Information Sciences 420 (2017) 386–402

Contents lists available at ScienceDirect

Information Sciences

journal homepage: www.elsevier.com/locate/ins

Temporal and spatial correlation-based reversible data hiding for RGB CFA videos☆

Kuo-Liang Chung[a], Chih-Yi Chiu[b,*], Tzu-Yu Yu[a], Po-Lin Huang[a]

[a] Department of Computer Science and Information Engineering, National Taiwan University of Science and Technology, No. 43, Section 4, Keelung Road, Taipei 10672, Taiwan, ROC

[b] Department of Computer Science and Information Engineering, National Chiayi University, No. 300, Xuefu Rd., Chiayi 60004, Taiwan, ROC

ARTICLE INFO

Article history:
Received 24 November 2016
Revised 25 June 2017
Accepted 20 August 2017
Available online 24 August 2017

Keywords:
Color filter array (CFA)
Difference expansion
Embedding capacity
Quality
Reversible data hiding

ABSTRACT

Considering a mosaic video, in which each mosaic image could be captured by any of the existing eleven RGB color filter array (CFA) structures and each mosaic pixel contains only one primary color, this paper proposes a novel spatial and temporal correlation-based reversible data hiding (RDH) method. In the proposed RDH method, an intellectual switching strategy to select the spatial correlation-based or the temporal prediction scheme to maximize the quality improvement of the marked mosaic videos under fixed embedding capacity. Based on eight test mosaic videos, the experimental results demonstrate that in terms of peak signal-to-noise ratio (PSNR), color PSNR (CPSNR), structure similarity (SSIM) index, motion-based video integrity evaluation (MOVIE) index, and human perceptual effect, the proposed RDH method clearly outperforms the spatial correlation-based RDH method by Yang et al. and the other four state-of-the-art RDH methods.

1. Introduction

Hiding data in a host image is an active research topic in the multimedia area. A large number of data hiding methods have been proposed for gray and color images. In particular, reversible data hiding (RDH) techniques that can completely recover the original image have attracted great interests for military, medical, and archival applications recently. The first RDH method was developed by Barton in 1997 [1]. Following Barton's concept, more and more studies [3–5,10,12,14,35] have focused on how to increase the embedding capacity and/or enhance the quality of marked images.

The development in RDH can be roughly divided into three categories, namely the compression domain-based category, the encrypted domain-based category, and the image domain-based category [24]. In the compression domain-based category, the hidden data is embedded into the compressed coefficients such as the discrete cosine transform (DCT) coefficients [3,7,13,17], wavelet transform coefficients [10], or vector quantization (VQ) indices [4,5]. In the encrypted domain-based category, the hidden data is embedded into the encrypted image, and the original image can be recovered from the decrypted image [38–40] directly. The image domain-based RDH methods embed the hidden data into image pixels. It generally yields

☆ The work of K.-L. Chung was supported by the Ministry of Science and Technology of the R. O. C. under the contract MOST104-2221-E-011-004-MY3. The work of C.-Y. Chiu was supported by the Ministry of Science and Technology of the R. O. C. under the contract MOST104-2628-E-415-002-MY2.

* Corresponding author.
E-mail addresses: klchung01@gmail.com (K.-L. Chung), cychiu@mail.ncyu.edu.tw (C.-Y. Chiu), tzeyu.yo@gmail.com (T.-Y. Yu), bmfish714285@gmail.com (P.-L. Huang).

http://dx.doi.org/10.1016/j.ins.2017.08.064
0020-0255/

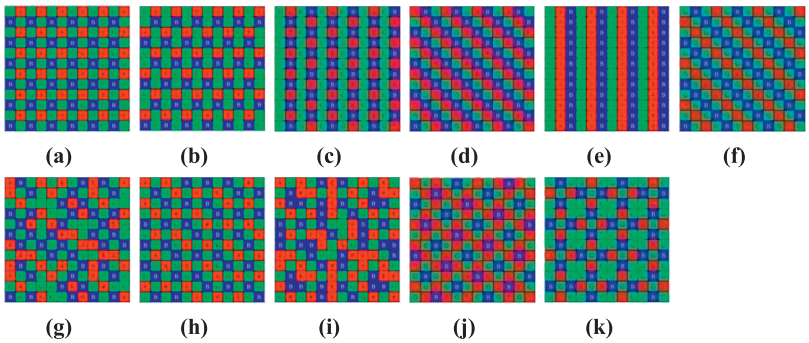

Fig. 1. Eleven typical RGB CFA structures. **(a)** Bayer CFA. **(b)** Lukac and Plataniotis CFA. **(c)** Yamanaka CFA. **(d)** diagonal stripe CFA. **(e)** vertical stripe CFA. **(f)** modified Bayer CFA. **(g)** HVS-based CFA. **(h)** type I pseudo-random CFA. **(i)** type II pseudo-random CFA. **(j)** type III pseudo-random CFA. **(k)** Fujifilm CFA. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

higher embedding capacity and better quality of marked images than the compression domain-based and encrypted domain-based methods. In the image domain-based RDH method, difference expansion (DE) approach [12,22,27,29] and histogram modification approach [14,21,26] are two popular approach. When compared with the histogram modification approach that might have the highest peak signal-to-noise ratio (PSNR) lower bound but might have limited embedding capacity, the DE approach embeds more hidden data into the images and can achieve acceptable quality of marked images.

With the advances in digital color camera technology, for reducing hardware cost, the single-sensor digital color cameras with various RGB color filter array (CFA) structures have been developed, where each mosaic pixel in the captured CFA image consists of only one primary color. The eleven typical RGB CFA structures [18,19] are shown in Fig. 1, in which the cameras Canon EOS-6D and Fujifilm X-Pro1 adopt the Bayer CFA structure [2] in Fig. 1(a) and the Fujifilm CFA structure in Fig. 1(k), respectively. By applying the universal demosaicking process [9,20,34] to an RGB CFA image, the reconstructed (or called demosaicked) RGB full-color image can be obtained. However, for the Bayer CFA image, a quality-oriented demosaicking process [31,37] is needed to reconstruct the full-color image.

1.1. The weakness in existing RDH methods for mosaic images and the motivation

Although the above-mentioned RDH methods are designed to deal with gray and color images, they can also be applied to the RGB mosaic images by embedding hidden data into each split color plane, e.g. the R color plane, B color plane, G_1 color plane constituted by the left-upper G pixels, and G_2 color plane constituted by the right-lower G pixels for the Bayer mosaic image. However, this straightforward approach often degrades the quality of the marked mosaic images. In the experiment section, some experimental results are shown to justify the above weakness for four related RDH methods. In the past, only two RDH methods [32,33] have been developed to deal with mosaic images. In [32], based on the color difference idea, Yang et al. presented a spatial correlation-based RDH method for Bayer CFA images and experimental results showed that their proposed method produces better quality of marked CFA images than the related methods [22,26,27] under the same embedding capacity. Later, in [33], Yang et al. slightly extended their previous RDH work from the Bayer CFA structure to four CFA structures and their RDH method can also be applied to the mosaic videos by embedding hidden data into each mosaic image of the mosaic video. The motivation of this research are threefold: (1) develop a novel spatial and temporal correlation-based RDH method for mosaic videos such that the proposed switching mechanism can select the best strategy, the spatial correlation-based RDH or the temporal correlation-based RDH, to minimize the prediction error, (2) determine the embedding extraction order and the extraction order in the mosaic video such that the embedding and extraction steps are assured to be correct, and (3) extend the proposed temporal and spatial correlation-based RDH method to handle more CFA structures, say 11 CFA structures, by exploiting the universal CFA properties.

1.2. Contributions

In this paper, we propose a novel spatial and temporal correlation-based RDH method to handle up to above-mentioned eleven existing CFA structures of mosaic videos. In addition, an intellectual switching mechanism is proposed to select which scheme, the spatial correlation-based prediction scheme or the temporal correlation-based prediction scheme, as the prediction scheme to minimize the prediction error such that the proposed RDH method can achieve the maximal quality improvement. Further, the proposed RDH method can assure the correctness of the embedding and extraction phases. Based on eight typical test videos, four are low-motion videos and the other four are high-motion ones, we first generate 88 CFA mosaic videos. The experimental results demonstrate that under the same embedding capacity, the proposed RDH method delivers significant quality improvement in terms of the peak signal-to-noise ratio (PSNR), color PSNR (CPSNR), structure similarity (SSIM) index [30], motion-based video integrity evaluation (MOVIE) index [23], and human perceptual effect, of the marked CFA images and reconstructed marked RGB full-color videos. For example, the average PSNR gain of the proposed

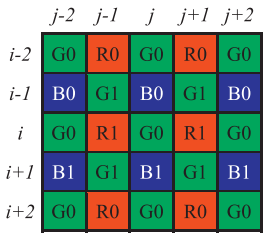

Fig. 2. The depiction of the 2-set partition for the Bayer CFA structure. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

method over Yang et al.'s method [33] is 3.67 dB and the average MOVIE improvement ratio is 36%. In addition, based on eight Bayer CFA test videos, the proposed RDH method is clearly superior to four related RDH methods [6,22,26,27].

The remainder of this paper is organized as follows. In Section 2, we present the proposed spatial and temporal correlation-based prediction scheme for eleven existing CFA structures. In Section 3, the embedding and extraction phases of the proposed RDH method are presented in detail. In addition, a real example is provided to assist the understanding of the proposed method. In Section 4, experimental results are presented to show the significant quality merits of the proposed RDH method. Concluding remarks are given in Section 5.

2. Proposed spatial and temporal correlation-based prediction scheme

In this section, firstly, we introduce the spatial correlation-based prediction scheme in [33] for the concerned CFA structures. Secondly, we propose the temporal correlation-based prediction scheme for the same CFA structures. Thirdly, we take a real video example to provide the statistic foundation to show why proposed spatial and temporal correlation-based prediction scheme has better prediction accuracy when compared with the spatial correlation-based prediction scheme.

