



A perceptual macroblock layer power control for energy scalable video encoder based on just noticeable distortion principle [☆]

Wen Ji ^{a,*}, Min Chen ^b, Xiaohu Ge ^c, Peng Li ^a, Yiqiang Chen ^a

^a Institute of Computing Technology, China

^b School of Computer Science & Engineering, Seoul National University, Seoul 151-744, Korea

^c Department of Electronics and Information Engineering, Huazhong University of Science and Technology, China

ARTICLE INFO

Available online 3 July 2010

Keywords:

Video encoding system
Power control
Just noticeable distortion

ABSTRACT

In most mobile video encoding systems, long battery life and high performance video encoding are competing design goals. This paper proposed a Macroblock (MB) level perceptual energy scalable video encoding method noted as PMP-ESVE, in which just noticeable distortion (JND) model is introduced as the perceptual cue. PMP-ESVE includes three parts, first, PMP-ESVE can dynamic adapt the variable energy resource budget in MB level, second, PMP-ESVE jointly consider the available energy resource and the perceptual feature in order to provide a MB level scalable video encoding method under variable energy consumption budget. Third, JND model, which refers to the maximum distortion that human visual system cannot perceive, is extended from spatial domain to temporal domain so as to determine perceptual cue in unit of MB. This provides the guideline of resource allocation in MB's. Finally, both objective and subjective quality evaluations are given to evaluate the proposed method. These experimental results demonstrate the efficiency of the proposed approach.

© 2010 Elsevier Ltd. All rights reserved.

1. Introduction

With the wide progress of wireless communication, video compression has recently become an important feature of 3G cell phones, portable terminals and other battery-powered devices. An urgent requirement for portable wireless video systems is low power dissipation. As a result, power control and energy control play important roles in these systems. In portable video encoding systems, designing the energy aware systems so as to extend the battery lifetime is an effective way. Since the large computational complexity, video encoding systems need flexible tradeoffs between encoding quality and power consumption. Lian et al. (2007) present a power aware video encoding system by embedding some reconfigurable points inside so as to provide system-level power-aware control. They give the sufficient condition of a power-aware encoding. From the relation among rate, distortion and power consumption aspects, power-rate-distortion (P-R-D) model (He et al., 2004) was

proposed. It analyzes the rate-distortion (R-D) behavior of video encoding system under the energy constraint. Based on this, further research in He et al. (2008) proves power is tightly coupled with rate, thus mapping bits to joules to perform energy minimization is a rapid method to achieve lower energy. These provide the comprehensive analysis for power scalable video encoding. From constructing a energy scalable video encoder aspect, De Schrijver et al. (2006) consider the memory, processing power, and bridge these with the amount of bandwidth which comes from video fragment. These researches demonstrate that energy consumption scalable video encoding becomes a tendency especially in energy constraint applications.

On the other hand, since the goal of video compression and coding aims at the lowest bit rate for signal representation at certain level of perceptual quality, or the highest perceptual quality with a given bit rate, video encoding system based on perceptual cues shows increasing potential. However, low power video encoding system with perceptual consideration has received relatively little attention. The reasons are as follows:

- (1) Perceptual video quality evaluation is a very difficult problem. Current two accepted judgement methods are objective measurement in terms of PSNR or MSE, and subjective measurement in ways of MOS or human vision systems. For the sake of implement difficulty and stability requirement, objective measure becomes the mainstream in practical measurement. This leads to an large improvement space in

[☆]This research was supported by the National Natural Science Foundation of China (NSFC), contract/Grant number: 60872007; National 863 High Technology Program of China, contract/Grant number: 2009AA01Z239; The Ministry of Science and Technology (MOST), China, International Science and Technology Collaboration Program, contract/Grant number: 0903.

* Corresponding author.

E-mail addresses: jiwen@ieee.org (W. Ji), minchen@ieee.org (M. Chen), xhge@mail.hust.edu.cn (X. Ge).

current video encoding systems. For example, both subjective measures and objective video quality measures with perceptual support are based on the human vision systems (HVS), which limits the applications on mobile and portable video systems due to intensive computation. Therefore, the perceptual processing of mobile video encoding over wireless network has not been well studied.

- (2) Even if the quality of mobile video users is defined, it is not clear how video processing energy budget should be allocated to match the HVS. Ideally, the most important goal in portable video encoding systems is low power design. The video encoder is better to produce both lower power consumption and lower bitrates output that is prepared to be transmitted or stored. There is a serious conflict. Hence, due to intensive complexity both in video encoding process and perceptual video quality acquirement, perceptual video encoding in mobile and portable terminals is not a low-hanging fruit.

Instead, the mainstream video coding standards such as H.264/MEPG4 and AVS, provide many flexibilities for each application field to design its own encoder according to its specific compression requirement. On the other hand, significant quality improvement in block-based coding is verified in Yang et al. (2005). It obtains optimal results in rate control through considering the local perceptual cues of input video. The perceptual process relies on the fact that human eyes cannot sense any changes below the just noticeable distortion (JND) threshold (Jayant et al., 1993). Furthermore, JND model is used in motion-compensated module in video encoder and can improve both objective coding quality and perceptual quality of decoded video for a given bit rate. Inevitable, the novel features of perceptual model can significantly improve the coding efficiency but at the expense of high computational complexity. Therefore, reasonable energy resource control in video encoding systems shows increasing requirements.

In this paper, we investigated the JND model in video processing systems and proposed a MB-level perceptual energy scalable video encoding (PMP-ESVE) for energy limited terminals. Since the scalable energy control in video encoding system has multiple reconfigurable encoding results, we investigated how to apply JND as a perceptual criterion in video encoding process, so as to keep low energy consumption in each scalable level while maintain optimal encoding results. We focus on building a perceptual video encoder architecture based on JND model under energy resource constraint, and extend JND to temporal domain so as to make it more suitable to video encoder. As a whole, a scalable and low energy control video encoding system based on perceptual cues is derived.

