Effective Gradient Descent-Based Chroma Subsampling Method for Bayer CFA Images in HEVC

Kuo-Liang Chung, *Senior Member, IEEE*, Yu-Ling Lee, and Wei-Che Chien

***Abstract*— The most widely used color filter array (CFA) pattern in commercial digital color cameras is the Bayer pattern, and the captured image is called the Bayer CFA image, in which each pixel contains only one color value and each image consists of 25% red, 50% green, and 25% blue color values. The chroma 4:2:2 or 4:2:0 subsampling of Bayer CFA images is a necessary process prior to compression. According to the block-distortion minimization principle, in this paper, we propose an effective gradient descent-based chroma subsampling (GDCS) method for Bayer CFA images. Based on the test Bayer CFA images collected from the Kodak and IMAX datasets, experimental results demonstrated that in high efficiency video coding, our GDCS method has better quality and quality-bitrate tradeoff performance of the reconstructed images when compared with the existing chroma subsampling methods.**

***Index Terms*— Bayer color filter array (CFA) image, chroma subsampling, gradient descent, high efficiency video coding (HEVC), quality, quality-bitrate tradeoff.**

I. Introduction

TO SAVE hardware costs, most digital color cameras are equipped with a single-sensor covered with a red-green-blue (RGB) Bayer color filter array (CFA) [1] such that each pixel in the captured Bayer CFA image I^{Bayer} contains only one color component and I^{Bayer} consists of 25% red, 50% green, and 25% blue color values. One 8×8 Bayer CFA image example is shown in Fig. 1(a), and Fig. 1(b) depicts its CFA module [G, R, B, G]. The other three CFA modules are depicted in Figs. 1(c)-(e). In the color digital cameras market, the Bayer CFA pattern with module [G, R, B, G] has been used in the Agfa DC-504, Agfa Sensor530s, Nikon D200, etc. For simplicity, in this paper, we only consider the Bayer CFA module [G, R, B, G], although our discussion is also applicable to the other three CFA modules. In the past years, several compression methods have been developed for encoding Bayer CFA images, and these methods can be divided into two approaches: the structure conversion-first compression approach [5], [6], [8], [11], [14] and the demosaicking-first compression approach [3], [13]. Due to the quality and quality-bitrate tradeoff merits, the demosaicking-first compression approach for encoding I^{Bayer} has been the main trend instead of the structure conversion-first compression approach. Given I^{Bayer}, the flowchart of the overall pipeline to reconstruct the RGB full-color image and Bayer CFA image is depicted in Fig. 2. Prior to compression, we perform a demosaicking method [9] on I^{Bayer} to obtain an RGB full-color image I^{RGB}. Next, I^{RGB} is transformed to a YUV image I^{YUV} by

$$\begin{bmatrix} Y \\ U \\ V \end{bmatrix} = \begin{bmatrix} 0.257 & 0.504 & 0.098 \\ -0.148 & -0.291 & 0.439 \\ 0.439 & -0.368 & -0.071 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix} + \begin{bmatrix} 16 \\ 128 \\ 128 \end{bmatrix} \quad (1)$$

where Y denotes the luma image and UV denotes the chroma image. The commonly used chroma subsampling formats are 4:4:4, 4:2:2, and 4:2:0. For each 2×2 UV block, the 4:4:4 scheme has no compression and maintains both luma and chroma data completely. To achieve the compression effect, the 4:2:2 scheme determines one subsampled (U, V)-pair for each row of the 2×2 UV block. The 4:2:2 format has been used in high-end digital videos and interfaces,

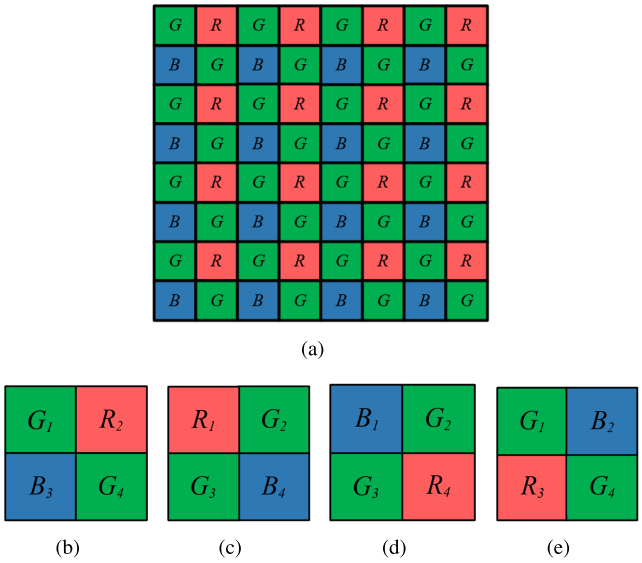


Fig. 1. One 8×8 Bayer CFA image example and the four possible CFA modules. (a) I^{Bayer} with size 8×8. (b) [G, R, B, G]. (c) [R, G, G, B]. (d) [B, G, G, R]. (e) [G, B, R, G].

Manuscript received June 14, 2018; revised September 18, 2018 and October 14, 2018; accepted October 30, 2018. Date of publication November 1, 2018; date of current version October 29, 2019. This work was supported by the contracts under Grant MOST-104-2221-E-011-004-MY3 and Grant MOST-107-2221-E-011-108-MY3. This paper was recommended by Associate Editor J. Xu. *(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, Taiwan (e-mail: klchung01@gmail.com).

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

Digital Object Identifier 10.1109/TCSVT.2018.2879095

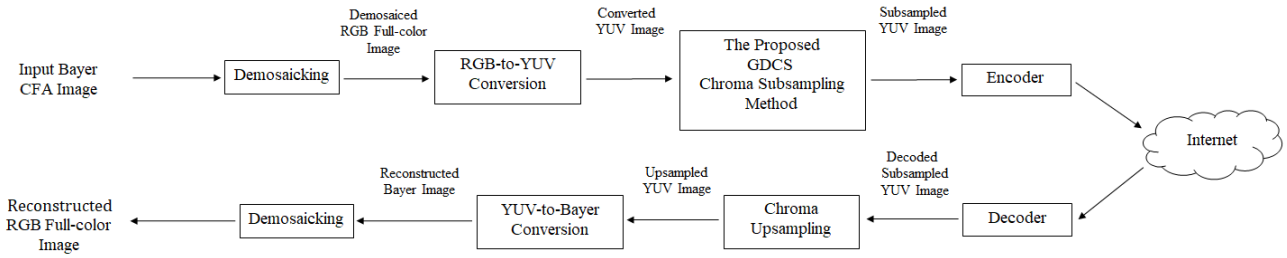


Fig. 2. Given an input Bayer CFA image, the flowchart of the overall pipeline to reconstruct RGB full-color images and Bayer CFA images.

such as AVC-Intra 100, Digital Betacam, Digital-S, Canon MXF HD422, and XDCAM HD422. To achieve a better compression effect, the 4:2:0 scheme determines one subsampled (U, V)-pair for each 2×2 UV block. The 4:2:0 format has been used in Blu-ray discs (BDs), digital versatile discs (DVDs), movies, sports, and TV shows.

After performing a chroma 4:2:2 or 4:2:0 subsampling on the UV chroma image, the subsampled YUV image is fed into the encoder for compression. At the client side, the decoded subsampled UV image is first upsampled, and based on the Bayer CFA module of I^{Bayer}, the YUV-to-Bayer conversion is performed by

$$\begin{bmatrix} R \\ G \\ B \end{bmatrix} = \begin{bmatrix} 1.164 & 0 & 1.596 \\ 1.164 & -0.391 & -0.813 \\ 1.164 & 2.018 & 0 \end{bmatrix} \begin{bmatrix} Y - 16 \\ U - 128 \\ V - 128 \end{bmatrix}, \quad (2)$$

Then, the RGB full-color image is reconstructed by demosaicking the reconstructed, i.e. converted, Bayer CFA image. In this paper, we focus on the design of more effective chroma 4:2:2 and 4:2:0 subsampling methods for Bayer images, making a contribution to this practically challenging research area.