2.1. Existing spatial correlation-based prediction scheme

Without loss of generality, we only present the spatial correlation-based prediction scheme for green pixels in a Bayer mosaic image. As shown in Fig. 2, the green pixels are partitioned into two disjoint sets, G_0 and G_1, and they alternate between two consecutive rows. We can predict the color difference values of the G_0 pixels by those in G_1 and vice versa. The merit of the 2-set partition strategy is that embedding hidden data in the G_0 pixels does not affect those in G_1. Therefore, we can recover the original predicted color difference values of the G_0 pixels through G_1's information in the extraction process.

Consider the pixel $G_0(i, j)$ at location (i, j). To predict the color difference value of $G_0(i, j)$, we utilize $G_0(i, j)$'s neighboring information within a $k \times k$ window W_{ij} centered at $G_0(i, j)$. Empirically, W_{ij} has to contain at least two G_1 pixels so that the prediction of the color difference value of the target pixel can be more effective; otherwise, we enlarge W_{ij} until it contains at least two G_1 pixels. Let $\mathcal{K}_{G_1}$, κ_R, and κ_B be the numbers of G_1, R (including R_0 and R_1), and B (including B_0 and B_1) in W_{ij}, respectively. The prediction errors of the G-R and G-B color differences are defined by

$$\begin{cases} \Delta D_{G-R}(i,j) = D_{G-R}(i,j) - \bar{D}_{G-R}(i,j) \\ \Delta D_{G-B}(i,j) = D_{G-B}(i,j) - \bar{D}_{G-B}(i,j) \end{cases}, \tag{1}$$

where

$$\begin{cases} D_{G-R}(p,q) = G(p,q) - \frac{1}{\mathcal{K}_R} \sum\limits_{(x,y)\in W_{pq}} R(x,y) \\ D_{G-B}(p,q) = G(p,q) - \frac{1}{\mathcal{K}_B} \sum\limits_{(x,y)\in W_{pq}} B(x,y) \\ \bar{D}_{G-R}(i,j) = \frac{1}{\mathcal{K}_{G_1}} \sum\limits_{(x,y)\in W_{ij}} D_{G-R}(x,y) \\ \bar{D}_{G-B}(i,j) = \frac{1}{\mathcal{K}_{G_1}} \sum\limits_{(x,y)\in W_{ij}} D_{G-B}(x,y) \end{cases}, \tag{2}$$

For $k = 3$, we have four G_1 pixels, two R pixels, and two B pixels in W_{ij}. Table 1 summarizes the precise numbers of G pixels, R pixels, and B pixels required in the $k \times k$ window W_{ij} when computing the prediction errors of the G-R and G-B color differences of the concerned nine CFA structures, except for the too complicated type II pseudo-random and HVS-based CFA structures shown in Fig. 1(i) and 1(g).

Since a smoother color difference local patch implies a better prediction, we use the local variances of the G-R and G-B color differences to determine the smoother color difference local patch at position (i, j). The local variances are calculated

by

$$\begin{cases} \sigma^2_{G-R}(i,j) = \frac{1}{K_G} \sum\limits_{(x,y)\in W_{ij}} \left(D_{G-R}(x,y) - \bar{D}_{G-R}(i,j)\right)^2 \\ \sigma^2_{G-B}(i,j) = \frac{1}{K_G} \sum\limits_{(x,y)\in W_{ij}} \left(D_{G-B}(x,y) - \bar{D}_{G-B}(i,j)\right)^2 \end{cases} \tag{3}$$

Thus, the final prediction error of pixel $G_0(i, j)$ is determined by selecting the one predicted by color difference local patch with a smaller local variance

$$\Delta D_G(i,j) = \begin{cases} \Delta D_{G-R}(i,j) \ if \ \sigma^2_{G-R}(i,j) < \sigma^2_{G-B}(i,j), \\ \Delta D_{G-B}(i,j) \ otherwise. \end{cases} \tag{4}$$

Similarly, the color difference prediction errors of pixel R(i, j) and pixel B(i, j) can be obtained by

$$\begin{cases} \Delta D_R(i,j) = \Delta D_{R-G}(i,j) = D_{R-G}(i,j) - \bar{D}_{R-G}(i,j) \\ \Delta D_B(i,j) = \Delta D_{B-G}(i,j) = D_{B-G}(i,j) - \bar{D}_{B-G}(i,j) \end{cases}. \tag{5}$$

Note that only the R-G (red-green) and B-G (blue-green) color difference local patches are considered for the R and B pixels, respectively, because G pixels are the majority in the three color channels.

2.2. Proposed spatial and temporal correlation-based prediction scheme

This section proposes a novel spatial and temporal correlation-based prediction scheme. For easy exposition, firstly, we present how to find the best matched block in the $(t+1)$th image from the current block in the tth image such that the matched block-pair has the minimal sum of absolute difference (SAD). Note that for assuring the correctness of the extracted hidden data, we search the best matched block in the next image, not in the previous one. Secondly, we propose a switching strategy to determine the spatial correlation-based prediction scheme or the temporal correlation-based prediction scheme for the embedding hidden bit.

Let $G_0^t(i, j)$ be the G_0 pixel at position (i, j) in the tth image of the CFA video, and let M_{ij}^t be the $u \times u$ mosaic block centered at position (i,j). For M_{ij}^t, we perform a block matching process to find the best matched block $M_{i^*j^*}^{t+1}$ within a $v \times v$ search window range centered at position (i^*, j^*) in the $(t+1)$th image such that the following SAD criterion is minimized and (i^*, j^*) is given by

$$(i^*, j^*) = \underset{(i',j')}{argmin} \sum_{i'=i-\lfloor v/2 \rfloor}^{i+\lfloor v/2 \rfloor} \sum_{j'=j-\lfloor v/2 \rfloor}^{j+\lfloor v/2 \rfloor} \mathrm{SAD}\left(M_{ij}^t, M_{i'j'}^{t+1}\right), \tag{6}$$

with

$$\mathrm{SAD}\left(M_{ij}^t, M_{i'j'}^{t+1}\right) = \sum_{\Delta i=-\lfloor u/2 \rfloor}^{\lfloor u/2 \rfloor} \sum_{\Delta j=-\lfloor u/2 \rfloor}^{\lfloor u/2 \rfloor} \left| C^t(i+\Delta i, j+\Delta j) - C^t\left(i'+\Delta i, j'+\Delta j\right) \right| \tag{7}$$

where $C^t(i, j)$ represents the color value at position (i, j) in the tth image. Here, the color patterns of M_{ij}^t and $M_{i^*j^*}^{t+1}$ must be the same. Besides, G_0 pixels are not included in the above SAD calculation process since the G_0 set is currently used for error prediction. Fig. 3 gives a real example to illustrate how to find the best matched block $M_{i^*j^*}^{t+1}$ of M_{ij}^t for the Fujifilm CFA structure. Empirically, setting $u = 3$ and $v = 32$ leads to good performance to all the test images. After determining the best matched position (i^*,j^*), the prediction error of the proposed temporal correlation-based prediction scheme is defined by

$$D_G(i,j) = G_0^t(i,j) - G_0^{t+1}(i^*, j^*) \tag{8}$$

Combining the temporal and spatial correlations, given a threshold θ_{SAD}, the final prediction error is determined by

$$\begin{cases} \Delta D_G(i,j) \text{ by Eq. (8)} & if \ \mathrm{SAD}\left(M_{ij}^t, M_{i'j'}^{t+1}\right) < \theta_{\mathrm{SAD}}, \\ & (\textbf{temporal correlation}) \\ \Delta D_G(i,j) \text{ by Eq. (4)} & otherwise. \\ & (\textbf{spatial correlation}) \end{cases} \tag{9}$$

In Eq. (9), if $\mathrm{SAD}(M_{ij}^t, M_{i'j'}^{t+1})$ is smaller than a specified threshold θ_{SAD}, which is set to be 20 empirically to achieve good performance, it indicates the time domain difference between the corresponding blocks is small, and the temporal correlation-based prediction scheme is selected; otherwise, the spatial correlation-based prediction scheme is selected. To help the reader understand the proposed process, the flowchart of the proposed spatial and temporal correlation-based prediction scheme is illustrated in Fig. 4.

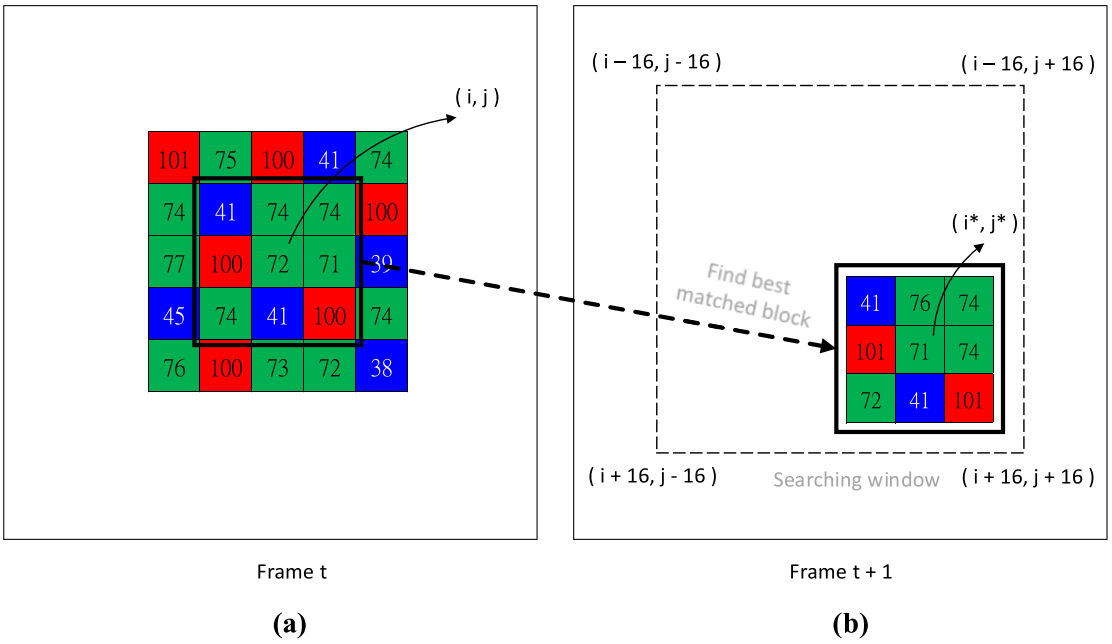

Fig. 3. The depiction of block matching to find the best matched block $M_{i^*j^*}^{t+1}$ of M_{ij}^{t} for the Fujifilm CFA structure. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

