A region-based intra-frame rate control scheme by jointing inter-frame dependency and inter-frame correlation

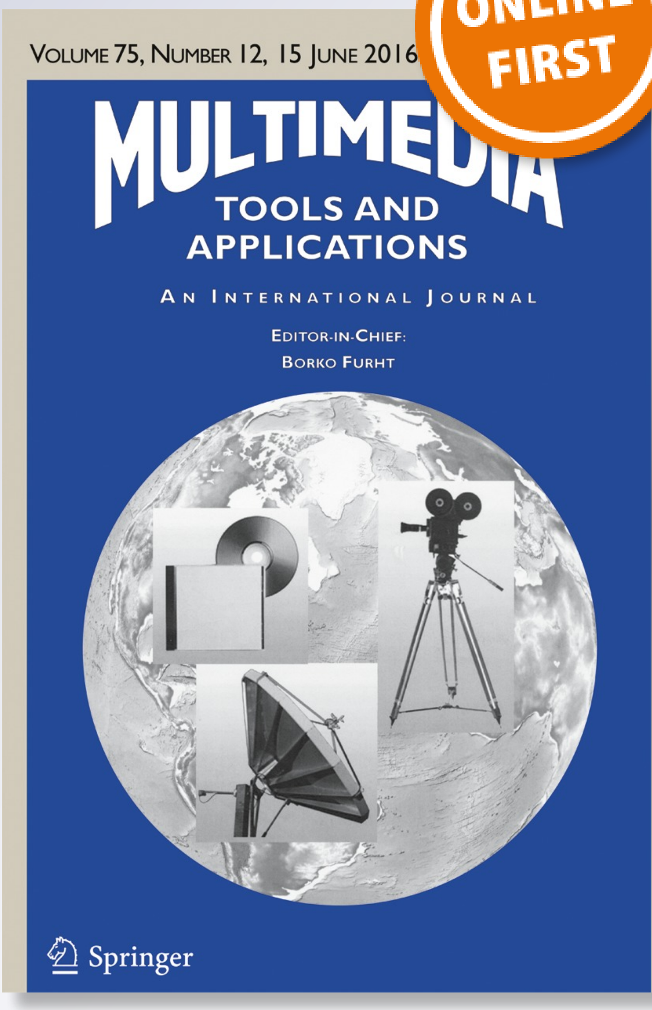
Hai-Miao Hu, Mingliang Zhou, Yang Liu & Naiyu Yin

Multimedia Tools and Applications
An International Journal

ISSN 1380-7501

Multimed Tools Appl
DOI 10.1007/s11042-016-3666-x

ONLINE FIRST



Springer

Multimed Tools Appl
DOI 10.1007/s11042-016-3666-x

A region-based intra-frame rate control scheme by jointing inter-frame dependency and inter-frame correlation

Hai-Miao Hu [1,2] **· Mingliang Zhou** [1] **· Yang Liu** [3] **·**
Naiyu Yin [1]

Received: 30 November 2015 / Revised: 18 April 2016 / Accepted: 3 June 2016

Abstract During I frame switching, the subjective quality between I frame and P frames usually have obvious fluctuation due to different coding methods. The periodic temporal visual fluctuation will cause video flicker. According to extensive experiments, we observe that I frame flicker possess a strong regional characteristics and different region have different degree of flicker. Based on this observation, a region-based I frame rate control scheme is proposed to suppress I frame flicker according to the different characteristics of the moving and non-moving regions. Firstly, by jointly considering the inter-frame dependency between I frame and subsequent un-encoded P frames and the inter-frame correlation between I frame and previous encoded P frame, an optimization model is proposed to achieve the optimal QPs for different regions. Secondly, a region-based inter-frame dependency model is proposed to separately describe the inter-frame dependency of different regions, which can accurately describe their description of the inter-frame dependency. The experimental results demonstrate that the proposed scheme can efficiently suppress I frame flicker and maintain the smoothness of subjective quality. Moreover, the proposed scheme can achieve a PSNR gain by 0.26 dB on average when compared with the rate control scheme adopted by the HEVC reference software HM15.0.

Keywords I frame rate control · Video flicker · Inter-frame dependency · Inter-frame correlation · Region-division · HEVC

✉ Hai-Miao Hu
frank0139@163.com

[1] Beijing Key Laboratory of Digital Media, School of Computer Science and Engineering, Beihang University, Beijing 100191, China

[2] State Key Laboratory of Virtual Reality Technology and Systems, Beihang University, Beijing 100191, China

[3] Beijing Institute of Graphics, Beijing 100029, China

Published online: 04 July 2016

1 Introduction

With the increasing demand of various video transmission applications over heterogeneous networks, video coding with a high efficiency and flexibility is important. The latest video coding standard HEVC [2, 10, 18, 52] can obtain a significant gain by employing more flexible quad-tree coding block partitioning structure, improved intra prediction, adaptive motion parameter prediction [2, 27, 42], etc.

In the HEVC reference software HM15.0, the R-lambda model is the recommended rate control scheme [6, 32, 42, 48]. However, the R-lambda model only consider the coding efficiency of I frames and the bitrates allocated to I frames are only according to bits per pixel, which fails to consider the contents of different frames. Although the quality of I frames can be guaranteed, I frames will consume much more bits which will cause the bit starvation of the following P frames. The quality of P frames will be unavoidably degraded. Especially, the quality degradation of last frames of one sequence is usually serious. Therefore, the R-lambda model can not achieve a smooth perceptual quality and will easily cause the perceptual quality fluctuation between I frames and P frames.

Intra-frame (I frame) is important and the rate control for I frame has been widely focused in the researches of video coding. Recently, many researches aiming at I frames focus on two major issues, namely the bit allocation between I frame and the following P frames and the rate-distortion (R-D) model establishment [4, 6, 11, 16, 17, 23, 24, 29, 38, 41, 45–47, 50, 53, 54, 58]. A novel quadratic ρ-domain rate model [41, 47] is proposed and a frame layer rate control scheme is proposed, which is proved to outperform the quadratic pixel-based unified rate-quantization (URQ) model [6] and can generate smoother coding quality when compared with URQ model [6]. However, this rate control scheme of HEVC is realized based on an entire frame, which will reduce the coding performance and will lead to the obvious perceptual quality fluctuation between I frame and P frames.

Due to the different encoding approaches of I frame and P frame, the video quality of I frame and the encoded P frame will be obviously fluctuated and this periodic temporal visual artifact will easily cause flicker. The flicker will undoubtedly affect the perceptual quality and can easily cause visual fatigue to human eyes, especially under the low bitrates or frequent I frame switching situations. This flicker is caused by intra-frame encoding mode and cannot be completely eliminated [19]. Although some rate control schemes have been proposed to suppress I frame flicker [13, 36], these algorithms have following two major limitations.

On the one hand, these rate control schemes are based on the frame level, assuming that as long as the overall objective quality can be consistently maintained, video flicker can be eliminated. However, I frame flicker has obvious regional characteristics, and different regions have different human visual characteristics and rate-distortion characteristics. Therefore, the treatment targeting at the entire frame cannot achieve the desired flicker suppression effect. Note that the regional characteristics of I frame flicker will be analyzed in depth through experiments in the next section.

On the other hand, there is a relatively strong inter-frame dependency between I frame and P frame [3, 20, 22, 28, 31, 33, 34, 39, 49]. Note that the quality of the subsequent un-encoded frames is dependent on the quality of the encoded frame, and the bit rate consumed by the encoded frame affects the bit allocation of the subsequent un-encoded frames in turn. This results in a strong dependency among frames. The dependency is obvious between I frame and P frame [3, 20, 22, 28, 33, 35, 39, 40, 49]. Since I frames will consume much more bits, which will easily cause the bit starvation of the following P frames. The quality of P frames will be unavoidably degraded. Therefore, an accurately description of the dependency between I frame and P frame is essential to improving encoding efficiency [3, 28, 31, 39, 49], and also, is important to maintain the smooth perceptual quality. However, the above I frame rate control schemes fail to take the inter-frame dependency into the consideration and to achieve smooth quality between I frames and P frames.

Therefore, this paper proposes a region-based I frame rate control scheme to suppress I frame flicker and maintain the smoothness of subjective quality. Firstly, based on our observations, a region-based inter-frame dependency model is proposed to separately describe the inter-frame dependency of moving and non-moving regions according to their different characteristics. The proposed model fully exploits the significant regional characteristics to achieve an accurate description of the inter-frame dependency. Secondly, an optimization model is proposed by jointly considering the inter-frame correlation between I frame and encoded P frame and the dependency between I frame and subsequent un-encoded P frames. The optimized QP can be achieved by solving the optimization model, which can effectively suppress I frame flicker. Finally, based on the fact that human visual system features varying levels of sensitivity to quality variations in different regions, the QP of the sensitive regions is adjusted to further suppress I frame flicker. The experimental results demonstrate that the proposed I frame rate control scheme can significantly reduce the fluctuation of perceptual quality and suppress I frame flicker by comparing the state-of-the art. Moreover, the proposed scheme can achieve a PSNR gain by around 0.26 dB on average when compared the rate control scheme adopted by the HEVC reference software HM15.0 [12].

The remainder of this paper is organized as follows. Some observations and justifications are provided in Section 2. Section 3 discusses the proposed optimization model for smooth quality. The proposed region-based inter-frame dependency model is presented in Section 4. The proposed approach is summarized in Section 5. Section 6 presents the experimental results. Finally, the paper is concluded in Section 7.

