PAPER

QP Selection Optimization for Intra-Frame Encoding Based on Constant Perceptual Quality

Chao WANG[†a)], ***Student Member***, **Xuanqin MOU**[†], ***and*** **Lei ZHANG**[††], ***Nonmembers***

SUMMARY In lossy image/video encoding, there is a compromise between the number of bits and the extent of distortion. Optimizing the allocation of bits to different sources, such as frames or blocks, can improve the encoding performance. In intra-frame encoding, due to the dependency among macro blocks (MBs) introduced by intra prediction, the optimization of bit allocation to the MBs usually has high complexity. So far, no practical optimal bit allocation methods for intra-frame encoding exist, and the commonly used method for intra-frame encoding is the fixed-QP method. We suggest that the QP selection inside an image/a frame can be optimized by aiming at the constant perceptual quality (CPQ). We proposed an iteration-based bit allocation scheme for H.264/AVC intra-frame encoding, in which all the local areas (which is defined by a group of MBs (GOMBs) in this paper) in the frame are encoded to have approximately the same perceptual quality. The SSIM index is used to measure the perceptual quality of the GOMBs. The experimental results show that the encoding performance on intra-frames can be improved greatly by the proposed method compared with the fixed-QP method. Furthermore, we show that the optimization on the intra-frame can bring benefits to the whole sequence encoding, since a better reference frame can improve the encoding of the subsequent frames. The proposed method has acceptable encoding complexity for offline applications.

key words: *bit allocation optimization, intra-frame encoding, constant quality, perceptual video coding*

1. Introduction

The goal of lossy image/video encoding is to minimize the loss on the quality of the reconstructed image/video under the constraint of bit budget, or to use as few as possible bits to encode the image/video at a certain level of quality. Generally, the more bits that an encoder uses to encode a source (such as a block in an image, or a frames in a video), the less distortion on the source it brings to, which is known as the rate-distortion (R-D) property of the source [1]. However, the R-D property varies with the content contained in the sources. Thus, it is necessary for an encoder to determine how many bits should be allocated to each source in the coding object so that the overall encoding can achieve the best performance [2]. Generally, bit allocation is controlled by the quantization parameters (QP) used to encode the sources [3].

In modern hybrid video coding standards (e.g. H.264/AVC [4] or HEVC [5]), intra-frame plays an important role, where the encoding of all the other frames in a group of picture (GOP) depends on it directly or indirectly due to the inter prediction. The optimization on intra-frame encoding can not only improve the encoding performance on itself, but also benefit the encoding of the whole GOP, because a better reference frame can further improve the performance of inter prediction [6]. In this sense, optimizing the bit allocation inside intra-frames is valuable.

Most of the existing works optimize the bit allocation in rate control algorithms [7]–[10]. However, all of these methods only control the QP for blocks in inter-frames, and use fixed-QP to encode the intra-frames. In these methods, the object for optimization is to minimize the average distortion of the sources (MINAVE) [11], and it is usually assume that the sources are encoded independently. When sources are encoded independently, the MINAVE based optimal bit allocation can be solved easily by using the Lagrangian multiplier method [12]. However, when sources are dependent, the Lagrangian multiplier method is quite difficult to use, details about this have been discussed in [13]. The much more complicated dynamic programming (DP) method is usually used to solve the optimization problem when sources are encoded dependently [6], [11], [14], [15]. It has been proved that the computational complexity of DP method increases exponentially with the number of dependent sources [11]. Due to the intra prediction, the number of dependent sources in an intra-frame in H.264/AVC is equal to the number of macro blocks (MBs) in a slice [4], which makes it very time consuming to use the DP method in intra-frame encoding. In this sense, MINAVE-based bit allocation methods are difficult to be used in intra-frame encoding. The fixed-QP method is still the commonly used method to encode intra-frames for now.

Another drawback of the MINAVE criterion could be its incompatible with the human perception. Since the encoded video is to be shown to the human, the optimization of the bit allocation should improve the perceptual quality of the reconstructed frames. However, it was pointed out in [11] that MINAVE may result in sudden variation in quality, which is annoying. The study in [16] had shown that the human vision prefers to pay more attention to the signal which is distinctly different from the others. As a result, it is better to encode all the sources to have similar level of distortion, so that no individual source is extremely distorted to attract the viewer's attention.

The MINMAX criterion aims to minimize the max-

Manuscript received May 11, 2015.
Manuscript revised September 15, 2015.
Manuscript publicized November 17, 2015.

[†]The authors are with the Institute of Image Processing and Pattern Recognition, Xi'an Jiaotong University, Xi'an, China.
[††]The author is with the Dept. of Computing, The Hong Kong Polytechnic University, Hong Kong, China.
a) E-mail: cwang.2007@stu.xjtu.edu.cn

DOI: 10.1587/transinf.2015EDP7183

imum distortion of the sources in bit allocation, and it generally leads to constant quality under a given bit constraint [17]. However, almost all the existing MINMAX-based methods [18]–[21] focus on minimizing the quality variation among frames, and none aims for constant quality inside a frame. This is understandable because the distortion/quality metrics that been used in these works are usually the mean square error (MSE) or the peak signal to noise ratio (PSNR). It is well known that MSE/PSNR can not represent the human visual perception [22]. Forcing the adjacent frames in a GOP to have similar MSE/PSNR is reasonable, because they contain almost the same scenes. But it is quite different for the blocks or local areas inside a frame, the contents of these sources can be very different even if they are adjacent to each other, and forcing them to have similar MSE/PSNR is meaningless.

In our opinion, the optimization of bit allocation in intra-frames should aim for constant perceptual quality (CPQ) among the sources inside the frame. Perceptual quality metrics are necessary to be used in this situation, and the great improvement of the research on image quality assessment (IQA) in the past decade creates opportunities for this purpose [23]. There are already many works that use IQA metrics in video encoding and achieves promising results [10], [24]–[28]. Our previous works have shown the advantage of using IQA metrics in intra-frame constant quality encoding [29]–[31]. In this paper, we propose an CPQ-based method for the bit allocation of intra-frames in H.264/AVC. The proposed method encodes all the local areas in an intra-frame, which are named by group of MBs (GOMBs), to have similar perceptual quality. The purpose of constant perceptual quality among GOMBs is achieved by using an iteration-based encoding process, and the distortion/quality of GOMBs are measured by the widely Structual SIMilarity (SSIM) index [32]. The proposed method can be used to replace the standard intra-frame encoding in H.264/AVC with acceptable encoding complexity. It is useful for the applications that can be encoded offline, such as still image encoding, video storage and video streaming, etc.

The rest of this paper is organized as follows: the CPQ-based bit allocation method is introduced in Sect. 2; the experiments and results are shown in Sect. 3; more discussions about the proposed method are given in Sect. 4; and finally we draw the conclusions in Sect. 5.

