Research Article

A Simple and High Performing Rate Control Initialization Method for H.264 AVC Coding Based on Motion Vector Map and Spatial Complexity at Low Bitrate

Yalin Wu and Sun-Woo Ko

Department of Smart Information Systems, Jeonju University, 303 Cheonjam-ro, Wansan-gu, Jeonju 560-759, Republic of Korea

Correspondence should be addressed to Sun-Woo Ko; godfriend@hanmail.net

The temporal complexity of video sequences can be characterized by motion vector map which consists of motion vectors of each macroblock (MB). In order to obtain the optimal initial QP (quantization parameter) for the various video sequences which have different spatial and temporal complexities, this paper proposes a simple and high performance initial QP determining method based on motion vector map and temporal complexity to decide an initial QP in given target bit rate. The proposed algorithm produces the reconstructed video sequences with outstanding and stable quality. For any video sequences, the initial QP can be easily determined from matrices by target bit rate and mapped spatial complexity using proposed mapping method. Experimental results show that the proposed algorithm can show more outstanding objective and subjective performance than other conventional determining methods.

1. Introduction

In the last decade, multimedia data has been applied to communication, security, entertainment, and military. Because multimedia data has the problem of large amount of data, it can be hardly stored and transmitted. Video coding can effectively solve the problem. With the development of terminal equipment and communication networks, the video coding standards have been continually established as MPEG-1 [1], MPEG-2 [2], MPEG-4 [3], H.261 [4], H.262, H.263 [5], H.264 [6], and H.265 [7]. Especially, H.264/AVC can be applied for extensive areas such as DVD and VOD over cable and bit streaming video at low bit rate with high quality.

In multimedia communication and transmission, rate control (RC) algorithm plays a crucial role. H.264/AVC includes an RC [8] which can be used to achieve optimal subjective quality given transmission bandwidth limit by regulating the encoding parameters. In RC, the optimal QP has been determined for every frame of video sequence. The large amount of encoded and good perceptual quality can be obtained by reducing the QP. In contrast, the QP increment can reduce the encoded bits and perceptual quality. The mean absolute difference (MAD) of residual MB can be used to determine the QP for MB. However, the chicken-and-egg dilemma [6] occurs in the process of determining QP. In JVT-G012 [9], the scheme of one-pass RC uses a linear MAD prediction model to solve the chicken-and-egg dilemma. According to the efficiency and simplicity of JVT-G012, the scheme of JVT-G012 has been widely applied in H.264 software and hardware. The algorithm of JVT-G012 includes the process of initialization for RC. In this process, the initial QP is determined for IDR picture in a video sequence. Therefore, the optimal MAD of IDR can be calculated using initial QP. The MAD of next frame is predictable using optimal MAD of IDR. The process of initial QP determination is an important part in RC. The value of initial QP can influence the RC performance. Conventionally, the value of initial QP can be determined according to the number of bits per pixel (BPP) which depends on the target bit rate, frame rate, and image size in JVT-G012. Although the scheme of BPP is easy to implement, it is very rough and imprecise. Generally, video sequence is two-element data which includes spatial and temporal components. According to information theory, the relatively large number of bits is used to represent

the video sequence which has highly complicated spatial and temporal features. In contrast, relatively small amount of bits can encode the low complex video sequence. In the situation of the restricted target bits, a large initial QP is expected for highly complicated video sequence. In spite of reducing the image quality, it is in the limited data range. For the low complex video sequence, a small initial QP is assigned; not only the target bits can be fulfilled but the excellent image quality can be also obtained. Using this idea, it is very important that the scheme of RC initialization is used to consider spatial and temporal complexities of video sequence. In order to complete this target, Wang and Kwong [10] and Wu and Kim [11] proposed the schemes which used the characteristics of video sequences as well as BPP to determine the initial QP. However, their methods do not consider quality balance of reconstructed video sequence and the provided parameters cannot apply any video sequence. The criterion of selecting the sample video sequences has not been provided. In addition, the algorithm of determining optimal initial QP has not been explained either. Moreover, Hu et al. [12] proposed a scheme of computing the initial QP for spatial scalable video coding (SVC). However, this scheme is only applicable to the SVC standard.

In order to solve the existing problems, we propose a simple and high performance method to determine an initial QP in given target bit rate. To obtain the initial QP for any video sequences, it is very important to measure the spatial and temporal complexities of the video sequences. In H.264/AVC [6] standard, the motion vector is proposed as a measure of temporal complexities of video sequences. Primarily, temporal complexities of ten video sequences are analyzed using motion vector and 4 video sequences are screened out as the samples. And then, the spatial complexities of samples are calculated using the rate of number of complex MBs, which are determined by variance of MB, in the IDR. In this paper, an algorithm for determining optimal initial QP is proposed for producing the reconstructed video sequences with high and stable qualities at the encoding bits which are very close to target bit rates within the range of 0.4 to 1.0 Mbps. Subsequently, optimal initial QPs of 4 samples constitute two-dimensional matrices basis of spatial complexity and target bit rate. Moreover, we propose a mapping method to determine the spatial complexity of tested video sequences by the two-dimensional matrices. For any tested video sequence, its optimal initial QP can be chosen from the built matrix using its mapped spatial complexity and target bit rate. Section 2 presents the related method of initialization process of RC. Section 3 interprets the evolution process of proposed method. The results of experiment are shown in Section 4. Finally, Section 5 shows the conclusion of this study.

2. Related Method of Initialization Algorithm of Rate Control

In this section, we review some initialization methods of RC which are used to decide initial QPs in the recent literatures. The method of JVT-G012 that can automatically conclude the value of initial QP for the IDR is the most traditional and coarse method. The advantages of JVT-G012 are easy implementation and low computational complexity. In the case of the poor performance hardware of terminal equipment, JVT-G012 is being of extensive usage. Various versions of the reference software of H.264 [7, 13] have adopted JVT-G012. The disadvantage of JVT-G012 is being not accurate enough. The method of JVT-G012 uses only BPP to determine an initial QP as follows:

$$\text{JVTQP} = \begin{cases} 35, & \text{BPP} \geq J_1, \\ 25, & J_1 < \text{BPP} \leq J_2, \\ 20, & J_2 < \text{BPP} \leq J_3, \\ 10, & \text{BPP} > J_3, \end{cases} \tag{1}$$

$$\text{BPP} = \frac{\text{BR}}{\text{FR} \times \text{VS}}, \tag{2}$$

where JVTQP is calculated by the initial QP of initialization process of RC. BPP, which is the bits per pixel, is computed using BR, FR, and VS which represent the target bit rate, the frame rate, and the size of frame in (2). J_1, J_2, and J_3 are thresholds. For QCIF video sequences, the thresholds are given as 0.1, 0.3, and 0.6. For CIF video sequences, the thresholds are recommended to be 0.6, 1.4 and 2.4 in JM9.3 [10].