2 Observation and justification

The video flicker is caused by the different encoding modes of I frame and P frame and it is obvious in the low bit-rate applications. The rate-distortion optimization aiming at intra-frame prediction is designed to minimize the distortion between the current reconstructed video and the original video, which does not consider the quality continuity between adjacent frames. Therefore, during I frame switching, the corresponding Coding Tree Unit (CTU) between I frame and P frame usually results in significant perceptual quality fluctuation. Even under the all I frame encoding configurations, the slight variation of content in spatial domain will result in the change of intra modes selection. The corresponding CTUs between two adjacent I frames are also significantly different in terms of objective quality. These periodic quality fluctuations will lead to video flicker [1, 7–9, 13, 19, 21, 25, 26, 36, 37, 55, 56].

Based on extensive experiments, we observe that I frame flicker possesses obvious regional characteristics and different kinds of region will results in different degree of flicker. Compared with the regions containing motions (it is called the moving region in this paper), there is obvious flicker in non-moving regions. One example is shown in Fig. 1. It can be observed from Fig. 1b that the labeled non-moving regions have obvious subjective quality fluctuation, while the subjective quality fluctuation of the moving regions is not as obvious as it.

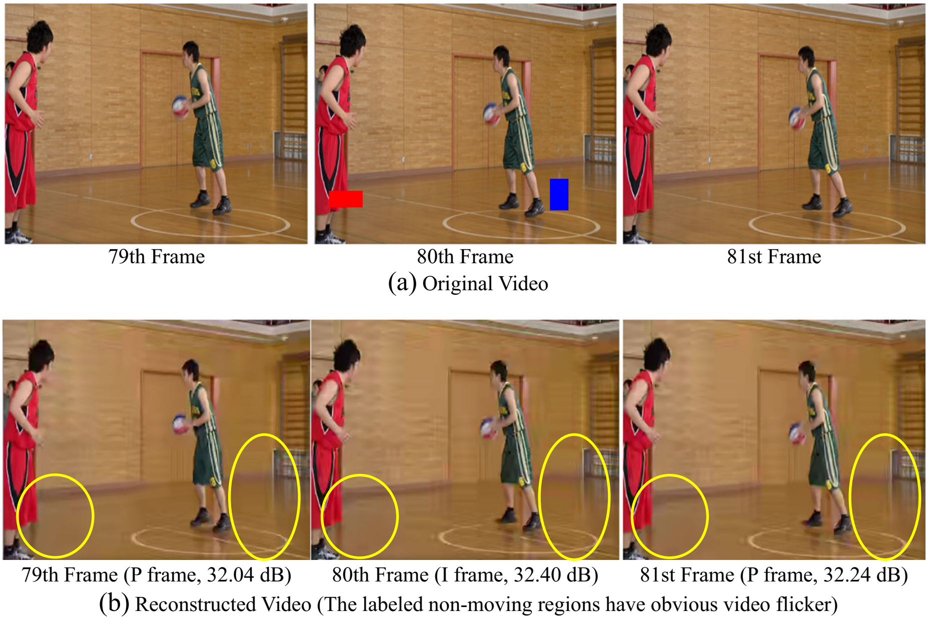


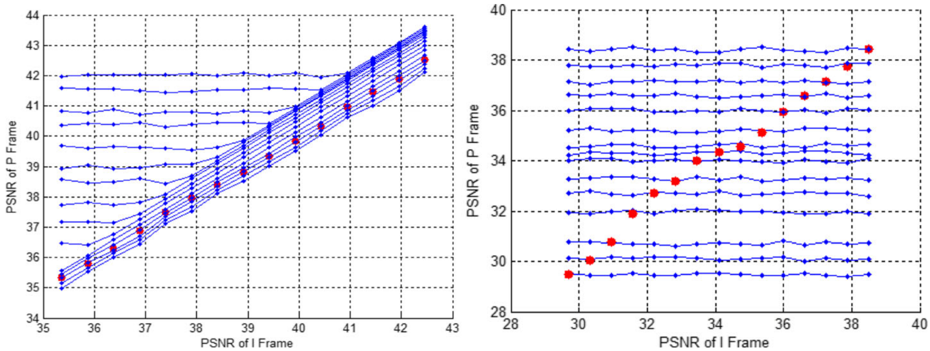
Fig. 1 A example of I frame flicker (BasketballPass, 416*240, Intra period = 20, IPP, QP = 38). **a** Original Video 79th Frame (P frame, 32.04 dB) 80th Frame (I frame, 32.40 dB) 81st Frame (P frame, 32.24 dB) **b** Reconstructed Video (The labeled non-moving regions have obvious video flicker)

There are mainly three reasons causing the regional characteristics of I frame flicker, which is discussed in detail as follows.

Firstly, different regions may have different inter-frame correlations and different rate-distortion characteristics [14]. Compared to the moving regions, the non-moving regions have relatively strong inter-frame correlations. Usually, the quality of non-moving regions in P frames can maintain smooth between I frame and its subsequent P frames. However, all these regions will be encoded by using Intra modes in the subsequent I frame. Thus, due to different encoding modes and quantization distortion, the quality of these regions will fluctuate. And this quality fluctuation will easily lead to I frame flicker. On the other hand, since the moving regions contain motions, the corresponding CTUs in the moving regions among different frames are usually encoded with different modes. Thus, the quality of these regions may be fluctuated among different frames (including I frames and P frames). Therefore, I frame flicker of moving regions is not as obvious as that of non-moving regions.

Secondly, different regions of one frame may have different inter-frame dependency. The dependency is obvious between I frame and P frame and an accurate description of the dependency between I frame and P frame is important to maintain the smooth perceptual quality [3, 28, 33, 35, 39, 40, 49]. In the existing researches, the inter-frame dependency models are established based on the entire frame [3, 28, 29, 31, 33–35, 39, 49]. As shown in Fig. 2, it can be observed that different video have different inter-frame dependency characteristics. In Fig. 2, Mobisode2 sequence has no camera motion, while Keiba sequence contains camera motions.

It should be noted that there are different kinds of regions in one frame and they may also have different inter-frame dependency characteristics. This conclusion can be confirmed by the following experiments. The distortion dependencies of different regions are shown in Fig. 3. It



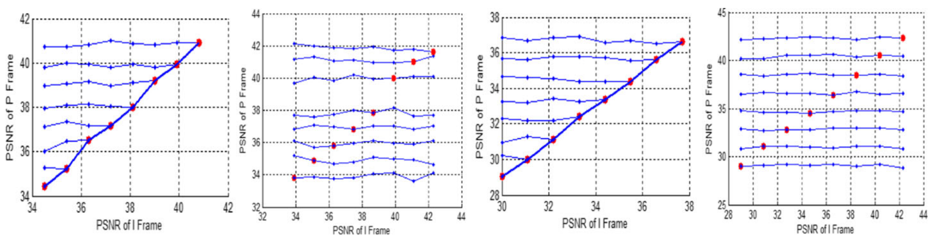
(a) Mobisode2 (without camera motion) (b) Keiba (with camera motion)

Fig. 2 The PSNR depenency relationship between I frame and P frame (416*240, Intra period = 10, IPP structure, QP = 28:1:42. The red dot is the output result when I frame and P frame adopt the same QP. Note that PSNR is used to describe the distortion depenecy relationship between I frame and P frame insteads of MSE in [33, 34]). **a** Mobisode2 (without camera motion) **b** Keiba (with camera motion)

indicates that for the moving regions, the quality of P frame is unaffected by that of I frame. While for the non-moving regions, the encoding quality of P frame will change along with that of I frame. Specifically, when the QP of I frame is smaller than that of P frame, the encoding quality of non-moving regions in P frame having a linear relationship with that of I frame. On the other hand, when the QP of I frame is greater than that of P frame, the encoding quality of non-moving regions in P frame is unaffected by that of I frame. Therefore, it is necessary to establish a region-based inter-frame dependency according to their different characteristics, which is important to improve encoding efficiency and maintain the smooth perceptual quality.

Thirdly, the video flicker is a subjective visual effect, and human eyes have different levels of sensitivity to quality variation in different regions. For an example, human eyes are less sensitive to especially bright or dark regions in a frame [5, 15, 43, 44, 51]. Additionally, compared with medium and low frequency, humans are less visually sensitive to information loss in high frequency range [1, 8, 21].

Based on the above observations, a region-based inter-frame dependency model is proposed to achieve an accurate description of inter-frame dependency to improve the coding efficiency. Correspondingly, an optimization model aiming at smooth quality is proposed by



CTU(X:3,Y:2) CTU(X:14,Y:11) CTU(X:9,Y:1) CTU(X:14,Y:17)

(a) Non-moving region (b) Moving region (c) Non-moving region (d) Moving region

Mobisode2 Keiba

Fig. 3 Distortion dependency relationship between I frame and P frame of different regions (416*240, Intra period = 10, IPP structure, QP = 28:1:42, Diff is calcuated through (1), the red points denotes that I frames and P frames have the same QP). **a** Non-moving region **b** Moving region **c** Non-moving region **d** Moving region jointing the inter-frame dependency and correlation to suppress I frame flicker, which will be presented in the next section.

3 Region-based inter-frame dependency model

An accurate description of inter-frame dependency can maintain subjective video quality continuity between I frame and P frame, and is important in smooth quality rate control. According to our observation in Section 2, this paper proposes a region-based inter-frame dependency model according to the different characteristics of the moving and non-moving regions. The proposed region-based inter-frame dependency model can achieve an accurate description of inter-frame dependency.

In this paper, by considering both the complexity and the accuracy, one frame is divided into two regions, namely the moving regions and the non-moving regions according to the difference in luminance value between two adjacent frames. The detailed region division refers to our previous work [14]. The proposed dependency model will be discussed in detail as follows.

Assume that the original I frame is $f_I(x, y)$, the corresponding reconstructed image is $f_I''(x, y)$, the first P frame after I frame is $f_P(x, y)$, and the corresponding predicted frame and reconstructed frame are $f_P'(x, y)$ and $f_P''(x, y)$, respectively. Then the frame error of P frame can be calculated by the following equation, and the detailed derivation refers to the Appendix of this paper.

$$E_P = f_P(x,y) - f_I(x,y) + \left(f_I(x,y) - f_I''(x+i, y+j)\right) - e_P''(x,y) \tag{1}$$

where (i, j) is the motion vector. From the above equation, the frame error of P frame depends on three aspects, namely the difference between the original I frame and the original P frame, the difference between the original I frame and the predicted P frame, and $e_P''(x, y)$ is the quantization distortion of P frame residuals.