A. Related Works

We first introduce seven existing chroma 4:2:0 subsampling methods, namely, 4:2:0(A), 4:2:0(L), 4:2:0(R), 4:2:0(DIRECT), 4:2:0(MPEG-B), Chen *et al.*'s [3] method, and Lin *et al.*'s [13] method, and point out their weaknesses. Second, four existing chroma 4:2:2 subsampling methods, namely, 4:2:2(A), 4:2:2(L), 4:2:2(R), and Chung *et al.*'s [4] method, are introduced and their weaknesses are highlighted. These weaknesses motivated us to develop more effective chroma 4:2:0 and 4:2:2 subsampling methods for Bayer images in order to achieve better quality improvement of the reconstructed images.

1) Existing Chroma 4:2:0 Subsampling Methods and Their Weaknesses: 4:2:0(A), 4:2:0(L), 4:2:0(R), 4:2:0(DIRECT), and 4:2:0(MPEG-B) are five traditionally used chroma 4:2:0 subsampling methods. For each 2×2 UV block in Fig. 3(a), 4:2:0(A) subsamples the (U, V)-pair by averaging the U and V components of the 2×2 chroma block. 4:2:0(L) and 4:2:0(R) subsample the (U, V)-pairs by averaging the chroma components in the left column and the right column of the 2×2 block, respectively. 4:2:0(DIRECT) subsamples the (U, V)-pair by selecting the top-left (U, V)-pair of the 2×2 block. 4:2:0(MPEG-B) [15] decides the subsampled (U, V)-pair by performing the 13-tap filter with the mask [2,0,-4,-3,5,19,26,19,5,-3,-4,0,2]/64 on the top-left location of

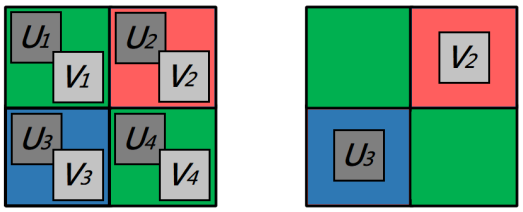


Fig. 3. The compression effect *et al.*'s method. (a) The 2×2 UV block. (b) The subsampled (U, V)-pair by Chen *et al.*'s method.

the 2×2 block. The main weakness of the above five traditional chroma subsampling methods is the lack of taking the Bayer CFA module information into account, leading to the quality degradation of the reconstructed images.

To overcome the weakness in the above mentioned chroma 4:2:0 subsampling methods for Bayer images, Chen *et al.* [3] observed that the R value in Eq. (2) is dominated by the Y and V values and the B value is dominated by the Y and U values. In addition to the above observation from Eq. (2), their chroma subsampling method also considered the Bayer CFA module information in connection with the 2×2 UV block. For example, suppose the Bayer CFA module is [G, R, B, G], as shown in Fig. 1(b). For the 2×2 UV block in Fig. 3(a), according to Chen *et al.*'s method, the subsampled (U, V)-pair is equal to (U_3, V_2), as shown in Fig. 3(b). Although Chen *et al.*'s method benefits the quality of the reconstructed R and B components, it does not benefit the quality of the reconstructed G components, leading to limited quality improvement because I^{Bayer} consists of 50% G values.

In order to evenly benefit the reconstruction of the R, G, and B components, Lin *et al.* [13] proposed an improved method. Let (U_s, V_s) be the subsampled (U, V)-pair of the 2×2 UV block. The Bayer CFA block-distortion used in [13] is defined by

$$\begin{aligned} &D^{Bayer}(U_s, V_s) \\ &= (G_1 - G_1')^2 + (R_2 - R_2')^2 + (B_3 - B_3')^2 + (G_4 - G_4')^2 \\ &= [(-0.391 \times U_1 - 0.813 \times V_1) \\ &\qquad - (-0.391 \times U_s - 0.813 \times V_s)]^2 \\ &\quad + [(1.596 \times V_2) - (1.596 \times V_s)]^2 \\ &\quad + [(2.018 \times U_3) - (2.018 \times U_s)]^2 \\ &\quad + [(-0.391 \times U_4 - 0.813 \times V_4) \\ &\qquad - (-0.391 \times U_s - 0.813 \times V_s)]^2 \end{aligned} \quad (3)$$

where G'_1, R'_2, B'_3, and G'_4 denote the reconstructed non-negative values of G_1, R_2, B_3, and G_4, respectively, which can be obtained via replacing U and V in Eq. (2) by U_s and V_s. Further, by performing the first derivative on Eq. (3) with respect to U_s and V_s, and then setting the two derivatives to zero, it yields

$$\frac{\partial D^{Bayer}(U_s, V_s)}{\partial V_s} = 0$$
$$\frac{\partial D^{Bayer}(U_s, V_s)}{\partial U_s} = 0. \tag{4}$$

The solution of Eq. (4), (U_{Lin}, V_{Lin}), is equal to Eq. (5), shown at the bottom of this page, where $a_1 = -0.391$, $a_2 = 0$, $a_3 = 2.018$, $a_4 = -0.391$, $b_1 = -0.813$, $b_2 = 1.596$, $b_3 = 0$, and $b_4 = -0.813$. The solution (U_{Lin}, V_{Lin}) is only a local optimal solution to minimize Eq. (3) because it may violate the constraint: the reconstructed G_1, R_2, B_3, and G_4 values in integer should be in the range [0, 255]. When this constraint is violated, they enforced the reconstructed color value(s) to 0 or 255, limiting the quality improvement. It is noticeable that the chroma upsampling process in Lin *et al.*'s method is the COPY reconstruction process, called COPY for short, in which the four reconstructed (U, V)-pairs of the current 2×2 UV block just copy the subsampled (U, V)-pair obtained by the concerned method.

2) Existing Chroma 4:2:2 Subsampling Methods and Their Weaknesses: 4:2:2(A), 4:2:2(L), and 4:2:2(R) are the three traditionally used chroma 4:2:2 subsampling methods. For each 2×2 UV block, 4:2:2(A) subsamples the two (U, V)-pairs, (U_{upp}, V_{upp})-pair and (U_{low}, V_{low})-pair, by averaging the U and V components of the upper 1×2 chroma sub-block and the lower 1×2 chroma sub-block, respectively, of the 2×2 UV block. The common weakness of the above three chroma 4:2:2 subsampling methods is to leave the Bayer CFA module information out of consideration, degrading the quality of the reconstructed images.

Chung *et al*. [4] observed that the coefficient of the V component in the first equation of Eq. (2) has a greater impact on reconstructing the R component than the G component; similarly, the coefficient of the U component in the third equation of Eq. (2) has a greater impact on reconstructing the B component than the G component. To determine better (U_{upp}, V_{upp})-pair and (U_{low}, V_{low})-pair, they preferentially considered the ordered significance of the U component for reconstructing the B and G pixels and the V component for reconstructing the R and G pixels. Experimental data demonstrated that Chung *et al.*'s method has better quality performance of the reconstructed images when compared with 4:2:2(A), 4:2:2(L), and 4:2:2(R). Chung *et al.*'s method is rather heuristic and lacks for mathematical analysis on minimizing the Bayer CFA 1×2 sub-block-distortion, limiting the quality improvement of the reconstructed images.

The above mentioned weaknesses in the previous works on 4:2:0 and 4:2:2 motivated us to develop more effective chroma 4:2:0 and 4:2:2 subsampling methods for Bayer images to improve the quality of the reconstructed images. In addition, besides the above COPY reconstruction process, we also consider the bilinear interpolation reconstruction process, called BILINEAR for short, which will be defined in the next subsection.