In the method of Wang, the value of initial QP is computed on the basis of BPP and spatial feature. The spatial feature can be computed using entropy information and the dc mode of INTRA 16×16 of the IDR.

Consider

$$\text{Entropy} = -\sum_{i=0}^{L-1} p(i) \log_2 [p(i)],$$

$$\text{MC} = \frac{\sum_{k=1}^{N} \sum_{t=0}^{15} \sum_{j=0}^{15} \left| M^k(t, j) - M_{\text{dc}}^k \right|}{1000 \times N}, \tag{3}$$

where $L - 1$ is maximum gray level value of pixel, $p(i)$ is the probability of gray level i, N is the number of MBs in IDR, $M^k(t, j)$ is used to denote the pixel value at (i, j) of the kth MB, and M_{dc}^k is used to denote the predicted compensation value computed from the dc mode of INTRA 16×16. Finally, the initial QP can be calculated by (4). Figure 1 shows the relationship of the best initial QP and BPP by News, Foreman, and Mobile. And then (5) can be computed by this relationship:

$$\text{initial QP} = F_1(\text{BPP}) + F_2(\text{Entropy, MC}), \tag{4}$$

$$F_1(\text{BPP}) = \begin{cases} c1 \cdot \text{BPP} + c2, & \text{BPP} \geq 0.4, \\ c3 \cdot \text{BPP} + c4, & 0.4 > \text{BPP} \geq 0.2, \\ c5 \cdot \text{BPP}^2 + c6 \cdot \text{BPP} + c7, & 0.2 \geq \text{BPP}, \end{cases} \tag{5}$$

$$F_2 = \max \left\{ t1\text{Entropy}^2 + t2\text{Entropy} + t3, \right.$$
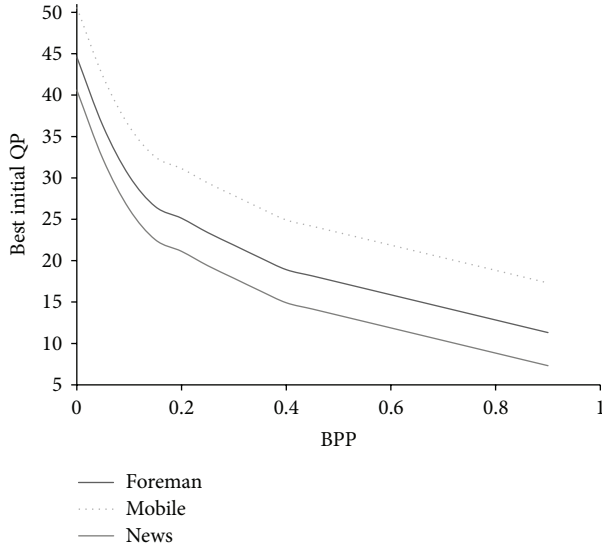$$\left. e1\text{MC}^2 + e2\text{MC} + e3 \right\}, \tag{6}$$

FIGURE 1: Relationship between the best initial QP and BPP according to News, Foreman, and Mobile.

where $\theta = \{c1, c2, c3, c4, c5, c6, c7, t1, t2, t3, e1, e2, e3\}$ can be defined as $\theta = \{-15.21, 25, -30.41, 31, 462.47, -189.86, 44.60, -15.25, 241.81, -944.42, -0.76, 13.59, -49.18\}$ for CIF video sequences. At QCIF video sequences, $\theta = \{-12.67, 24, -25.34, 29, 546.59, -221.55, 46.41, 14.25, -184.76, 595.95, 0.67, -7.23, 18.50\}$. The value of initial QP is computed using (4).

In the scheme of Wu, initial QP is calculated on the basis of BPP, the MAD value of the IDR, and the average MAD value of the 2nd, 3rd, and 4th frames which are encoded using intermode as follows:

$$\text{initial QP} = F_1\left(\text{BPP}\right) + F_3\left(\text{MAD}_I, \text{MAD}_{AP}\right), \quad (7)$$

$$
\begin{aligned}
F_3\left(\text{MAD}_I, \text{MAD}_{AP}\right) = p1 \cdot \left[a \cdot \text{MAD}_I \right. \\
\left. + (1-a) \cdot \text{MAD}_{AP}\right]^2 \\
+ p2 \cdot \left[a \cdot \text{MAD}_I \right. \\
\left. + (1-a) \cdot \text{MAD}_{AP}\right] + p3,
\end{aligned}
\quad (8)
$$

where the value of a is 0.05 and $\theta = \{c1, c2, c3, c4, c5, c6, c7, p1, p2, p3\}$ can be determined as $\theta = \{-15.21, 25, -30.41, 31, 462.47, -189.86, 44.60, -0.78, 8.48, -10.90\}$ for CIF video sequences and $\theta = \{-12.67, 24, -25.34, 29, 546.59, -221.55, 46.41, 0.36, 1.30, -2.78\}$ for the size of QCIF. The value of initial QP is computed using (7).

The parameters used in the method of Wang and Wu are calculated using three types of tested video sequences. In Wu and Wang, the extracting method of sample video sequences has not been shown. It is difficult to say that they can represent the various spatial and temporal complexities video sequences. Moreover, Wang and Wu have not taken into account the quality consistency of recovered video sequences. Furthermore, the scheme of the best initial QP determination is not explained.
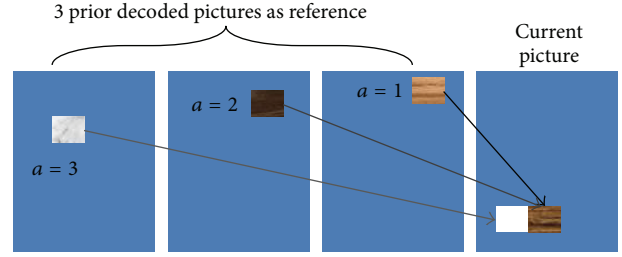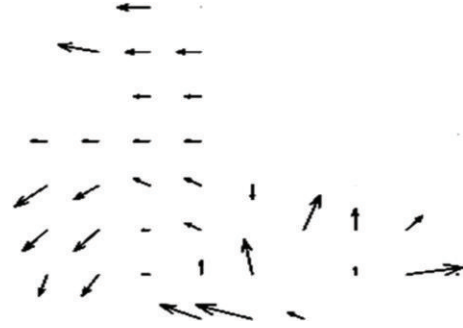

FIGURE 2: The process of ME.