This paper is organized as follows. After briefly reviewing, Section 2 provide power scalable method in video encoder, which is the foundation of the PMP-ESVE; Section 3 gives the temporal JND model addressed in Section 1; Section 4 builds the perceptual energy scalable video encoder architecture, PMP-ESVE, and gives a solution in video encoding systems; Section 5 evaluates a PMP-ESVE from both subject and objective aspects; and Section 6 concludes.

2. Preliminary of power scalable control in video encoder

Video encoding system is a hybrid architecture. Video encoder operates by removing redundancy in the temporal, spatial and/or frequency, and data domains, separately. Advanced video coding standards including MPEG-4, H.264, AVS and so on are designed for the coded representation. These standards specifically do not define an encoder, they define the output that an encoder must produce, instead. Thus, similar architecture is adopted in major

video encoding standards. It based on the same generic design that incorporates a motion estimation and compensation front end, a transform stage and an entropy encoder. Undoubtedly, advanced codec provides improved compression at the expense of increased complexity, which results in high power and/or energy consumption. For a typical video codec, it has been commonly recognized that the most computational consumption comes from the major modules including motion estimation (ME), motion vector resolution (MVR), inter prediction mode decision (InterMD) with variable block sizes, intra prediction mode decision (IntraMD), discrete cosine transform (DCT), deblocking filter (DF), and entropy coding (EC). Therefore, from functional level, we can summarize the complexity into the following eight aspects: variable block sizes, Hadamard transform, RD-Lagrangian optimization, displacement vector resolution, displacement vector resolution, search range, multiple reference frames, and deblocking filter. For the sake of clarity, in the following discussion, we analyze the general advanced video codec from three aspects, motion estimation, transform and mode decision. For the trust, we employ the video codec based on the AVS standard, the recent video coding standard developed by the Audio and Video Coding Standard Workgroup of China, which promises similar performance but lower complexity compared with H.264. And the conclusion from this discussion is easily applied or transferred to the other video standards such as H.264, MPEG4/2, etc.

2.1. Complexity scalable video encoder

For a given codec, four scalable modules including ME, MVR, InterMD and IntraMD, cover nearly the main features of computation and memory complexity, and the distortion introduced by these modules have close relation with the configuration parameters. Empirical method is adopted in this paper. We estimate the relation between encoding effect and corresponding complexity from the statistic distribution in experiments. The analysis method and corresponding conclusion are easily extended to other video sequences. Next, we give the detail in complexity scalable video encoder through adjusting these four major modules, where each module can provide complexity scalable output.

(1) *Complexity scalable motion estimation module*: It has been well recognized that motion estimation is the most complex and with high computation consumption module. According the previous analysis, ME module can be summarized into two behaviors:

- (1) search range (SR): as we know, better ME search quality need more computation and corresponding power consumption. To search for the best matched MB, full search ME, which search for all possible candidates in a search range, can guarantee the smallest sum of absolute differences (SAD) value, but the complexity of exhausted search is the highest. Different ME algorithms have different search patterns and flows, but they can share the same processing unit of parallel tree to accumulate the SAD (Lian et al., 2007). Increasing both reference frame numbers and search size leads to near 60 times access frequency, while it has a minimal impact on PSNR and bit rate performances. Thus, adjusting SR range in motion estimation module is a rapid complexity-control method.
- (2) motion vector resolution (MVR): the accuracy of motion vectors has three case, these are integer-pixel, fractional pixel in 1/2 and fractional pixel in 1/4. For the encoder, 1/2 pixel search results in a serious increase of access frequency processing time, and 1/4 pixel accuracy increases the processing time about 10% when reduces the bit-rates up to 30%

except at the very low bit-rates. MVR is regarded as an important behavior feature for nearly all ME algorithms. This constitutes the second complexity adjusting way.

(2) *Complexity scalable mode decision module*: It is quite common that new techniques used in video encoder, such as the spatial prediction in intra mode coding, leads to the increase in computational complexity. Mode decision is classified into two categories: inter and intra prediction decisions. For the inter prediction, most of the computations consumed in this stage lies in the variable block size ME employed. In H.264, seven different block sizes (16×16 , 16×8 , 8×16 , 8×8 , 8×4 , 4×8 and 4×4) are supported in inter mode decision. In addition, the SKIP mode, direct mode and two intra modes (INTRA4 and INTRA16) are also supported in H.264. In AVS, there are four different block sizes (16×16 , 16×8 , 8×16 , 8×8) in inter mode decision, besides, the SKIP mode and five intra modes are also employed, including vertical (v) mode, horizontal (h) mode, DC (dc) mode, down-left (dl) mode, and down-right (dr) mode. Thus, to achieve the best encoding efficiency, the encoder usually tries all these possible modes and select the best. Many researches devote in efficient and low complexity mode decision, they often rely on zero blocks detection and early termination techniques or direction detection method so as to decrease the number of mode decisions. As we know, the key of MD module is how to adaptively select the candidate modes before an MB is actually coded. The other fact is that the complexity increases in direct proportion to the number of mode candidates. Consider the case of a mode decision that has total n individual candidates, the mode decision module can produce $\sum_{i=1}^n C_n^i$ output. Therefore, it is possible to reduce the computational complexity without sacrificing the R-D performance by early termination after going through only a few modes in MD module, if the actual best matched mode is in the candidates. Then the solution of MD module comes up to two points: (1) providing a scalable MD output through control the number of candidates; (2) trying to make the actual best matched mode in the candidates. On the other hand, it is noticed that for the mode decision module, the best matched mode does not comply with equal probability distribution. Then, acquire some statistics of the mode distribution beforehand can help to solve the point (2). To solve the two points, we use two steps: (1) orderly arrange the element in each set according to the statistics of the mode distribution; (2) orderly select the modes in each set, under the computation constraint. There are:

- (1) Intra mode decision. The combination of possible intra predictions: the number of candidate intra mode is k , (i.e. $k=5$ in AVS and $k=2$ in H.264). The number of possible outputs are $\sum_{i=1}^k C_k^i$. For example in AVS, we build a simulation platform for intra prediction analysis. All the results of each mode are in worst-case analysis that only one mode is accepted in intra-code, while best mode represents that the encoding are using best match intra-prediction through fully computing all the five modes.
- (2) Inter mode decision. Following the definition of the above case, this case is the combination of possible inter predictions. Given k candidate inter mode, (i.e. $k=4$ in AVS). The number of possible outputs are $\sum_{i=1}^4 C_4^i$. The possible value is the combination of the four different block sizes (16×16 , 16×8 , 8×16 , 8×8).