B. Contributions

In this paper, for chroma 4:2:0 subsampling of one 2×2 UV chroma block, i.e. for 4:2:0, we first analyze the convex property of the surface generated from the Bayer block-distortion function in Eq. (3) under the assumption: U and V are real and in the interval [0, 255]; the reconstructed R, G, and B color values are in the real domain. For convenience, this assumption is called $assumption_1$. In addition, we provide a convex property analysis of the Bayer 1×2 sub-block-distortion function for chroma 4:2:2 subsampling under $assumption_1$ for COPY. Further, we exploit the room to improve Lin *et al.*'s method [13] for 4:2:0 and Chung *et al.*'s [4] method for 4:2:2. Finally, an effective gradient descent-based chroma subsampling (GDCS) method is proposed to tackle the two kinds of chroma subsampling, 4:2:0 and 4:2:2, under COPY and BILINEAR. In BILINEAR, the left-top chroma pixel of the current 2×2 chroma block B_c is interpolated by performing the bilinear interpolation on the subsampled chroma (U, V)-pairs of the western, north-western, northern neighboring 2×2 chroma blocks of B_c, which have been determined by GDCS, and the subsampled (U, V)-pair located at the center of B_c. As to the left-bottom chroma pixel of B_c, it can be interpolated by referring to the subsampled chroma (U, V)-pair of the western neighboring 2×2 chroma block of B_c, which has been determined by GDCS, the subsampled chroma (U, V)-pairs of the south-western and southern neighboring 2×2 chroma blocks of B_c, which are determined by Lin *et al.*'s method, and the subsampled (U, V)-pair located at the center of B_c. Following the similar way, the right-top and right-bottom chroma pixels of B_c can be reconstructed.

Based on the test Bayer CFA images generated from the Kodak [10] and IMAX datasets [18], experimental results

$$U_{Lin} = \frac{(\sum_{k=1}^{4} b_k^2)\cdot(\sum_{k=1}^{4} a_k^2 U_k + a_k b_k V_k) - (\sum_{k=1}^{4} a_k b_k)\cdot(\sum_{k=1}^{4} a_k b_k U_k + b_k^2 V_k)}{(\sum_{k=1}^{4} a_k^2)\cdot(\sum_{k=1}^{4} b_k^2) - (\sum_{k=1}^{4} a_k b_k)^2}$$
$$V_{Lin} = \frac{(\sum_{k=1}^{4} a_k b_k)\cdot(\sum_{k=1}^{4} a_k^2 U_k + a_k b_k V_k) - (\sum_{k=1}^{4} a_k^2)\cdot(\sum_{k=1}^{4} a_k b_k U_k + b_k^2 V_k)}{(\sum_{k=1}^{4} a_k b_k)^2 - (\sum_{k=1}^{4} a_k^2)\cdot(\sum_{k=1}^{4} b_k^2)} \tag{5}$$

demonstrated that in the High Efficiency Video Coding (HEVC) standard, our GDCS methods for 4:2:0 and 4:2:2 clearly outperform the seven existing 4:2:0 methods and four existing 4:2:2 methods, respectively. Here, the evaluation items used in the comparison include the peaksignal-to-noise-ratio (PSNR), color PSNR (CPSNR), structure similarity index (SSIM) [17], quality-bitrate tradeoff, and visual effect.

In addition, we also provide the execution time comparison among the concerned methods and our GDCS method. For completeness, we also compare our GDCS method with the interpolation dependent image downsampling (IDID) method [19] and the $CS_{BILINEAR}$ method [16] under the new edge-directed interpolation (NEDI) [12] and BILINEAR reconstruction processes, respectively. The experimental results demonstrated that the combination GDCS-BILINEAR has better PSNR, CPSNR, and SSIM quality performance relative to the two combinations, IDID-NEDI and $CS_{BILINEAR}$-BILINEAR, indicating the quality superiority of GDCS over IDID and $CS_{BILINEAR}$.

The rest of this paper is organized as follows. In Section II, the convex property of the Bayer block-distortion functions and our GDCS method for 4:2:0 and 4:2:2 on Bayer CFA images are presented. In Section III, the experimental results are demonstrated to show the quality and quality-bitrate tradeoff merits of our GDCS method. In Section IV, some concluding remarks are addressed.

II. The Proposed Chroma Subsampling Method for Bayer CFA Images: GDCS

In the first subsection, we prove that the Bayer 2×2 block-distortion function in Eq. (3) for 4:2:0 and the corresponding 1×2 sub-block-distortion function for 4:2:2 under $assumption_1$ are convex. Then, we exploit the room to improve Lin *et al.*'s method for 4:2:0 and Chung *et al.*'s method for 4:2:2. In the second subsection, we propose a GDCS method to improve the quality of the reconstructed images. In the third subsection, taking the exhaustive search (ES) method as the comparison base, the accuracy merit of our GDCS method over the concerned methods is provided.

A. The Convex Property of the Bayer Block-Distortion Functions and the Room to Improve Previous Methods for 4:2:0 and 4:2:2

For 4:2:0, the block-distortion function in Eq. (3) is a quadratic function in terms of the parameters U_s and V_s. In Appendix, we have proved that under $assumption_1$, the Bayer 2×2 block-distortion function is convex. From Eq. (11) in Appendix, we know that the block-distortion function for [G, R, B, G] is the same for the other three Bayer CFA modules [R, G, G, B], [B, G, G, R], and [G, B, R, G]. Therefore, under $assumption_1$, the block-distortion function is always convex, no matter what the Bayer CFA module is. We thus have the following proposition.

Proposition 1: The Bayer 2×2 block-distortion function in Eq. (3) for any one of the four CFA modules in Figs. 1(b)-(e) is convex under $assumption_1$.

For 4:2:2 under $assumption_1$, let (U'_s, V'_s) denote the subsampled (U, V)-pair of either the upper 1×2 UV sub-block or the lower 1×2 UV sub-block. Considering the 1×2 Bayer CFA sub-module [G, R] (see the upper and lower 1×2 Bayer CFA sub-modules of Fig. 1(b) and Fig. 1(d), respectively), the 1×2 sub-block-distortion can be expressed as

$$\begin{aligned} &D^{Bayer}(U'_s, V'_s) \\ &= (G_1 - G'_1)^2 + (R_2 - R'_2)^2 \\ &= [(-0.391 \times U_1 - 0.813 \times V_1) \\ &\qquad - (-0.391 \times U'_s - 0.813 \times V'_s)]^2 \\ &\quad + [(1.596 \times V_2) - (1.596 \times V'_s)]^2 \end{aligned} \tag{6}$$

Following the similar proving technique in Appendix, we can prove that the corresponding determinant of $H(D^{Bayer}(U'_s, V'_s))$ in Eq. (6) is expressed as

$$\begin{aligned} \det H(D^{Bayer}) &= 2(\sum_{k=1}^{2} a_k^2 \sum_{k=1}^{2} b_k^2 - \sum_{k=1}^{2} a_k b_k \sum_{k=1}^{2} a_k b_k) \\ &= 10.7667 > 0 \end{aligned} \tag{7}$$

where a_i and b_i, $1 \le i \le 4$, have been defined before, implying the Bayer CFA 1×2 sub-block-distortion function in Eq. (6) is convex. Considering the 1×2 CFA sub-module [R, G] (see the upper and lower 1×2 Bayer CFA sub-modules of Fig. 1(c) and Fig. 1(e), respectively), we can also prove that the determinant of $H(D^{Bayer}(U'_s, V'_s))$ is equal to 1.5577, indicating the convex property of the Bayer CFA 1×2 sub-block-distortion function. By the same arguments, considering the 1×2 CFA sub-modules [B, G] and [G, B], we can prove that the determinants of $H(D^{Bayer}(U'_s, V'_s))$ are equal to 1.5577, indicating the convex property of their corresponding Bayer CFA sub-block-distortion functions. Further, we have the following proposition.