FIGURE 3: Motion vector map.

3. Proposed Initialization Algorithm of Rate Control

3.1. Spatial and Temporal Complexities. In the video sequence, the content of adjacent frames has not any significant difference. In order to save amount of bits, only difference is encoded. For finding difference, most video coding standards support the method named motion estimation (ME) [6]. ME is used to investigate the 16×16 at the objective region in reference frames that closely matches the current MB. ME is improved by H.264/AVC such as variable block size, multiple reference frames, and optimization algorithm. Figure 2 shows the process of ME. Motion vectors are used to compress video by storing the changes to an image from one frame to the next. Figure 3 shows motion vector.

By the number and magnitude of the motion vectors, the temporal complexity of video sequences can be measured and predicted. In H.264/AVC, most frames are encoded using intermode [4–6]. For the video coding performance, the temporal complexity is more important than spatial complexity.

Figure 4 shows the result of motion vector maps for ten different types of sample video sequences [14]. In Figure 4, the ten video sequences can be simply divided into two categories. The Bus, Flower, and Mobile video sequences are classified as complex case. On the contrary, Foreman, Waterfall, Silent, Paris, Bridge-far, Mother-daughter, and News video sequences are relatively simple case. Complexity means that the number of motion vectors is high and vice versa. In complex videos, Mobile video is captured by fixed camera and its motion vectors exist in each MB. Flower video shows the movement of the objects located at the lower part of the video while Bus video has the movement on the all-region.

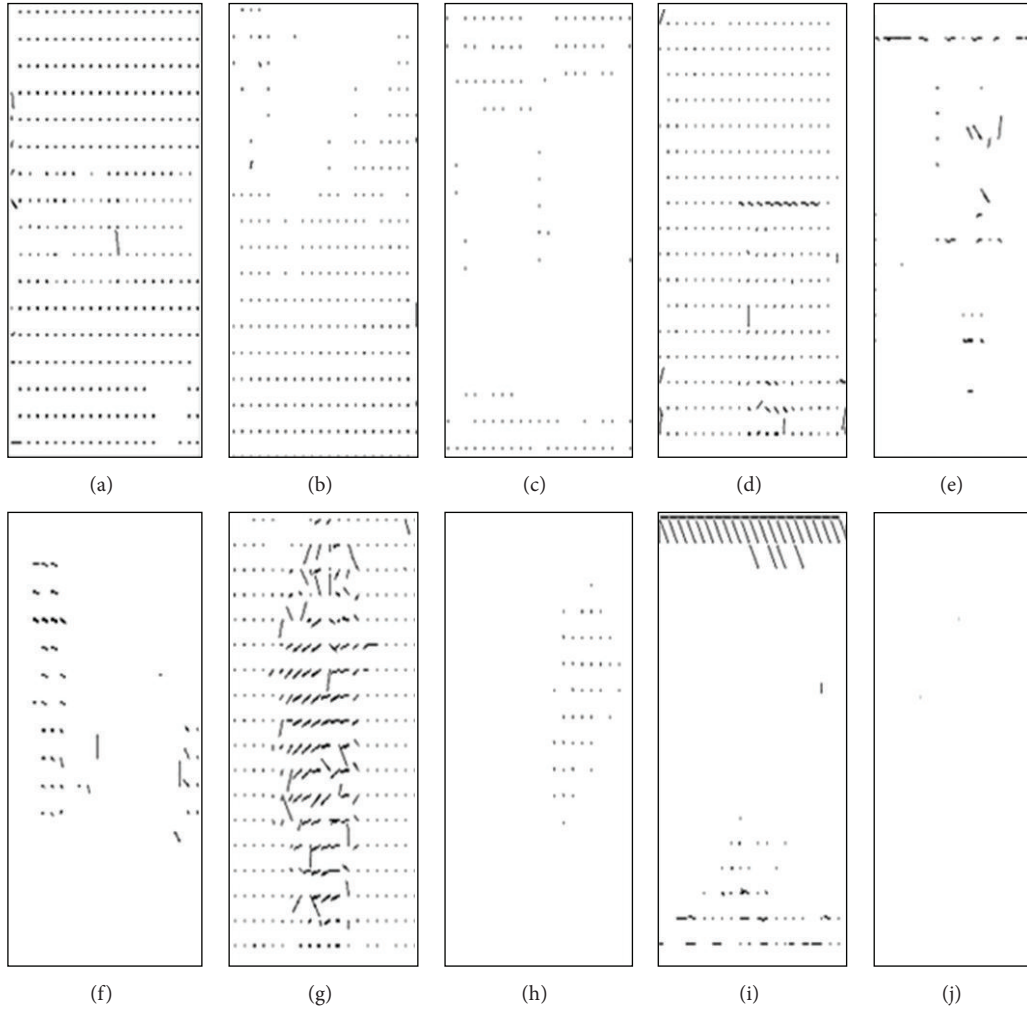| (a) | (b) | (c) | (d) | (e) |

| (f) | (g) | (h) | (i) | (j) |

FIGURE 4: Motion vectors obtained from ten video sequences.

In simple videos, Waterfall is a typical case which applies the zoom-out photography technology and Foreman, which is classic interview type, has motion vectors in the middle of the image. Moreover, News, Paris, Mother-daughter, and Silent are also interview style video sequences. According to their motion vectors, various types of video sequences have different patterns. The patterns can be used to distinct typical video sequences. We use 4 video sequences which are Mobile, Flower, Waterfall, and Foreman as the sample video sequence to analyse any video sequences because they have distinctive features.

In proposed initialization algorithm of RC, the value of initial QP is computed based on the spatial and temporal complexities at the given target bit rate. As the sample video sequences are selected, the calculating method of spatial complexity of sample video sequences must be provided.

In H. 264, the smallest encoding unit is MB. MB includes two INTRA prediction modes which are INTRA 16×16 [4] and INTRA 4×4 [4]. Figure 5 shows the INTRA prediction modes. In Figure 5(a), the INTRA 4×4 has 9 prediction modes. In Figure 5(b), the INTRA 16×16 has 4 prediction modes.

