Joint Chroma Downsampling and Upsampling for Screen Content Image

Shiqi Wang, *Member, IEEE*, Ke Gu, Siwei Ma, *Member, IEEE*, and Wen Gao, *Fellow, IEEE*

***Abstract*— Screen content images are originally captured in a full-chroma format. The chroma downsampling, which is commonly applied to the chroma component in screen content image representation and processing (e.g., YUV4:2:0 compression), will significantly degrade the image quality and create annoying artifacts such as blur and color shifting. To tackle this problem, in this paper we propose luma aware chroma downsampling and upsampling algorithms to jointly improve the quality of the chroma image reconstruction. Guided by the luma information, the chroma upsampling algorithm is proposed with the utilization of major color and index map representation. The geometric information-based linear mapping is developed to transfer the structure of luma to the interpolated chroma. Subsequently, the error sensitivity of the upsampling method is analyzed, and content dependent downsampling algorithm is presented to minimize the error sensitivity function. We further explore the applicability of the proposed scheme in the scenario of screen content compression, targeting at improving the decoded chroma image quality for display. Extensive experimental results demonstrate the viability and efficiency of the proposed scheme.**

***Index Terms*— Chroma component, downsampling, screen content compression, screen content image, upsampling.**

I. Introduction

RECENTLY, there is a striking rise in the popularity of the screen virtualization. A variety of remote processing and virtual desktop applications emerge with the purpose of accessing and controlling the remote data and computational resources through the network, such as cloud–mobile convergence [1], cloud gaming [2], [3], remote computing platform [4], and remote desktop sharing system [5], [6]. In these applications, the updated screen is generated, compressed, and transmitted to the display side. The screen virtualization can be realized by the users' interaction with the local display interface, which is a typical computer generated screen content image.

The quality of screen content images directly determines the interactivity performance and user experience of the remote system. In general, the human visual system (HVS) is much more sensitive to the variations in luma than chroma, and the compression performance can be improved by downsampling the chroma to reduce the consumed bits. This introduces the YUV4:2:0 codecs that are commonly used in current image and video systems. However, in addition to blur, it is realized that chroma downsampling for screen content may create annoying artifacts such as color shifting [7]. This is because of the anisotropic features in textual content [8]. Being aware of this problem, full chroma format-based screen content compression is taken into account in the development of High Efficiency Video Coding (HEVC) range extension [9]. Nevertheless, it is usually inefficient to compress the entire image with full chroma [4], as natural image content is usually involved in the captured screen content image. To tackle this dilemma, mixed format compression algorithms are proposed in [7], [10], and [11], where the pure screen content is compressed with YUV4:4:4 format and meanwhile the natural image content is compressed with YUV4:2:0 format. However, the requirements of standard and backward-compatibilities in real applications may not be satisfied by simply combining two chroma formats together. To support YUV4:2:0 format representation in these applications, such as screen virtualization and remote control, high efficiency chroma upsampling algorithm is an urgent need.

Recently, spatial adaptive upsampling algorithms have been intensively studied in the literature, such as the new edge-directed interpolation method [12], regularized local linear regression [13], joint bilateral upsampling [14], and guided filter [15]. However, to the best of the author's knowledge, few of them are specifically designed for the upsampling of screen content. The screen content images may not always share the same properties of natural images. For example, the discontinuous-tone computer generated textual content features sharp edges and thin lines with few colors, while the natural images usually have continuous-tone content with smooth edges, thick lines, and more colors [8], [10].

In general, there is a high correlation among the color channels. To reduce the inter-channel redundancy, inter-channel prediction techniques have been intensively studied in video compression. For example, in [16] and [17], the blue and red channels are linearly predicted from the colocated green channel. Alternatively, this method can be applied from blue or red to the other two channels as well. In the luma–chroma spaces, efforts have also been devoted to

Manuscript received April 14, 2015; accepted July 25, 2015. Date of publication July 29, 2015; date of current version September 1, 2016. This work was supported in part by the National Basic Research Program of China (973 Program) under Grant 2015CB351800; in part by the National Science Foundation of China under Grant 61322106 and Grant 61421062; and in part by the Cooperative Medianet Innovation Center. This paper was recommended by Associate Editor P. L. Callet.

S. Wang, S. Ma, and W. Gao are with the Institute of Digital Media, School of Electronic Engineering and Computer Science, Peking University, Beijing 100871, China (e-mail: sqwang1986@gmail.com; swma@pku.edu.cn; wgao@pku.edu.cn).

K. Gu is with the Institute of Image Communication and Information Processing, Shanghai Jiao Tong University, Shanghai 200240, China (e-mail: gukesjtuee@gmail.com).

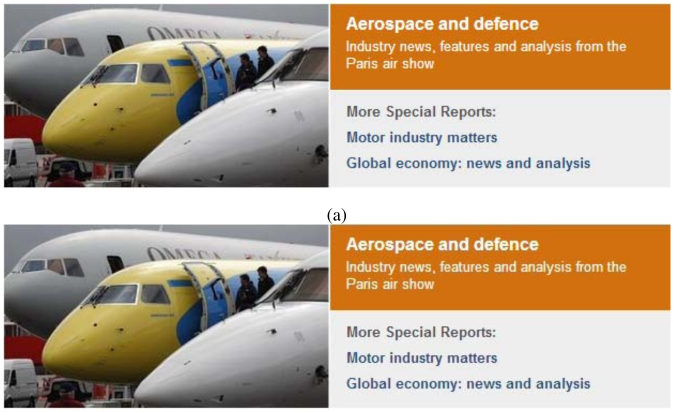
Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TCSVT.2015.2461891

developing linear models from luma to chroma. In [18], the reconstructed luma blocks are used to predict the chroma samples for YUV4:2:0 coding. In [19], an effective intra-coding mode called linear mode is developed to improve the chroma coding performance in HEVC. In inter-channel prediction, the most important issue is how to establish the relationship between luma and chroma and derive the parameters of the regression model. This process should be highly adaptive to the luma content, which well preserves the image structure. Inspired by this, in this paper, we target at developing luma guided downsampling and upsampling algorithms for screen content image chroma format conversion. The unique characteristics such as limited colors and sharp edge in screen content are employed in the adaptive sampling process. The contributions of our work are as follows.

1) A luma information guided chroma upsampling algorithm is proposed. The local major colors are extracted from luma component to exploit the structure geometric mapping between luma and chroma. The low complexity local linear transform is further employed to derive the chroma components, which produces more structured chroma information.
2) An adaptive downsampling algorithm is devised to deliver more information in the coarse resolution chroma image. The luma pixels that are involved in the upsampling process are identified to extract the base colors. Based on the analysis of the error sensitivity function, the downsampled chroma pixels are derived by minimizing the differences between the interpolated pixels and the filtering output of chroma image.
3) With the proposed downsampling and upsampling algorithms, high efficiency screen content image compression scheme that supports YUV4:2:0 format codec is proposed. The advantages of the proposed scheme are that it achieves standard and backward compatibilities with the existing natural image/video codecs as well as maintains high quality chroma reconstruction.

The remainder of this paper is organized as follows. In Section II, we analyze and demonstrate the artifacts created by color downsampling. Section III elaborates the luma guided chroma upsampling and downsampling algorithms, and the complexity analysis is provided as well. Section IV demonstrates the high efficiency screen content image compression scheme that supports YUV4:2:0 format codec. Experimental results are presented in Section V. Finally, this paper is concluded in Section VI.

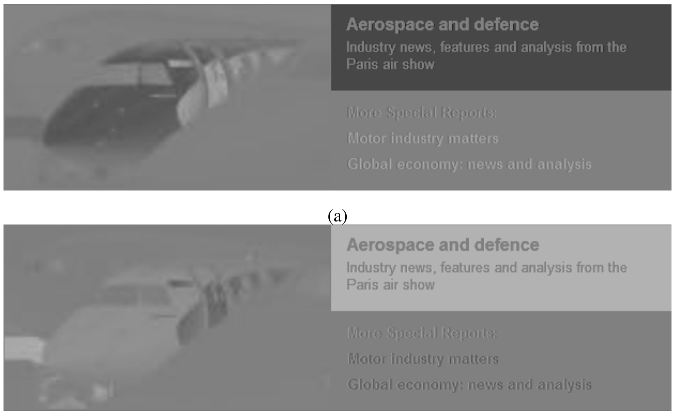

Aerospace and defence
Industry news, features and analysis from the Paris air show
More Special Reports:
Motor industry matters
Global economy: news and analysis

(a)

Aerospace and defence
Industry news, features and analysis from the Paris air show
More Special Reports:
Motor industry matters
Global economy: news and analysis

(b)

Fig. 1. Visualization of the artifacts created by chroma downsampling. (a) Raw image (original RGB input). (b) YUV4:2:0 image.


Aerospace and defence
Industry news, features and analysis from the Paris air show
More Special Reports:
Motor industry matters
Global economy: news and analysis

(a)

Aerospace and defence
Industry news, features and analysis from the Paris air show
More Special Reports:
Motor industry matters
Global economy: news and analysis

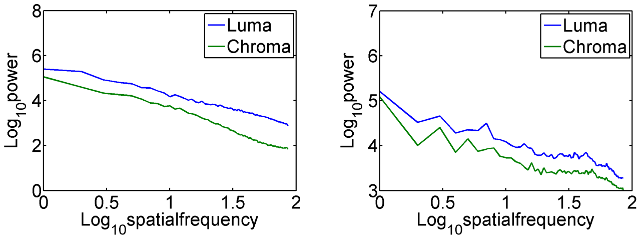
(b)

Fig. 2. Illustration of the chroma components. (a) Cb component. (b) Cr component.

II. Analysis on Screen Chroma Downsampling

For natural images, the pixel values in chroma channel are usually much smoother than that in the luma channel. In contrast, due to the discontinuous-tone features of the screen content, sharp edges often exist in chroma channel of the textual regions. Therefore, the natural image can be converted into YUV4:2:0 format for compression, which not only maintains the visual quality but also achieves high compression performance. However, for screen content, the visual quality may be distorted after chroma downsampling because of the obvious visual artifacts such as blurring and color shifting. In Fig. 1, we provide an example showing the effect of chroma downsampling on the textual and natural image contents, respectively. The YUV4:2:0 image is generated by first downsampling the chroma components with an average filter that computes the mean value of the four corresponding samples, and then performing bilinear method on the downsampled chroma components for interpolation. It can be observed that the color of the textual content is significantly degraded and obvious artifacts are observed, while little visual quality loss is observed in the natural image. To further investigate the cause of the visual artifacts created by chroma downsampling, chroma components in Fig. 1 are extracted and demonstrated in Fig. 2. It is observed that the textual structure is well kept in the chroma component, while the chroma content in natural images is smoother and less structured.

From natural scene statistics, the amplitude spectrum that can be characterized by the amplitude after the discrete Fourier transform of natural images falls with the spatial frequency approximately proportional to $1/f^p$ [20]–[22], where f is the spatial frequency and p is an image dependent constant. The textual and natural subimages in Fig. 1 are decomposed



(a) (b)

Fig. 3. Amplitude spectrum falloffs for natural and textual content. (a) Natural subimage in Fig. 1(a). (b) Textual subimage in Fig. 1(a).

with Fourier transform and the amplitude spectrum is then computed, respectively, as demonstrated in Fig. 3. It is observed that the falloffs for natural images are approximately straight lines in log–log scale, which is consistent with the $1/f^p$ relationship. However, for the textual image, this law is no longer obeyed for both luma and chroma. Moreover, the energy of textual content at high frequency is larger than that of natural image content, especially for the chroma component, which verifies the observation that chroma content in textual content is sharper and more structured.