The moving region and the non-moving region have different frame errors for P frame. For the non-moving regions, the difference of corresponding positions among adjacent frames is relatively small, and the motion vector *(i, j)* can be approximated to *(0,0)*, and the quantization distortion of P frame prediction residual can be ignored. Then according to (1), the frame error E_P^{NM} of the non-moving regions in P frame can be represented as follows.

$$\begin{aligned} E_P^{NM} &\approx f_P(x,y) - f_I(x,y) + \left(f_I(x,y) - f_I''(x,y)\right) \\ E_P^{NM} &\approx f_P(x,y) - f_I(x,y) \; + \; E_I^{NM} \approx E_I^{NM} \end{aligned} \tag{2}$$

where E_I^{NM} is the error of non-moving regions in I frame. Therefore, the error of moving regions in P frame is closely related to the error of non-moving regions in I frame.

Furthermore, according to the analysis of the literatures [33, 34], when the quantization stepsize of I frame is smaller than that of P frame (namely $Q_I < Q_P$), the use of a smaller quantization stepsize (Q_I) can produce smaller distortion than a bigger quantization stepsize (Q_P). Thus, the quantization distortion of E_P^{NM} varies linearly with the quantization distortion of E_I^{NM}. Conversely, when the quantization stepsize of I frame is larger than that of P frame (namely $Q_I > Q_P$), the distortion of P does not depend on that of I frame and the distortion of P frame only determined by the smaller quantization stepsize (Q_P). Therefore, taking into account both scenarios of $Q_I < Q_P$ and $Q_I > Q_P$, the distortion of P frame is only determined by the smaller quantization stepsize of I frame and P frame, i.e., $\min\{Q_I, Q_P\}$.

Suppose that P frame takes a fixed quantization parameter, the distortion of the non-moving regions in P frame varies with the changes in I frame quantization parameter, and the inter-frame distortion dependency can be modeled as follows (as shown in Fig. 4a, where the PSNR value is used to measure the distortion).

$$PSNR_k^{P,NM}\left(Q_k^I, Q_k^P\right) = \begin{cases} \alpha + \beta \times PSNR_k^I\left(Q_k^I\right), & Q_k^I < Q_k^P + \Delta \\ \alpha + \beta \times PSNR_k^I\left(Q_k^P + \Delta\right) & Q_k^I \geq Q_k^P + \Delta \end{cases} \tag{3}$$

where Q_k^I represents the QP of the kth CTU in the current I frame, α and β are two parameters and they can be calculated by the linear regression. Δ is a constant. $PSNR_k^{P,NM}(Q_k^I, Q_k^P)$ is the PSNR value of the kth CTU in the subsequent P frame, and $PSNR_k^I(Q_k^I)$ is the PSNR value of the kth CTU in current I frame. $PSNR_k^I(Q_k^I)$ is calculated by the linear PSNR-QP model proposed in our previous work [57].

$$PSNR_k^I\left(Q_k^I\right) = \gamma \times C_k^I \times Q_k^I + \mu \tag{4}$$

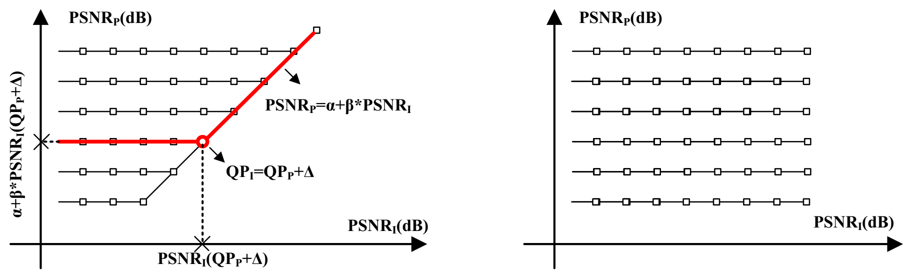
where C_k^I is the complexity of the kth CTU in I frame, γ and μ are two parameters, which can be determined by linear regression based on encoding results.

In (3), when I frame is set to a quantization parameter Q_k^I, and the PSNR value of kth CTU of P frame, $PSNR_k^{P,NM}(Q_k^I, Q_k^P)$, is changed along with different Q_k^P. In (3), $Q_k^P + \Delta$ divides the inter-frame distortion dependency model into two stages. As shown in Fig. 4a, these two stages have significantly different characteristics of distortion dependency. According to our experiments, Δ adopts −2 in this paper.

For the moving regions, the pixel value of the corresponding CTU in adjacent frames is more significantly different from each other, and the motion vector (i,j) can not be ignored. In addition, for regions with intense movements, motion estimation is unable to obtain accurate motion vectors, and the residual $e_P(x,y)$ between the predicted image and the original image is not negligible. Therefore, in accordance with (1), the error of the moving regions in P frame can be represented as follows.

$$\begin{aligned} E_P^M &= f_P(x,y) - f_I(x,y) + \left(f_I(x,y) - f_I''(x,y) + f_I''(x,y) - f_I''(x+i,y+j)\right) - e_P''(x,y) \\ E_P^M &= f_P(x,y) - f_I(x,y) + E_I^M + \left(f_I''(x,y) - f_I''(x+i,y+j)\right) - e_P''(x,y) \end{aligned} \tag{5}$$

For $f_P(x,y) - f_I(x,y), f_I''(x,y) - f_I''(x+i,y+j)$, and $e_P''(x,y)$, E_I^M is negligible. Whereas for the moving regions, E_P^M is unaffected by E_I^M. In particular, the regions with irregular movements in



(a) Inter-frame dependency of non-moving regions (b) Inter-frame dependency of moving regions

Fig. 4 The inter-frame depenecny model of different regions. **a** Inter-frame dependency of non-moving regions **b** Inter-frame dependency of moving regions

P frame may apply intra-frame encoding, and these regions only use the spatial correlation information in the current frame for encoding, without using inter-frame correlation information. So the distortion of these regions is completely independent of that of the encoded I frame. Therefore, the distortion of moving regions in P frame does not change with the variation in the QP of I frame (Q_k^I), and instead, depends only on the rate-distortion characteristics of P frame. Based on the above analysis, the inter-frame distortion dependency of the moving regions ($PSNR_k^{P,M}(Q_k^I, Q_k^P)$) can be simplified as the following model (as shown in Fig. 4b).

$$PSNR_k^{P,M}\left(Q_k^I, Q_k^P\right) = PSNR_k^P\left(Q_k^P\right) \tag{6}$$

In summary, the distortion inter-frame dependency model for the non-moving regions is given in (3), while the distortion inter-frame dependency model for the moving regions is given (6). For the non-moving regions, the parameters of the distortion dependency model can be updated by means of linear regression based on the encoded frames. Since the non-moving regions can maintain relatively strong inter-frame correlation, the statistical information used for updating parameters can be obtained from the encoding results of the CTU in the corresponding position of the previous encoded P frames.

The proposed inter-frame dependency model, compared with the models proposed in [33, 34], the biggest difference is that while the models proposed in [33, 34] is frame-oriented without distinguishing among different regions, while the proposed region-based inter-frame dependency model is established upon the inter-frame dependencies of different regions. The inter-frame dependency model for different region can achieve an accurate description of inter-frame dependency.

4 Optimization model by jointing inter-frame dependency and correlation

The key to suppressing video flicker is to maintain quality continuity between the corresponding regions in adjacent frames. The continuity itself is a relative concept, referring to a relatively small gap between the two objects under comparison. As shown in Fig. 5, during intra period switching, not only the initial I frame (the k th frame) in the current intra period is adjacent to the last P frame (the $k-1$ th frame) in the preceding intra period, but its quality also affects the quality of the subsequent P frame (the $k+1$ th frame), thereby transferring the impact to the last P frame in the current intra period the $k+N$ th frame). Meanwhile, the $k+N$ th frame, serving as the last P frame in the current intra period (its position is equivalent to that of the $k-1$ th frame in the preceding intra period), is adjacent to the initial I frame in the next

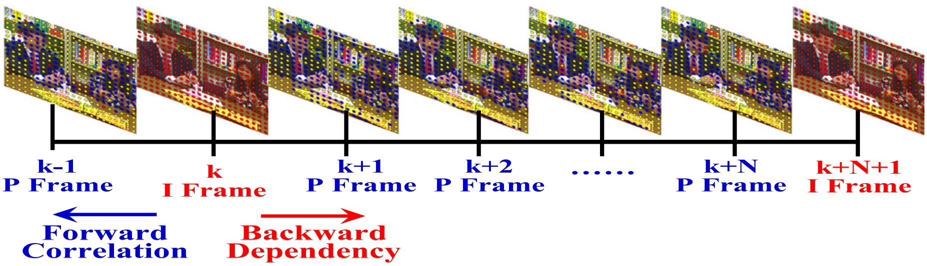


Fig. 5 Optimization model by jointing inter-frame dependency and correlation

intra period, and thus the fluctuations between frames within the current intra period will affect the quality of the next intra period. When a significant quality gap exists between the I frame and its subsequent P frame within a intra period, these intra quality disparities will diffuse into subsequent intra periods, resulting in "ups and downs" of encoding quality between the subsequent intra periods. Therefore, I frame should be treated as a frame key to transition; when creating an I frame rate-distortion optimization model for smooth video quality, we need to take into account the correlation between the preceding encoded frames and I frame, so as to maintain the quality continuity between the current I frame (the *k*th frame) and its preceding encoded P frame (the *k-1*th frame). In addition, the dependency between I frame and the subsequent un-encoded frames should be taken into account in order to avoid sharp quality gaps between I frame (the *k*th frame) and its subsequent P frame (the $k+1$ th frame).