Generally, the INTRA 16×16 mode is used in the MBs which are in the homogeneous regions of image. However, the INTRA 4×4 mode is used in the MBs which are in the object and edge parts of image [4]. In Figure 6, the nonhomogeneous MBs have the same value of pixels. The INTRA 16×16 mode is used to process the homogeneous MBs. The reasons to use INTRA 16×16 blocks are to save the computing time and to maintain the image quality. The INTRA 4×4 mode is used to process the non-homogeneous MBs which have the different pixel values. The reason to use INTRA 4×4 blocks is to maintain the quality.

Since the variance of an MB, which includes 256 pixels, is the equal of the total information of the DC and AC coefficients of the MB, the spatial complexity of the MB can be estimated using variance [15].

Consider

$$\mathrm{MB_{var}} = \frac{1}{256} \left(\sum_{i=0}^{15}\sum_{j=0}^{15}[Y(i,j)]^2 - \left[\sum_{i=0}^{15}\sum_{j=0}^{15}Y(i,j) \right]^2 \right), \quad (9)$$

where the variance of MB is computed by $\mathrm{MB_{var}}$ and $Y(i,j)$ is value of the pixel which is luminance signal at (i, j). An MB
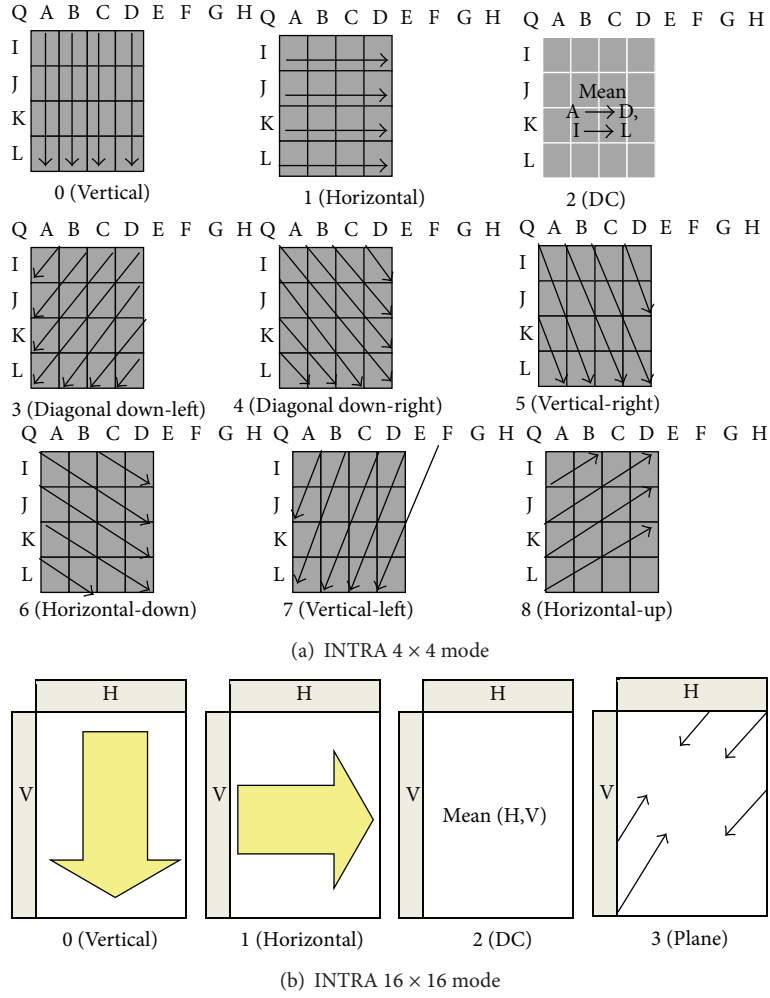
(a) INTRA 4×4 mode

(b) INTRA 16×16 mode

FIGURE 5: INTRA prediction modes.

(a) Homogeneous MB
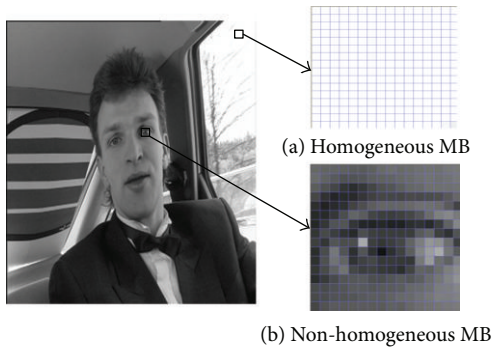
(b) Non-homogeneous MB

FIGURE 6: Type of MBs.

is classified to be of high or low complexity by threshold of variance as follows:

$$\text{MB}_{\text{Complexity}} = \begin{cases} \text{high}, & \text{MB}_{\text{var}} > T, \\ \text{low}, & \text{MB}_{\text{var}} \leq T, \end{cases} \tag{10}$$

where T is the threshold value defined as 92735 [15]. Figure 7 shows the classification process of MB according to the value of variance. The spatial complexity of the IDR is measured according to the proportion of the number of complex MBs as follows:

$$\text{Frame}_{\text{complex}} = \frac{\text{MB}_{\text{Complex}}}{\text{MB}_{\text{Frame}}} \times 100 \, (\%), \tag{11}$$

where the $\text{Frame}_{\text{complex}}$ is the rate of the number of complex MBs in the IDR. $\text{MB}_{\text{Complex}}$ is the number of the complex MBs of IDR and MB_{Frame} is the total number of the MBs of IDR.

The $\text{Frame}_{\text{complex}}$ can be a measure of the spatial complexity.

3.2. Proposed Algorithm for Determining the Optimal Initial QP.
According to the given target bit rate, the value of initial QP is directly related to the performance of encoding at the H.264/AVC. The performance of encoding can be evaluated using the quality of reconstructed video sequences and total bits. In other words, the objectives are to satisfy target bit rate to ensure the best quality of reconstructed video sequences.
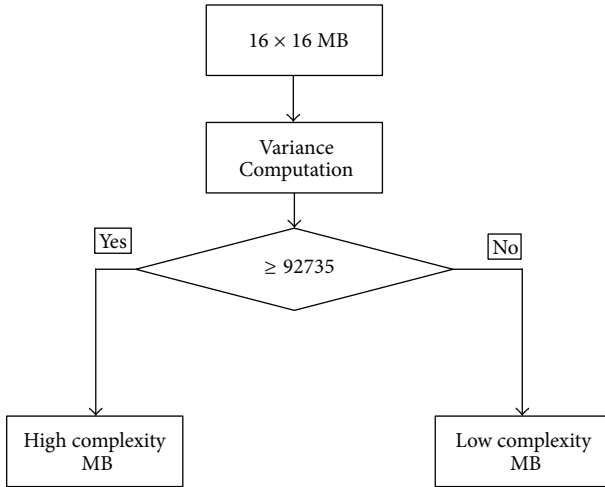
Figure 7: Classification process of MB.

