Circuits Syst Signal Process (2014) 33:939–957
DOI 10.1007/s00034-013-9675-3

# A Segmentation-Based Chroma Intra Prediction Coding Scheme for H.264/AVC

**Qingbo Wu · Jian Xiong · Bing Luo · Zhengning Wang**

Received: 12 December 2012 / Revised: 4 September 2013 / Published online: 11 October 2013


**Abstract** In this paper, we propose a novel segmentation-based intra prediction coding scheme for low-bitrate video coding. Different coding schemes are separately designed for the luma and chroma components in our proposed method. The traditional block-based coding scheme is still used for the luma components, and the segmentation-based coding scheme is developed for the chroma components. The segmentation operation is used for the reconstructed luma components, which groups similar pixels together and produces a set of homogenous regions. Here, these local and homogenous regions are referred to superpixels. By utilizing the spatial correlation between the luma and chroma planes, we transfer the segmentation result of the luma components to the chroma components, which will not induce any side information in the chroma intra prediction coding. Instead of using the macroblock (MB) as the coding unit, the proposed method implements the chroma intra prediction in each superpixel, and the original pixels in each superpixel are employed to substitute the neighboring reconstructed samples in the prediction process. The experimental results show that the proposed method can achieve an average 0.20 dB and up to 0.63 dB coding gains in comparison to the directional intra prediction scheme for H.264/AVC low-bitrate video coding.

**Keywords** Segmentation · Intra prediction · Video coding · H.264/AVC · HEVC

Q. Wu (✉) · J. Xiong · B. Luo · Z. Wang
School of Electronic Engineering, University of Electronic Science and Technology of China,
No. 2006, Xiyuan Ave, West Hi-Tech Zone, Chengdu 611731, China
e-mail: wqb.uestc@gmail.com

J. Xiong
e-mail: jxiong219@gmail.com

B. Luo
e-mail: Mathild1987@163.com

Z. Wang
e-mail: zhengning.wang@uestc.edu.cn

## 1 Introduction

Intra prediction is an essential technology in video coding standards. As discussed in [27, 34], H.264/AVC is superior to JPEG2000 and JPEG by employing efficient directional intra prediction (DIP) scheme. In H.264/AVC [30, 33, 37], similar intra-frame coding strategies are designed for the luma and chroma components separately. For the luma components, there are three intra-frame prediction types, called Intra-4 $\times$ 4, Intra-8 $\times$ 8, and Intra-16 $\times$ 16, and nine spatial directional intra prediction modes. For the chroma components, only one prediction type and four similar prediction modes like Intra-16 $\times$ 16 type are supported. The fixed reference samples and extrapolation methods in DIP scheme efficiently trade-off the decorrelation performance and computational complexity. This also facilitates some fast intra mode selection algorithms developed based on the edge or directional information [20, 35]. However, this inflexible design also hinders improvement of intra prediction precision.

There are two limitations in DIP scheme for both luma and chroma components. First, the DIP scheme is practically suitable for the block with simple contours when the selected intra mode is consistent with the orientation of the contour. However, it fails in the regions with more complex textures. Second, because of the progressive block coding, only the reconstructed upper and left pixels are used to generate the predictor for current block. The low-quality reference pixels would degrade the performance of DIP especially for the case that the coarse quantization is used. To improve the intra prediction performance, many block-based intra prediction methods have been proposed. Two strategies are employed in these methods, refined prediction and flexible partition. For the first strategy, a bidirectional intra prediction scheme is developed in [39], which adds more directional prediction modes. The template matching method is proposed in [31], which utilizes the nonlocal spatial correlation in the intra-frame. The line-by-line scheme is proposed to decrease the spatial distance for large block intra prediction [6, 7]. A position-dependent intra prediction scheme is developed in [41] to update the extrapolation coefficients with least square method. In [21], the edge information is extracted to enhance the intra prediction precision. For the second strategy, a segmentation tree is derived by Multidimensional Multiscale Parser (MMP) algorithm to further partition the residual block into more refined scales [10]. In [8], the geometry-adaptive block partitioning method is proposed to implement the geometric partition in an MB. In [28], the intra-frame is partitioned into two subframes, and different coding schemes are separately employed for each of them.

For the methods mentioned above, each color plane is encoded independently. The correlation between the color planes is not exploited. In [14, 15], some interesting color set partitioning in hierarchical tree (CSPIHT) schemes is discussed. The interdependency between luma and chroma components is utilized to code the color video sequences. This inspires us to improve the chroma intra coding by using the correlation between luma and chroma planes. In addition, the block-based intra prediction methods cannot avoid high-texture areas in the regular block partition. The similar MBs also cannot be merged to save the bits for intra modes, MB types, and so on. The geometry-adaptive partition scheme [8] tries to improve the fixed block partition with a parametric line partition to facilitate the intra prediction in some more

homogenous areas. Particularly, a similar target can be found in many object or region segmentation methods [1, 11, 17, 19, 23–25, 29], which try to cluster the pixels with the same or similar features and separate the ones that are different. A local and coherent region that contains a set of similar pixels is referred to as a superpixel in these segmentation methods. Unlike the regular MB partition, the superpixels could better preserve the contour of the object in partitioning an image. The content in a superpixel shows obvious homogenous characteristics. This inspires us to improve the image partition with segmentation result.

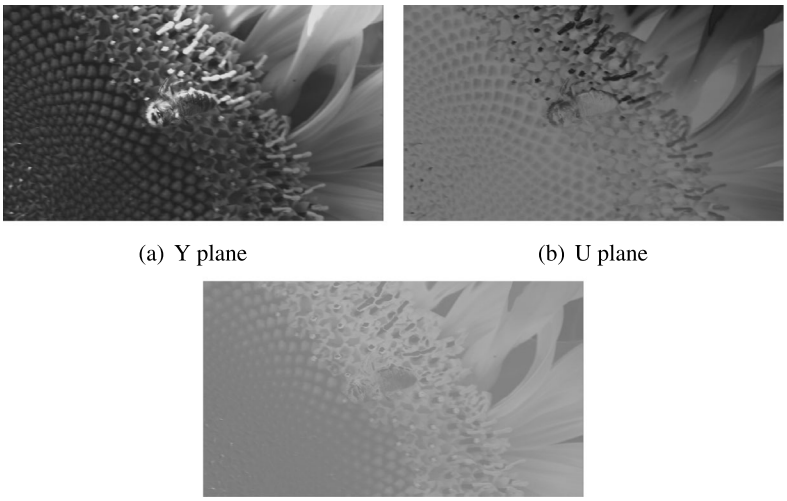
In this paper, we propose a segmentation-based chroma intra prediction coding scheme. Unlike the conventional video coding schemes, which encode the luma and chroma components independently, we utilize the reconstructed luma components to improve the performance of chroma intra coding. Based on the spatial correlation between the luma components and the chroma components, we first implement the segmentation operation in the reconstructed luma image, and then the superpixel segmentation result is transferred to the chroma components. For the YUV input with 4:2:0 chroma subsampling, a down-sampled superpixel with $1/2$ sampling rate in horizontal and vertical directions is shared for both the U and V components. In chroma intra prediction, the prediction unit is a superpixel instead of the conventional MB. In view of the reference pixels' quality decreasing under coarse quantization, we directly employ original pixels in each superpixel to implement the intra prediction. Because of the homogenous property in each superpixel, the proposed method facilitates the decorrelation task in chroma components. The segmentation-based partition efficiently merges similar regions together. Comparing with the MBs partition, we can use much less superpixels to achieve efficient intra prediction, which further improves the coding efficiency by saving the bits for the syntaxes like intra modes, MB types, etc.

