# Joint Chroma Subsampling and Distortion-Minimization-Based Luma Modification for RGB Color Images With Application

Kuo-Liang Chung, *Senior Member, IEEE*, Tsu-Chun Hsu, and Chi-Chao Huang

***Abstract*— In this paper, we propose a novel and effective hybrid method, which joins the conventional chroma subsampling and the distortion-minimization-based luma modification together, to improve the quality of the reconstructed RGB full-color image. Assume the input RGB full-color image has been transformed to a YUV image, prior to compression. For each 2×2 UV block, one 4:2:0 subsampling is applied to determine the one subsampled U and V components, $U_s$ and $V_s$. Based on $U_s$, $V_s$, and the corresponding 2×2 original RGB block, a main theorem is provided to determine the ideally modified 2×2 luma block in constant time such that the color peak signal-to-noise ratio (CPSNR) quality distortion between the original 2×2 RGB block and the reconstructed 2×2 RGB block can be minimized in a globally optimal sense. Furthermore, the proposed hybrid method and the delivered theorem are adjusted to tackle the digital time delay integration images and the Bayer mosaic images whose Bayer CFA structure has been widely used in modern commercial digital cameras. Based on the IMAX, Kodak, and screen content test image sets, the experimental results demonstrate that in high efficiency video coding, the proposed hybrid method has substantial quality improvement, in terms of the CPSNR quality, visual effect, CPSNR-bitrate trade-off, and Bjøntegaard delta PSNR performance, of the reconstructed RGB images when compared with existing chroma subsampling schemes.**

***Index Terms*— Bayer mosaic image, BD-PSNR, chroma subsampling, CPSNR, distortion-minimization, HEVC, luma modification, rate-distortion curve, RGB full-color image.**

## I. Introduction

PRIOR to compression, the input red-green-blue (RGB) full-color image is first transformed to a YUV image by (1), where $(R_i^{ori},G_i^{ori},B_i^{ori})$ and $(Y_i,U_i,V_i)$ denote the original RGB triple value and the co-located YUV triple value, respectively, at pixel location $i$, $1 \le i \le 4$, in each 2×2 RGB and YUV block-pair based on row-major order. Next, a chroma 4:2:0 subsampling scheme is adopted to subsample the one sampled U and V components, $U_s$ and $V_s$, for each 2×2 UV block. As shown in Fig. 1, the subsampled YUV image is then conveyed into the encoder for compression. Throughout this paper, we only discuss the chroma 4:2:0 subsampling, although our discussion is also applicable to chroma 4:2:2 subsampling. The commonly used seven chroma subsampling schemes are the 4:2:0(L), 4:2:0(R), 4:2:0(A), 4:2:0(DIRECT), 4:2:0(BRIGHT), 4:2:0(BRIGHT_MEAN) [13], and 4:2:0(MPEG-B) [16].

$$\begin{bmatrix} Y_i \\ U_i \\ V_i \end{bmatrix} = \begin{bmatrix} 0.257 & 0.504 & 0.098 \\ -0.148 & -0.291 & 0.439 \\ 0.439 & -0.368 & -0.071 \end{bmatrix} \begin{bmatrix} R_i^{ori} \\ G_i^{ori} \\ B_i^{ori} \end{bmatrix} + \begin{bmatrix} 16 \\ 128 \\ 128 \end{bmatrix} \quad (1)$$

### A. Existing Chroma Subsampling Schemes

The 4:2:0(L) and 4:2:0(R) determine their own $U_s$ and $V_s$ by averaging the U and V components in the left column and right column, respectively, of each 2×2 chroma block. The 4:2:0(A) determines $U_s$ and $V_s$ by averaging all the U and V components of the block, respectively. The 4:2:0(DIRECT) sets $U_s = U_1$ and $V_s = V_1$. The 4:2:0(MPEG-B) applies the 13-tap filter with mask [2,0,-4,-3,5,19,26,19,5,-3,-4,0,2]/64 to estimate $U_s$ and $V_s$. The 4:2:0(BRIGHT) determines $U_s$ and $V_s$ from the position with the largest luma value in the 2×2 luma block. The 4:2:0(BRIGHT_MEAN) is equal to the 4:2:0(BRIGHT) when the ratio of the largest luma value over the smallest one in the 2×2 luma block is larger than 2; otherwise, it is equal to the 4:2:0(A).

### B. Motivation and Contributions

As shown in Fig. 2, this paper proposes an effective hybrid method, which joins the conventional chroma subsampling scheme and the RGB pixel distortion-minimization based luma modification together such that a substantially better reconstructed RGB full-color image is followed. In the proposed method, we retain the subsampled UV image, but modify the Y image to substantially improve the quality of the reconstructed RGB images. Although it causes the loss in the YUV domain due to the luma modification, the gain in the RGB domain makes the reconstructed RGB full-color image better. As a target image for the purpose of display on screens or human visual inspection, increasing the quality of the reconstructed RGB full-color image is important. Based on $U_s$ and $V_s$, the 2×2 original RGB block, and the minimal RGB pixel-distortion criterion, determining the ideally modified

Manuscript received August 18, 2016; revised December 27, 2016, March 12, 2017, and May 4, 2017; accepted June 19, 2017. Date of publication June 26, 2017; date of current version July 18, 2017. This work was supported by the Ministry of Science and Technology of Taiwan under Contract MOST 104-2221-E-011-004-MY3. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Wen Gao. *(Corresponding author: Kuo-Liang Chung.)*

The authors are with the Department of Computer Science and Information Engineering, National Taiwan University of Science and Technology, Taipei 10672, Republic of China (e-mail: klchung01@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/TIP.2017.2719945

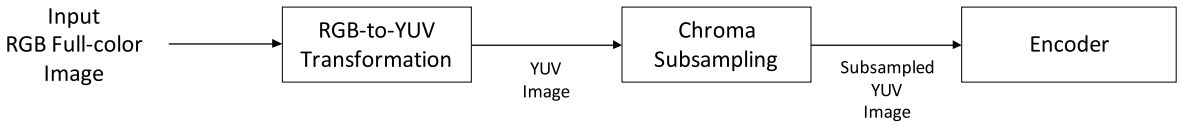



Fig. 1. The conventional chroma subsampling scheme.

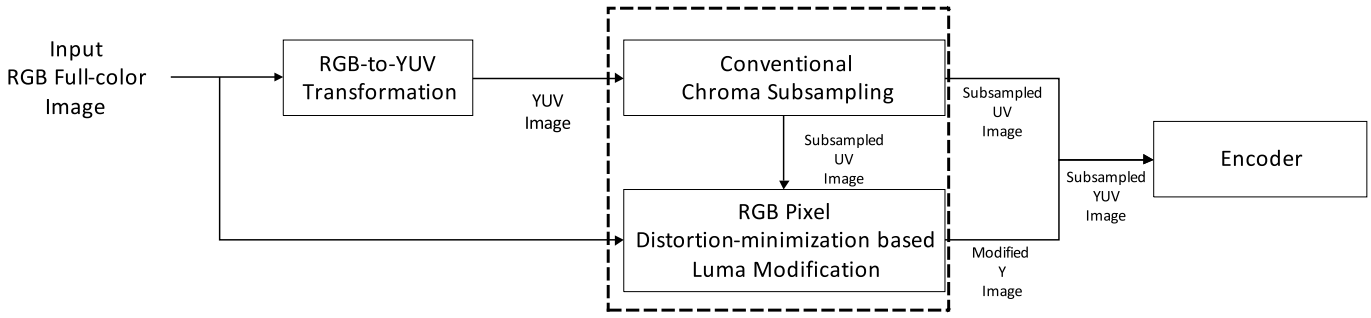



Fig. 2. The proposed hybrid method combining the conventional chroma subsampling and RGB pixel distortion-minimization based luma modification.

luma value for each luma component in a globally optimal sense is a constraint minimization problem. We first apply the differentiation technique to obtain a near-optimal solution of the constraint minimization problem. Further, we investigate the distribution of the error between the near-optimal solution and the ideal, i.e. globally-optimal, one. To remedy the error which occurs in the near-optimal solution and to determine the ideal luma modification, a main theorem and its correctness proof are provided to exactly solve the constraint minimization problem in O(1) time; for the definition of the big-O notation, it is suggested to refer to [6]. Furthermore, the proposed hybrid method and the delivered theorem are adjusted to handle the digital time delay integration (DTDI) image [2], in which each pixel contains only two primary colors, and the Bayer mosaic image [1] in which each pixel contains only one primary color and the Bayer color filter array (CFA) structure in Fig. 6 has been widely used in modern commercial color digital cameras, e.g. the Canon EOS series cameras. Based on the three test image sets, the IMAX test set with 18 images [8], the Kodak test set with 24 images [9], and the other 16 test screen content images [12], several experiments in HEVC are carried out to demonstrate the quality and bitrate-distortion trade-off merits of the proposed hybrid method. When setting the quantization parameter (QP) to be zero in high efficiency video coding (HEVC) [4], we take the COPY and BILINEAR [10], [16] as the chroma reconstruction processes. The average color peak signal-to-noise ratio (CPSNR) gains of the reconstructed RGB full-color images, DTDI images, and Bayer mosaic images by the proposed hybrid method are 1.01dB, 2.25dB, and 14.58dB when compared with the conventional chroma subsampling schemes mentioned in Subsection IV-A, the one by Chung *et al.* [7], and the one by Lin *et al.* [11], respectively. Furthermore, for different QP values, in terms of the visual effect, the CPSNR-bitrate trade-off represented by the rate-distortion (RD) curve, and the Bjøntegaard delta PSNR (BD-PSNR) performance [3], we demonstrate the merits of the proposed hybrid method when compared with the concerned existing chroma subsampling schemes.

The rest of this paper is organized as follows. In Section II, we first transform the ideal luma modification problem to a constraint minimization problem. Then a near-optimal solution by differentiation technique is proposed. In addition, the distribution of the error between the near-optimal solution and the ideal solution is plotted. Further, we point out the challenging problem to obtain the ideal solution in constant time for the constraint minimization problem. In Section III, a main theorem and its correctness proof are provided to determine the ideal solution in O(1) time such that the distortion between the reconstructed RGB pixel value and the original RGB pixel value can be minimized. Furthermore, we describe how to adjust the proposed hybrid method and theorem to tackle the DTDI image and the Bayer mosaic image. In Section IV, several experiments are carried out to demonstrate the objective and subjective quality, quality-bitrate trade-off, and bitrate-distortion trade-off merits of the proposed hybrid method. In Section V, some concluding remarks are addressed.

