Optimal model-based complexity control for H.264 video encoding

M.-C. Chien P.-C. Chang

Department of Communication Engineering, National Central University, Taoyuan 320, Taiwan
E-mail: pcchang@ce.ncu.edu.tw

Abstract: An H.264 video encoder adopts multiple encoding tools to achieve high coding efficiency at the expense of high computational complexity. The allowable computational complexity for real-time video encoding, however, is generally limited in a wireless handset. This research proposes a complexity control mechanism that is composed of two algorithms to minimise the distortion of each encoded video frame under the computational complexity constraint and the rate constraint. The first proposed algorithm performs optimal complexity allocation among encoding tools based on a new complexity–rate–distortion (C–R–D) model. This model precisely describes how each encoding tool influences the C–R–D performance of the encoder with concise formulas. Accordingly, the algorithm obtains the optimal complexity of each encoding tool by a closed-form solution with small complexity overhead. Based on a new C–D model of motion estimation, this work proposes the second algorithm that performs optimal complexity allocation among macro-blocks to further allocate suitable complexity to each macro-block. Experiments performed on a software-optimised source code show that these two algorithms yield superior performance to the existing algorithms.

1 Introduction

1.1 Research problem

Applications of real-time video encoding, such as video recording and video conference, are widely equipped in modern wireless handsets. An H.264 video encoder uses multiple encoding tools to achieve superior rate–distortion (R–D) performance at the expense of high computational complexity [1]. The allowable computational complexity of real-time video encoding, however, is generally limited in a wireless handset because the processor has limited computation capability. The computation capability of the processor is further limited if power saving is considered [2, 3]. In addition, the allowable bit rate is also limited due to limited transmission bandwidth or limited storage space. Therefore a complexity control mechanism that well allocates the computational complexity of video encoding under the complexity constraint and the rate constraint is important.

Video quality is composed of temporal quality and spatial quality. The former is determined by the frame rate and the latter is determined by the distortion of each frame. To keep a high enough frame rate for acceptable temporal quality in a complexity constrained environment [4], the allowable computational complexity of encoding a video frame, C_{FCA}, is restricted. This complexity control mechanism aims to control the video encoder so that the distortion of each frame is minimised under the given complexity limit and the given rate limit R_{FC} is expressed as

$$\begin{aligned} &\min\ D \\ &\text{s.t.} \\ &c_F \le C_{FCA} \\ &R_F \le R_{FC} \end{aligned} \tag{1}$$

where D, R_F and c_F denote the distortion, bit rate and computational complexity, respectively, of a frame.

For an H.264 video encoder, a frame is partitioned into a number of macro-blocks (MBs) while an MB is the basic encoding unit. The encoding block diagram is plotted in Fig. 1 [5]. A video encoder mainly adopts motion estimation (ME), motion compensation (MC), intra-prediction, transform (T), quantisation (Q) and entropy coding to encode each MB. T, Q, Q^{-1} and T^{-1} have been collectively denoted as PRECODING [6]. After encoding all MBs in a frame, the video encoder performs deblocking filtering if it is enabled in the frame layer [7].

Each encoding tool has different encoding efficiency from each other [1, 5]. Accordingly, complexity allocation among encoding tools (CAET) is the first key problem for complexity-constrained video coding. After the complexity of each encoding tool for all MBs in a frame is determined, complexity allocation among MBs (CAMB) is the second key

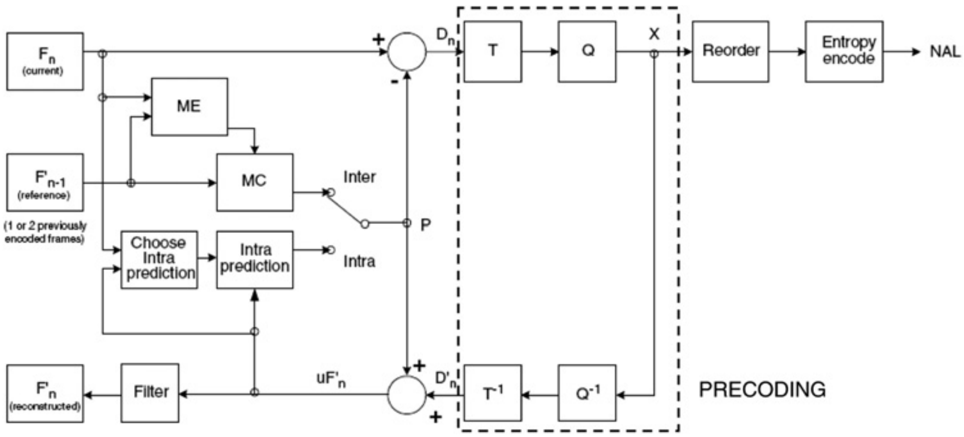

Fig. 1 *Block diagram of the H.264 video encoder*

problem because each MB has different motion and context, and hence deserves different computational complexity.

1.2 Related work

He *et al.* [6] proposed a complexity–rate–distortion (C–R–D) model to address the CAET problem for power saving. This is the first model that can be used to control complexity while maintaining the video quality. However, this model is very complicated. A closed-form solution of the optimisation problem cannot be found. Alternatively, a global search which requires a large computational overhead needs to be conducted to find the solution. The computation overhead of a global search might be acceptable for handset power control which only needs to be performed once per few seconds. For complexity control in real-time video encoding, the complexity allocation should be performed for each frame. The computation overhead of a global search becomes a serious problem. Therefore a simpler C–R–D model which results in low computation overhead is needed in practical applications.

The study [6] also proposed an algorithm of CAMB for ME based on motion history matrix (MHM). This algorithm uses MHM to record the static probability of each MB and then allocates complexity to each MB based on MHM. This algorithm works effectively for low motion videos, but inefficiently for high motion videos because most MBs have zero static probability that results in inaccuracy of the algorithm.

Kannangara *et al.* [4, 8] developed a complexity control algorithm by identifying the MBs that are likely to be skipped prior to motion estimation. By adjusting the SKIP mode proportion, it can satisfy arbitrary computation constraints. However, as Section 5 in this paper will present, the first few searches of ME generally have high coding efficiencies. Therefore skipping an MB sacrifices these coding efficiencies and could hurt R–D performance. A joint complexity–distortion optimisation approach for H.264 under complexity-constrained environment has been proposed [9]. The work proposed a virtual leaky bucket model to prevent encoding a frame from consuming too much or too little computational complexity. The complexity models of the H.264 video decoder, which is a part of encoder, have also been studied [10–14].

1.3 Proposed work

This work proposes a new C–R–D model of the H.264 video encoder. It uses concise mathematical equations to describe how each encoding tool influences the C–R–D performance of the encoder with high accuracy. The optimal CAET problem based on our proposed C–R–D model has a closed-form solution. Accordingly, the optimal complexity of each encoding tool can be obtained with low computation overhead. In addition, this work proposes an optimal algorithm of CAMB for ME based on the modelling of ME behaviour and optimisation theory [15].

1.4 Organisation of paper

This paper is organised as follows. In Section 2, we analyse complexity consumption of each major tool and briefly review the existing C–R–D model for comparison. Section 3 presents our C–R–D model. Section 4 proposes the optimal algorithm of CAET and Section 5 proposes the optimal algorithm of CAMB, respectively. Section 6 describes how to encode video with our complexity control method. Section 7 presents experimental results. Finally, Section 8 draws conclusions.