Based on the proposed inter-frame dependency model, an optimization model for smooth video quality is proposed by considering not only the inter-frame correlation between I frame and its previous encoded P frames, but also the inter-frame dependency between I frame and its subsequent un-encoded P frames.

In the non-moving regions, adjacent frames can maintain good similarity and the current frame can be directly predicted by the corresponding position in the previous P frame encoding. Therefore, when there is not scene change, the video quality of the encoded frame will impact that of the subsequent un-encoded frame. On the other hand, according to our observation, the non-moving regions have a relatively strong inter-frame dependency between I frame and its subsequent un-encoded P frames (as shown in Fig. 3). Therefore, the video quality of non-moving regions in I frame needs to be consistent with that of the corresponding regions in not only the previous encoded P frame, but also the previous encoded I frame.

Considering the influence of the previous inter-frame correlation and the subsequent inter-frame dependency, the optimization model aiming at the non-moving regions is established to guarantee quality continuity between I frame and the adjacent P frames.

$$\begin{aligned}
\mathbf{Q}_k^I &= \underset{Q_k^I \in Q}{\operatorname{argmin}} \left\{ D_{\mathrm{var}}^{Fw}(Q_k^I) + D_{\mathrm{var}}^{Bw}(Q_k^I) \right\} \\
D_{\mathrm{var}}^{Fw}(Q_k^I) &= \nu \times \left| PSNR_k^I(Q_k^I) - PPSNR_k^I \right| + \nu \times \left| PSNR_k^I(Q_k^I) - PPSNR_k^P \right| \\
D_{\mathrm{var}}^{Bw}(Q_k^I) &= \left| PSNR_k^I(Q_k^I) - PSNR_k^{P,NM}(Q_k^I, Q_k^P) \right|
\end{aligned} \tag{7}$$

where Q_k^I represents the QP of the *k*th CTU in the current I frame. $PSNR_k^I(Q_k^I)$ is the PSNR value of the *k*th CTU in the current I frame and it can be calculated by our linear PSNR-QP model [57]. $D_{\mathrm{var}}^{Fw}(Q_k^I)$ is the quality difference between the current I frame and the previous encoded frames (including the previous P frame, $PPSNR_k^P$, and the previous I frame, $PPSNR_k^I$). Both $PPSNR_k^P$ and $PPSNR_k^I$ can be determined before the encoding of the current I frame. $D_{\mathrm{var}}^{Bw}(Q_k^I)$ represents the quality difference between the current I frame and the subsequent un-encoded P frame, where $PSNR_k^{P,NM}(Q_k^I, Q_k^P)$ is the PSNR value of the subsequent P frame and is predicted based on the proposed region-based inter-frame dependency model (3). v is a weighted factor determined by the interrelationship between the encoding quality of the current I frame and the previous frame. Our experimental results show that different trend of quality variation (becoming better or worse) will cause different degree of flicker. Compared with the situation that the quality is changed from bad to good, the degree of flicker is more serious when the quality is changed from good to bad. This is because the human eyes have different levels of sensitivity to the high- and low-frequency component of video and the information loss in the low-frequency component can be more easily perceived when the quality is degraded. Therefore, in this paper, when $PSNR_k^I(Q_k^I)$ is less than $PPSNR_k^P$ or $PPSNR_k^I$, ν is assigned a larger number, i.e., $\nu = 1.5$, otherwise a smaller number, i.e., $\nu = 1.0$ is adopted.

In the moving regions, the pixel differences in the corresponding CTUS are relatively significant in adjacent frames, and motion estimation is required when inter-frame encoding is applied. Since motion estimation and compensation may not obtain accurate results, compared with the non-moving regions, the moving regions usually have more evident quantization distortion when the same QP is adopted. Furthermore, the moving regions have more significant fluctuations in video quality. Therefore, the smooth quality optimization model of the moving regions is only required to maintain quality consistency between the current I frame and previous frames. Meanwhile, the interdependency between I frame and its subsequent P frame also needs to be taken into account. The smooth quality model of moving regions is shown in (8).

$$\begin{aligned} \mathbf{Q}_k^I &= \underset{Q_k^I \in Q}{\operatorname{argmin}} \left\{ D_{\mathrm{var}}^{Fw}\left(Q_k^I\right) + D_{\mathrm{var}}^{Bw}\left(Q_k^I\right) \right\} \\ D_{\mathrm{var}}^{Fw}\left(Q_k^I\right) &= \nu \times \left| PSNR_k^I\left(Q_k^I\right) - PPSNR_k^P \right| \\ D_{\mathrm{var}}^{Bw}\left(Q_k^I\right) &= \left| PSNR_k^I\left(Q_k^I\right) - PSNR_k^{P,M}\left(Q_k^I, Q_k^P\right) \right| \end{aligned} \tag{8}$$

where $PSNR_k^{P,M}(Q_k^I, Q_k^P)$ is calculated through (6) and $PSNR_k^I(Q_k^I)$ is calculated by our linear PSNR-QP model [57]. By solving the optimization model of both moving and non-moving regions ((7) and (8)), the optimal QPs can be achieved for the moving and the non-moving regions, respectively. The calculated QP can efficiently suppress I frame video flicker.

5 The summary of the proposed scheme for I frame flicker suppression

In this paper, the proposed I frame rate control scheme for smooth video quality mainly includes the following five steps, and all the other remaining steps are identical to those of the algorithm adopted by the HEVC reference software HM15.0 [12]. The flowchart of the proposed rate-control scheme is shown in Fig. 6.

5.1 Region division and image complexity measurement

The image is divided into moving regions and non-moving regions, and the gradient information for each CTU is calculated to measure the image complexity.

5.2 Optimization and solving of QP

Based on the results of region division and the gradient information for each CTU, the QP of each CTU are derived from the proposed optimization model for smooth quality. (7) and (8) are optimization models for non-moving region and moving region, respectively.

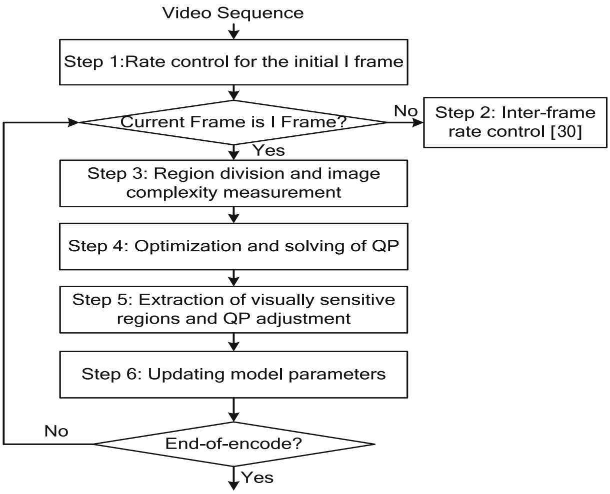


Fig. 6 The flowchart of the proposed rate control schemes

5.3 Extraction of visually sensitive regions and QP adjustment

Human eyes have different visual sensitivity to distortions of different regions. To further suppress video flicker, it is needed to distinguish the visually sensitive regions and insensitive regions according to the features of human visual system, and make adjustments to QP accordingly for different regions. According to [5, 15, 43, 44, 51], the visual sensitivity of human eyes is determined by two aspects. One is the characteristics of the image itself, such as luminance, contrast sensitivity, edges, etc. The other depends on the viewing distance, eye movement, etc. Since the viewing distance is full of uncertainty, this paper only takes into account the characteristics of the image itself when extracting visually sensitive regions.

For simplicity, this paper only incorporates image luminance and edges for extracting sensitive regions.

Luminance information According to human visual system characteristics, human eyes are only sensitive to luminance within a certain range and cannot easily detect the distortion in regions that are either too bright or too dark. Therefore, two threshold values are defined as ThL and ThH, and the normalized luminance value is divided into three intervals, where human eyes are sensitive to the luminance value that falls into the range of $[ThL, ThH]$. When the normalized luminance value is less than ThL and greater than ThH, the luminance sensitivity St_L is deemed as 0. Otherwise, when the normalization value falls between ThL and ThH, the luminance sensitivity St_L is equal to the normalized luminance value, i.e., $St_L = Lum/255$. This paper sets ThL to 0.25 and ThH to 0.75.

Edge information in regions with edges, distortions of either the background region or the foreground region are more likely to be visually perceived [44]. Therefore, the paper uses Sobel operator to determine whether each pixel is an edge point. The number of edge points in each CTU is represented by SbN, then the edge sensitivity of the CTU, St_S, is equal to the normalized SbN, i.e., $St_S = \frac{(SbN - SbN_{Min})}{(SbN_{Max} - SbN_{Min})}$, where SbN_{Max} and SbN_{Min} are the maximum SbN and the minimum SbN in the current frame, respectively.

Taking into account both of the above factors, the sensitivity factor of the CTU can be defined as the square root of the product of St_L and St_S, i.e., the sensitivity factor of the kth CTU is $St_k = \sqrt{St_L \times St_S}$. When St_k of the CTU is less than the preset threshold value, it belongs to the insensitive region. Otherwise it belongs to the sensitive region. This paper sets a threshold value of 0.1 for dividing sensitive regions, and the extraction result is shown in Fig. 7.