2.2. Scalable control in video encoder

Therefore, a general video encoder contains a number of configurations for enable scalable complexity control. Each

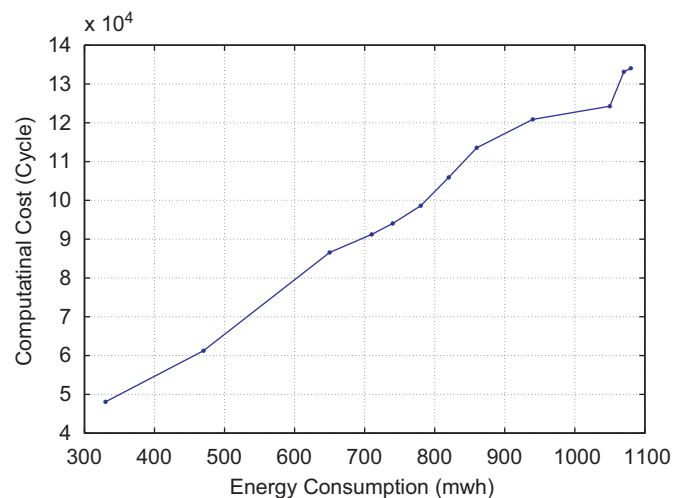


Fig. 1. The relationship between Energy consumption and Computational cost.

configuration represents a working state and corresponding output for encoder. Let $S = \{s_{M_1}, s_{M_2}, \dots, s_{M_m}\}$ denote the set of working state for an encoder. s_{Mi} is a vector and represents there is $s_{Mi} = \{1, \dots, k_{Mi}\}$ working states for module M_i . For example, from the encoder behavior analysis on motion estimation module, if the SR can be adjusted as one of $\{4, 8, 16, 32\}$ and MVR can be adjusted as one of $\{\text{integer pixel}, 1/2 \text{ pixel}, 1/4 \text{ pixel}\}$, then there is 12 working state for motion estimation module. $|s_{ME}| = k_{ME} = 12$.

2.3. The relation between energy consumption and computational cost

It is well accepted that there is relationship among power consumption, energy consumption and computational cost. Since video belongs to period sources, the results on power consumption are closely equivalent to energy consumption in certain aspect. Power constraints are often translated to the encoding computation costs when design power efficient video encoder. Processing unit often use one term of SAD operation or processing cycle measurement. In this paper, we use the latter as the processing unit. Besides, empirical method is introduced to get the relationship based on a given encoder and certain platform. Fig. 1 shows the statistical results, the relation between energy consumption and complexity is given by linear approximation $P = c \times PU$, where PU is in term of processing unit. At the same time, video is periodic signal, such as in 30 or 15 frame/s. Hence, the power consumption is easily transferred to energy consumption during fixed period.

3. Perceptual sensitivity model based on just-noticeable-differences principle

3.1. Stimulus driven sensitivity

Since the ultimate criterion of video signal quality is usually judged or measured by the human receiver, reducing the perceptual redundancy which is inherent in video signals is a trend in video encoding systems. And the human vision is more sensitive to the luminance difference. Based on these facts, a metric of JND (Jayant et al., 1993) is proposed to measure the perceptual distortion caused by the compression algorithm. Chou and Li (1995) incorporates the properties of the HVS into the estimation of the JND profile for measuring the perceptual

redundancies inherent in an image. It gives the expression of JND of the pixel at (x,y) , that is

$$JND(x,y) = \max\{F_1[bg(x,y),mg(x,y)],F_2[bg(x,y)]\} \quad (1)$$

where the subfunction in equal (1) is given by

$$F_1[bg(x,y),mg(x,y)] = mg(x,y) \cdot \alpha[bg(x,y)] + \beta[bg(x,y)] \quad (2)$$

$$F_2[bg(x,y)] = \begin{cases} 17 \cdot [1 - (bg(x,y)/127)^{0.5}] + 3, & bg(x,y) \leq 127 \\ \frac{3}{128} \cdot [bg(x,y) - 127] + 3, & bg(x,y) \geq 127 \end{cases} \quad (3)$$

$$\alpha[bg(x,y)] = bg(x,y) \cdot 0.0001 + 0.115 \quad (4)$$

$$\beta[bg(x,y)] = \frac{1}{2} - bg(x,y) \cdot 0.01, x \in [0, H], y \in [0, W] \quad (5)$$

where $bg(x,y)$ and $mg(x,y)$ represent the average background luminance and the maximum weighted average of luminance differences around the pixel at (x,y) , respectively. H and W are the height and width of the picture, respectively. The calculations of $bg(x,y)$ and $mg(x,y)$ refer to Chou and Li (1995).

3.2. Perceptual integration in MB level

(1) *Extending JND to MB level:* Considering a block-based video encoding system, perceptual control is inclined to be characterized in unit of block or macroblock (MB). Energy control is equivalent to do decisions under certain conditions in unit of MB or in unit of frame, and finally receive benefit or tradeoff under energy resource constraint. JND is introduced in this framework for modeling such a strategic situation with perceptual control. Due to encoding process is partitioned in unit of MB, we extend the JND model around pixel $\{x,y\}$ into MB (i,j) . The expression of the JND of MB (i,j) is

$$JND_{MB}(i,j) = \frac{\sum_{x=1}^{width_{MB}} \sum_{y=1}^{height_{MB}} JND(x,y)}{width_{MB} \cdot height_{MB}} \quad (6)$$

For the sake of clarity, we compare the JND_{MB} results with the original picture and show JND computing results through converting them to gray scale in unit of MB. Fig. 2 shows that the value of JND_{MB} can reflect the local perceptual cues of picture contents.