2 Complexity analysis and existing C–R–D model

2.1 Complexity analysis of encoding tools

Theoretically, a video encoder with higher computational complexity will yield lower video distortion. Therefore the optimal c_F in (1) is presumably set to the limit C_{FCA}. Let c_{MEs}, c_{MCs}, c_{INTRAs}, c_{PRECs} and c_{ENCs} represent the complexity of ME, MC, intra-prediction, PRECODING and entropy coding, respectively, for all MBs in a frame. Also, let c_{DF} and c_{INTP} represent the complexity of deblocking filtering and interpolation, respectively. The allowable computational complexity C_{FCA} can be distributed by

$$C_{FCA} = c_{MEs} + c_{MCs} + c_{INTRAs} + c_{PRECs} + c_{ENCs} + c_{DF} + c_{INTP} \tag{2}$$

Each tool consumes distinct complexity and its complexity control method varies. This work first analyses the complexity consumption of each tool without complexity control. With experiment options listed in Table 1, the average complexity share of each tool is shown in Table 2. To reflect the realistic complexity consumption, we utilise a software optimised code x264 rather than JM code [5, 16, 17]. Rate–distortion optimised (RDO) mode decision is turned off to reduce the complexity [18]. In this work, before the half-pel ME is performed, all the half-pel points

Table 1 Options for complexity analysis

CPU	Intel Pentium 4 2.66 GHz
RAM	512 M bytes
MMX tech.	on for SAD, SATD computation, and 6 tap interpolation filter
source code of H.264	x264
fast ME	diamond
number of reference frames	1
GOP type	IPPPP
RDO mode decision	off
frame rate	30
test sequences	Akiyo, Suzie, Carphone, Foreman, Stefan, Mobile QCIF

in a frame are obtained using 6-tap interpolation filter, which typically consumes significant but almost constant complexity [9]. Therefore the half-pel interpolation is separately counted. On the other hand, the process of quarter-pel interpolation is relatively simple. A quarter-pel point is generated only when it is going to be used for quarter-pel search. Therefore the quarter-pel interpolation is included as a part of ME. As observed in many previous works [9], ME consumes most of the total complexity. MC complexity, which is significant in the decoder [10, 14], is less significant in the encoder.

Deblocking filtering has high coding efficiency as we ever proposed [19], and is beneficial for subjective quality [1]. Thus deblocking filtering is recommended to be utilised. Its complexity c_{DF} can be estimated from the previous frame. Intra-prediction is associated with significant complexity. However, intra-prediction has relatively low coding efficiency in P frames because most MBs in P frames choose inter-mode unless scene change happens. As we also proposed [19], Intra_4 $\times$ 4 prediction has very low coding efficiency for QCIF video sequence. This work recommends using it only when complexity is allowable after inter-prediction and Intra_16 $\times$ 16 prediction in P frames. Since Intra_16 $\times$ 16 prediction is important to scene-changed P frames, Intra_16 $\times$ 16 prediction which consumes roughly 4% complexity is recommended to be used. Therefore in solving CAET problem, we simplify the problem by focusing on complexity allocation among ME, PRECODING and entropy coding

$$C_{\mathrm{FC}} = c_{\mathrm{MEs}} + c_{\mathrm{PRECs}} + c_{\mathrm{ENCs}} \tag{3}$$

where C_{FC} denotes the result of subtracting c_{MCs}, c_{INTRAs}, c_{DF} and c_{INTP} from C_{FCA}.

2.2 *Existing C–R–D model*

According to the work proposed by He *et al.* [6] for MPEG4 video encoding, c_{MEs} can be scaled by the number of search points x as

$$c_{\mathrm{MEs}} = C_1 x \tag{4}$$

where C_1 represents the complexity of searching one point. In other words, the complexity of ME can be represented by the number of search points.

Since the PRECODING process for each MB is identical, the PRECODING complexity of an MB, C_2, is treated as a constant. Accordingly, [6] scaled the PRECODING complexity of a frame, c_{PRECs}, by the number of MBs to be encoded by PRECODING, l, as

$$c_{\mathrm{PRECs}} = C_2 l \tag{5}$$

that is, c_{PRECs} can be represented by the number of MBs that need PRECODING, and all other MBs will have zero PRECODING output. The complexity of entropy coding, c_{ENCs}, was modelled proportional to the bit rate R as

$$c_{\mathrm{ENCs}} = C_3 R \tag{6}$$

where C_3 represents the complexity of encoding one bit by entropy coding. We show that this model is still accurate for both CABAC and CAVLC of the H.264 encoder in Fig. 2. C_1, C_2 and C_3 can be regarded as constants for a video encoding system.

Let M denotes the number of MBs in a frame. Let σ_i^2 represent the sum of square difference (SSD) after ME for ith MB and $\{\sigma_i^2 | i = 1, 2, \ldots, M\}$ represent the set of SSD for all MBs in a video frame in ascending order. A C–R–D

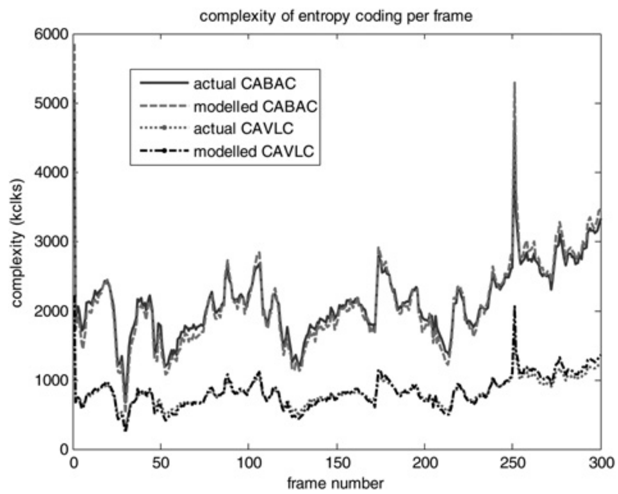

Fig. 2 *Actual and modelled complexity of entropy coding for Stefan sequence when QP is 32*

Table 2 Complexity share (CS%) of each encoding tool

	ME	Intra-prediction	MC	PRECODING	CABAC	Deblocking filter	6-Tap interpolation
CS(QP = 28)	57	9	2	14	10	5	4
CS(QP = 32)	58	11	2	14	6	5	4
CS(QP = 28, no Intra_4 $\times$ 4)	61	4	2	14	10	5	4
CS(QP = 32, no Intra_4 $\times$ 4)	63	4	2	15	7	5	4

doi: 10.1049/iet-ipr.2010.0149

model of PRECODING is proposed as [6]

$$D = \left(\sum_{i=1}^{M} \sigma_i^2\right)\left[\left(1-\frac{l}{M}\right)^2 + \frac{2l}{M}\left(1+\frac{a_0 l}{M}\right)\cdot 2^{-2\gamma(MR_F/l)}\right]$$

$$\text{where } a_0 = \frac{1}{e} + \frac{1}{e^3} - 1 \tag{7}$$

where γ is a model constant. This model is complicated in analysis because the PRECODING complexity l appears in the exponential term.

The work [6] also proposed a C–D model of ME as

$$\sum_{i=1}^{M} \sigma_i^2 = \alpha + \beta e^{-\tau x} \tag{8}$$

where α, β and τ are model constants.

By combining (7) and (8), the first C–R–D model of the video encoder proposed in [6] is

$$D = (\alpha + \beta e^{-\tau x})\left[\left(1-\frac{l}{M}\right)^2 + \frac{2l}{M}\left(1+\frac{a_0 l}{M}\right)\cdot 2^{-2\gamma(MR_F/l)}\right] \tag{9}$$

The optimal CAET problem based on (9) is

$$\min D = \min\left\{(\alpha + \beta e^{-\tau x})\left[\left(1-\frac{l}{M}\right)^2 + \frac{2l}{M}\left(1+\frac{a_0 l}{M}\right)\cdot 2^{-2\gamma(MR_F/l)}\right]\right\}$$

$$\text{s.t.}\quad C_1 x + C_2 l + C_3 R_F = C_{FC} \tag{10}$$

By solving this problem, the optimal ME complexity and PRECODING complexity can be obtained. However, this problem is too complicated for a closed-form solution and should be solved by global searching.

3 Proposed C–R–D model

Since the encoding tools in H.264 is much more complicated than that in MPEG4, the assumptions used in MEPG4 cannot be hold in H.264. We need to verify each step and incorporate the effects of additional tools into the new model.