And also the stability of reconstructed video sequences is a very important quality measure in multimedia broadcasting and transmission. Thus the optimal initial QP algorithm has the following properties: (1) maximizing PSNR(peak signal-to-noise ratio) that means the best quality of reconstructed video sequences, (2) maximizing stability that is defined as the differences of QPs, and (3) minimizing total real bits under satisfying the target bit rate.

To find out the optimal initial QP, all potential 52 initial QPs are calculated. For each initial QP, the results which include PSNR, real bits of prior 60 frames, and the differences of QPs in a group of pictures (GOP) are recorded at the given target bit rate. So an initial QP, which produces the maximum PSNR and the minimum of real bits under satisfying target bit rate and minimum differences of QPs, is the optimal initial QP. Since the numeric value of real bits is much larger than PSNR and differences of QPs and also PSNR, real bits, and differences of QPs have different dimensions, PSNR, real bits, and differences of QPs are normalized as follows:

$$
\begin{aligned}
\overline{\text{PSNR}_{\text{initial QP}}} = {} & \left(\text{PSNR}_{\text{Initial QP}} - \text{MIN}_{\text{Initial QP}=0,\ldots,51}\left(\text{PSNR}_{\text{Initial QP}}\right)\right) \\
& \times \left(\text{MAX}_{\text{Initial QP}=0,\ldots,51}\left(\text{PSNR}_{\text{Initial QP}}\right)\right. \\
& \left. - \text{MIN}_{\text{Initial QP}=0,\ldots,51}\left(\text{PSNR}_{\text{Initial QP}}\right)\right)^{-1}, \\
& \quad\quad\quad\quad\quad\quad \text{Initial QP} = 0,\ldots,51,
\end{aligned}
\tag{12}
$$

$$
\begin{aligned}
\overline{\text{BIT}_{\text{Initial QP}}} = {} & \left(\text{BIT}_{\text{Initial QP}} - \text{MIN}_{\text{Initial QP}=0,\ldots,51}\left(\text{BIT}_{\text{Initial QP}}\right)\right) \\
& \times \left(\text{MAX}_{\text{Initial QP}=0,\ldots,51}\left(\text{BIT}_{\text{Initial QP}}\right)\right. \\
& \left. - \text{MIN}_{\text{Initial QP}=0,\ldots,51}\left(\text{BIT}_{\text{Initial QP}}\right)\right)^{-1}, \\
& \quad\quad\quad\quad\quad\quad \text{Initial QP} = 0,\ldots,51,
\end{aligned}
\tag{13}
$$

$$
\begin{aligned}
\overline{\text{DQP}_{\text{Initial QP}}} = {} & \left(\text{DQP}_{\text{Initial QP}} - \text{MIN}_{\text{Initial QP}=0,\ldots,51}\left(\text{DQP}_{\text{Initial QP}}\right)\right) \\
& \times \left(\text{MAX}_{\text{Initial QP}=0,\ldots,51}\left(\text{DQP}_{\text{Initial QP}}\right)\right. \\
& \left. - \text{MIN}_{\text{Initial QP}=0,\ldots,51}\left(\text{DQP}_{\text{Initial QP}}\right)\right)^{-1}, \\
& \quad\quad\quad\quad\quad\quad \text{Initial QP} = 0,\ldots,51,
\end{aligned}
\tag{14}
$$

where $\overline{\text{PSNR}_{\text{Initial QP}}}$, $\overline{\text{BIT}_{\text{Initial QP}}}$, and $\overline{\text{DQP}_{\text{Initial QP}}}$ represent the values of normalized PSNR, real bits, and differences of QPs, respectively. $\text{PSNR}_{\text{Initial QP}}$, $\text{BIT}_{\text{Initial QP}}$, and $\text{DQP}_{\text{Initial QP}}$ are actual unnormalized values. The optimal initial QP is calculated as follow:

$$
\begin{aligned}
\text{Optimal Initial QP} = {} & \\
\underset{\text{Initial QP}=0,\ldots,51}{\text{argmin}} & \left(\alpha\left(1 - \overline{\text{PSNR}_{\text{Initial QP}}}\right)\right. \\
& \left. + \beta\overline{\text{BIT}_{\text{Initial QP}}} + \gamma\overline{\text{DQP}_{\text{Initial QP}}}\right),
\end{aligned}
\tag{15}
$$

where Optimal Initial QP is the optimal initial QP, $0 \le \alpha, \beta, \gamma \le 1$, and $\alpha + \beta + \gamma = 1$. In our research, the values of α, β, γ are set to $1/3$.

3.3. Mapping Method of Spatial Complexity for Generalization.
For the video sequences which have the similar complexity, the initial QPs should be similar. A mapping method of spatial complexity is proposed for any tested video sequence. The spatial complexities of selected sample video sequences can be calculated using (11), according to the given target bit rate. For any video sequence, its spatial complexity is computed using (11) too. The spatial complexity of a video sequence can be mapped by selecting the nearest matching spatial complexity from a set of sample proportion spaces as follows:

$$
\text{Difference}_i = \left|\text{SSC}_i - \text{TSC}\right|, \quad \text{for } i = 1, 2, \ldots, 4,
\tag{16}
$$

where Difference_i is the absolute value of difference between the spatial complexity of ith sample video and a tested one. SSC_i is the ith spatial complexity among 4 sample video sequences. TSC is the spatial complexity of the tested video sequence.

Let *Mapping Sample* denote the sample video that has minimum Difference_i, $i = 1, \ldots, 4$.

In Table 1, MSC is the spatial complexity of sample video. Type 1, Type 2, Type 3, and Type 4 are 4 sample video sequences, which are Waterfall, Foreman, Flower, and Mobile, respectively. Table 1 shows that Mobile has the highest spatial complexity and Waterfall has the lowest spatial complexity, which is in agreement with Figure 4. In addition, Table 1 includes 4 groups of the optimal initial QPs which are 4 affiliated sample video sequences at low bit rate, respectively. The initial QP of a test video sequence can be determined by only selecting an element from Table 1 based on *Mapping Sample* and given target bit rate.

TABLE 1: Lookup table for proposed initial QP.