Proposition 2: The Bayer 1×2 block-distortion function in Eq. (6) for any one of the four CFA modules, [G, R], [R, G], [B, G], and [G, B], is convex under $assumption_1$.

One 2×2 Bayer CFA block example, its corresponding demosaicked RGB full-color block, and the transformed 2×2 YUV block are shown in Figs. 4(a)-(c), respectively. Under $assumption_1$, the Bayer CFA block-distortion convex function of Fig. 4(a) is plotted in Fig. 4(d). We now put the constraint "U_s, V_s, and the reconstructed color values should be integer and must be within the interval [0, 255]" back into the Bayer CFA block-distortion in Eq. (3) for 4:2:0. Accordingly, the resultant grid plot of Fig. 4(d) is depicted in Fig. 4(e) where the figure shape is discretely convex-like, although it is not strictly convex, leading to the room to improve Lin *et al.*'s method for 4:2:0. Because it has been verified that the Bayer CFA 1×2 sub-block-distortion function for 4:2:2 is convex, its corresponding grid plot and the figure shape are similar to that for 4:2:0.

For Fig. 4(a), the subsampled (U, V)-pair by Lin *et al.*'s method for 4:2:0, (U_{Lin}, V_{Lin}), is denoted by the red-marked point in Fig. 5(a), while a better subsampled (U, V)-pair with much less Bayer CFA block distortion is denoted by the yellow-marked point in Fig. 5(a), clearly indicating the room to improve Lin *et al.*'s method. How to design an effective chroma 4:2:0 subsampling method to find the solution path

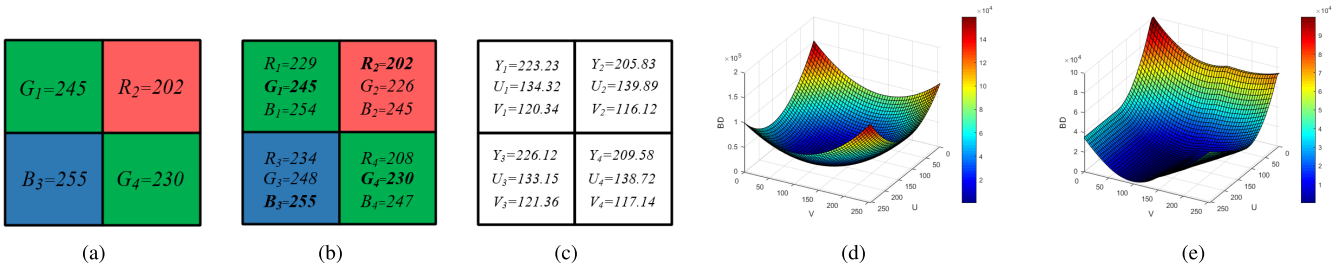


Fig. 4. One 2×2 block example, the convex function of the Bayer CFA 2×2 block-distortion in Eq. (3) under $assumption_1$, and the corresponding grid plot without $assumption_1$. (a) The 2×2 Bayer CFA block. (b) The 2×2 demosaicked RGB full-color block of Fig. 3(a). (c) The transformed 2×2 YUV block. (d) The convex function of the 2×2 CFA block-distortion in Eq. (3) under $assumption_1$. (e) The grid plot of the 2×2 Bayer CFA block-distortion in Eq. (3) without $assumption_1$.

Procedure: GDCS

Input: 2×2 Bayer CFA block and the transformed 2×2 YUV block.

Output: Subsampled (U, V)-pair, (U_{GDCS}, V_{GDCS}).

Step 1: $k = 0$.

Step 2: Select (U_{Lin}, V_{Lin}) as the initial solution $(U_s^{(k)}, V_s^{(k)})$ and calculate the Bayer CFA block-distortion, $D^{Bayer}(U_s^{(k)}, V_s^{(k)})$.

Step 3: Under the constraint: $0 \le G_1', R_2', B_3'$, and $G_4' \le 255$, we calculate all the eight neighboring Bayer CFA block-distortions, $D^{Bayer}(U_s^{(k)} + i, V_s^{(k)} + j)s$ for (i, j) $\in \{(0, 1), (0, -1), (1, 0), (-1, 0), (1, 1), (1, -1), (-1, 1), (-1, -1)\}$; we select the minimal one among the eight distortions and set the corresponding (U, V)-pair to be the temporary subsampled (U, V)-pair.

Step 4: If $D^{Bayer}(U_s^{(k+1)}, V_s^{(k+1)})$ is greater than or equal to $D^{Bayer}(U_s^{(k)}, V_s^{(k)})$, we stop GDCS and output $(U_s^{(k)}, V_s^{(k)})$ as the final subsampled (U, V)-pair (U_{GDCS}, V_{GDCS}); otherwise, we perform the assignment operation $k := k + 1$ and go to Step 3.

from the red-marked circle to the yellow-marked circle is challenging and important.

B. The Proposed Gradient Descent-Based Chroma Subsampling Method: GDCS

According to the observation on the path from the red-marked point (U_{Lin}, V_{Lin}) to the yellow-marked point, by selecting the point (U_{Lin}, V_{Lin}) as the initial solution point, we propose an effective gradient descent-based chroma subsampling (GDCS) method to improve the quality of the reconstructed images by Lin *et al.*'s method for 4:2:0. The whole GDCS procedure, in which 'k = 0' denotes the initial iteration, is listed below.

For the example in Fig. 4(a), using our GDCS procedure, the solution path from the red circle located at (U_{Lin}, V_{Lin}) to the yellow circle located at (U_{GDCS}, V_{GDCS}) is depicted in Fig. 5(b), leading to the reduction of the Bayer CFA block-distortion of (U_{Lin}, V_{Lin}), i.e. improving the quality of the reconstructed images. Our GDCS procedure mentioned above can be slightly modified to tackle the 4:2:2 case. To save space, we omit the detail. The available execution codes of our GDCS procedures for 4:2:0 and 4:2:2 can be accessed from the website in [7].

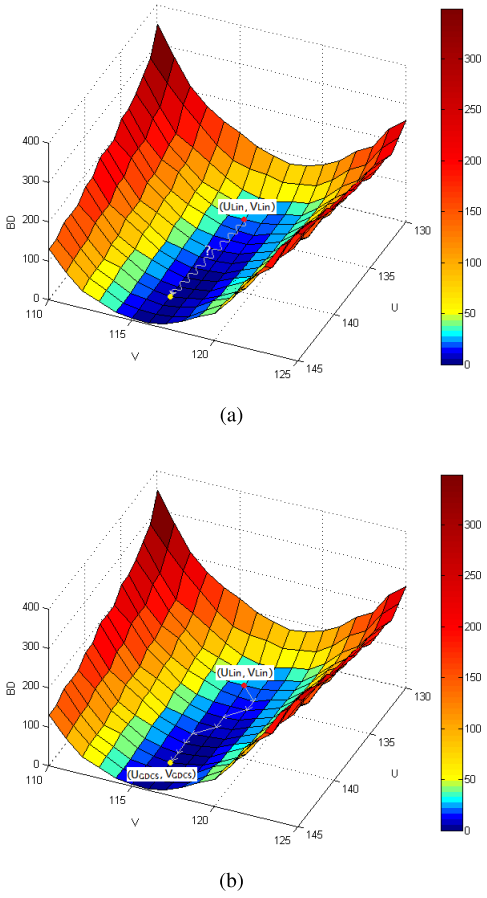


Fig. 5. For Fig. 4(a), the depiction to show the room to improve Lin *et al.*'s method for 4:2:0. (a) The solution path to be determined for improving Lin *et al.*'s method. (b) The solution path determined by our GDCS method for Fig. 5(a).

C. Accuracy Analysis

Taking the subsampled (U, V)-pairs determined by the exhaustive search (ES) method as the base, which will be described in the next paragraph, we take the Kodak and IMAX test datasets to analyze the accuracy of the subsampled (U, V)-pairs determined by Lin *et al.*'s method and our GDCS method for 4:2:0. In addition, the accuracy analysis of the subsampled (U, V)-pairs determined by the eight concerned methods for 4:2:0 and the five concerned methods for 4:2:2.