3.1 C–R–D model of PRECODING

According to the classical R–D theory [20], the R–D model of a quantiser can be expressed by

$$D = \sigma^2 2^{-\gamma R_F} \tag{11}$$

where σ^2 represents the variance of a frame and γ is a model constant. By testing various video sequences, we discover that the R–D model of PRECODING can be modelled as

$$D = \left(\sum_{i=1}^{M} \sigma_i^2\right) 2^{-\gamma R_F} \tag{12}$$

where $\Sigma_{i=1}^{M} \sigma_i^2$ represents the total residual signals measured by SSD in a frame. Fig. 3 shows the R–D curves of Foreman sequence. The PRECODING process for each MB whose prediction mode is not Intra_16 × 16 is the same. For the Intra_16 × 16 MB, additional 4 × 4 luma DC coefficients should be transformed and quantised. We observe that this additional process just increases 4% complexity. Therefore the PRECODING complexity of an MB is almost a constant. Thus, encoding an MB with larger residual signal by PRECODING is more efficient since larger distortion can be avoided with the same computational complexity. In the condition that only l MBs is allowed for PRECODING, an effective strategy is to select the l MBs with greater residual signals for PRECODING [6, 21]. The MB without PRECODING is still inter-predicted and intra-predicted but no bit is allocated to the residual signal. Therefore the distortion of the $M-l$ MBs with smaller residual signals can be degenerated to $\Sigma_{i=1}^{M-l} \sigma_i^2$. This research proposes an R–D model of PRECODING for a frame as

$$D = \sum_{i=1}^{M-l} \sigma_i^2 + \left(\sum_{i=M-l+1}^{M} \sigma_i^2\right) 2^{-\gamma R_F} \tag{13}$$

which can be rewritten as

$$D = \left(\sum_{i=1}^{M} \sigma_i^2\right)\left[\frac{\Sigma_{i=1}^{M-l}\sigma_i^2}{\Sigma_{i=1}^{M}\sigma_i^2} + \left(1 - \frac{\Sigma_{i=1}^{M-l}\sigma_i^2}{\Sigma_{i=1}^{M}\sigma_i^2}\right) 2^{-\gamma R_F}\right] \tag{14}$$

The curve of σ^2 against MB index has been modelled as a linear equation [6]

$$\sigma_i^2 \simeq Ai \tag{15}$$

where A is a constant. The linear relationship of this model is confirmed by our experimental results shown in Fig. 4 except for those MBs with the largest SSD. Note that the ratio of the residual of MBs without PRECODING to the residual of

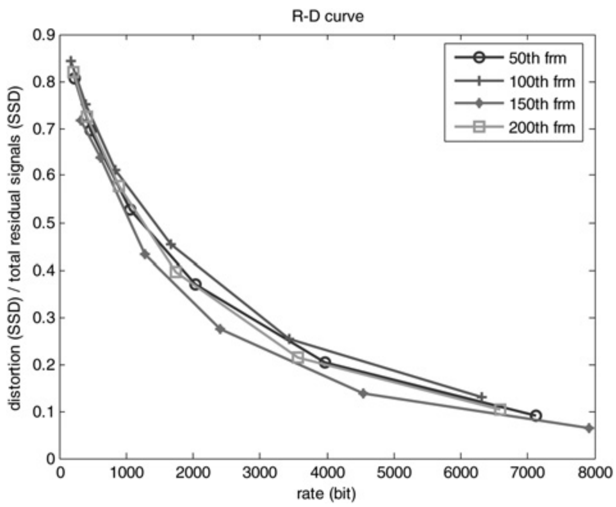

Fig. 3 *R–D curves of PRECODING for various frames in Foreman*

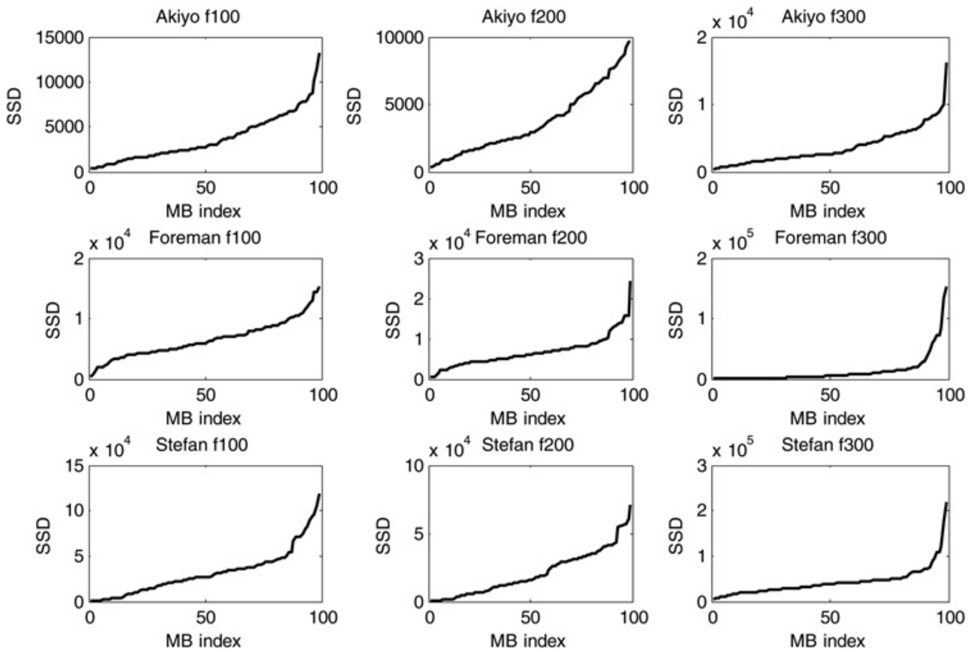

Fig. 4 *Sorted residual signals against MB index in a frame*

total MBs can be formulated as a quadratic function of l

$$\frac{\Sigma_{i=1}^{M-l}\sigma_i^2}{\Sigma_{i=1}^{M}\sigma_i^2} = \frac{A\Sigma_{i=1}^{M-l} i}{A\Sigma_{i=1}^{M} i} = \frac{(M-l)(M-l+1)/2}{M(M+1)/2} \simeq \frac{(M-l)^2}{M^2} = \left(1-\frac{l}{M}\right)^2 \tag{16}$$

Then the distortion in (14) can be expressed as

$$D = \left(\sum_{i=1}^{M}\sigma_i^2\right)\left[\left(1-\frac{l}{M}\right)^2 + \left(\frac{2l}{M}-\frac{l^2}{M^2}\right)2^{-\gamma R_\mathrm{F}}\right] \tag{17}$$

This is our C–R–D model for PRECODING. As Fig. 5 shows, the accuracy of this model is similar to the existing model of (7) but the complexity of this model is substantially reduced.

3.2 C–D model of ME

As mentioned in Section 2, ME complexity was scaled by the number of search points for MPEG4 video encoder [6]. For the H.264 video encoder, the fractional-pel search is recommended to use sum of absolute transformed difference (SATD) to obtain the cost, whereas the full-pel search uses sum of absolute difference (SAD) [18]. In addition, a fractional-pel search needs interpolation. Therefore the complexity of a fractional-pel search is greater than that of a full-pel search. Using distinct interpolation filters, a quarter-pel search consumes different complexity than a half-pel search does. In our work, we follow (4) but redefine C_1, the ME complexity unit, as the complexity of a full-pel search (CFPS) and x as the number of CFPS. The complexity of a half-pel search and that of a quarter-pel search can then be expressed as a number of CFPS.

By observing the C–D curves of various frames in various video sequences shown in Fig. 6, we find the total residual signals measured by SSD is approximately inversely proportional to ME complexity and propose a concise ME model as

$$\sum_{i=1}^{M}\sigma_i^2 = \alpha + \beta x^{-1} \tag{18}$$

where α and β are model constants. As Fig. 6 shows, the existing model (8) underestimates the coding efficiency when ME complexity is low for many video sequences such as Akiyo. The proposed model (18) is more accurate and simpler in derivation.

3.3 Proposed overall C–R–D model

By combining (17) and (18), we propose a new C–R–D model of the H.264 video encoder as

$$D = (\alpha + \beta x^{-1})\left[\left(1-\frac{l}{M}\right)^2 + \left(\frac{2l}{M}-\frac{l^2}{M^2}\right)\cdot 2^{-\gamma R_\mathrm{F}}\right] \tag{19}$$