## II. Transform Ideal Luma Modification to a Constraint Minimization Problem and a Near-optimal Solution

In the first subsection, we transform the determination of the distortion-minimization based ideal luma modification to a constraint minimization problem. Then, based on the differentiation technique, a closed form is derived to determine the modified luma value for each luma component. In the second subsection, we explain why the modified luma value determined by the closed form is only near-optimal. The distribution of the error between the near-optimal solution and the ideal solution, which is obtained by the exhaustive search method, is also plotted. Finally, we highlight the challenge to determine the ideal solution of the constraint minimization problem in constant time.

### A. Transform to a Constraint Minimization Problem

For each 2×2 subsampled YUV block, based on the two subsampled chroma values, $U_s$ and $V_s$, and the modified luma value $Y_i'$ to be determined, the reconstructed RGB triple value $(R_i^{rec}, G_i^{rec}, B_i^{rec})$ can be obtained by

$$\begin{bmatrix} R_i^{rec} \\ G_i^{rec} \\ B_i^{rec} \end{bmatrix} = \begin{bmatrix} 1.164 & 0 & 1.596 \\ 1.164 & -0.391 & -0.813 \\ 1.164 & 2.018 & 0 \end{bmatrix} \begin{bmatrix} Y_i' - 16 \\ U_s - 128 \\ V_s - 128 \end{bmatrix} \quad (2)$$

Since the original chroma components $U_i$ and $V_i$ have been replaced by the subsampled components $U_s$ and $V_s$, the three equalities, $R_i^{rec}$=$R_i^{ori}$,$G_i^{rec}$=$G_i^{ori}$, and $B_i^{rec}$=$B_i^{ori}$, would be violated. In the proposed luma modification, our idea is to view $Y_i'$ as a parameter and our goal is to minimize the following objective function in terms of in the pixel distortion (PD):

$$\begin{aligned} &PD(Y_i') \\ &= (R_i^{ori} - R_i^{rec})^2 + (G_i^{ori} - G_i^{rec})^2 + (B_i^{ori} - B_i^{rec})^2 \\ &= \left\{R_i^{ori} - [1.164(Y_i' - 16) + 1.596(V_s - 128)]\right\}^2 \\ &\quad + \left\{G_i^{ori} - [1.164(Y_i' - 16) - 0.391(U_s - 128) \right. \\ &\qquad \left. -0.813(V_s - 128)]\right\}^2 \\ &\quad + \left\{B_i^{ori} - [1.164(Y_i' - 16) + 2.018(U_s - 128)]\right\}^2, \end{aligned} \quad (3)$$

subject to the color constraint $0 \le R_i^{rec}, G_i^{rec}, B_i^{rec} \le 255$. Although the distortion between the adjusted Y image and the original Y image becomes worse, the modified Y image makes the quality of the reconstructed RGB full-color images, i.e. the target images, substantially better. Our goal is to determine the ideally modified luma value $Y_i^{ideal}$ to solve this constriant minimization problem such that besides the color constraint, it must satisfy

$$\begin{aligned} PD(Y_i^{ideal}) &= \arg\min_{Y} PD(Y) \\ &= \arg\min_{\hat{Y}} \left\{(C_1 - \hat{Y})^2 + (C_2 - \hat{Y})^2 + (C_3 - \hat{Y})^2\right\} \\ &= \arg\min_{\hat{Y}} \left\{PD(\hat{Y})\right\} \end{aligned} \quad (4)$$

where $\hat{Y} = 1.164(Y - 16)$, $C_1 = R_i^{ori} - 1.596(V_s - 128)$, $C_2 = G_i^{ori} + 0.391(U_s - 128) + 0.813(V_s - 128)$, and $C_3 = B_i^{ori} - 2.018(U_s - 128)$.

Equation (3) is similar to the term CMSE in the CPSNR quality metric mentioned in (14). The smaller the value of $CMSE$ is, the higher the value of CPSNR will be. That's the main reason why the ideal luma modification in (4) can improve the quality of the reconstructed RGB full-color images.

### B. Near-Optimal Solution and its Error Distribution

We differentiate the quadratic function $PD(\hat{Y})$ in (4) with respect to $\hat{Y}$ to zero in order to find the critical point of $PD(\hat{Y})$. It yields

$$\frac{\partial PD(\hat{Y})}{\partial \hat{Y}} = 6\hat{Y} - 2(C_1 + C_2 + C_3) = 0. \quad (5)$$

From (5), the critical point occurs at $\hat{Y} = \frac{C_1+C_2+C_3}{3}$. According to the relation $\hat{Y} = 1.164(Y - 16)$, the closed form of the near-optimal solution $Y_i^{near-opt}$, which doesn't consider the color constraint $0 \le R_i^{rec}, G_i^{rec}, B_i^{rec} \le 255$, is equal to

$$\begin{aligned} Y_i^{near-opt} &= (\frac{\hat{Y}}{1.164} + 16) \\ &= \frac{(R_i^{ori} + G_i^{ori} + B_i^{ori})}{3.492} \\ &\quad + (104.339 - 0.446U_s - 0.224V_s). \end{aligned} \quad (6)$$

The near-optimal solution $Y_i^{near-opt}$ in (6) may cause some of the reconstructed RGB triple value, $(R_i^{rec}, G_i^{rec}, B_i^{rec})$, to be out of the range [0,255]. Unfortunately, even if we force each of the infeasible reconstructed RGB triple value to be within the range [0,255], the pixel-distortion minimization criterion in (4) would be violated. This is why $Y_i^{near-opt}$ is only local-optimal, but not ideal.

Here we take the 640th 2×2 RGB block in the first image of the IMAX test set as the example to demonstrate the intuition of the proposed luma modification and the quality merit of the reconstructed RGB values. As shown in Fig 3(a), the original RGB values and the transformed YUV values are given. After running the 4:2:0(A) chroma subsampling and the COPY chroma reconstruction on the UV block, the chroma values have been changed. We consider the upper-left pixel in the reconstructed YUV block, as shown in Fig. 3(b). Since the original chroma values $U_1$ and $V_1$ have been replaced by the subsampled values $U_s$ and $V_s$, respectively, the reconstructed RGB triple value will be undoubtedly different from the original RGB triple value, (136, 253, 188). The reconstructed RGB triple value is (134, 255, 178), in which by (2), the originally transformed G value is larger than 255 and we force the value of $G_1^{rec}$ to 255. Then, by (3), the resultant RGB pixel-distortion is 108, i.e. $PD(Y_1)$ =108. Fig. 3(c) illustrates the reconstructed triple RGB value, (136, 255, 181), by using the reconstructed chroma values as well as the modified luma value determined by the proposed near-optimal luma modification. Similarly, the originally reconstructed G value $G_1^{rec}$ is also out of the range and is forced to 255. Interestingly, it yields a much less RGB pixel-distortion, $PD(Y_1^{near-opt})$=53, which is clearly less than $PD(Y_1)$=108. From Fig. 3(b) and Fig. 3(c), we find that although the distortion between the modified luma value and the original luma value becomes worse, the proposed near-optimal luma modification in (6) makes the reconstructed RGB triple value much better.

In the following, we analyze the distance between the near-optimal pixel-distortion $PD(Y_i^{near-opt})$ and the ideal pixel-distortion $PD(Y_i^{ideal})$, i.e. the value of $(PD(Y_i^{near-opt}) - PD(Y_i^{ideal}))$, according to the exhaustive search method over the range [0,255]. As the first possible ideal solution $Y_i^{ideal}$=0, we put it into (2) and calculate the reconstructed RGB triple value $(R_i^{rec}, G_i^{rec}, B_i^{rec})$. Further, the corresponding pixel-distortion $PD(0)$ is followed by (3). Next, we try $Y_i^{ideal}$=1 and repeat the above process until all the possible ideal solutions over [0,255] have been examined. Finally, among these 256 pixel-distortion values, $PD(0), PD(1), \cdots$, and $PD(255)$, we select the smallest one, say $PD(k)$ for

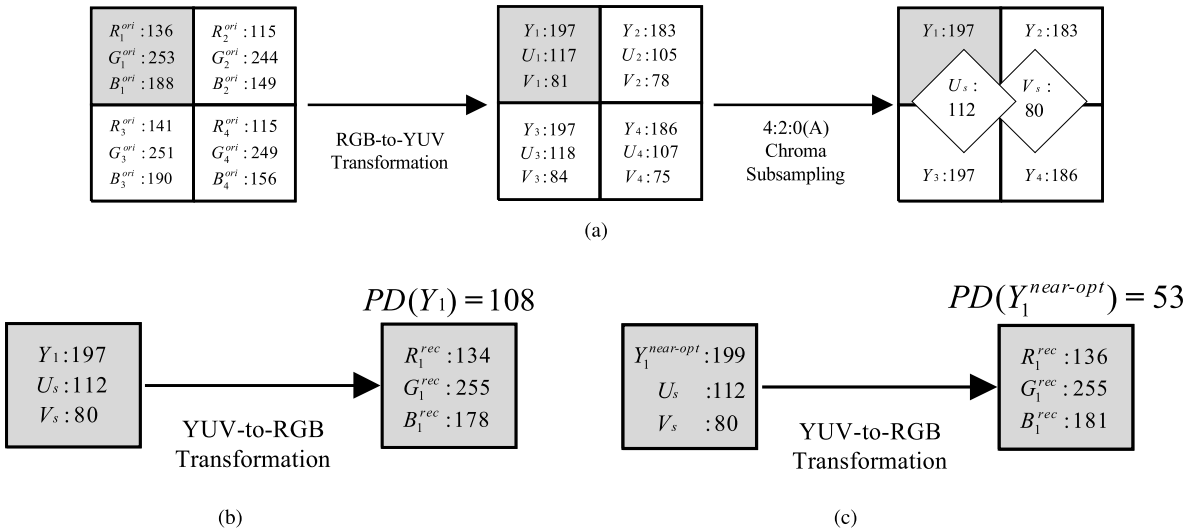



Fig. 3. An example to demonstrate the quality merit of the reconstructed RGB values by the proposed near-optimal luma modification. (a) RGB-to-YUV transformation and 4:2:0(A) chroma subsampling. (b) Reconstructed YUV and RGB values of the upper-left pixel by the conventional method. (c) Reconstructed YUV and RGB values of the upper-left pixel by the proposed near-optimal luma modification.