(2) *Spatial bandpass filtering operation:* Since there are strong relation in spatial domain in image and the JND in unit of MB is the statistical average of pixels, we use a bandpass filtering to continue closed results on a reasonable scale and permit an alias-free reconstruction. It has been found that the bandpass modeled by $f_1=1$ and $f_2=7$ is useful to typical applications of image compression. As shown in Fig. 3 are the JND MB filtering results.

(3) *Extending JND-MB to temporal domain:* Since we aim at the video encoding pre-processing, in order to extend the spatial JND MB profile to video sequence by considering the temporal effects,

we further extend the JND-MB into temporal domain. Considering temporal JND-MB in unit of pixels in DCT domain (Wei and Ngan, 2008), we present temporal JND-MB profile to give a simple but effective prediction. We use $T_{JND_{MB}}$ to represent the temporal variable result and the corresponding temporal contrast sensitivity function is defined as

$$T_{JND_{MB}}(i,j,n) = |T_{JND_{MB}}(i,j,n) - T_{JND_{MB}}(i,j,n-1)| \quad (7)$$

where the (i,j) represents the corresponding index of MB and the n represents the frame number which the MB belongs to. Fig. 4 shows the temporal varieties in unit of frames. Since there are fixed MBs in a frame, we use the statistic results in unit of frame. Fig. 5 provides the gray scale results. Figs. 4 and 5 shows that the temporal domain JND_{MB} results will further reflect the perceptual cues of adjacent frames, which is useful for temporal compression design based on perceptual cues. For example, in sequence 'Foreman', the most fluctuant part begins from 180th frame, which is reflected by $T_{JND_{MB}}$, because the values are larger others. While there are little motion and corresponding little temporal varieties, which is also shown by $T_{JND_{MB}}$ through lower numerical results.

As mentioned above, perceptual energy scalable video encoder relies on the following three facts and a deduction: (1) video encoder has multiple reconfigurable encoding results through adjusting the reconfigurable modules (Lian et al., 2007); (2) basic video encoder consists of finite functional units. Reconfigurable function can be implemented through changing the relationship among complexity, rate and performance in these modules (Saponara et al., 2004; He et al., 2004); (3) for block-based encoder, each frame's encoding process can be regarded as the combination of N MBs' sub-encoding; and the conclusion in Section 2, it is: (4) energy scalable control is transferred into making the encoding into different working states, and each working state corresponds to a different energy consumption budget. Besides, since the basic unit of this energy scalable control is MB, the corresponding allocation scheme is designed among these MBs.

Then each MB is regarded as a sub-unit in energy consumption budget. These MBs compete for the use of a fixed energy resource, which is the target energy consumption budget. Thus, we propose the perceptual energy allocation scheme, in which each MB seeks to choose its encoding energy from whole frame budget so as to maximize the overall utility. The concept of utility refers to the level of satisfaction from that the decision-taker receives. As mentioned above, the criterion of the satisfaction is usually judged or measured by human receiver, thus the effect of human visual properties is considered as a dominant criterion. In many video applications, clients would pay more attention to the visual sensitive regions or pay less attention to the regions under the amplitude of visibility threshold. Yang et al. (2005) developed a new rate control scheme that takes the unequal perceptual

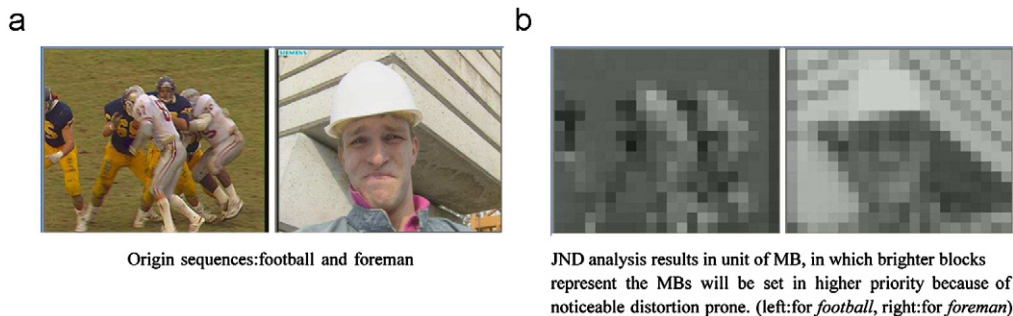


Fig. 2. Illustration of JND analysis masking in unit of MB, by brightness representation.

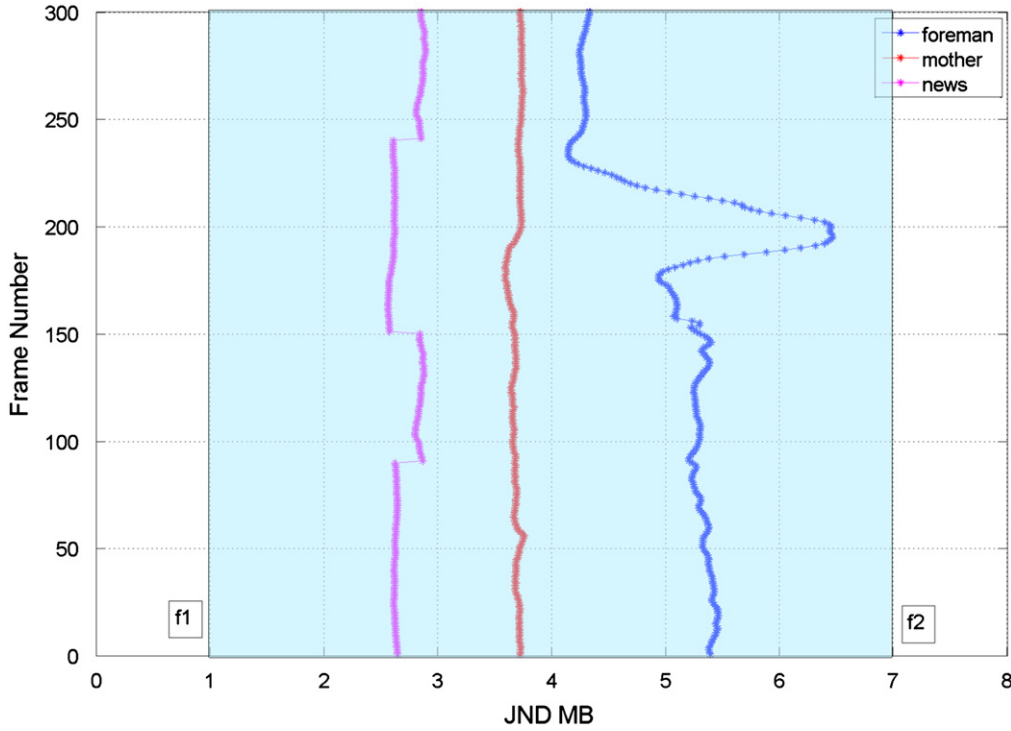


Fig. 3. Band-pass filter of JND_{MB} .