As Fig. 7 shows, this model is more accurate and simpler in derivation than the previous model presented by (9). The optimal CAET problem based on the new model is

$$\begin{aligned} &\min D = \min\left\{(\alpha+\beta x^{-1})\left[\left(1-\frac{l}{M}\right)^2 + \left(\frac{2l}{M}-\frac{l^2}{M^2}\right)2^{-\gamma R_\mathrm{F}}\right]\right\} \\ &\quad \text{s.t.} \\ &C_1 x + C_2 l + C_3 R_\mathrm{F} = C_\mathrm{FC} \\ &x > 0 \\ &l \geq 0 \\ &l \leq M \\ &R_\mathrm{F} \geq 0 \\ &R_\mathrm{F} \leq R_\mathrm{FC} \end{aligned} \tag{20}$$

doi: 10.1049/iet-ipr.2010.0149

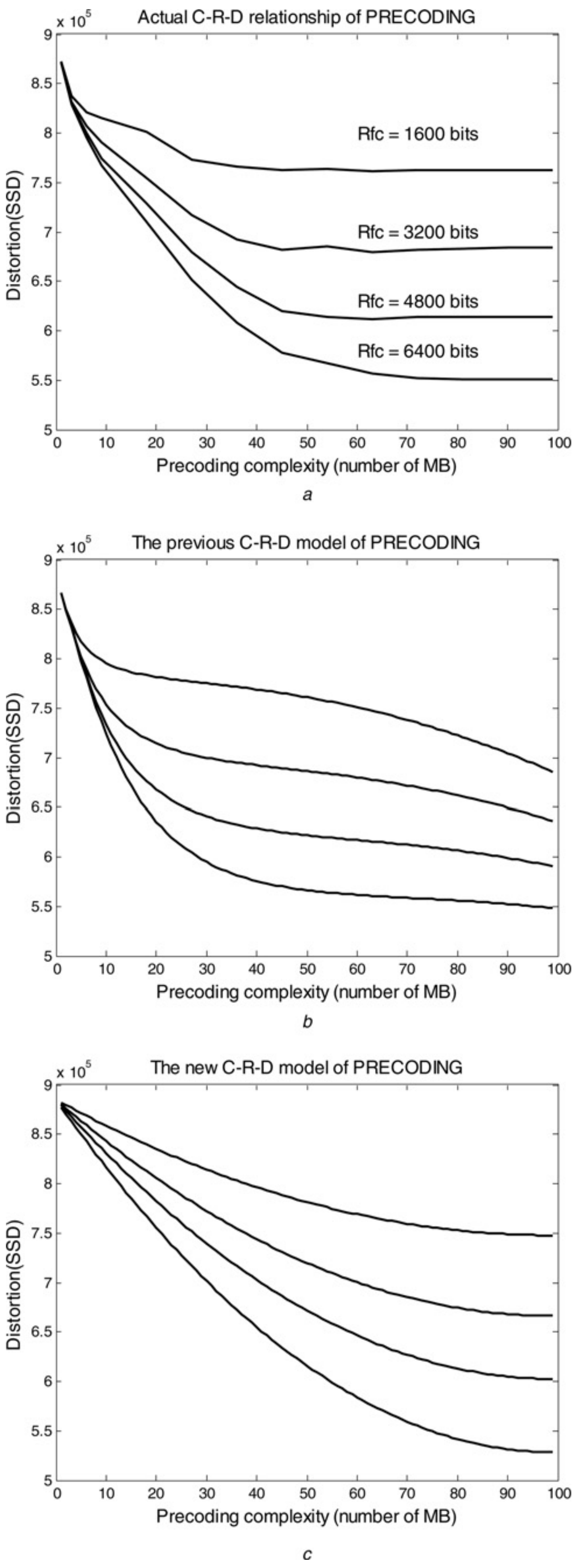

Fig. 5 *C–R–D models of PRECODING for frame 50 in Foreman sequence*

a Actual curve
b Estimated by the model in [6]
c Estimated by our proposed model

Problem (20) is still complex for a closed-form solution because the rate R_F is a variable in the exponential term. In complexity-unconstrained case, the minimal D can be obtained when R_F equals R_{FC}. In complexity-constrained case, a higher bit rate corresponds to higher complexity of entropy coding as described in (6) and the allowable complexity of ME and PRECODING become less. To obtain the optimal rate, experiments are conducted with options shown in Table 1 and the optimum solutions are obtained by global searches [15].

The minimal distortion, optimal x, l and R are plotted in Fig. 8. According to the simulations above, the optimal rate R_F equals R_{FC} if C_{FC} is not exceptionally small, for example, greater than 2 MHz in this simulation. If C_{FC} is exceptionally small, it is not suitable to run a video encoding process in a portable device anyway. Therefore the optimal CAET problem can be simplified as

$$D = (\alpha + \beta x^{-1})\left[\left(1 - \frac{l}{M}\right)^2 + \left(\frac{2l}{M} - \frac{l^2}{M^2}\right)2^{-\gamma R_{FC}}\right]$$
$$\text{s.t.}$$
$$C_1 x + C_2 l + C_3 R_{FC} = C_{FC} \tag{21}$$
$$g_1 \equiv x > 0$$
$$g_2 \equiv l \geq 0$$
$$g_3 \equiv l - M \leq 0$$

Problem (21) can be solved into a closed form as Section 4 presents.

4 Optimal algorithm of CAET

This section solves the optimisation problem (21) according to optimisation theory. The problem is a convex optimisation problem as Fig. 7*c* shows. Because of the convex property, a local minimiser is definitely a global minimiser. To obtain a local minimiser, we should consider that each inequality constraint is active ($g_i = 0$, $i = 1, 2, 3$) or inactive separately [15]. As Fig. 8 shows, x is greater than 0 in general cases. l is also greater than 0 if C_{FC} is not exceptionally small. l is smaller than M if C_{FC} is not sufficiently large and is equal to M otherwise. Accordingly, two general cases are discussed as follows.

4.1 *Case 1: $x > 0$, $l > 0$ and $l < M$*

In this case, the available complexity for encoding a frame is not sufficiently large to encode all MBs with PRECODING. With the derivations for optimal solution, a cubic equation is derived as

$$l^3 - \left(M + \frac{2b}{k} + \frac{\beta}{2\alpha k}\right)l^2 + \left(\frac{b^2}{k^2} + \frac{\beta b}{\alpha k^2} + \frac{2bM}{k}\right)l - \left[\frac{b^2 M}{k^2} + \frac{\beta b M}{\alpha k^2} - \frac{\beta M^2}{(1-a)2\,\alpha k}\right] = 0 \tag{22}$$

where

$$a \equiv 2^{-\gamma R_{FC}} \tag{23}$$

doi: 10.1049/iet-ipr.2010.0149

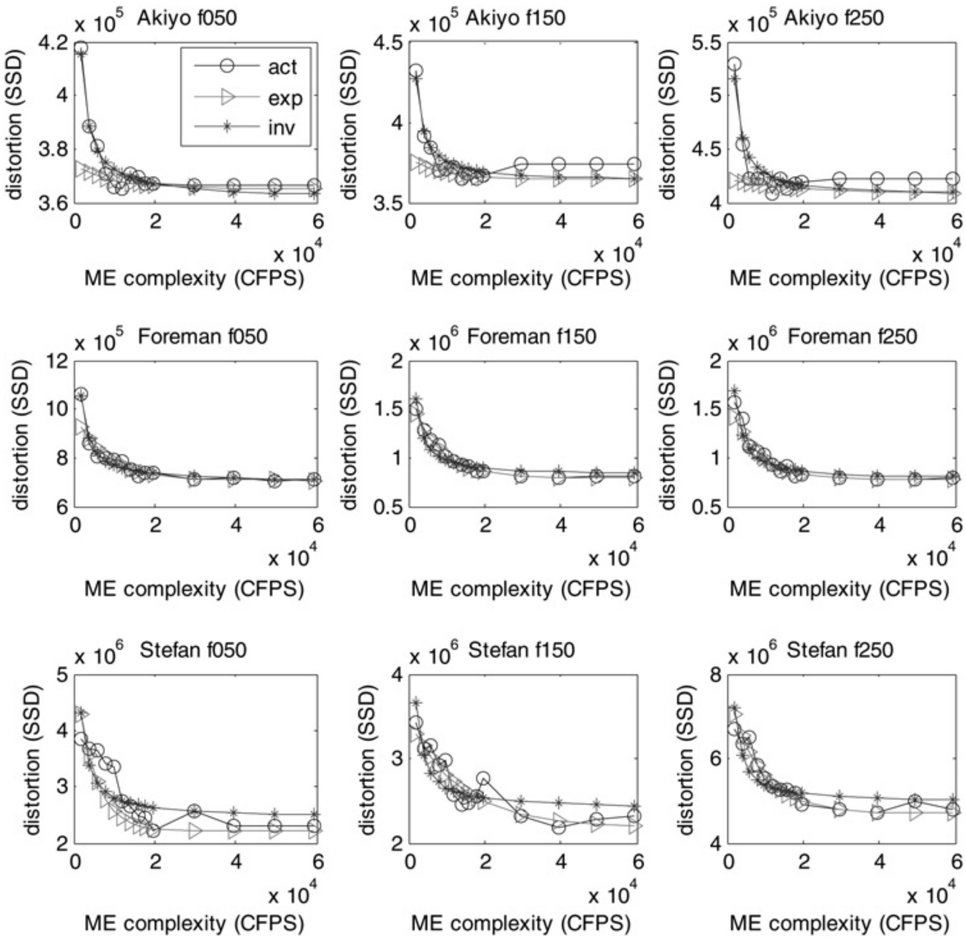

Fig. 6 *C–D curves of ME*

'act': actual curve; 'exp': the curve estimated by the model in [6]; 'inv': the curve estimated by our proposed model