2. Bit Allocation Based on Constant Perceptual Quality

In this section, we propose a bit allocation method for intra-frame encoding in H.264/AVC, which is called constant perceptual quality (CPQ). The idea is to encode the local areas in the intra-frame to similar perceptual quality under given bitrate constraint. We define the local area first, and then propose an iteration-based bit allocation method aiming for CPQ. The SSIM index is used to evaluate the quality of the local areas.

2.1 Group of Macro Blocks

To design the method for achieving constant perceptual quality inside a frame, we first define the concept of a local area. Since the basic encoding unit of H.264/AVC is an MB, We define a local area as a group of macro blocks (GOMB). An GOMB is a square area consists of several adjacent MBs. Let $MB_{i,j}$ be the j^{th} MB of the i^{th} row in a frame, we define the GOMB whose upper left is $MB_{i,j}$ as

$$GOMB_{i,j} = \{MB_{u,v} \mid u = i, \ldots, i+N-1, v = j, \ldots, j+N-1\}, \quad (1)$$

where, N is the size of GOMB. We set N to 4 in this paper. A legal GOMB should satisfy the following conditions: $1 \leq i \leq H_{MB} + 1 - N$ and $1 \leq j \leq W_{MB} + 1 - N$, where H_{MB} and W_{MB} are the frame height and width in MB, respectively. We denote the set of all the legal GOMBs as S_G.

Figure 1 illustrates the definition of $GOMB_{i,j}$ and shows the example of three other legal GOMBs, where the dashed boxes refer to the MBs.

2.2 CPQ-Based Bit Allocation

An iteration-based bit allocation method for intra-frame encoding is developed to pursue constant perceptual quality among GOMBs. The flowchart of the method is shown in Fig. 2. The idea is to find a proper target quality Q_T that when all the GOMBs in the intra-frame are encoded to Q_T, the encoding results (total number of bits or frame quality)

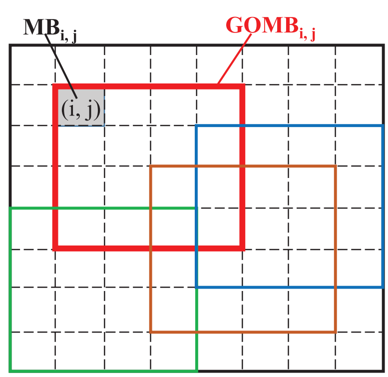


Fig. 1 The definition of $GOMB_{i,j}$ (the area outlined by the red square) and three legal GOMBs (the areas outlined by the green, brown and blue square).

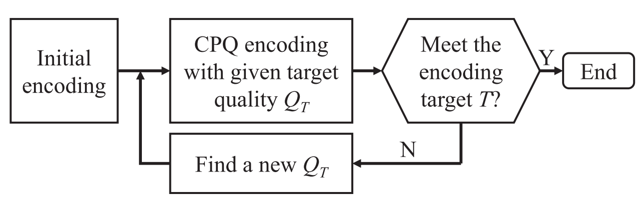


Fig. 2 Flowchart of the CPQ-based bit allocation method for intra-frame encoding.

meet the given constraint T. The encoding is called rate constrained when T is the constraint on bit number for encoding the frame, and quality constrained when T is the constraint on frame quality.

In this paper, we try to improve the perceptual quality of the reconstructed frames, and the object of constant perceptual quality is for this purpose. So, we use Multi-scale SSIM (MS-SSIM) index [33] to measure the perceptual quality, which has been proved to have better correlation with human assessment.

2.2.1 Initial Encoding

In the initial encoding, the fixed-QP method is used to encode the frame, in which all the MBs are encoded by a fixed QP QP_0. Suppose the quality of $GOMB_{i,j}$ after the initial encoding is $Q_G^{(0)}(i,j)$, the initial target quality for the following CPQ encoding is calculated as

$$Q_T^{(0)} = \frac{\sum_{GOMB_{i,j} \in S_G} Q_G^{(0)}(i,j)}{numel(S_G)}, \tag{2}$$

where $numel(S)$ is a function to get the number of elements in set S. QP_0 should be carefully determined so that the result (bits or quality) of the initial encoding is as close to T as possible. Given T, there exist a lot of R-Q (or D-Q) models that can be used to help determining the QP_0 [7], [8], [34].

2.2.2 CPQ Encoding with Given Q_T

Given a target quality Q_T for the GOMBs, the CPQ encoding is performed to pursue the constant perceptual quality. The process of CPQ encoding is shown in Algorithm 1. In the algorithm, θ is the adjusting step of the QPs for the MBs in an GOMB. Function $direct(\cdot)$ is used to determine to which direction the QPs should be adjusted, which is defined as

$$direct(q, Q_T) = \begin{cases} 1 & q > Q_T + \Delta Q_T \\ -1 & q < Q_T - \Delta Q_T \\ 0 & \text{other} \end{cases} \tag{3}$$

Algorithm 1 CPQ Encoding

Require: Q_T, Q_G, QP_B
Ensure: T_{act}, Q_G, QP_B
$it \Leftarrow 1$, $H_{QPB} \Leftarrow \{QP_B\}$
while do
  $it \Leftarrow it + 1$
  $\forall (i,j) \in S_G$, $M_G(i,j) \Leftarrow \theta \cdot direct(Q_G(i,j), Q_T)$
  $\forall (u,v)$, $\Delta QP(u,v) \Leftarrow round\left(\frac{\sum_{GOMB_{i,j} \in S_{u,v}} M_G(i,j)}{numel(S_{u,v})}\right)$
  $\forall (u,v)$, $QP_B^{(1)}(u,v) \Leftarrow clip3(QP_B(u,v) + \Delta QP(u,v), 0, 51)$
  if $QP_B^{(1)} \in H_{QPB}$ **then**
    break
  else
    $QP_B \Leftarrow QP_B^{(1)}$, $H_{QPB} \Leftarrow \{H_{QPB}, QP_B\}$
  end if
  Encode the frame with QP_B, and get T_{act}, Q_G
end while

Where, q states for the quality (SSIM) of an GOMB, and ΔQ_T is the threshold to terminate the CPQ encoding. Set $S_{u,v}$ contains the GOMBs that $MB_{u,v}$ belongs to, and it is defines as

$$S_{u,v} = \{GOMB_{i,j} | MB_{u,v} \in GOMB_{i,j}\} \tag{4}$$

Function $round(x)$ gets the nearest integer of x, and $clip3(x, a, b)$ clips x into the range of $[a, b]$.