The rest of the paper is organized as follows. Section 2 makes the rate-distortion analysis for the segmentation-based coding scheme. Then, the proposed approach is described in detail in Sect. 3. The experimental results are presented in Sect. 4. Finally, we draw the conclusion of this paper in Sect. 5.

## 2 Rate-Distortion Analysis for Segmentation-Based Coding Scheme

As discussed in [5, 40], although there is no relation between luma and chroma intensity values, the spatial correlation between luma and chroma components exist. To visualize this correlation, we show the Y, U, and V components of the sequence *sunflower* in Fig. 1. We can see that similar object and contour locations can be found in Figs. 1(a)–(c). This inspires us to implement more efficient region partitions for chroma components by directly utilizing the spatial distribution information in the luma components. In this way, a pixel-wise partition can be implemented in the chroma components. For example, a segmentation transferring result for the sequence *sunflower* is shown in Fig. 2.

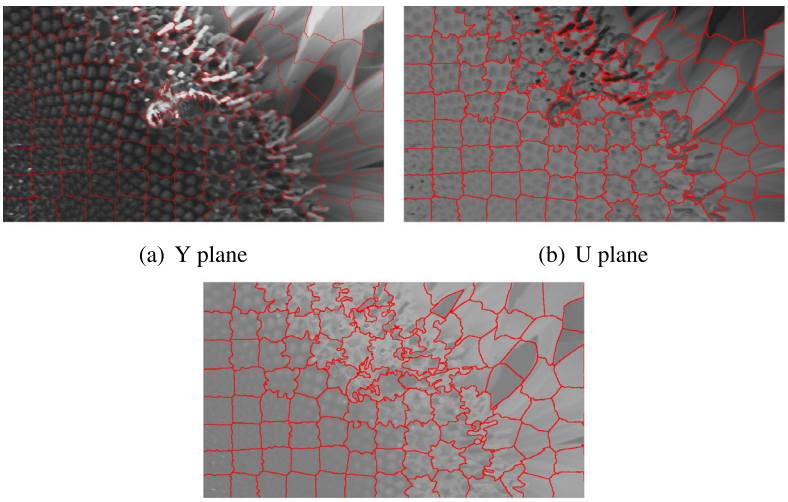
In the same way as discussed in [8, 9], we further analyze the rate-distortion performance of the segmentation-based image partition scheme. For a classic quadtree



(a) Y plane (b) U plane

(c) V plane

**Fig. 1** Spatial correlation between luma and chroma components for the sequence *sunflower*



(a) Y plane (b) U plane

(c) V plane

**Fig. 2** Spatial correlation between luma and chroma components for the sequence *sunflower*

partition coding scheme, the asymptotic dependence of distortion on the rate can be formulated as

$$D_{QT}(R) \sim \frac{\log R}{R}, \tag{1}$$

where $D_{QT}$ denotes the distortion for the quadtree-based coding scheme, and $R$ denotes the corresponding rate. This is justified based on the following analysis. For

the tree with $J$ levels, let $N_J$ denote the number of the quadtree leaves that cross the smooth contour. Then we can deduce that

$$N_J \sim \sum_{j=0}^{J} 2^j \sim 2^J, \tag{2}$$

and the total bit cost can be represented as

$$R(J) \sim N_J J \sim J2^J. \tag{3}$$

Then, the distortion can be further deduced as

$$D_{QT}(J) \sim N_J 2^{-2J} \sim 2^{-J}. \tag{4}$$

However, in our proposed segmentation-based partition method, the regular blocks corresponding to the quadtree leaves are further subdivided into different areas according to the pixels' similarity. Since the chroma segmentation result comes from the luma components, there is only negligible side information, such as segmentation parameter for the luma image, induced to obtain the partition. It has been verified that when the wedgelet-like geometry partition is employed, the $R$–$D$ asymptotic behavior will approach

$$D_{\text{Wedge}}(R) \sim \frac{\log R}{R^2}, \tag{5}$$

where $D_{\text{Wedge}}$ denotes the distortion of the wedge-like partition scheme.

Since the wedge-like geometry separation lines can be approximated by quanting the object boundary, we can conclude that the segmentation-based block partition method produces a similar $R$–$D$ asymptotic behavior like the wedge-like partition scheme, i.e.,

$$D_{\text{proposed}}(R) \sim \frac{\log R}{R^2} \tag{6}$$

where $D_{\text{proposed}}$ denotes the distortion of our proposed method. Since the decay of distortion with the rate in (6) is faster than the one in (1), the proposed methods can achieve more efficient coding relative to the quadtree partition.

## 3 Proposed Algorithm

The framework of our proposed method is illustrated in Fig. 3. The input video is first divided into the separate luma intra-frame $I_{\text{Y}}$ and chroma intra-frames $I_{\text{U}}/I_{\text{V}}$. For luma component $I_{\text{Y}}$, the conventional directional intra prediction coding is employed. Following the luma components coding pass, the segmentation operation is implemented on the reconstructed luma components to produce the superpixel segmentation result. Elaborate prediction and coding schemes are designed for the chroma components, and the detail descriptions are presented in the following.

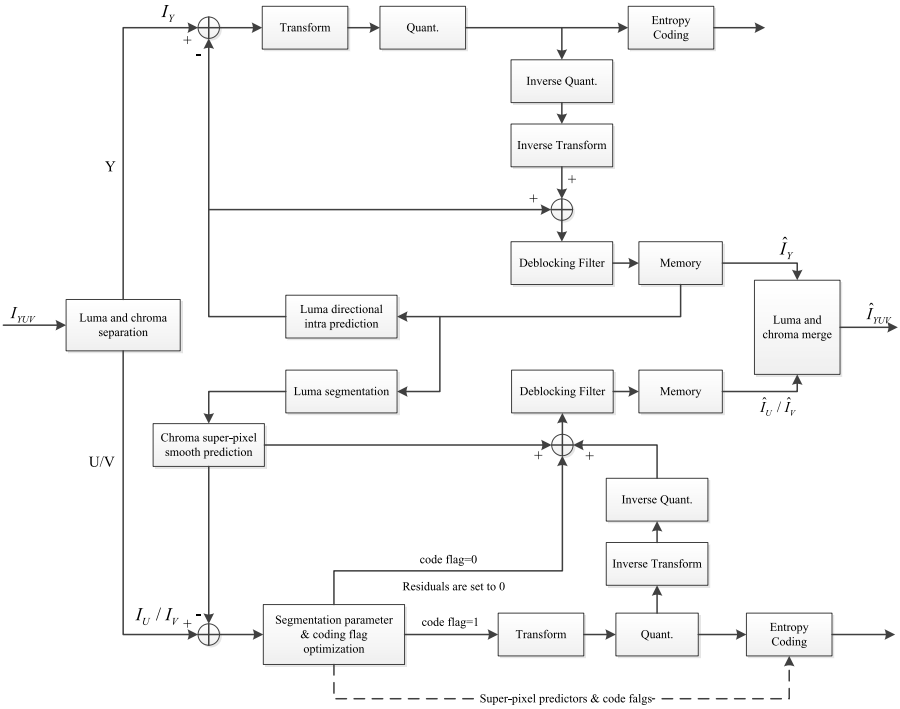



**Fig. 3** The framework of the segmentation-based chroma intra prediction coding

### 3.1 Review of Segmentation Method

Unlike the regular grid partition, the segmentation methods try to cluster the pixels into different perceptually meaningful regions. The pixels in each segmented region share the similar image features, which significantly facilitates the redundancy reduction task for video coding.