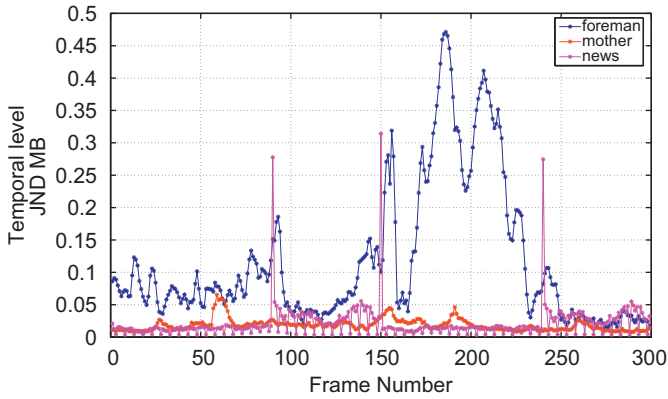


Fig. 4. Temporal JND_{MB} results.

importance between foreground and background into account, and achieved effective quality improvement in block-based coding for bandwidth-hungry applications. This work provides an evidence for the perceptual video encoding based on JND. Due to a tight energy consumption budget, it is anticipant that the more perceptible regions should have higher priority resource allocation under limited energy resource.

4. MB-level power allocation using perceptual cue model

4.1. Problem statement

Our goal includes two parts, one is to provide a scalable energy control encoder so as to adapt the system resource; the other is to find a good configuration that has optimal visual quality under energy constraint. Perceptual control based on JND is introduced in these two part so as to realize the perceptual energy control encoder.

In a video encoder, there are various parameters that can be adjusted to lower and grade the energy consumption (Lian et al., 2007). As mentioned above, a basic video encoder consists of finite functional units: motion estimation, inter prediction, DCT and so on. These module can be analyzed from basic functional units.

In power-rate-distortion optimized video coding, the optimal objective is to find a suitable working state with maximum encoding effect in terms of lower bitrate and lower distortion, while minimum power consumption is in term of computational cost. Given N MBs in a frame, this optimization problem is

$$\begin{cases} \min \sum_N E_{MB}(R, C) \\ \min \sum_N D_{MB}(R, C) \\ \text{s.t. } \sum_N E_{MB}(R, C) \leq E_{budget} \end{cases} \quad (8)$$

where $\sum E_{MB}(R,C)$ represents the total energy consumption of MBs sub-encoding in a frame; $\sum D_{MB}(R,C)$ is the distortion which is equal to frame distortion; and E_{budget} is the power consumption budget for encoding a frame. Obviously, it is a complex multiple objective optimization problem. Here, we divide this complex problem into the combination of single objective problem. Then, when an encoder provides perceptual energy scalable output under the energy consumption budget, it is decomposed to the following steps:

4.2. Energy budget mapping

Energy budget includes two parts. Part 1. manual adjust the working states of video service. These states include ‘Maximum battery time mode’, ‘Battery optimized mode’, ‘Maximum performance mode’ and so on. Each state corresponds to a battery working mode of the device. These states are widely used in mobile devices and terminals. Part 2. automatic aware the working state of video services according to the perception of remaining battery capacity. Since users can manually specify the

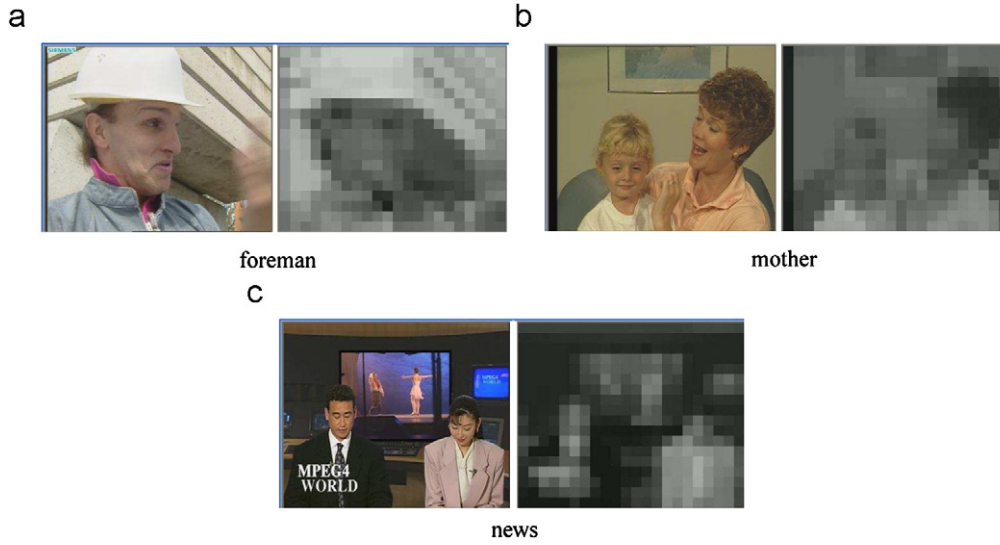


Fig. 5. Illustration of JND analysis masking in unit of MB, by brightness representation.

working state only when the device resource is sufficient, Part 2. has high priority than Part 1. in practice.