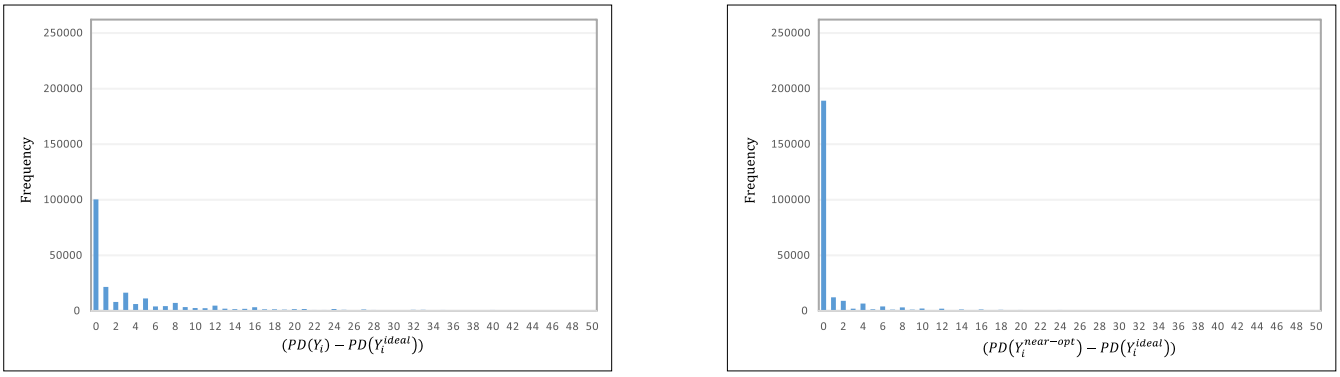



Fig. 4. The distributions of $(PD(Y_i) - PD(Y_i^{ideal}))$ and $(PD(Y_i^{near\text{-}opt}) - PD(Y_i^{ideal}))$. (a) The distribution of $(PD(Y_i) - PD(Y_i^{ideal}))$. (b) The distribution of $(PD(Y_i^{near\text{-}opt}) - PD(Y_i^{ideal}))$.

$0 \le k \le 255$, and $k$ is set to the truly ideal solution, i.e. $Y_i^{ideal}$=k. Taking the upper-left pixel in Fig. 3(a) as the example, it is easy to yield $PD(Y_1)$=$PD(197)$=108, as shown in Fig. 3(b). By (6), we have $Y_1^{near-opt}$=199, and then we have $PD(Y_1^{near-opt})$=$PD(199)$=53, as shown in Fig. 3(c). Consequently, the RGB pixel-distortion reduction, i.e. the gain in the RGB domain, is 55 (=108-53), although the loss in the YUV domain is 4 ($=(199-197)^2$). Most important of all, as a target image for the purpose of display on screens or human visual inspection, increasing the gain in the reconstructed RGB full-color image is the most important thing. By the exhaustive search method, it yields $Y_1^{ideal}$=202 and $PD(Y_1^{ideal})$=$PD(202)$=36. On the other hand, for the upper-left pixel in Fig. 3(a), the RGB pixel-distortion reduction by using the exhaustive search based ideal luma modification is 72 (=108-36). Therefore, the gap between $PD(Y_1^{near-opt})$ and $PD(Y_1^{ideal})$ in this example is 17 (=53-36). It means there is room to improve the near-optimal solution to achieve the ideal solution. In the following, the distributions of the above two gaps are investigated.

Based on the first image in the IMAX test set, Fig. 4(a) and Fig. 4(b) illustrate the distributions of $(PD(Y_i)-PD(Y_i^{ideal}))$ and $(PD(Y_i^{near-opt})-PD(Y_i^{ideal}))$, respectively. In Fig. 4(a) and Fig. 4 (b), the x-axes denote the values of $(PD(Y_i) - PD(Y_i^{ideal}))$ and $(PD(Y_i^{near-opt}) - PD(Y_i^{ideal}))$, respectively, and the y-axes denote the corresponding frequency. Although Fig. 4(b) is more sharp than Fig. 4(a), we clearly observe that in Fig. 4(b), there is still room for improvement. It is a challenge to develop an effective method to narrow the search range such that determining the ideal solution $Y_i^{ideal}$ can be done in constant time rather than that required in the terribly time-consuming exhaustive search method.

## III. Main Theorem for Ideal Luma Modification and Appliction

In this section, a main theorem and its correctness proof are provided to practically determine the ideally modified luma value $Y_i^{ideal}$ in O(1) time for RGB full-color images. In addition, we adjust the theorem to handle the ideal luma

TABLE I
AVERAGE LENGTH OF THE SEARCH RANGE EXAMINED FOR IDEAL LUMA MODIFICATION BASED ON THREE TEST RGB FULL-COLOR IMAGE SETS

| | Average length of the search range |
|---|---|
| IMAX | 6.05 |
| Kodak | 3.29 |
| SCI | 4.63 |
| Average | 4.66 |

modification for the DTDI images, which are captured by line-scan cameras and each pixel contains two primary colors, and the Bayer mosaic images associated with the well-known Bayer CFA structure in which each pixel contains only one primary color.

### A. Main Theorem for Ideal Luma Modification and Correctness Proof

The RGB-to-YUV transformation has the following relation:

$$\begin{bmatrix} R_i^{ori} \\ G_i^{ori} \\ B_i^{ori} \end{bmatrix} = \begin{bmatrix} 1.164 & 0 & 1.596 \\ 1.164 & -0.391 & -0.813 \\ 1.164 & 2.018 & 0 \end{bmatrix} \begin{bmatrix} Y_i - 16 \\ U_i^{ori} - 128 \\ V_i^{ori} - 128 \end{bmatrix} \quad (7)$$

Taking chroma subsampling into account, if the two original chroma components $U_i^{ori}$ and $V_i^{ori}$ in (7) are replaced by $U_s$, and $V_s$, respectively, a unique Y which satisfies (7) does not exist. However, there do exist $Y_i^R$, $Y_i^G$, and $Y_i^B$, which satisfy

$$\begin{aligned} R_i^{ori} &= 1.164(Y_i^R - 16) + 1.596(V_s - 128) \\ G_i^{ori} &= 1.164(Y_i^G - 16) - 0.391(U_s - 128) \\ &\quad -0.813(V_s - 128) \\ B_i^{ori} &= 1.164(Y_i^B - 16) + 2.018(U_s - 128) \end{aligned} \quad (8)$$

After solving each equality in (8), we have

$$\begin{aligned} Y_i^R &= \frac{[R_i^{ori} - 1.596(V_s - 128)]}{1.164} + 16 \\ Y_i^G &= \frac{[G_i^{ori} + 0.391(U_s - 128) + 0.813(V_s - 128)]}{1.164} + 16 \\ Y_i^B &= \frac{[B_i^{ori} - 2.018(U_s - 128)]}{1.164} + 16 \end{aligned} \quad (9)$$

Let $Low_i = \lfloor Min(Y_i^R, Y_i^G, Y_i^B) \rfloor$ and $High_i = \lceil Max(Y_i^R, Y_i^G, Y_i^B) \rceil$, where '$\lfloor\cdot\rfloor$' and '$\lceil\cdot\rceil$' denote the floor function and ceiling function, respectively. We have the following main theorem to determine the ideal solution $Y_i^{ideal}$ in the search range $[Low_i, High_i]$. Based on the three test image sets, Table I indicates that on average, the length of the search range examined is 4.66. Therefore, in practice, the ideal luma modification for each luma pixel can be done in O(1) time.

*Theorem 1: By our proposed hybrid method for RGB full-color image, for each 2×2 luma block, the ideally modified luma value $Y_i^{ideal}$, $1 \le i \le 4$, can be determined in O(1) time in practice.*

*Proof:* We prove it by contradiction. Without loss of generality, let $Y_i^R < Y_i^G < Y_i^B$. Assume the value of $Y_i^{ideal}$ is out of the range $[Low_i, High_i]$ where $Low_i = \lfloor Y_i^R \rfloor$ and $High_i = \lceil Y_i^B \rceil$, and for simplicity, assume $Y_i^{ideal} < \lfloor Y_i^R \rfloor$. The following proof is also applicable to $Y_i^{ideal} > \lceil Y_i^B \rceil$. Our target aims to show that the assumption '$Y_i^{ideal} < \lfloor Y_i^R \rfloor$' will be false since it will violate the minimal pixel-distortion criterion. Because the ideally modified luma value $Y_i^{ideal}$ must satisfy the minimal pixel-distortion criterion, we have $PD(\lfloor Y_i^R \rfloor) > PD(Y_i^{ideal})$, i.e. $PD(\lfloor Y_i^R \rfloor) - PD(Y_i^{ideal}) > 0$. By (3), (9), and some algebraic manipulations, we have

$$\begin{aligned} &\left[(Y_i^R - \lfloor Y_i^R \rfloor)^2 + (Y_i^G - \lfloor Y_i^R \rfloor)^2 + (Y_i^B - \lfloor Y_i^R \rfloor)^2\right] \\ &\quad - \left[(Y_i^R - Y_i^{ideal})^2 + (Y_i^G - Y_i^{ideal})^2 + (Y_i^B - Y_i^{ideal})^2\right] \\ &= \left[(Y_i^R - \lfloor Y_i^R \rfloor)^2 - (Y_i^R - Y_i^{ideal})^2\right] \\ &\quad + \left[(Y_i^G - \lfloor Y_i^R \rfloor)^2 - (Y_i^G - Y_i^{ideal})^2\right] \\ &\quad + \left[(Y_i^B - \lfloor Y_i^R \rfloor)^2 - (Y_i^B - Y_i^{ideal})^2\right] \\ &> 0 \end{aligned} \quad (10)$$

Based on the assumption '$0 \le Y_i^{ideal} < \lfloor Y_i^R \rfloor \le Y_i^R < Y_i^G < Y_i^B$', we have $(Y_i^C - \lfloor Y_i^C \rfloor)^2 - (Y_i^C - Y_i^{ideal})^2 < 0$ for $C \in \{R, G, B\}$. We conclude that

$$\begin{aligned} &\left[(Y_i^R - \lfloor Y_i^R \rfloor)^2 - (Y_i^R - Y_i^{ideal})^2\right] \\ &\quad + \left[(Y_i^G - \lfloor Y_i^R \rfloor)^2 - (Y_i^G - Y_i^{ideal})^2\right] \\ &\quad + \left[(Y_i^B - \lfloor Y_i^R \rfloor)^2 - (Y_i^B - Y_i^{ideal})^2\right] \\ &< 0 \end{aligned} \quad (11)$$

(10) and (11) conflict, which means that the initial assumption '$Y_i^{ideal} < \lfloor Y_i^R \rfloor$' is not true. By the same arguments, we can prove that the assumption '$Y_i^{ideal} > \lceil Y_i^B \rceil$' is still not true. Consequently, the ideally modified luma component $Y_i^{ideal}$ should be in the search range $[Low_i, High_i]$. Because by Table I, the value of $(High_i - Low_i + 1)$ is a constant, 4.66, the ideally modified luma value $Y_i^{ideal}$ can be determined in O(1) time by search over the range $[Low_i, High_i]$. **We complete the proof**.