For the sensitive regions, the QP derived from (7) or (8) needs to deduct 1 to further guarantee the video quality in this region. Assume that Q_k^I is the QP of the kth CTU, which is derived from (7) or (8), and St_k is the sensitivity factor of the kth CTU. Then the adjusted QP of the kth CTU is calculated as follows.

$$\tilde{Q}_k^I = \begin{cases} Q_k^I - 1 & St_k > Th_S \\ Q_k^I & St_k \le Th_S \end{cases} \tag{9}$$

where Th_S is the threshold value for the extraction of sensitive regions, and in this paper Th_S is equal to 0.1. It should be noted that this implementation is a simple but effective method for extracting sensitive regions. Many other technologies [5, 15, 43, 44, 51], such as Temporal Contrast Sensitive Function, Spatial Contrast Sensitive Function, and Contrast Masking, can be employed to achieve a more accurate extraction of the sensitive regions. However, this is not the focus of this paper and it is one of our ongoing works.

Meanwhile, in order to ensure the smooth video quality and to limit the calculated QP within the [18, 44] interval, the finalized QP of the kth CTU, $\hat{Q}_k^I$, is:

$$\hat{Q}_k^I = \min\left\{\max\left\{\tilde{Q}_k^I, \hat{Q}_k^P - 3, 1\right\}, \hat{Q}_k^P + 3, 51\right\} \tag{10}$$

where, $\hat{Q}_k^P$ represents the QP of the corresponding CTU in the previous P frame.

5.4 Updating model parameters

Since the rate-distortion characteristics of the current I frame and the distortion dependency between I frame and P frame are unknown when deriving the QP for the current I frame, this paper predicts their parameters based on the encoding information of the previous I frame. After the encoding of I frame, the parameters in (3) are updated by means of linear regression. The details refer to our previous work [57]. Similarly, before coding the next I frame, the parameters in (6) and (8) are updated by the linear regression.

Original Video Sensitive Region Orignal Video Sensitive Region

(a) Hall (40th, CIF) (b) News (40th, CIF)

Fig. 7 The extracted sensitive regions (The white regions denotes the extracted sensitive regions). **a** Hall (40th, CIF), **b** News (40th, CIF)

5.5 Rate control for the initial I frame

The proposed rate control scheme in this paper needs to refer to encoding results of previous I frame. Therefore, in order to avoid affecting the subsequent I frame because of the inaccurate rate control of the initial I frame, and to obtain accurate parameters, this paper applies the following rate control scheme for first two I frames of the sequences. Since the rate-distortion characteristics of the initial I frame are unknown prior to encoding, this paper adjusts the QP of the initial I frame Q^I_{First}, heuristically.

$$\hat{Q}^I_{First} = \begin{cases} Q^I_{First} - \Delta, & Grad < 5 \\ Q^I_{First} - 0.5 \times \Delta, & 5 \le Grad < 11 \\ Q^I_{First}, & 11 \le Grad < 17 \\ Q^I_{First} + 0.5 \times \Delta, & 17 \le Grad < 23 \\ Q^I_{First} + \Delta, & 23 \le Grad < 29 \\ Q^I_{First} + 2 \times \Delta, & 29 \le Grad \end{cases} \tag{11}$$

where, $\hat{Q}^I_{First}$ is the adjusted QP of the initial I frame, *Grad* is the average gradient for all the CTUs in the current image, ε is the adjustment factor that can be determined by Q^I_{First} as follows:

$$\varepsilon = \begin{cases} 0.5, & Q^I_{First} = 10 \\ 1, & Q^I_{First} = 20 \\ 1.5, & Q^I_{First} = 25 \\ 2, & Q^I_{First} = 35 \end{cases} \tag{12}$$

Since the parameters of I frame distortion model and the distortion dependency model need to be determined by two sampling points, the QP of the second I frame of the sequence is equal to the average QP of both the initial I frame and the adjacent P frame. And it is calculated as follow.

$$Q^I_{Sec,k} = \left(\hat{Q}^I_{First} + \hat{Q}^P_k\right)/2 \tag{13}$$

where
$Q^I_{Sec,k}$ is the QP of the *k*th CTU in the second I frame.

6 Experiments and discussions

In this section, the performance of the proposed rate control scheme for I Frame is fully evaluated from two aspects, namely the subjective quality and the objective quality. The experiments are carried out based on the HEVC reference software HM15.0 [12]. Three similar schemes, namely the rate-rate control scheme used by HM15.0 (labeled as "HM") [12], our previous work (labeled as "APSIPA" [57]), and the latest Intra-frame rate control scheme proposed by S. Wang [47] (labeled as "Wang"), are adopted to compare with the proposed scheme. To ensure the fairness, we apply method "APSIPA" and "Wang" to the HM 15.0 platform for comparison. The objective quality comparison is carried out based on 11 test sequences. These video sequences which cover a wide range of resolution and motion characteristics are representative. The GOP structure for low delay applications is IPPP. The intra interval adopts 12 and 20, respectively. RDO is enabled, CTU size is set 64×64 (16×16-sized CTU is used for the sequences of 416×240 size). The frame rate is set to 30 fps (50fps for 1080p). The rest of the encoder settings is configured the same as default HM configuration. More detail information is shown in Table 1.

The experimental results are shown in Tables 2 and 3. The average gain in the table is calculated by comparing with HM. The flicker measurement in the table is calculated through the following equation [19, 57].

$$FlickerMeasurement = \underset{i,j(SSD(O(i-1,j),O(i,j))<\varepsilon)}{AVG}\{SSD(R(i-1,j)-R(i,j),O(i-1,j)-O(i,j))\} \quad (14)$$

where *SSD* denotes the sum of squared difference, *O (i, j)* denotes the original pixel value of the *j*th CTU in *i*th frame, *R (i, j)* denotes the reconstructed pixel value of the *j*th CTU in *i*th frame. Since the proposed scheme aims at I frames, the statistics of the flicker measurement only considers I frames.

As compared to the Wang algorithm, which yields the best results among the other three comparable schemes, the proposed scheme can obtain 0.146 dB (Intra period:12) and 0.127 (Intra period:20) in the average PSNR, respectively. Through experiments, it can be observed that our algorithm have an obvious improvements not only on the small image, but also on the high resolution image. At the same time, Tables 2 and 3 show the proposed scheme can efficiently suppress I frame flicker and maintain the smoothness of subjective quality.

The big value of the flicker measurement indicates a serious fluctuation of the perceptual quality and implies a serious video flicker as well. The frame-by-frame flicker measurement comparison is shown in Fig. 8.

Note that the proposed I frame rate control scheme is aiming at suppressing I frame flickers. In order to measure I frame flicker, a conventional measurement is employed by referring to [19, 57]. Using a more effective method to measure Mean Opinion Score (MOS) for subjective evaluation is our next research work.

Some more subjective quality compressions are shown in Figs. 9 and 10. It can be easily observed that the proposed scheme can effectively maintain the smoothness between I frame and its neighboring P frames. As shown in Fig. 9, when compared with HM, the proposed scheme can not only achieve a clearer the right wall, but also maintain a smooth quality of the right wall. Although Wang can also achieve a clearer the right wall when compared with HM,

Table 1 Experimental Conditions

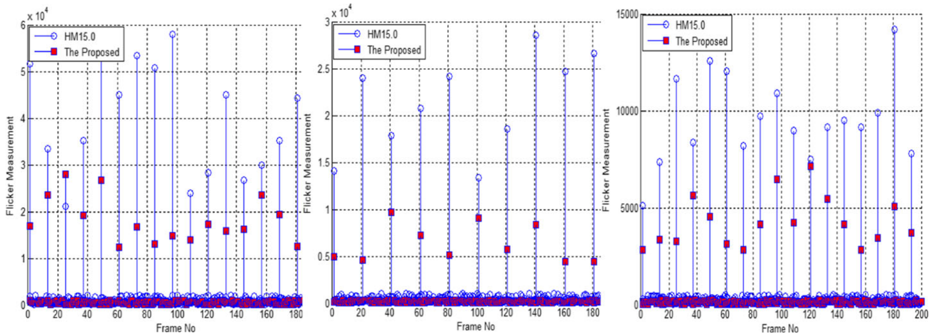
Parameter	Value
Platform	HM15.0
Coding structure	LP-main (IPPP)
CTU size	64×64 (16×16 for 416×240)
Intra period	12,20
Total number	200
Frame rate	30 fps (50fps for 1080p)
SAO	On
CABAC	On
RDO	On

Table 2 Objective qualtiy comparision (Intra period = 12)