4.3. Energy scalable analysis

From the analysis in Section 2, consider the video encoder in unit of MB, the energy consumption is decomposed into each scalable module: (1) complexity scalable motion estimation module; (2) complexity scalable mode decision module; and the residual modules. The energy consumed by each part relies on its working state. There is $S = \{s_{ME}, s_{MD}\}$. As a results, the energy consumption of the MB is

$$E_{MB} = E_{S_{ME}} + E_{S_{MD}} + E_0 \quad (9)$$

where $E_{S_{ME}}$ and $E_{S_{MD}}$ represent the energy consumption of scalable motion estimation module and scalable mode decision module, respectively. E_0 represents the residual energy consumption of unscalable modules. Obviously, each configuration of the modules leads to a corresponding working state. For a general video encoder, the size of the set S trends to very huge value. Therefore, we introduce the pro-processing method through building a encoder information database which stores the best working state of video encoder in statistical. The rules of building this info database include: (1) the evaluation depends on its RD cost; (2) the number of working states for each Qp is in finite and sparse, and sort the energy consumption of these working states in ascending while descending to the values Qp .

4.4. RD-cost analysis in encoder information database

Since scalable control are focus on motion estimation module and mode decision module, the judgment of each adjusting result relies on the relationship between RD-cost function and corresponding computational cost. Sullivan and Wiegand (1998), Hu et al. (2006), Tu et al. (2006) show that the RD-cost functions of motion estimation J_{ME} and mode decision J_{MD} are obtained from:

$$\begin{cases} J_{ME} = D_{DFD} + \lambda_{ME} R_{ME} \\ J_{MD} = D_{Rec} + \lambda_{MD} R_{Rec} \\ \lambda_{ME} = \sqrt{\lambda_{MD}} \\ \lambda_{MD} = 0.85Qp^2 \end{cases} \quad (10)$$

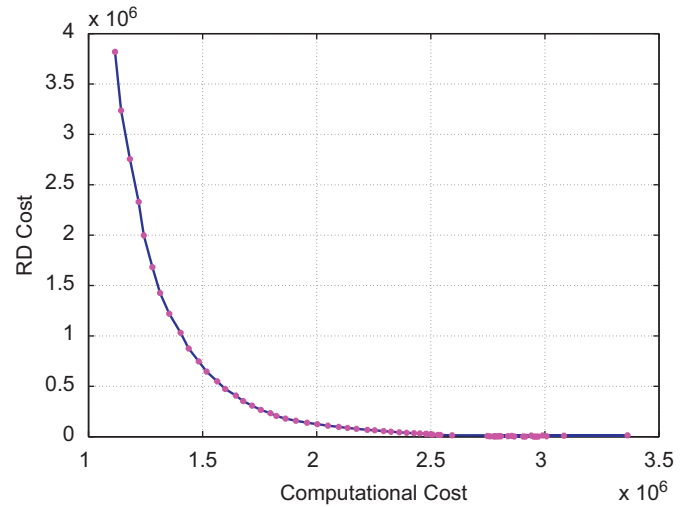


Fig. 6. The relationship between RD-cost and computational cost.

where D_{DFD} is the displaced block difference between current block and reference block; D_{Rec} is the difference between current and reconstructed MB; λ_{ME} and λ_{MD} are the Lagrange multipliers. From the principle of building encoder info database, the relation between RD-cost and computational cost conform to uniform monotonicity. Therefore, the power scalable control can be transferred to $\lambda_{(.)}$ on the criterion of RD-cost, which is easy mapping to the function of Qp . Fig. 6 shows this relation from sequence 'Foreman'. Experiments on other video frames yield similar results. We build the RD-cost and computational cost function from empirical curve fitting method. It is reasonable to model the J_{RD} with deterministic function.

4.5. Energy allocation scheme of MB level based on perceptual cues

As mentioned above, most portable terminals provide many working states such as 'Maximum battery life mode', 'Battery optimized mode', 'Maximum performance mode' and so on. Accordingly, supposing that video encoder can provide corresponding encoding output to match these working states. Each energy profile corresponds to a encoding level and an energy

consumption budget. Then the goal is to adjust the encoding state in unit of MB to obtain best encoding quality under energy consumption budget $E_{budget(k)}, k \in M$. To be simplified, we introduced two categories to present energy allocation scheme. The first category includes the MBs which belong to important perceptual scale and the second category includes the MBs which are with uneasy perceived distortion. In fact the number of category will be many more along with more fine granularity. The partition of the categories keeps consistent with the JND_{MB} results. Let m_f be the number of MBs in first category and m_s be the number of second category. Then the energy allocation under resource constraint problem is formulated as

$$m_f \cdot E_f + m_s \cdot E_s \rightarrow E_{budget(k)} \quad (11)$$

where E_f is the energy consumption budget of a MB in first category, and E_s is the one in second category. On the other hand, even though the resource is limited, we also anticipate that all the MBs in the different perceptual category would have better encoding effect.

For dynamic control, best fair energy allocation can be achieved by accurately calculating the whole JND_{MB} in each frame, and then get the exact E_f and E_s . But it is undeniable because these will lead to additional computational complexity. To be further simplified, we let the relation between E_f and E_s be $E_f = AE_s$, A represents the proportion between these two levels, and $A \geq 1$. From Equal (11), we get

$$\begin{cases} E_f = E_{budget(k)} / \left(m_f + \frac{1}{A} m_s \right) \\ E_s = E_{budget(k)} / (A m_f + m_s) \end{cases} \quad (12)$$

5. Experiment

To evaluate the performance of the proposed method, we implement the proposed perceptual MB Layer power control for energy scalable video encoder (PMP-ESVE) in the AVS (Gao et al.) encoder. In which, perceptual cue is based on the proposed temporal JND principle. Similar performance is expected for the other coding systems and the conclusion from this platform is easy applied or transferred to the other video standards such as H.264/MPEG4, MPEG2. We compare PMP-ESVE with average MB-level power control algorithm (AMP). The AMP always averagely allocates the computational resources to each MB. For the sake of fair comparison, both PMP-ESVE and AMP are all based on our previous research results energy scalable video encoder (Ji et al., 2009). This encoder can provide scalable video encoding output based on computational resource aware and can adapt different power consumption budget. From the experiments, we can observe the performance of MB Layer power control scheme based on perceptual cues. The experiment includes three parts: (1) Encoder information database, which is shared by PMP-ESVE and AMP; (2) The objective test and corresponding influence on PSNR, bitrates and power consumption.