Bit rate	(MSC) Video			
	(32%) Type 1	(44%) Type 2	(53%) Type 3	(90%) Type 4
0.4 (Mbps)	30	38	42	45
0.5 (Mbps)	33	32	39	42
0.6 (Mbps)	30	30	36	40
0.7 (Mbps)	28	27	35	36
0.8 (Mbps)	28	25	37	37
0.9 (Mbps)	25	24	34	33
1.0 (Mbps)	25	23	35	32

4. Experimental Results

The proposed algorithm and existing methods, which are JVT-G012 and Wu, are implemented on JM9.3 [13] which is the reference software for H.264. In addition, the experiment uses 7 various CIF (352 × 288) standard video sequences that are Waterfall, Foreman, Flower, Mobile, Bus, City, and Stefen, respectively. In the standard video sequences, Waterfall, Foreman, Flower, and Mobile are selected samples, and Bus, City, and Stefen are used to tested generalized quality of proposed algorithm. According to the proposed scheme which is the spatial complexity calculation and mapping, Mobile is the *Mapping Sample* of Stefen and Bus and City are correspondent to Flower and Foreman, respectively. The experimental conditions are as follows.

(a) The system platform is Intel (R) Core(TM)2 Duo CPU E7400 2.80 GHZ, 2.00 GB RAM, and the OS is Microsoft Windows XP professional 2002 Service Pack 3.

(b) JM 9.3 is implemented at the Visual Studio 6.0.

(c) The profile baselines are used; one GOP has 15 frames which includes that the 1st frame is encoded by intra and others are encoded as interframes; the B-picture is not adopted. The item of "Rate Control Enable" is enabled, the item of "Initial QP" is set to 0, and the target bit rates are limited to range that is from 0.4 to 1.0 (units: Mbps).

(d) The proposed initial QP is determined using Table 1.

(e) As for the standard video sequence, the number of frames is 60, the frame rate is 30.

4.1. Objective Evaluation. The three methods which are proposed algorithm, JVT-G012, and Wu are compared in terms of PSNR and the difference of real bits. These indicators of performance can be quantized as follows:

$$\text{PSNR} = 10 \log_{10} \frac{(2^n - 1)^2}{\text{MSE}}, \qquad (17)$$

where $\text{MSE} = \sum_{x=0}^{X-1} \sum_{Y=0}^{Y-1} (\widehat{p}(x, y) - p(x, y))^2$ and $\widehat{p}$ is the pixel value of the reconstructed video sequence. p is the original pixel value.

TABLE 2: Comparison of coding performance.

(a) $\text{PSNR}_{\text{Average}}$ (db)

Video sequence	JVT-G012	Wu	Proposed
Waterfall	36.04	35.83	36.12
Foreman	37.78	37.68	37.81
Flower	29.11	29.14	29.16
Mobile	28.12	28.28	28.37
Bus	30.12	30.18	30.24
City	34.72	34.63	34.75
Stefen	31.48	31.46	31.59
Average	32.48	32.46	32.58

(b) ΔR_{total} (bit)

Video sequence	JVT-G012	Wu	Proposed
Waterfall	0	−2288	14048
Foreman	0	−2272	2152
Flower	0	62336	69072
Mobile	0	128888	154184
Bus	0	5944	5456
City	0	−1576	−1904
Stefen	0	−1096	24
Total	0	189936	243032

Consider

$$\Delta R = R_{\text{JVT}} - Rb_{\text{JVT \& TEST}}, \qquad (18)$$

where R_{JVT} is the real bits of the JVT-G012 and $Rb_{\text{JVT \& TEST}}$ is the real bits of the algorithms which are proposed method, Wu, and JVT-G012.

Table 2 shows that the simulation results for the average PSNR and the total ΔR, which can be calculated at the range of the target bit from 0.4 to 1.0 Mbps for each video sequence, are indicated by $\text{PSNR}_{\text{Average}}$ and ΔR_{total}, respectively.

In Table 2(a), the proposed method obtains more better quality performance than JVT-G012 and Wu, although improvement is not obvious. However, ΔR_{total} shows that the proposed algorithm has quite prominent bit rate performance. The proposed method reduces 243032 bits than JVT-G012 in total although each $\text{PSNR}_{\text{Average}}$ is similar. This illustrates that the reconstructed video of proposed method has minimum actual total bits at the same or similar quality in almost all of simulations except City.

One of the important quality measures of a video is that the quality of each frame should be uniform. The existing methods have not involved this issue. The proposed method has solved this issue by selecting initial QP according to the highest and stablest quality as well as the lowest actual bits in (15). In Table 3, the quality of each frame shows an example of extreme changes. According to the JVT-G012, the value of initial QP is 25 at the target bit rate 0.7 Mbps. The maximum QP is 51 which can generate abominable quality and the minimum QP is 25 which can generate good quality. So the difference between the maximum and minimum of QPs in GOP can be a good measure for the stability of video quality. Figure 8 shows this stability performance
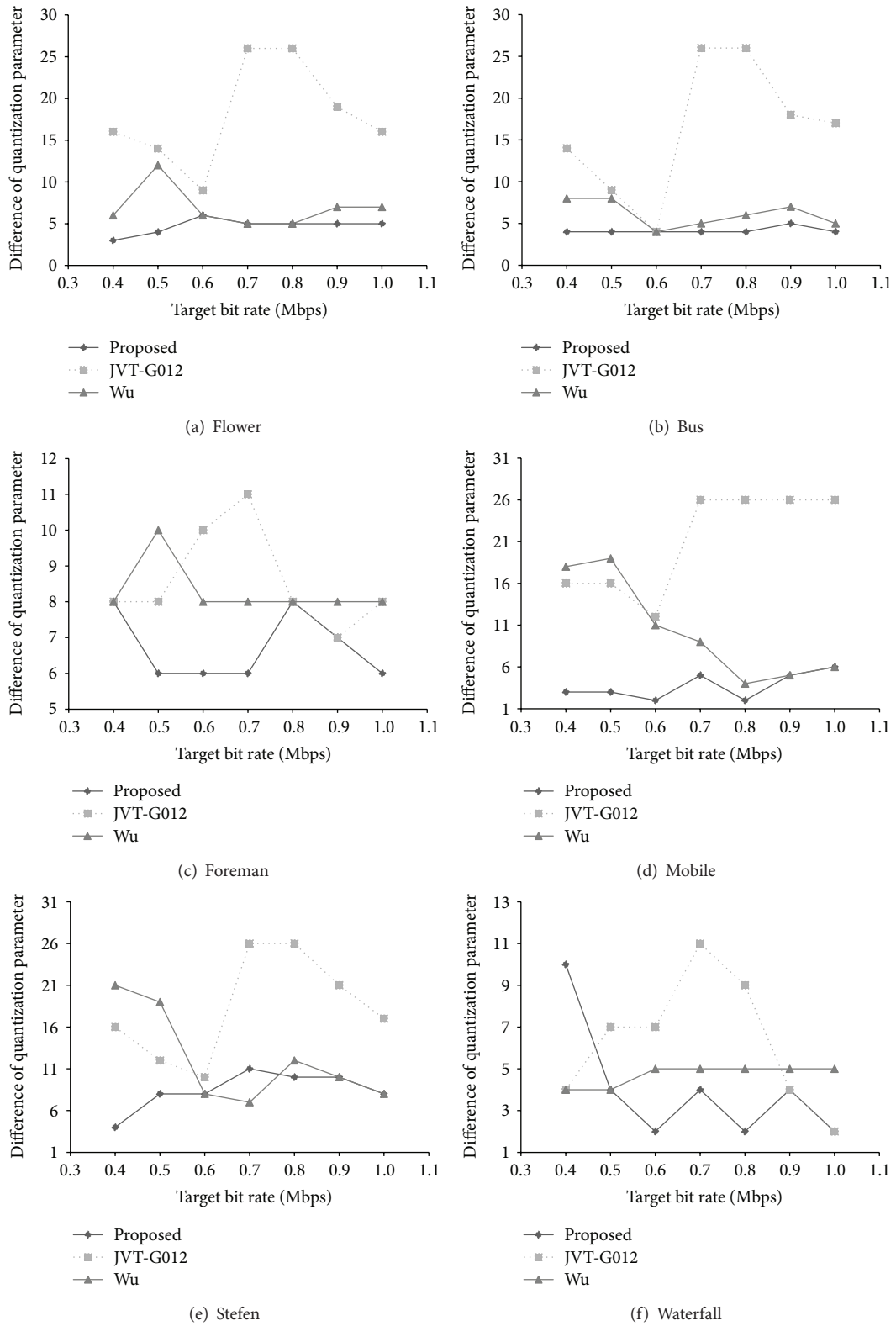
(a) Flower