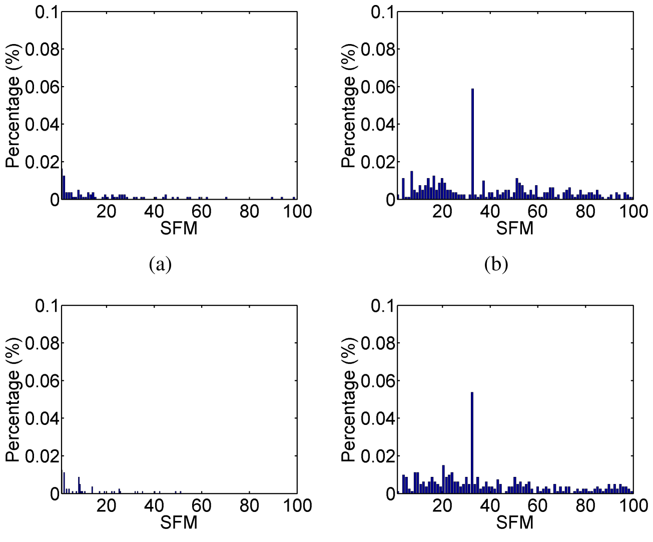
We further examine the chroma component by the block-wise spatial frequency measure (SFM) [23], which is defined as

$$\begin{aligned} \text{SFM} &= \sqrt{G_H^2 + G_V^2} \\ G_H &= \sqrt{\frac{1}{m \times n} \sum_{i=1}^{m-1} \sum_{j=1}^{n-1} \left(\tilde{C}_{i,j} - \tilde{C}_{i-1,j}\right)^2} \\ G_V &= \sqrt{\frac{1}{m \times n} \sum_{i=1}^{m-1} \sum_{j=1}^{n-1} \left(\tilde{C}_{i,j} - \tilde{C}_{i,j-1}\right)^2} \end{aligned} \tag{1}$$

where $\tilde{C}_{i,j}$ indicates the chroma pixel value in the block with size $m \times n$. In Fig. 4, the SFM distributions in two chroma channels of the natural and textual subimages in Fig. 2 are illustrated (the whole left and right subimages are treated as natural and textual images, respectively). The probability of zero bin is ignored for demonstration. It is observed that the SFM values concentrate on low values for natural image, while it has a scattered distribution at high values for textual content. The high gradient pixels appear at edge boundaries of textual content. As the HVS is more sensitive to the artifacts around edge boundaries, after chroma upsampling, these artifacts can be easily detected and cause the visual experience degradation.

III. Chroma Upsampling and Downsampling

In this section, we first describe the upsampling method with the utilization of base color and index map (BCIM) representation. Subsequently, the downsampling algorithm based on the upsampling process is detailed. Finally, the computational complexity is analyzed for both upsampling and downsampling. In the following description, only one chroma channel is considered. Without loss of generality, the algorithm is also valid for the other one.



(a) (b) (c) (d)

Fig. 4. Histograms of chroma SFM on natural and textual images. (The first bin is ignored for demonstration.) (a) and (c) SFM of Cb and Cr components for natural content. (b) and (d) SFM of Cb and Cr components for textual content.

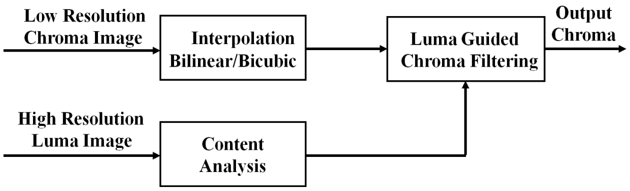

Low Resolution Chroma Image
Interpolation Bilinear/Bicubic
Luma Guided Chroma Filtering
Output Chroma
High Resolution Luma Image
Content Analysis

Fig. 5. Flowchart of the proposed chroma upsampling.

A. Chroma Upsampling

The proposed upsampling process is illustrated in Fig. 5, which includes chroma interpolation, content analysis, and luma guided chroma filtering. Considering real-time application scenarios, simple interpolation method such as bicubic or bilinear is first employed to generate the initial full resolution chroma image. Low complexity content analysis is then conducted to identify the high gradient screen content and extract major colors. Following this process, efficient filtering that linearly transforms the luma samples to chroma components is finally performed.

The block-wise content analysis and chroma filtering is performed by dividing the image into nonoverlapping blocks. The default block size is 16 × 16. To ensure low complexity upsampling, there is no need to perform luma guided filtering on the natural content. Therefore, we classify the block type by the number of high gradient pixels [4], [11]

$$N_S = \sum_{i=1}^{m-1} \sum_{j=1}^{n-1} u(|I_{i,j} - I_{i-1,j}| > I_{\text{th}} \vee |I_{i,j} - I_{i,j-1}| > I_{\text{th}}) \tag{2}$$

where I_{th} denotes the predefined threshold and $I_{i,j}$ denotes the luma samples at the location (i, j). The symbol $\vee$ denotes the logical or operation. The function u equals to one when

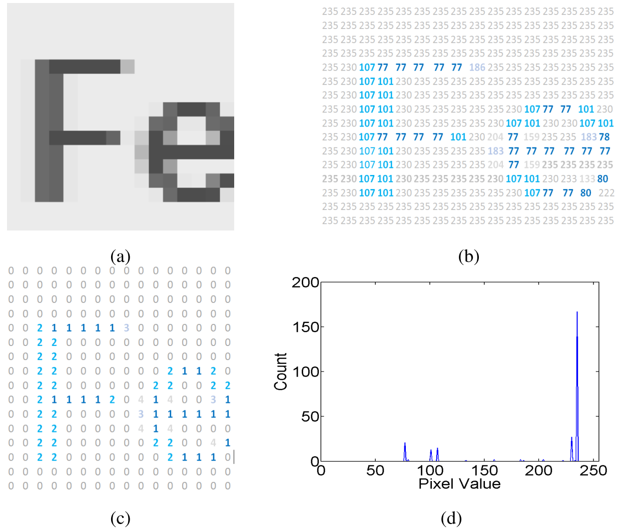


Fig. 6. BCIM representation. (a) Amplified luma block. (b) Luma pixel values. (c) Major color index map. (d) Pixel value histogram.

the input logical value is true. In general, there is a high probability that blocks with many high gradient pixels belong to textual content. As such, if the number N_S exceeds a certain threshold N_{ST}, the current block is identified to be a high gradient block that is subject to further processing. The default values of I_{th} and N_{ST} are set to be 32 and 3, respectively.

Due to the features of sharp edge and limit colors in textual content, the BCIM representation has been a powerful tool in screen content compression and various coding algorithms based on BCIM were proposed in [4], [8], [11], [24], and [25]. In this paper, this method is employed to establish the geometric structure mapping between luma and chroma components. One example of high gradient textual block and its corresponding luma pixel histogram are shown in Fig. 6, and it is observed that the block only contains limited number of sample values. In general, algorithms such as k-means and vector quantization can be used to extract the base colors. However, high computational complexity will be involved when applying these methods. To obtain the base colors with low complexity, the pixel value histogram is established and consequently a majority of pixels are convergent to a small number of colors, as illustrated in Fig. 6(d). In this paper, the number of base colors is limited to four. These colors are extracted as base colors and equal size windows are used to range the extracted major colors. Colors that cannot be represented by base colors are identified as escape colors. The corresponding index map in Fig. 6(c) represents the geometric information extracted from the block, and different colors indicate different indices. This method has been proven useful in BCIM-based screen content compression [4], [11].

The corresponding chroma component of the block in Fig. 6(a) is shown in Fig. 7, and it is observed that there is a high correlation between the luma and chroma structure. The relationship between luma and chroma of two major colors is further demonstrated in Fig. 8. The corresponding chroma pixels converge to two points, indicating that there almost exists a perfect linear correlation between two base colors.

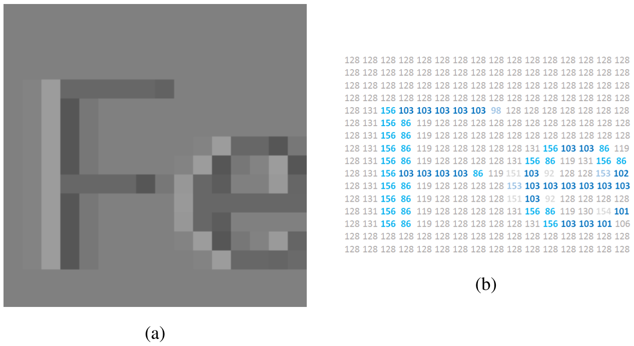


Fig. 7. Exemplified chroma content. (a) Amplified chroma block. (b) Pixel values.

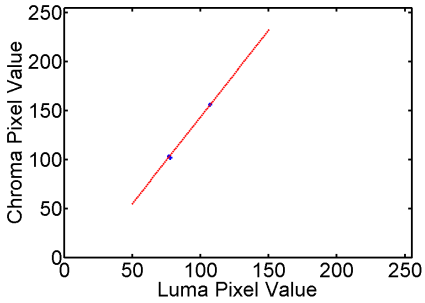


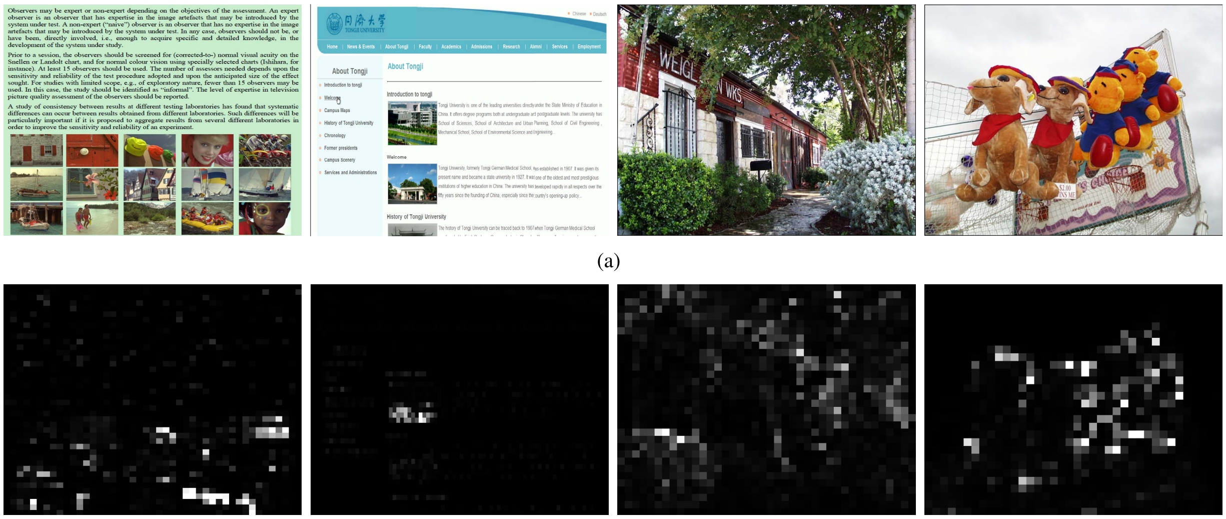
Fig. 8. Relationship between luma and chroma for two major colors.

This motivated us to apply the luma index map to identify the major colors in the interpolated chroma component. Subsequently, BCIM-based linear transform is performed after the major color's identification. To achieve a good balance between complexity and performance, the four most important base colors that compose the major part of blocks are extracted and divided into two pixel sets according to the number of pixels corresponding to each base color. Specifically, the base colors are ranked by the number of corresponding pixels. The first pixel set corresponds to the two most important colors and the second pixel set corresponds to the rest two. Under special circumstances, there may exist only three base colors in a block, and the second set is formed by the last two base colors. In this case, the two pixel sets may include duplicated pixels. As such, in the second pixel set reconstruction, only the chroma samples that belong to the third base color are reconstructed.

Assuming the ith pixel set that corresponds to the two neighboring base colors within one block is s_i, to derive the chroma component, the linear transform that establishes the relationship between luma and chroma is formulated as follows:

$$P_k = \alpha_i \cdot I_k + \beta_i, \quad k \in s_i \tag{3}$$

where I_k indicates the luma pixels in the set s_i and P_k represents the filter output. This linear model ensures that the chroma edge corresponds to the luma edge, and has been successfully applied in image filtering [15] and matting [26]. Parameters α_i and β_i are the coefficients that are derived locally from the I_k and the initially interpolated chroma



(a)

(b)

Fig. 9. Demonstration of the error sensitivity e_l' computed for both natural and screen content images. (a) Test images. (b) Error sensitivity maps (scaled for visualization, light/dark regions represent high/low error values).

value C_k, which can be easily computed by minimizing the mean squared error between the transformed sample of I_k and C_k according to

$$E_{rr} = \sum_{k \in s_i} (\alpha_i \cdot I_k + \beta_i - C_k)^2. \tag{4}$$

Finally, the parameters are calculated as

$$\begin{aligned} \alpha_i &= \frac{\frac{1}{|s_i|}\sum_{k \in s_i} I_k C_k - \mu_I \mu_C}{\sigma_I^2} \\ \beta_i &= \mu_C - \alpha_i \mu_I \end{aligned} \tag{5}$$

where μ_I and μ_C indicate the mean of I and C in the current set, and σ_I^2 represents the variance of luma pixel values.