In general, the segmentation methods can be divided into three categories according to the availability of the object prior, i.e., supervised, weakly supervised, and unsupervised methods. Here, the object prior is necessary for the supervised segmentation. Many algorithms have been proposed to obtain accurate object prior, such as saliency detection-based method [17, 18] and learning-based methods [19, 29]. For the weakly supervised methods, only partial prior information is available. For instance, the cosegmentation problem tries to segment the common objects in a pair of images with the knowledge that there are common objects in the image pairs, but the accurate objects locations are unknown. The graph-based shortest path algorithm [23] and an active contour-based model [24] were proposed to solve this task. Although the supervised and weakly supervised methods could achieve accurate segmentation result, they require some mutual manipulations or prior information that is not suitable for the video coding application. In contrast, the unsupervised methods do not need any prior information. In [25], Moore et al. proposed a greedy reg-

ular lattice method to generate the superpixel segmentation. In [1], Achanta et al. proposed a fast superpixel segmentation algorithm, which is termed as SLIC. The SLIC is based on an adaption of the $k$-means cluster method, where all pixels are grouped to $k$ classes, and each pixel is classified to its nearest cluster center based on the feature distance. Each superpixel consists of a cluster of pixels in the same class. In view of the advantage in complexity and without prior information, in this paper, we employ the SLIC algorithm to generate the superpixel segmentation result.

### 3.2 Superpixel-Based Smooth Prediction

As discussed in Sect. 1, the low reconstruction quality of reference pixels limits the prediction performance for DIP scheme. To overcome this issue, we employ original pixels in each superpixel as the reference samples instead of the neighboring reconstructed ones. Let $\boldsymbol{S}$ denote the set of all pixels in a superpixel, and $N_s$ denote the number of pixels in $\boldsymbol{S}$. Then, the intra prediction optimization problem in terms of mean square error (MSE) can be formulated as

$$\begin{aligned} &\min \sum_{i=1}^{N_s} \| I_i - \mathbf{w}_i \cdot \mathbf{R}_i \|_2^2 \\ &\text{s.t.} \quad \| \mathbf{w}_i \|_1 = 1, \quad \mathbf{w}_i \geq 0, \end{aligned} \tag{7}$$

where $I_i$ denotes the $i$th pixel in $\boldsymbol{S}$, $\mathbf{R}_i$ denotes the vector of $I_i$'s neighboring pixels in $\boldsymbol{S}$, $\mathbf{w}_i$ is the weights vector for the elements of $\mathbf{R}_i$, and the symbol "$\cdot$" represents the dot product of two vectors.

In our proposed method, the additional predictor is induced since the original samples are not available at the decoder. This pixel-wise prediction scheme in (7) brings too much overhead comparing with the block-wise intra modes in DIP. However, unlike the fixed block partition in DIP scheme, there is a more homogenous property in each superpixel that facilitates the decorrelation task of intra prediction. To tradeoff the prediction precision and the side information for signaling the predictors, we simplify the pixel-wise prediction scheme in (7) with a superpixel-wise prediction scheme, that is,

$$\begin{aligned} &\min \sum_{i=1}^{N_s} \| I_i - \mathbf{w}_s \cdot \mathbf{I}_s \|_2^2 \\ &\text{s.t.} \quad \| \mathbf{w}_s \|_1 = 1, \quad \mathbf{w}_s \geq 0, \end{aligned} \tag{8}$$

where $\mathbf{I}_s$ denotes the vector composed of all pixels in $\boldsymbol{S}$, and $\mathbf{w}_s$ is the weight vector for $\boldsymbol{S}$.

Let $E(\cdot)$ denote the expectation operator, and $P_s$ denote the predictor in $\boldsymbol{S}$, where $P_s = \mathbf{w}_s \cdot \mathbf{I}_s$. The MSE of intra prediction in (8) can be represented as the dependent variable of $P_s$, that is,

$$
\begin{aligned}
f(P_s) &= \sum_{i=1}^{N_s} \|I_i - P_s\|_2^2 \\
&= \sum_{i=1}^{N_s} P_s^2 - 2P_s \sum_{i=1}^{N_s} I_i + \sum_{i=1}^{N_s} I_i^2 \\
&= N_s\left[P_s^2 - 2E(\mathbf{I}_s)P_s + E\left(\mathbf{I}_s^2\right)\right]. \qquad (9)
\end{aligned}
$$

From the definition in (9) we can deduce that the intra prediction MSE satisfies the inequality $f(P_s) \geq 0$. Then, the optimal solution to minimize the $f(P_s)$ can be given by $P_s = E(\mathbf{I}_s)$. That is, the predictor for each superpixel is just the mean value of the corresponding region in the original picture. Correspondingly, each weighting factor in $\mathbf{w}_s$ can be derived as $\mathbf{w}_s(i) = 1/N_s$ $(i = 1, 2, \ldots, N_s)$ according to the constraint in (7). In this way, only one additional predictor $P_s$ is needed for each superpixel, and we name this method as the superpixel-based smooth prediction in our proposed algorithm.

After the superpixel-based smooth prediction, the MB-based coding operation will be implemented on the chroma residual image. Since the proposed prediction method can efficiently achieve the decorrelation task, we only encode the residual for the MB that achieves better rate distortion performance. Detailed descriptions will be given in the following.

### 3.3 Optimization Scheme for Segmentation-Based Chroma Coding Pass

As discussed in Sect. 3.1, we design a superpixel-wise prediction method to trade-off the prediction precision and the overhead for signaling the predictor. Accordingly, the optimization scheme for selecting appropriate segmentation parameters and MB coding flags is essential in our proposed method. By employing the classic rate distortion optimization (RDO) framework, we incorporate the luma components segmentation process into the chroma coding pass.

