H.264/SVC parameter optimization based on quantization parameter, MGS fragmentation, and user bandwidth distribution

Xu CHEN[1*], Ji-hong ZHANG[2], Wei LIU[2], Yong-sheng LIANG[2] and Ji-qiang FENG[1]

Abstract

In the situation of limited bandwidth, how to improve the performance of scalable video coding plays an important role in video coding. The previously proposed scalable video coding optimization schemes concentrate on reducing coding computation or trying to achieve consistent video quality; however, the connections between coding scheme, transmission environments, and users' accesses manner were not jointly considered. This article proposes a H.264/SVC (scalable video codec) parameter optimization scheme, which attempt to make full use of limited bandwidth, to achieve better peak signal-to-noise ratio, based on the joint measure of user bandwidth range and probability density distribution. This algorithm constructs a relationship map which consists of the bandwidth range of multiple users and the quantified quality increments measure, QP_e, in order to make effective use of the video coding bit-stream. A medium grain scalability fragmentation optimization algorithm is also presented with respect to user bandwidth probability density distribution, encoding bit rate, and scalability. Experiments on a public dataset show that this method provides significant average quality improvement for streaming video applications.

Keywords: H.264/SVC, Bandwidth distribution, Quality increments, Quantization parameter, MGS fragmentation

1. Introduction

Network bandwidth and error rate changes frequently in wireless networks because of user mobility and dynamic channel conditions. These can have critical impact on video streaming applications because video data are generally very sensitive to delay and error. For this reason, video codecs need to be more aware of the network conditions and should have adaptive bit-rate functions [1]. To address these issues, the Telecommunication Standardization Sector of the International Telecommunication Union (ITU-T) and ISO/IECMPEG (Moving Picture Experts Group) have published a draft H.264/SVC (scalable video codec) standard. The main feature of H.264/SVC is that it provides bandwidth-optimized transmission for video streaming by observing current network conditions [2,3]. There are three types of scalability for H.264/SVC: quality, spatial, and temporal. The codec

provides quality scalability by using medium or coarse grain scalability (MGS/CGS) that delivers quality refinements to a preceding layer representation. Fine grain scalability has not been adopted in the scalable baseline profile, high profile, and high intra-profiles of SVC [4-6] due to its complexity. In terms of coding mechanisms, MGS is almost the same as CGS, except that the coded data corresponding to the quantization step size can be fragmented into 15 layers with MGS, while CGS only provides 1 layer [7].

The objective of H.264/SVC is to achieve scalability without a significant loss in coding efficiency. As the scalable extension of H.264/AVC, the SVC codec inherits all of the coding tools of H.264/AVC. However, the SVC extension distinguishes itself from the scalable systems in prior video coding standards by incorporating an adaptive inter-layer prediction mechanism that has superior coding efficiency. It is worth noting that, although the SVC encoder's operations are non-normative, the codec implementation is still quite flexible [8,9] as long as its bit-streams conform to the specifications.

* Correspondence: chenxu_email@163.com
[1]School of Information Engineering, Shenzhen University, Shenzhen, Guangdong, China
Full list of author information is available at the end of the article

Springer

Thus, selecting an appropriate coding scheme to improve peak signal-to-noise ratio (PSNR) performance of the SVC codec for different applications is a strong research focus in the field of video encoding.

In an existing work on MGS configurations, Gupta et al. [10] examine and analyze the rate distortion (R-D) performance of MGS with different weights and quantization parameters, but did not give a suitable coding scheme instruction for different applications. Burak et al. [11] compared the performance of various MGS fragmentation configurations and recommended combining the use of up to five MGS fragments. They showed that the extraction performance improved at low bandwidth values as the number of MGS fragments in the configuration increased, even as the number of fragments had a negative impact on PSNR performance due to fragmentation overhead. Others have presented R-D models for MGS scalable video coded streams [12,13]. Based on the statistical proprieties of the DCT model, Michele et al. [12] estimated the PSNR of MGS once the rate is given, which requires the knowledge of the two extreme points of the enhancement part of the stream and a measure of the temporal complexity calculated over the raw luminance sequence. On the other hand, Long et al. [13] provided a different R-D model for each kind of picture based on picture type or temporal decomposition level using statistical data of the residual frames.

Recently, some researchers have been taking transmission effects and user environments into consideration for parameter optimization. Dalei and Song [14] Constructed a relationship between the coding scheme and the optimal routing path for transmission against the select coding quantization parameter by using transmission delay and end-to-end distortion, while Haechul and Jung Won [15] proposed a dynamic adaptation scheme of the SVC bit stream that used MPEG-21 DIA (Digital Item Adaptation). Another method transmits the user's environment as represented using the UED tool defined in MPEG-21 DIA—including user information, usage preference, history, accessibility, location, and so on—to a server with an ADTE (Adaptation Decision Taking Engine) [16]. Based on the ADTE's decision, the dynamic extractor can drop and/or crop network abstraction layer units (NALU) by examining only the network abstraction layer (NAL) header. Sketch map of our approach is shown as Figure 1.

In this article, we translate the problem of selecting a suitable coding scheme and making effective use of the video coding bit steam into a problem of overall quality optimization for the user. Our proposed framework consists of two main modules: parameter calculation and performance analysis. First, we construct a mapping relationship between the bandwidth range and quantization parameter QP_e to calculate the appropriate parameters, and then propose a coding rate calculation module for different MGS fragmentation configurations. We then find an optimal solution by iterating a proposed object function to quantify the user's overall quality optimization.

This article has the following sections: Section 2 outlines our proposed framework for parameter optimization. Section 3 describes calculating QP_e and the rate estimation for different number of MGS fragments. Section 4 describes the coding performance analysis for parameter selection. Finally, Section 5 demonstrates the experimental results of our proposed framework with conclusions provided in Section 6.

2. Problem formulation

SVC provides a good solution platform for different applications. In the H.264/SVC coding framework, the quality increment as defined by the quantization parameter QP_e correlates with the quality fineness that the enhancement layer bit stream can get to, while MGS fragmentation correlates to scalability and PSNR performance. Therefore, it is important to select a suitable QP_e and set an appropriate number of MGS fragmentations in the codec. Official and operator testing, and statistics [17,18] provide the typical user bandwidth ranges and probability density distribution functions for different cities and applications. Because bit streams have similar scalability and performance when coding with the same number of MGS fragments (such as 3, 7, 6 and
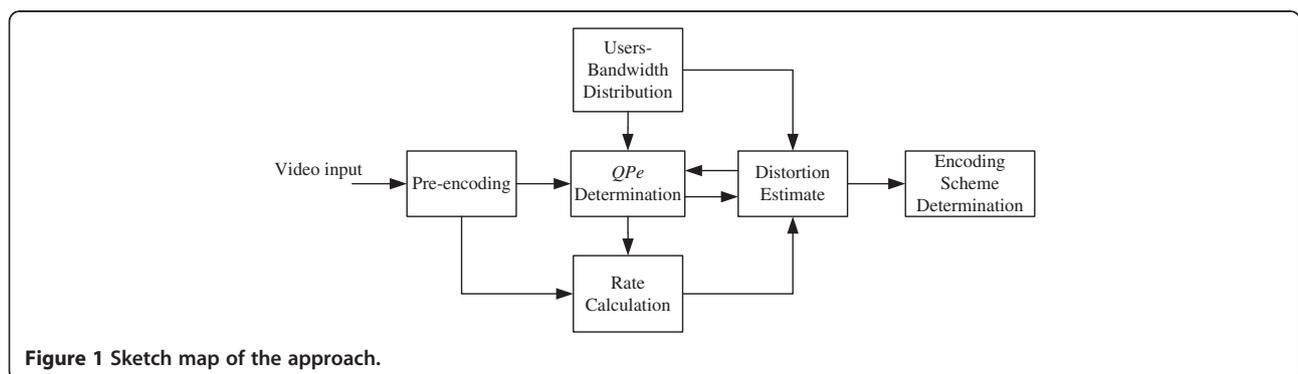

Figure 1 Sketch map of the approach.