As for the execution time complexity comparison, let $T_{CLE}$ denote the execution time, in terms of seconds, required to complete the proposed hybrid method, e.g. joining the 4:2:0(A) and the ideal luma modification, and the encoding process; let $T_{CE}$ denote the execution time required to complete the chroma 4:2:0(A) subsampling and the encoding process. Based on the three test image sets, we have $T_{CLE} = 3.47$ and $T_{CE} = 3.26$. Therefore, the execution time overhead required in the ideal luma modification is only 6.4% (=0.21/3.26), which is rather small. It is easily verified that the ratio of the average search range considered in the exhaustive search method over the one considered in Theorem 1 is 55 (=256/4.66), and it indicates the impracticability of the exhaustive search method. On the contrary, the proposed ideal luma modification in Theorem 1 is fast and could meet the real-time demand. At the end of this subsection, we report the CPSNR gain of the ideal luma modification over the near-optimal luma modification when we set QP=0 in the compression and the chroma reconstruction is the COPY.

TABLE II

AVERAGE CPSNR GAINS OF THE NEAR-OPTIMAL AND IDEAL SOLUTION OVER 4:2:0(A) FOR THE THREE TEST RGB FULL-COLOR IMAGE SETS

| | Conventional 4:2:0(A) | Near-optimal solution | Ideal solution |
|---|---|---|---|
| IMAX | 37.62 | 38.16 | **38.47** |
| Kodak | 43.95 | 44.46 | **44.84** |
| SCI | 35 | 35.22 | **35.89** |
| Average CPSNR | 38.86 | 39.28 | **39.73** |
| Average gain | | 0.42 | **0.87** |

Based on the three test image sets and the combination 4:2:0(A)-COPY, Table II indicates that the CPSNR gains of the proposed near-optimal luma modification against the conventional one and the proposed ideal luma modification against the conventional one are 0.42 (=39.28-38.86)dB and 0.87 (=39.73-38.86)dB, respectively. In Section IV, for different QP values, the CPSNR, PSNR, MSE (mean square error), visual effect, RD curves, and BD-PSNR performance will be illustrated to demonstrate the related advantages of the proposed hybrid method when compared with the conventional and related chroma subsampling schemes.

### *B. Application to DTDI and Bayer Mosaic Images*

In this subsection, the proposed hybrid method and Theorem 1 are adjusted to handle the DTDI and Bayer mosaic images such that their ideally modified luma values can also be determined in O(1) time and can be used to substantially improve the quality of the reconstructed images.

*1) Application to DTDI Images:* As shown in Fig. 5, the DTDI image [2] has 4:2:2 (G:B:R) color component composition in each 2×2 block. For each 2×2 DTDI block, the four pixels contain (G,B), (G,R), (G,B), and (G,R) color-compositions based on the row-major order. In each block, the two B colors are arranged in the left column of the block and the two R colors are arranged in the right column. The DTDI structure is the image format used in high-speed scan-line industry printers. According to the RGB-to-YUV transformation in (2), we observe that the B color is affected by the Y and U components and the R color is affected by the Y and V components. Therefore, in [7], the subsampled U component, $U_s$, is preferable for performing 4:2:0(L) on the 2×2 U block, while the subsampled component, $V_s$, is preferable for running 4:2:0(R) on the 2×2 V block. Following the similar discussion in the last subsection, let

$$low_i = \begin{cases} \lfloor \min(Y_1^G, Y_1^B) \rfloor & \text{for } i = 1 \\ \lfloor \min(Y_2^G, Y_2^R) \rfloor & \text{for } i = 2 \\ \lfloor \min(Y_3^G, Y_3^B) \rfloor & \text{for } i = 3 \\ \lfloor \min(Y_4^G, Y_4^R) \rfloor & \text{otherwise} \end{cases}$$

$$high_i = \begin{cases} \lceil \max(Y_1^G, Y_1^B) \rceil & \text{for } i = 1 \\ \lceil \max(Y_2^G, Y_2^R) \rceil & \text{for } i = 2 \\ \lceil \max(Y_3^G, Y_3^B) \rceil & \text{for } i = 3 \\ \lceil \max(Y_4^G, Y_4^R) \rceil & \text{otherwise} \end{cases} \quad (12)$$

Based on the three DTDI test image sets, $\text{DTDI}_{\text{IMAX}}$, $\text{DTDI}_{\text{Kodak}}$, and $\text{DTDI}_{\text{SCI}}$, which are generated from the IMAX, Kodak, and SCI test image sets, respectively, according to the DTDI structure, Table III indicates that the average value of $(High_i - Low_i + 1)$ is 3.67. Following the similar proving technique as in Theorem 1, we have the following ideal luma modification result for DTDI images.

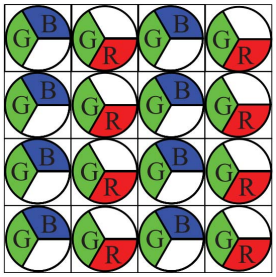


Fig. 5. DTDI structure.

TABLE III

AVERAGE LENGTH OF THE SEARCH RANGE EXAMINED FOR IDEAL LUMA MODIFICATION BASED ON THREE TEST DTDI IMAGE SETS

| | Average length of the search range |
|---|---|
| $\text{DTDI}_{\text{IMAX}}$ | 4.31 |
| $\text{DTDI}_{\text{Kodak}}$ | 3.11 |
| $\text{DTDI}_{\text{SCI}}$ | 3.59 |
| Average | 3.67 |

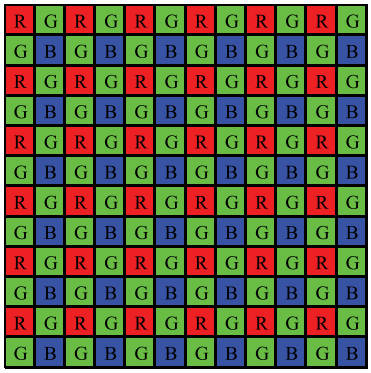


Fig. 6. Bayer CFA structure.

*Theorem 2: By our proposed hybrid method for DTDI image, for each 2×2 luma block, the ideally modified luma value $Y_i^{ideal}$, $1 \le i \le 4$, can be determined in O(1) time in practice.*

*2) Application to Bayer Mosaic Images:* As shown in Fig. 6, for saving hardware costs and preserving more brightness, the Bayer CFA structure has 2:1:1 (G:B:R) color component composition in each 2×2 block. In the demosaicking-first mosaic compression scheme, Yang *et al.* [14] presented a chroma subsampling strategy to determine $U_s$ and $V_s$ for each 2×2 chroma block. The main idea in their proposed strategy is that the significance of the U and V components for reconstructing R and B pixels is considered according to the CFA structure. Later, Lin *et al.* [11] presented a chroma subsampling strategy to determine $U_s$ and $V_s$ by minimizing the quality distortion between the original 2×2 Bayer CFA block and the reconstructed 2×2 Bayer CFA block. The experimental results show that the chroma subsampling strategy by

Lin *et al.* has better quality when compared with the one by Yang *et al.*

According to the 2×2 Bayer CFA structure and the similar discussion in Subsection III-A, let

$$low_i = \begin{cases} \lfloor Y_1^R \rfloor & \text{for } i = 1 \\ \lfloor Y_2^G \rfloor & \text{for } i = 2 \\ \lfloor Y_3^G \rfloor & \text{for } i = 3 \\ \lfloor Y_4^B \rfloor & \text{otherwise} \end{cases}$$

$$high_i = \begin{cases} \lceil Y_1^R \rceil & \text{for } i = 1 \\ \lceil Y_2^G \rceil & \text{for } i = 2 \\ \lceil Y_3^G \rceil & \text{for } i = 3 \\ \lceil Y_4^B \rceil & \text{otherwise} \end{cases} \quad (13)$$

By (13), we have $(High_i - Low_i + 1) = (\lceil Y_i^C \rceil - \lfloor Y_i^C \rfloor + 1) = 2$ for $C \in \{R, G, B\}$. Following the similar proving technique in Theorem 1, we have the following result. Besides the Bayer CFA structure, the proposed result of this part can be applied to the other CFA structures, such as the modified Bayer CFA structure, HVS CFA structure, etc.

*Theorem 3: By our proposed hybrid method for Bayer mosaic image, for each 2×2 luma block, the ideally modified luma value $Y_i^{ideal}$, $1 \le i \le 4$, can be determined in the range $[Low_i, High_i]$, as shown in (13), and completed in O(1) time in practice.*

The proposed ideal luma modification in Theorem 3 does improve the CPSNR performance of the reconstructed Bayer mosaic image by Lin *et al.* [11]. Besides, the proposed method also outperforms the one by Chiu *et al.* [5], whose luma modification formula like (9) cannot guarantee that the modified luma value is ideal, although it is in the range $[Low_i, High_i]$. In addition, Chiu *et al.*'s result cannot be extended to tackle both the DTDI images and the RGB full-color images.

Based on the three DTDI and Bayer test image sets, the average CPSNR and PSNR values of the proposed hybrid method and the two state-of-the-art chroma subsampling strategies are illustrated in Table IV. For DTDI images, we observe that when setting QP=0 in the compression, the CPSNR gains of our proposed hybrid method, 'Proposed', are 2.31dB and 2.19dB, respectively, when compared with the method by Chung *et al.* under the COPY and BILINEAR chroma reconstruction processes. For Bayer mosaic images, when compared with Lin *et al.*'s chroma subsampling scheme, our proposed hybrid method has PSNR gains, 12.5dB and 16.66dB, respectively.