Furthermore, the error sensitivity of the proposed model is analyzed to show the benefits of BCIM-based linear transform. Let $\tilde{C}_k$ denote the original full-chroma component and parameters α_i' and β_i' denote the coefficients derived from the full luma and chroma using (5). Subsequently, we compute the error sensitivity function of the luma–chroma linear transform for both screen content and natural images

$$e_l' = E\big[\big(\tilde{C}_k - \alpha_i' \cdot I_k - \beta_i'\big)^2\big]. \tag{6}$$

For each block, the error sensitivity maps that are computed within the two major colors that take the largest proportion of the total pixels are demonstrated in Fig. 9, where we can observe that the e_l' of natural images is relatively larger than that of screen content images. Moreover, within a screen content image, the error sensitivity is higher in natural regions. This further provides useful evidence that there is a high correlation between luma and chroma for textual content and it is therefore effective to transfer the structure of luma to chroma via linear mapping.

The error sensitivity function in (6) can also be written as

$$\begin{aligned} e_l' &= E\big[\big(\tilde{C}_k - \alpha_i' \cdot I_k - \beta_i'\big)^2\big] \\ &= E\big[\big(\tilde{C}_k - \alpha_i' \cdot I_k - \big(\mu_{\tilde{C}} - \alpha_i' \mu_I\big)\big)^2\big] \\ &= E\big[\big(\big(\tilde{C}_k - \mu_{\tilde{C}}\big) - \alpha_i' \cdot (I_k - \mu_I)\big)^2\big] \\ &= \sigma_{\tilde{C}}^2 + \alpha_i'^2 \sigma_I^2 - 2\alpha_i' \rho_{I\tilde{C}} \sigma_{\tilde{C}} \sigma_I \end{aligned} \tag{7}$$

where $\rho_{I\tilde{C}}$ denotes the correlation coefficient between the input chroma and luma components.

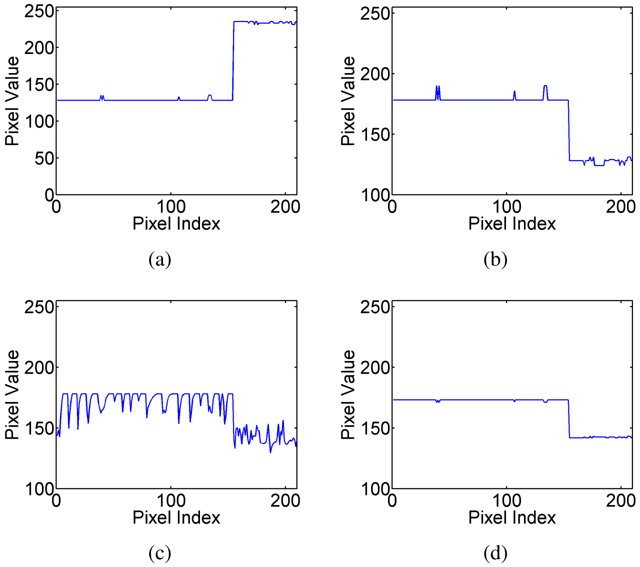
From (5), we have

$$\alpha_i' = \rho_{I\tilde{C}} \cdot \frac{\sigma_{\tilde{C}}}{\sigma_I}. \tag{8}$$

With (7) and (8), the error sensitivity function is given by

$$\begin{aligned} e_l' &= \sigma_{\tilde{C}}^2 + \alpha_i'^2 \sigma_I^2 - 2\alpha_i' \rho_{I\tilde{C}} \sigma_{\tilde{C}} \sigma_I \\ &= \sigma_{\tilde{C}}^2 \big(1 - \rho_{I\tilde{C}}^2\big). \end{aligned} \tag{9}$$

This implies that the higher correlation between the two channels, the better performance we can achieve with the proposed model. This explains why we employ the index map to exploit the geometric relationship between luma and chroma. With the BCIM representation, the geometric mapping between luma and chroma can be easily obtained and the structure from luma content can be efficiently transformed to chroma. As only two base colors are employed in each optimization, this model ensures that the chroma edge exactly matches the luma edge, so as to eliminate the color shifting and blurring artifacts. The illustration of the filtering process is shown in Fig. 10, where the step edge due to the abrupt pixel value change on the text boundary is demonstrated. It is observed that the advantages of the BCIM-based linear transform are twofold. Within each major color, the chroma component is smoothen to unify the color. In contrast, near edge boundaries, the chroma step edge is further sharpen to eliminate the blurring artifacts.



(a) (b) (c) (d)

Fig. 10. Illustration of the filtering process. (a) Luma samples. (b) Original chroma samples. (c) Initially interpolated chroma samples. (d) Filtering output.

It is noted that the minimal error can only be achieved when α_i and β_i derived from the upsampled chroma exactly equal to the parameters computed from the input chroma. This is impossible in general. In the following section, we will show how to minimize the error sensitivity function with adaptive downsampling.

The proposed scheme is applicable to any codec as a postprocessing component. Moreover, it can be well incorporated into the palette mode-based screen coding techniques. In this way, the BCIM information can be directly obtained from the decoding information, which not only speeds up the upsampling process but also maintains accurate structure mapping. Application of the proposed scheme in YUV4:2:0 coding is discussed in Section IV.

B. Adaptive Downsampling

The downsampling process gives rise to the loss of information. Therefore, the low-resolution image should be able to imply as much information as possible. Interpolation-based image downsampling was proposed in [27], in which the downsampled pixels are calculated based on the interpolation methods. In this paper, we further study the adaptive downsampling with the combination of the proposed BCIM-based upsampling scheme.

The aim of the adaptive downsampling is to make use of the interpolated chroma and full resolution luma to derive more accurate parameters α and β. In general, the more accurate these coefficients are, the better upsampling quality will be achieved. Assuming the zero mean difference between the input chroma and upsampled chroma is n_k

$$n_k = C_k - \tilde{C}_k \tag{10}$$

and we will have

$$\begin{aligned} \mu_C &= \mu_{\tilde{C}} \\ \rho_{IC} &= \rho_{I\tilde{C}} \cdot \frac{\sigma_{\tilde{C}}}{\sigma_C} \\ \alpha_i &= \rho_{IC} \cdot \frac{\sigma_C}{\sigma_I}. \end{aligned} \tag{11}$$

Therefore, when considering the interpolation chroma in calculating the parameters α_i and β_i, the error sensitivity function in (7) is reformulated as

$$\begin{aligned} e_l &= E[(\tilde{C}_k - \alpha_i \cdot I_k - \beta_i)^2] \\ &= E\big[\big((\tilde{C}_k - \mu_C) - \alpha_i \cdot (I_k - \mu_I)\big)^2\big] \\ &= \sigma^2_{\tilde{C}} + \alpha_i^2 \sigma_I^2 - 2\alpha_i \rho_{IC} \sigma_C \sigma_I \\ &= e'_l. \end{aligned} \tag{12}$$

This implies that with the zero mean assumption, the error sensitivity derived from the interpolated chroma is equal to the input chroma. Moreover, it is also observed that the downsampling error increases monotonously with the local variance. Therefore, the optimization objective in adaptive downsampling should be well structured with less noise. To approach this, luma guided chroma filtering is performed before downsampling. More specifically, the coefficients α'_i and β'_i in the local set s_i are first computed with the input luma and chroma, and given the input chroma block $\tilde{C}$, the filter output is obtained by linearly transforming the corresponding luma pixels to chroma, resulting in the filtering output $\tilde{C}'$

$$\tilde{C}'_k = \alpha'_i \cdot I_k + \beta'_i. \tag{13}$$

In set s_i, the difference between $\tilde{C}$ and $\tilde{C}'$ is calculated as

$$\begin{aligned} E_{\tilde{C}\tilde{C}'} &= E\big[(\tilde{C}_k - \alpha'_i \cdot I_k - \beta'_i)\big] \\ &= E\big[\big((\tilde{C}_k - \mu_{\tilde{C}}) - \alpha_i \cdot (I_k - \mu_I)\big)\big] \\ &= 0. \end{aligned} \tag{14}$$

Therefore, to ensure a better mapping relationship when upsampling, the filtering output $\tilde{C}'$ is treated as the optimization objective. In particular, the block-wise adaptive downsampling is performed by dividing each image into $m \times n$ nonoverlapping block. Assuming the interpolation filter in upsampling (such as bilinear and bicubic) is F with matrix size $(m \times n) \times (m/2 \times n/2)$ [27]

$$F = \begin{bmatrix} f_{0,0} & f_{0,1} & \cdots & f_{0,m/2\times n/2-1} \\ f_{1,0} & f_{1,1} & \cdots & f_{1,m/2\times n/2-1} \\ \cdots & \cdots & \cdots & \cdots \\ f_{m\times n-1,0} & f_{m\times n-1,1} & \cdots & f_{m\times n-1,m/2\times n/2-1} \end{bmatrix}. \tag{15}$$

Here, the matrix element $f_{k,l}$ indicates the interpolation coefficient contributed by the lth downsampled pixel when performing the interpolation of the kth pixel.

Assuming the downsampled chroma block is Z of size $(m/2 \times n/2)$, then the objective function is defined as follows:

$$D_S = ||FZ + \Phi - \tilde{C}'||^2 \tag{16}$$

where Φ accounts for the upsampling contribution of boundary pixels that lie outside the current block [27]. The optimization process is performed by setting the derivative of D_S to zero, leading to

$$\frac{dD_S}{dZ} = 2F^T(FZ - (\tilde{C}' - \Phi)) = 0. \tag{17}$$

Finally, the optimal downsampled block Z^* is derived as follows:

$$Z^* = (F^T F)^{-1} F^T (\tilde{C}' - \Phi). \tag{18}$$

It performs as the left-inverse operator for a full-rank interpolation operator.

C. Complexity Analysis

In applications such as screen virtualization, real-time decoding and processing should be strictly satisfied to ensure the interactivity performance and user experience. The downsampling is usually performed in the cloud, which can provide more powerful computational resources for visual processing tasks. The complexity of the proposed upsampling algorithm is crucial, as it is usually performed at the thin client. In addition to the direct interpolation, the additional computational complexity brought by the upsampling algorithm originates from content analysis, parameter calculation, and linear transform. Assuming the block size is $N = m \times n$, and we are particularly interested in the computational complexity at block level. For content analysis, the complexity of gradient calculation and histogram establishment is $O(N)$ (constant per pixel). The linear least square method that solves (4) and the linear mapping are also $O(N)$ complexity algorithm. Therefore, the overall complexity of the upsampling method is $O(N)$, or equivalently $O(1)$ per pixel. The low complexity algorithm can be well incorporated into the screen virtualization applications to generate high quality screen content images. For the downsampling process, the complexity of the BCIM filtering is also $O(N)$ as the upsampling process. Another operation is to compute Φ and derive the downsampled pixels with (18). As the operator $(F^T F)^{-1} F^T$ can be precalculated, the complexity of this process is $O(N^2)$. Therefore, the overall complexity of the proposed downsampling algorithm is $O(N^2)$.

IV. Application to Screen Content Image Compression

Dong and Ye [28] proposed a practical approach that applies optimal downsampling ratio prior to encoding and upsampling after decoding, which has been demonstrated to significantly improve the rate–distortion (RD) performance over a wide range of bit rates. To further support the YUV4:2:0 screen content image compression with the proposed scheme, the downsampling is performed as a preprocessing process at the encoder, and adaptive upsampling is executed to generate the full chroma for display at the decoder. This process is shown in Fig. 11, where typical YUV4:2:0 codec is employed, such as H.264/AVC or HEVC. In this case, not only downsampling, but also compression will incur chroma distortion.

The main task of the codec is to convey the images with minimum possible distortion within available bit rate. Therefore, the downsampling and upsampling processes are better to be optimized in the framework of RD optimization (RDO) that attempts to optimize the reconstruction quality D subject to the constraint R_c

$$\min\{D\} \quad \text{s.t.} \quad R \le R_c. \tag{19}$$

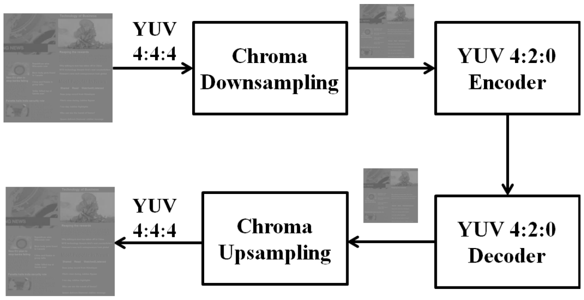


Fig. 11. Compression framework with the proposed joint downsampling and upsampling schemes.