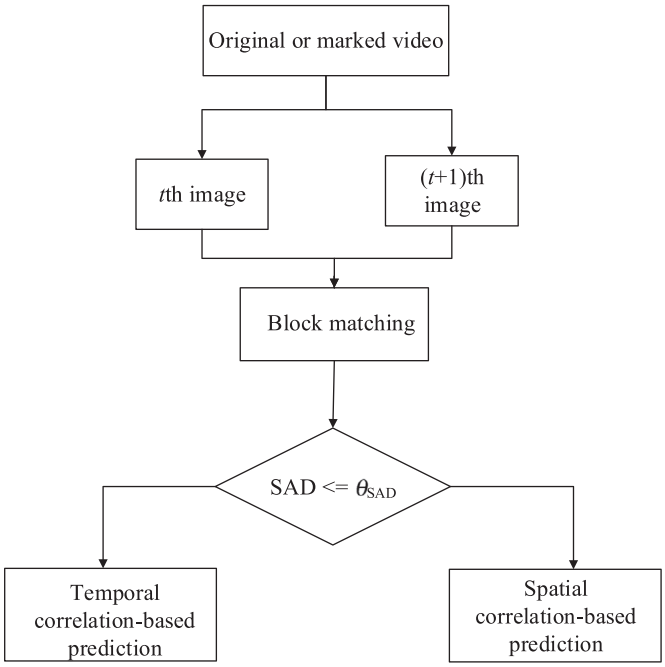

Fig. 4. The flowchart of the spatial and temporal correlation-based prediction scheme.

2.3. Statistic distribution to show the prediction error reduction merit of the proposed scheme

We first give a real example, as shown in Fig. 3, to illustrate the prediction error reduction merit of the proposed error prediction scheme. It's known that the computation for SAD has to exclude the pixels belonging to the G_0 set. In Fig. 3(b), we obtain $\mathrm{SAD}(M_{ij}^{t}, M_{i^*j^*}^{t+1}) = 6\ (=|41-41|+|74-76|+|74-74|+|100-101|+|74-72|+|41-41|+|100-101|)$. According to Eq. (9), because $\mathrm{SAD}(M_{ij}^{t}, M_{i^*j^*}^{t+1})(=6) < \theta_{\mathrm{SAD}}(=20)$, we select the temporal correlation-based prediction scheme and the prediction error is $\Delta D_G(i,j) = 1(=72-71)$. For this example, if we select the spatial correlation-based prediction

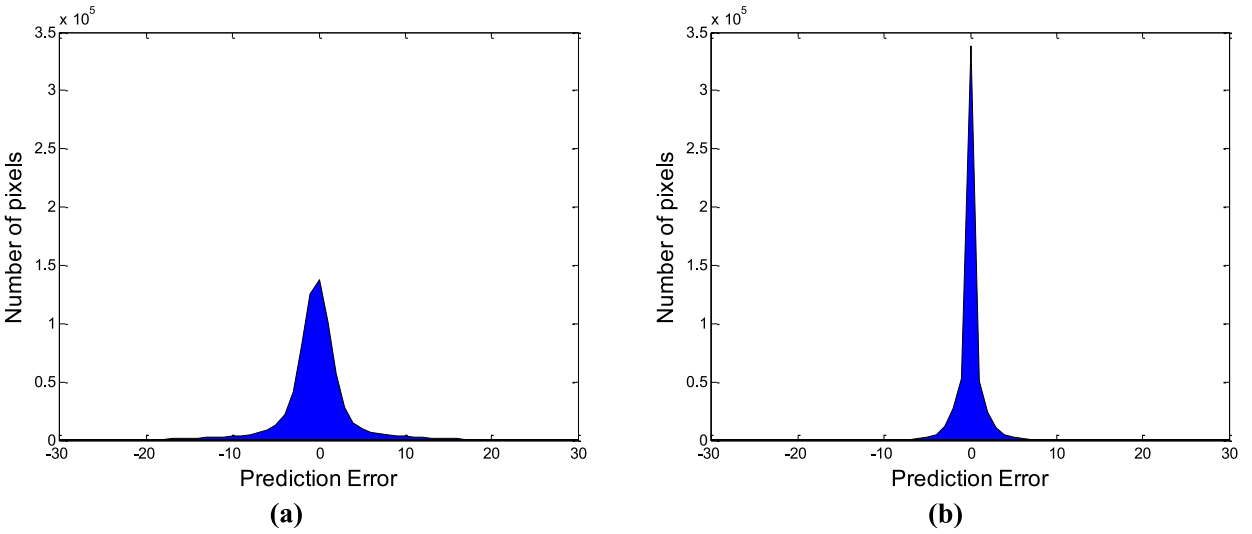

Fig. 5. Two Laplacian distributions for **(a)** Yang et al.'s spatial domain-based prediction scheme and **(b)** the proposed temporal-spatial based prediction scheme (source data from "Akiyo" Bayer CFA video, bpp = 0.1–0.7).

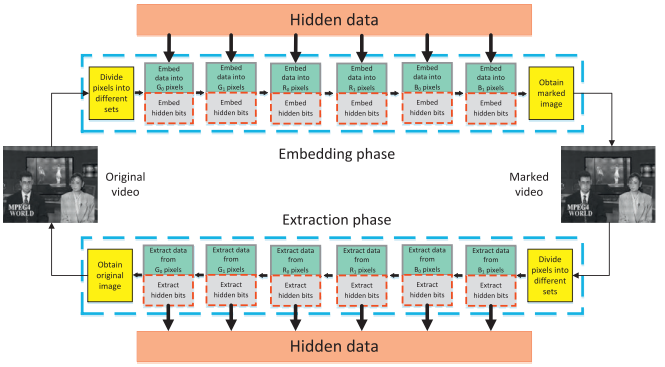

Fig. 6. The flow of the proposed RDH method.

scheme [28], by Eqs. (1) and (2), it yields two larger prediction errors, $\Delta D_{G-R}(i,j) = D_{G-R}(i,j) - \tilde{D}_{G-R}(i,j) = -28 - (-26) = -2$ and $\Delta D_{G-B}(i,j) = -1.67$.

Further, we take the Akiyo video, which is a low-motion video, as the example. Therefore, we expect the prediction error between adjacent frames in the video is low, and it is suitable for embedding data by exploiting the temporal correlation-based prediction approach. Fig. 5(a) plots the Laplacian distribution of the prediction errors for Yang et al.'s prediction scheme [33], in which X-axis denotes the prediction error and the Y-axis denotes the number of pixels occurred. Fig. 5(b) illustrates the Laplacian distribution of the prediction errors for the proposed prediction scheme. We clearly observe that Fig. 5(b) is much sharper than Fig. 5(a), and the sharper statistic distribution of the proposed prediction scheme implies a better prediction error reduction effect. On the other hand, it leads to the embedding capacity and marked video quality merits of the proposed prediction scheme.

3. The propsed RDH method

In this section, we present the embedding and extraction phases of the proposed RDH method. Fig. 6 illustrates the flow of the proposed RDH method. In the embedding phase, the hidden data are embedded in the CFA images according to the order: $I_1 \rightarrow I_2 \rightarrow \ldots \rightarrow I_t \rightarrow \ldots \rightarrow I_n$, where I_t is the tth image and n is the number of images. For the tth image I_t, the color sets are processed in the following order: $G_0^t \rightarrow G_1^t \rightarrow R_0^t \rightarrow R_1^t \rightarrow B_0^t \rightarrow B_1^t$. In the extraction phase, the embedded data are extracted from the CFA images in the reverse order: $I_n \rightarrow I_{n-1} \rightarrow \ldots \rightarrow I_t \rightarrow \ldots \rightarrow I_1$, and the color sets to be processed in each frame are also arranged in the reverse order: $B_1^t \rightarrow B_0^t \rightarrow R_1^t \rightarrow R_0^t \rightarrow G_1^t \rightarrow G_0^t$. In the embedding phase, when embedding data into the current color set, e.g. G_0, the proposed prediction scheme, as shown in Fig. 4, is performed on the current color set.

Algorithm 1. Hidden data embedding.

Input: The original tth CFA image and a hidden data set H.

Output: The marked tth CFA image.

Step 1: For the pixel $G_0^t(i, j)$, apply Eq. (7) to find M_{ij}^t's best matched block $M_{i^*j^*}^{t+1}$ at position (i^*, j^*) in the $(t+1)$th image.

Step 2: Determine $\Delta D_G(i, j)$ based on the temporal and spatial correlations given in Eq. (9).

Step 3: Embed H into the pixel $G_0^t(i, j)$ based on the difference expansion concept as follows

$$\Delta\overline{D_G}(i,j) = \begin{cases} \Delta D_G(i,j) + \theta_{err} + 1 \ if\ \Delta D_G(i,j) > \theta_{err} \\ \Delta D_G(i,j) - \theta_{err} - 1 \ if\ \Delta D_G(i,j) < -\theta_{err} - 1 \\ 2 \cdot \Delta D_G(i,j) + h \ otherwise \end{cases} \quad (10)$$

where $h \in \{0, 1\}$ is a binary bit of H to be embedded, and θ_{err} is a prediction error threshold to decide whether h is embedded in $G_0^t(i, j)$ or not.