2.2.3 Searching for the Best Q_T

Generally, the result of the CPQ encoding (the number of frame bits and frame quality) changes monotonously with Q_T. Figure 3 shows the relation between Q_T and the result the of first frame of "*bus*" encoded by the CPQ encoding. We can find that, the higher Q_T the GOMBs are encoded to, the higher frame quality the encoding achieves, and the more bits the encoder spends on the frame. Based on the monotonicity discussed above, we propose an adaptive-step method to search for the best Q_T to meet the target T, which is shown in Algorithm 2. In the algorithm, S_{stp} is a set of available Q_T adjusting steps for the encoder, B_{QT} is the boundary condition to terminate the searching of Q_T. Func-

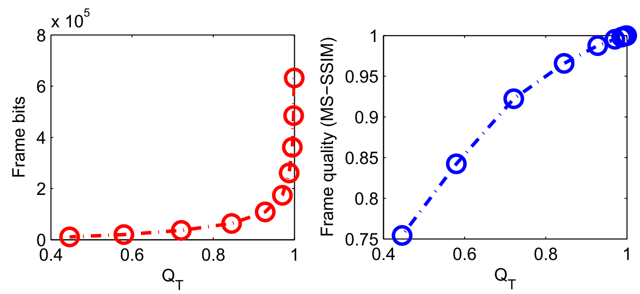


Fig. 3 The results of the 1st frame of "*bus*" encoded by the CPQ encoding with different Q_T.

Algorithm 2 Adaptive-step Method

Require: T, $Q_T^{(0)}$, $Q_G^{(0)}$, QP_0, S_{stp}, B_{QT}
Ensure: Q_T, T_{act}
$Q_T \Leftarrow Q_T^{(0)}$, $Q_G \Leftarrow Q_G^{(0)}$, $QP_B \Leftarrow QP_0$
$H_{QT} \Leftarrow \{\}$, $k \Leftarrow 1$
$[T_{act}, Q_G, QP_B] \Leftarrow CPQ(Q_T, Q_G, QP_B)$
while do
  $sn \Leftarrow comp(T_{act}, T, B_{QT})$
  if $sn == 0$ **then**
    break
  end if
  $Q_T^{(1)} \Leftarrow Q_T + sn \cdot S_{stp}(k)$
  if $Q_T^{(1)} \in H_{QT}$ **then**
    $k \Leftarrow k + 1$
  else
    $Q_T \Leftarrow Q_T^{(1)}$, $H_{QT} \Leftarrow \{H_{QT}, Q_T\}$
    $[T_{act}, Q_G, QP_B] \Leftarrow CPQ(Q_T, Q_G, QP_B)$
  end if
  if $k > numel(S_{stp})$ **then**
    break
  end if
end while

Table 1 Parameter setting for the CPQ-based method.

Parameters	Value
θ	2
ΔQ_T	0.01
S_{stp}	[0.01, 0.005, 0.0025]
B_{QT}	Rate constrained: 0.02 Quality constrained: 0.0005

tion $comp(T_{act}, T, B_{QT})$ is defined in different form depends on the type of constraint. In rate constraint encoding, it is defined as

$$comp(T_{act}, T, B_{QT}) = \begin{cases} 1 & T_{act} < T \cdot (1 - B_{QT}) \\ -1 & T_{act} > T \\ 0 & \text{other} \end{cases} \quad (5)$$

And in quality constraint encoding, it is defined as

$$comp(T_{act}, T, B_{QT}) = \begin{cases} 1 & T_{act} < T \\ -1 & T_{act} > T + B_{QT} \\ 0 & \text{other} \end{cases} \quad (6)$$

2.3 Parameters Setting

There are several encoding parameters in the CPQ-based method, and the parameters setting should give consideration to both the encoding performance and the encoding complexity. In this paper, the parameters are set empirically as it is shown in Table 1. We found that, smaller adjusting steps or more strict thresholds for terminating could not improve more on the encoding performance, but brings much more encoding complexity. More details about the parameter setting will be discussed in Sect. 4.

3. Experiments and Results

The proposed method is tested on JM18.6 [35], the reference software of H.264/AVC. Since there is barely no optimal bit allocation for intra-frame encoding, we compare the proposed method with the recommended fixed-QP method in JM18.6.

3.1 Rate Constrained Encoding

We first evaluate the performance of CPQ-based method in rate constrained intra-frame encoding, where the bits used by the CPQ-based method is constrained to be not more than that used by the fixed-QP method.

3.1.1 Rate-Distortion Performance

Figure 4 shows the R-D curves of the four intra-frames encoded by the CPQ-based method and the fixed-QP method, respectively. For each frame, it is first encoded by the fixed-QP method with QPs of 25, 30, and 35, respectively. And then the CPQ-based encoding is performed based on the results of the fixed-QP method. It is clear that the CPQ-based method can achieve better R-D performance than the fixed-QP method. We can also find that, the performance of the

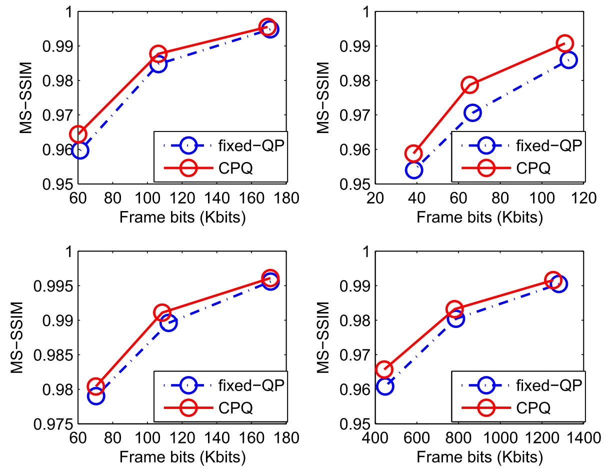


Fig. 4 The R-D curves of four frames which are encoded as intra-frames by the CPQ-based method and the fixed-QP method, respectively. From left to right and from top to bottom: the 1st frame of "*bus*(CIF)", "*container*(CIF)", "*paris*(CIF)" and "*parkjoy*(720P)", respectively.

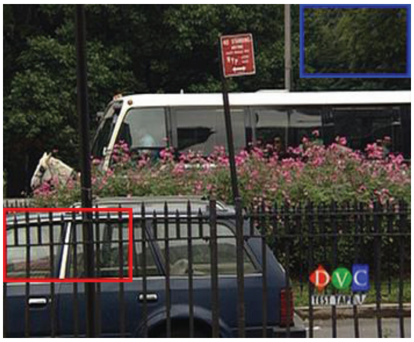
CPQ-based method varies with the frames. We will discuss more about this phenomenon in Sect. 4.

3.1.2 Subjective Quality

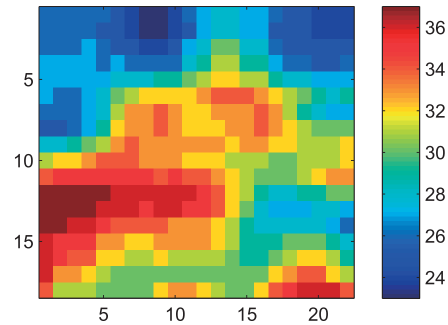
To illustrate the quality improvement by the CPQ-based method more clearly, we show more details of the encoded results of the 1st frame of "*bus*" and "*paris*" in Figs. 5 and 6, respectively. In both the figures, the QP used by the fixed-QP method is 30, and the CPQ-based encoding is setup correspondingly.