From the observation in Tables II and IV, we conclude that the fewer the number of colors in each pixel, the greater the quality improvement by the proposed hybrid method.

## IV. Experimental Results

The three test image sets are the IMAX test set with 18 images, of which each is of size 500×500, the Kodak test set with 24 images, of which each is of size 768×512, and the 16 screen content images (SCIs), each with size 1024×512, as shown in Fig. 7. All the concerned experiments are implemented on a computer with an Intel Core i7-6700 CPU 3.4 GHz and 8 GB RAM. The operating system is the Microsoft Windows 7 64-bit operating system. The program development environment is Visual C++ 2013. The HEVC reference software used for image compression is HM-16.4.

TABLE IV

Average CPSNR and PSNR Gains of the Proposed Hybrid Method for DTDI and Bayer Images when Compared With Two State-of-the-Art Chroma Subsampling Schemes

| DTDI | | |
|---|---|---|
| | Chung *et al.* [7]-COPY | Proposed-COPY |
| $DTDI_{IMAX}$ | 41.64 | **43.69** |
| $DTDI_{Kodak}$ | 47.48 | **49.51** |
| $DTDI_{SCI}$ | 38.94 | **41.79** |
| Average CPSNR | 42.69 | **45.00** |
| Average gain | | **2.31** |

| DTDI | | |
|---|---|---|
| | Chung *et al.* [7]-BILINEAR | Proposed-BILINEAR |
| $DTDI_{IMAX}$ | 41.28 | **43.18** |
| $DTDI_{Kodak}$ | 48.33 | **50.29** |
| $DTDI_{SCI}$ | 38.96 | **41.69** |
| Average CPSNR | 42.86 | **45.05** |
| Average gain | | **2.19** |

| Bayer | | |
|---|---|---|
| | Lin *et al.* [11]-COPY | Proposed-COPY |
| $Bayer_{IMAX}$ | 43.14 | **55.59** |
| $Bayer_{Kodak}$ | 48.26 | **56.87** |
| $Bayer_{SCI}$ | 40.39 | **56.81** |
| Average PSNR | 43.93 | **56.43** |
| Average gain | | **12.50** |

| Bayer | | |
|---|---|---|
| | Lin *et al.* [11]-BILINEAR | Proposed-BILINEAR |
| $Bayer_{IMAX}$ | 37.06 | **55.35** |
| $Bayer_{Kodak}$ | 44.50 | **56.77** |
| $Bayer_{SCI}$ | 35.57 | **54.98** |
| Average PSNR | 39.04 | **55.70** |
| Average gain | | **16.66** |

In Subsection IV-A, when setting QP=0, we demonstrate the CPSNR gains and the color mean square error (CMSE) reductions of the proposed hybrid method when compared with the conventional chroma subsampling schemes under the COPY and BILINEAR chroma reconstruction processes for the RGB full-color images. For the Bayer mosaic images, the PSNR gains and MSE reductions of the proposed hybrid method are also illustrated when compared with Lin *et al.*'s subsampling scheme. In Subsection IV-B, we demonstrate the visual merit of the proposed hybrid method. In Subsection IV-C, for different QPs, the experimental results are reported to demonstrate that, in terms of the RD curves and BD-PSNR performance, the proposed hybrid method also has better performance when compared with the conventional subsampling schemes.

### A. PSNR Gain and MSE Reduction Merits

To evaluate the quality of the reconstructed RGB full-color images, the CPSNR metric is defined by

$$\text{CPSNR} = \frac{1}{N}\sum_{n=1}^{N} 10\log_{10}\frac{255^2}{CMSE} \quad (14)$$

with $CMSE = \frac{1}{3WH}\sum_{p\in P}\sum_{C\in\{R,G,B\}}[I_{n,C}^{ori,RGB}(p) - I_{n,C}^{rec,RGB}(p)]^2$ in which $\text{P} = \{(l,m)|1 \le l \le H, 1 \le m \le W\}$

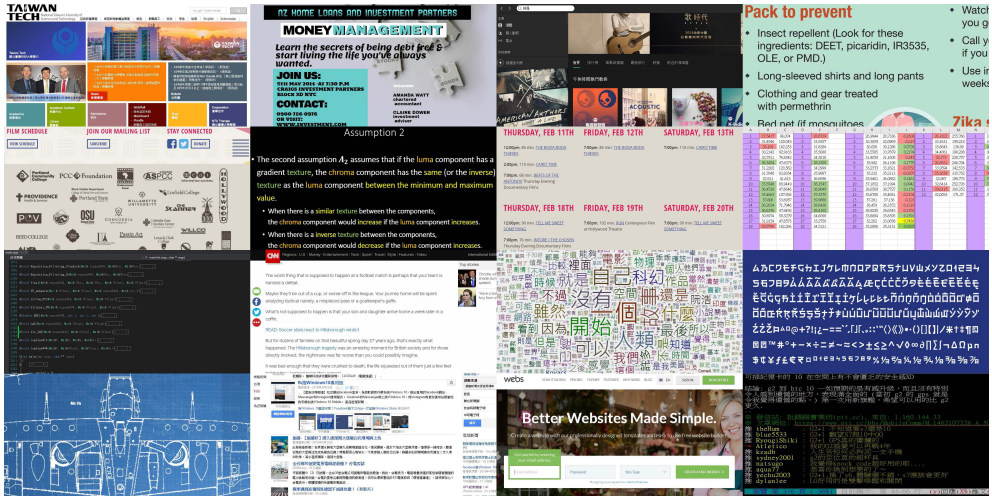



Fig. 7. Sixteen screen content test images.

TABLE V

AVERAGE CPSNR GAINS AND CMSE REDUCTIONS (IN PARENTHESES) OF THE PROPOSED HYBRID METHOD FOR RGB FULL-COLOR IMAGES WHEN COMPARED WITH SEVEN EXISTING CHROMA SUBSAMPLING SCHEMES

| COPY | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | 4:2:0(A) | 4:2:0(L) | 4:2:0(R) | 4:2:0 (DIRECT) | 4:2:0 (BRIGHT) | 4:2:0 (BRIGHT_MEAN) | 4:2:0 (MPEG-B) | Average CPSNR gain (Average CMSE reduction) |
| IMAX | Conventional | 37.62 (15.71) | 36.12 (22.29) | 36.07 (22.52) | 34.80 (30.15) | 33.24 (31.87) | 37.40 (16.38) | 35.19 (27.82) | |
| | Proposed | **38.47 (12.85)** | **37.03 (18.02)** | **36.95 (18.27)** | **35.72 (24.26)** | **35.66 (24.88)** | **38.34 (13.12)** | **36.08 (22.53)** | **1.11 (-4.69)** |
| Kodak | Conventional | 43.95 (2.89) | 42.63 (3.95) | 42.58 (4.00) | 41.22 (5.51) | 40.85 (6.06) | 43.54 (3.23) | 41.87 (4.73) | |
| | Proposed | **44.84 (2.35)** | **43.54 (3.20)** | **43.40 (3.30)** | **42.12 (4.47)** | **41.42 (5.31)** | **44.34 (2.66)** | **42.74 (3.87)** | **0.83 (-0.74)** |
| SCI | Conventional | 35.00 (30.10) | 33.51 (42.57) | 33.41 (43.83) | 32.36 (55.50) | 32.38 (57.56) | 33.62 (42.72) | 32.98 (46.10) | |
| | Proposed | **35.89 (24.82)** | **34.43 (35.01)** | **34.34 (35.89)** | **33.27 (45.58)** | **33.66 (48.37)** | **35.03 (34.16)** | **33.88 (37.89)** | **1.04 (-8.09)** |
| Average | Conventional | 38.86 (16.23) | 37.42 (22.94) | 37.35 (23.45) | 36.12 (30.39) | 35.49 (31.83) | 38.19 (20.78) | 36.68 (26.21) | |
| | Proposed | **39.73 (13.34)** | **38.33 (18.74)** | **38.23 (19.15)** | **37.04 (24.77)** | **36.92 (26.19)** | **39.24 (16.65)** | **37.56 (21.43)** | **0.99 (-4.51)** |

| BILINEAR | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | 4:2:0(A) | 4:2:0(L) | 4:2:0(R) | 4:2:0 (DIRECT) | 4:2:0 (BRIGHT) | 4:2:0 (BRIGHT_MEAN) | 4:2:0 (MPEG-B) | Average CPSNR gain (Average CMSE reduction) |
| IMAX | Conventional | 36.35 (20.74) | 37.36 (16.24) | 34.30 (33.94) | 38.50 (11.99) | 33.25 (31.99) | 36.28 (21.09) | 38.23 (13.02) | |
| | Proposed | **37.19 (16.99)** | **38.19 (13.31)** | **35.17 (27.62)** | **39.32 (09.82)** | **35.58 (25.39)** | **37.16 (17.09)** | **39.06 (10.66)** | **1.06 (-4.02)** |
| Kodak | Conventional | 44.33 (2.65) | 45.66 (1.97) | 41.72 (4.91) | 47.69 (1.29) | 41.75 (4.94) | 43.98 (2.92) | 47.28 (1.39) | |
| | Proposed | **45.32 (2.09)** | **46.83 (1.47)** | **42.56 (4.05)** | **49.30 (0.85)** | **42.37 (4.29)** | **44.88 (2.35)** | **48.60 (0.98)** | **1.06 (-0.57)** |
| SCI | Conventional | 34.06 (37.37) | 34.63 (34.61) | 32.09 (57.42) | 35.11 (33.94) | 32.18 (57.14) | 33.08 (45.82) | 35.10 (32.45) | |
| | Proposed | **34.96 (30.98)** | **35.53 (28.70)** | **33.02 (47.18)** | **35.99 (28.25)** | **33.23 (47.98)** | **34.26 (36.99)** | **35.99 (26.89)** | **0.96 (-7.40)** |
| Average | Conventional | 38.25 (20.25) | 39.22 (17.60) | 36.03 (32.09) | 40.43 (15.74) | 35.73 (31.36) | 37.78 (23.28) | 40.21 (15.62) | |
| | Proposed | **39.15 (16.69)** | **40.18 (14.49)** | **36.91 (26.28)** | **41.54 (12.97)** | **37.06 (25.89)** | **38.77 (18.81)** | **41.22 (12.84)** | **1.03 (-3.99)** |

denotes the set of pixel coordinates in one W×H image; N denotes the number of test images used, e.g. N=18 for the IMAX image set; $I_{n,C}^{ori,RGB}(p)$ and $I_{n,C}^{rec,RGB}(p)$ denote the C-color value of the pixel at position $p$ in the $n$th original RGB full-color image and the reconstructed one, respectively.