Before estimating the accuracy, we first determine the global optimal solution by the ES method over the

range [0, 255] × [0, 255] as the comparison base. In ES, as the first possible solution, we try $(U, V) = (0, 0)$ and calculate its Bayer block-distortion value. If the reconstructed color value is greater than 255 (or less than 0), it is forced to 255 (or 0). Next, we try $(U, V) = (1, 0)$ and repeat the above process. After all the possible (U, V)-pairs in the whole search area [0, 255] × [0, 255] have been examined, among the values of $D^{Bayer}(0, 0)$, $D^{Bayer}(1, 0)$, $D^{Bayer}(2, 0)$, ... , and $D^{Bayer}(255, 255)$, we select the minimal one and its corresponding subsampled (U, V)-pair is taken as the global optimal solution. According to the above description of ES, the computational complexity of ES is dependent on the search area with size 256^2, and it is terribly time-consuming.

Let (U_{CS}, V_{CS}) denote the subsampled (U, V)-pair by the concerned chroma subsampling method CS. For example, when $CS = ES$, it means that the chroma subsampling method is the exhaustive search method; when $CS = Lin$, it means that the chroma subsampling method is Lin *et al.*'s method. To measure the accuracy of one concerned chroma subsampling method, for any 2×2 UV block, we perform the following assignment:

$$f(CS) := f(CS) + 1, \quad \text{if } (U_{CS}, V_{CS}) = (U_{ES}, V_{ES}),$$
$$f(CS) := f(CS); \quad \text{otherwise.} \tag{8}$$

where the frequency function $f(CS)$ is added by one when the two subsampled (U, V)-pairs, (U_{CS}, V_{CS}) and (U_{ES}, V_{ES}), are equivalent; otherwise, we do nothing. After processing all the 2×2 UV blocks by Eq. (8), the accuracy of the chroma subsampling method CS can be estimated by the resultant frequency function $f(CS)$.

Based on the Kodak and IMAX datasets, for 4:2:0 under COPY, it yields to f(4:2:0(A)) = 42.28%, f(4:2:0(L)) = 39.02%, f(4:2:0(R)) = 36.34%, f(DIRECT) = 30.85%, f(MPEG-B) = 25.02%, f(Chen) = 60.86%, f(Lin) = 73.31%, and f(GDCS) = 99.64%; under BILINEAR, it yields to f(4:2:0(A)) = 30.54%, f(4:2:0(L)) = 28.54%, f(4:2:0(R)) = 27.82%, f(DIRECT) = 25.46%, f(MPEG-B) = 25.28%, f(Chen) = 30.19%, f(Lin) = 31.96%, and f(GDCS) = 76.31%. Clearly, for 4:2:0, our method has the highest accuracy among the eight concerned methods. Under COPY, it indicates that 99.64% of the subsampled (U, V)-pairs determined by GDCS are global optimal. On the contrary, only 0.36% of the subsampled (U, V)-pairs by GDCS are not equal to those by ES, and the experimental results showed that most of the corresponding 2×2 blocks in this case appear on textural parts of the image.

For 4:2:2 under COPY, it yields to f(4:2:2(A)) = 45.14%, f(4:2:2(L)) = 42.83%, f(4:2:0(R)) = 40.94%, f(Chung) = 47.12%, and f(GDCS) = 99.72%; under BILINEAR, it yields to f(4:2:2(A)) = 42.33%, f(4:2:2(L)) = 38.82%, f(4:2:0(R)) = 37.62%, f(Chung) = 40.56%, and f(GDCS) = 88.97%, indicating the accuracy superiority of our GDCS method relative to the other four concerned methods.

III. Experimental Results

Based on the test Bayer CFA images collected from the Kodak dataset with 24 RGB full-color images [10] and the IMAX dataset with 18 RGB full-color images [18], the comprehensive experiments were carried out using the HEVC reference software platform HM-16.18 to compare the quality and quality-bitrate tradeoff performance among the concerned methods. The execution time comparison of the concerned eight methods are also investigated. All the concerned experiments are implemented on a computer with an Intel Core i7-3770 CPU 3.4 GHz and 7.68 GB RAM. The operating system is the Microsoft Windows 10 64-bit operating system. The program development environment is Visual C++ 2017.

A. Quality Merit

PSNR, CPSNR, and SSIM are used to justify the quantitative quality merit of our GDCS method among the concerned methods. In the first set of experiments, the quantization parameter (QP) in the encoder is set to zero. Therefore, the encoder passes each concerned method. In Subsection III.B, the second set of experiments is carried out under different QP values to compare the quality-bitrate tradeoff performance among the concerned methods. Let the reconstructed Bayer CFA image be denoted by I'^{Bayer}.

The PSNR value of the reconstructed Bayer CFA image is calculated by

$$\text{PSNR} = \frac{1}{N}\sum_{n=1}^{N} 10\log_{10}\frac{255^2}{MSE} \tag{9}$$

with $MSE = \frac{1}{WH}\sum_{p\in P}[I_n^{Bayer}(p) - I_n^{'Bayer}(p)]^2$ in which *'p'* denotes the position of the Bayer CFA image pixel in one W×H image. *'N'* denotes the number of test images. To measure the quality of the reconstructed RGB full-color images I'^{RGB}, the used CPSNR metric is defined by

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

with $CMSE = \frac{1}{3WH}\sum_{p\in P}[I_n^{RGB}(p) - I_n^{'RGB}(p)]^2$. The SSIM metric is expressed as the product of the luminance mean similarity, the contrast similarity in terms of variance, and the structure similarity in terms of co-variance between I^{RGB} and I'^{RGB}. We omit the detailed definition of SSIM; the reader can refer to the paper by Wang *et al.* [17].

The average PSNR, CPSNR, and SSIM values of the concerned 4:2:0 and 4:2:2 methods are tabulated in Table I in which the three average values under BILINEAR are listed in the parentheses. Table I indicates the quality improvement of our GDCS method over seven and four concerned methods for 4:2:0 and 4:2:2, respectively.

Based on the two test datasets, the mean square errors (MSEs) of GDCS-COPY and GDCS-BILINEAR for 4:2:0 are 8.1223 and 6.0497, respectively; the MSEs of GDCS-COPY and GDCS-BILINEAR for 4:2:2 are 0.4745 and 1.2064, respectively. This can explain why in Table I, for 4:2:0, the CPSNR performance of GDCS-BILINEAR is better than GDCS-COPY and for 4:2:2, the CPSNR performance of GDCS-BILINEAR is worse than GDCS-COPY. Furthermore, based on the same datasets, the MSEs of ES-COPY and ES-BILINEAR for 4:2:0 are 8.0376 and 4.9017, respectively;

TABLE I

QUALITY COMPARISON (QP = 0) AMONG OUR GDCS METHOD AND THE CONCERNED METHODS FOR KODAK AND IMAX

	4:2:0(A)	4:2:0(L)	4:2:0(R)	4:2:0(DIRECT)	4:2:0(MPEG-B)	Chen *et al.* [3]	Lin *et al.* [13]	GDCS
PSNR (dB)	40.2984 [40.9342]	39.4902 [41.4668]	39.7161 [41.0732]	38.7300 [40.3216]	38.9551 [40.1901]	44.3682 [41.7213]	45.2406 [42.3105]	**45.6626** **[46.6642]**
CPSNR (dB)	40.1903 [41.5291]	39.5803 [42.0338]	39.7654 [40.2882]	39.0456 [40.8551]	39.0923 [40.6492]	43.4585 [41.7409]	44.1822 [42.4149]	**44.5032** **[45.5077]**
SSIM	0.9577 [0.9605]	0.9860 [0.9883]	0.9810 [0.9869]	0.9798 [0.9855]	0.9809 [0.9849]	0.9912 [0.9870]	0.9935 [0.9895]	**0.9938** **[0.9946]**
PSNR Gain (dB)	5.3642 [5.7300]	6.1724 [5.1973]	5.9465 [5.5910]	6.9327 [6.3426]	6.7076 [6.4740]	1.2944 [4.9429]	0.4220 [4.3537]	
CPSNR Gain (dB)	4.3128 [3.9786]	4.9229 [3.4739]	4.7377 [5.2195]	5.4576 [4.6527]	5.4109 [4.8585]	1.0446 [3.7668]	0.3210 [3.0928]	
SSIM Gain	0.0361 [0.0341]	0.0079 [0.0063]	0.0128 [0.0077]	0.0140 [0.0092]	0.0130 [0.0098]	0.0027 [0.0076]	0.0004 [0.0051]	