This can be converted into an unconstraint optimization problem by [29]

$$\min\{J\} \quad \text{where} \quad J = D + \lambda \cdot R \tag{20}$$

where J is called the RD cost and λ is known as the Lagrange multiplier that controls the tradeoff between R and D.

Central to such an optimization problem is the way in which the distortion D is defined, because the quality of video can only be as good as it is optimized for. To achieve high coding efficiency, the distortion D should accurately reflect the ultimate distortion. For example, in [30] and [31], perceptual RDO scheme is proposed by employing more subjective-equivalent distortion models. In this context, as the final displayed image is rendered by the interpolated chroma components, the interpolation filter should be considered in the RDO framework as well.

In [32] and [33], the distortion model in HEVC is defined as

$$D = D_Y + \omega \cdot D_C + \lambda \cdot R \tag{21}$$

where D_Y and D_C denote the distortion of luma and chroma components in terms of sum of squared error (SSE). Parameter ω is the relative weight for chroma, and can be computed according to the difference between the quantization parameter (QP) of luma and chroma components, or the SSE of them in the previous frames. However, since the chroma frame has to be upsampled before display when using the YUV4:2:0 codec, a new distortion model is defined following the proposed upsampling processing to improve the coding efficiency.

Assuming the quantization error is ε, analogies to (16), the total distortion when considering the quantization process in encoding is formulated as

$$D_C = ||F(Z + \varepsilon) + \Phi - \tilde{C}'||^2. \tag{22}$$

In classical RD models [34], [35], the quantization error is a function of the quantization step. For example, one general model to estimate $||\varepsilon||^2$ is $(\Delta^2/12)$ [35], where Δ denotes the quantization step. Therefore, $FZ + \Phi - \tilde{C}'$ and $F\varepsilon$ can be approximated to be uncorrelated [36]. Finally, D_C can be approximated as

$$D_C = ||FZ + \Phi - \tilde{C}'||^2 + ||F\varepsilon||^2. \tag{23}$$

As the first term, $||FZ + \Phi - \tilde{C}'||^2$ is independent of quantization and optimized in the downsampling process, the chroma

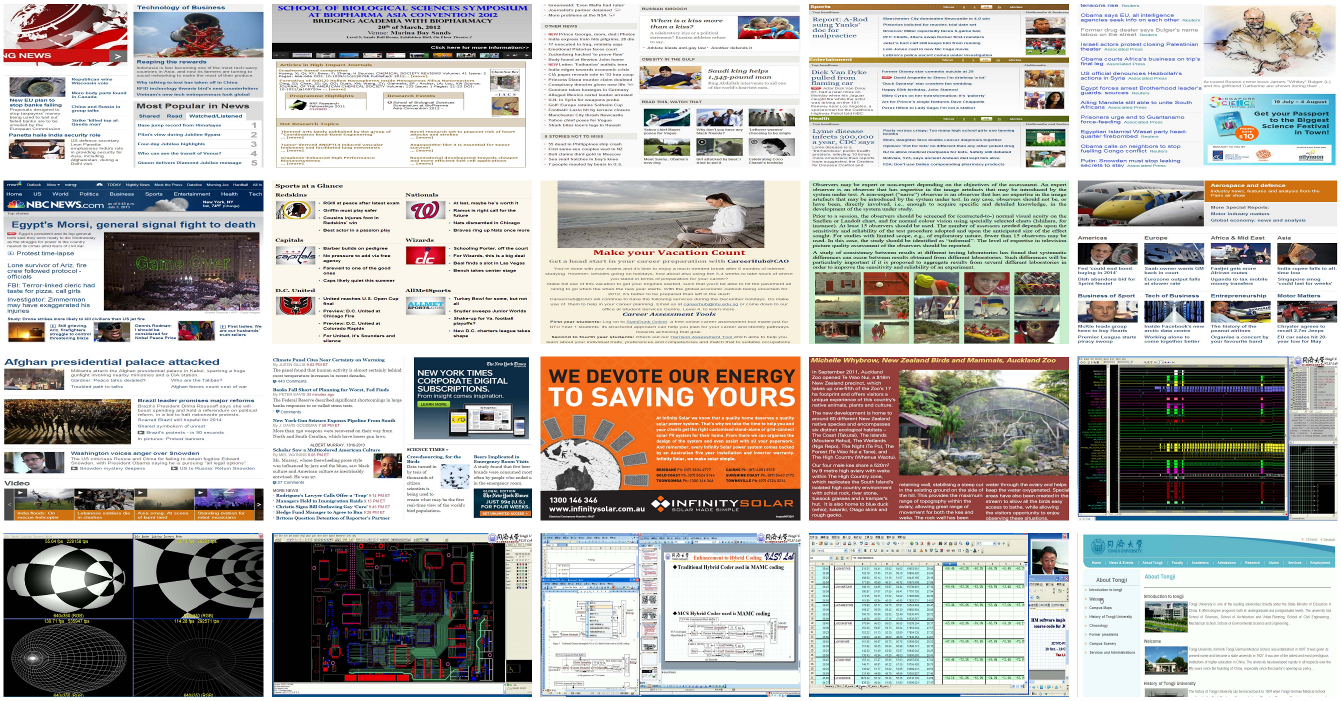


Fig. 12. Test screen content images (image1–image20).

distortion in the YUV4:2:0 encoder is finally defined as

$$D_C = ||F\varepsilon||^2. \quad (24)$$

As such, the RDO process in the encoder takes both the distortion and the interpolation process into account, so that more efficient partitions and coding modes can be selected.

V. Experimental Results

In this section, extensive experiments are carried out to evaluate the performance of the proposed algorithm. As shown in Fig. 12, 20 images from both the screen content quality assessment database [37] and the HEVC range extension test sequences are used for testing. The resolutions are from 624 × 624 to 1920 × 1080. Application scenarios of these test images include Web browsing, cloud computer aided design (CAD), and word editing. All of them are in YUV4:4:4 format. In the first experiment, we verify the performance of the downsampling and upsampling algorithms in terms of both objective quality assessment measures and subjective testing. In the second experiment, the complexity of the proposed method is evaluated. From the third to the fifth experiments, the influences of various parameter settings regarding the block size, block classification, and major color number on the final performance are investigated. Finally, the proposed method is incorporated into the YUV4:2:0 coding framework to demonstrate its applicability in screen content image compression.

A. Performance of Downsampling and Upsampling

In this experiment, the chroma components are first downsampled to YUV4:2:0 format and then upsampled to full chroma. In the proposed scheme, bilinear and bicubic are used for the initial interpolation, respectively. With regard to the downsampling process, in addition to the interpolation-dependent downsampling, two downsampling methods are employed, including average downsampling and MPEG-B downsampling [38]. In average downsampling, the downsampled value is the average of the four corresponding values. In MPEG-B downsampling, each image is filtered and then the upper left one in the four corresponding pixel values is used. The filter coefficient is set to be [2, 0, −4, −3, 5, 19, 26, 19, 5, −3, −4, 0, 2]/64. The combination of the downsampling and upsampling methods used for comprehensive comparison, include the following.

1) *D1-Direct:* Directly interpolating the chroma samples using bilinear/bicubic with average downsampling.
2) *D2-Direct:* Directly interpolating the chroma samples using bilinear/bicubic with MPEG-B downsampling.
3) *D1-GF:* Guided filter [15] with average downsampling.
4) *D2-GF:* Guided filter [15] with MPEG-B downsampling.
5) *IDID:* Directly interpolating the chroma samples using bilinear/bicubic with interpolation-dependent downsampling [27].
6) *ID-GF:* Guided filter [15] with interpolation-dependent downsampling.

In this experiment, the default settings of parameters are applied, and the average quality of the two chroma components is evaluated in terms of peak signal-to-noise ratio (PSNR) and structural similarity (SSIM) index [39], respectively. The results for each image are demonstrated in Tables I and II, from which we can observe that the proposed scheme achieves the best performance among these methods. This is because that the proposed method takes advantages of the properties of screen content to exploit the performance of joint downsampling and upsampling. Moreover, one can discern that simply

TABLE I

PERFORMANCE COMPARISON OF DIFFERENT COMBINATIONS IN TERMS OF PSNR

Image	Bicubic							Bilinear						
	D1-Direct	D2-Direct	D1-GF	D2-GF	ID-GF	IDID	Proposed	D1-Direct	D2-Direct	D1-GF	D2-GF	ID-GF	IDID	Proposed
Image1	33.46	32.21	32.42	31.93	32.64	33.76	34.24	33.07	32.09	32.20	31.81	32.64	33.74	34.23
Image2	32.05	30.88	31.38	31.24	31.62	32.39	33.13	31.60	30.74	31.09	31.01	31.63	32.37	33.11
Image3	33.10	32.03	31.50	31.38	31.57	33.37	33.80	32.78	31.94	31.44	31.34	31.57	33.35	33.79
Image4	30.92	30.07	30.39	30.03	30.56	31.12	31.69	30.69	29.99	30.21	29.92	30.56	31.10	31.67
Image5	35.15	33.13	36.93	35.28	37.05	35.79	37.37	34.30	32.88	35.45	34.37	37.10	35.75	37.38
Image6	33.35	32.12	32.07	31.60	32.31	33.72	34.63	32.90	31.97	31.84	31.45	32.31	33.69	34.61
Image7	32.36	31.08	30.24	30.06	30.33	32.75	33.70	31.90	30.92	30.16	30.01	30.34	32.73	33.68
Image8	33.56	32.72	32.96	32.71	33.11	33.81	34.50	33.28	32.63	32.79	32.59	33.11	33.79	34.48
Image9	31.26	30.23	29.84	29.64	29.97	31.62	31.86	30.88	30.08	29.72	29.57	29.97	31.60	31.84
Image10	32.35	31.18	31.60	31.16	31.86	32.65	33.32	31.98	31.07	31.35	31.02	31.85	32.63	33.29
Image11	32.36	31.54	31.79	31.47	31.98	32.58	33.40	32.11	31.46	31.59	31.34	31.98	32.56	33.39
Image12	30.54	29.64	29.69	29.35	29.88	30.86	31.70	30.20	29.50	29.52	29.25	29.88	30.84	31.68
Image13	33.79	31.16	35.26	34.06	35.44	34.54	35.66	32.76	30.89	34.53	33.07	35.45	34.49	35.65
Image14	38.86	36.51	39.34	37.78	39.15	39.60	40.24	37.91	36.26	38.03	36.99	39.17	39.56	40.23
Image15	25.65	25.15	26.49	25.83	26.71	25.86	27.80	25.52	25.07	26.06	25.55	26.71	25.84	27.79
Image16	38.45	36.60	36.70	36.10	37.35	39.08	41.52	37.64	36.32	36.15	35.68	37.36	39.03	41.49
Image17	28.27	27.21	26.33	26.03	26.56	28.62	31.02	27.88	27.07	26.11	25.89	26.56	28.59	31.02
Image18	33.53	32.56	32.43	32.19	32.67	33.81	34.92	33.19	32.44	32.19	32.01	32.68	33.79	34.91
Image19	32.42	31.65	31.80	31.59	32.00	32.65	33.35	32.16	31.55	31.59	31.44	32.00	32.62	33.33
Image20	36.58	35.45	35.30	35.09	35.42	36.94	37.92	36.21	35.31	35.18	35.00	35.42	36.92	37.91
Average	**32.90**	**31.66**	**32.22**	**31.73**	**32.41**	**33.28**	**34.29**	**32.45**	**31.51**	**31.86**	**31.47**	**32.41**	**33.25**	**34.27**