5, 3, 8, where the MGS stack is split into 3 MGS fragments), in this article we concentrate on optimizing the quantization parameter QP_e and n, which denotes as the number of MGS fragments.

Given a user's bandwidth probability density distribution $f(x)$, we can select the appropriate coding scheme for an application by multi-coding with a different quantization parameter and number of MGS fragments. The quality evaluation criterion for multiple users can be calculated according to the average video quality of all the users

$$Q = \frac{1}{N} \sum_{i=1}^{N} Q_i \qquad (1)$$

where Q_i is the video quality of user i, and N is the number of users in the bandwidth range. If the coding scheme results in a higher Q using (1), it means that the coding scheme is more suitable for users.

However, because of the high degree of flexibility in the values of QP_e and number of MGS fragments n in the selection, it is difficult to select a suitable coding scheme by multi-coding that does not also dramatically increase encoding time and computational complexity. To solve this problem, we propose an effective algorithm in this article that can find suitable values of QP_e and n for applications, while still achieving low complexity.

3. Calculation of parameter

Pre-encoding framework is shown in Figure 2. At the beginning of several frames (taking the first group of pictures (GOP) as a unit), we set the $n = 15$ for the initial configuration. A 16th fragment is added to the MGS stake that is coding independent, and includes all transformation coefficients data; i.e., the quality increment versus coding using CGS. We can determine the details of QP_e and rate estimation for each different number of MGS fragments n using the coding information of the GOP, which is useful for the PSNR performance calculation conducted in Section 4.

3.1. Qe Determination

The quantization parameter proposed by Liu et al. [19] describes the relationship between the total amount of bits of both texture and non-texture information. The statistical information of the residual frame, such as the actual mean absolute difference (MAD), changes with the QP value adjustment even as the QP value influences the motion information. Through extensive experiments, Long et al. [13] formulated quadratic (I/P frame) and linear (B frame) rate quantization (R-Q) models for the rate estimation of these increments

$$\begin{cases} R_{IP} = \left(\dfrac{c_1}{QP} + \dfrac{c_2}{QP^2} \right) \times MAD \\[2mm] R_B = \dfrac{c}{QP} \times MAD \end{cases} \qquad (2)$$

In the first GOP encoding, we set the n=15. c_1, c_2, and c are models coefficients which need to be initialized for the I/P/B frame R-Q models by using (2), the coding information of the I/P/B coding rate, and the MAD of the increments. Thus, the coding rate is the sum of the coding rates of the I/P and B frames.
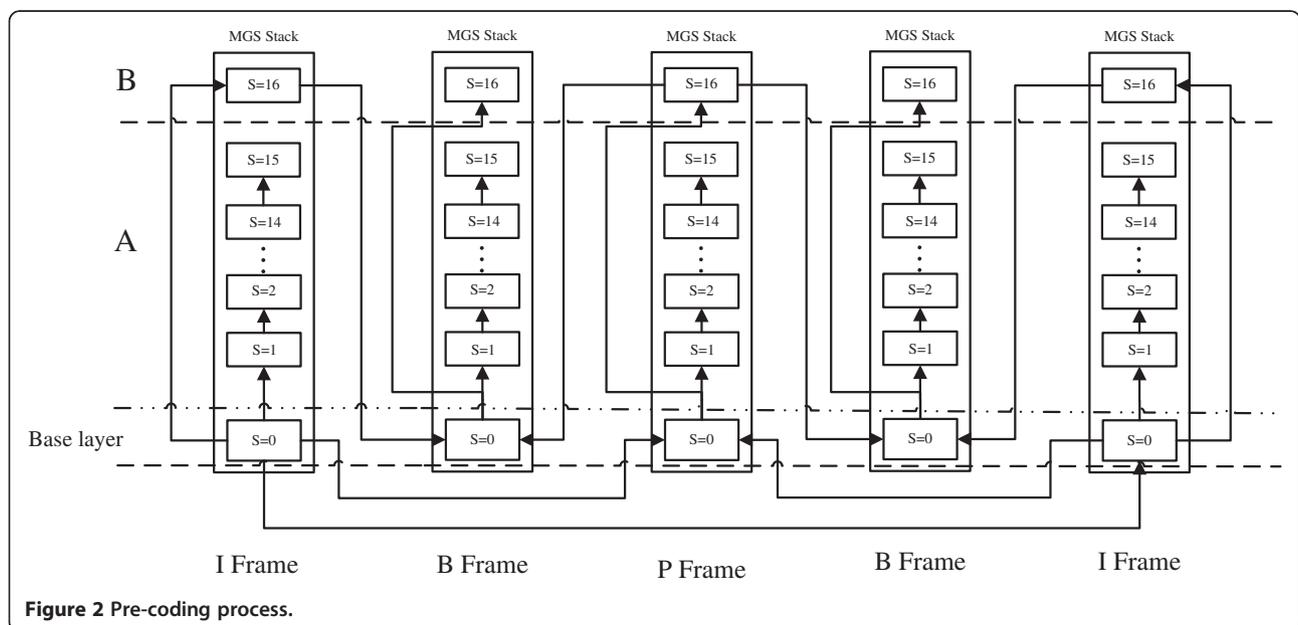

Figure 2 Pre-coding process.

We can calculate the video coding rate R_{IPB} as

$$R_{IPB} = R_{base} + R_{Menh} \qquad (3)$$

$$R_{Menh} = R_{Ienh} + R_{Penh} + R_{Benh}$$

$$
\begin{aligned}
= \sum_u &\left[\sum_d \left(\frac{c_{1d}}{QP_e} + \frac{c_{2d}}{QP_e^2} \right) \times MAD_{ed} \right] \\
&+ \sum_v \left[\sum_q \left(\frac{c_{1q}}{QP_e} + \frac{c_{2q}}{QP_e^2} \right) \times MAD_{eq} \right] \\
&+ \sum_h \left[\sum_s \frac{c_s^e}{QP_e} \times MAD_e \right]
\end{aligned}
\qquad (4)
$$

Here, u, v, and h are the coding numbers of the I, P, and B frame increments, respectively, and R_{base} is the coding rate size of the base layers and is closely related to QP_b. Note that we should select a suitable QP_b that satisfies all users. Therefore, we select QP_b so as to make R_{base} close to the lower bound of the bandwidth distribution of the users. R_{Menh} is the coding rate of the quality increments, which is equivalent to the sum of the coding rate of the quality increments of the I/P/B frames ($R_{Ienh}/R_{Penh}/R_{Benh}$). In (4), c_{1d} and c_{2d} are quadratic R-Q model coefficients for different coding region units of the I frame, while c_{1q} and c_{2q} are quadratic R-Q model coefficients for different coding region units of the P frame. Finally, QP_e is the quantization step of the quality increments and d, q, and s are the coding region numbers of the I/P/B frames, respectively.