				4:2:2(A)	4:2:2(L)	4:2:2(R)	Chung *et al.* [4]	GDCS
PSNR (dB)				43.9747 [44.7642]	42.9571 [43.1985]	41.5321 [42.3896]	45.7886 [44.3662]	**54.1003** **[50.7068]**
CPSNR (dB)				43.7107 [45.0104]	43.3227 [43.5725]	41.6067 [42.5876]	45.0457 [44.0609]	**52.2649** **[49.4076]**
SSIM				0.9911 [0.9931]	0.9918 [0.9914]	0.9854 [0.9891]	0.9943 [0.9923]	**0.9982** **[0.9969]**
PSNR Gain (dB)				10.1256 [5.9426]	11.1432 [7.5083]	12.5682 [8.3172]	8.3117 [6.3406]	
CPSNR Gain (dB)				8.5542 [4.3972]	8.9423 [5.8351]	10.6582 [6.8200]	7.2192 [5.3467]	
SSIM Gain				0.0071 [0.0038]	0.0064 [0.0055]	0.0129 [0.0078]	0.0039 [0.0046]	

TABLE II

QUALITY COMPARISON (QP = 0) AMONG IDID, $CS_{BILINEAR}$, AND OUR GDCS METHOD UNDER NEDI, BILINEAR, AND BILINEAR RECONSTRUCTION, RESPECTIVELY, FOR KODAK AND IMAX

	IDID [19]	$CS_{BILINEAR}$ [16]	GDCS
PSNR	44.4049	43.0246	**46.5938**
CPSNR	44.1269	42.7403	**45.5077**
SSIM	0.9916	0.9929	**0.9946**

the MSEs of ES-COPY and ES-BILINEAR for 4:2:2 are only 0.4437 and 0.5937, respectively. This can explain why in Table I, the CPSNR values via 4:2:2 are much higher than that via 4:2:0.

Previously, Zhang *et al.* [19] proposed an IDID method which was designed under the NEDI reconstruction process. Wang *et al.* [16] proposed a $CS_{BILINEAR}$ method which is designed for screen content images under the BILINEAR reconstruction process. We only consider three combinations, IDID-NEDI, $CS_{BILINEAR}$-BILINEAR, and GDCS-BILINEAR, which denote Zhang *et al.*'s method, Wang *et al.*'s method, and our method for 4:2:0 under the NEDI, BILINEAR, and BILINEAR reconstruction processes, respectively. Table II indicates that the combination GDCS-BILINEAR has the best PSNR, CPSNR, and SSIM quality performance in boldface for the reconstructed images, implying the quality superiority of GDCS over IDID and $CS_{BILINEAR}$.

B. Quality-Bitrate Tradeoff and Visual Effect Merits

1) Quality-Bitrate Tradeoff Merit: We depict the quality-bitrate tradeoff performance in a rate-distortion (RD) curve for each combination based on IMAX. Here, we consider these distinct QP values: 0, 4, 8, 12, 16, 20, ..., and 48. For 4:2:0, within the QP intervals, [0, 20] and [0, 24], under COPY and BILINEAR, respectively, as shown in Figs. 6(a)-(b) and Figs. 6(c)-(d), our GDCS method delivers the best quality-bitrate tradeoff performance among the concerned methods. However, within the QP intervals, [21, 48] and [25, 48], under COPY and BILINEAR, respectively, GDCS delivers similar quality-bitrate tradeoff performance relative to the other methods. For 4:2:2, within the QP intervals, [0, 16] and [0, 18], under COPY and BILINEAR, respectively, as shown in Figs. 6(e)-(f) and Figs. 6(g)-(h), our GDCS method delivers the best quality-bitrate tradeoff performance among the concerned methods.

From Figs. 6(a)-(b), for 4:2:0 under COPY, our GDCS method demonstrates its quality-bitrate tradeoff superiority mainly under very high bitrate setting, and GDCS is quite competitive with the other seven methods under the other bitrate setting. From Figs. 6(c)-(d), for 4:2:0 under BILINEAR, our GDCS method demonstrates its performance superiority mainly within the QP interval [0, 24]; under the other QP values, GDCS is quite competitive with the concerned comparative methods. Accordingly, our GDCS method is recommended to be used in the chroma 4:2:0 subsampling for the applications in BDs, DVDs, movies, sports, and TV shows, as mentioned in the third paragraph of Section I.

From Figs. 6(e)-(f), for 4:2:2 under COPY, our GDCS method demonstrates its performance superiority mainly within the QP interval [0, 16]; from Figs. 6(g)-(h), for 4:2:2 under BILINEAR, GDCS demonstrates its performance superiority mainly within the QP interval [0, 20]. However, GDCS has some performance loss cases under low bitrate setting relative to the four comparative methods. Therefore, under high bitrate setting, our GDCS method is recommended to be used in the chroma 4:2:2 subsampling for the applications in AVCIntra 100, Digital Betacam, Digital-S, Canon MXF HD422, and XDCAM HD422, as mentioned in the third paragraph of Section I. However, for 4:2:2 and low bitrate setting, Chung *et al.*'s [4] method and 4:2:2(A) are recommended under COPY and BILINEAR, respectively.