Sequenc (Target Bitrates, kbps)		Y-PSNR				Bit rate (kbps)				Flicker measurement			
		HM	Wang	APSIPA	Our	HM	Wang	APSIPA	Our	HM	Wang	APSIPA	Our
416*240	BasketballPass (300kbps)	33.61	33.80 (+0.19)	33.82 (+0.21)	33.92 *(+0.31)*	303	306	304	303	6039.97	5886.37 (−2.54%)	4083.55 (−32.39%)	4004.26 *(−33.70%)*
	Keiba (1000kbps)	40.80	41.02 *(+0.22)*	40.78 (−0.02)	40.86 (+0.06)	1002	1004	1002	1001	4178.41	3317.22 (−20.61%)	2227.35 (−46.69%)	2155.34 *(−48.42%)*
832*480	RaceHorses (2000kbps)	32.60	32.81 (+0.21)	32.80 (+0.20)	32.93 *(+0.33)*	2007	2010	2007	2003	1178.78	1294.82 (+9.84%)	1208.31 (+2.51%)	979.17 *(−16.93%)*
	PartyScene (4000kbps)	32.90	33.08 (+0.18)	33.04 (+0.14)	33.19 *(+0.29)*	4021	4017	4033	4025	894.36	1056.64 (+18.14%)	915.33 (+2.34%)	742.49 *(−16.98%)*
HD (720p)	FourPeople (2000kbps)	40.98	41.17 *(+0.19)*	40.99 (+0.01)	40.99 (+0.01)	2026	2036	2007	2003	1346.37	1414.17 (+5.04%)	1264.21 (−6.10%)	1102.26 *(−18.13%)*
	KristenAndSara (3000kbps)	43.59	43.68 (+0.09)	43.54 (−0.05)	43.99 *(+0.40)*	3025	3020	3041	3026	1057.37	1100.55 (+4.08%)	949.59 (−10.19%)	840.34 *(−20.53%)*
FHD (1080-p)	Cactus (8000kbps)	34.76	34.86 (+0.10)	34.98 (+0.22)	35.09 *(+0.33)*	8060	8088	8094	8071	2056.76	2246.58 (+9.23%)	1939.36 (−5.71%)	1760.74 *(−14.39%)*
	PartyScene (12000kpbs)	39.15	39.34 (+0.19)	39.32 (+0.17)	39.48 *(+0.33)*	12050	12056	12009	12001	1400.18	1498.26 (+7.00%)	1375.28 (−1.78%)	1120.55 *(−19.97%)*
2k	PeopleOnstreet (12000kbps)	33.45	33.42 (−0.03)	33.59 (+0.14)	33.65 *(+0.20)*	12100	12070	12058	12036	1834.59	1674.33 (−8.74%)	1492.54 (−18.64%)	1277.86 *(−30.35%)*
	RaceHorse (12000kbps)	38.14	38.20 (+0.06)	38.24 (+0.10)	38.48 *(+0.34)*	12099	12064	12031	12007	1177.96	1066.58 (−9.46%)	1002.54 (−14.89%)	897.99 *(−23.77%)*
4k	PeopleOnstreet (24000kbps)	32.04	32.13 (+0.09)	32.12 (+0.08)	32.40 *(+0.36)*	24107	24150	24079	24035	1999.36	1864.54 (−6.74%)	1762.55 (−11.84%)	1683.28 *(−15.81%)*
	Traffic (24000kbps)	37.18	37.29 (+0.11)	37.33 (+0.15)	37.45 *(+0.27)*	24099	24108	24085	24021	1677.35	1596.53 (−4.82%)	1503.46 (−10.37%)	1407.58 *(−16.08%)*
8k	PeopleOnstreet (50000kbps)	30.56	30.65 (+0.09)	30.67 (+0.11)	30.75 *(+0.19)*	50200	50180	50100	50055	2655.34	2436.82 (−8.23%)	2407.33 (−9.34%)	2301.69 *(−13.32%)*
	Traffic (50000kbps)	35.96	35.99 (+0.03)	36.07 (+0.11)	36.30 *(+0.34)*	50150	50155	50090	50040	2206.55	2114.37 (−4.18%)	2099.47 (−4.85%)	1980.51 *(−10.24%)*
	Avg.		*+0.123*	*+0.112*	*+0.269*			*−0.856%*	*−11.996%*	*−21.33%*			

Table 3 Objective qualtiy comparision (Intra period = 20)

Sequenc (Target Bitrates, kbps)		Y-PSNR				Bit rate (kbps)				Flicker measurement			
		HM	Wang	APSIPA	Our	HM	Wang	APSIPA	Our	HM	Wang	APSIPA	Our
416*240	BasketballPass (300 kbps)	33.73	33.93 (+0.20)	33.92 (+0.19)	34.08 *(+0.35)*	304	305	305	303	5789.33	5844.38 (+0.95 %)	3686.52 (−36.32 %)	3445.36 *(−40.49 %)*
	Keiba (1000 kbps)	40.94	41.15 *(+0.21)*	40.88 (−0.06)	40.99 (+0.05)	1004	1004	1004	1000	3715.28	3122.58 (−15.95 %)	2487.79 (−33.04 %)	2007.87 *(−45.96 %)*
832*480	RaceHorses (2000 kbps)	32.78	33.00 (+0.22)	32.97 (+0.19)	33.12 *(+0.34)*	2011	2016	2012	2005	937.46	944.36 (+0.74 %)	877.88 (−6.36 %)	795.66 *(−15.13 %)*
	PartyScene (4000 kbps)	33.06	33.26 (+0.20)	33.18 (+0.12)	33.39 *(+0.33)*	4028	4022	4041	4029	811.33	853.78 (+5.23 %)	821.56 (+1.26 %)	648.58 *(−20.06 %)*
HD (720p)	FourPeople (2000 kbps)	41.14	41.30 *(+0.16)*	41.14 (0.00)	41.19 (+0.05)	2034	2041	2014	2007	989.56	1112.33 (+12.41 %)	927.41 (−6.28 %)	775.23 *(−21.66 %)*
	KristenAndSara (3000 kbps)	43.75 (+0.03)	43.78 (−0.06)	43.69 *(+0.20)*	43.95	3034	3022	3047	3033	902.36 (−3.51 %)	870.71 (−8.94 %)	821.69 *(−22.71 %)*	697.46
FHD (1080p)	Cactus (8000 kbps)	35.20	35.30 (+0.10)	35.44 (+0.24)	35.55 *(+0.35)*	8120	8148	8130	8099	1891.55	2121.52 (+12.16 %)	1808.28 (−4.40 %)	1645.94 *(−12.98 %)*
	PartyScene (12000kpbs)	39.59	39.78 (+0.19)	39.81 (+0.22)	39.88 (+0.29)	12120	12110	12077	12054	1693.33	1323.49 (−21.84 %)	1269.22 (−25.05 %)	1081.36 *(−36.14 %)*
2 k	PeopleOnstreet (12000 kbps)	33.77	33.86 (+0.09)	33.90 (+0.13)	33.99 *(+0.22)*	12096	12068	12055	12017	1724.58	1534.26 (−11.04 %)	1324.86 (−23.18 %)	1177.33 *(−31.73 %)*
	RaceHorse (12000 kbps)	38.46	38.58 (+0.12)	38.60 (+0.14)	38.71 *(+0.25)*	12091	12055	12030	12001	1065.85	986.43 (−7.45 %)	922.56 (−13.44 %)	855.31 *(−19.75 %)*
4 k	PeopleOnstreet (24000 kbps)	32.35	32.46 (+0.11)	32.51 (+0.16)	32.69 *(+0.34)*	24105	24140	24076	24021	1807.47	1744.39 (−3.49 %)	1632.89 (−9.66 %)	1541.33 *(−14.72 %)*
	Traffic (24000 kbps)	37.50	37.60 (+0.10)	37.64 (+0.14)	37.79 *(+0.29)*	24089	24106	24081	24017	1544.77	1477.26 (−4.37 %)	1324.56 (−14.26 %)	1276.35 *(−17.38 %)*
8 k	PeopleOnstreet (50000 kbps)	31.01	31.09 (+0.08)	31.11 (+0.10)	31.33 *(+0.32)*	50180	50170	50090	50044	2506.36	2379.55 (−5.06 %)	2284.96 (−8.83 %)	2199.37 *(−12.25 %)*
	Traffic (50000 kbps)	36.29	36.38 (+0.09)	36.41 (+0.12)	36.59 *(+0.30)*	50130	50145	50080	50020	2103.65	2065.32 (−1.82 %)	1979.76 (−5.89 %)	1866.33 *(−11.28 %)*
	Avg.		*+0.136*	*+0.116*	*+0.263*			*−3.074 %*	*−13.885 %*	*−23.017 %*			



(a) Blowing Bubbles (416*240,200kbps,Intra period =12)
(b) BQMall (832*480,500 kbps,Intra period =20)
(c) FourPeople (720p,1000 kbps,Intra period =12)

Fig. 8 Frame-by-frame flicker measurement comparision (Flicker measurement is calculated by (14)). **a** Blowing Bubbles (416*240,200 kbps,Intra period =12) **b** BQMall (832*480,500 kbps,Intra period =20) **c** FourPeople (720p,1000 kbps,Intra period =12)

but it fails to maintain its smoothness when I frame switch (between 47th frame and 48th frame).

Also, it can be observed from Fig. 10 that HM can not achieve a clear face of the left girl in the 59 h frame (P frame), but it achieve a clearer face of the left girl in the 60th frame (I frame). Thus, this leads to an obvious video flicker during I frame switches. Although Wang can achieve a clearer face of the left girl when compared with HM, it can not maintain the smooth quality. As shown in Fig. 10b, the 60th frame (I frame) is clearer than the other two P frames (59th frame and 60st frame). It can be observed from Fig. 10c that the proposed scheme can achieve a smooth subjective quality when compared with the other two schemes.

Therefore, the above experiments results demonstrate that the proposed scheme can effectively suppress perceptual flicker and maintain a smooth subjective quality between I frame and P frames. Moreover, when compared with the rate control scheme adopted in HM15.0, the proposed scheme can also improve the subjective quality in terms of average PSNR.

7 Conclusion and discussion

This paper proposes a region-based I frame rate control scheme. According to different characteristics of different regions, a region-based inter-frame dependency model is proposed to separately describe the inter-frame dependency of different regions, which can achieve an accurate description of the inter-frame dependency. Moreover, by jointly considering the inter-frame correlation between I frame and encoded P frame and the dependency between I frame and subsequent un-encoded P frames, an optimization model is proposed to achieve the optimal QP to effectively suppress I frame flicker. The experimental results demonstrate that the proposed scheme can not only significantly reduce the fluctuation of perceptual quality, but also can achieve a PSNR gain by around 0.26 dB on average when compared the state-of-the-art. Therefore, the proposed scheme can efficiently suppress I frame flicker and maintain the smoothness of subjective quality.