To make effective use of the video coding bit steam, we build a mapping relationship between QP_e and the coding rate by selecting a suitable QP_e such that the coding rate is close to the upper end of the users' bandwidth R_{up}

$$R(QP_e, MGS) \rightarrow R_{up} \qquad (5)$$

Note that QP_e of the fragments in the MGS stack should be the same because of the 4×4 integer transformation.

Also note that, for different MGS stacks, QP_e can be different. Mathias Wien found a difference of 1 to 2 between successive temporal levels and a refinement for the key pictures [9]. In this case, the mapping relationship is defined as

$$
\begin{cases}
R\left(QP_e^0 \ldots QP_e^n, MGS\right) \rightarrow R_{up} \\
QP_x = QP_0 + (t \neq 0 ? 2 : 0) - 1.7 \times (N - 1 - i) \\
QP_e^i = min(51, max(0, round(QP_x))) \\
\qquad\qquad 0 \leq i \leq N - 1
\end{cases}
$$
$$(6)$$

Where QP_0 is a given quantization parameter, N is the number of temporal layers, and QP_e^i is the quantization parameter for temporal level t ($0 \leq i \leq N - 1$).

3.2. Rate Estimation for Different n

After determining QP_e, we must optimize the number of MGS fragments n by first calculating the coding rate when coding with different number of MGS fragments. In the MGS coding mechanism, the 16 coefficients in every macro block are determined by means of a 4×4 macro-block data integer transformation, so the data importance of each of the coefficients will be different [20]. We define the vector $K = [k_1, k_2, \ldots k_i, \; i = 1, 2, \ldots 15]$ according to the coefficient number (weight) that each slice contains, which represents the importance of the fragment in quality increments. The vector element is a percentage of the fragment quality in quality increments, and the calculation process, shown in Figure 3, is defined aswhere the PSNR is the image reconstruction quality of the bit stream including the base and increment coding bit rates (R_{IPB}), $PSNR_b$ is the image reconstruction quality of the bit-stream that only includes the base layer coding bit rate (R_{base}), $PSNR_i$ is the bit stream quality without slice i after bit extraction, and k_i is the percentage of fragment i quality in quality increments.
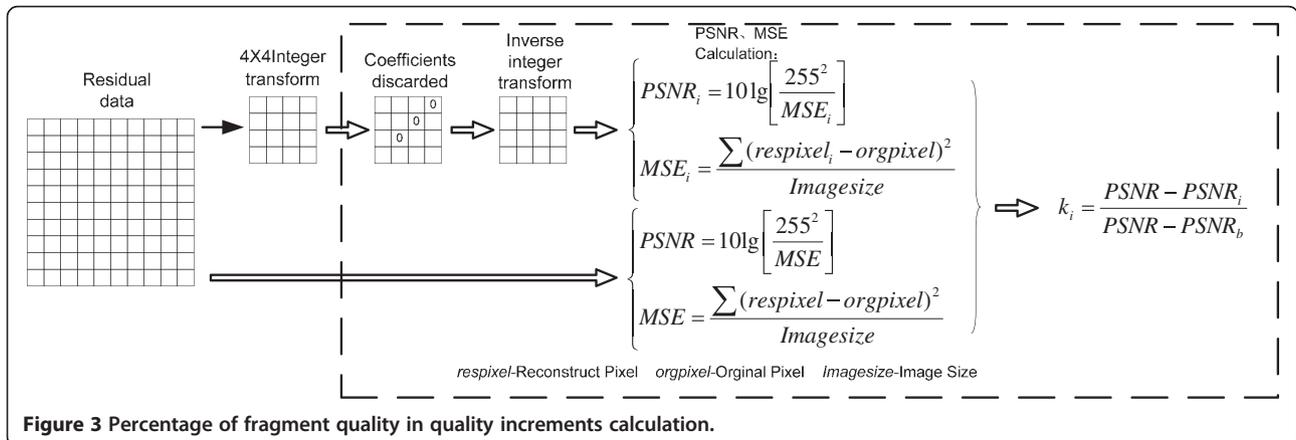

Figure 3 Percentage of fragment quality in quality increments calculation.

Through extensive experiments, we have determined that there is a linear relationship between the coding bit rate and the percentage of fragment quality in quality increments. This correlation is due to the sub-streams having the same coding mechanism and configuration (MGS weight) as the original bit stream. According to this linear relationship, the coding rate R_{mgsi} with the number of MGS fragments n can be calculated as

$$\frac{R_{top} - R(Sub(R_i))}{R_{top} - R(Sub(R_1))} = \frac{R_{top} - R_{mgsi}}{R_{top} - R_{cgs}} \quad (7)$$

where R_{top} is the coding bit rate when coding with $n = 15$, R_{cgs} is the coding bit rate when adopting a CGS coding mechanism (equivalent to $n = 1$), and $R(Sub(R_i))$ is the bit rate of the sub-stream that contains all increments with quality level $<= i$.

Actually, we initial QPe which makes the coding rate close to the upper of users' bandwidth region when $n = 15$ in Section 3.1. The number of MGS fragments n selection after QP_e determination will result in coding rate decrease.

4. Performance for various coding schemes
4.1. Performance analysis
To optimize overall video quality for all users, the optimization function can be written as

$$max\left[\frac{1}{N}\sum_{i=1}^{N}Q(R, R_i)|M\right] \quad M{\in}K \quad (8)$$

Where N is the number of users in the bandwidth range, K is the set of coding configurations, $Q(.)$ is each user's video quality, and R is the encoding bit rate with R_i the bandwidth restriction of user i when adopting coding configuration M. To maximize this objective function, we must consider the scalability and video quality when adopting different number of MGS fragments n by using the coding bit rate determined in Section 3.2. We accomplish this by calculating the scalability of the bit streams for different temporal levels according to the number of coding frames and quality increments. By combining this information with bandwidth probability density distribution of the users, we can optimize the number of MGS fragments n through the following process:

Step 1: Set i as the temporal level and H_i as the coding frame corresponding to this temporal level. This means that M is the number of quality increments and we can calculate the fidelity scalability L of the coding bit streams as

$$L = \sum_{i=1}^{Tn} H_i \times M \quad (9)$$

Because of inter-frame prediction, the data importance of the different temporal levels is different. T_n is the maximum temporal level. As usual, the data level priorities should be $I > P > B_1 > B_2 \ldots$ where B_1 and B_2 denote increments of temporal levels 2 and 3, respectively.