Step 4: Let $\overline{G_0^t}(i, j)$ be the marked pixel value at position (i, j) after embedding. Compute $\overline{G_0^t}(i, j)$ based on the following two cases

Case 1: (temporal correlation)

If $SAD(M_{ij}^t, M_{i^*j^*}^{t+1}) < \theta_{SAD}$, $\overline{G_0^t}(i, j) = \Delta\overline{D_G}(i, j) + G_0^{t+1}(i^*, j^*)$. (11)

Case 2: (spatial correlation)

Otherwise, $\overline{G_0^t}(i, j) =$

$$\begin{cases} \Delta\overline{D_G}(i,j) + \frac{1}{K_R} \sum_{(x,y)\in W_{ij}} R(x,y) + \frac{1}{K_G} \sum_{(x,y)\in W_{ij}} D_{G-R}(x,y) \ if\ \sigma_{G-R}^2(i,j) < \sigma_{G-B}^2(i,j) \\ \Delta\overline{D_G}(i,j) + \frac{1}{K_B} \sum_{(x,y)\in W_{ij}} B(x,y) + \frac{1}{K_G} \sum_{(x,y)\in W_{ij}} D_{G-B}(x,y) \ otherwise \end{cases} \quad (12)$$

3.1. The embedding phase

Given the tth image, we apply the difference expansion technique [27] to embed a hidden data set H in the color set G_0^t first. If G_0^t cannot consume all hidden data, we embed the remainder of H into the next color set G_1^t, and so on. The embedding procedure is detailed in Algorithm 1 as follows. For a concise presentation, only G_0^t is addressed in the proposed method. Other color sets are processed in the same way. After processing all the color sets in the tth frame, we then proceed to the $(t+1)$th image and repeat the above embedding procedure starting from the color set G_0^{t+1} to embed the remainder.

The prediction error threshold θ_{err} shown in Eq. (10) controls the embedding capacity in RDH. Using a higher θ_{err} can increase the probability to embed hidden data into pixels. θ_{err} is set to 0 initially. If we cannot embed all hidden data into the CFA video, θ_{err} is increased by one and restart the embedding procedure again. The process is repeated until all hidden data are embedded into the CFA video.

3.2. The extraction phase

In the extraction phase, we start from the nth image to process the CFA video in the reverse order, where in the nth image, we only take the spatial correlation-based prediction scheme to extract the hidden data because there is no next image frame for reference. In the tth image frame, we process the color sets in the following order: $B_1^t \to B_0^t \to R_1^t \to R_0^t \to G_1^t \to G_0^t$, and then continue to process the $(t-1)$th image frame. To be consistent, we again take G_0^t for illustration in the extraction procedure, as listed in Algorithm 2.

3.3. Real embedding and extraction example

We utilize the same example in Section 2.3 to explain how the proposed embedding and extraction algorithm works. Consider to embed the hidden data $h = 1$ into the pixel $G_0^t(i, j)$. Assume the temporal correlation-based embedding scheme is applied and $\Delta D_G(i, j) = 1$, $\theta_{err} = 2$. According to Step 3 of Algorithm 1, since $-\theta_{err} - 1 < \Delta D_G(i, j) \le \theta_{err}$, h can be embedded in the pixel $G_0^t(i, j)$ and we have $\Delta\overline{D_G}(i, j) = 3$. At Step 4, the marked pixel value becomes $\overline{G_0^t}(i, j) = \Delta\overline{D_G}(i, j) + G_0^{t+1}(i^*, j^*) = 3 + 71 = 74$. To extract the hidden data and recover from the marked pixel $\overline{G_0^t}(i, j)$, we find that the temporal correlation-based scheme is selected based on the SAD criterion and we calculate $\Delta\overline{D_G}(i, j) = \overline{G_0^t}(i, j) - G_0^{t+1}(i^*, j^*) = 74 - 71 = 3$ at Step 2 of Algorithm 2. Since $-2\theta_{err} - 2 \le \Delta\overline{D_G}(i, j) \le 2\theta_{err} + 1$, it means the hidden data is embedded in the pixel. At Step 3, the hidden data is $h = 1$ $(= \Delta\overline{D_G}(i, j)\ mod\ 2 = 3\ mod\ 2 = 1)$. Finally, the recovered pixel value $G_0^t(i, j)$ is equal to 72 $(= \Delta D_G(i, j) + G_0^{t+1}(i^*, j^*) = 1 + 71 = 72)$.

3.4. Computational complexity analysis

The time complexity of the two algorithms are proportional to the size of the hidden data set. For simplicity, we assume one binary bit to be processed, and give the following analysis for the embedding algorithm, i.e., Algorithm 1. At Step 1, the block matching technique is performed for a pixel. Recall the block size and search window range are $u \times u$ and $v \times v$,

Algorithm 2. Hidden data extraction.

Input: The marked tth CFA image.

Output: The hidden data set H and recovered tth CFA image.

Step 1: For the marked pixel $\overline{G_0^t}(i,j)$, apply Eq. (7) to find the best matched block of M_{ij}^t at position (i^*, j^*) in the $(t+1)$th image, namely $M_{i^*j^*}^{t+1}$.

Step 2: Determine $\Delta\overline{D_G}(i,j)$ based on the following two cases

Case 1: (temporal correlation)

$$\text{If } \mathrm{SAD}(M_{ij}^t, M_{i^*j^*}^{t+1}) < \theta_{\mathrm{SAD}},\ \Delta\overline{D_G}(i,j) = \overline{G_0^t}(i,j) - G_0^{t+1}(i^*,j^*). \quad (13)$$

Case 2: (spatial correlation)

Otherwise, $\Delta\overline{D_G}(i,j) =$

$$\begin{cases} \overline{G_0^t}(i,j) - \frac{1}{K_R}\sum_{(x,y)\in W_{ij}} R(x,y) - \frac{1}{K_G}\sum_{(x,y)\in W_{ij}} D_{G-R}(x,y) \ if\ \sigma_{G-R}^2(i,j) < \sigma_{G-B}^2(i,j) \\ \overline{G_0^t}(i,j) - \frac{1}{K_B}\sum_{(x,y)\in W_{ij}} B(x,y) - \frac{1}{K_G}\sum_{(x,y)\in W_{ij}} D_{G-B}(x,y) \ otherwise \end{cases} \quad (14)$$

Step 3: If $-2\theta_{err} - 2 \le \Delta\overline{D_G}(i,j) \le 2\theta_{err} + 1$, extract the hidden data by

$$h = \Delta\overline{D_G}(i,j) \bmod 2\ . \quad (15)$$

Step 4: Recover the prediction error of the original pixel $\Delta D_G(i,j)$ by

$$\Delta D_G(i,j) = \begin{cases} \Delta\overline{D_G}(i,j) - \theta_{err} - 1 \ if\ \Delta\overline{D_G}(i,j) > 2\theta_{err} + 1 \\ \Delta\overline{D_G}(i,j) + \theta_{err} + 1 \ if\ \Delta\overline{D_G}(i,j) < -2\theta_{err} - 2 \\ \lfloor \Delta\overline{D_G}(i,j)/2 \rfloor \ \text{otherwise} \end{cases} \quad (16)$$

Step 5: Recover the marked pixel to the original pixel value $G_0^t(i,j)$ based on the following two cases

Case 1: (temporal correlation)

$$\text{If } \mathrm{SAD}(M_{ij}^t, M_{i^*j^*}^{t+1}) < \theta_{\mathrm{SAD}},\ G_0^t(i,j) = \Delta D_G(i,j) + G_0^{t+1}(i^*,j^*). \quad (17)$$

Case 2: (spatial correlation)

Otherwise, $G_0^t(i,j) =$

$$\begin{cases} \Delta D_G(i,j) + \frac{1}{K_R}\sum_{(x,y)\in W_{ij}} R(x,y) + \frac{1}{K_G}\sum_{(x,y)\in W_{ij}} D_{G-R}(x,y) \ if\ \sigma_{G-R}^2(i,j) < \sigma_{G-B}^2(i,j) \\ \Delta D_G(i,j) + \frac{1}{K_B}\sum_{(x,y)\in W_{ij}} B(x,y) + \frac{1}{K_G}\sum_{(x,y)\in W_{ij}} D_{G-B}(x,y) \ otherwise \end{cases} \quad (18)$$

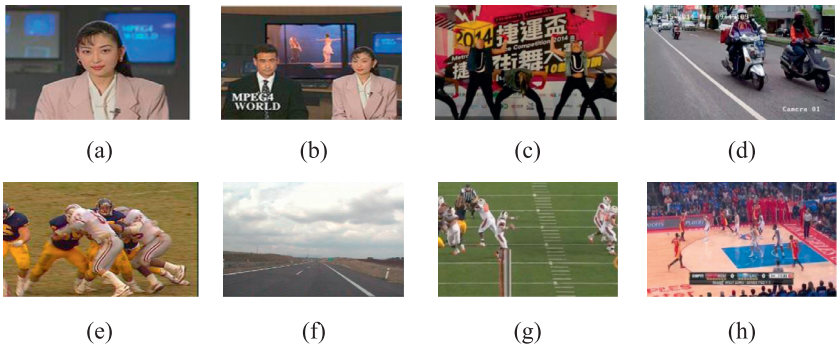

Fig. 7. Eight test videos. (a) Akiyo. (b) News. (c) Dance. (d) Supervisor. (e) Football. (f) Highway. (g) NFL. (h) Basketball.

respectively; the time complexity is thus $O(u^2v^2)$. Steps 2–4 all take a constant time $O(1)$ despite which correlation-based prediction scheme is used. Similarly, Algorithm 2 takes $O(u^2v^2)$ time for block matching in Step 1 and $O(1)$ in remaining steps. It is clear that both algorithms have the same time complexity $O(u^2v^2)$, which is dominated by the block matching technique. The selection of u and v makes the tradeoff between the algorithm accuracy and efficiency. In this study, we favor the efficiency and prefer to use the small block size, i.e. $u = 3$, and use the typical search window size, i.e. $v = 32$, in the experiments.