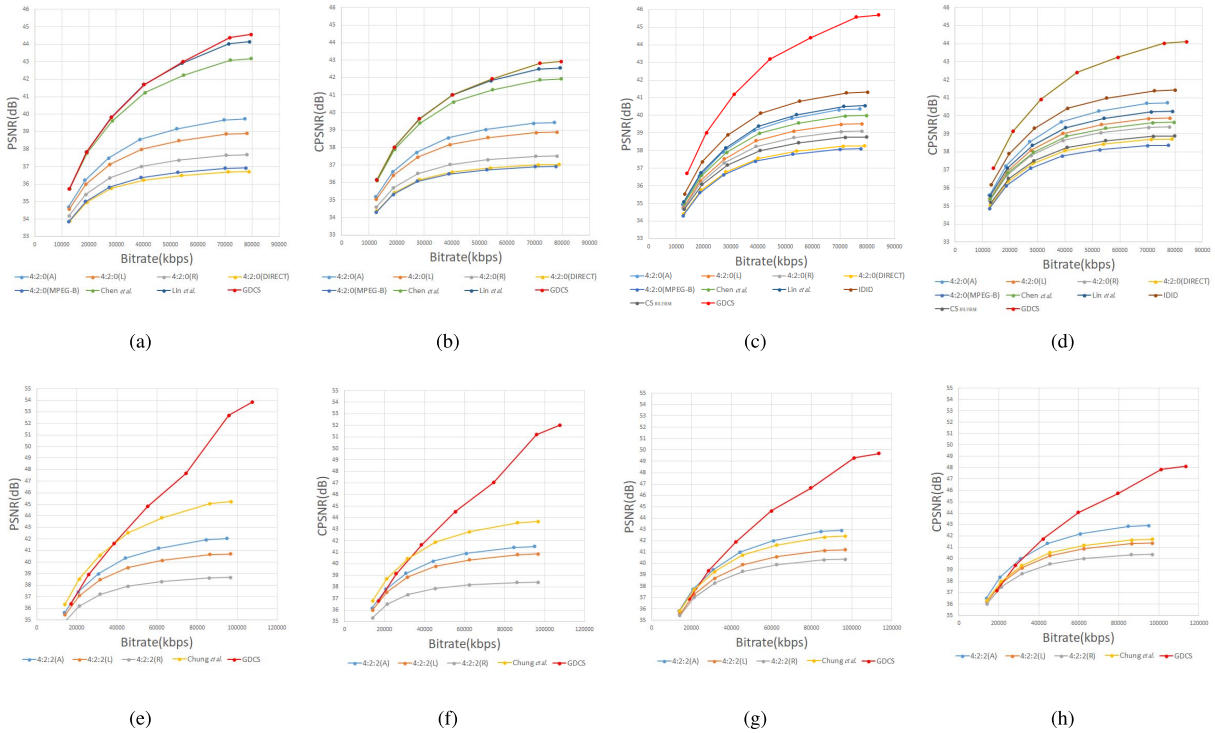


Fig. 6. Quality-bitrate performance comparison for 4:2:0 and 4:2:2. (a) For reconstructed Bayer CFA images under COPY for 4:2:0. (b) For reconstructed RGB full-color images under COPY for 4:2:0. (c) For reconstructed Bayer CFA images under BILINEAR for 4:2:0. (d) For reconstructed RGB full-color images under BILINEAR for 4:2:0. (e) For reconstructed Bayer CFA images under COPY for 4:2:2. (f) For reconstructed RGB full-color images under COPY for 4:2:2. (g) For reconstructed Bayer CFA images under BILINEAR for 4:2:2. (h) For reconstructed RGB full-color images under BILINEAR for 4:2:2.

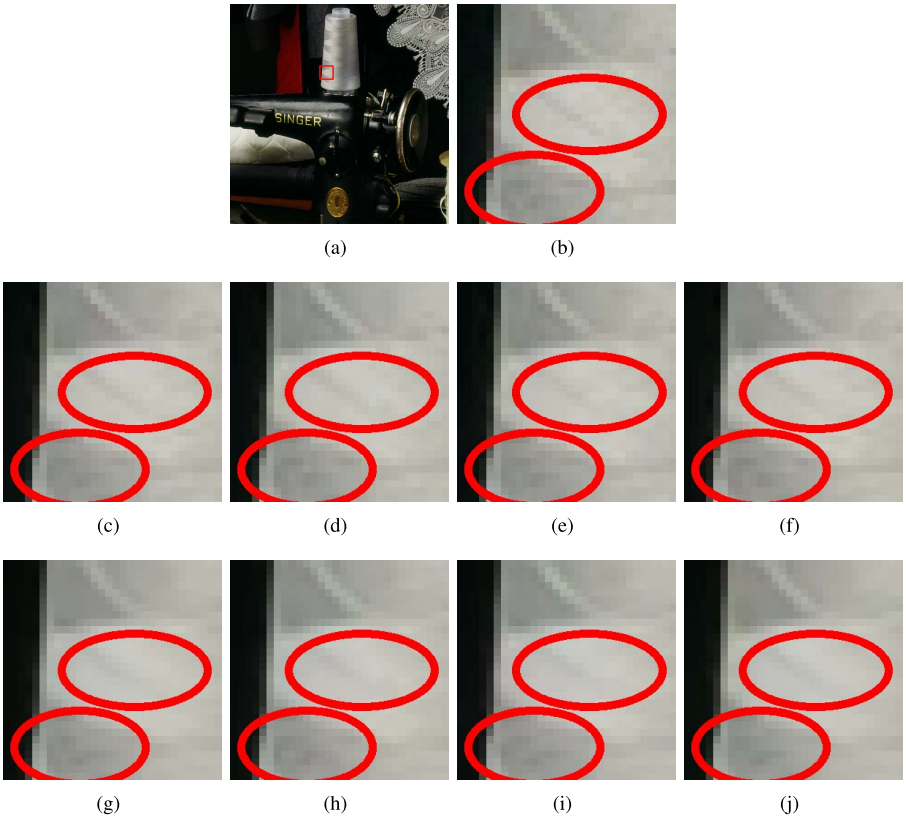


Fig. 7. Visual effect comparison. (a) The 8th image in IMAX. (b) The magnified subimage decoupled from (a). (c) 4:2:0(A)-COPY. (d) 4:2:0(DIRECT)-COPY. (e) Lin *et al.*-COPY for 4:2:0. (f) GDCS-COPY for 4:2:0. (g) 4:2:0(A)-BILINEAR. (h) 4:2:0(DIRECT)-BILINEAR. (i) Lin *et al.*-BILINEAR for 4:2:0. (j) GDCS-BILINEAR for 4:2:0.

2) Visual Effect Merit: As shown in Fig. 7(a), we take the 8th image from IMAX as the example to compare the visual effect. For visual comparison, as shown in Fig. 7(b), the magnified subimage is decoupled from the character-containing and color-containing regions in Fig. 7(a). Here, for 4:2:0, we consider the four methods: 4:2:0(A),

TABLE III
EXECUTION TIME COMPARISON (QP = 0) AMONG OUR GDCS METHOD AND THE CONCERNED METHODS FOR KODAK AND IMAX

	4:2:0(A)	4:2:0(L)	4:2:0(R)	4:2:0(DIRECT)	4:2:0(MPEG-B)	Chen *et al.* [3]	Lin *et al.* [13]	GDCS
Time	0.0031 [0.0031]	**0.0018** [0.0019]	**0.0018** [0.0019]	0.0019 [0.0019]	0.0035 [0.0034]	0.0019 [**0.0018**]	0.0036 [0.0037]	0.0258 [0.0619]
				4:2:2(A)	4:2:2(L)	4:2:2(R)	Chung *et al.* [4]	GDCS
Time				0.0026 [0.0025]	0.0024 [**0.0023**]	**0.0023** [**0.0023**]	0.0349 [0.0348]	0.0341 [0.0673]

4:2:0(DIRECT), Lin *et al.*'s method, and our GDCS method. Because the CPSNR values of the concerned methods for 4:2:2 are too high to compare the visual effect, we omit the visual effect comparison for 4:2:2.

For 4:2:0, we consider 4:2:0(A), 4:2:0(DIRECT), Lin *et al.*'s method, and our GDCS method for QP = 20 and QP = 24 under COPY and BILINEAR, respectively. After performing all the eight combinations, 4:2:0(A)-COPY, 4:2:0(DIRECT)-COPY, Lin *et al.*-COPY, and GDCS-COPY for QP = 20; 4:2:0(A)-BILINEAR, 4:2:0(DIRECT)-BILINEAR, Lin *et al.*-BILINEAR, and GDCS-BILINEAR for QP = 24, on Fig. 7(b), the eight reconstructed RGB full-color images are shown in Figs. 7(c)-(f) and Figs. 7(g)-(j), respectively. From Figs. 7(c)-(j), we observe that our GDCS method delivers the best visual effect among the concerned methods. In particular, 4:2:0(A), 4:2:0(DIRECT), and Lin *et al.*'s method have worse feature preservation effect.

C. Execution Time Comparison

Based on the same test datasets, for one test image, the average execution time requirements of the eight 4:2:0 methods of interest, namely, 4:2:0(A), 4:2:0(L), 4:2:0(R), 4:2:0(DIRECT), 4:2:0(MPEG-B), Chen *et al.*, Lin *et al.*, and our GDCS method, are listed in Table III in which the execution time is measured in terms of seconds. From these actual time costs, although GDCS takes more time than the other methods, as described in Subsection III.A, our GDCS method has better quality merit. For 4:2:2, we have the same conclusion.