TABLE II

PERFORMANCE COMPARISON OF DIFFERENT COMBINATIONS IN TERMS OF SSIM

Image	Bicubic							Bilinear						
	D1-Direct	D2-Direct	D1-GF	D2-GF	ID-GF	IDID	Proposed	D1-Direct	D2-Direct	D1-GF	D2-GF	ID-GF	IDID	Proposed
Image1	0.9134	0.8812	0.8781	0.8651	0.8842	0.9225	0.9257	0.9016	0.8758	0.8721	0.8616	0.8843	0.9221	0.9253
Image2	0.8933	0.8567	0.8650	0.8591	0.8725	0.9048	0.9215	0.8788	0.8500	0.8575	0.8530	0.8726	0.9043	0.9212
Image3	0.8879	0.8506	0.8274	0.8240	0.8297	0.8995	0.9068	0.8746	0.8454	0.8253	0.8227	0.8297	0.8990	0.9063
Image4	0.8623	0.8241	0.8327	0.8160	0.8405	0.8724	0.8861	0.8484	0.8177	0.8248	0.8115	0.8405	0.8720	0.8856
Image5	0.9528	0.9256	0.9694	0.9558	0.9606	0.9605	0.9726	0.9410	0.9190	0.9573	0.9451	0.9708	0.9602	0.9726
Image6	0.9222	0.8933	0.8893	0.8752	0.8966	0.9321	0.9433	0.9098	0.8865	0.8823	0.8706	0.8967	0.9318	0.9431
Image7	0.9189	0.8872	0.8589	0.8524	0.8624	0.9303	0.9437	0.9056	0.8808	0.8559	0.8509	0.8625	0.9299	0.9435
Image8	0.9037	0.8804	0.8778	0.8707	0.8813	0.9118	0.9213	0.8947	0.8766	0.8743	0.8690	0.8813	0.9114	0.9210
Image9	0.8498	0.7999	0.7777	0.7666	0.7868	0.8698	0.8762	0.8284	0.7890	0.7701	0.7617	0.7868	0.8692	0.8757
Image10	0.9107	0.8761	0.8792	0.8641	0.8867	0.9210	0.9266	0.8981	0.8708	0.8719	0.8603	0.8868	0.9206	0.9262
Image11	0.8931	0.8657	0.8648	0.8544	0.8702	0.9011	0.9130	0.8822	0.8608	0.8588	0.8506	0.8702	0.9007	0.9127
Image12	0.8623	0.8232	0.8222	0.8078	0.8304	0.8776	0.8966	0.8454	0.8146	0.8146	0.8030	0.8306	0.8772	0.8962
Image13	0.9471	0.9076	0.9661	0.9509	0.9563	0.9552	0.9614	0.9316	0.9005	0.9541	0.9404	0.9564	0.9547	0.9612
Image14	0.9512	0.9185	0.9628	0.9459	0.9549	0.9607	0.9696	0.9376	0.9111	0.9495	0.9346	0.9550	0.9604	0.9696
Image15	0.8033	0.7479	0.8129	0.7836	0.8225	0.8066	0.8739	0.7871	0.7418	0.7947	0.7704	0.8225	0.8060	0.8736
Image16	0.9874	0.9818	0.9828	0.9814	0.9846	0.9890	0.9941	0.9848	0.9805	0.9810	0.9798	0.9846	0.9890	0.9941
Image17	0.8546	0.8184	0.8198	0.7981	0.8345	0.8678	0.9104	0.8390	0.8103	0.8069	0.7905	0.8344	0.8673	0.9102
Image18	0.9200	0.9014	0.8992	0.8935	0.9040	0.9261	0.9418	0.9124	0.8977	0.8946	0.8902	0.9041	0.9258	0.9416
Image19	0.8884	0.8619	0.8687	0.8606	0.8767	0.8973	0.9191	0.8777	0.8565	0.8607	0.8546	0.8767	0.8968	0.9189
Image20	0.9391	0.9179	0.9131	0.9095	0.9155	0.9461	0.9563	0.9312	0.9142	0.9107	0.9077	0.9156	0.9459	0.9561
Average	**0.9031**	**0.8710**	**0.8784**	**0.8667**	**0.8826**	**0.9126**	**0.9280**	**0.8905**	**0.8650**	**0.8709**	**0.8614**	**0.8831**	**0.9122**	**0.9277**

applying the GF method in screen content will not bring better performance, as the unique characteristics of screen content are not considered. This implies that BCIM-based luma to chroma mapping plays an important role in the upsampling process. It is not surprising that IDID can improve the performance compared with the direct and GF approaches, as it performs optimization to deliver more information in the downsampled image. However, the proposed method significantly outperforms IDID in terms of both PSNR and SSIM, which shows the superior performance of the proposed algorithm.

In Figs. 13 and 14, we demonstrate the cropped images from original, D1-Direct interpolation, state-of-the-art schemes (D1-GF and IDID), and the proposed method. Since our proposed scheme is based on the structure mapping between luma and chroma, better subjective quality is achieved. It can be observed that more structure information have been preserved, and the color shifting artifacts are efficiently reduced. The visual quality improvement is due to the fact that the efficient representation of the geometric relationship with BCIM, resulting in more accurate structure transfer from luma to chroma.

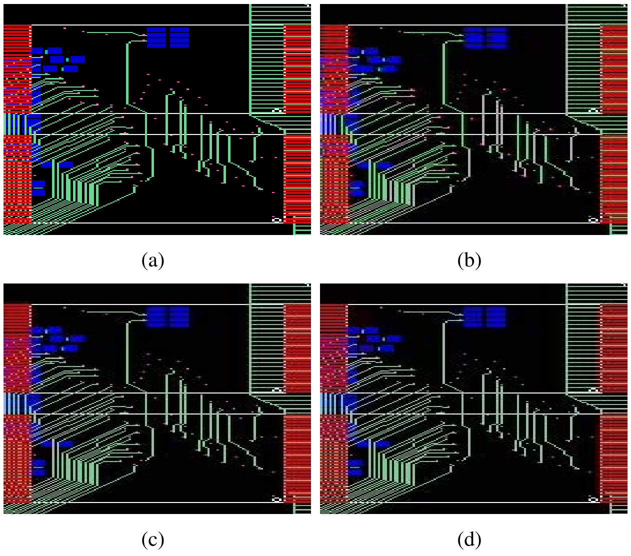
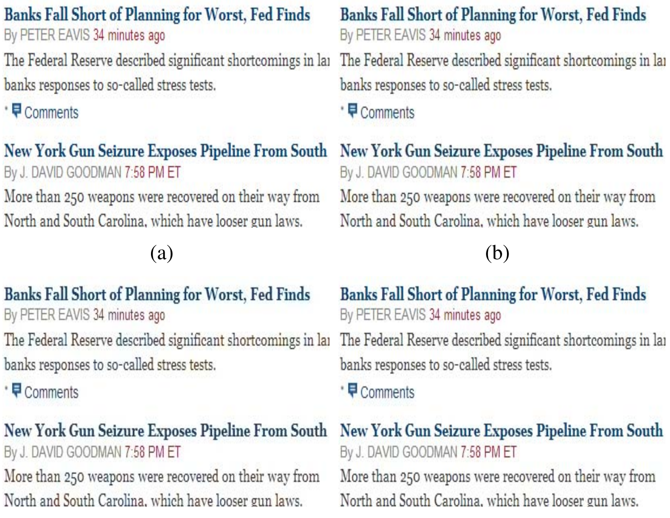

(a) (b) (c) (d)

Fig. 13. Visual quality comparison (cropped for visualization). (a) Original. (b) D1-GF. (c) IDID. (d) Proposed scheme.

To further validate the proposed scheme, we have conducted a subjective experiment, where 16 subjects were invited to rank six sets of images based on bilinear and six sets of images based on bicubic, respectively. It is noted that the original image sets for bilinear and bicubic cases are exactly the same. A general introduction was given at the beginning of the whole test, and more specific instructions and training session were given afterward. In particular, each screen content image is viewed at least for 10 s and zooming is allowed. This is



Banks Fall Short of Planning for Worst, Fed Finds
By PETER EAVIS 34 minutes ago
The Federal Reserve described significant shortcomings in la
banks responses to so-called stress tests.
Comments

New York Gun Seizure Exposes Pipeline From South
By J. DAVID GOODMAN 7:58 PM ET
More than 250 weapons were recovered on their way from
North and South Carolina, which have looser gun laws.

(a)

Banks Fall Short of Planning for Worst, Fed Finds
By PETER EAVIS 34 minutes ago
The Federal Reserve described significant shortcomings in la
banks responses to so-called stress tests.
Comments

New York Gun Seizure Exposes Pipeline From South
By J. DAVID GOODMAN 7:58 PM ET
More than 250 weapons were recovered on their way from
North and South Carolina, which have looser gun laws.

(b)

Banks Fall Short of Planning for Worst, Fed Finds
By PETER EAVIS 34 minutes ago
The Federal Reserve described significant shortcomings in la
banks responses to so-called stress tests.
Comments

New York Gun Seizure Exposes Pipeline From South
By J. DAVID GOODMAN 7:58 PM ET
More than 250 weapons were recovered on their way from
North and South Carolina, which have looser gun laws.

(c)

Banks Fall Short of Planning for Worst, Fed Finds
By PETER EAVIS 34 minutes ago
The Federal Reserve described significant shortcomings in la
banks responses to so-called stress tests.
Comments

New York Gun Seizure Exposes Pipeline From South
By J. DAVID GOODMAN 7:58 PM ET
More than 250 weapons were recovered on their way from
North and South Carolina, which have looser gun laws.

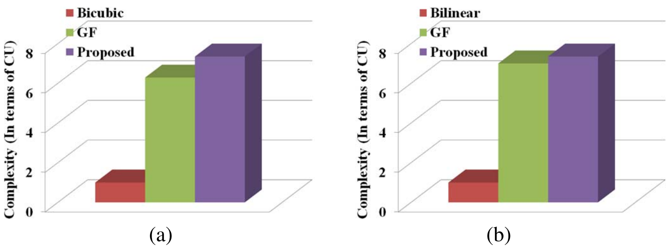
(d)

Fig. 14. Visual quality comparison (cropped for visualization). (a) Original. (b) D1-Direct method (bilinear). (c) D1-GF. (d) Proposed scheme.

TABLE III
AVERAGE RANKINGS BY SUBJECTIVE TESTS (THE TOP TWO METHODS ARE IN BOLDFACE)

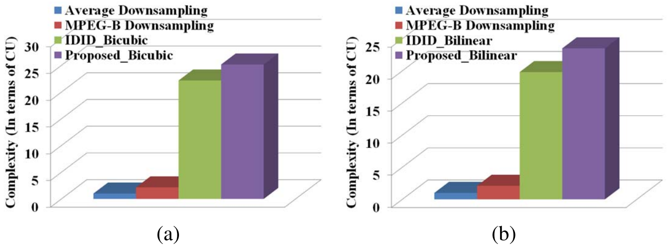
Number	Org	D1-Direct	IDID	ID-GF	D1-GF	Proposed
1	**1.44**	4.31	2.50	4.81	5.88	**2.06**
2	**1.38**	5.25	**2.13**	4.50	5.00	2.75
3	**1.13**	4.25	2.81	5.44	5.25	**2.13**
4	**2.06**	3.56	3.69	4.63	5.50	**1.56**
5	**1.56**	4.19	4.13	4.31	5.25	**1.56**
6	**1.44**	4.00	2.75	4.94	5.63	**2.25**
7	**1.25**	4.00	4.13	4.75	4.88	**2.00**
8	**1.13**	4.69	3.81	3.94	5.38	**2.06**
9	**1.63**	4.56	2.81	4.94	5.19	**1.88**
10	**2.00**	4.13	2.56	5.13	5.56	**1.63**
11	**1.44**	4.25	2.94	4.81	5.44	**2.13**
12	**1.88**	3.25	3.50	5.06	5.38	**1.94**
13	**1.56**	4.25	**2.25**	5.19	5.31	2.44
14	**1.38**	4.38	**2.38**	5.00	5.38	2.50
15	**1.31**	4.75	**2.31**	5.00	4.94	2.69
16	**1.50**	4.38	2.44	4.81	5.56	**2.31**
17	**1.38**	4.06	3.44	4.94	5.44	**1.75**
18	**1.69**	4.75	3.19	4.63	5.19	**1.56**
19	**1.31**	3.94	2.81	5.31	5.19	**2.44**
20	**1.38**	4.75	2.81	4.75	5.44	**1.88**

because that the major task of the HVS when viewing an image is to act as an optimal information extractor [40], and the screen content image usually contains richer information than natural images. The principle is to rank the images based on the visual quality. The results for bilinear and bicubic are averaged for demonstration. In Table III, the average rankings are shown, where lower values correspond to better quality. It is observed that, the proposed scheme has obtained outstanding results and 16 of 20 images are among the top two (in most cases, the best one is the original image without any distortion). These results provide proof of the superiority of the proposed scheme in chroma sampling applications of screen content images.