4. Experimental results

To evaluate the performance among the proposed RDH method and some related RDH methods, thorough experiments are carried out to justify the quality merit of the proposed RDH method. As shown in Fig. 7, eight videos are taken from the website in [36] and YouTube as the test videos. They are further divided as low-motion videos (Akiyo, News, Dance, and Supervisor) and high-motion videos (Football, Highway, NFL, and Basketball). For each video, we excerpt ten consecutive images and generate eleven CFA videos based on Fig. 1, and a total of 88 videos are served as the test videos. The frame

Table 1
The values of $\mathcal{K}_{G_1}$, $\mathcal{K}_R$, and $\mathcal{K}_B$ used in Eq. (2) for nine CFA structures ('2/3' means it can be 2 or 3 depending on the structure).

RGB CFA	$\mathcal{K}_{G_1}$	$\mathcal{K}_R$	$\mathcal{K}_B$
Bayer	4	2	2
Lukac and Plataniotis	3	2/3	2/3
Yamanaka	2	3	3
Diagonal stripe	2	3	3
Vertical stripe	2	3	3
Modified Bayer	4	2	2
Type I pseudo-random	4	2	2
Type III pseudo-random	3	3	2
Fujifilm	4	2	2

resolutions vary in different videos: for Akiyo, News, Highway, and Football, each frame is of size 352×288; for Dance and NFL, each frame is of size 768×512; for Supervisor and Basketball, each frame is of size 1280×720. The proposed method was implemented by C++ and run on a Windows 7 operating system with Intel Core i7 4790 CPU and 4GB RAM.

4.1. PSNR and SSIM quality comparison of marked CFA images

Under the same embedding capacity, two image quality metrics, PSNR and SSIM, are used to measure the quality between the marked CFA image and the original CFA image. For a marked image I with size $M \times N$, the embedding capacity is measured by

$$\text{Capacity}(I) = \frac{\#\{\textit{hidden bits}\}}{MN}, \tag{19}$$

where #{*hidden bits*} denotes the number of hidden bits embedded in the marked image. The PSNR metric is calculated by

$$\text{PSNR}(I) = 10\log_{10}\frac{255^2 MN}{\sum_{i=0}^{M-1}\sum_{j=0}^{N-1}\left(V(i,j)-\bar{V}(i,j)\right)^2}, \tag{20}$$

where $V(i, j)$ and $\bar{V}(i, j)$ are the original and marked CFA pixel values at position (i, j).

Recall that in Section 2.2, the threshold θ_{SAD} is used to determine which correlation, temporal or spatial, is selected as the prediction scheme. The threshold θ_{SAD} is actually a key parameter influencing the image quality of the marked mosaic video. We have tried various θ_{SAD} ranging from 0 to 60. Particularly, when setting $\theta_{\text{SAD}} = 0$, all mosaic pixels are predicted by the spatial correlation-based prediction scheme. On the contrary, setting a larger θ_{SAD} will increase the probability to select the temporal correlation-based prediction scheme. Fig. 8 plots PSNR under various θ_{SAD} for the eight test videos. Each subfigure shows eleven curves corresponding to the eleven CFA structures of the test video, each of which represents the average PSNR from 0.1bpp to 0.7bpp capacity. In each test video, these CFA structures demonstrate close trends and variations along with θ_{SAD}. In fact, we find θ_{SAD} has different impacts on low-motion videos in Figs. 7(a)–(d) and high-motion videos in Figs. 7(e)–(h). We see a burst improvement on PSNR when $\theta_{\text{SAD}} > 0$ for low-motion videos, and it generally converges to a stable status when $\theta_{\text{SAD}} \geq 20$, indicating a good positive effect by integrating the spatial and temporal correlations. Such an integration still works well for high-motion videos, but the effect is not as much as that in the low-motion videos. The reason is that one block in the high-motion video may move and vary rapidly, so it is difficult to find the best matched block such that the SAD is small enough. Hereinafter, we set $\theta_{\text{SAD}} = 20$ for the remaining experiments. Although it is not optimal choice for every video, it leads to the best quality in average in our experiment.

In what follows, under the embedding capacity ranging from 0.1 bpp to 0.7 bpp, we first compare the PSNR performance among the concerned RDH methods. For fairness, we extend Yang et al.'s spatial correlation-based RDH method [33] so that it could handle the concerned eleven CFA structures. Tables 2 and 3 list the PSNR results of the proposed method and Yang et al.'s method for low-motion video "Akiyo" and high-motion video "NFL". The PSNR improvement of the proposed method is 0.82–9.8 dB for "Akiyo" and 0.7 dB–1.7 dB for "NFL". Table 4 shows that the average PSNR gain of the proposed method is 3.67 dB when compared to Yang et al.'s method. Moreover, Tables 5 and 6 indicate that the proposed method yields significant PSNR improvement, 7.24 dB, for the four low-motion videos, but the proposed method yields only 0.11 dB PSNR improvement for the four high-motion videos.

Besides PSNR, the SSIM index is also included to evaluate the structure similarity quality of the marked CFA images. The SSIM index is defined by

$$\text{SSIM}(x, y) = \frac{(2\mu_x\mu_y + c_1)(2\sigma_{xy} + c_2)}{\left(\mu_x^2 + \mu_y^2 + c_1\right)\left(\sigma_x^2 + \sigma_y^2 + c_2\right)} \tag{21}$$

where x and y denote the marked CFA image and the original CFA image, respectively; μ_x and μ_y denote the average pixel values of x and y, respectively; σ_x^2 and σ_y^2 denote the variances of x and y, respectively; σ_{xy} is the co-variance of x and y,

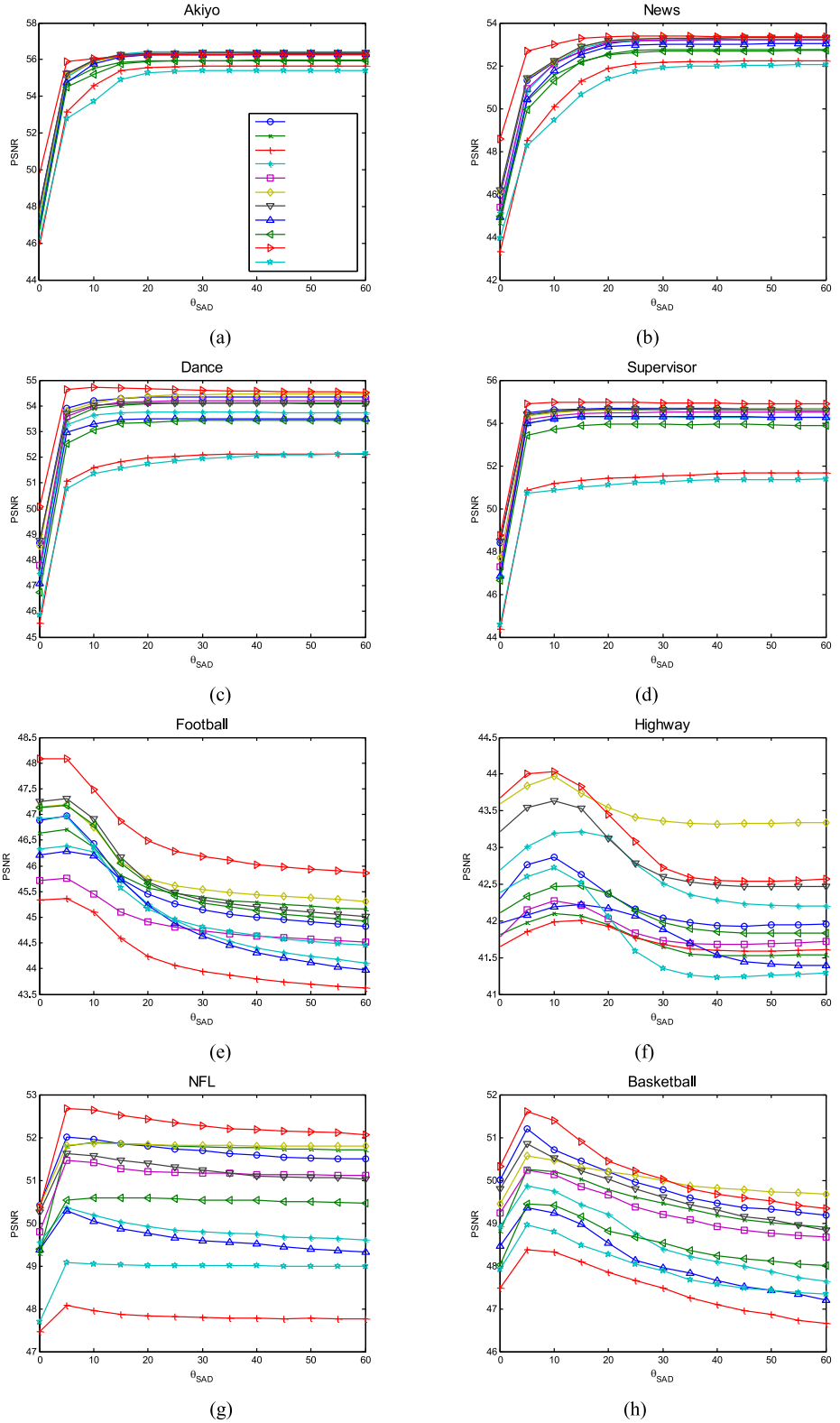

Fig. 8. The PSNR against θ_{SAD} for the eight test videos.

Table 2
The PSNRs for video "Akiyo".