IV. CONCLUSION

For 4:2:0 and 4:2:2, we have presented our GDCS method for Bayer CFA images. The accuracy analysis has indicated that the accuracy of our GDCS method is much higher than the other concerned methods. Under COPY and BILINEAR with QP = 0, for 4:2:0, our GDCS method achieves clear quality gains over the five commonly used methods, Chen *et al.*'s [3] method, Lin *et al.*'s [13] method, Zhang *et al.*'s [19] method, and Wang *et al.*'s [16] method. For 4:2:2 with QP = 0, our GDCS method also achieves clear quality gains over the four previous methods, 4:2:2(A), 4:2:2(L), 4:2:2(R), and Chung *et al.*'s [4] method. In general, our GDCS method is recommended to be used in 4:2:0 under COPY or BILINEAR for the applications in BDs, DVDs, movies, sports, and TV shows. Under high bitrate setting for 4:2:2, our GDCS method is recommended to be used to the applications in AVCIntra 100, Digital Betacam, Digital-S, Canon MXF HD422, and XDCAM HD422. However, under low bitrate setting for 4:2:2, Chung *et al.*'s method and 4:2:2(A) are recommended under COPY and BILINEAR, respectively.

APPENDIX

By Eq. (3), the Bayer CFA block-distortion function is rewritten as

$$\begin{aligned} D^{Bayer}(U_s, V_s) &= (G_1 - G_1')^2 + (R_2 - R_2')^2 + (B_3 - B_3')^2 + (G_4 - G_4')^2 \\ &= \sum_{k=1}^{4} [a_k(U_s - U_k) + b_k(V_s - V_k)]^2 \end{aligned} \tag{11}$$

where a_k and b_k, $1 \le k \le 4$, have been defined below Eq. (4). Under $assumption_1$, the Hessian matrix of $D^{Bayer}(U_s, V_s)$ is expressed as

$$H(D^{Bayer}) = \begin{bmatrix} \dfrac{\partial^2 D^{Bayer}}{\partial U_s^2} & \dfrac{\partial^2 D^{Bayer}}{\partial U_s \partial V_s} \\ \dfrac{\partial^2 D^{Bayer}}{\partial V_s \partial U_s} & \dfrac{\partial^2 D^{Bayer}}{\partial V_s^2} \end{bmatrix} \tag{12}$$

with

$$\begin{aligned} \frac{\partial^2 D^{Bayer}}{\partial U_s^2} &= \sum_{k=1}^{4} 2a_k^2, \\ \frac{\partial^2 D^{Bayer}}{\partial V_s^2} &= \sum_{k=1}^{4} 2b_k^2, \\ \frac{\partial^2 D^{Bayer}}{\partial U_s \partial V_s} &= \frac{\partial^2 D^{Bayer}}{\partial V_s \partial U_s} = \sum_{k=1}^{4} 2a_k b_k. \end{aligned} \tag{13}$$

The determinant of $H(D^{Bayer})$ is calculated by

$$\begin{aligned} \det H(D^{Bayer}) &= \sum_{k=1}^{4} 2a_k^2 \sum_{k=1}^{4} 2b_k^2 - \sum_{k=1}^{4} 2a_k b_k \sum_{k=1}^{4} 2a_k b_k \\ &= 4\Big(\sum_{k=1}^{4} a_k^2 \sum_{k=1}^{4} b_k^2 - \sum_{k=1}^{4} a_k b_k \sum_{k=1}^{4} a_k b_k\Big) \\ &= 66.1412 > 0 \end{aligned} \tag{14}$$

Because of $\det H(D^{Bayer}) > 0$, it means that $D^{Bayer}(U_s, V_s)$ is semi-positive definite [2]. We complete the proof that the Bayer CFA block distortion function in Eq. (11) is a convex function under $assumption_1$.

ACKNOWLEDGEMENT

The authors appreciate the valuable comments of the Associate Editor and three reviewers to improve the paper and the proofreading help of Ms. C. Harrington.

REFERENCES

[1] B. E. Bayer, "Color imaging array," U.S. Patent 3 971 065 A, Jul. 20, 1976.

[2] K. Binmore and J. Davies, *Calculus: Concepts and Methods*, 2nd ed. Cambridge, U.K.: Cambridge Univ. Press, 2012, p. 190.

[3] H. Chen, M. Sun, and E. Steinbach, "Compression of Bayer-pattern video sequences using adjusted chroma subsampling," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 19, no. 12, pp. 1891–1896, Dec. 2009.

[4] K.-L. Chung, Y.-H. Huang, and C. H. Lin, "Improved universal chroma 4:2:2 subsampling for color filter array video coding in HEVC," *Signal, Image Video Process.*, vol. 11, no. 6, pp. 1041–1048, Jan. 2017.

[5] C. Doutre, P. Nasiopoulos, and K. N. Plataniotis, "H.264-based compression of Bayer pattern video sequences," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 18, no. 6, pp. 725–734, Jun. 2008.

[6] C. Doutre and P. Nasiopoulos, "Modified H.264 intra prediction for compression of video and images captured with a color filter array," in *Proc. 16th IEEE Int. Conf. Image Process.*, Nov. 2009, pp. 3401–3404.

[7] *Execution Codes of Our GDCS Procedures for 4:2:0 and 4:2:2*. Accessed: Sep. 2018. [Online]. Available: ftp://140.118.175.164/GDCS/code

[8] F. Gastaldi, C. C. Koh, M. Carli, A. Neri, and S. K. Mitra, "Compression of videos captured via Bayer patterned color filter arrays," in *Proc. 13th Eur. Signal Process Conf.*, Sep. 2005, pp. 1–4.

[9] D. Kiku, Y. Monno, M. Tanaka, and M. Okutomi, "Residual interpolation for color image demosaicking," in *Proc. IEEE Int. Conf. Image Process.*, Sep. 2013, pp. 2304–2308.

[10] *Kodak True Color Image Collection*. Accessed: Aug. 2018. [Online]. Available: http://r0k.us/graphics/kodak/

[11] S.-Y. Lee and A. Ortega, "A novel approach of image compression in digital cameras with a Bayer color filter array," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Thessaloniki, Greece, Oct. 2001, pp. 482–485.

[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] 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.

[14] H. S. Malvar and G. J. Sullivan, "Progressive-to-lossless compression of color-filter-array images using macropixel spectral-spatial transformation," in *Proc. Data Compress. Conf.*, Apr. 2012, pp. 3–12.

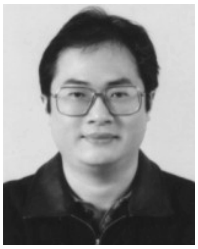
[15] *Spatial Scalability Filters*, document ISO/IEC JTC1/SC29/WG11 ITU-T SG 16 Q.6, Jul. 2005.

[16] 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.

[17] 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.

[18] L. Zhang, X. Wu, A. Buades, and X. Li, "Color demosaicking by local directional interpolation and nonlocal adaptive thresholding," *J. Electron. Imag.*, vol. 20, no. 2, pp. 023016-1–023016-16, Jun. 2011.

[19] 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 has been the Chair Professor with the Department of Computer Science and Information Engineering, National Taiwan University of Science and Technology, Taipei, since 2009. His research interests include deep learning, image processing, and video compression. He was a recipient of the Distinguished Research Award (2004 to 2007) and the Distinguished Research Project Award (2009 to 2012) from the Ministry of Science and Technology, China. In 2017, he received the Scientific Paper Award from the Far Eastern Y. Z. Hsu Science and Technology Memorial Foundation. He has been an Associate Editor of the *Journal of Visual Communication and Image Representation* since 2011.

Yu-Ling Lee received the B.S. degree in computer science and information engineering from Chang Gung University, Taoyuan, Taiwan, in 2017. 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.

Wei-Che Chien received the B.S. degree in computer science and information engineering from National Taiwan Ocean University, Keelung, Taiwan, in 2016. He is currently pursuing the M.S. degree in computer science and information engineering with the National Taiwan University of Science and Technology, Taipei, Taiwan. His research interests include image processing and video compression.