5.1. Encoder information database

It is accepted that there are various configurations or parameters which can be adjusted to lower the energy consumption in video encoder. These include adjusting the search window (SW) in motion estimation, Lagrangian multiplier (Kwon et al., 2006), DCT and so on. We introduce the prediction mode decision to extend the adjusting scope. For the sake of clarity, we use the search window and inter prediction (IP) selections as the main adjustable parameters. We build an information database to collect the variation of PSNR and energy consumption in finite

encoding states. From the aspect of inter prediction selection, it is partitioned into five types: 16^*16 , 16^*8 , 8^*16 , 8^*8 and the four partition integration, dividually. From the aspect of SW in motion estimation, we test the variable region of 2^k , $k \in [2,5]$. Each adjustment in inter prediction and SW motion estimation corresponds to a working state. Obviously, there are above twenty working states through the configuration combination of these two modules.

From the aspect of energy consumption measurement, we use the clock cycles to evaluate energy consumption in our studies. To calculate the energy consumption in these sub-functions which lead to encoder modes, we assume that all other possible operations among the sub-functions are running. That is, if we test the effect of inter prediction selection with 16^*16 partition or 16^*8 partition, the motion estimation module will work in uniform configuration. Besides, we use the typical standard video sequences with different features as the test video set. The format is CIF and coded in AVS standard. We recycle the encoding process and repeat the experiment many times to reduce the occasionality, when test on the influence on energy consumption, PSNR, bitrates.

Based on these two principles, the encoder information database is built as follows:

- (1) Power consumption scaling based on RDC model: The power consumption scaling operations are the heart of the power scalable control paradigm. The detailed description of the principles of scalable coding in each composing modules is in and Section 2.
- (2) Best working state in each power consumption level: Since the number of working state maybe a huge number, best working state in each level is pre-computed so as to speedup direct decision. The detailed description of the decision process follows Section 4.4.

5.2. Objective test

(1) *Testing materials*: For the sake of fairness, three standard video sequences are used in the experiment. These sequences are in the same resolution and frequencies. The frame rate is kept in 30 fps.

For the power consumption measurement aspect, the power constraints are translated to computational costs, which is measured by the number of processing units (PUs). PUs is further scaled by processing cycles. For fixed 30 fps video encoding, since the PUs are employed to facilitate the quantitative measurement of the computation costs, the results are also effective in power and energy consumption combined with Section 2.3.

(2) *Testing method*: In conventional method, the encoder is divided into a number of basic operation units. Then, compute the number of PUs that each operation unit contains. However, the actual number of anchor PUs varies with different video sequences so that the actual PUs differ with the testing sequences even for the same encoder. Since the goal of this experiment lies in comparing the effect of proposed PMP-ESVE, content-related testing are introduced. First, we operate the whole encoder as a single module and run the encoder in different sequences many times, then denote the overall PUs number of the anchor as 100% computational costs. The following test related with its sequence is set in specific percentages of this anchor PUs, from 20% to 100%, which is in ascend proportion to energy consumption.

5.3. Analysis of objective testing results

(1) *Influence on power consumption*: The AMP allocates the available resources to each module rapidly, and does fast decision

among adjacent encoding states in each module. Both PMP-ESVE and AMP work under the same power consumption budget so as to obtain fair evaluation. We give the results when the two encoders work in the situations of 0%~20%, 20%~40%, 40%~60%, 60%~80% and 80%~100% of power budget. We test the performances of PMP-ESVE and AMP in these four power situations.

We analysis two cases: (1) the effect on the bitrates among each work mode; (2) the effect on the energy consumption among each work mode. Worst-case method is used to estimate these effects.

(2) *Influence on PSNR and bitrate*: Based on the testing materials and method, we test the encoders in four modes of 20%~100% of energy consumption. Here, for the sake of clarity, energy consumption is measured in direct proportion to computational cost.

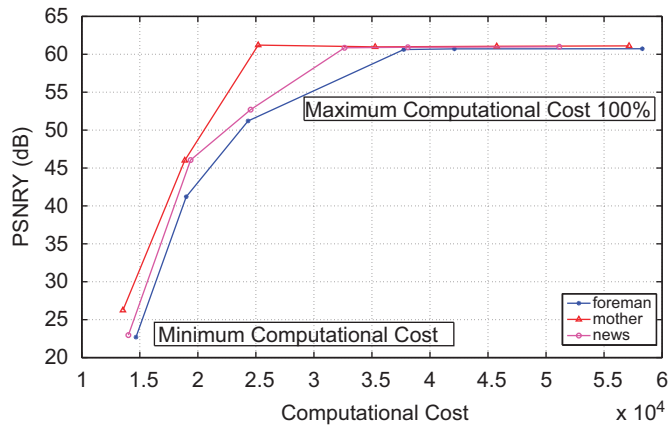


Fig. 7. The energy adjusting scope.

These two encoders also show better scalable performance in each working mode as in Fig. 7. Besides, the PMP-ESVE shows the advantages in encoding quality effect which achieve the same, sometimes even better PSNR with much lower bit rates at the same power level. Fig. 8 gives the statistic objective quality of encoders in term of PSNR and bit rate in case 1 and case 2. The experiment adopts three sequences ‘Foreman’, ‘mother’ and ‘news’, which represent three different video features. The dashed lines represent the results in AMP scheme, while the solid lines are the results in proposed AMP-ESVE scheme. Since video quality and low bit rate are both important considerations when measure the encoding states, Fig. 8(a) presents the PSNR results while Fig. 8(b) presents the bitrates results. Based on scalable video encoder (Ji et al., 2009), when the encoder work under different scalable computational cost, PMP-ESVE can make the encoder work with similar computational consumption as AMP, while have better encoding effects. This is reflected by (1) the PSNR results under PMP-ESVE are close and even higher than those under AMP scheme; (2) the bitrates under PMP-ESVE are lower than those under AMP. The reason lies in that under different cost budget, PMP-ESVE helps the encoder to choose the best parameter configurations and decide the best modes in its composed modules so as to keep the encoder work under best states in variable resource circumstance. Fig. 9 provides the subjective quality of encoders. Figs. 8 and 9 show that PMP-ESVE can reach same subjective quality as the direct energy control scheme, sometimes even better, while PMP-ESVE can save much bit rate and power consumption, as shown in Fig. 8(a) and (b).