To make the proposed method tractable, some predefined segmentation parameters for different resolutions are specified. Let $N_c$ denote the number of chroma MBs, $N_p$ denote the number of candidate segmentation parameters, $T_c$ denote the set of candidate segmentation parameters, and $o_c$ denote the selected segmentation parameter. Let $T_f$ denote all available coding flags combination, and $o_f(i)$ denote the selected coding flag for the $i$th chroma MB. Let $o_F$ denote selected chroma MB's coding flags combination, that is, $o_F = o_f(1) \cup o_f(2) \cup \cdots \cup o_f(N_c)$. The competing process to minimize the rate distortion cost for coding the chroma components can be formulated as

$$
\begin{aligned}
&\min_{o_c \in T_c, o_F \in T_f} J(o_F, o_c) \\
&\text{with } J(o_F, o_c) = D(o_F, o_c) + \lambda R(o_F, o_c), \qquad (10)
\end{aligned}
$$

where $J(o_F, o_c)$, $D(o_F, o_c)$, and $R(o_F, o_c)$ denote the overall rate distortion (RD) cost, reconstruction distortion, and rate for the chroma components. It should be

noted that the values of $J(o_F, o_c)$, $D(o_F, o_c)$, and $R(o_F, o_c)$ represent the sums for both of the chroma components U and V. For clarity, the reconstruction distortion and rate are further expanded as follows:

$$\begin{aligned} D(o_F, o_c) &= \sum_{i=1}^{N_c} D_i(o_F, o_c), \\ R(o_F, o_c) &= R^p(o_F, o_c) + R^f(o_F, o_c) + \sum_{i=1}^{N_c} R_i^r(o_F, o_c), \end{aligned} \tag{11}$$

where $D_i(o_F, o_c)$ denotes the reconstruction distortion of the $i$th chroma MB under the coding option $o_f$ and $o_c$. $R^p(o_F, o_c)$ denotes the rate for the predictors of all superpixels, and $R^f(o_F, o_c)$ denotes the rate for the flags that indicate whether the MB's residual is encoded or not. $R_i^r(o_F, o_c)$ is the rate for encoding the residual of the $i$th chroma MB under the coding options $o_f$ and $o_c$. In our proposed method, the eight-bit fixed-length coding is employed for the predictor of each superpixel. For the MB's coding flags, a PAQ8 [22] encoder is employed to further compress these binary data.

In the whole coding pass, both the optimal segmentation parameter and coding flags for each chroma MB are determined according to the RDO criteria. To simplify the multiparameter global optimization problem in (10), we implement the RDO process in each MB unit according to the independent assumption in [36]. Particularly, the segmentation parameter is used as a priori in determining the coding flag in each MB.

Let $o_c(i)$ denote the $i$th segmentation parameter, and $o_f(j)$ denote the coding flag of the $j$th chroma MB. Let $J_{\min}^{\mathrm{MB}}$ denote the minimum RD cost in an MB, and $J_{\min}^{T}$ denote the minimum RD cost for the whole chroma images. Let $o_c^{\mathrm{opt}}$ denote the optimal segmentation parameters, and $o_f^{\mathrm{opt}}(j)$ denote the optimal coding flag for the $j$th MB. We further give the detailed descriptions for the proposed coding pass in Fig. 4. In this way, the segmentation parameters and coding flags can be interactively optimized for the chroma components.

## 4 Experimental Results

To verify the efficiency and extensibility of our proposed method, we implement the segmentation-based chroma intra coding scheme on both the H.264/AVC reference software VCEG KTA2.4r1 [13] and HEVC reference software HM4.1 [12]. Since the proposed method focuses on improving the intra prediction under coarse quantization, the quantization parameters for both of the reference softwares are set to QP $= \{44, 46, 48, 50\}$. The SLIC method [1] is employed in our proposed method to obtain the superpixel partitions. All the candidate parameters are listed in Table 1, where $m$ represents the compactness parameter, and $k$ denotes the desired superpixels' number. For the H.264/AVC intra-only simulation, the other experimental conditions are set according to the common simulation conditions [32]. The

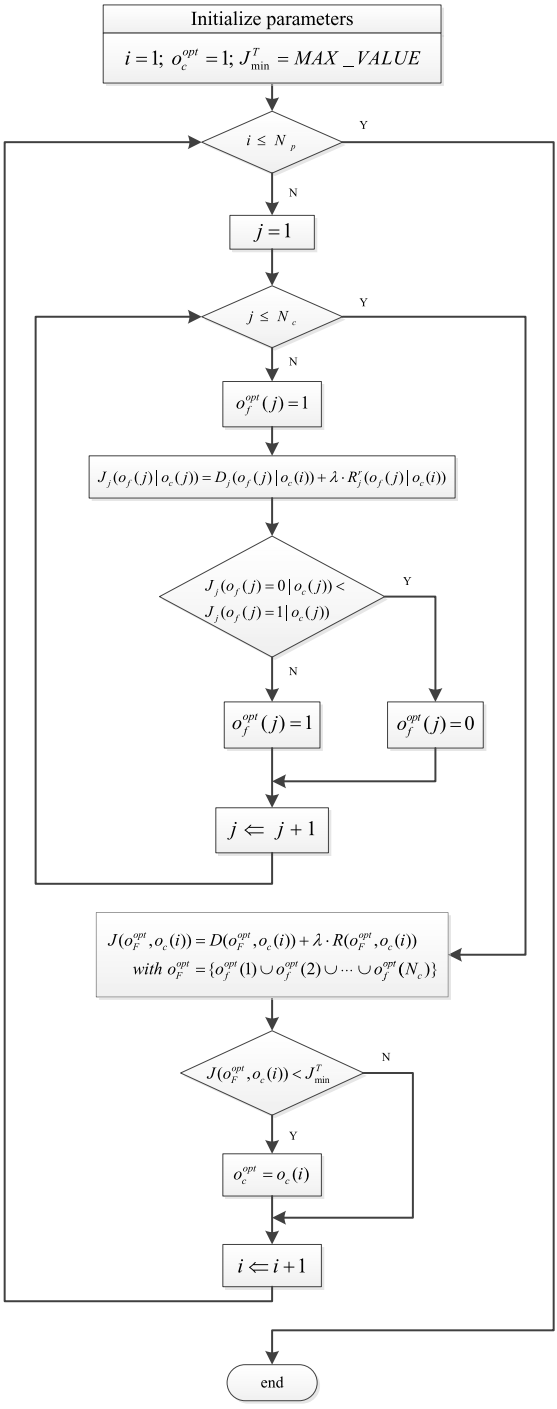



**Fig. 4** Determination of segmentation parameters and coding flags

**Table 1** Segmentation parameters

| Index | 720p | | 1080p | |
|---|---|---|---|---|
| | Compactness ($m$) | Number ($k$) | Compactness ($m$) | Number ($k$) |
| 1 | 10 | 500 | 10 | 600 |
| 2 | 5 | 500 | 5 | 600 |
| 3 | 10 | 450 | 1 | 600 |
| 4 | 5 | 450 | 10 | 550 |
| 5 | 10 | 400 | 5 | 550 |
| 6 | 5 | 400 | 1 | 550 |
| 7 | 10 | 350 | 10 | 500 |
| 8 | 5 | 350 | 5 | 500 |
| 9 | 10 | 300 | 1 | 500 |
| 10 | 5 | 300 | 10 | 450 |
| 11 | 10 | 250 | 5 | 450 |
| 12 | 5 | 250 | 1 | 450 |
| 13 | 10 | 200 | 10 | 400 |
| 14 | 5 | 200 | 5 | 400 |
| 15 | 10 | 150 | 1 | 400 |
| 16 | 5 | 150 | 10 | 350 |
| 17 | 10 | 100 | 5 | 350 |
| 18 | 5 | 100 | 1 | 350 |
| 19 | 10 | 50 | 10 | 300 |
| 20 | 5 | 50 | 5 | 300 |