(b) Bus

(c) Foreman

(d) Mobile

(e) Stefen

(f) Waterfall

Figure 8: Stability performances.

(a) Minimum QP of GOP based on JVT-G012

(b) Maximum QP of GOP based on JVT-G012

(c) Minimum QP of GOP based on Wu

(d) Maximum QP of GOP based on Wu

(e) Minimum QP of GOP based on proposed method

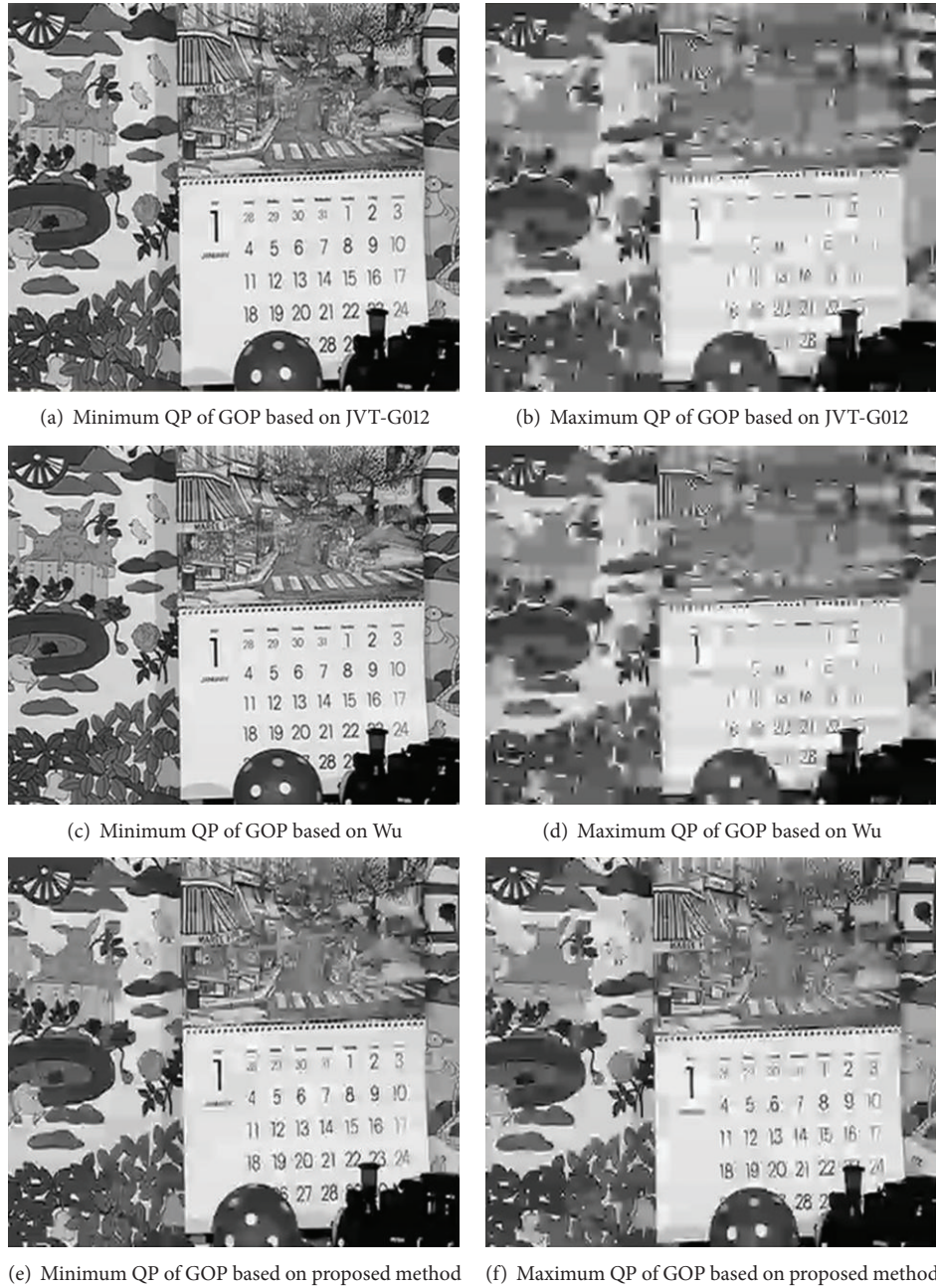(f) Maximum QP of GOP based on proposed method

Figure 9: Result of subjective evolution.

for proposed method, Wu, and JVT-G012. In Figure 8, the proposed method obtains the lower difference of QPs than others in all the cases.