CFA	Method	Capacity (bpp)						
		0.1	*0.2*	*0.3*	*0.4*	*0.5*	*0.6*	*0.7*
Bayer	*Yang*	55.3698	52.0354	49.9321	47.5723	45.2523	43.849	42.2859
	Proposed	**63.6153**	**58.5352**	**57.1967**	**55.3092**	**54.0431**	**53.0548**	**52.2126**
Lukac and Plataniotis	*Yang*	55.5298	52.0996	49.7759	47.3777	44.9787	43.4319	41.3304
	Proposed	**63.6408**	**58.5700**	**57.2282**	**55.3258**	**54.0378**	**53.0217**	**52.1217**
Yamanaka	*Yang*	55.5415	52.0042	49.0016	46.283	43.8024	41.1208	35.9728
	Proposed	**63.6657**	**58.5850**	**57.1776**	**55.211**	**53.8854**	**51.2516**	**47.2764**
Diagonal stripe	*Yang*	54.6208	50.9973	49.1998	46.6418	44.0846	41.6438	39.7468
	Proposed	**62.4897**	**58.8470**	**56.7308**	**55.0605**	**53.6583**	**52.7231**	**51.9646**
Vertical stripe	*Yang*	56.6619	53.1071	51.0226	49.7826	48.0578	45.6301	44.3647
	Proposed	**62.6653**	**59.0248**	**56.9039**	**55.4028**	**54.0825**	**53.1996**	**52.483**
Modified Bayer	*Yang*	55.0420	51.7199	49.7815	47.3582	45.0258	43.4685	41.3731
	Proposed	**63.5830**	**58.5051**	**57.1687**	**55.2769**	**54.0377**	**53.0643**	**52.2089**
HVS-based	*Yang*	54.5315	51.2136	49.3164	46.8354	44.4392	42.9433	40.8411
	Proposed	**63.4503**	**58.9740**	**57.1741**	**55.4411**	**54.1808**	**53.2067**	**52.3148**
Type I Pseudo-random	*Yang*	54.1689	50.9892	49.2169	46.7084	44.3291	42.8849	40.8021
	Proposed	**62.8725**	**59.0472**	**56.8763**	**55.5353**	**54.0652**	**53.1575**	**52.2052**
Type II Pseudo-random	*Yang*	55.2171	51.8609	49.8999	47.5398	45.2624	43.8804	42.3426
	Proposed	**63.6536**	**58.5448**	**57.2329**	**55.3376**	**54.1220**	**53.1694**	**52.3558**
Type III Pseudo-random	*Yang*	54.6039	51.3913	48.4467	46.9534	44.0144	42.307	40.1128
	Proposed	**63.2238**	**58.9374**	**56.5662**	**55.4690**	**54.0266**	**52.8301**	**50.1648**
Fuji RGB	*Yang*	54.8481	51.4262	48.4193	45.7955	43.3298	40.9832	36.5279
	Proposed	**63.7390**	**58.7503**	**57.1166**	**55.4059**	**53.9436**	**51.4782**	**48.6385**
Average	*Yang*	55.1127	51.7148	49.3870	47.0703	44.7117	42.7676	40.2770
	Proposed	**63.3545**	**58.7409**	**57.0437**	**55.3279**	**54.0117**	**52.7597**	**51.1696**

Table 3
The PSNRs for video "NFL".

CFA	Method	Capacity (bpp)						
		0.1	*0.2*	*0.3*	*0.4*	*0.5*	*0.6*	*0.7*
Bayer	*Yang*	57.1167	53.9446	52.0734	50.5676	48.2910	46.6722	44.2470
	Proposed	**57.8780**	**55.3970**	**53.7458**	**52.2552**	**50.8657**	**47.8759**	**44.5701**
Lukac and Plataniotis	*Yang*	56.9071	53.7095	51.8467	50.2188	47.9125	45.3176	42.6720
	Proposed	**57.6893**	**55.1412**	**53.5223**	**51.9851**	**50.1105**	**46.5227**	**43.4816**
Yamanaka	*Yang*	56.8675	53.5273	51.7606	48.9290	46.1404	41.3071	35.3411
	Proposed	**57.8494**	**55.3174**	**53.625**	**51.3401**	**46.8744**	**42.1599**	**35.8343**
Diagonal stripe	*Yang*	57.0272	53.2382	51.0718	49.5737	47.2715	44.5959	42.2278
	Proposed	**58.5256**	**55.4072**	**53.5294**	**52.125**	**50.8197**	**47.7797**	**44.5928**
Vertical stripe	*Yang*	57.8429	54.1928	51.9997	50.5393	48.3071	45.6304	44.0496
	Proposed	**58.7969**	**55.7081**	**53.891**	**52.5397**	**51.3269**	**48.6009**	**46.0874**
Modified Bayer	*Yang*	57.0846	53.8254	52.1123	50.6785	48.1973	46.7756	44.0930
	Proposed	**57.7374**	**55.1452**	**53.5837**	**52.2281**	**51.0640**	**48.1853**	**44.8811**
HVS-based	*Yang*	56.8625	53.4713	51.3073	49.9768	47.4990	44.9983	42.3929
	Proposed	**56.9654**	**53.9026**	**52.0737**	**50.6948**	**47.7816**	**45.1414**	**42.4811**
Type I Pseudo-random	*Yang*	57.1354	53.9040	52.0172	50.5311	48.1530	45.8549	44.2098
	Proposed	**57.6355**	**54.9612**	**53.2725**	**51.8848**	**50.2922**	**47.2904**	**44.4589**
Type II Pseudo-random	*Yang*	56.7861	53.1230	51.1810	49.8416	47.3940	44.9612	**42.3883**
	Proposed	**57.1466**	**53.8692**	**51.9511**	**50.5135**	**47.5596**	**44.9952**	42.2797
Type III Pseudo-random	*Yang*	57.0013	53.5102	51.5724	50.2025	47.8082	44.5347	41.1164
	Proposed	**57.8886**	**54.8113**	**53.0135**	**51.7353**	**48.8014**	**45.7635**	**42.0782**
Fuji RGB	*Yang*	56.4877	53.3415	51.5574	48.7221	**45.8585**	41.4106	34.9120
	Proposed	**56.7966**	**54.1561**	**52.4979**	**49.1926**	45.5641	**41.4730**	**35.1121**
Average	*Yang*	56.9772	53.5892	51.6485	49.8662	47.3642	44.5102	41.1916
	Proposed	**57.6855**	**54.8617**	**53.1399**	**51.4673**	**49.0456**	**45.7759**	**41.9519**

Table 4
The average PSNRs for the eight test videos.

Capacity (bpp)	*Yang*	*Proposed*
0.1	55.0760	**57.9804**
0.2	51.5567	**54.4516**
0.3	49.1051	**52.1575**
0.4	46.8115	**50.3450**
0.5	44.3870	**48.5479**
0.6	41.9303	**46.5296**
0.7	39.2397	**43.8163**
Average	46.8723	**50.5469**

Table 5
The average PSNRs for low-motion videos.

Capacity (bpp)	*Yang*	*Proposed*
0.1	55.1337	**60.7897**
0.2	51.8958	**57.4763**
0.3	49.4780	**55.5056**
0.4	46.9777	**53.9877**
0.5	44.3466	**52.5564**
0.6	41.6155	**50.7484**
0.7	38.5073	**47.5808**
Average	46.8507	**54.0921**

Table 6
The average PSNRs for high-motion videos.

Capacity (bpp)	*Yang*	*Proposed*
0.1	55.0183	**55.1711**
0.2	51.2176	**51.4270**
0.3	48.7322	**48.8094**
0.4	46.6453	**46.7024**
0.5	44.4275	**44.5394**
0.6	42.2452	**42.3109**
0.7	39.9720	**40.0518**
Average	46.8940	**47.0017**

Table 7
The mean SSIM indexes of Yang et al.'s method and the proposed method for the eleven CFA structures (embedding capacity = 0.7 bpp).

RGB CFA Structure	Yang	Proposed
Bayer	0.9787	**0.9841**
Lukac and Plataniotis	0.9737	**0.9823**
Yamanaka	0.9696	**0.9772**
Diagonal stripe	0.9707	**0.9827**
Vertical stripe	0.9833	**0.9873**
Modified Bayer	0.9800	**0.9869**
HVS-based	0.9757	**0.9850**
Type I Pseudo-random	0.9814	**0.9865**
Type II Pseudo-random	0.9741	**0.9838**
Type III Pseudo-random	0.9736	**0.9830**
Fuji RGB	0.9637	**0.9764**
Average	0.974946	**0.983202**

and c_1 and c_2 are small constants. Under the embedding capacity with 0.7 bpp, Table 7 lists the mean SSIM indices of the original and marked CFA videos. When compared with Yang et al.'s method, the average SSIM improvement of the proposed method is 0.008256 (= 0.983202–0.974946) for the concerned eleven CFA patterns.

4.2. CPSNR and MOVIE quality comparison of reconstructed RGB full-color images and videos

For screen display, the marked CFA images and videos should be transformed to the reconstructed RGB full-color ones. Therefore, the CPSNR and MOVIE metrics are used to evaluate the quality of the reconstructed full-color image and video,

Table 8
The mean CPSNRs of Yang et al.'s method and the proposed method for the eleven CFA structures (embedding capacity = 0.7 bpp).

CFA Structure	Yang	Proposed
Bayer	37.7438	**38.3932**
Lukac and Plataniotis	35.8715	**36.6332**
Yamanaka	33.5685	**35.0750**
Diagonal stripe	35.8993	**36.9699**
Vertical stripe	29.5055	**29.5386**
Modified Bayer	36.6755	**37.5132**
HVS-based	34.8291	**35.4294**
Type I Pseudo-random	37.4683	**38.0836**
Type II Pseudo-random	34.5453	**35.1181**
Type III Pseudo-random	35.3593	**36.4185**
Fuji RGB	33.2689	**35.0023**
Average	34.9759	**35.8341**

Table 9
The mean MOVIE indexes of Yang et al.'s method and the proposed method for the eleven CFA structures (embedding capacity = 0.7 bpp).