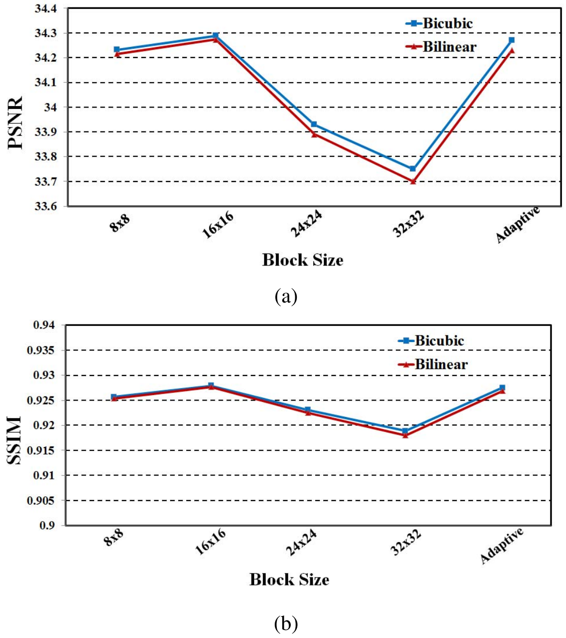
(a) (b)

Fig. 15. Comparison of the computational complexity for upsampling. (a) Bicubic case. (b) Bilinear case.



(a) (b)

Fig. 16. Comparison of the computational complexity for downsampling. (a) Bicubic case. (b) Bilinear case.



(a)

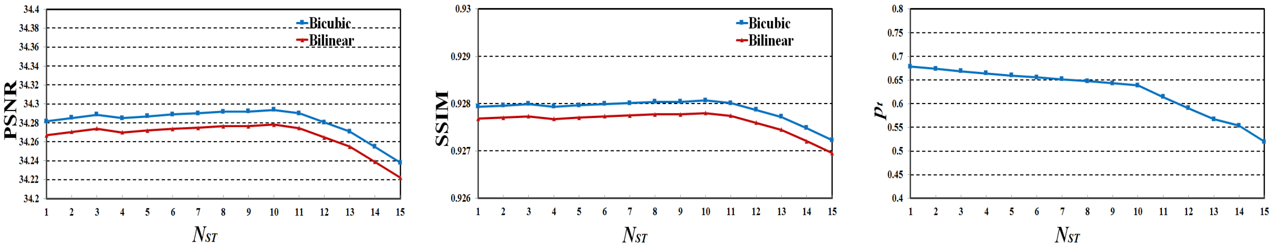
(b)

Fig. 17. Influence of the block partition size on the final performance. (a) Performance evaluated in terms of PSNR. (b) Performance evaluated in terms of SSIM.

B. Complexity Comparison

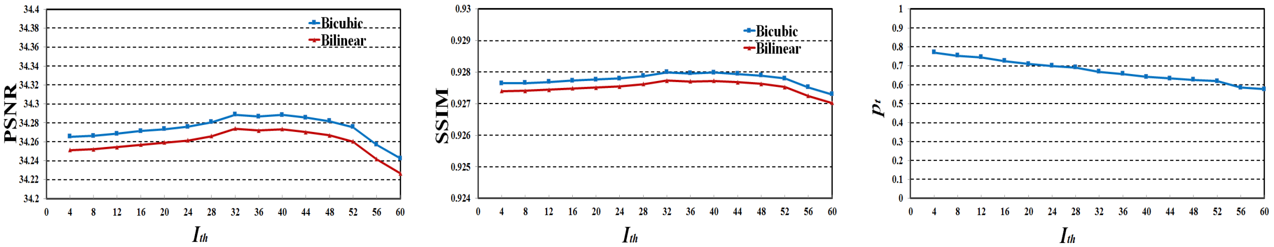
In this section, we evaluate the complexity of the proposed downsampling and upsampling methods, respectively. In particular, we employ the computation unit (CU) to facilitate the quantitative measurement of the computational consumptions. For downsampling, we choose the computation time of average downsampling as the CU. For upsampling, we choose the computation time of direct bilinear/bicubic as the CU. Furthermore, we scale the computation consumptions of other downsampling/upsampling methods as the multiple of CU. To obtain more accurate complexity comparison, we perform the program in MATLAB for 100 times with Intel 3.40-GHz Core processor and 8-GB random access memory, and the average results are recorded.

As indicated in Section III-C, the computational overhead for downsampling is $O(N^2)$ while for upsampling algorithm



(a) (b) (c)

Fig. 18. Influence of parameter N_{ST} on the final performance. (a) PSNR. (b) SSIM. (c) High gradient block ratio.



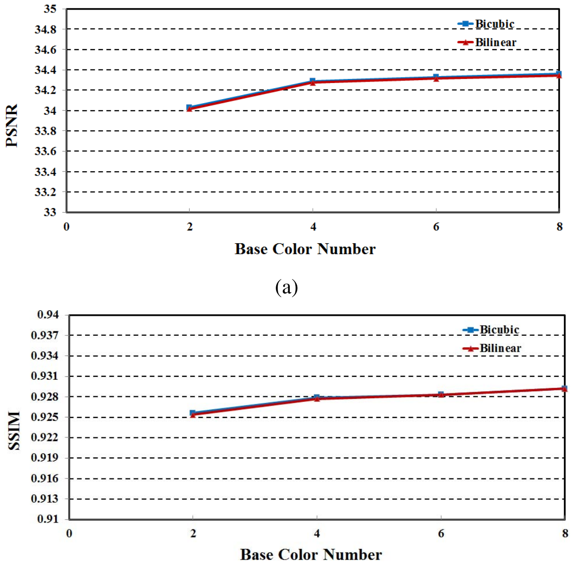
(a) (b) (c)

Fig. 19. Influence of parameter I_{th} on the final performance. (a) PSNR. (b) SSIM. (c) High gradient block ratio.

is $O(N)$. As demonstrated in Figs. 15 and 16, it is observed that the complexity of upsampling processing is around 7 CUs and that of the downsampling processing is 25 CUs. It is noted that for IDID and proposed methods, they require to specify the upsampling method to generate the downsampling pixels. Therefore, the downsampling complexity levels in the scenarios of both bicubic and bilinear are demonstrated. Moreover, it is interesting to find that the proposed approach has a comparable complexity level with the GF method, of which the computational time is also $O(N)$. The relative low complexity upsampling ensures its applicability in thin clients for efficient screen rendering.

C. Impact of Block Partition Size

In Section V-A, the default block size 16 × 16 is employed to evaluate the performance. In general, either too large or too small block size may introduce performance loss, as small block size may not fully capture the major colors for chroma pixel reconstruction, while large block may destroy the structure mapping between luma and chroma. To further investigate this issue, the block size is set to be 8 × 8, 16 × 16, 24 × 24, and 32 × 32. Moreover, an adaptive block size selection strategy is applied, where the block size is set to be equal to the coding unit size in HEVC ($QP = 20$) [41]. The average PSNR and SSIM scores for 20 images are demonstrated in Fig. 17. It is observed that the default setting achieves comparable performance with the adaptive block size selection strategy. Moreover, smaller or larger block partitions may lead to performance degradation in terms of both PSNR and SSIM. However, when comparing the reconstruction quality in Tables I and II of other methods such as IDID and GF, they are still able to achieve superior performance, which further illustrates the robustness of the proposed algorithm.



(a)

(b)

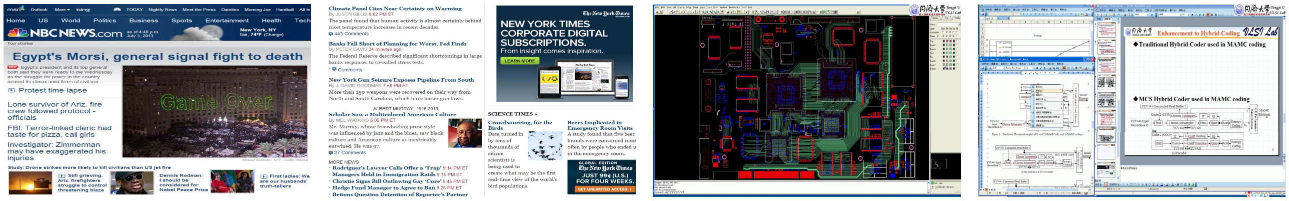
Fig. 20. Influence of the number of base colors on the final performance. (a) Performance evaluated in terms of PSNR. (b) Performance evaluated in terms of SSIM.

D. Impact of Block Classification

To exhibit the sensitivity of parameter setting in classifying the block types, we conduct a detailed study to investigate the impact of the variations of N_{ST} and I_{th} on the final performance. In (2), I_{th} is defined as the threshold for classifying high gradient pixels and N_{ST} indicates the threshold on the number of high gradient pixels to identify the high gradient blocks.

In particular, in addition to the reconstruction quality, the high gradient block ratio is examined as well

$$p_t = \frac{\text{Num}_t}{\text{Num}_w} \tag{25}$$



msn Outlook More bing TODAY Nightly News Meet the Press Dateline Morning Joe Hardball All In

Home US World Politics Business Sports Entertainment Health Tech

NBCNEWS.com as of 4:48 p.m. July 3, 2013 New York, NY fair, 74°F (Change)

Top stories

Egypt's Morsi, general signal fight to death

Egypt's president and its top general both said they were ready to die Wednesday as the struggle for power in the country neared its climax amid fears of civil war.

Protest time-lapse

Lone survivor of Ariz. fire crew followed protocol - officials

FBI: Terror-linked cleric had taste for pizza, call girls

Investigator: Zimmerman may have exaggerated his injuries

Khaled Desouki / AFP - Getty Images

Study: Drone strikes more likely to kill civilians than US jet fire

Still grieving, Ariz. firefighters struggle to control threatening blaze

Dennis Rodman: I should be considered for Nobel Peace Prize

First ladies: We are our husbands' truth-tellers

(a)

Climate Panel Cites Near Certainty on Warming
By JUSTIN GILLIS 9:02 PM ET
The panel found that human activity is almost certainly behind most temperature increases in recent decades.
443 Comments

Banks Fall Short of Planning for Worst, Fed Finds
By PETER EAVIS 34 minutes ago
The Federal Reserve described significant shortcomings in large banks responses to so-called stress tests.
Comments

New York Gun Seizure Exposes Pipeline From South
By J. DAVID GOODMAN 7:58 PM ET
More than 250 weapons were recovered on their way from North and South Carolina, which have looser gun laws.

ALBERT MURRAY, 1916-2013
Scholar Saw a Multicolored American Culture
By MEL WATKINS 6:05 PM ET
Mr. Murray, whose freewheeling prose style was influenced by jazz and the blues, saw black culture and American culture as inextricably entwined. He was 97.
27 Comments

MORE NEWS
- Rodriguez's Lawyer Calls Offer a 'Trap' 9:14 PM ET
- Managers Held in Immigration Raids 9:15 PM ET
- Christie Signs Bill Outlawing Gay 'Cure' 9:45 PM ET
- Hedge Fund Manager to Agree to Ban 9:26 PM ET
- Britons Question Detention of Reporter's Partner

The New York Times
NEW YORK TIMES CORPORATE DIGITAL SUBSCRIPTIONS.
From insight comes inspiration.
LEARN MORE

SCIENCE TIMES »

Crowdsourcing, for the Birds
Data turned in by tens of thousands of citizen scientists is being used to create what may be the first real-time view of the world's bird populations.

Beers Implicated in Emergency Room Visits
A study found that five beer brands were consumed most often by people who ended u in the emergency room.

GLOBAL EDITION The New York Times JUST 99¢ (U.S.) FOR FOUR WEEKS. GET UNLIMITED ACCESS ›

(b)

(c)

Enhancement to Hybrid Coding

◆Traditional Hybrid Coder used in MAMC coding

◆MCS Hybrid Coder used in MAMC coding

(d)

Fig. 21. Illustration of the test images in screen image compression. (a) WebPage1. (b) WebPage2. (c) PCB_Layout. (d) PPT_DOC_XLS.