Step 2: Obtain the average interval of fidelity scalability for the different temporal levels by mapping the video quality mapping and calculating the fidelity scalability using

$$\Delta = \begin{cases} \Delta_{Tn} = \dfrac{R_M R_{Tn}}{R_{top}K_{Tn}} \\[2mm] \Delta_{Tn-1} = \dfrac{R_M R_{Tn-1}}{R_{top}K_{Tn-1}} \\ \quad\vdots \\ \Delta_1 = \dfrac{R_M R_1}{R_{top}K_1} \\[2mm] K_{Ti} = H_{Ti} \times M \end{cases} \quad (10)$$

where R_i is the coding bit-rate of temporal level i when coding with $n = 15$, K_i is the fidelity scalability of temporal level i when $n = M$. T_n is the maximum temporal level and R_M is the bit-rate when coding with $n = M$. R_{top} is the bit-rate when $n = 15$, and Δ_i is the average interval of fidelity scalability for temporal level i when $n = M$.

Step 3: The optimization of the number of MGS fragments n is a balance between the coding bit rate and the fidelity scalability according to the users' bandwidth probability density distribution function $f(x)$. Thus, (8) can be rewritten as

$$\begin{cases} max(\sum_{m=1}^{B-1}\int_{x=D}^{D+\Delta_{m+1}} f(x)Q_M(D)dx|M) \\[2mm] Q_M(s) = Q\left[Stm\left(R_{top} - \dfrac{(R_M - s)(R_{top} - R_b)}{R_M - R_b}\right)\right] \\[2mm] D = R_b + \sum_{i=1}^{m}\Delta_i \\[2mm] 1 \le M \le 15 \end{cases} \quad (11)$$

Where $Q_M(s)$ is the video quality for $n = M$ when the user's bandwidth $<= s$, R_b is the coding bit-rate of the base layer, and D is the interval between points of the scalable fidelity. $Stm(v)$ is the sub-stream when coding

bit rate of the sub-stream close to v. The algorithm iterates step 3 until the algorithm reaches a predefined quality requirement.

The block diagram visualizing the entire algorithm described above is shown in Figure 4. The method updates the quantization parameter QP_e and iteratively solves (11) to optimize the number of MGS fragments n to determine the coding configuration that maximizes overall quality for all users.

4.2. Computational complexity analysis

Compared to traditional methods, coding configuration optimization inevitably increases complexity. In our algorithm, the cost of this complexity is mainly reflected in calculation time and added coding. The cost of the added coding is in coding the CGS layer and the QP_e initialization of the first GOP coding of each intra-period. The proportion in times PT can be written as

$$PT = \frac{\frac{F}{JM} \times (16 - L)M \times T_e}{F \times (LT_e + T_b)} = \frac{T_e \times (16 - L)}{J \times (LT_e + T_b)} \quad (12)$$

where T_e is the coding time of an increment and T_b is the coding time of the base layer. M is the frame number of a GOP, J is GOP number of an intra-period, F is the number that frames to be encoded, L is the number of MGS fragments when coding in traditional.

Compared with the cost of added code, the calculation cost is more trivial to determine since there is no need to transform this process. According to (7), (10), and (11), the cost of calculation is

$$O\left[(K + 5)\frac{UF}{JM} + (K + 15)\frac{VF}{JM}\right] \quad (13)$$

where K is the point number of scalability, and U and V are addition and multiplication operations, respectively.

5. Experimental results

In this section, we evaluate our approach using a public video dataset [21] consisting of 28 video clips (QCIF, CIF, and VGA), and the joint scalable video model (JSVM) platform [22]. As shown in Table I, these video clips mainly contain video sequences in genres such as "Foreman," "Bus," and "Mobile." The main objective of the three experiments conducted using these clips is to evaluate whether our approach achieves an appropriate coding configuration to improve overall video quality for users. We randomly selected nine video clips from the public video dataset to perform the process test, and selected the "Mobile" and "Bus" video sequences for
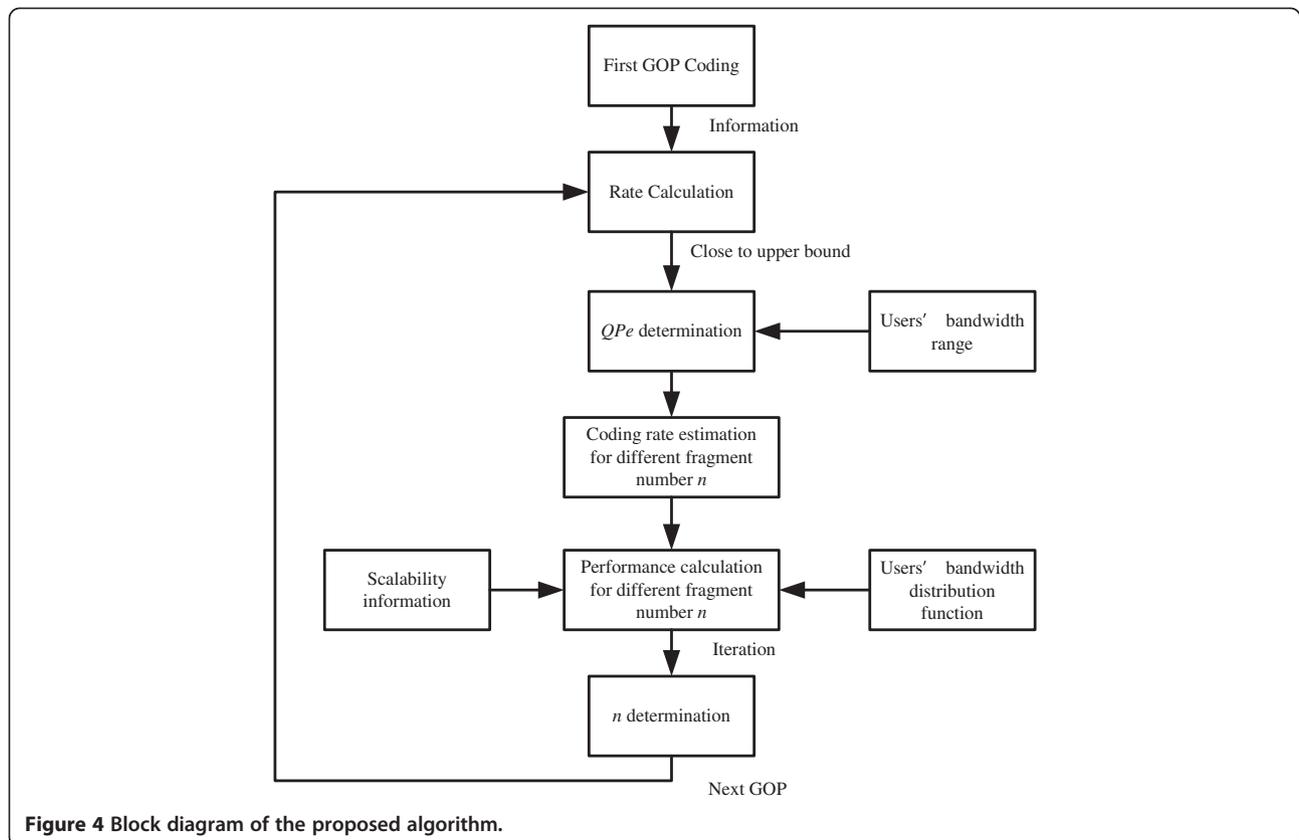
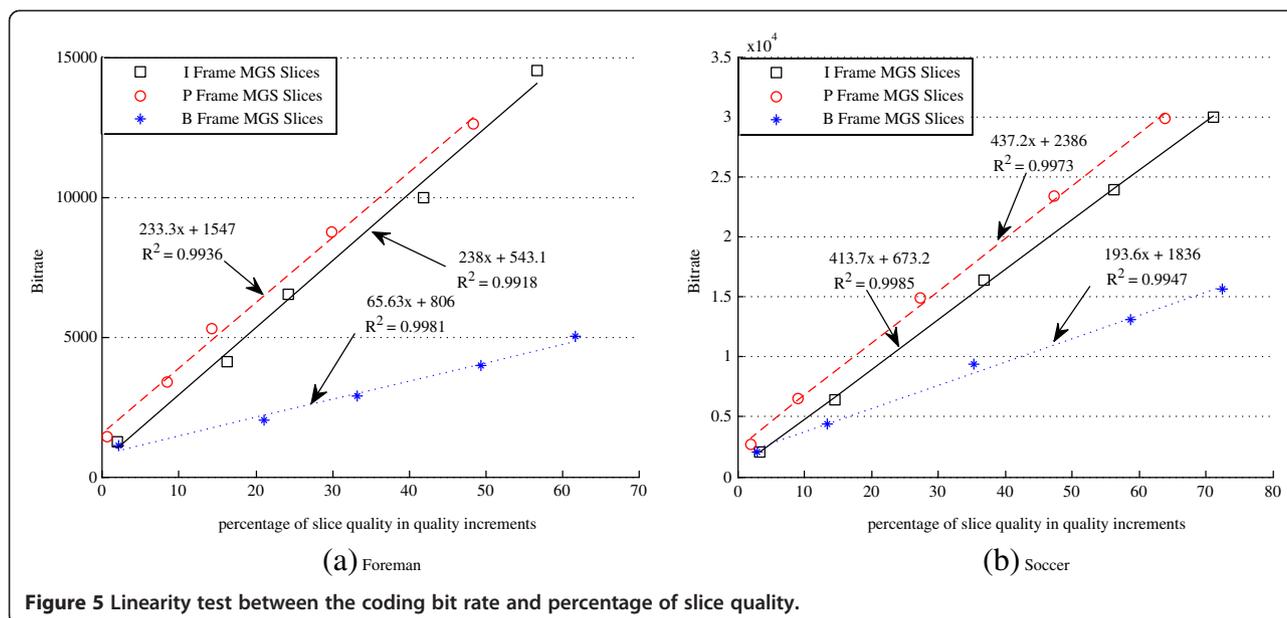