Different from the fixed-QP method, the CPQ-based method uses different QPs for different MBs in a frame. The final QPs used for the MBs by the CPQ-based method are show as a map in Figs. 5 (b) and 6 (b), in which the value of the QPs are shown by different colors: 'blue' states for smaller QPs, and 'red' states for larger QPs. Since the total number of bits used by the two methods are the same, we can deduce that the CPQ-based method allocates more bits to the MBs with 'blue' color and less bits to the MBs with 'red' color, compared with the fixed-QP method. It is notable that the bit allocation based on CPQ correlates strongly with the contents of the MBs.

We outline the areas with smaller QPs (and thus are encoded with more bits) by blue rectangles, as shown in Figs. 5 (a) and 6 (a), and name them *Area-blue*. The areas with larger QPs (encoded with less bits) are also outlined by red rectangles, and named by *Area-red*. The reconstructed results of these two kinds of areas by the two methods are compared in Figs. 5 (c), 5 (d), 6 (c) and 6 (d), respectively. We can see that, more details in the *Area-blue*s are retained by the CPQ-based method than the fixed-QP method (e.g. the leaves in Fig. 5 (c), the texture on the table and the wrinkle of the sleeve in Fig. 6 (c)). On the other hand, we can hardly observe any difference between the reconstructed *Area-red*s encoded by the two methods (see Figs. 5 (d) and 6 (d)), though the CPQ-based method uses larger QP (less

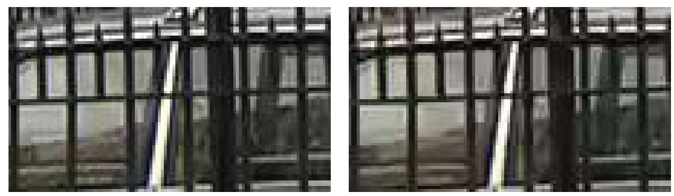


(a) Original frame



(b) QP of MBs used by the CPQ-based method

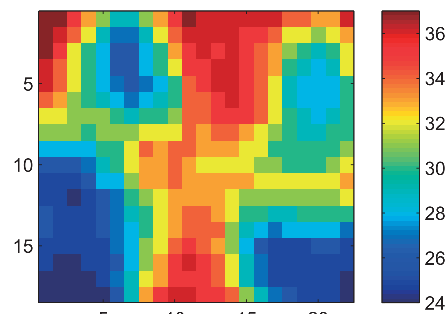
(c) Zoom-in of *Area-blue*. (Left: fixed-QP, SSIM = 0.8335. Right: CPQ, SSIM = 0.9245.)



(d) Zoom-in of *Area-red*. (Left: fixed-QP, SSIM = 0.9526. Right: CPQ, SSIM = 0.9464.)

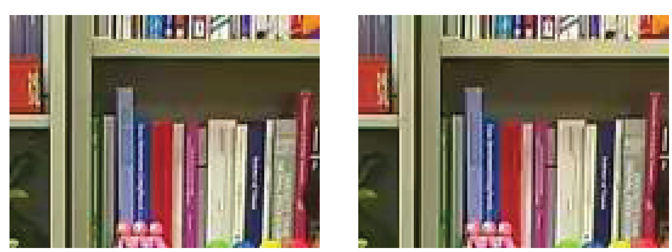
Fig. 5 Comparing the subjective quality of the 1st frame of "*bus*". The QP used by the fixed-QP method is 30, and the number of bits used by the two methods are 108970 (fixed-QP) and 108882 (CPQ), respectively.

(a) Original frame



(b) QP of MBs used by the CPQ-based method

(c) Zoom-in of *Area-blue*. (Left: fixed-QP, SSIM = 0.8972. Right: CPQ, SSIM = 0.9458)



(d) Zoom-in of *Area-red*. (Left: fixed-QP, SSIM = 0.9622. Right: CPQ, SSIM = 0.9531)

Fig. 6 Comparing the subjective quality of the 1st frame of "*paris*". The QP used by the fixed-QP method is 30, and the number of bits used by the two methods are 114924 (fixed-QP) and 111159 (CPQ), respectively.

bits) on this areas. The results illustrate that, the CPQ-based method can achieve better perceptual quality on intra-frames than the fixed-QP method while using the same number of bits.

3.1.3 Quality Variation

Figure 7 shows the quality (SSIM) of all the GOMBs of the 1st frames of “*bus*” and “*pairs*” encoded by the CPQ-based method and fixed-QP method, respectively. Where, the QP used by the fixed-QP method is 30. It is clear that the quality of GOMBs encoded by the fixed-QP method spreads in a wide range, while the CPQ-based method achieves almost constant quality.

In order to compare the quality variation of the results encoded by the two methods numerically, we use the standard deviation of the GOMBs’ quality to describe the extent of the quality variation in a frame. The quality variation achieved by the CPQ-based method and the fixed-QP method are noted as σ_{cpq} and σ_{fix}, respectively. Then, the reduction on quality variation by the CPQ-based method is defined as

$$\Delta\sigma_{cpq} = \frac{\sigma_{cpq} - \sigma_{fix}}{\sigma_{fix}} \times 100\%. \tag{7}$$

The results are shown in Table 2. Thirteen frames from 9 different CIF sequences and 2 720P sequences are used in the test. The QPs used by the fixed-QP method are 25, 30 and 35, and the rate constrained CPQ encoding is setup correspondingly. We can find that the CPQ-based method can greatly reduce the quality variation among GOMBs (45% $\sim$ 60% in average and up to nearly 80%) compared with the fixed-QP method.

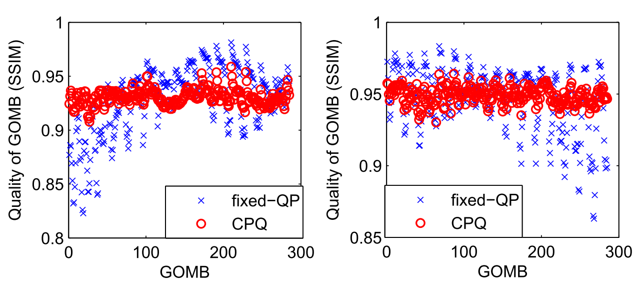


Fig. 7 The quality (SSIM) of the GOMBs in intra-frame encoded by the CPQ-based method and the fixed-QP method (QP = 30), respectively. Left: “*bus*”. Right: “*paris*”.

3.2 Quality Constrained Encoding

In quality constrained encoding, the CPQ-based method is constrained to encode an intra-frame with quality not lower than that encoded by the fixed-QP method. The results are shown in Table 3, in which R states for the number of bits used in the encoding, Q is the quality of the reconstructed frame (measured by the MS-SSIM index). ΔR and ΔQ state for the percentage of bit saving and quality improvement by the CPQ-based method compared with the fixed-QP method, respectively. From Table 3 we can see that, the CPQ-based method uses much fewer bits on most of the frames than the fixed-QP method (4% $\sim$ 9% in average and up to 24.22%), while maintaining the same or even higher frame quality.