In fact, PSNR can objectively and effectively assess quality of one frame. However, PSNR is not a perfect measure to evaluate the qualities of video sequences which have multiple frames. In Table 3, although the average PSNR is good, the quality of reconstructed video sequences is very low because the change of frame quality is very high. Extremely high fluctuation of the frame qualities means not only the repetition of fuzzy and vivid images but also video broken phenomenon.

So in order to evaluate video sequences, average PSNR and stability should be considered simultaneously in a GOP.

The quality of video sequences that have high complexity is very sensitive to the value of initial QP at low bit rate. For a given target bit, low initial QP can lead the large bits assignment to the first frame and insufficient bits to the other frames in a GOP to maintain qualities. Therefore, the stability of a reconstructed video sequence is a very important quality performance. Proposed optimal initial QP is determined under consideration of stability in (15). In Figure 8, the proposed method shows better results for all test

TABLE 3: Reconstructed video with extremely changing quality.

Frame number	Bit/frame	QP	Frame PSNR
0	291952	25	38.084
1	95904	25	36.900
2	42176	31	34.305
3	24272	32	32.891
4	17208	34	32.058
5	12504	36	30.613
6	10280	38	30.614
7	8848	40	29.980
8	7776	42	29.338
9	7184	44	28.591
10	6408	46	27.687
11	5720	48	26.682
12	4976	50	25.367
13	4856	51	24.405
14	5432	51	24.615

video sequences at low bitrate. Especially, in high complexity video sequences, the proposed method guarantees excellent results.

4.2. Subjective Evaluation. General video user evaluates video sequences by just looking but not by calculating PSNR or stability. This implies the importance of subjective evaluation. Frankly speaking the relationship between the results of objective evaluation and subjective one is not known. Therefore, it is not easy to convert the difference of objective evaluation results to the differences in subjective evaluation.

In this part, three methods, proposed method, JVT-G012, and Wu, are subjectively evaluated using the objective evaluation results, that is, the maximum and minimum QPs at 0.4 Mbps target bitrate. The Mobile sample video sequence is tested. Figure 9 shows result of this situation.

The average PSNRs of the proposed method, JVT-G012, and Wu are very similar at objective assessment, which are 25.22, 25.03, and 24.83, respectively.

Although Figures 9(a) and 9(c) show better quality than Figure 9(e), all objects in Figure 9(e) are vividly identified at minimum QPs. Although Figures 9(b) and 9(d) are very fuzzy and some details of frame are lost, the quality of Figure 9(f) is excellent compared with Figures 9(b) and 9(d).

In subjective evaluation, we can see also the importance of stability from the gaps in Figures 9(a), 9(b), 9(c), 9(d), 9(e), and 9(f), respectively.

5. Conclusions

In order to obtain the optimal initial QP for the various video sequences which have different spatial and temporal complexities, we propose a simple and high performance initial QP determining method based on motion vector map and temporal complexity to decide an initial QP in given target bit rate. Four sample video sequences are selected according to the temporal complexity which is measured

using motion vectors map, and the spatial complexities of four sample video sequences are computed according to proposed method. For any video sequences, the initial QP can be easily determined from matrices by target bit rate and mapped spatial complexity using proposed mapping method. Experimental results show that the proposed algorithm can obtain more outstanding objective and subjective performance than other conventional determining methods. In the future, one of the further research areas will be the development of quantitative measure for the temporal complexity. The study on temporal complexity will provide a hint to explain or to solve the exceptional case for H.264 AVC coding.

Conflict of Interests

The authors declare that they have no financial or personal relationships with other people or organizations that can inappropriately influence their work; there is no professional or other personal interests of any nature or kind in any product, service, and/or company that could be construed as influencing the position presented in, or the review of, the paper entitled.

References

[1] MPEG-2: ISO/IEC JTC1/SC29/WG11 and ITU-T, ISO/IEC, 13818-2, Information Technology_Generic Coding of Moving Pictures and Associated Audio Information: Video, ISO/IEC and ITU-T International Standard Draft, 1994.

[2] ITU-T Rec. H. 264/ISO/IEC, 11496-10, Advanced Video Coding, Final Committee Draft, Document JVT-F100, December 2002.

[3] H. Kalva, "Standards: the H.264 video coding standard," *IEEE Multimedia*, vol. 13, no. 4, pp. 86–90, 2006.

[4] T. Wiegand, G. J. Sullivan, G. Bjøntegaard, and A. Luthra, "Overview of the H.264/AVC video coding standard," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 13, no. 7, pp. 560–576, 2003.

[5] H. Schwarz, D. Marpe, and T. Wiegand, "Overview of the scalable video coding extension of the H.264/AVC standard," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 17, no. 9, pp. 1103–1120, 2007.

[6] G. J. Sullivan and T. Wiegand, "Video compression-from concepts to the H.264/AVC standard," *Proceedings of the IEEE*, vol. 93, no. 1, pp. 18–31, 2005.

[7] G. J. Sullivan, J. Ohm, H. Woo-Jin, and T. Wiegand, "Overview of the high efficiency video coding (HEVC) standard," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 22, no. 12, pp. 1649–1668, 2012.

[8] T. Wiegand, H. Schwarz, A. Joch, F. Kossentini, and G. J. Sullivan, "Rate-constrained coder control and comparison of video coding standards," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 13, no. 7, pp. 688–703, 2003.

[9] Z. Chen and K. N. Ngan, "Recent advances in rate control for video coding," *Signal Processing: Image Communication*, vol. 22, no. 1, pp. 19–38, 2007.

[10] H. Wang and S. Kwong, "Rate-distortion optimization of rate control for h.264 with adaptive initial quantization parameter determination," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 18, no. 1, pp. 140–144, 2008.

[11] W. Wu and H. K. Kim, "A novel rate control initialization algorithm for H.264," *IEEE Transactions on Consumer Electronics*, vol. 55, no. 2, pp. 665–669, 2009.

[12] S. Hu, H. Wang, S. Kwong, and C.-C. J. Kuo, "Novel rate-quantization model-based rate control with adaptive initialization for spatial scalable video coding," *IEEE Transactions on Industrial Electronics*, vol. 59, no. 3, pp. 1673–1684, 2012.

[13] "JM Reference Software Version 9.3," http://iphome.hhi.de/suehring/download.

[14] "YUV Video Sequences," http://trace.eas.asu.edu/yuv/index.html.

[15] Y.-H. Huang, T.-S. Ou, and H. H. Chen, "Fast decision of block size, prediction mode, and intra block for H.264 intra prediction," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 20, no. 8, pp. 1122–1132, 2010.