Figure 4 Block diagram of the proposed algorithm.

Figure 5 Linearity test between the coding bit rate and percentage of slice quality.

performance testing. In the first experiment, we analyze the use of different coding bit-rate weights while using the same number of MGS fragments. The second experiment tests the proposed mapping relationship between the sub-stream bit rate and the actual coding bit-rate for bit rate estimation. Finally, in the third experiment, we compare our approach with results using standard compression schemes to demonstrate the suitability of this new method.

5.1. Coding bit rate comparison with same number of MGS fragments

The first experiment is designed to evaluate the effect of the coding bit rate with the same n (for $n = 3/5/7$), but with different MGS weights. For this article, we assume that when we adopt the same number of MGS fragments n, the coding bit rate is similar regardless of

the distribution of the transformation coefficients in the different fragments. Because we use the same coding mechanism and number of MGS fragments n, bit-streams with different distribution of transformation coefficients in the fragments obtain the same number of NALU and coding scalability. Bit streams that have same number of NALUs have similar bit rates for their NAL headers, and the bit rate of the payload data should be similar because of the same coding mechanism and quantization parameters. Experimental results and Equation (7) show that the relationship between the MGS fragment coding bit rate and the percentage of fragment quality in quality increments is approximately linear, as shown in Figure 5. Therefore, as long as we adopt the same n, the coding bit rate should be similar for all distribution of transformation coefficients in the fragments. Table 1 shows the results of this experiment, and validates this theoretical analysis.

Table 1 Coding bit-rate state of different coding schemes when adopting the same n (QPb=45, QPe=35, coding frames=48, frame rate=30, GOPsize=8)

Kbps	[3 8 5]	[6 3 7]	[5 3 1 4 3]	[2 5 6 1 2]	[1 3 2 4 1 2 3]	[4 1 2 1 1 3 4]
Foreman	561.6	571.3	623.2	625	685.9	677.9
Soccer	692.4	702	750.2	755.5	847.2	828.4
Bus	1115.3	1121.4	1176.6	1182	1299.9	1294.9
Coastguard	903.8	900.2	942.7	987.4	1054.6	1030.9
Container	494.9	499.2	549.4	550.4	616.9	614.8
Stefan	1169	1172.9	1222.3	1232.3	1361.1	1363.4
Highway	316.6	324.1	375.6	373.2	429.6	425.3
Mobile	1491.4	1492.5	1566.5	1557.1	1699	1712
Bridge	534.8	540.8	592	594.8	669.4	653.9

5.2. Coding bit-rate estimation with different number of MGS fragments

Bit rate estimation is a significant step in n optimization. In this experiment, we test the mapping relationship between the sub-stream bit rate and the actual coding rate with different n. The experiment uses the "Mobile" and "Bus" video sequences as the test sets and extracts the quality sub-streams. Figure 6 shows the resulting coding bit-rate curve and percentage of fragment quality in quality increments as a function of n. The results show that the proportional relationship between the percentage of fragment quality in quality increments and the sub-stream bit rate is similar to the corresponding relationship between the number of MGS fragments and the actual coding bit rate.

Our analysis shows this linear relationship holds true for a number of different correlations. First, experiment in Section 5.1 and a subsequent correlation analysis demonstrate that there is an approximately linear mapping relationship between MGS fragment coding bit rate and percentage of fragment quality in quality increments. Thus, payload data interception due to sub-stream extraction results in a mainly linear descending bit rate. Second, the sub-stream and the actual coding have the same NALUs and similar bit rate sizes for their respective NAL headers. Finally, as the percentage of fragment quality in quality increments increases, the error drift that results due to inter-frame prediction will also increase in a linear fashion.

From Figure 6b,d we can see the proportional relationship between the sub-stream bit rate and the percentage of fragment quality in quality increments. Figure 6a,c shows a comparison between the calculated coding bit rate after transformation using (7) and the actual coding bit rate. We found that there is less than 5% deviation between the actual and calculated coding bit rates.