We can also see that, as it is found in the rate constrained encoding, the performance of the CPQ-based method varies with diffrent frame contents. Though the CPQ-based method saves bits in most of the conditions and has extremely better performance on several frames (see the data that bolded), it has lower performance on some frames (see the data that underlined).

3.3 Encoding Performance on Sequence

In prediction-based video coding, a better reference frame usually improves the encoding performance of the following frames [6], and thus can improve the encoding performance on the whole sequence. Since it has been proved that the CPQ-based method can achieve better intra-frames, we want to see how much improvement it can bring to the whole sequence encoding.

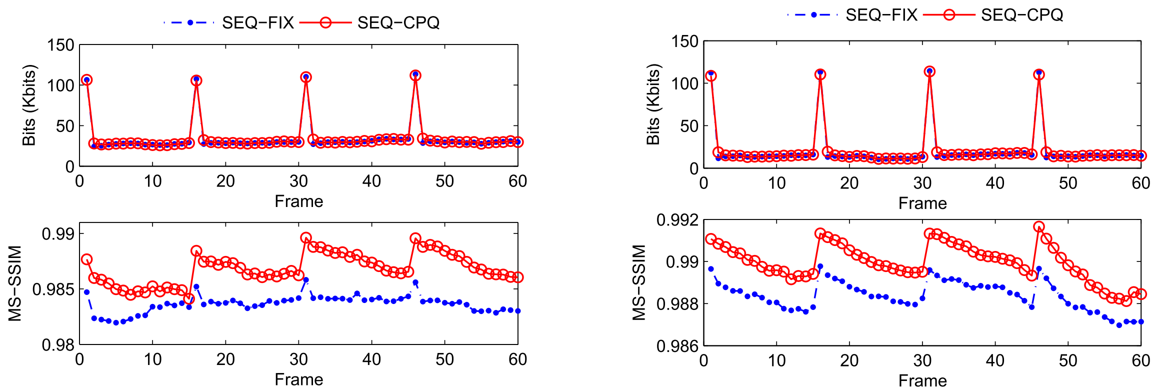
The baseline encoder of H.264/AVC without rate control [35] is used as the benchmark for the comparing, we call it SEQ-FIX because it encode all the frames in the sequence by the fixed-QP method. In the CPQ-based sequence encoding, only the intra-frames are encoded by the CPQ-based method, and the other frames are encoded by the fixed-QP method with QP the same with that used in SEQ-FIX, we

Table 2 Comparing the quality variation on the intra-frames (the 1st frame of different sequences) in rate-constrained encoding.

Sequences	QP = 25		QP = 30		QP = 35	
	σ_{fix}	$\Delta\sigma_{cpq}(\%)$	σ_{fix}	$\Delta\sigma_{cpq}(\%)$	σ_{fix}	$\Delta\sigma_{cpq}(\%)$
bus(CIF)	0.014	−62.1	0.032	−77.5	0.063	−76.7
foreman(CIF)	0.018	−42.5	0.030	−43.6	0.039	−35.7
container(CIF)	0.036	−71.0	0.060	−70.5	0.076	−65.7
flower(CIF)	0.006	−5.0	0.015	−36.7	0.031	−22.9
news(CIF)	0.008	−27.6	0.015	−53.6	0.029	−58.7
pairs(CIF)	0.011	−44.5	0.023	−73.9	0.033	−72.8
silent(CIF)	0.026	−76.6	0.034	−63.7	0.045	−49.9
stefan(CIF)	0.008	−19.8	0.009	−40.2	0.014	−67.6
tempete(CIF)	0.009	−44.4	0.016	−66.2	0.029	−77.3
parkjoy(720P)	0.020	−41.7	0.046	−70.5	0.088	−44.0
mobcal(720P)	0.033	−62.7	0.044	−69.9	0.053	−73.3
Average		−45.3		−60.6		−58.6

Table 3 Comparing the R-D performance on the intra-frames (the 1st frame of different sequences) in quality-constrained encoding.

Sequences	*QP* = 25				*QP* = 30				*QP* = 35			
	fixed-QP		CPQ		fixed-QP		CPQ		fixed-QP		CPQ	
	R	*Q*	Δ*R*(%)	Δ*Q*	*R*	*Q*	Δ*R*(%)	Δ*Q*	*R*	*Q*	Δ*R*(%)	Δ*Q*
bus(CIF)	174869	0.9948	−4.60	0.0003	108970	0.9847	**−10.11**	0.0001	62775	0.9597	**−10.66**	0.0005
foreman(CIF)	87050	0.9918	−1.38	0.0002	49796	0.9828	−1.83	0.0001	29442	0.9706	9.87	−0.0006
container(CIF)	115645	0.9859	**−20.67**	0.0003	68364	0.9706	**−24.22**	0.0001	39598	0.9540	**−16.93**	0.0000
flower(CIF)	213278	0.9983	1.47	0.0002	148217	0.9953	−1.62	0.0002	94914	0.9876	1.10	0.0002
news(CIF)	91587	0.9943	−0.72	0.0003	58910	0.9885	−7.80	0.0000	37844	0.9791	−3.68	0.0000
pairs(CIF)	175108	0.9956	−5.79	0.0000	114924	0.9896	**−10.11**	0.0002	72153	0.9790	−5.88	0.0001
silent(CIF)	126724	0.9915	−7.88	0.0002	68996	0.9784	−6.20	0.0001	37403	0.9541	−0.18	−0.0002
stefan(CIF)	175407	0.9962	−5.50	0.0003	114484	0.9928	−5.39	0.0001	75978	0.9855	4.37	0.0003
tempete(CIF)	201078	0.9966	−0.93	0.0003	129495	0.9916	−5.00	0.0002	78003	0.9779	−5.37	0.0002
parkjoy(720P)	1312366	0.9905	−9.35	0.0005	806815	0.9804	**−12.10**	0.0001	458791	0.9608	**−10.03**	0.0009
mobcal(720P)	1749813	0.9919	**−14.88**	0.0002	1030794	0.9812	**−12.62**	0.0008	582448	0.9601	−7.86	0.0004
Average			−6.38	0.0003			−8.82	0.0002			−4.11	0.0002



(a) Bus (b) Paris

Fig. 8 Frame bits and quality (MS-SSIM) of each frame in sequence "*bus*" and "*paris*".

call this scheme SEQ-CPQ. The intra-frame encoding in SEQ-CPQ is constrained by the bits used by the fixed-QP method on the corresponding frame.

The first 60 frames of each test sequence are encoded, and the coding configuration are set as follows: all intra and inter modes are enabled; one reference frames; each GOP has 15 frames, with one I frame and follows by 14 P frames; high complexity RDO.