H.264/AVC High Profile ("Anchor-1") scheme is utilized as the benchmark. Both the luma chroma separation (LCS) scheme and the line-based intra $16 \times 16$ prediction scheme [7] are compared with our proposed method in the H.264/AVC simulation. For the HEVC intra-only simulation, the other configuration parameters are set according to the common test condition [3]. The All Intra-High efficiency (AI-HE) profile with LMChroma tool off is treated as the benchmark, which is labeled by "Anchor-2." Then, we compare our proposed method with the linear model chroma (LMChroma) prediction scheme [4] in the HEVC simulation. As a metric to evaluate coding efficiency, BD-BR (Bjonteggard Delta Bit-Rate) and BD-PSNR (Bjonteggard Delta PSNR) [2] are used here. Particularly, since we modify the luma and chroma coding passes simultaneously, the average YUV PSNR [16, 26] is employed to calculate the overall coding performance. It should be noted that we only compute the rate-distortion cost of the luma block in determining its coding option, where there is no syntax changes for the luma coding pass. The definition of the average YUV PSNR ($\mathrm{PSNR}_{\mathrm{YUV}}$) is given by

$$\mathrm{PSNR}_{\mathrm{YUV}} = (6 \cdot \mathrm{PSNR}_{\mathrm{Y}} + \mathrm{PSNR}_{\mathrm{U}} + \mathrm{PSNR}_{\mathrm{V}})/8, \tag{12}$$

where $\mathrm{PSNR}_{\mathrm{Y}}$, $\mathrm{PSNR}_{\mathrm{U}}$, and $\mathrm{PSNR}_{\mathrm{V}}$ denote the PSNR for Y, U, and V components, respectively.

**Table 2** Coding gain of different schemes in terms of BDBR (%) and BDPSNR (dB) for H.264/AVC simulation

| Sequence | | LCS | | Choi et al. [7] | | Proposed | |
|---|---|---|---|---|---|---|---|
| | | BDBR | BDPSNR | BDBR | BDPSNR | BDBR | BDPSNR |
| 1080p | pedestrian_area | −2.01 | 0.11 | −2.00 | 0.11 | −3.92 | 0.22 |
| | rush_hour | −1.63 | 0.10 | −2.61 | 0.17 | −5.13 | 0.33 |
| | sunflower | 2.08 | −0.15 | −1.42 | 0.10 | −8.62 | 0.63 |
| | tractor | 2.60 | −0.10 | −1.87 | 0.08 | −2.74 | 0.11 |
| | blue_sky | 2.36 | −0.12 | −0.18 | 0.01 | −4.76 | 0.23 |
| Average 1080p | | 0.68 | −0.032 | −1.62 | 0.09 | −5.03 | 0.30 |
| 720p | ducks_take_off | −2.73 | 0.09 | −0.67 | 0.02 | −7.94 | 0.27 |
| | old_town_cross | −2.74 | 0.05 | −1.22 | 0.02 | −6.85 | 0.14 |
| | park_joy | 0.47 | −0.01 | 0.48 | −0.01 | −5.78 | 0.12 |
| | parkrun | −1.24 | 0.02 | −0.52 | 0.01 | −4.51 | 0.09 |
| | shields | 3.23 | −0.08 | −0.80 | 0.02 | −1.48 | 0.04 |
| | stockholm | −1.71 | 0.04 | −1.68 | 0.04 | −2.50 | 0.06 |
| Average 720p | | −0.79 | 0.02 | −0.74 | 0.02 | −4.84 | 0.12 |
| Average Total | | −0.12 | −0.01 | −1.14 | 0.05 | −4.93 | 0.20 |

In addition, to further evaluate the computational complexity of different schemes, a quantitative comparison in terms of time consumption is made in this section. Here, we use the percentage of coding time difference ($\Delta T\%$) as the metric defined as

$$\Delta T = \frac{T_{\text{cand}} - T_{\text{anc}}}{T_{\text{anc}}} \times 100, \tag{13}$$

where $T_{\text{cand}}$ and $T_{\text{anc}}$ denote the coding time of the candidate scheme and the anchor, respectively.

## 4.1 Coding Performance

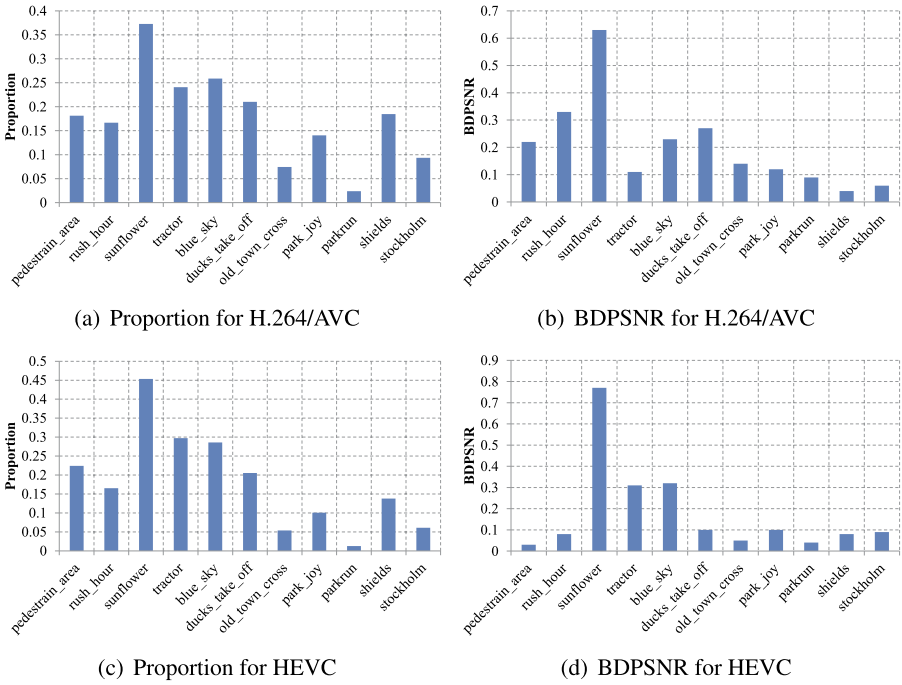
The detailed coding gains for the H.264/AVC and HEVC simulations have been shown in Tables 2 and 3, respectively. It can be seen that the proposed method achieves a consistent coding gain for all of the test sequences. For both H.264/AVC and HEVC simulations, on average, 0.20 dB and 0.18 dB coding gains are achieved. Especially for the sequence *sunflower*, up to 0.63 dB and 0.77 dB PSNR improvements can be achieved in the H.264/AVC and HEVC simulations, respectively. To investigate the BDPSNR distribution for different sequences, we further count the chroma coding coefficients' proportion with respect to the total bitrate in the anchor scheme. The results have been shown in the Fig. 5. Here, Figs. 5(a) and (c) show the chroma coefficients' proportions in H.264/AVC and HEVC, respectively.