5.3. Q_e and n Determination

In this experiment, we analyze coding performance when using the optimized coding configuration determined
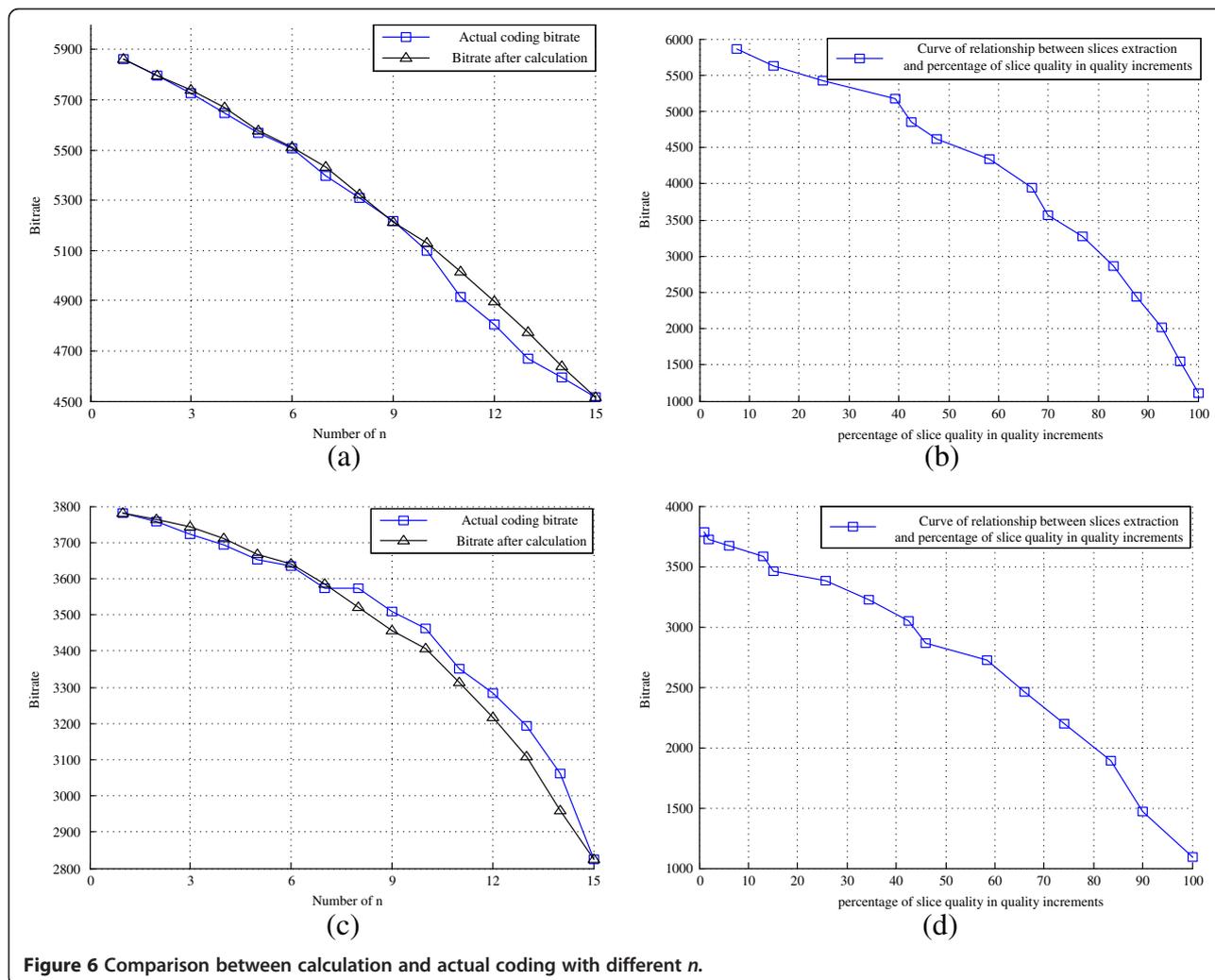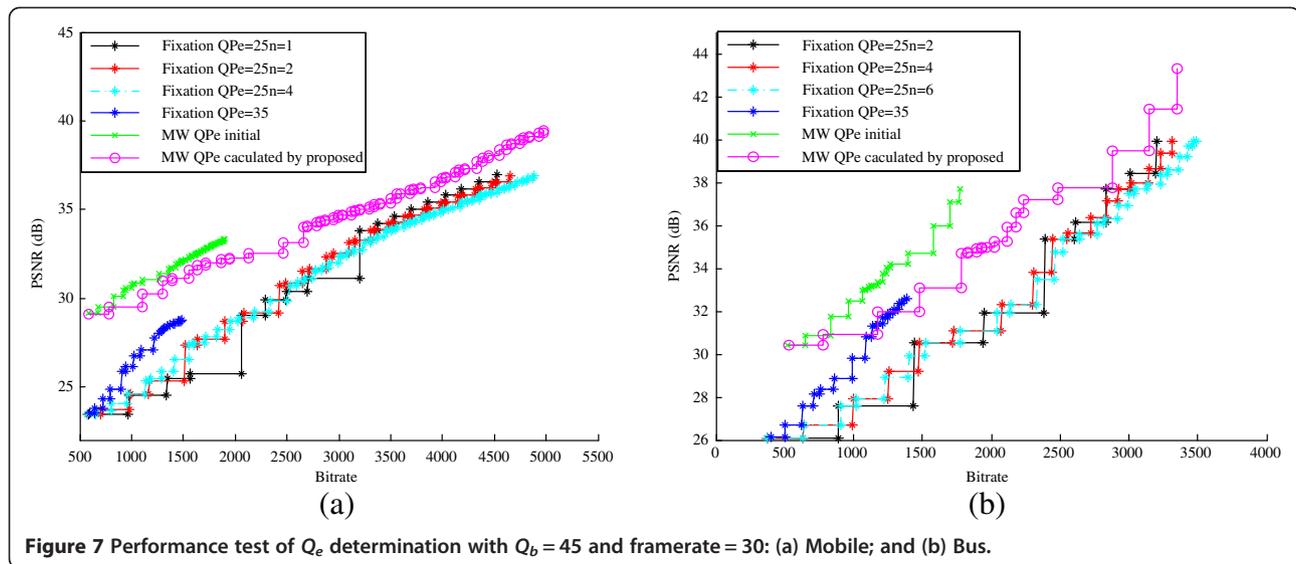

Figure 6 Comparison between calculation and actual coding with different *n*.

Figure 7 Performance test of Q_e determination with $Q_b = 45$ and framerate = 30: (a) Mobile; and (b) Bus.

through our method, and compare these results against the coding performance when using other coding configurations. This experiment continues to use the "Mobile" and "Bus" (CIF format) video sets as the test objects. Coding configuration optimization is mainly reflected in QP_e and n determination for a given users' bandwidth distribution.