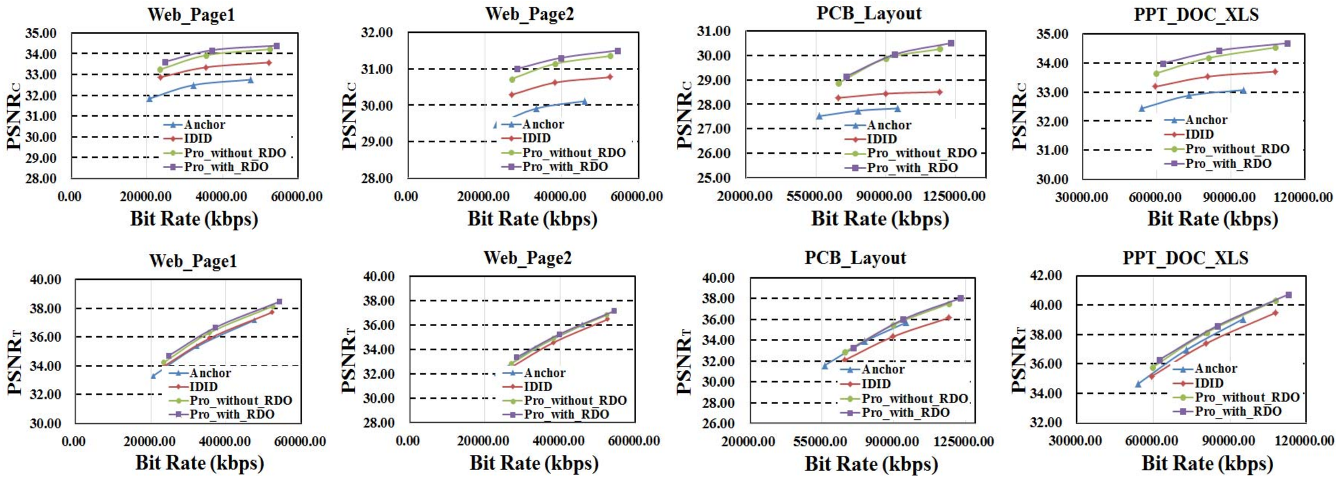


Fig. 22. RD performance comparison in terms of PSNR_C and PSNR_T (interpolation method: bilinear).

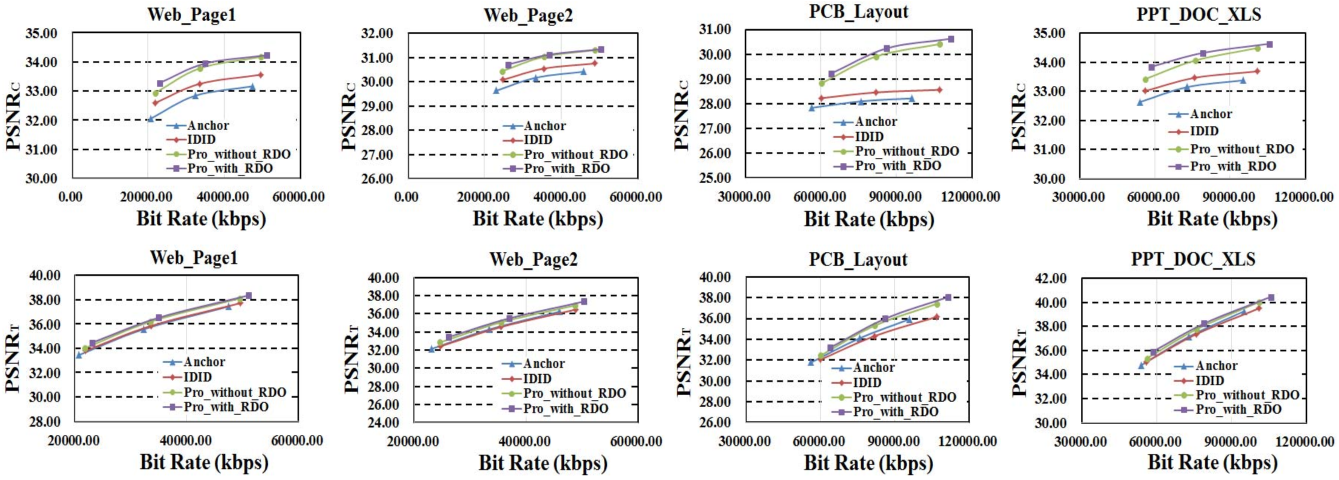


Fig. 23. RD performance comparison in terms of PSNR_C and PSNR_T (interpolation method: bicubic).

where Num_t and Num_w indicate the number of high gradient blocks and total blocks, respectively.

The average of PSNR, SSIM as well as the high gradient block ratio p_t for 20 images in terms of different N_{ST} settings are demonstrated in Fig. 18, where it is shown that the proposed method produces reasonably stable performance when N_{ST} is set within a certain range. The performance starts to degrade when $N_{\text{ST}} > 10$, resulting from the reason that less blocks are classified into high gradient blocks for processing. A similar trend can be observed for various settings of I_{th} in Fig. 19 as well. When $I_{\text{th}} > 52$, the performance begin to significantly degrade. Moreover, it is also observed that introducing more natural image blocks into the BCIM-based upsampling process by setting small values of N_{ST} and I_{th} will not lead to significant performance variation, as we restrict the number of major colors to accommodate the limited color property of textual blocks.

E. Impact of Base Colors

In this section, we conduct a study to investigate the impact of extracted base color number on the final performance.

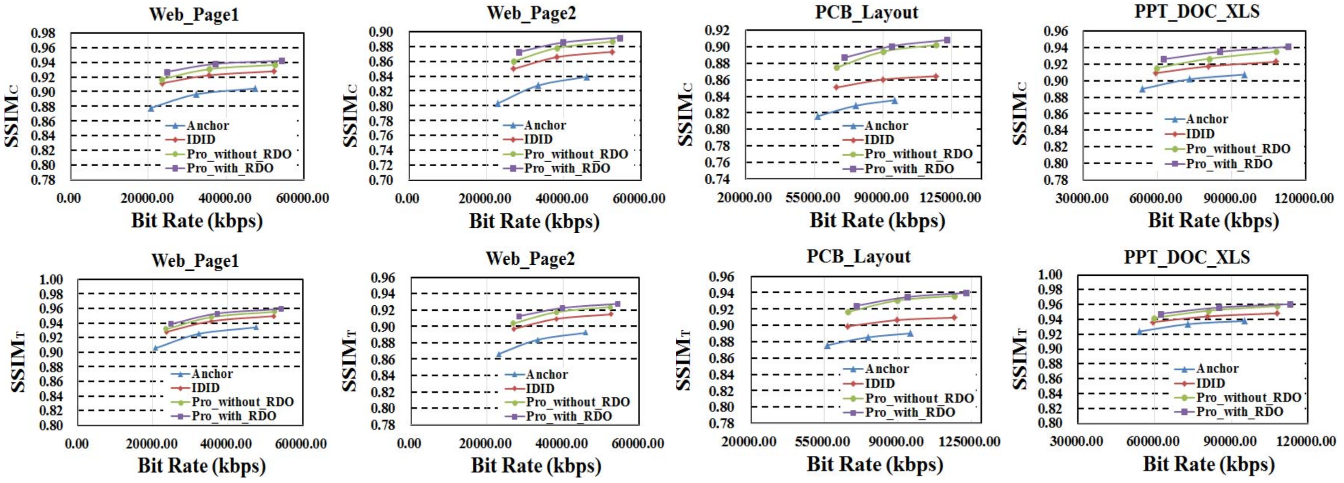


Fig. 24. RD performance comparison in terms of SSIM_C and SSIM_T (interpolation method: bilinear).

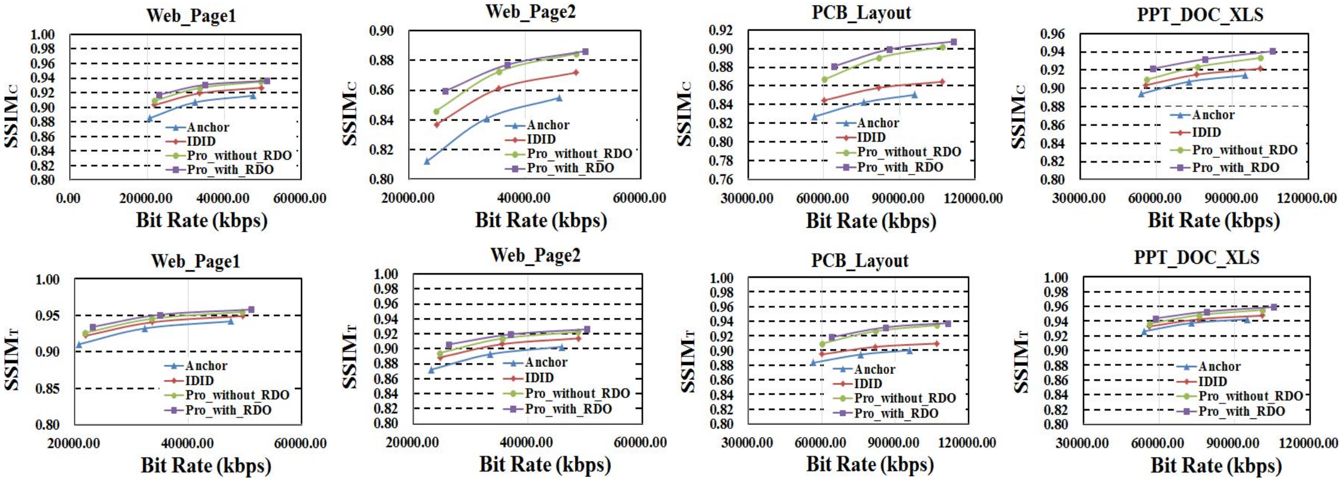


Fig. 25. RD performance comparison in terms of SSIM_C and SSIM_T (interpolation method: bicubic).

In the proposed scheme, to achieve a good balance between the performance and complexity, the default setting on the number of employed base colors is limited to four. To further study the optimal base color number setting, we vary the settings by including two, four, six, and eight base colors in the BCIM representation. If the maximum number of base colors within a block is less than the restricted number, all the base colors are employed. The performance is demonstrated in Fig. 20, where it is observed that when the number of employed base colors is larger than four, little performance improvement is observed. This can be explained by the property of limited colors in screen content, such that increasing the number of base colors beyond the inherent color number in textual block has little influence on the final performance.

F. Performance of Screen Content Image Compression

In this section, the proposed scheme implemented with default parameter settings is incorporated into the YUV4:2:0 coding framework. More specifically, HEVC codec (HM13.0) [42] with intra configuration is employed to compress the screen content images. The RD performance is evaluated in terms of the PSNR and SSIM (PSNR_C and SSIM_C) of the chroma component as well as the average quality of all the channels

$$\mathrm{PSNR}_T = \frac{1}{3}(\mathrm{PSNR}_Y + \mathrm{PSNR}_U + \mathrm{PSNR}_V) \tag{26}$$

$$\mathrm{SSIM}_T = \frac{1}{3}(\mathrm{SSIM}_Y + \mathrm{SSIM}_U + \mathrm{SSIM}_V). \tag{27}$$

We compare the RD performance of the proposed scheme with the methods such as bicubic/bilinear and IDID. In particular, the performance of the RDO algorithm that considers the final sampling process is evaluated as well. Typical application scenarios of screen content compression are considered, including Web browsing, word editing, and cloud CAD, as demonstrated in Fig. 21. For each comparison, bicubic or bilinear interpolation method with average downsampling is considered as anchor. The results are shown from Figs. 22–25, which demonstrate that the proposed scheme with RDO outperforms the others.[1] Moreover, the proposed scheme without RDO still achieves significant improvement

[1] The coding bits are converted to bit rate (kbits/s) in RD curves.

over the other methods except the approach with RDO. This verifies the effectiveness of the proposed joint upsampling and downsampling schemes in the application of screen content image compression. It is also observed that the bit rate saving at high bit rate is more significant. This is because that at high bit rate, the quantization error is almost ignorable compared with the distortion introduced by downsampling.