**Table 3** Coding gain of different schemes in terms of BDBR (%) and BDPSNR (dB) for HEVC simulation

| Sequence | | LMChroma | | Proposed | |
|---|---|---|---|---|---|
| | | BDBR | BDPSNR | BDBR | BDPSNR |
| 1080p | pedestrian_area | −1.31 | 0.05 | −0.67 | 0.03 |
| | rush_hour | −0.95 | 0.04 | −1.68 | 0.08 |
| | sunflower | −22.91 | 1.70 | −12.63 | 0.77 |
| | tractor | −16.69 | 0.72 | −7.97 | 0.31 |
| | blue_sky | −6.14 | 0.29 | −6.97 | 0.32 |
| Average 1080p | | −9.6 | 0.56 | −5.98 | 0.30 |
| 720p | ducks_take_off | −15.29 | 0.56 | −2.82 | 0.10 |
| | old_town_cross | −1.87 | 0.03 | −2.56 | 0.05 |
| | park_joy | −10.75 | 0.20 | −5.07 | 0.10 |
| | parkrun | −0.82 | 0.01 | −2.81 | 0.04 |
| | shields | −9.15 | 0.15 | −4.83 | 0.08 |
| | stockholm | −5.83 | 0.08 | −6.15 | 0.09 |
| Average 720p | | −7.28 | 0.17 | −4.04 | 0.08 |
| Average Total | | −8.33 | 0.35 | −4.92 | 0.18 |

Figures 5(b) and (d) show the BDPSNR improvements in H.264/AVC and HEVC, respectively. It can be seen that although there are some fluctuations between the distributions of the chroma coefficients proportions and the coding gains, a larger chroma coefficients' proportion usually produces the larger coding gain. It is reasonable because the proposed method focuses on chroma intra coding improvement.

In the H.264/AVC simulation, since the LCS scheme still employs the same intra prediction methods in the Anhcor-1 scheme, its coding performance does not significantly change. As shown in Table 2, the average BDPSNR of LCS scheme is only −0.01 dB. The coding gain for the method proposed in [7] is a little better, whose BDPSNR is 0.05 dB on average. However, this coding gain is not as high as the result reported in [7]. It should be noted that the method proposed in [7] focuses on high-bitrate environment, which is different from the low-bitrate coding issue in this paper. The performance of the line-based intra prediction is based on the assumption that the correlation between two pixels is inversely relative to the distance between them. As discussed in [38], this assumption is valid for the fine quantization. For the low-bitrate video coding, coarse quantization significantly reduces the correlation of the neighboring pixels, which degrades the performance of the line-based intra prediction method [7]. In the HEVC simulation, the LMChroma scheme achieves higher coding gains with respect to our proposed method. Since the proposed method is initially designed for the H.264/AVC encoder, some new characteristics of HEVC (e.g., larger prediction unit size and transform size) are not considered in our experiment. In summary, there are three factors accounting for the comparison result in the HEVC simulation. First, the segmentation parameters in Table 1 are initially designed for the



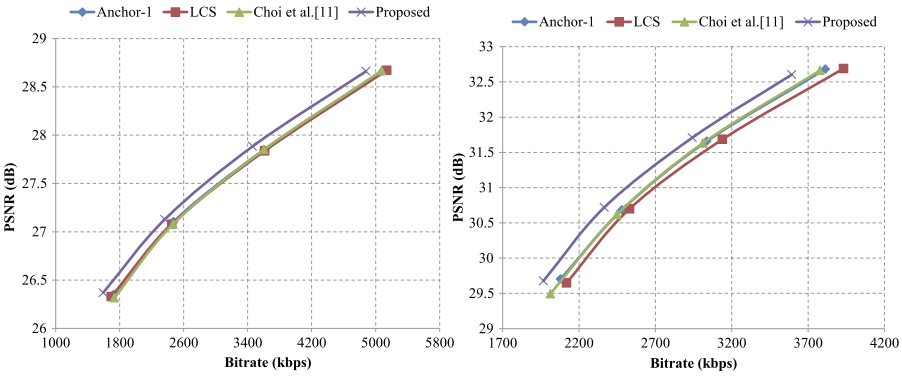

(a) Proportion for H.264/AVC (b) BDPSNR for H.264/AVC

(c) Proportion for HEVC (d) BDPSNR for HEVC

**Fig. 5** Chroma coefficients' proportions and coding gains in H.264/AVC and HEVC simulations

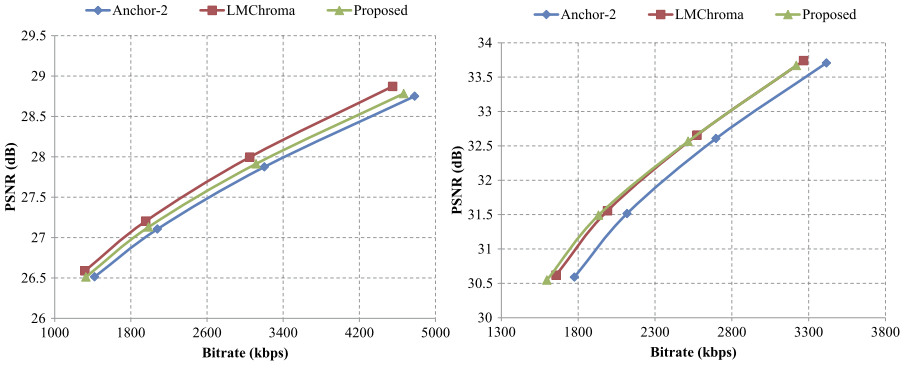
H.264/AVC encoder, which is not completely suitable for the HEVC codec. Second, we only use the luma component to determine the CU depth, which produces a sub-optimal CU coding option relative to the LMChroma scheme. Third, only a simple fixed-length code is employed to encode the prediction value for each superpixel, which limits the coding performance of our proposed method. Although the performance of our proposed method is prior to the state-of-the-art LMChroma scheme, we bring forward a new idea to improve the MB-based coding scheme. Comparing with the conventional MB partition, the superpixel provides us a novel and efficient coding unit in the video coding application. In our future work, the statistic characteristics of the superpixel's prediction value will be studied. Some more efficient variable length coding methods and segmentation parameters settings will be developed for the H.264/AVC and HEVC, respectively.

To visualize the coding gains, we illustrate some sequences' RD curves in both of the H.264/AVC and HEVC simulations. As shown in Fig. 6, the RD-curves of the LCS and line-based intra prediction schemes are always close to the Anchor-1 in the H.264/AVC simulation. However, the RD curves of our proposed scheme are significantly higher than all the other schemes. In addition, as shown in Fig. 7, both the RD curves of the LMChroma and our proposed methods are above the Anchor-2. For the sequence *park_joy*, the LMChroma scheme outperforms our method as shown in the Fig. 7(a). For the sequence *blue_sky*, the RD curves of the LMChroma and the proposed methods are very close as shown in Fig. 7(b).