3.3.1 Encoding Performance on Each Frame

Figure 8 shows the results on each frame of the CIF sequences "*bus*" and "*paris*", encoded by SEQ-CPQ and SEQ-FIX, respectively. Where the QP used by the SEQ-FIX and the initial QP for CPQ in SEQ-CPQ is 30. We can find that, though we only control the bit number that used on the intra-frame, the SEQ-CPQ encodes the other frames with almost the same number of bits to those used in SEQ-FIX. On the other hand, all the quality of frames are improved obviously by the SEQ-CPQ scheme than SEQ-FIX, though we only optimize the bit allocation on the intra-frames. The results show clearly the benefits that CPQ-based intra-frame encoding can bring to the whole sequence encoding.

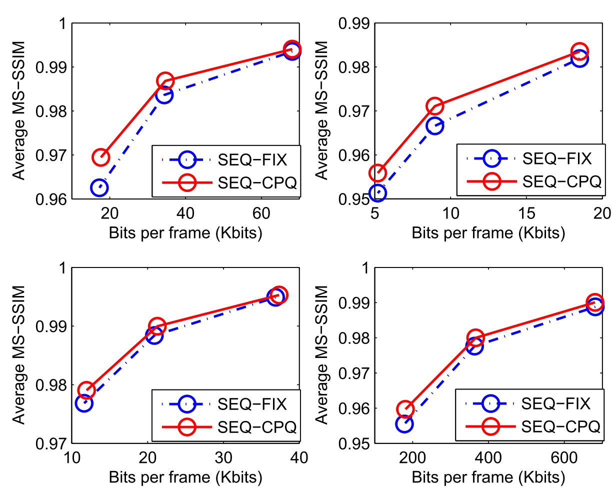


Fig. 9 R-D performance of the encoded sequences. From left to right and from top to bottom: "*bus*(CIF)", "*container*(CIF)", "*paris*(CIF)" and "*parkjoy*(720P)".

3.3.2 R-D Performance

To compare the R-D performance of the two encoding schemes, we draw the R-D curves of their results in Fig. 9.

The QPs used by the SEQ-FIX scheme are 25, 30 and 35. We can see it clearly that the SEQ-CPQ scheme achieves better R-D performance than the SEQ-fixedQP scheme.

3.4 Encoding Complexity

From the algorithms we can see that, the CPQ-based method needs to encode a frame multi-times to find the final solution, this brings much complexity to the encoding. In this subsection, we show the encoding complexity of the proposed method.

3.4.1 Comparing with the DP Method

In Table 4, we count the encoding times of each intra-frame by the CPQ-based method. We can see that, the average encoding times of a frame is about 20 ~ 40.

We can compare the proposed method with the DP method. It has been pointed out that the encoding complexity of DP method is $O(N \cdot |x|^M)$ [11], where N is the number of encoding units, $|x|$ is the number of QPs can be used, and M is the number of dependent encoding units. In intra-frame encoding, no single MB in a Slice is encoded independently, except the first one. In this case, N and M are both the number of MBs in a Slice. As a result, it needs $|x|^S$ times of MB encoding for each Slice by the DP method, where S is the number of MBs in a Slice, and needs about $|x|^S/S$ times of frame encoding for the DP method on each intra-frame.

In the proposed method, the deference between the maximum and minimum QP used for the MBs is about 15, this corresponds to that $|x| = 15$; the size of a Slice is set as a whole frame, in the CIF sequences this corresponds to that $S = 396$. We can find it clearly that, the DP method is too complicate to be used in intra-frame encoding, while the proposed method is a much better choice.

3.4.2 Encoding Complexity on the Whole Sequence

The results in Table 4 show that, the proposed method is still quite time consuming to encode a single frame. However, the total complexity for the whole sequence encoding is not so unacceptable. In sequence encoding, only the intra-frames are encoded multi-times, the other frames are encoded only once. In this sense, the whole complexity for sequence encoding is much lower, which is depends on the size of GOP.

Suppose the average encoding times needed by an intra-frame is N_I, the size of GOP is G, and then the average encoding times N_s needed by a frame in the sequence can be estimated as

$$N_s = \frac{N_I + G - 1}{G}. \tag{8}$$

Based the results in Table 4, when $G = 15$, the encoding complexity of SEQ-CPQ is about 2 ~ 4 times of SEQ-FIX.

Further, since intra-frame encoding doesn't need to do the motion estimation (ME), the encoding time of intra-frames are much lower than that of inter-frames, this can be seen clearly in Table 5. In this sense, the encoding complexity of the proposed method can be even lower.

4. Discussion

The experimental results have shown the advantage of the proposed CPQ-based bit allocation method. In this section, we discuss more about the method in two aspects, the parameters setting and the variational performance on different contents.

4.1 Parameter Setting

The parameters (θ, ΔQ_T, S_{stp} and B_{QT}) in the CPQ-based method are used to control the iteration and the coding performance. Generally, the smaller adjusting steps and thresholds are, the better results will be found by the iteration, but the more times of iteration will be needed. We have given the empirically selected parameters for the method in Table 1, we call this set of parameters as SET-I. To compare, we chose a set of parameters for rate constrained encoding with much smaller adjusting steps and stricter thresholds, which is set as $\theta = 1$, $S_{stp} = [0.01, 0.005, 0.002, 0.001]$, $\Delta Q_T = 0.005$, $B_{QT} = [0.005, 0]$, and we call it SET-II.

Figure 10 shows the R-D curves of the results of the 1st

Table 4 The encoding times of the intra-frames (the 1st frame of different sequences) by the CPQ-based method in rate constrained encoding.

Sequences	$QP = 25$	$QP = 30$	$QP = 35$
bus(CIF)	5	12	57
foreman(CIF)	23	23	38
container(CIF)	37	53	25
flower(CIF)	15	27	36
news(CIF)	6	30	48
pairs(CIF)	5	36	10
silent(CIF)	33	14	33
stefan(CIF)	23	18	37
tempete(CIF)	5	7	73
parkjoy(720P)	29	36	57
mobcal(720P)	26	26	46
Average	18.8	25.6	41.8

Table 5 The average encoding time of intra-frames and inter-frames in sequence encoding with $QP = 30$.

Sequences	I frame (ms)	P frame (ms)
bus(CIF)	181	2712
foreman(CIF)	137	2800
container(CIF)	139	2786
flower(CIF)	164	2881
news(CIF)	136	2797
pairs(CIF)	157	2783
silent(CIF)	143	2733
stefan(CIF)	158	2785
tempete(CIF)	163	2843
parkjoy(720P)	1228	26633
mobcal(720P)	1147	25124

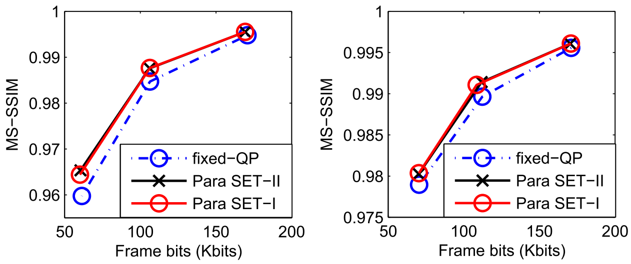


Fig. 10 R-D performance of the intra-frames encoded by CPQ-based method with different parameters in rate-constrained encoding. Left:"*bus*". Right:"*paris*".