VI. Conclusion

The novelty of this paper lies in the joint design of adaptive upsampling and downsampling algorithms, which seamlessly work together for chroma format conversion of screen content image. The upsampling filter is developed with the utilization of major colors extracted from the luma component. By means of the major color and index map representation, the structure information is transferred from luma to chroma component with a linear transform. To further enable high efficiency upsampling, an adaptive downsampling filter is developed by involving the BCIM-based inverse operation of interpolation. This scheme is further incorporated into screen content compression framework to verify its efficiency. The superior performance of the proposed was demonstrated, which offered significant quality improvement as well as bit rate reduction in terms of the chroma quality.

Acknowledgment

The authors would like to thank the anonymous reviewers for their valuable comments that significantly helped in improving the presentation of this paper.

References

[1] Y. Lu, S. Li, and H. Shen, “Virtualized screen: A third element for cloud—Mobile convergence,” *IEEE MultiMedia*, vol. 18, no. 2, pp. 4–11, Feb. 2011.

[2] *Onlive Cloud Gaming*. [Online]. Available: https://games.onlive.com/, accessed Oct. 2015.

[3] C.-Y. Huang, C.-H. Hsu, Y.-C. Chang, and K.-T. Chen, “GamingAnywhere: An open cloud gaming system,” in *Proc. 4th ACM Multimedia Syst. Conf.*, 2013, pp. 36–47.

[4] H. Shen, Y. Lu, F. Wu, and S. Li, “A high-performanance remote computing platform,” in *Proc. IEEE Int. Conf. Pervasive Comput. Commun.*, Mar. 2009, pp. 1–6.

[5] *Virtual Network Computing (VNC)*. [Online]. Available: http://www.realvnc.com/, accessed Oct. 2015.

[6] Microsoft. *Remote Desktop Protocol (RDP)*. [Online]. Available: http://msdn.microsoft.com/en-us/library/aa383015(v=vs.85).aspx, accessed Oct. 2015.

[7] Z. Pan, H. Shen, Y. Lu, S. Li, and N. Yu, “A low-complexity screen compression scheme for interactive screen sharing,” *IEEE Trans. Circuits Syst. Video Technol.*, vol. 23, no. 6, pp. 949–960, Jun. 2013.

[8] C. Lan, G. Shi, and F. Wu, “Compress compound images in H.264/MPGE-4 AVC by exploiting spatial correlation,” *IEEE Trans. Image Process.*, vol. 19, no. 4, pp. 946–957, Apr. 2010.

[9] H. Yu, R. Cohen, A. Duenas, D.-K. Kwon, T. Lin, and J. Xu, *JCT-VC AHG Report: Screen Content Coding (AHG8)*, document JCTVC-Q0008, 2014.

[10] T. Lin, P. Zhang, S. Wang, K. Zhou, and X. Chen, “Mixed chroma sampling-rate High Efficiency Video Coding for full-chroma screen content,” *IEEE Trans. Circuits Syst. Video Technol.*, vol. 23, no. 1, pp. 173–185, Jan. 2013.

[11] S. Wang, J. Fu, Y. Lu, S. Li, and W. Gao, “Content-aware layered compound video compression,” in *Proc. IEEE Int. Symp. Circuits Syst.*, May 2012, pp. 145–148.

[12] X. Li and M. T. Orchard, “New edge-directed interpolation,” *IEEE Trans. Image Process.*, vol. 10, no. 10, pp. 1521–1527, Oct. 2001.

[13] X. Liu, D. Zhao, R. Xiong, S. Ma, W. Gao, and H. Sun, “Image interpolation via regularized local linear regression,” *IEEE Trans. Image Process.*, vol. 20, no. 12, pp. 3455–3469, Dec. 2011.

[14] J. Kopf, M. F. Cohen, D. Lischinski, and M. Uyttendaele, “Joint bilateral upsampling,” *ACM Trans. Graph.*, vol. 26, no. 3, p. 96, 2007.

[15] K. He, J. Sun, and X. Tang, “Guided image filtering,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 6, pp. 1397–1409, Jun. 2013.

[16] B. C. Song, Y. G. Lee, and N. H. Kim, “Block adaptive inter-color compensation algorithm for RGB 4:4:4 video coding,” *IEEE Trans. Circuits Syst. Video Technol.*, vol. 18, no. 10, pp. 1447–1451, Oct. 2008.

[17] L. Zhao and M. Ai, “Region adaptive inter-color prediction approach to RGB 4:4:4 intra coding,” in *Proc. 4th Pacific-Rim Symp. Image Video Technol.*, Nov. 2010, pp. 203–207.

[18] S. H. Lee and N. I. Cho, “Intra prediction method based on the linear relationship between the channels for YUV 4:2:0 intra coding,” in *Proc. IEEE Int. Conf. Image Process.*, Nov. 2009, pp. 1037–1040.

[19] X. Zhang, C. Gisquet, E. François, F. Zou, and O. C. Au, “Chroma intra prediction based on inter-channel correlation for HEVC,” *IEEE Trans. Image Process.*, vol. 23, no. 1, pp. 274–286, Jan. 2014.

[20] D. J. Field and N. Brady, “Visual sensitivity, blur and the sources of variability in the amplitude spectra of natural scenes,” *Vis. Res.*, vol. 37, no. 23, pp. 3367–3383, Dec. 1997.

[21] D. L. Ruderman and W. Bialek, “Statistics of natural images: Scaling in the woods,” *Phys. Rev. Lett.*, vol. 73, no. 6, pp. 814–817, Aug. 1994.

[22] E. P. Simoncelli and B. A. Olshausen, “Natural image statistics and neural representation,” *Annu. Rev. Neurosci.*, vol. 24, no. 1, pp. 1193–1216, 2001.

[23] M. Mrak, S. Grgic, and M. Grgic, “Picture quality measures in image compression systems,” in *Proc. Int. IEEE Region 8 Conf. Comput. Tool EUROCON*, vol. 1. Sep. 2003, pp. 233–236.

[24] Y.-W. Huang, P. Onno, R. Joshi, R. Cohen, X. Xiu, and Z. Ma, *SCCE3: Summary Report of CE on Palette Mode*, document JCTVC-R0033, Sapporo, Japan, Jul. 2014.

[25] W. Zhu, W. Ding, J. Xu, Y. Shi, and B. Yin, “Screen content coding based on HEVC framework,” *IEEE Trans. Multimedia*, vol. 16, no. 5, pp. 1316–1326, Aug. 2013.

[26] A. Levin, D. Lischinski, and Y. Weiss, “A closed-form solution to natural image matting,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 30, no. 2, pp. 228–242, Feb. 2008.

[27] Y. Zhang, D. Zhao, J. Zhang, R. Xiong, and W. Gao, “Interpolation-dependent image downsampling,” *IEEE Trans. Image Process.*, vol. 20, no. 11, pp. 3291–3296, Nov. 2011.

[28] J. Dong and Y. Ye, “Adaptive downsampling for high-definition video coding,” *IEEE Trans. Circuits Syst. Video Technol.*, vol. 24, no. 3, pp. 480–488, Mar. 2014.

[29] G. J. Sullivan and T. Wiegand, “Rate-distortion optimization for video compression,” *IEEE Signal Process. Mag.*, vol. 15, no. 6, pp. 74–90, Nov. 1998.

[30] Y.-H. Huang, T.-S. Ou, P.-Y. Su, and H. H. Chen, “Perceptual rate-distortion optimization using structural similarity index as quality metric,” *IEEE Trans. Circuits Syst. Video Technol.*, vol. 20, no. 11, pp. 1614–1624, Nov. 2010.

[31] S. Wang, A. Rehman, Z. Wang, S. Ma, and W. Gao, “SSIM-motivated rate-distortion optimization for video coding,” *IEEE Trans. Circuits Syst. Video Technol.*, vol. 22, no. 4, pp. 516–529, Apr. 2012.

[32] B. Li, G. J. Sullivan, and J. Xu, “Compression performance of High Efficiency Video Coding (HEVC) working draft 4,” in *Proc. IEEE Int. Symp. Circuits Syst.*, May 2012, pp. 886–889.

[33] B. Li, J. Xu, and H. Li, “Rate-distortion optimization with adaptive weighted distortion in High Efficiency Video Coding,” in *Proc. IEEE Int. Symp. Circuits Syst.*, May 2013, pp. 481–484.

[34] Z. He, J. Cai, and C. W. Chen, “Joint source channel rate-distortion analysis for adaptive mode selection and rate control in wireless video coding,” *IEEE Trans. Circuits Syst. Video Technol.*, vol. 12, no. 6, pp. 511–523, Jun. 2002.

[35] H.-M. Hang and J.-J. Chen, “Source model for transform video coder and its application. I. Fundamental theory,” *IEEE Trans. Circuits Syst. Video Technol.*, vol. 7, no. 2, pp. 287–298, Apr. 1997.

[36] Y. Zhang, X. Ji, H. Wang, and Q. Dai, “Stereo interleaving video coding with content adaptive image subsampling,” *IEEE Trans. Circuits Syst. Video Technol.*, vol. 23, no. 7, pp. 1097–1108, Jul. 2013.

[37] H. Yang, W. Lin, C. Deng, and L. Xu, “Study on subjective quality assessment of digital compound images,” in *Proc. IEEE Int. Symp. Circuits Syst.*, Jun. 2014, pp. 2149–2152.

[38] *Spatial Scalability Filters*, document ISO/IEC JTC1/SC29/WG11 ITU-T SG 16 Q.6, Jul. 2005.

[39] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, Apr. 2004.

[40] Z. Wang and Q. Li, "Information content weighting for perceptual image quality assessment," *IEEE Trans. Image Process.*, vol. 20, no. 5, pp. 1185–1198, May 2011.

[41] I.-K. Kim, J. Min, T. Lee, W.-J. Han, and J. Park, "Block partitioning structure in the HEVC standard," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 22, no. 12, pp. 1697–1706, Dec. 2012.

[42] *HM13.0*. [Online]. Available: https://hevc.hhi.fraunhofer.de/svn/svn_HEVCSoftware/tags/HM-13.0/, accessed Oct. 2015.

Shiqi Wang (M'15) received the B.S. degree in computer science from Harbin Institute of Technology, Harbin, China, in 2008 and the Ph.D. degree in computer application technology from Peking University, Beijing, China, in 2014.

He was an Intern with Microsoft Research Asia, Beijing, in 2011. He is currently a Post-Doctoral Fellow with the Department of Electrical and Computer Engineering, University of Waterloo, Waterloo, ON, Canada. His research interests include video compression and image/video quality assessment.

Ke Gu received the B.S. degree in electronic engineering from Shanghai Jiao Tong University, Shanghai, China, in 2009, where he is currently working toward the Ph.D. degree.

He was a Visiting Student with the Department of Electrical and Computer Engineering, University of Waterloo, Waterloo, ON, Canada, in 2014. His research interests include quality assessment and contrast enhancement.

Siwei Ma (S'03–M'12) received the B.S. degree from Shandong Normal University, Jinan, China, in 1999 and the Ph.D. degree in computer science from Institute of Computing Technology, Chinese Academy of Sciences, Beijing, China, in 2005.

He held a post-doctoral position with University of Southern California, Los Angeles, CA, USA, from 2005 to 2007. He joined the School of Electronic Engineering and Computer Science, Institute of Digital Media, Peking University, Beijing, where he is currently a Professor. He has authored over 100 technical articles in refereed journals and proceedings in image and video coding, video processing, video streaming, and transmission.

Wen Gao (M'92–SM'05–F'09) received the Ph.D. degree in electronics engineering from The University of Tokyo, Tokyo, Japan, in 1991.

He was a Professor of Computer Science with Harbin Institute of Technology, Harbin, China, from 1991 to 1995, and the Institute of Computing Technology, Chinese Academy of Sciences, Beijing, China. He is currently a Professor of Computer Science with Peking University, Beijing. He has authored five books and over 600 technical articles in refereed journals and conference proceedings in image processing, video coding and communication, pattern recognition, multimedia information retrieval, multimodal interface, and bioinformatics.

Dr. Gao served or serves on the Editorial Board of several journals, such as IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY, IEEE TRANSACTIONS ON MULTIMEDIA, IEEE TRANSACTIONS ON IMAGE PROCESSING, IEEE TRANSACTIONS ON AUTONOMOUS MENTAL DEVELOPMENT, *EURASIP Journal of Image Communications*, and *Journal of Visual Communication and Image Representation*. He was a Chair of a number of prestigious international conferences on multimedia and video signal processing, such as the IEEE International Conference on Multimedia and Expo and ACM Multimedia, and served on the Advisory and Technical Committee of numerous professional organizations.