CFA Structure	Yang	Proposed
Bayer	0.000084	**0.000050**
Lukac and Plataniotis	0.000142	**0.000069**
Yamanaka	0.000415	**0.000177**
Diagonal stripe	0.000208	**0.000115**
Vertical stripe	**0.000577**	0.000586
Modified Bayer	0.000129	**0.000083**
HVS-based	0.000218	**0.000148**
Type I Pseudo-random	0.000099	**0.000067**
Type II Pseudo-random	0.000273	**0.000199**
Type III Pseudo-random	0.000212	**0.000118**
Fuji RGB	0.000445	**0.000181**
Average	0.000255	**0.000163**

Table 10
The average PSNRs of eight Bayer CFA test videos among the concerned RDH methods.

Capacity (bpp)	Sachnev et al. [22]	Tai et al. [26]	Thodi and Rodriguez [27]	Chang et al. [6]	Yang et al. [33]	Proposed
0.1	56.7578	52.8378	53.6566	58.0388	55.3350	**58.2564**
0.2	51.5576	49.5941	50.3604	53.9380	51.7355	**54.6003**
0.3	49.8357	46.8960	47.4325	51.2114	49.3930	**52.3412**
0.4	47.5602	44.6994	45.2981	48.4898	47.2407	**50.5659**
0.5	46.0171	42.0940	43.8529	45.4099	45.1871	**48.9758**
0.6	43.9662	40.4649	42.2391	42.9044	43.1278	**47.2808**
0.7	41.2378	37.4706	40.4969	39.6645	41.0738	**45.1148**
Average	48.1333	44.8652	46.1909	48.6480	47.5847	**51.0193**

respectively. The CPSNR metric is calculated by

$$\mathrm{CPSNR} = \frac{1}{n}\sum_{t=1}^{n} 10\log_{10}\frac{255^2}{\mathrm{CMSE}(t)}, \tag{22}$$

where

$$\mathrm{CMSE}(t) = \frac{1}{3WH}\sum_{p=1}^{WH}\sum_{C\in\{\mathrm{R,G,B}\}}\left[V_C^t(p) - \tilde{V}_C^t(p)\right]^2, \tag{23}$$

where $V_C^t(p)$ and $\tilde{V}_C^t(p)$ denote pixel value of the component $C \in \{\mathrm{R, G, B}\}$ of the pth pixel in the tth image for the original RGB full-color video and the reconstructed RGB full-color video demosaicked by [34], respectively. Table 8 lists the mean CPSNRs for Yang et al.'s and the proposed methods under the embedding capacity with 0.7 bpp. It indicates that the proposed method has 0.8582 (35.8341–34.9759) dB CPSNR improvement.

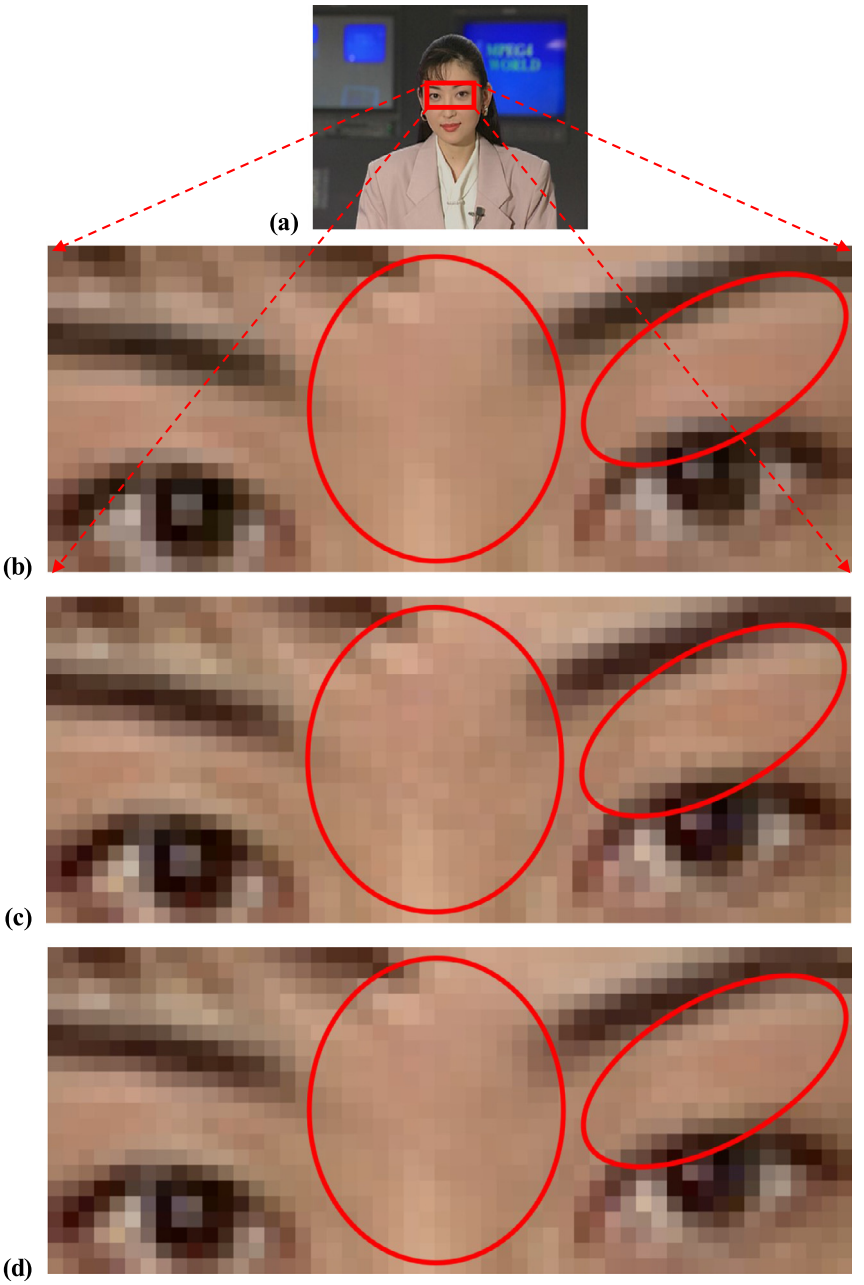

Fig. 9. The visual effect comparison for the reconstructed full-color marked images for "Akiyo" (Bayer CFA, embedding capacity 0.7 bpp): **(a)** the original image; **(b)** the magnified sub-image of the original image; **(c)** the magnified sub-image of the reconstructed full-color marked image by Yang et al.'s method; **(d)** the magnified sub-image of the reconstructed full-color marked image by the proposed method. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Finally, we adopt the MOVIE index metric [23] to evaluate the quality performance of the reconstructed full-color videos by the concerned RDH methods. The MOVIE index can be used to calculate the spatial and temporal distortion with respect to the reconstructed full-color video and the original full-color video. The reconstructed and original full-color videos are decomposed using spatio-temporal Gabor filters. The spatial distortion for expressing the image quality and the temporal distortion for expressing the motion quality are measured separately, and then are combined to obtain an overall assessment score as the MOVIE index. A lower MOVIE index means a better video quality of the reconstructed full-color video. In Table 9, under the embedding capacity with 0.7 bpp, the proposed RDH method does show the substantial video quality improvement of the reconstructed full-color videos, and the MOVIE index improvement ratio is 36% $(=(0.000255-0.000163)/0.000255)$.