Table 6 Encoding times on intra-frames by the CPQ-based method with different parameter sets.

Sequence		SET-II	SET-I
bus	$QP = 25$	45	5
	$QP = 30$	13	12
	$QP = 35$	104	57
paris	$QP = 25$	10	5
	$QP = 30$	50	36
	$QP = 35$	17	10

frame of sequences "*bus*" and "*paris*" encoded by the fixed-QP method, CPQ-based method with SET-I and SET-II, respectively. We can see that the though SET-II uses much finer adjusting steps and stricter thresholds than SET-I, it does not improve the encoding performance more than the latter. On the other hand, the times of iteration that used to encode a frame by the two set of parameters are shown in Table 6. It is clear that, by using parameters SET-II, the CPQ-based method needs much more encoding times of the frame than by using SET-I.

4.2 Performance Depends on the Contents

The results of the experiments have shown that the performance of the CPQ-based method varies with the different contents and QPs, this phenomenon is caused by the intrinsic property of the CPQ-based method.

The results in Figs. 5 and 6 have shown that the bit allocation by the CPQ-based method relates strongly with the frame contents. Survey the four different areas that compared in the figures, we can find that the contents in the *Area-red*s have much stronger structures than those in the *Area-blue*s. This illustrates that the CPQ-based method prefers to allocate more bits to the areas with abundant textures, which are easily to be distorted by quantization, and use larger QPs on the areas with strong structures, which are hard to be distorted. The reduction of bits from the hard to be distorted *Area-red*s brings no more perceptual distortion to the frame, but adding them to encode the *Area-blue*s can easily improve the perceptual quality. As a result, in the frames which have both abundant textures and lots of strong structures, the CPQ-based method can play a role in improving the encoding performance. But, when the frame has few of either the two kinds of areas, the proposed method could not work. This is why the proposed method has lower performance on scenes such as "*foreman*" "*flower*" and "*silent*". Fortunately, most of the natural scenes contains both these two kinds of areas.

5. Conclusion

In this paper, we proposed a CPQ-based method for intra-frame bit allocation, in which all the GOMBs, which are used to define the local areas in a frame, are encoded to have as similar a perceptual quality as possible under the given encoding constraint. The perceptual quality of the GOMBs is measured by the SSIM index. The results show that in most of the frames with different natural scenes, the encoding performance can be improved greatly by the proposed method compared to the commonly used fixed-QP method in intra-frame encoding. Further experiments show that the intra-frame encoding optimized by the CPQ-based method can bring benefits to the whole sequence encoding. The complexity of the proposed method is acceptable for offline video compression applications.

Acknowledgments

This work was supported in part by the National Natural Science Foundation of China under Grant 61172163 and Grant 90920003, and in part by the Research Grants Council, Hong Kong, under Grant PolyU-5315/12E.

References

[1] T. Berger and J.D. Gibson, "Lossy source coding," IEEE Trans. Inf. Theory, vol.44, no.6, pp.2693–2723, 1998.
[2] A. Ortega and K. Ramchandran, "Rate-distortion methods for image and video compression," IEEE Signal Process. Mag., vol.15, no.6, pp.23–50, 1998.
[3] G.J. Sullivan and T. Wiegand, "Rate-distortion optimization for video compression," IEEE Signal Process. Mag., vol.15, no.6, pp.74–90, 1998.
[4] T. Wiegand, G.J. Sullivan, G. Bjontegaard, and A. Luthra, "Overview of the H.264/AVC video coding standard," IEEE Trans. Circuits Syst. Video Technol., vol.13, no.7, pp.560–576, 2003.
[5] G.J. Sullivan, J.-R. Ohm, W.-J. Han, and T. Wiegand, "Overview of the high efficiency video coding (HEVC) standard," IEEE Trans. Circuits Syst. Video Technol., vol.22, no.12, pp.1649–1668, 2012.
[6] K. Ramchandran, A. Ortega, and M. Vetterli, "Bit allocation for dependent quantization with applications to multi resolution and MPEG video coders," IEEE Trans. Image Process., vol.3, no.5, pp.533–545, 1994.
[7] Z. He and S.K. Mitra, "Optimum bit allocation and accurate rate control for video coding via ρ-domain source modeling," IEEE Trans. Circuits Syst. Video Technol., vol.12, no.10, pp.840–849, 2002.
[8] Z.G. Li, F. Pan, K.P. Lim, and G.N. Feng, "Adaptive basic unit layer rate control for JVT," Joint Video Team of ISO/IEC MPEG and ITU-T VCEG, JVT-G012, 7th Meeting, pp.7–14, 2003.
[9] D.-K. Kwon, M.-Y. Shen, and C.-J. Kuo, "Rate control for H.264 video with enhanced rate and distortion models," IEEE Trans. Circuits Syst. Video Technol., vol.17, no.5, pp.517–529, 2007.
[10] T.-S. Ou, Y.-H. Huang, and H.H. Chen, "SSIM-based perceptual rate control for video coding," IEEE Trans. Circuits Syst. Video Technol., vol.21, no.5, pp.682–691, 2011.