(a) park_joy (b) blue_sky

**Fig. 6** RD-curves for different sequences in H.264/AVC simulation




(a) park_joy (b) blue_sky

**Fig. 7** RD-curves for different sequences in HEVC simulation

## 4.2 Computational Complexity

In this section, we quantitatively evaluate the complexity of different schemes in terms of execution time. The experiment is implemented on a PC. The system platform is the Intel Xeon Processor of speed 2.3 GHz, 4 GB DDR2 RAM, and Microsoft Windows 7. It should be noted that due to the independence of each coding pass with different segmentation parameters, the complexity of the proposed method can be significantly reduced with parallel operation. Thus, we analyze the execution time of proposed method in two modes. Let $T_{\text{proposed-S}}$ denote the execution time in serial mode, and $T_{\text{proposed-P}}$ denote the execution time in parallel mode. We give their definitions as follows:

$$T_{\text{proposed-S}} = T_{\text{luma}} + T_{\text{chroma}}, \tag{14}$$

$$T_{\text{proposed-P}} = T_{\text{luma}} + \frac{T_{\text{chroma}}}{N_p}, \tag{15}$$

**Table 4** Average encoding time comparison in terms of $\Delta T$ (%) for H.264/AVC simulation

| Sequence | | LCS | Choi et al. [7] | Proposed-S | Proposed-P |
|---|---|---|---|---|---|
| 1080p | pedestrian_area | −61.07 | 1.15 | 826.85 | −25.85 |
| | rush_hour | −61.05 | 1.65 | 822.47 | −26.30 |
| | sunflower | −60.79 | 2.17 | 817.20 | −26.62 |
| | tractor | −61.07 | 2.48 | 813.99 | −26.88 |
| | blue_sky | −60.96 | 2.34 | 838.68 | −24.91 |
| Average 1080p | | −60.99 | 1.96 | 823.84 | −26.11 |
| 720p | ducks_take_off | −66.83 | 4.10 | 784.92 | −29.21 |
| | old_town_cross | −65.97 | 4.50 | 829.81 | −25.62 |
| | park_joy | −66.53 | 4.50 | 796.43 | −28.29 |
| | parkrun | −67.13 | 4.57 | 747.20 | −32.22 |
| | shields | −66.60 | 4.44 | 794.70 | −28.42 |
| | stockholm | −66.48 | 4.60 | 811.95 | −27.04 |
| Average 720p | | −66.59 | 4.45 | 794.17 | −28.47 |
| Average Total | | −64.04 | 3.32 | 807.65 | −27.40 |

**Table 5** Average encoding time comparison in terms of $\Delta T$ (%) for HEVC simulation

| Sequence | | LMChroma | Proposed-S | Proposed-P |
|---|---|---|---|---|
| 1080p | pedestrian_area | 5.01 | 213.98 | −21.50 |
| | rush_hour | 6.38 | 217.92 | −20.52 |
| | sunflower | 5.14 | 204.78 | −23.80 |
| | tractor | 4.17 | 202.16 | −24.46 |
| | blue_sky | 4.47 | 215.40 | −21.15 |
| Average 1080p | | 5.03 | 210.85 | −22.29 |
| 720p | ducks_take_off | 4.02 | 185.20 | −28.70 |
| | old_town_cross | 4.50 | 203.75 | −24.06 |
| | park_joy | 4.75 | 202.53 | −24.37 |
| | parkrun | 4.57 | 205.95 | −23.51 |
| | shields | 5.20 | 207.15 | −23.21 |
| | stockholm | 4.74 | 208.83 | −22.79 |
| Average 720p | | 4.63 | 202.24 | −24.44 |
| Average Total | | 4.81 | 206.15 | −23.46 |

where $T_{\text{luma}}$ and $T_{\text{chroma}}$ denote the execution time of our proposed method for luma and chroma coding pass, respectively. $N_p$ is the number of candidate segmentation parameters.

The detailed comparison results for the H.264/AVC and HEVC simulations are shown in Table 4 and Table 5, respectively. It can be seen that for both the H.264/AVC

and HEVC simulations, the proposed-S scheme induces a high complexity increase due to the multiple-parameter segmentation operation. Since the computational complexity of H.264/AVC codec is much lower than the HEVC one, this increase is more significant in the H.264/AVC simulation. On average, the execution time increase of the proposed-S scheme is up to 807.65 % and 206.15 % for the H.264/AVC and HEVC simulations, respectively. In contrast, the proposed-P scheme even reduces the execution time in both simulations. To explain this result, we analyze the coding process for the H.264/AVC and HEVC encoders, respectively.

In the H.264/AVC encoder, the intra prediction modes for the luma and chroma components are determined by exploring all available intra mode combinations. Let $M$ denote the number of intra modes for the luma, and $N$ denote the intra modes number for the chroma. The candidate intra mode combinations are up to $M \times N$. In the LCS scheme, the luma and chroma components are separated into two independent coding passes. There are only $M + N$ candidate intra mode combinations left, which significantly reduces the number of RDO loops. Thus, the LCS scheme can save 64.04 % execution time on average. In the proposed-P scheme, the intra mode decision for the luma and chroma is also independent. Except for exploring $M$ candidate intra modes for luma component, an additional segmentation operation is needed for the chroma component. With the help of the parallel implementation, only one group of segmentation parameters are explored in the chroma coding pass each time. Thus, the proposed-P scheme also achieves 27.40 % time saving on average. For the method proposed in [7], there is no change for the number of the coding pass. Thus, its execution time is similar with the anchor scheme, whose $\Delta T$ is 3.32 % on average. Unlike the H.264/AVC encoder, the complexity of the HEVC codec mainly comes from the recursive coding unit (CU) depth decision operation. Although the luma and chroma intra mode decision is also independent in HEVC, this decision operation needs to be recursively implemented in each available CU depth, which significantly increases the number of RDO loops. In the proposed-P scheme, the optimal CU depth is only determined by the rate-distortion cost of the luma component, where the chroma component is not included in the recursive process. The CU depth option is directly used for the following chroma coding, which significantly reduces the number of RDO loops. Although the additional segmentation operation induces a little complexity increase in the parallel mode, we can still save 23.46 % execution time on average. For the LMChroma [4] scheme, an additional LM chroma prediction mode is included in the intra mode decision process. Thus, its execution time increases 4.81 % on average.

## 5 Conclusion

In this paper, we propose a segmentation-based chroma intra prediction coding scheme. By exploiting the spatial correlation between the luma and chroma images, the segmentation result derived from the reconstructed luma image is transferred to the chroma image. Since the segmentation-based areas partition method produces more homogenous local image distribution in each superpixel, the proposed method achieves better coding efficiency for the chroma intra coding. By incorporating the segmentation parameters determining process into the RDO framework with

the chroma coding pass, the robustness of the proposed method is further improved. Although a significant complexity increase is induced by our proposed method, it can be efficiently solved by parallel operation. Experimental results show that the proposed method achieves a superior performance relative to the H.264/AVC and HEVC encoders at low-bitrate video coding. An average, 0.20 dB and 0.18 dB PSNR improvements have been achieved for the H.264/AVC and HEVC, respectively.

In our future work, a more flexible coding scheme corresponding to the superpixel partition structure will be studied. A more elaborate prediction method for this irregular partition will also be researched.

**Acknowledgements** This work was partially supported by NSFC (Nos. 61179060 and 61101091), National High Technology Research and Development Program of China (863 Program, No. 2012AA011503), and Fundamental Research Funds for the Central Universities (ZYGX2012YB007).