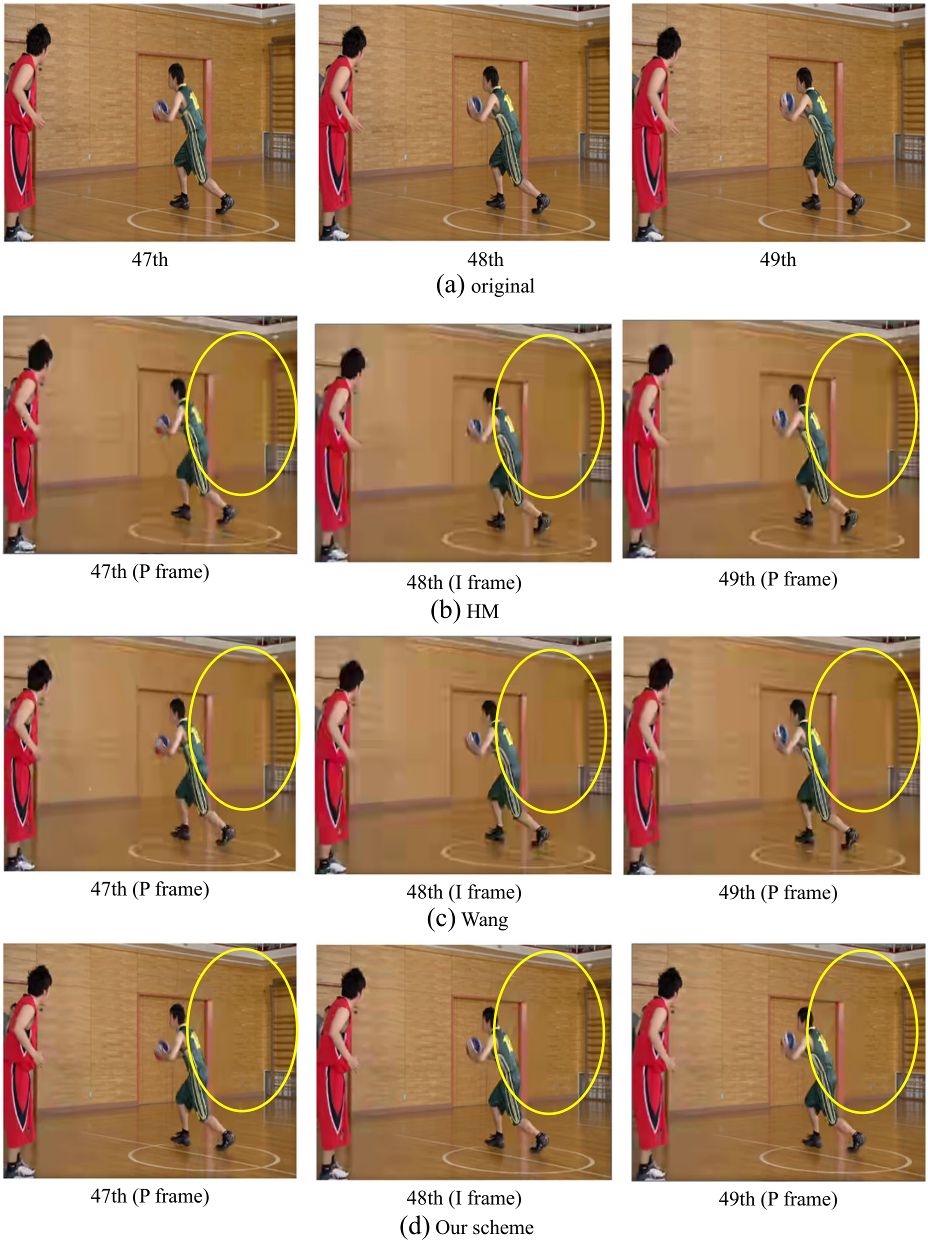


Fig. 9 The subjective quality comparison (BasketballPass, 416*240, frame rate =50,Intra period = 12, 100 kbps.) **a** original 47th (P frame) 48th (I frame) 49th (P frame) **b** HM 47th (P frame) 48th (I frame 49th (P frame) **c** Wang 47th (P frame) 48th (I frame) 49th (P frame) **d** Our scheme

Although the proposed scheme is aiming at achieving a smooth quality I frame rate control, it can be extended for P frame rate control to further improve the coding efficiency. The proposed region-based inter-frame dependency is helpful to carry out an accurate bit-allocation between I frame and P frames. This is one of our ongoing works.

Fig. 10 The subjective quality comparison (BlowingBubbles, 416*240, frame rate =50,Intra period = 20, 200 kbps.) . **a** original 59th (P frame) 60th (I frame) 61st (P frame) **b** HM 59th (P frame) 60th (I frame) 61st (P frame) **c** Wang 59th (P frame) 60th (I frame) 61st (P frame) **d** Our scheme

Note that *ThL* and *ThH* are chosen according to the statistics from the experimental results. Our experimental results show that the threshold values we choose can achieve good results for all test sequences. These thresholds could be adaptively adjusted according to different contents (e.g., taking more HVS features into the consideration or employing video content analysis methods) to achieve a more accurate region division. In order to further suppress video flicker, one frame is divided into sensitive regions and insensitive regions according to the features of human visual system, since human eyes have different visual sensitivity to distortions of different regions. In this paper, the sensitive regions are extracted for QP adjustment. Since QP adjustment is an amendment of the proposed algorithm and it is not the major contribution of this paper. For simplicity, this paper only incorporates image luminance and edges for extracting sensitive regions. Note that the more comprehensive sensitive region extraction method can be easily combined with the proposed rate-control scheme, which will be also our future work.

Acknowledgments This work was partially supported by the National Natural Science Foundation of China (No. 61370121), and the National Hi-Tech Research and Development Program (863 Program) of China (No.2014AA015102).

Appendix: derivation of P frame error

Suppose that the original image of I frame is $f_I(x, y)$, and the corresponding reconstructed image is $f_I''(x, y)$. The original image of the first P frame after I frame is $f_P(x, y)$, and its predicted image and reconstructed image are $f_P'(x, y)$ and $f_P''(x, y)$, respectively. Then the frame error between I frame and P frame (the error between the reconstructed frame and the original frame) can be defined as:

$$\begin{aligned} E_I &= f_I(x,y) - f_I''(x,y) \\ E_P &= f_P(x,y) - f_P''(x,y) \end{aligned} \qquad (A-1)$$

where, E_I is the frame error of I frame, and E_P is the frame error of P frame. Assume that $e_P(x, y)$ is the residual between the predicted image and the original image, i.e., $e_P(x, y) = f_P(x, y) - f_P'(x, y)$, and the quantized value is $e_P''(x, y) = Quant(e_P(x, y))$. The reconstructed frame of P frame can be represented as:

$$f_P''(x,y) = f_P'(x,y) + e_P''(x,y) \qquad (A-2)$$

where, the predicted image of P frame can be further represented as:

$$f_P'(x,y) = f_I''(x+i, y+j) \qquad (A-3)$$

where, (i, j) is the motion vector. Then after incorporating (A-2) and (A-3) into (A-1), the frame error of P frame can be rewritten as:

$$\begin{aligned} E_P &= f_P(x,y) - \left(f_P'(x,y) + e_P''(x,y)\right) \\ E_P &= f_P(x,y) - f_I(x,y) + \left(f_I(x,y) - f_P'(x,y)\right) - e_P''(x,y) \\ E_P &= f_P(x,y) - f_I(x,y) + \left(f_I(x,y) - f_I''(x+i, y+j)\right) - e_P''(x,y) \end{aligned} \qquad (A-4)$$

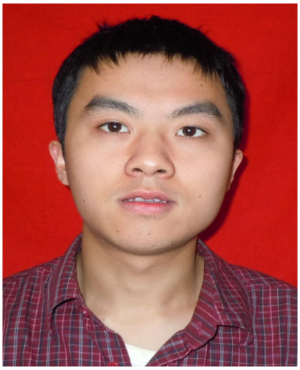
References

1. Becker A, Chan W, Poulouin D (2004) Flicker reduction in intra frame codecs. IEEE data compression conference (DCC), pp 251–261