[11] G.M. Schuster, G. Melnikov, and A.K. Katsaggelos, "A review of the minimum maximum criterion for optimal bit allocation among dependent quantizers," IEEE Trans. Multimedia, vol.1, no.1, pp.3–17, 1999.
[12] Y. Shoham and A. Gersho, "Efficient bit allocation for an arbitrary set of quantizers," IEEE Trans. Acoust., Speech, Signal Process., vol.36, no.9, pp.1445–1453, 1988.
[13] C. An and T.Q. Nguyen, "Iterative rate-distortion optimization of H.264 with constant bit rate constraint," IEEE Trans. Image Process., vol.17, no.9, pp.1605–1615, 2008.
[14] J. Wen, M. Luttrell, and J. Villasenor, "Trellis-based R-D optimal quantization in H.263," IEEE Trans. Image Process., vol.9, no.8, pp.1431–1434, 2000.
[15] Y. Yang and S.S. Hemami, "Generalized rate-distortion optimization for motion-compensated video coders," IEEE Trans. Circuits Syst. Video Technol., vol.10, no.6, pp.942–955, 2000.
[16] J.M. Henderson, P.A. Weeks Jr., and A. Hollingworth, "The effects of semantic consistency on eye movements during complex scene viewing," Journal of experimental psychology: Human perception and performance, vol.25, no.1, pp.210–228, 1999.
[17] G.M. Schuster and A.K. Katsaggelos, "The minimum-average and minimum-maximum criteria in lossy compression," Vistas in Astronomy, vol.41, no.3, pp.427–437, 1997.
[18] M.F. Ringenburg, R.E. Ladner, and E.A. Riskin, "Global MINMAX interframe bit allocation for embedded video coding," IEEE Data Compression Conference, pp.222–231, 2004.
[19] H.B. Yin, X.Z. Fang, L. Chen, and J. Hou, "A practical consistent-quality two-pass VBR video coding algorithm for digital storage application," IEEE Trans. Consum. Electron., vol.50, no.4, pp.1142–1150, 2004.
[20] N. Cherniavsky, G. Shavit, M.F. Ringenburg, R.E. Ladner, and E.A. Riskin, "Multistage: A MINMAX bit allocation algorithm for video coders," IEEE Trans. Circuits Syst. Video Technol., vol.17, no.1, pp.59–67, 2007.
[21] K. Wang and J.W. Woods, "MPEG motion picture coding with long-term constraint on distortion variation," IEEE Trans. Circuits Syst. Video Technol., vol.18, no.3, pp.294–304, 2008.
[22] Z. Wang and A.C. Bovik, "Mean squared error: Love it or leave it? A new look at signal fidelity measures," IEEE Signal Process. Mag., vol.26, no.1, pp.98–117, 2009.
[23] W. Lin and C.-C.J. Kuo, "Perceptual visual quality metrics: A survey," Journal of Visual Communication and Image Representation, vol.22, no.4, pp.297–312, 2011.
[24] Y.-H. Huang, T.-S. Ou, P.-Y. Su, and H.H. Chen, "Perceptual rate distortion optimization using structural similarity index as quality metric," IEEE Trans. Circuits Syst. Video Technol., vol.20, no.11, pp.1614–1624, 2010.
[25] S.Q. Wang, A. Rehman, Z. Wang, S.W. Ma, and W. Gao, "SSIM-motivated rate-distortion optimization for video coding," IEEE Trans. Circuits Syst. Video Technol., vol.22, no.4, pp.516–529, 2012.
[26] C. Yeo, H.-L. Tan, and Y.H. Tan, "On rate distortion optimization using SSIM," IEEE Trans. Circuits Syst. Video Technol., vol.23, no.7, pp.1170–1181, 2013.
[27] S.Q. Wang, A. Rehman, Z. Wang, S.W. Ma, and W. Gao, "Perceptual video coding based on SSIM-inspired divisive normalization," IEEE Trans. Image Process., vol.22, no.4, pp.1418–1429, 2013.
[28] Z.-Y. Mai, C.-L. Yang, K.-Z. Kuang, and L.-M. Po, "A novel motion estimation method based on structural similarity for H.264 inter prediction," Proc. IEEE International Conference on Acoustics, Speech and Signal Processing, pp.II-913–II-916, 2006.
[29] C. Wang, X. Mou, and L. Zhang, "Block-layer optimal bit allocation based on constant perceptual quality," Proc. SPIE-IS&T Electronic Imaging, vol.8305, 83050E, pp.1–9, 2012.
[30] C. Wang, X. Mou, W. Hong, and L. Zhang, "Block-layer bit allocation for quality constrained video encoding based on constant perceptual quality," Proc. SPIE-IS&T Electronic Imaging, SPIE, vol.8666, 86660J, pp.1–8, 2013.
[31] C. Wang, X. Mou, and L. Zhang, "A study on consistency between MINAVE and MINMAX in SSIM based independent perceptual video coding," IEICE Trans. Inf. & Syst., vol.E98-D, no.7, pp.1417–1421, July 2015.
[32] Z. Wang, A.C. Bovik, H.R. Sheikh, and E.P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," IEEE Trans. Image Process., vol.13, no.4, pp.600–612, 2004.
[33] Z. Wang, E.P. Simoncelli, and A.C. Bovik, "Multiscale structural similarity for image quality assessment," IEEE Conference Record of the Thirty-Seventh Asilomar Conference on Signals, Systems and Computers, vol.2, pp.1398–1402, 2003.
[34] N. Kamaci, Y. Altunbasak, and R.M. Mersereau, "Frame bit allocation for H.264/AVC video coder via Cauchy-density-based rate and distortion models," IEEE Trans. Circuits Syst. Video Technol., vol.15, no.8, pp.994–1006, 2005.
[35] JVT, "H.264/AVC reference software JM18.6." http://iphome.hhi.de/suehring/tml

Chao Wang received the B.Sc. degree in information engineering from the School of Electronic and Information Engineering, Xi'an Jiaotong University, Xi'an, China, in 2007. He is currently pursuing the Ph.D. degree with the Institute of Image Processing and Pattern Recognition, Xi'an Jiaotong University. His research interest focuses on video encoding and perceptual quality of visual signals.

Xuanqin Mou has been with the Institute of Image Processing and Pattern Recognition (IPPR), Electronic and Information Engineering School, Xi'an Jiaotong University, since 1987. He has been an Associate Professor since 1997, and a Professor since 2002. He is currently the Director of IPPR. Dr. Mou served as the member of the 12th Expert Evaluation Committee for the National Natural Science Foundation of China, the Member of the 5th and 6th Executive Committee of China Society of Image and Graphics, the Vice President of Shaanxi Image and Graphics Association. He has authored or co-authored more than 200 peer-reviewed journal or conference papers. He has been granted as the Yung Wing Award for Excellence in Education, the KC Wong Education Award, the Technology Academy Award for Invention by the Ministry of Education of China, and the Technology Academy Awards from the Government of Shaanxi Province, China.

Lei Zhang received the B.Sc. degree in 1995 from Shenyang Institute of Aeronautical Engineering, Shenyang, P.R. China, the M.Sc. and Ph.D degrees in Control Theory and Engineering from Northwestern Polytechnical University, Xi'an, P.R. China, respectively in 1998 and 2001. From 2001 to 2002, he was a research associate in the Dept. of Computing, The Hong Kong Polytechnic University. From Jan. 2003 to Jan. 2006 he worked as a Postdoctoral Fellow in the Dept. of Electrical and Computer Engineering, McMaster University, Canada. In 2006, he joined the Dept. of Computing, The Hong Kong Polytechnic University, as an Assistant Professor. Since Sept. 2010, he has been an Associate Professor in the same department. His research interests include Image and Video Processing, Computer Vision, Pattern Recognition and Biometrics, etc. Dr. Zhang has published about 200 papers in those areas. Dr. Zhang is currently an Associate Editor of IEEE Trans. on CSVT and Image and Vision Computing. He was awarded the 2012–13 Faculty Award in Research and Scholarly Activities. More information can be found in his homepage http://www4.comp.polyu.edu.hk/~cslzhang/.