Based on the three RGB test image sets, Table V indicates that under the COPY and the BILINEAR chroma reconstruction processes, the proposed hybrid method always has better CPSNR performance marked in boldface when compared with any of the conventional ones. On average, the CPSNR gains (CMSE reductions) of the two combinations, Proposed-COPY and Proposed-BILINEAR, are 0.99 (4.51) dB and 1.03 (3.99) dB, respectively, when compared with the conventional combinations. Table VI shows the average MSE reductions of the proposed hybrid method for Bayer mosaic

TABLE VI

AVERAGE MSE REDUCTIONS OF THE PROPOSED HYBRID METHOD FOR BAYER MOSAIC IMAGES WHEN COMPARED WITH LIN *et al.*'S CHROMA SUBSAMPLING SCHEME

| | | COPY | | BILINEAR | |
|---|---|---|---|---|---|
| | | MSE | Average MSE reduction | MSE | Average MSE reduction |
| $\text{Bayer}_{\text{IMAX}}$ | Lin *et al.* | 3.74 | | 17.87 | |
| | Proposed | **0.16** | **-3.58** | **0.19** | **-17.67** |
| $\text{Bayer}_{\text{Kodak}}$ | Lin *et al.* | 0.98 | | 2.55 | |
| | Proposed | **0.13** | **-0.84** | **0.14** | **-2.41** |
| $\text{Bayer}_{\text{SCI}}$ | Lin *et al.* | 10.60 | | 23.97 | |
| | Proposed | **0.22** | **-10.38** | **0.53** | **-23.44** |
| Average | Lin *et al.* | 5.11 | | 14.79 | |
| | Proposed | **0.17** | **-4.93** | **0.29** | **-14.51** |

images when compared with the chroma subsampling scheme by Lin *et al.* Since the average PSNR gains of the proposed hybrid method over Lin *et al.*'s subsampling scheme for the reconstructed Bayer mosaic images have been shown in Table IV, we only demonstrate the MSE reductions of the proposed hybrid method in Table VI.

Here, the PSNR metric is defined by

$$\text{PSNR} = \frac{1}{N}\sum_{n=1}^{N} 10\log_{10}\frac{255^2}{MSE} \tag{15}$$

with $MSE = \frac{1}{WH}\sum_{p\in P}[I_n^{ori,Bayer}(p) - I_n^{rec,Bayer}(p)]^2$ in which $\text{P} = \{(l,m)|1 \le l \le H, 1 \le m \le W\}$ denotes the set of pixel coordinates in one W×H Bayer mosaic image, $I_n^{ori,Bayer}(p)$ and $I_n^{rec,Bayer}(p)$ denote the color values of the pixels at position $p$ in the $n$th original Bayer mosaic image and the reconstructed analogue.

In (14), the CPSNR definition reveals that all the three components, R, G, and B, have been considered to be of the same importance. However, sometimes people consider one component, say G component, to be more important than the other two components, i.e. R and B components. Consequently, based on this color component preference, the traditional CPSNR metric in (14) can be extended to

$$\text{CPSNR}^P = \frac{1}{N}\sum_{n=1}^{N} 10\log_{10}\frac{255^2}{CMSE^P} \tag{16}$$

with $CMSE^P = \sum_{l\in L}\alpha[I_n^{ori,R}(l) - I_n^{rec,R}(l)]^2 + \beta[I_n^{ori,G}(l) - I_n^{rec,G}(l)]^2 + \gamma[I_n^{ori,B}(l) - I_n^{rec,B}(l)]^2$ in which $\text{L} = \{(i,j)|1 \le i \le H, 1 \le j \le W\}$ denotes the set of pixel coordinates in one W×H image, and $1 = \alpha + \beta + \gamma$ for $0 < \alpha, \beta, \gamma < 1$. When setting $\alpha = \beta = \gamma = 1/3$ in (16), we have CPSNR=$\text{CPSNR}^P$. Suppose one prefers G component with weight 0.6 rather than R and B components with weight 0.3 and 0.1, respectively, and then one can set $\alpha$=0.3, $\beta$=0.6, and $\gamma$=0.1. Generally, following the color component preference, the pixel distortion term in (3) can be modified to

$$PD^P(Y_i') = 3\alpha(R_i^{ori} - R_i^{rec})^2 + 3\beta(G_i^{ori} - G_i^{rec})^2 + 3\gamma(B_i^{ori} - B_i^{rec})^2. \tag{17}$$

In Appendix, the value of $Y_i^{ideal}$ can be determined in constant time such that the pixel distortion with color component preference $PD^P(Y_i')$ can be minimized. To be nothing surprising, the value of the finally determined $Y_i^{ideal}$ and the reconstructed image pixel with color component preference would be different from the ones without color component preference. Based on the three color component preferences, say ($\alpha$, $\beta$, $\gamma$)=(0.3, 0.6, 0.1), ($\alpha$, $\beta$, $\gamma$)=(0.4, 0.4, 0.2), and ($\alpha$, $\beta$, $\gamma$)=($0.\bar{3}$, $0.\bar{3}$, $0.\bar{3}$), Table VII illustrates the performance of each color component for the three color component preferences. It will be an interesting research issue to model the color component preference based on the social media to satisfy both the learning-based color component preference and the distortion-minimization criterion.

TABLE VII

THE PERFORMANCE OF EACH COLOR COMPONENT FOR THREE COLOR COMPONENT PREFERENCES

| $\alpha = 0.3, \beta = 0.6, \gamma = 0.1$ | | | | | | |
|---|---|---|---|---|---|---|
| | Conventional | | | Proposed | | |
| | R | G | B | R | G | B |
| IMAX | 38.23 | 43.38 | 35.13 | 38.67 | 43.73 | 35.1 |
| Kodak | 44.02 | 49.13 | 41.95 | 44.28 | 49.63 | 42.23 |
| SCI | 35.18 | 42.05 | 33.07 | 35.92 | 43.16 | 33.31 |

| $\alpha = 0.4, \beta = 0.4, \gamma = 0.2$ | | | | | | |
|---|---|---|---|---|---|---|
| | Conventional | | | Proposed | | |
| | R | G | B | R | G | B |
| IMAX | 38.23 | 43.38 | 35.13 | 39.97 | 40.68 | 36 |
| Kodak | 44.02 | 49.13 | 41.95 | 45.33 | 47.1 | 42.89 |
| SCI | 35.18 | 42.05 | 33.07 | 36.56 | 40.05 | 33.68 |

| $\alpha = 0.\bar{3}, \beta = 0.\bar{3}, \gamma = 0.\bar{3}$ | | | | | | |
|---|---|---|---|---|---|---|
| | Conventional | | | Proposed | | |
| | R | G | B | R | G | B |
| IMAX | 38.23 | 43.38 | 35.13 | 38.73 | 39.72 | 37.54 |
| Kodak | 44.02 | 49.13 | 41.95 | 44.15 | 46.63 | 44.39 |
| SCI | 35.18 | 42.05 | 33.07 | 35.09 | 39.56 | 35.2 |

### *B. Visual Effect Merit*

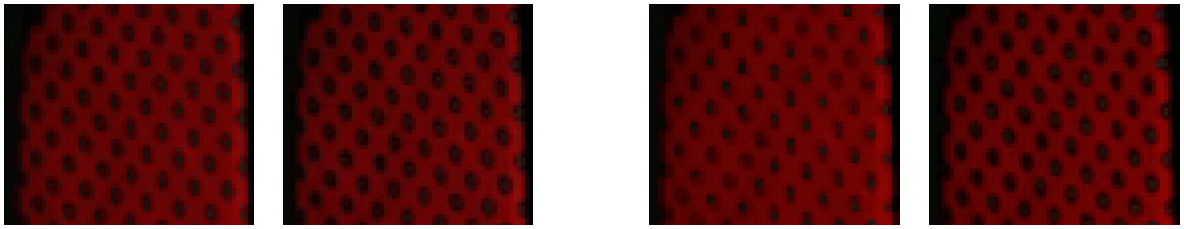
As shown in Fig. 8(a), we take the amplified sub-image of one IMAX sub-image highlighted by the red rectangle as the comparison basis. Figs. 8(b)-(c) depict the reconstructed RGB full-color sub-images by the traditional combination 'Conventional 4:2:0(A)-COPY' and our proposed combination 'Proposed 4:2:0(A)-COPY', respectively, where 'Conventional 4:2:0(A)-COPY' denotes the combination 4:2:0(A)-COPY and 'Proposed 4:2:0(A)-COPY' denotes the combination joining the proposed luma modification based 4:2:0(A) and the COPY reconstruction process. We observe that the proposed combination ensures that the black spots in the reconstructed RGB sub-image are almost the same as those in the original RGB sub-image, while the traditional combination suffers from the blurred black spots. As shown in Fig. 8(d), the traditional