$$b \equiv \frac{C_{\mathrm{FC}}}{C_1} - \frac{C_3}{C_1} R_{\mathrm{FC}} \tag{24}$$

$$k \equiv \frac{C_2}{C_1} \tag{25}$$

Set

$$s \equiv -\left(M + \frac{2b}{k} + \frac{\beta}{2\alpha k}\right) \tag{26}$$

$$u \equiv \left(\frac{b^2}{k^2} + \frac{\beta b}{\alpha k^2} + \frac{2\,bM}{k}\right) \tag{27}$$

$$v \equiv -\left[\frac{b^2 M}{k^2} + \frac{\beta\, bM}{\alpha k^2} - \frac{\beta M^2}{(1-a)2\alpha k}\right] \tag{28}$$

$$p \equiv u - \frac{s^2}{3} \tag{29}$$

$$q \equiv \frac{2}{27}s^3 - \frac{s}{3}u + v \tag{30}$$

One real root of (22) can be obtained [22] as

$$l_1 = \left(\frac{-q + \sqrt{q^2 + (4/27)p^3}}{2}\right)^{(1/3)} + \left(\frac{-q - \sqrt{q^2 + (4/27)p^3}}{2}\right)^{(1/3)} - \frac{s}{3} \tag{31}$$

Obtain x_1 by putting l_1 to the first constraint in (21)

$$x = \frac{C_{\mathrm{FC}}}{C_1} - \frac{C_3}{C_1} R_{\mathrm{FC}} - \frac{C_2}{C_1} l \tag{32}$$

If (x_1, l_1) satisfies $x > 0$, $l > 0$ and $l < M$, it is the solution. Else, the discriminate of (22) defined as (33) is checked.

$$\Delta = 4p^3 + 27q^2 \tag{33}$$

If $\Delta \geq 0$, l_1 is the only real root, go to case 2. Else, divide (22) by $(l - l_1)$, which results in a quadratic equation. Find the two roots l_2 and l_3 of the equation by quadratic formula and obtain x_2 and x_3 by (32). If either pair of them meets $x > 0$, $l > 0$ and $l < M$, it is a global minimiser. Else, go to case 2.

doi: 10.1049/iet-ipr.2010.0149

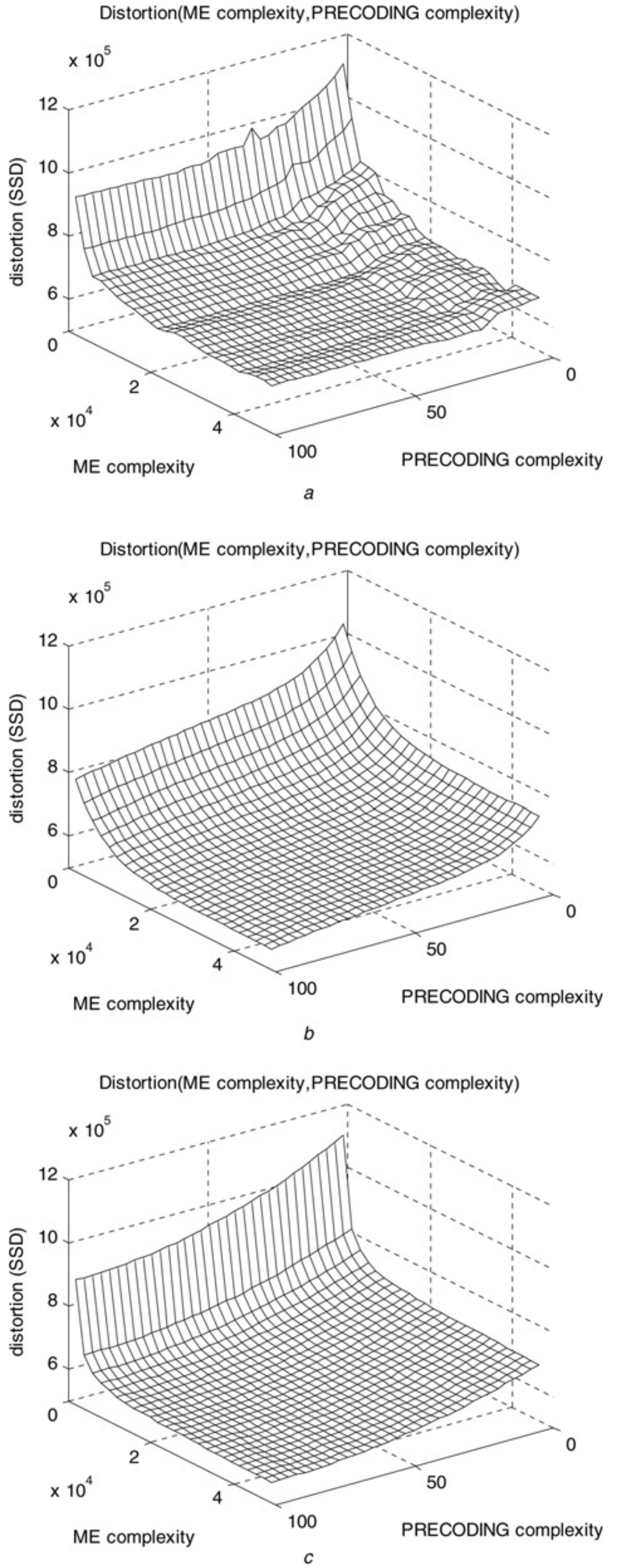

Fig. 7 *C–R–D curves of the overall video encoder for frame 50 in Foreman sequence with $R_{FC} = 3200$ bits*

a Actual curve
b Estimated by the model in [6]
c Estimated by our model

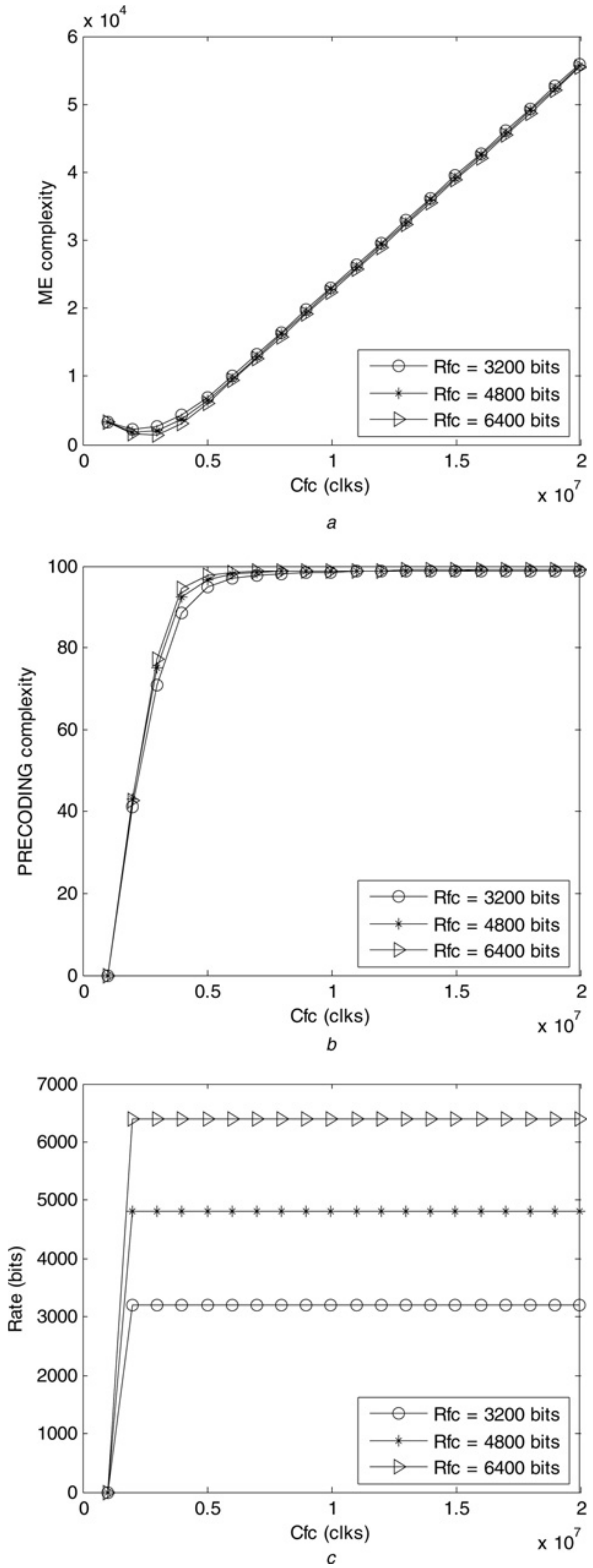

Fig. 8 *Optimal ME complexity, PRECODING complexity and rate under various complexity and rate constraints*