2. Bross B, Han WJ, Ohm JR, Sullivan GJ and Wiegand T (2012) High efficiency video coding (HEVC) text specification draft 8. JCT-VC document, JCTVC-J1003
3. Changuel N, Sayadi B (2010) H.264/AVC inter-frame rate-distortion dependency analysis based on independent regime-switching AR models. IEEE international conference on acoustics speech and signal processing (ICASSP), pp 914–917
4. Chao J, Huitl R, Steinbach E (2015) A novel rate control framework for SIFT/SURF feature preservation in H.264/AVC video compression. IEEE Trans Circuits Syst Video Technol (TCSVT 25(6):958–972
5. Chen Z, Guillemot C (2010) Perceptually-friendly H.264/AVC video coding based on foveated just-noticeable-distortion model. IEEE Trans Circuits Syst Video Technol (TCSVT) 20(6):806–819
6. Choi H, Nam J, Yoo J, Sim D and Bajić VI (2012) Rate control based on unified RQ model for HEVC. Joint collaborative team on video coding (JCT-VC) of ITU-T SG16 WP3 and ISO/IECJTC1/SC29/WG118th meeting
7. Chono K, Senda Y, Miyamoto Y (2006) Detented quantization to suppress flicker artifacts in periodically inserted Intra-coded pictures. In: IEEE international conference on image processing,H.264 video coding, (ICIP), pp 1713–1716
8. Chun SS, Kim JR (2006) Intra prediction mode selection for flicker reduction in H.264/AVC. IEEE Trans Consum Electron (TCE) 52(4):1303–1310
9. Chun S, Kim JR, Sull S (2006) Flicker reduction in intra coded frames of H.264/AVC. In: IEEE international conference on image processing (ICIP), pp 2025–2028
10. Coding of audio-visual objects—Part 2: Visual (1999) in ISO/IEC 14 496–2 (MPEG-4 Visual Version 1), Apr
11. H.264/AVC Reference Software (JM16.0) (2009) Available: http://iphome.hhi.de/suehring/tml
12. HEVC Test Model (HM15.0) (2014) Available: http://hevc.hhi.fraunhofer.de/svn/svn_HEVCSoftware
13. Hong SH, Yoo SJ, Lee SW et al (2003) Rate control of MPEG video for consistent picture quality. IEEE Trans Broadcast (TB) 49(1):1–13
14. Hu HM, Li B, Lin W, Li W, Sun MT (2012) Region-based rate control for H.264/AVC for Low Bit-rate applications. IEEE Trans Circuits Syst Video Technol (TCSVT) 22(11):1564–1576
15. Jia Y, Lin W, Kassim AA (2006) Estimating just-noticeable distortion for video. IEEE Trans Circuits Syst Video Technol (TCSVT) 16(7):820–829
16. Jing X, Chau L (2006) A novel Intra-rate estimation method for H.264 rate control. In: IEEE international symposium on circuits and systems (ISCAS), pp 5019–5022
17. Jing X, Chau LP, Siu W (2008) Frame complexity-based rate-quantization model for H.264/AVC Intraframe rate control. IEEE Signal Process Lett 15:373–376
18. JVT, Advanced Video Coding (AVC) (2004) ITU-T Rec.H.264 and ISO/IEC 14496–10 (MPEG-4 Part 10), 3rd ed
19. JVT-E070 (2002) Flicking Reduction in All Intra Frame Coding, Switzerland
20. JVT-G012, Li Z G, Pan F et al (2003) Adaptive basic unit layer rate control for JVT. Thailand
21. Kuge T (2002) Wavelet picture coding and its several problems of the application to the interlace HDTV and the ultra-high definition images. In: IEEE international conference on image processing (ICIP), pp III-217-220
22. Lee B, Kin M, Nguyen TQ (2014) A frame-level rate control scheme based on texture and Nontexture rate models for high efficiency video coding. IEEE Trans Circuits Syst Video Technol (TCSVT) 24(3):465–479
23. Lee J, Shin I, Park H (2006) Adaptive intra-frame assignment and bit-rate estimation for variable GOP length in H.264. IEEE Trans Circuits Syst Video Technol (TCSVT) 16(10):1271–1279
24. Lee YG, Song BC (2009) An intra-frame rate control algorithm for ultralow delay H.264/advanced video coding (AVC). IEEE Trans Circuits Syst Video Technol (TCSVT) 19(5):747–752
25. Leontaris A, Tonomura Y, Nakachi T et al (2006) Rate control for flicker artifact suppression in motion JPEG2000. IEEE international conference on acoustics, speech, and signal processing (ICASSP), pp II-41-44
26. Leontaris A, Tonomura Y, Nakachi T (2007) Flicker suppression in JPEG2000 using segmentation-based adjustment of block truncation lengths. In: IEEE international conference on acoustics, speech, and signal processing (ICASSP), pp 1117–1120
27. Li B (2012) Rate control by R-lambda model for HEVC. In: Proceedings of the 11th meeting joint collaborative team video coding (JCT-VC) of ITU-T1/SC 29/WG 11, China
28. Li X, Hutter A, Kaup A (2009) One-pass frame level budget allocation in video coding using inter-frame dependency, IEEE Intl. Multimedia Signal Processing (MMSP), pp: 1–6
29. Li Y, Jia H, Zhu C, Yang M, Xie X and Gao W (2014) Window-based rate control for video quality optimization with a novel INTER-dependent rate-distortion model. In: Signal processing: image communication (SPIC)
30. Li B, Li H, Li L, Zhang J (2014) λ domain rate control algorithm for high efficiency video coding. IEEE Trans Image Process (TIP) 23(9):3841–3854

31. Li X, Wien M, Ohm JR (2011) Rate-complexity-distortion optimization for hybrid video coding. IEEE Trans Circuits Syst Video Technol (TCSVT) 21(7):957–970
32. Li S, Xu M, Deng X, Wang Z (2015) Weight-based R- lambda rate control for perceptual HEVC coding on conversational videos. Signal Process Image Commun 38:127–140
33. Lin LJ (1997) Video bit-rate control with spline approximated rate-distortion characteristics. University of Southern California, California
34. Lin LJ, Ortega A (1998) Bit-rate control using piecewise approximated rate-distortion characteristics. IEEE Trans Circuits Syst Video Technol (TCSVT) 8:446–459
35. Liu S, Kuo CJ (2005) Joint temporal-spatial bit allocation for video coding with dependency. IEEE Trans Circuits Syst Video Technol (TCSVT) 15:15–26
36. Matsumura A, Naito S, Kawada R (2005) Effective rate control method for minimizing temporal fluctuations in picture quality applicable for MPEG-4 AVC/H.264 encoding. IEEE international conference on image processing (ICIP), pp 569–572
37. Nguyen TQ, Yea S, Vetro A (2009) Adaptive fuzzy filtering for artifact reduction in compressed images and videos. IEEE Trans Image Process (TIP) 18:1166–1178
38. Pan F, Li ZG, Lim KP (2004) Adaptive intra-frame quantization for very low bit rate video coding. In: IEEE international symposium on circuits and systems (ISCAS), pp 781–784
39. Pang C, Au OC, Dai J (2011) Frame-level dependent bit allocation via geometric programming. In: IEEE international symposium on circuits and systems (ISCAS), pp 2149–2152
40. Pang C, Au OC, Zou F (2011) An analytical framework for frame-level dependent bit allocation in hybrid video coding, IEEE Intl. Workshop on Multimedia Signal Processing (MMSP), pp: 1–4
41. Seo CW, Moon JH, Han JK (2013) Rate control for consistent objective quality in high efficiency video coding. IEEE Trans Image Process (TIP) 22(6):2442–2454
42. Sullivan GJ, Ohm J, Han WJ, Wigand T (2012) Overview of the high efficiency video coding (HEVC) standard. IEEE Trans Circuits Syst Video Technol (TCSVT) 22(12):1649–1668
43. Tang CW (2007) Spatiotemporal visual considerations for video coding. IEEE Trans Multimedia (TMM) 9(2):231–238
44. Tang CW, Chen CH, Yu YH et al (2006) Visual sensitivity guided bit allocation for video coding. IEEE Trans Multimedia (TMM) 8(1):11–18
45. Tian L, Sun Y, Ahmad I (2009) Effective intra-only rate control for H.264/AVC. In: IEEE international conference on image processing (ICIP), pp 3445–3448
46. Tsai WJ, Chou TL (2010) Scene change aware intra-frame rate control for H.264/AVC. IEEE Trans Circuits Syst Video Technol (TCSVT) 20(12):1882–1886
47. Wang S, Ma S, Wang S, Zhao D, Gao W (2013) Rate-GOP based rate control for high efficiency video coding. IEEE J Sel Top Sign Proces 7(6):1101–1111
48. Wang M, Ngan KN, Li H (2015) An efficient frame-content based intra frame rate control for high efficiency video coding. IEEE Signal Process Lett 22(7):896–900
49. Wang Y, Sun J, Ma S et al (2009) Theoretic analysis of Inter frame dependency in video coding. In: 9th Pacific Rim conference on multimedia (PCM) [C], pp 935–939
50. Wang P, Zhang Y, Hu HM (2013) Region classification-based rate control for flicker suppression of I frames in HEVC. In: IEEE international conference on image processing (ICIP), pp 1986–1990
51. Wei Z, Ngan KN (2009) Spatio-temporal just noticeable distortion profile for grey scale image/video in DCT domain. IEEE Trans Circuits Syst Video Technol (TCSVT) 19(3):337–346
52. Wiegand T, Sullivan GJ, Bjøntegaard G, Luthra A (2003) Overview of the H.264/AVC video coding standard. IEEE Trans Circuits Syst Video Technol (TCSVT) 13(7):560–576
53. Yan Y, Sun K (2012) Joint complexity estimation of I-frame and P frame for H.264/AVC rate control. IEEE Trans Circuits Syst Video Technol (TCSVT) 22(5):790–798
54. Yan B, Wang M (2009) Adaptive distortion-based intra-rate estimation for H.264/AVC rate control. IEEE Signal Process Lett (SPL) 16(3):145–148
55. Yang H, Boyce JM (2008) Effective flicker removal from periodic Intra frames and accurate flicker measurement. In: IEEE international conference on image processing (ICIP), pp 2868–2871
56. Yang JX, Wu HR (2010) Robust filtering technique for reduction of temporal fluctuation. IEEE Trans Circuits Syst Video Technol (TCSVT) 20(3):458–462
57. Zhou M, Hu HM and Zhang Y (2014) Region-based intra-frame rate control Scheme for high efficiency video coding. Asian-Pacific signal and information processing association annual summit and conference (APSIPA ASC)
58. Zhou Y, Sun Y, Feng Z et al (2009) New rate-distortion modeling and efficient rate control for H.264/AVC video coding. Signal Process Image Commun (SPIC) 24:345–356

Hai-Miao Hu received the B.S. degree from Central South University, Changsha, China, in 2005, and the Ph.D. degree from Beihang University, Beijing, China, in 2012, all in computer science. He was a visiting student at University of Washington from 2008 to 2009. Currently, he is an associate professor of Computer Science and Engineering at Beihang University. His research interests include video coding and networking, image/video processing, and video analysis and understanding.



Mingliang Zhou received the B.S. degree in applied the School of Computer Science, Sichuan University of Science and Engineering, Zigong, China, in 2008, and the M.S. degree from the School of Computer Science, Guizhou University, Guiyang, China, in 2011. He is currently pursuing the Ph.D degree in computer science and engineering from Beihang University, Beijing, China. His current research interests include video processing and high-efficiency video coding.

Yang Liu received the B.S. degree in applied mathematics from National University of Defense Technology in 1998, the M.S. degree in management science from National University of Defense Technology in 2001, and Ph.D. degree in management science from National University of Defense Technology in 2004. Currently, he is a senior engineer of Beijing Institute of Graphics. His research interests include system engineering, video processing and video networking.

Naiyu Yin is currently pursuing the B.S. degree of Beihang University, Beijing. She is an intern on the subject of video coding and networking in Beijing Key Laboratory of Digital Media.