TABLE VIII

AVERAGE BD-PSNR COMPARISON OF THE PROPOSED HYBRID METHOD FOR RGB FULL-COLOR IMAGES OVER THE SEVEN EXISTING CHROMA SUBSAMPLING SCHEMES

| COPY | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Image test set | QP interval | 4:2:0(A) | 4:2:0(L) | 4:2:0(R) | 4:2:0 (DIRECT) | 4:2:0 (BRIGHT) | 4:2:0 (BRIGHT_MEAN) | 4:2:0 (MPEG-B) |
| IMAX | 0-12 | **0.6961** | **0.7766** | **0.7561** | **0.8147** | **0.9922** | **0.7797** | **0.7764** |
| | 12-24 | **0.3260** | **0.4050** | **0.3964** | **0.4631** | **0.6011** | **0.3689** | **0.4453** |
| | 24-36 | **0.0801** | **0.0728** | **0.0639** | **0.0625** | **0.0720** | **0.0546** | **0.1031** |
| | 36-48 | **0.0229** | **0.0148** | **0.0006** | **0.0116** | -0.0135 | **0.0014** | **0.0056** |
| | 40-51 | **0.0171** | **0.0127** | -0.0089 | **0.0104** | -0.0174 | -0.0131 | -0.0086 |
| Kodak | 0-12 | **0.5992** | **0.6539** | **0.5865** | **0.6906** | **0.4246** | **0.5512** | **0.6450** |
| | 12-24 | **0.1924** | **0.2125** | **0.2048** | **0.2384** | **0.1320** | **0.1760** | **0.2307** |
| | 24-36 | **0.0349** | **0.0276** | **0.0427** | **0.0248** | **0.0082** | **0.0303** | **0.0370** |
| | 36-48 | **0.0056** | -0.0037 | **0.0073** | -0.0099 | **0.0056** | -0.0024 | **0.0070** |
| | 40-51 | **0.0014** | -0.0104 | **0.0001** | -0.0111 | **0.0058** | -0.0057 | **0.0023** |
| SCI | 0-12 | **0.8343** | **0.8730** | **0.8951** | **0.8834** | **1.2549** | **1.3612** | **0.8529** |
| | 12-24 | **0.6415** | **0.7181** | **0.7449** | **0.7683** | **1.0910** | **1.1518** | **0.7074** |
| | 24-36 | **0.3244** | **0.3444** | **0.3742** | **0.4081** | **0.5730** | **0.6017** | **0.3986** |
| | 36-48 | **0.0995** | **0.1043** | **0.1147** | **0.0899** | **0.1226** | **0.1367** | **0.1187** |
| | 40-51 | **0.0417** | **0.0502** | **0.0602** | **0.0511** | **0.0658** | **0.0689** | **0.0695** |


(a)



(b) (c) (d) (e)

Fig. 8. The visual effect comparison for the reconstructed RGB color images by 'Conventional 4:2:0(A)-COPY', 'Proposed 4:2:0(A)-COPY', 'Conventional 4:2:0(BRIGHT)-COPY', and 'Proposed 4:2:0(BRIGHT)-COPY'. (a) The original IMAX sub-image and the amplified one. (b) The reconstructed RGB sub-image 'Conventional 4:2:0(A)-COPY'. (c) The reconstructed RGB sub-image 'Proposed 4:2:0(A)-COPY'. (d) The reconstructed RGB sub-image 'Conventional 4:2:0 (BRIGHT)-COPY'. (e) The reconstructed RGB sub-image 'Proposed 4:2:0(BRIGHT)-COPY'.

combination 'Conventional 4:2:0(BRIGHT)-COPY' suffers from shrunken and blurred black spots in the reconstructed RGB sub-image, while the reconstructed RGB sub-image created by our proposed combination 'Proposed 4:2:0(BRIGHT)-COPY' is quite similar to the original one.

### *C. Quality-Bitrate Trade-off Merit*

Based on the three test image sets and the COPY chroma reconstruction process, the RD curves are plotted to demonstrate the quality-bitrate trade-off merit of the proposed hybrid method. First, the bitrate of one compressed image set is defined by

$$\text{bitrate} = \frac{B}{N} \tag{18}$$

where B denotes the total number of bits used to compress one test image set with N images, e.g. N=18 for the IMAX test set. When setting QP=0, 4, 8, 12, 16, 20, 24, 28, 32, 36, 40, 44, 48, and 51 in the compression, the RD curves in Figs. 9-10 demonstrate the quality-bitrate trade-off superiority of the proposed hybrid method for the reconstructed RGB full-color and Bayer mosaic images, respectively, when compared with the conventional chroma subsampling schemes and Lin *et al.*'s scheme. In Figs. 9-10, for QP$\leq$20, the quality gain is clearly visible.

Besides the RD curves comparison in Figs. 9-10, we further provide the BD-PSNR comparison [2] to indicate the average PSNR difference over the same bitrate interval for any two concerned chroma subsampling schemes. For the reconstructed RGB full-color images, we take the conventional chroma subsampling schemes to be the comparison basis and then tabulate the BD-PSNR results of the proposed hybrid method in Table VIII. For the reconstructed Bayer mosaic images, we take Lin *et al.*'s subsampling scheme to be the comparison basis, and Table IX demonstrates the BD-PSNR result of the proposed hybrid method. In Tables VIII-IX, a positive

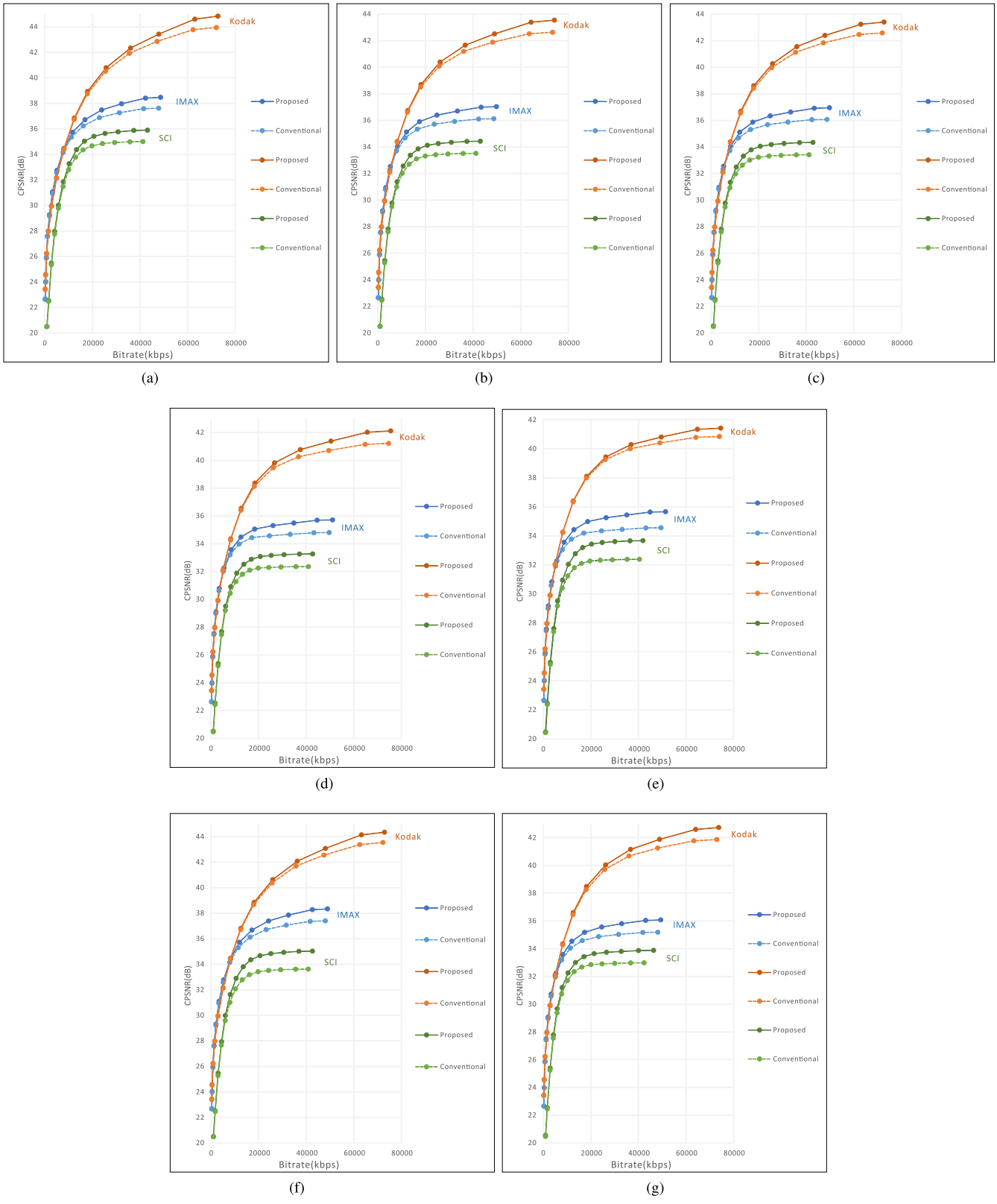



Fig. 9. RD curves of the seven conventional chroma 4:2:0 subsampling schemes and the proposed ones for the three RGB full-color test image sets. (a) RD curves for 4:2:0(A)-COPY. (b) RD curves for 4:2:0(L)-COPY. (c) RD curves for 4:2:0(R)-COPY. (d) RD curves for 4:2:0(DIRECT)-COPY. (e) RD curves for 4:2:0(BRIGHT)-COPY. (f) RD curves for 4:2:0(BRIGHT_MEAN)-COPY. (g) RD curves for 4:2:0(MPEG-B)-COPY.

value of BD-PSNR in boldface indicates the quality superiority of the proposed hybrid method under a fixed QP interval. In Figs. 9-10 and Tables VIII-IX, we observe that under low and middle QP values, the proposed hybrid method outperforms the conventional and Lin *et al.*'s subsampling schemes.

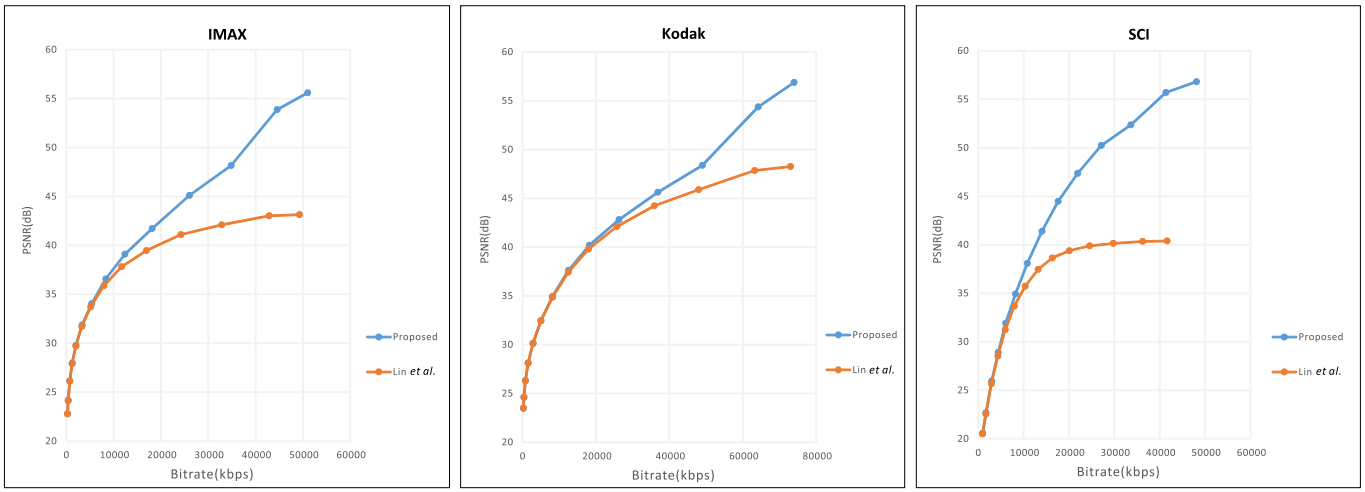



Fig. 10. RD curves of Lin *et al.*'s scheme and the proposed one for the three Bayer test image sets. (a) RD curves for IMAX test image set. (b) RD curves for Kodak test image set. (c) RD curves for SCI test image set.