Cfc and Rfc in the figure mean complexity constraint C_{FC} and rate constraint R_{FC} in the article, respectively

4.2 Case 2: $x > 0$, $l > 0$ and $l = M$

In this case, the available complexity for encoding a frame is sufficiently large to encode all MBs with PRECODING, that is, l is equal to M. Obtain x by (32). After CAET is performed for a frame, CAMB needs to be performed to allocate complexity to each MB.

5 Optimal algorithm of CAMB

5.1 CAMB for PRECODING and entropy coding

The complexity of PRECODING of a frame c_{PRECs}, obtained by CAET, must be allocated to the l MBs with larger residual signals before the first MB is encoded by PRECODING. For a general H.264 video encoder, however, the actual residual signals of all MBs cannot be obtained until the last MB is motion compensated and the other MBs are encoded completely. This work discovers that the residual signal of the MB is proportional to that of the co-located MB in the previous frame. Accordingly, this work suggests the l MBs, whose co-located MBs in the previous frame have larger residual signals, are selected for PRECODING.

The residual signal of the l MBs should be quantised with a proper QP to meet the target rate. Once the bit rate is controlled, the complexity of entropy coding is also controlled. The complexity allocation of ME to each MB is the most important part because ME consumes most complexity.

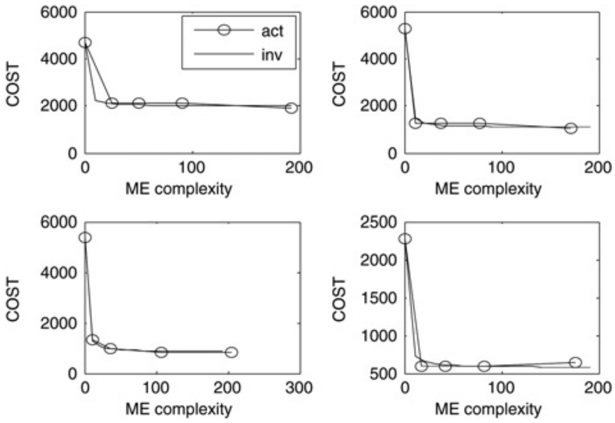

Fig. 9 *Cost against ME complexity curves of various MBs*

'act': actual curve; 'inv': estimated by our model

5.2 CAMB for ME

The general objective of ME is to minimise the R–D cost for each MB [18, 23]. In the complexity-constrained case, rather than minimising the cost individually, the optimal ME is to minimise the cost of all MBs in a frame, which can be formulated as

$$\min f = \min \sum_{i=1}^{M} \operatorname{cost}(i)$$
$$\text{s.t.} \qquad (34)$$
$$\sum_{i=1}^{M} c_i = c_{\mathrm{MEs}}$$

where cost(i) and c_i denote the R–D cost and the ME complexity, respectively, of ith MB, and c_{MEs} is obtained by CAET.

By testing various video sequences, we discover that the C–D model of ME for an MB is also suitable to be

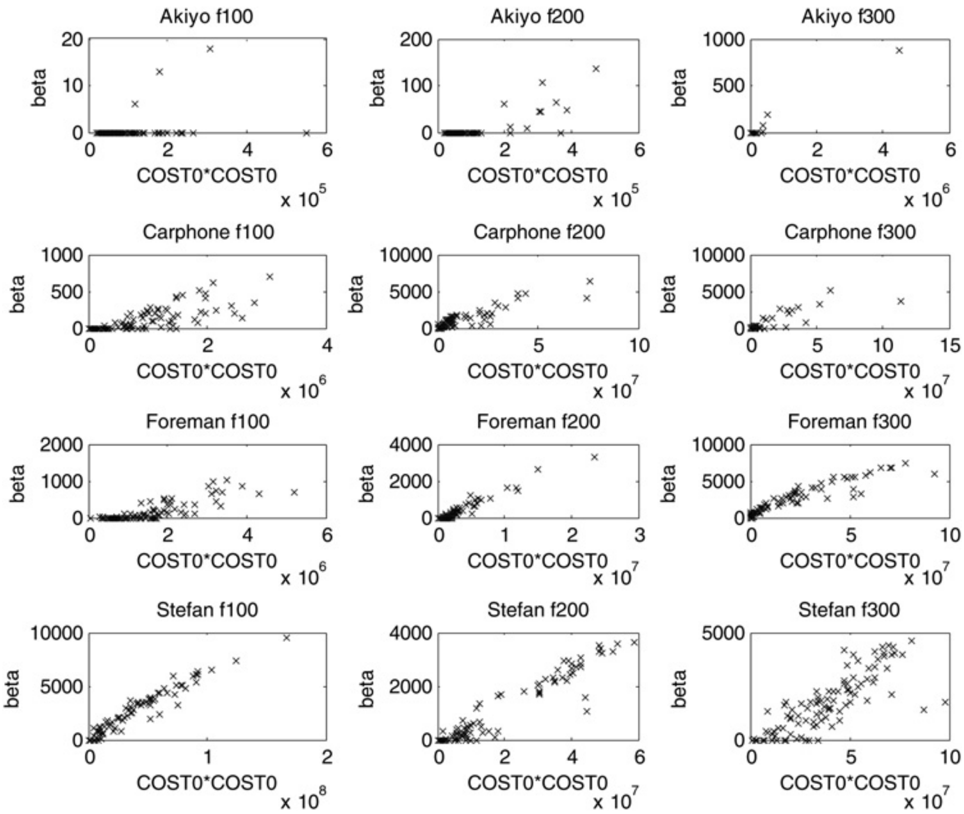

Fig. 10 *β against COST 0^2 curves in a frame*

doi: 10.1049/iet-ipr.2010.0149

modelled as

$$\text{cost}\,(i) = \alpha_i + \beta_i c_i^{-1} \tag{35}$$

Fig. 9 shows the C–D model of Foreman sequence. The parameters α_i, β_i are model constants. α_i represents information and β_i represents redundancy. The optimal complexity allocation can be formulated as

$$\begin{aligned} &\min\ f\ (\boldsymbol{cv}) = \min \sum_{i=1}^{M} (\alpha_i + \beta_i c_i^{-1}) \\ &\text{s.t.} \\ &\sum_{i=1}^{M} c_i = c_{\text{MEs}} \\ &c_i \geq 0, \qquad i = 1, 2, \ldots, M \\ &\text{where} \quad \boldsymbol{cv} \equiv [c_1, c_2, \ldots, c_M] \end{aligned} \tag{36}$$

Equation (36) is a convex optimisation problem. Owing to this property, a local minimiser is a global minimiser which can be obtained as follows

$$\frac{\partial f}{\partial c_i} + \lambda \frac{\partial(\Sigma_{i=1}^{M} c_i - c_{\text{MEs}})}{\partial c_i} = 0, \quad i = 1, 2, \ldots, M \tag{37}$$

which can be derived to

$$-\beta_i c_i^{-2} + \lambda = 0, \quad i = 1, 2, \ldots, M \tag{38}$$

Therefore

$$\lambda = \beta_i c_i^{-2} = \beta_1 c_1^{-2}, \quad i = 1, 2, \ldots, M \tag{39}$$

$$c_i = \left(\frac{\beta_i}{\beta_1}\right)^{(1/2)} c_1, \quad i = 1, 2, \ldots, M \tag{40}$$