Sub-stream performance for different QP_e was calculated using the R-D optimized extraction method [23] and is illustrated in Figure 7a,b. When QP_e is fixed, We see that $QP_e = 35$ (blue line) can be calculated by 3.1 and, when user bandwidth is 500–1500 kbps, it is more suitable than other selections such as $QP_e = 25$ (cyan, red, and black lines). In fact, its performance can only be improved up to 3 dB. However, when user bandwidth is 600–5300 kbps (Figure 7a) or 500–3600 kbps (Figure 7b), users can not achieve better performance when $QP_e = 35$ if their bandwidth capacity >= 1500 kbps due to coding restrictions. On the other hand, selecting a value of $QP_e = 25$ allows different performance and scalability results based on the n. This result is shown in Figure 7a, with the

black, red, and cyan lines corresponding to $n = 1$, $n = 2$, and $n = 4$, respectively, and in Figure 7b, where the black, red, and cyan lines correspond to $n = 2$, $n = 4$, and $n = 6$, respectively. Mathias Wien [2] found similar results by adjusting rate allocation in frames to improve coding efficiency rather than fixing QP_e as was done for this work. Note that if we initially set $QP_e = 30$ (green line), analysis shows that a $QP_e = 22$ (pink line) provides better quality when user bandwidth is in the range of 600–5300 kbps (Figure 7a) or 500–3600 kbps (Figure 7b). Therefore, it is important to select a QP_e appropriate to the users' bandwidth probability density distribution to improve overall video quality for all users.

After determining a suitable QP_e, we next looked at n selection under different user bandwidth probability density distributions by using the algorithm described in Sections 3.2 and 4 to optimize n, then comparing the coding performance results to actual coding performance produced using other configurations. For this experiment, we conducted six tests using three different bandwidth probability density distributions with multiple users.

Table 2 Performance of different n A (normal distribution): QP_e=25; mobile mean=600 with standard deviation=1000; bus mean=500 with standard deviation=650

Sequence	n Num	Calculation (dB)	Actual (dB)	Sequence	n Num	Calculation (dB)	Actual (dB)
Mobile Sequence (600–5300 Kb)	1	25.86	25.34	Bus Sequence (500–3600 Kbps)	1	24.83	26.45
	2	26.18	25.98		2	25.99	27.12
	3	**26.24**	26.04		3	26.69	27.28
	4	26.228	**26.11**		4	26.73	27.39
	5	26.18	26.05		5	**27.43**	27.41
	6	26.11	26.06		6	27.42	**27.444**
	7	26.01	26.06		7	27.42	27.457
	8	25.98	26.03		8	27.4	27.448

Table 3 Performance of different *n* B (uniform distribution)

Sequence	n Num	Calculation (dB)	Actual (dB)	Sequence	n Num	Calculation (dB)	Actual (dB)
Mobile Sequence (600–5300 Kb)	1	31.32	31.06	Bus Sequence (500–3600 Kbps)	1	32.26	31.65
	2	**31.43**	**31.38**		2	32.47	32.29
	3	31.39	31.28		3	**32.49**	32.27
	4	31.25	31.06		4	32.31	**32.4**
	5	31.12	31.07		5	32.46	32.17
	6	30.9	30.88		6	32.22	32.18
	7	30.74	30.78		7	32.14	32.16
	8	30.63	30.63		8	32	32.03

Take Mobile sequence coding for an example to analyze the process of our algorithm. In pre-encoding, we first determine a suitable $QP_e = 25$ by (2)–(6) when users' bandwidth is limited in 600–5300 kbps (Figure 7a). By setting frame rate = 15, GOP size = 16 and coding schemes of IBBBPBBB, we know that temporal level = 3 in the bit stream. Level of GOP quality scalability can also be calculated by (9) where M is the number of quality increments. In pre-encoding, do sub-stream extraction to get bit-rate information of different temporal level when pre-encoding with $n = 15$. According to the bit-rate calculation with different n in 3.2, build linear mapping relation to calculate the bit rate of different temporal level with $n = M$. With temporal levels of GOP quality scalability (calculated by Equation 10), bit rate of different temporal scalability and (11), we calculate average interval Δ_i between two scalable points of temporal levels. According to Δ_i, bit rate D of every scalable point can be calculated. Take curve relation of *PSNR–Bitrate* when pre-encoding with $n = 15$ as reference, we estimate $Q_M(s)$ with $n = M$ when the user's bandwidth <=s according to the linear mapping relation. Combined with the situation that $f(x)$ which denotes as probability density distribution functions of users bandwidth subject to normal distribution (mean = 600, standard deviation = 1000), we can calculate max(.) with (13), which means overall video quality for all users with

$n = M$. Find the maximum value when $1 \leq M \leq 15$ and take M as optimum value n.

As shown in Tables 2, 3, and 4, we select the maximum value to determine the number of MGS fragments n. Our approach accurately optimizes coding configuration depending on the transmission rate and user environment, and the different user bandwidth probability density distributions for the two different videos. For the "Mobile" video, our method gives values of $n = 3, 2, 1$ as optimal, while actual best results occur when $n = 4, 2, 1$ in this test. For the "Bus" video, the proposed optimized n values are $n = 5, 3, 1$, while actual performance is best when $n = 6, 4, 1$. However, it is worth noting that there is little PSNR difference between the actual and calculated values, which suggests that our optimization method does provide sufficient quality improvement to users.

6. Conclusion and discussion

In this article, we proposed a parameters' optimization scheme in order to eliminate uncertainties when selecting coding parameters. Compare with existing methods, our approach builds a relationship map of the quantization parameter QP_e, user bandwidth range, and an optimum number of fragments n based on the bandwidth probability density distribution of the users. The

Table 4 Performance of different *n* C (normal distribution): $QP_e = 25$; mobile mean = 5300 with standard deviation = 1000; bus mean = 3600 with standard deviation = 650

Sequence	n Num	Calculation (dB)	Actual (dB)	Sequence	n Num	Calculation (dB)	Actual (dB)
Mobile Sequence (600–5300 Kb)	1	**36.01**	**36.21**	Bus Sequence (500–3600 Kbps)	1	**39.13**	**38.33**
	2	35.96	36.11		2	38.75	38.16
	3	35.87	35.95		3	38.42	37.92
	4	35.67	35.94		4	38.19	38.02
	5	35.49	35.55		5	38	37.57
	6	35.15	35.15		6	37.62	37.5
	7	34.92	34.96		7	37.42	37.44
	8	34.74	34.63		8	37.19	37.21

algorithm first calculates the bit rate of the coding bit streams according to I/P/B frame R-D characteristics. Then, based on the result of this calculation, a coding bit rate estimation method when coding with different n is proposed for analyses the enhancement layer bit rates. Finally, an optimized coding configuration is calculated and combined with the user bandwidth probability density distribution and a bit-stream fidelity scalability analysis. The experimental results show that our approach can significantly improve average video quality for users.

Our method differs from existing approaches in three ways:

(1) it constructs a mapping relationship between the bandwidth range of the users and quantization parameter QP_e by presenting a coding rate calculation module that correlates coding rate and percentage of fragment quality in quality increments;

(2) The scheme takes overall quality for the users as a metric standard, and translates the problem of coding optimization into a problem of overall user quality optimization and effectively optimizing the number of MGS fragments;

(3) Provides a coding configuration optimization approach suitable for different bandwidth ranges and user bandwidth probability density distributions.