TABLE IX
AVERAGE BD-PSNR COMPARISON OF LIN *et al.*'S SCHEME AND THE PROPOSED HYBRID METHOD FOR BAYER MOSAIC IMAGES

| COPY | | |
|---|---|---|
| Image test set | QP interval | Proposed |
| $\text{Bayer}_{\text{IMAX}}$ | 0-12 | **6.9061** |
| | 12-24 | **1.4590** |
| | 24-36 | **0.1046** |
| | 36-48 | **0.0031** |
| | 40-51 | -0.0032 |
| $\text{Bayer}_{\text{Kodak}}$ | 0-12 | **3.6778** |
| | 12-24 | **0.4717** |
| | 24-36 | **0.0479** |
| | 36-48 | **0.0021** |
| | 40-51 | **0.0012** |
| $\text{Bayer}_{\text{SCI}}$ | 0-12 | **12.3132** |
| | 12-24 | **6.1124** |
| | 24-36 | **1.4357** |
| | 36-48 | **0.1655** |
| | 40-51 | **0.0734** |

Although the RD curves in Figs. 9-10 and the BD-PSNR performance of the proposed hybrid method decay in high QP, it seldom occurs in practice because encoding images under high QP values often leads to poor visual perception.

## V. CONCLUSION

In this paper, we have presented a novel and effective hybrid method joining the conventional chroma subsampling scheme and luma modification for RGB full-color, Bayer mosaic, and DTDI images, prior to compression. For RGB full-color images, we first explain why determining the ideally modified luma value for each pixel is a constraint minimization problem. Next, a near-optimal solution is presented based on the differentiation technique. To solve the challenging problem, i.e. removing the gap between the near-optimal solution and the ideal solution, a theorem and its correctness are provided to guarantee the modified, i.e. globally optimal, luma value can be determined in constant time practically. Furthermore, we adjust the proposed results to tackle the DTDI and Bayer mosaic images whose CFA structure has been widely used in modern color digital cameras to substantially increase the quality of the reconstructed images. To the best of our knowledge, this is the first time that such a hybrid method is proposed to substantially increase the quality performance of the reconstructed images for the three kinds of images. Based on the IMAX, Kodak, and SCI test image sets, the experimental results demonstrated that, in terms of the CPSNR, MSE, visual effect, RD curve, and BD-PSNR of the reconstructed images, the proposed hybrid method always has the best performance when compared with the related state-of-the-art chroma subsampling schemes. One future research work is to study the effect of using the refined luma component to further refine the co-located chroma component. The other future research work is to model the color component preference based on the social media to achieve the distortion-minimization criterion. The other future research work is to incorporate the results of this paper into the chroma reconstruction issue studied in [15].

## APPENDIX
## THE IDEALLY MODIFIED LUMA VALUE IN (17) CAN BE DETERMINED IN O(1) TIME.

*Proof:* Similar to the proof in Theorem 1, let $Y_i^R < Y_i^G < Y_i^B$, $Low_i = \lfloor Y_i^R \rfloor$, and $High_i = \lceil Y_i^B \rceil$. Assume $Y_i^{ideal} < \lfloor Y_i^R \rfloor$. In what follows, we show that the assumption '$Y_i^{ideal} < \lfloor Y_i^R \rfloor$' is not true. Because $Y_i^{ideal}$ must satisfy the minimal pixel-distortion criterion with color component preference in (17), we have

$$PD^P(\lfloor Y_i^R \rfloor) > PD^P(Y_i^{ideal}). \tag{19}$$

Therefore, it yields

$$\begin{aligned}
&\left[\alpha(Y_i^R - \lfloor Y_i^R \rfloor)^2 + \beta(Y_i^G - \lfloor Y_i^R \rfloor)^2 + \gamma\,(Y_i^B - \lfloor Y_i^R \rfloor)^2\right] \\
&\quad - \left[\alpha(Y_i^R - Y_i^{ideal})^2 + \beta(Y_i^G - Y_i^{ideal})^2 + \gamma\,(Y_i^B - Y_i^{ideal})^2\right] \\
&= \alpha\left[(Y_i^R - \lfloor Y_i^R \rfloor)^2 - (Y_i^R - Y_i^{ideal})^2\right] \\
&\quad + \beta\left[(Y_i^G - \lfloor Y_i^R \rfloor)^2 - (Y_i^G - Y_i^{ideal})^2\right] \\
&\quad + \gamma\left[(Y_i^B - \lfloor Y_i^R \rfloor)^2 - (Y_i^B - Y_i^{ideal})^2\right] \\
&> 0
\end{aligned} \tag{20}$$

Unfortunately, (19) and (20) conflict, and it indicates that the original assumption '$Y_i^{ideal} < \lfloor Y_i^R \rfloor$' is not true. Similarly, we can prove that the assumption '$Y_i^{ideal} > \lceil Y_i^B \rceil$' is still not true. By Table I, we conclude that $Y_i^{ideal}$ should be in a small search range with constant length. In other words, $Y_i^{ideal}$ can be determined in O(1) time by search over the range$[Low_i, High_i]$. **We complete the proof**.

## ACKNOWLEDGMENT

The authors appreciate the proofreading help of Ms. C. Harrington and the valuable comments of the three anonymous reviewers.

## REFERENCES

[1] B. E. Bayer, "Color imaging array," U.S. Patent 3 971 065 A, Jul. 20, 1976.
[2] E. Bodenstorfer *et al.*, "High-speed line-scan camera with digital time delay integration," *Proc. SPIE*, vol. 6496, pp. 64960I-1–64960I-10, Feb. 2007.
[3] G. Bjøntegaard, *Calculation of Average PSNR Differences Between RD-Curves*, document VECG-M33, Austin, TX, USA, Apr. 2001.
[4] B. Bross, W. J. Han, J. R. Ohm, G. J. Sullivan, and T. Wiegand, *High Efficiency Video Coding (HEVC) Text Specification Draft 10*, document JCTVC-L1003, Jan. 2013.
[5] Y. H. Chiu, K. L. Chung, and C. H. Lin, "An improved universal subsampling strategy for compressing mosaic videos with arbitrary RGB color filter arrays in H.264/AVC," *J. Vis. Commun. Image Represent.*, vol. 25, no. 7, pp. 1791–1799, Oct. 2014.
[6] T. H. Cormen, C. E. Leiserson, R. L. Rivest, and C. Stein, *'Asymptotic notation' in Introduction to Algorithms*, 3rd ed. London, U.K.: MIT Press, 2009, sec. 3.1.
[7] K. L. Chung, W. J. Yang, C. H. Chen, H. Y. M. Liao, and S. M. Zeng, "Efficient chroma subsampling strategy for compressing DTDI mosaic video sequences in H.264/AVC," *J. Electron. Imag.*, vol. 20, no. 2, pp. 023011-1–023011-15, 2011.
[8] *IMAX Image Database*, accessed on Jan. 10, 2016. [Online]. Available: http://www4.comp.polyu.edu.hk/~cslzhang/CDM_Dataset.htm
[9] *Kodak Image Database*, accessed on Jan. 10, 2016. [Online]. Available: http://www.math.purdue.edu/~lucier/PHOTO_CD/BMP_IMAGES/
[10] J. Korhonen, "Improving image fidelity by luma-assisted chroma subsampling," in *Proc. IEEE Int. Conf. Multimedia Expo*, Jun./Jul. 2015, pp. 1–6.
[11] C. H. Lin, K. L. Chung, and C. W. Yu, "Novel chroma subsampling strategy based on mathematical optimization for compressing mosaic videos with arbitrary RGB color filter arrays in H.264/AVC and HEVC," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 26, no. 9, pp. 1722–1733, Sep. 2016.
[12] *SCI Image Database*, accessed on Jan. 10, 2016. [Online]. Available: ftp://140.118.175.164/SCI
[13] H. Thoma, M. de Frutos-López, and J. Auer, "Chroma subsampling for HDR video with improved subjective quality," in *Proc. Picture Coding Symp.*, Dec. 2013, pp. 345–348.
[14] W. J. Yang, K. L. Chung, W. N. Yang, and L. C. Lin, "Universal chroma subsampling strategy for compressing mosaic video sequences with arbitrary RGB color filter arrays in H.264/AVC," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 23, no. 4, pp. 591–606, Apr. 2013.
[15] S. Wang, K. Gu, S. Ma, and W. Gao, "Joint chroma downsampling and upsampling for screen content image," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 26, no. 9, pp. 1595–1609, Sep. 2016.
[16] 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.

**Kuo-Liang Chung** (SM'01) received the B.S., M.S., and Ph.D. degrees from National Taiwan University, Taipei, Taiwan, in 1982, 1984, and 1990, respectively. He is currently the Chair Professor with the Department of Computer Science and Information Engineering, National Taiwan University of Science and Technology, Taipei, Taiwan. His current research interests include video coding and image processing. He was a recipient of the Distinguished Research Award from 2004 to 2007 and the Distinguished Research Project Award from 2009 to 2012 from the National Science Council of Taiwan. He has been serving as an Associate Editor of the *Journal of Visual Communication and Image Representation* since 2011.

**Tsu-Chun Hsu** received the B.S. degree in computer science and information engineering from National Taiwan Ocean University, Keelung, Taiwan, in 2015. She is currently pursuing the M.S. degree in computer science and information engineering with the National Taiwan University of Science and Technology, Taipei, Taiwan. Her research interests include image processing and video compression.

**Chi-Chao Huang** received the B.S. degree in computer science and information engineering from National Cheng-Chi University, Taipei, Taiwan, in 2015. He is currently pursuing the M.S. degree in computer science and information engineering with the National Taiwan University of Science and Technology, Taipei. His research interests include camera image processing and video compression.