Since

$$\sum_{i=1}^{M} c_i = c_{\text{MEs}} \tag{41}$$

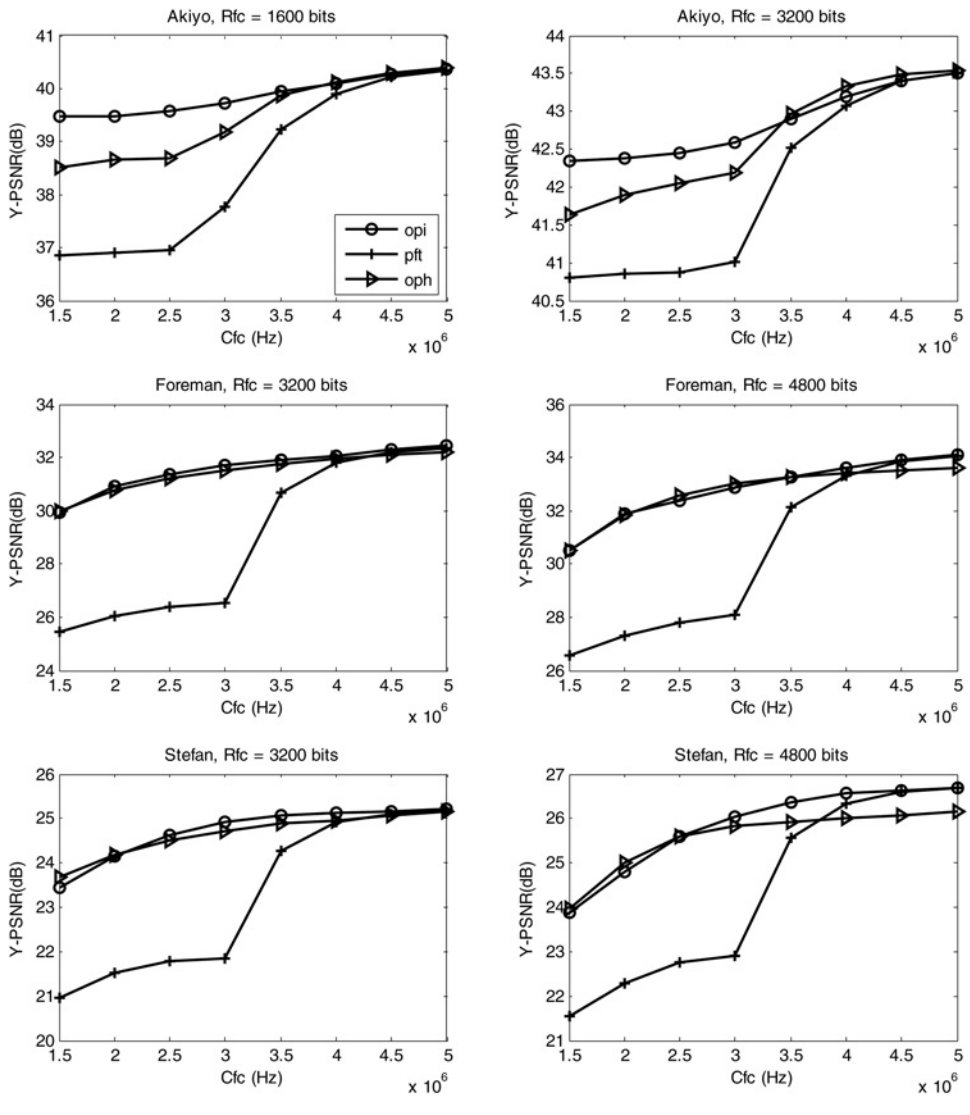

Fig. 11 *Comparison of CAET algorithms for sequences Akiyo, Foreman and Stefan*

Cfc in the figure means complexity constraint C_{FC}

the complexity of an MB can be expressed as

$$c_j = \frac{\beta_j^{(1/2)}}{\sum_{i=1}^{M}\beta_i^{(1/2)}} c_{\text{MEs}}, \quad j = 1, 2, \ldots, M \tag{42}$$

Equation (42) is our optimal algorithm of CAMB for ME. However, obtaining β_i takes much complexity overhead, which is not good for real-time video encoding. Let COST 0_i denotes the R–D cost of ith MB with zero motion vector. According to our extensive research shown in Fig. 10, β_i can be approximated in terms of COST 0_i for most frames in most video sequences as

$$\beta_i \simeq B \times (\text{COST}\,0_i)^2 \tag{43}$$

where B is a constant. Even though (43) is not exactly accurate for a few frames in Akiyo, it is still acceptable because β_i is still roughly proportional to COST 0^2 for MBs in these frames. Experimental results shown in Section 7 reveal that (43) is still a good approximation for these frames. Therefore (42) can be derived to

$$c_j = \frac{\text{COST}\,0_j}{\sum_{i=1}^{M}\text{COST}\,0_i} c_{\text{MEs}}, \quad j = 1, 2, \ldots, M \tag{44}$$

COST 0_i can be obtained simply by comparing the current MB to the co-located MB in the previous frame which can be performed before encoding the first MB in the current frame and the result can be saved for ME. Therefore (44) consumes very little overhead and is a very practical algorithm.

After all inter-prediction modes have been done, Intra_4 × 4 prediction can be performed if the consumed complexity is less than the budget. An H.264 video encoder provides nine modes for Intra_4 × 4 prediction. After each prediction mode is performed, the consumed complexity is checked to determine whether to perform the next mode.

6 Complexity-aware video encoding process

The video encoding with the proposed complexity control method operates as follows:

- *Step 1. Determine the complexity of each essential tool:* The complexity of each essential tool c_{MCs}, c_{INTRAs}, c_{DF} and c_{INTP} are measured by offline test. The values of C_1, C_2 and C_3, the number of CFPS a half-pel search consumes and that a quarter-pel search consumes are also measured by offline test. The subtotal complexity for ME, PRECODING and entropy coding for all MBs, C_{FC}, is obtained according to (3).
- *Step 2. Determine the model parameters:* Before encoding each frame, the ME model constants α and β in (21) are estimated from the statistics of previous frames using linear regression [24]. The R–D model constant γ is also determined from the statistics of previous frames.
- *Step 3. Determine the complexity for ME, PRECODING and entropy coding for all MBs:* Before encoding each frame, the ME complexity for all MBs and the number of MBs to be encoded by PRECODING, l, are determined using our CAET algorithm. The complexity of entropy coding is allocated using (6).
- *Step 4. Determine the ME complexity and PRECODING complexity for each MB:* Before encoding the first MB in each frame, COST 0 of each MB is obtained. The ME complexity budget of each MB is allocated using (44). The l MBs, whose co-located MBs in the previous frame have larger residual signals, are selected for PRECODING. Using rate control, a proper QP is determined for the target rate. Once the bit rate for this frame is controlled, the entropy coding complexity is also controlled.
- *Step 5. Complexity control:* During ME for each MB, the total complexity consumption after each search should be computed. If the total complexity consumption is less than the budget, ME proceeds. Otherwise, ME stops. After ME process, Intra_4 × 4 prediction is executed if the total complexity consumption is less than the budget. Intra_16 × 16 prediction should be executed because its complexity is reserved at Step 2.

7 Experimental results

7.1 Comparison between the proposed and existing CAET algorithms

Let 'opi' denotes the optimal algorithm proposed by this research and 'oph' denotes the optimal algorithm proposed by He's work [6]. Let 'pft' denotes the simple algorithm that allocates the complexity to PRECODING and entropy coding first. If there is more complexity available, then it is allocated to ME. All the three algorithms are accompanied with the method of CAMB for ME proposed in Section 5, because it yields the best performance as Section 7.2 will present. The experimental environment is shown in Table 1. According to the experimental results, the computation overhead of 'oph' takes 1/40 s per frame. The computation overhead of 'opi' takes only 1/330 s per frame. With the computation overhead being ignored, the experimental results in Fig. 11 reveal the optimal algorithm proposed by this research is superior to other existing algorithms.

7.2 Comparison between CAMB algorithms for ME

Let 'ome' denotes our optimal algorithm of CAMB for ME presented by (44) and 'mhm' denotes the MHM-based method [6]. This experiment compares these two complexity allocation methods with the options shown in Table 3. As experimental results in Fig. 12 show, 'ome' yields better R–D performance than 'mhm', especially when the complexity constraint is strict. The peak signal-to-noise ratio improvement measured at the same rate can be as high as 1 dB for Akiyo sequence and 0.5 dB for Foreman and Stefan sequences.