In future work, we will extend this approach to other coding parameter configurations that take transmission effects and user environments into consideration. Moreover, we will also explore ways to reduce computational complexity.

Competing interest

The authors declare that they have no competing financial interests.

Author details

Xu Chen: received the B.S. and M.S. degree from Guilin University of Electronic Technology, Guilin, China, in 2006 and 2009, respectively and received the PH.D. degree in signal and information processing from Shenzhen University, Shenzhen, China, in 2012. He is currently working in the Research Department of Hisilicon Semiconductor and Component Business Department, Huawei Technologies, focused on video coding.
Ji-hong Zhang: received the B.S., M.S. and PH.D. degree in information and communication engineering from Southeast University, Nanjing, China, in 1986, 1989 and 1992, respectively. He has been a professor at Shenzhen Institute of Information Technology and Shenzhen University, Shenzhen, China. His current research is focused on intelligent information processing.
Wei Liu: received the B.S., M.S. and PH.D. degree in system engineering from National University of Defense Technology, Changsha, China, in 1996, 1999 and 2005 respectively. He has been a senior engineer at Shenzhen Institute of Information Technology, Shenzhen, China. His research interests include image processing, video encoding and decoding, multimedia database
Yong-sheng Liang:received the B.S. degree from Shanxi University of Science and Technology, Xi'an, China, and the PH.D. Degree from Harbin Institute of Technology, Harbin, China, in 1993 and 1999 respectively. He completed the post doctoral research in information and communication engineering Station of Harbin Institute of Technology, Harbin, China, in 2005. He has

been a professor at Shenzhen Institute of Information Technology, Shenzhen, China. He is now a professor at the School of Information Engineering of Shenzhen University, Shenzhen, China. He is an IEEE ComSoc member. His research interests include computer network and data communication, signal processing and pattern recognition.
Ji-qiang Feng: received the M.S. degree of applied mathematics and the PH. D. degree from Shenzhen University, Shenzhen, China, in 2008 and 2011, respectively. His research is on computational intelligence both the theoretical and methodological aspects, with a special application in image processing.

Author details

[1]School of Information Engineering, Shenzhen University, Shenzhen, Guangdong, China. [2]Shenzhen Key Lab of Visual Media Processing and Transmission, Shenzhen Institute of Information Technology, Shenzhen, Guangdong 518029, China.

References

1. T Schierl, T Stockhammer, T Wiegand, Mobile video transmission using scalable video coding. IEEE Trans. Circuits Syst. Video Technol. **17**(9), 1204–1217 (2007)
2. *Joint Scalable Video Model, Joint Video Team(JVT) of ISO/IEC MPEG & ITU-T VCEG, Doc. JVT-X202, Joint Scalable Video Model, Joint Video Team(JVT) of ISO/IEC MPEG & ITU-T VCEG, Doc. JVT-X202* (2007). July
3. Z Chen, C Lin, X Wei, Enabling on-demand internet video streaming services to multi-terminal users in large scale. IEEE Trans. Consum. Electron. **55**(4), 1988–1996 (2009)
4. J Ridge, X Wang, I Amonou, N Cammas, *Simplification and unification of FGS, in Joint Video Team of ISO/IEC MPEG and ITU-T VCEG, JVT-S077* (April, Geneva, 2006)
5. I Amonou, N Cammas, Complexity reduction of FGS passes, in *Joint Video Team of ISO/IEC MPEG and ITU-T VCEG, JVT-R069* (, 2006). January
6. T Wiegand, G Sullivan, J Reichel, H Schwarz, M Wien, Study text (version 3) of ISO/IEC 14496–10: 2005/FPDAM3 scalable video coding, in *Joint Video Team of ISO/IEC and ITU-T VCEG, N8962* (2007). April
7. H Schwarz, D Marpe, T Wiegand, Overview of the scalable video coding extension of the H.264/AVC standard. IEEE Trans. Circuits Syst. Video Technol. **17**(9), 1103–1120 (2007)
8. H Schwarz, D Marpe, T Wiegand, Overview of the scalable extension of the H.264/MPEG-4 AVC video coding standard. IEEE Trans. Circuits Syst. Video Technol. **17**(9), 1174–1185 (2007)
9. M Wien, H Schwarz, T Oelbaum, Performance analysis of SVC. IEEE Trans. Circuits Syst. Video Technol. **17**(9), 1194–1203 (2007)
10. R Gupta, A Pulipaka, P Seeling et al., H.264 coarse grain scalable (CGS) and medium grain scalable (MGS) encoded video: a trace based traffic and quality evaluation. IEEE Trans. Broadcast. **58**(3), 428–439 (2012)
11. G Burak, V Sadi, AM Tekalp, Effects of MGS fragmentation, slice mode and extraction strategies on the performance of SVC with medium-grained scalability, in *Proceeding of International Conference on Image Processing*, 2010, pp. 4201–4204. September
12. C Michele, F Lorenzo, F Marco, Quality modeling for the medium grain scalability option of H.264/SVC, in *Proceeding of the 5th International ICST Mobile Multimedia Communications Conference*, 2009, pp. 1–6. September
13. X Long, W Gao, X Ji, A rate control algorithm for scalable video coding. Chin. J. Comput. **31**(7), 1175–1184 (2008)
14. W Dalei, C Song, Application-centric routing for video streaming over multihop wireless networks. IEEE Trans. Circuits Syst. Video Technol. **20**(12), 1721–1734 (2010)
15. C Haechul, K Jung Won, Dynamic and interoperable adaptation of SVC for Qos-enabled streaming. IEEE Trans. Consum. Electron. **53**(2), 384–389 (2007)
16. ISO/IEC, ISO/IEC TR 21000–7:2004, *Information technology–Multimedia framework (MPEG-21)-Part 7: Digital Item Adaptation* (ITU-T, Geneva, 2004)
17. China Internet Network Information Center, *Statistical Report on Internet Development in China* (CNNIC, China, 2011)

18. R Pingdom, *Statistical Report on Internet Connection Speeds Worldwide* (Akamai Technologies, Inc, USA, 2010)
19. Y Liu, GG Zheng, S Yeng Chai, A novel rate control scheme for low delay video communication of H.264/AVC standard. IEEE Trans. Circuits Syst. Video Technol. **17**(1), 68–78 (2007)
20. EY Lam, JW Goodman, A mathematical analysis of the DCT coefficient distributions for images. IEEE Trans. Image Process. **9**(10), 1661–1666 (2000)
21. *"YUV video sequences"*. http://trace.eas.asu.edu/ yuv/
22. J Reichel, H Schwart, M Wien, *Joint Scalable Video Model. (JSVM), Joint Video Team, doc* (JVT-AA202, Shenzhen, China, 2008). January
23. I Amonou, N Cammas, S Kervadec, Optimized rate-distortion extraction with quality layers in the scalable extension of H.264/AVC. IEEE Trans. Circuits Syst. Video Technol. **17**(9), 116–121 (2007)