6. Conclusion

This paper proposed PMP-ESVE method in perceptual energy scalable video encoding systems. There are three major

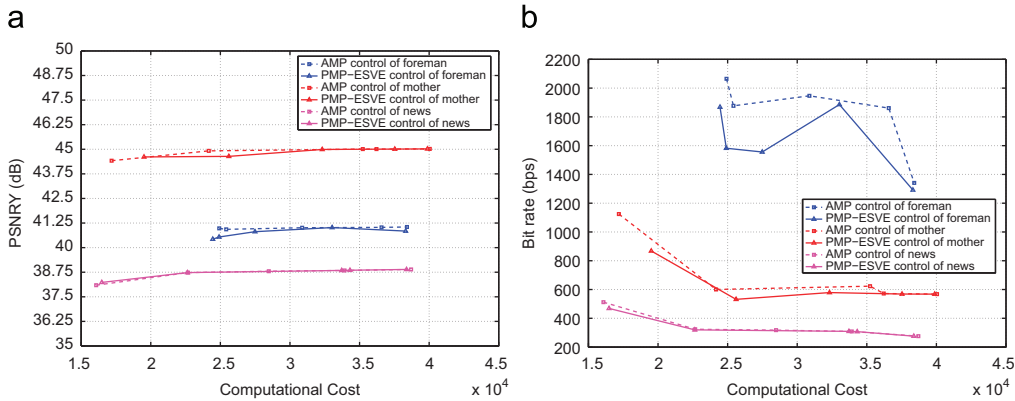


Fig. 8. Objective quality in case 1 and case 2.

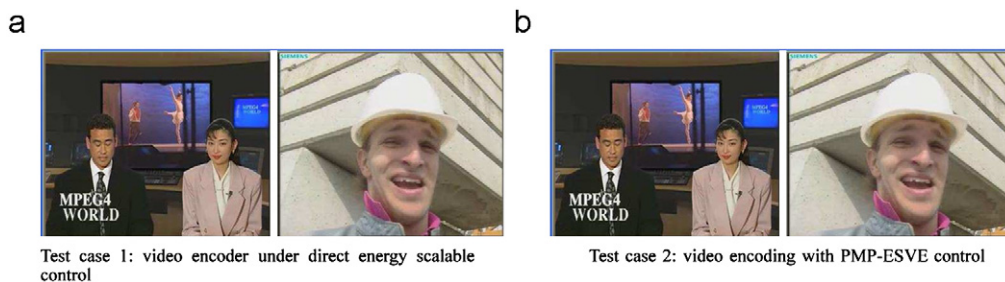


Fig. 9. Subjective comparison results. (left: for news, right: for foreman).

contributions in this work. First, video encoder can work under variable energy resource constraint marked with different energy consumption budgets; Second, the paper use perceptual method to obtain the quality optimization under energy consumption constraint. Third, JND model is extended to temporal domain so as to be more effective in video encoder. The experiment provide the results of two modules of video encoder. In fact, PMP-ESVE will be fine scalable when introduce more modules. Certainly, the accurate range of lower power consumption lies in not only PMP-ESVE but also the algorithm of each modules. However, PMP-ESVE can help the encoder provide many low power working modes while keep better performance, whatever the algorithm of modules will be designed. Therefore, we will expand and practice this method into more encoding modules so that finely scale the energy consumption control range in future research.

References

- Chou C-H, Li Y-C. A perceptually tuned subband image coder based on the measure of just-noticeable-distortion profile. *IEEE Trans CSVT* 1995;5(6):467–76.
- De Schrijver D, et al. MPEG-21 bitstream syntax descriptions for scalable video codes. *Multimedia Syst* 2006;11(5):403–21.
- Gao W, et al. AVS—the Chinese next-generation video coding standard, 2004. <<http://www.avs.org.cn/>>.
- He Z, et al. Power-rate-distortion analysis for wireless video communication under energy constraints. *IEEE Trans CSVT*, 2004.
- He Z, et al. Energy minimization of portable video communication devices based on power-rate-distortion optimization. *IEEE Trans CSVT* 2008;18(5).
- Hu Y, Li Q, et al. Joint rate-distortion-complexity optimization for H.264 motion search. In: *IEEE Conference ICME*. 2006. p. 1949–52.
- Jayant N, et al. Signal compression based on models of human perception. *Proc IEEE* 1993;81(10):1385–422.
- Ji W, Li P, Chen M, Chen Y. Power scalable video encoding strategy based on game theory. In: *IEEE Conference PCM*. 2009. p. 1237–43.
- Kwon DN, et al. Performance and computational complexity optimization in configurable hybrid video coding system. *IEEE Trans CSVT* 2006;16(1):11.
- Lian Jr. C, et al. Power-aware multimedia: concepts and design perspectives. *IEEE Circuits Syst Mag* 2007;7(2):26–34.
- Saponara S, et al. Performance and complexity co-evaluation of the advanced video coding standard for cost-effective multimedia communications. *EURASIP J Appl Signal Process* 2004;2:220–35.
- Sullivan GJ, Wiegand T. Rate-distortion optimization for video compression. *IEEE Signal Process Mag* 1998;15(6):74–90.
- Tu Y-K, Yang J-F, Sun M-T. Efficient rate-distortion estimation for H.264, AVC coders. *IEEE Trans CSVT* 2006;16(5):600–11.
- Wei Z, Ngan KN. A temporal just-noticeable distortion profile for video in DCT domain. In: *IEEE Conference ICIP*. October 2008. p. 1336–9.
- Yang X, et al. Rate control for videophone using local perceptual cues. *IEEE Trans CSVT* 2005;15(4):496–507.