Table 3 Options for complexity allocation for ME

Sequences	Foreman, Stefan, Akiyo QCIF
C_{MEs} complexity unit	20^*M, 40^*M, 60^*M CFPS
fast ME	Diamond
QP	27,29,31,33,35 for Foreman and Stefan 25,27,29,31,33 for Akiyo

doi: 10.1049/iet-ipr.2010.0149

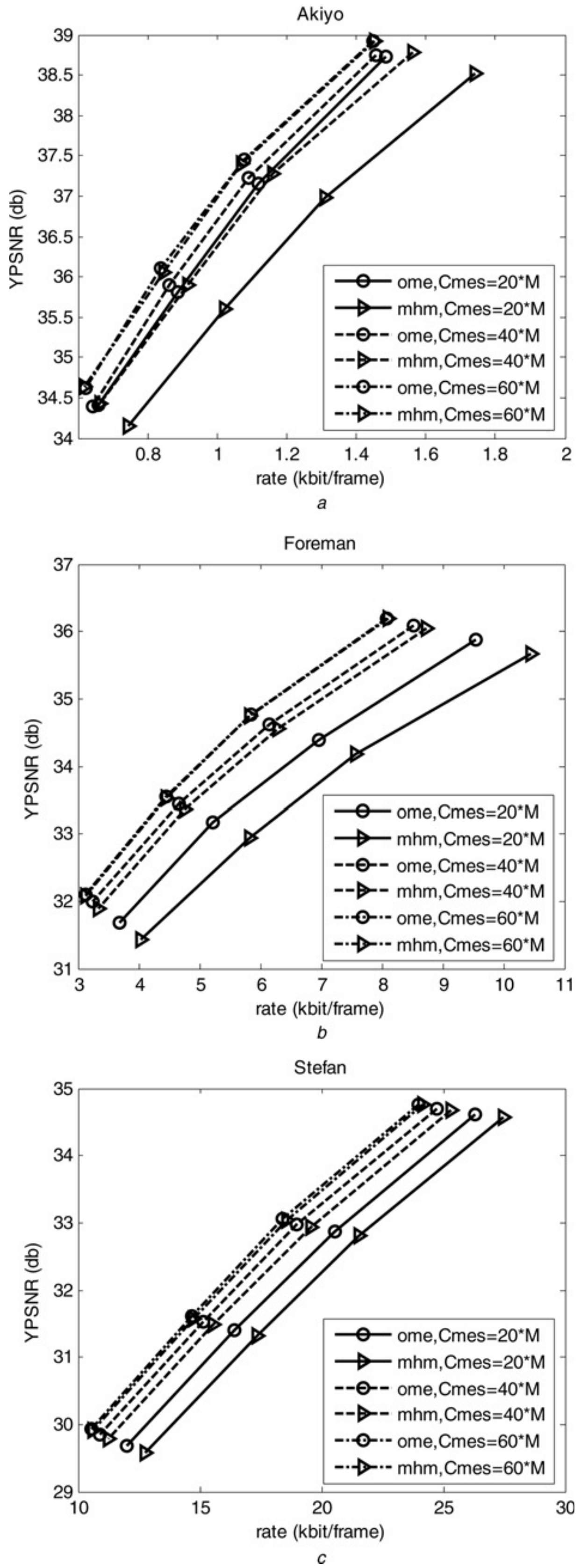

Fig. 12 *Comparisons of the algorithms of CAMB for ME for sequences Akiyo, Foreman and Stefan*

8 Conclusion

This work proposes a concise and accurate C–R–D model for real-time video encoding. Based on this model, we propose an optimal algorithm of CAET which yields better performance than the existing algorithms do with much less complexity overhead. We also propose an optimal algorithm of CAMB for ME which significantly increases the encoder performance under the complexity constraint.

9 References

1 Ostermann, J., Bormans, J., List, P., *et al.*: 'Video coding with H.264/AVC: tools, performance, and complexity', *IEEE Circuits Syst. Mag.*, 2004, **4**, (1), pp. 7–28

2 Min, R., Furrer, T., Chandrakasan, A.: 'Dynamic voltage scaling techniques for distributed microsensor networks'. Proc. IEEE Computer Society Workshop VLSI (WVLSI'00), April 2000, pp. 43–46

3 Chien, M.C., Huang, J.Y., Chang, P.C.: 'Complexity control for H.264 video encoding over power-scalable embedded systems'. Proc. IEEE Symp. on Consumer Electronics, April 2009, pp. 43–46

4 Kannangara, S., Richardson, I.E.G., Miller, A.J.: 'Computational complexity management of a real-time H.264/AVC encoder', *IEEE Trans. Circuits Syst. Video Technol.*, 2008, **18**, (9), pp. 1191–1200

5 Richardson, I.E.G.: 'H.264 and MPEG-4 video compression' (Wiley, 2003)

6 He, Z., Liang, Y., Chen, L., Ahmad, I., Wu, D.: 'Power–rate–distortion analysis for wireless video communication under energy constraints', *IEEE Trans. Circuits Syst. Video Technol.*, 2005, **15**, (5), pp. 645–658

7 ISO/IEC ITU-T Rec. H264: Advanced Video Coding for Generic Audiovisual Services, Joint Video Team (JVT) of ISO-IEC MPEG & ITU-T VCEG, Int. Standard, May 2003

8 Kannangara, S., Richardson, I.E.G., Bystrom, M., Zhao, Y.: 'Complexity control of H.264 based on mode-conditional cost probability distributions', *IEEE Trans. Multimedia*, 2009, **11**, (3), pp. 433–442

9 Su, L., Lu, Y., Wu, F., Li, S., Gao, W.: 'Complexity-constrained H.264 video encoding', *IEEE Trans. Circuits Syst. Video Technol.*, 2009, **19**, (4), pp. 1–14

10 Lee, S.W., Jay Kuo, C.C.: 'Complexity modeling for motion compensation in H.264/AVC decoder'. Proc. IEEE Int. Conf. Image Processing, October 2007, pp. 313–316

11 Lee, S.W., Jay Kuo, C.C.: 'Complexity modeling of spatial and temporal compensations in H.264/AVC decoding'. Proc. IEEE Int. Conf. Image Processing, October 2008, pp. 2504–2507

12 Lee, S.W., Jay Kuo, C.C.: 'Complexity modeling of H.264/AVC CAVLC/UVLC entropy decoders'. Proc. IEEE Int. Symp. Circuits and Systems (ISCAS), June 2008, pp. 1616–1619

13 Kontorinis, N., Andreopoulos, Y., Van Der Schaar, M.: 'Statistical framework for video decoding complexity modeling and prediction', *IEEE Trans. Circuits Syst. Video Technol.*, 2009, **19**, (7), pp. 1000–1013

14 Zia, W., Shafait, F.: 'Complexity reduction techniques for long-term memory motion compensated prediction based on spectral distortion analysis', *PCS*, 2006

15 Chong, K.P., Zak, H.: 'An introduction to optimization' (Wiley, 2001)

16 x264, Available at: http://developers.videolan.org/x264.html

17 Joint Model reference software version 10. Available at: http://iphome.hhi.de/suehring/tml/index.htm

18 'Text description of joint model reference encoding methods and decoding concealment methods'. Munich, Germany, 2004, Joint Video Team (JVT) of ISO/IEC MPEG & ITU-T VCEG Doc.JVT-K049

19 Chien, M.C., Chen, Z.Y., Chang, P.C.: 'Coding-gain-based complexity control for H.264 video encoder'. Proc. IEEE Int. Conf. Image Processing, October 2008, pp. 2136–2139

20 Berger, T.: 'Rate distortion theory' (Prentice-Hall, Englewood Cliffs, NJ, 1984)

21 Chen, Z., Zhou, P., He, Y.: 'Fast integer pel and fractional pel motion estimation in for JVT'. JVT-F017r1.doc, Joint Video Team (JVT) of ISO/IEC MPEG & ITU-T VCEG, 6th Meeting, Awaji, Island, JP, 5–13 December 2002

22 Goldstein, L.J.: 'Abstract algebra' (Prentice-Hall, 1973)

23 Wiegand, T., Schwarze, H., Joch, A., Kossentini, F., Sullivan, G.J.: 'Rate-constrained coder control and comparison of video coding standards', *IEEE Trans. Circuits Syst. Video Technol.*, 2003, **13**, pp. 688–703

24 Lapin, L.L.: 'Modern engineering statistics' (International Thomson Publishing Asia, 1997)