## References

1. R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, S. Susstrunk, SLIC superpixels compared to state-of-the-art superpixel methods. IEEE Trans. Pattern Anal. Mach. Intell. **34**(11), 2274–2282 (2012)
2. G. Bjontegaard, Calculation of average PSNR differences between RD-curves, in *ITU-T VCEG*, (2001). VCEG-M33
3. F. Bossen, Common test conditions and software reference configurations, in *JCT-VC Meeting*, Torino (2011). JCTVC-F900
4. J. Chen, V. Seregin, W.J. Han, J. Kim, B. Jeon, CE6.a. 4: chroma intra prediction by reconstructed luma samples, in *JCT-VC Meeting*, Geneva (2011). JCTVC-E266
5. I.H. Cho, J.H. Lee, W.H. Lee, D.S. Jeong, New intra luma prediction mode in H.264/AVC using collocated weighted chroma pixel value, in *Advanced Concepts for Intelligent Vision Systems*, vol. 4179, (2006), pp. 344–353
6. J.A. Choi, Y.S. Ho, Line-by-line intra 16 × 16 prediction for high-quality video coding, in *IEEE International Conference on Multimedia and Expo* (2010), pp. 1281–1286
7. J.A. Choi, Y.S. Ho, Implicit line-based intra 16 × 16 prediction for H.264/AVC high-quality video coding. Circuits Syst. Signal Process. **31**(5), 1829–1845 (2012)
8. O. Divorra Escoda, P. Yin, C. Dai, X. Li, Geometry-adaptive block partitioning for video coding, in *IEEE International Conference on Acoustics, Speech and Signal Processing*, vol. 1 (2007), pp. 657–660
9. P.L. Dragotti, M.N. Do, R. Shukla, M. Vetterli, On the compression of two-dimensional piecewise smooth functions, in *IEEE International Conference on Image Processing* (2001), pp. 14–17
10. N.C. Francisco, N.M.M. Rodrigues, E.A.B. da Silva, M.B. de Carvalho, S.M.M. de Faria, V.M.M. da Silva, M.J.C.S. Reis, Multiscale recurrent pattern image coding with a flexible partition scheme, in *IEEE International Conference on Image Processing* (2008), pp. 141–144
11. B. Fulkerson, A. Vedaldi, S. Soatto, Class segmentation and object localization with superpixel neighborhoods, in *IEEE International Conference on Computer Vision* (2009), pp. 670–677
12. HEVC Model 4.1 (2011). https://hevc.hhi.fraunhofer.de/svn/svn_HEVCSoftware/branches/HM-4.1-dev/
13. ITU-T VCEG KTA Reference Software (2011). http://iphome.hhi.de/suehring/tml/download/KTA
14. A. Kassim, L. Siong, Performance of the color set partitioning in hierarchical tree scheme (CSPIHT) in video coding. Circuits Syst. Signal Process. **20**, 253–270 (2001)
15. A. Kassim, E. Tan, W. Lee, 3D color set partitioning in hierarchical trees. Circuits Syst. Signal Process. **28**, 41–53 (2009)
16. B. Li, G.J. Sullivan, J. Xu, Common test conditions and software reference configurations, in *JCT-VC Meeting*, Torino (2011). JCTVC-F900
17. H. Li, K.N. Ngan, Saliency model based face segmentation in head-and-shoulder video sequences. J. Vis. Commun. Image Represent. **19**(5), 320–333 (2008)
18. H. Li, K.N. Ngan, A co-saliency model of image pairs. IEEE Trans. Image Process. **20**(12), 3365–3375 (2011)

19. H. Li, K.N. Ngan, Q. Liu, FaceSeg: automatic face segmentation for real-time video. IEEE Trans. Multimed. **11**(1), 77–88 (2009)
20. H. Li, K.N. Ngan, Z. Wei, Fast and efficient method for block edge classification and its application in H.264/AVC video coding. IEEE Trans. Circuits Syst. Video Technol. **18**(6), 756–768 (2008)
21. D. Liu, X. Sun, F. Wu, Y.Q. Zhang, Edge-oriented uniform intra prediction. IEEE Trans. Image Process. **17**(10), 1827–1836 (2008)
22. M. Mahoney, Data Compression Programs (2007)
23. F. Meng, H. Li, G. Liu, K.N. Ngan, Object co-segmentation based on shortest path algorithm and saliency model. IEEE Trans. Multimed. **14**(5), 1429–1441 (2012)
24. F. Meng, H. Li, G. Liu, K.N. Ngan, Image cosegmentation by incorporating color reward strategy and active contour model. IEEE Trans. Syst. Man Cybern. **43**(2), 725–737 (2013)
25. A. Moore, S. Prince, J. Warrell, U. Mohammed, G. Jones, Superpixel lattices, in *IEEE Conference on Computer Vision and Pattern Recognition* (2008), pp. 1–8
26. J. Ohm, G. Sullivan, H. Schwarz, T.K. Tan, T. Wiegand, Comparison of the coding efficiency of video coding standards—including high efficiency video coding (hevc). IEEE Trans. Circuits Syst. Video Technol. **22**(12), 1669–1684 (2012)
27. M. Ouaret, F. Dufaux, T. Ebrahimi, On comparing JPEG2000 and intraframe AVC, in *SPIE Applications of Digital Image Processing XXIX*, vol. 6312 (2006), p. U3120
28. Y. Piao, H. Park, Adaptive interpolation-based divide-and-predict intra coding for h.264/avc, in *IEEE Trans. Circuits Syst. Video Technol.* (2010), pp. 1915–1921
29. X. Ren, J. Malik, Learning a classification model for segmentation, in *IEEE International Conference on Computer Vision* (2003), pp. 10–17
30. I.E. Richardson, *The H.264 Advanced Video Compression Standard*, 2nd edn. (Wiley, New York, 2010)
31. T.K. Tan, C.S. Boon, Y. Suzuki, Intra prediction by template matching, in *IEEE International Conference on Image Processing* (2006), pp. 1693–1696
32. T.K. Tan, G. Sullivan, T. Wedi, Recommended simulation common conditions for coding efficiency experiments rev. 1, in *ITU-T Q.6/SG16*, Marrakech, Morocco (2007). VCEG-AE010
33. ITU-T Recommendation H.264 and ISO/IEC 14496-10 (MPEG-4) AVC, Advanced Video Coding for Generic Audiovisual Services (2005)
34. P. Topiwala, T. Tran, W. Dai, Performance comparison of jpeg2000 and h. 264/avc high profile intra-frame coding on hd video sequences, in *SPIE Applications of Digital Image Processing XXIX*, vol. 6312 (2006), p. T3120
35. Z. Wei, K.N. Ngan, H. Li, An efficient intra mode selection algorithm for H.264 based on edge classification and rate-distortion estimation. Signal Process. Image Commun. **23**(9), 699–710 (2008)
36. T. Wiegand, B. Girod, Lagrange multiplier selection in hybrid video coder control, in *IEEE International Conference on Image Processing*, vol. 3 (2001), pp. 542–545
37. T. Wiegand, G. Sullivan, G. Bjontegaard, A. Luthra, Overview of the H.264/AVC video coding standard. IEEE Trans. Circuits Syst. Video Technol. **13**(7), 560–576 (2003)
38. Q. Wu, H. Li, Mode dependent down-sampling and interpolation scheme for high efficiency video coding. Signal Process. Image Commun. **28**(6), 581–596 (2013)
39. Y. Ye, M. Karczewicz, Improved H.264 intra coding based on bi-directional intra prediction, directional transform, and adaptive coefficient scanning, in *IEEE International Conference on Image Processing* (2008), pp. 2116–2119
40. C. Yeo, Y.H. Tan, Z. Li, S. Rahardja, Chroma intra prediction using template matching with reconstructed luma components, in *IEEE International Conference on Image Processing* (2011), pp. 1637–1640
41. L. Zhang, S. Ma, W. Gao, Position dependent linear intra prediction for image coding, in *IEEE International Conference on Image Processing* (2010), pp. 2877–2880