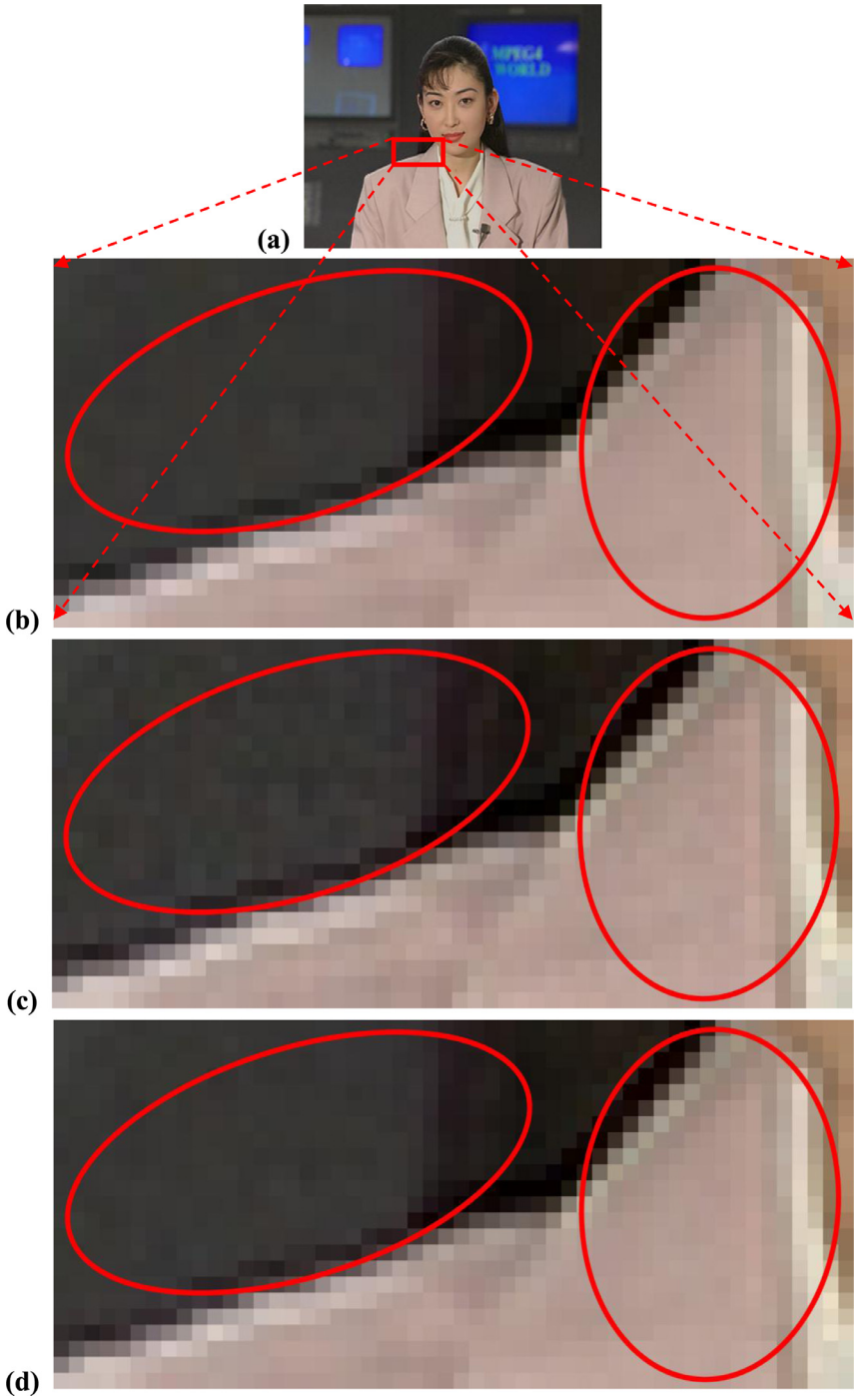

Fig. 10. The second visual effect comparison for the reconstructed full-color marked images for "Akiyo" (Bayer CFA, embedding capacity 0.7 bpp): **(a)** the original image; **(b)** the magnified sub-image of the original image; **(c)** the magnified sub-image of the reconstructed full-color marked image by Yang et al.'s method; **(d)** the magnified sub-image of the reconstructed full-color marked image by the proposed method. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

4.3. Quality assessment discussion on vertical stripe and bayer CFA patterns

From the above quality assessments, we find that the vertical stripe CFA usually obtains the best PSNR and SSIM performance for the marked CFA images while the Bayer CFA wins, in terms of CPSNR and MOVIE, for the reconstructed full-color images and videos, respectively. To investigate this phenomenon, we first examine the error prediction scheme for the two CFA structures and find that the vertical stripe CFA generally yields smaller prediction errors than the Bayer CFA in average. However, the vertical stripe CFA sometimes obtains very high prediction errors, for example, when it meets slanted lines in images. On the other hand, the Bayer CFA refers to more pixels at various locations in error prediction, so its prediction errors may be a little higher but less variant. The characteristics thus make the two CFA structures having distinguishing performances in data hiding and demosaicking. In data hiding, it embeds data in partial image pixels with small prediction errors (i.e., smaller than θ_{err}). On the contrary, demosaicking employs all image pixels despite their prediction errors. This is why the vertical stripe CFA behaves better in data hiding, and the Bayer CFA outperforms in demosaicking.

4.4. PSNR vs. embedding capacity comparison among the proposed RDH method and the related five methods for Bayer CFA structure

Besides Yang et al.'s RDH method [33], the other four RDH methods [6,22,26,27] are included to compare the proposed RDH method. Although the four RDH methods are originally designed to deal with gray images, we have made some modification so that they can tackle the Bayer CFA images. We partition each image of the Bayer CFA video into four color planes, the R (including R_0 and R_1) plane, B (including B_0 and B_1) plane, G_0 plane, and G_1 plane. The concerned comparative methods are applied to each color plane individually, and then the four marked color planes are combined to generate a whole marked mosaic image. Table 10 lists the average PSNRs of eight Bayer CFA test videos among the concerned RDH methods, showing the proposed method has the best average PSNR performance when compared with the other RDH methods.

4.5. Visual effect comparison of reconstructed full-color marked images

A subjective evaluation is also carried out to demonstrate the visual superiority of the proposed method. Fig. 9 gives the first visual example of the reconstructed full-color marked images, i.e. the demosaicked marked images, by Yang et al.'s method and the proposed method under the embedding capacity 0.7 bpp. Fig. 9(b) illustrates the magnified one of the sub-image highlighted by a red rectangle in Fig. 9(a), and the corresponding reconstructed full-color sub-image by Yang et al.'s method and the proposed method are shown in Fig. 9(c) and (d), respectively. The regions marked by red ellipses in Fig. 9(c) and (d) reveal that the proposed method displays fewer color artifacts than those of Yang et al.'s method. In other words, the proposed method can achieve better visual quality and its reconstructed full-color marked image is closer to the original image. For Fig. 10, the proposed method has the same visual benefits. The readers are suggested to refer to the related experiment results [28].

5. Conclusion

In this paper, we have presented the proposed novel RDH method combining spatial prediction and temporal prediction for mosaic videos with eleven CFA structures. Based on eight training CFA videos, a suitable threshold θ_{SAD} is determined and is used as a switch to intellectually select the temporal correlation-based prediction scheme or the spatial correlation-based prediction scheme in the proposed RDH method. Experimental results show that the proposed method has significant objective and subjective quality improvement on the marked CFA images and on the reconstructed full-color marked videos when compared with the spatial correlation-based method. In addition, based on eight test Bayer CFA videos, the proposed RDH method is clearly superior to five related RDH methods. One future research issue is to dynamically determine an optimal θ_{SAD} for each test mosaic video. Another future research issue is to extend the results of this paper to other CFA structures with four color channels, such as Sony RGBW CFA [25], Kodak CFA [8], and Hirakawa CFA [11]. The other research issue is to apply the feature selection mechanisms [15,16] to identify the smooth areas and the textural areas in the mosaic image to assist the RDH method for increasing the quality improvement.

Acknowledgment

The authors appreciate the proofreading help of Ms. C. Harrington.

References

[1] J.M. Barton, Method And Apparatus For Embedding Authentication Information Within Digital Data, U.S. Patent 5 646 997, 1997.
[2] B.E. Bayer, Color imaging array, U.S. Patent 3 971 065, 1976.
[3] C.C. Chang, C.C. Lin, C.S. Tseng, W.L. Tai, Reversible hiding in DCT-based compressed images, Inf. Sci. 177 (13) (2007) 2768–2786.
[4] C.C. Chang, T.D. Kieu, W.C. Wu, A lossless data embedding technique by joint neighboring coding, Pattern Recognit. 42 (7) (2009) 1597–1603.
[5] C.C. Chang, W.L. Tai, C.C. Lin, A reversible data hiding scheme based on side match vector quantization, IEEE Trans. Circuits Syst. Video Technol. 16 (10) (2006) 1301–1308.

[6] C.C. Chang, Y.H. Huang, T.C. Lu, A difference expansion based reversible information hiding scheme with high stego image visual quality, Multimed. Tools Appl. 76 (10) (2017) 12659–12681.
[7] P.C. Chang, K.L. Chung, J.J. Chen, C.H. Lin, T.J. Lin, A DCT/DST-based error propagation-free data hiding algorithm for HEVC intra-coded frames, J. Vis. Commun. Image Represent. 25 (2) (2014) 239–253.
[8] J. Compton, and J. Hamilton, Image Sensor With Improved Light Sensitivity, U.S. Patent 8 139 130 B2, 2012.
[9] L. Condat, A generic variational approach for demosaicking from an arbitrary color filter array, in: IEEE Int. Conf. on Image Processing, 2009, pp. 1625–1628.
[10] G. Gao, Y.Q. Shi, Reversible data hiding using controlled contrast enhancement and integer wavelet transform, IEEE Sig. Process. Lett. 22 (11) (2015) 2078–2082.
[11] K. Hirakawa, P.J. Wolfe, Spatio-spectral color filter array design for optimal image recovery, IEEE Trans. Image Process. 17 (10) (2008) 1876–1890.
[12] Y. Hu, H.K. Lee, J. Li, DE-based reversible data hiding with improved overflow location map, IEEE Trans. Circuits Syst. Video Technol. 19 (2) (2009) 250–260.
[13] F. Huang, X. Qu, H. Kim, J. Huang, Reversible data hiding in JPEG, IEEE Trans. Circuits and Systems for Video Technology (2015) in early access.
[14] X. Li, B. Li, B. Yang, T. Zeng, General framework to histogram-shifting-based reversible data hiding, IEEE Trans. Image Process. 22 (6) (2013) 2181–2191.
[15] Z. Li, J. Tang, Unsupervised feature selection via nonnegative spectral analysis and redundancy control, IEEE Trans. Image Process. 24 (12) (2015) 5343–5355.
[16] Z. Li, J. Liu, J. Tang, H. Lu, Robust structured subspace learning for data representation, IEEE Trans. Pattern Anal. Mach. Intell. 37 (10) (2015) 2085–2098.
[17] T.J. Lin, K.L. Chung, P.C. Chang, Y.H. Huang, H.Y.M. Liao, C.Y. Fang, An improved DCT-based perturbation scheme for high capacity data hiding in H.264/AVC intra frames, J. Syst. Softw. 86 (3) (2013) 604–614.
[18] R. Lukac, Single-sensor imaging: Methods and Applications For Digital Cameras, CRC Press/Taylor and Francis, Boca Raton, FL, 2008.
[19] R. Lukac, K.N. Plataniotis, Color filter arrays: design and performance analysis, IEEE Trans. Consum. Electron. 51 (4) (2005) 1260–1267.
[20] R. Lukac, K.N. Plataniotis, Universal demosaicking for imaging pipelines with an RGB color filter array, Pattern Recognit. 38 (11) (2005) 2208–2212.
[21] Z. Ni, Y.Q. Shi, N. Ansari, W. Su, Reversible data hiding, IEEE Trans. Circuits Syst. Video Technol. 16 (3) (2006) 354–362.
[22] V. Sachnev, H.J. Kim, J. Nam, S. Suresh, Y.Q. Shi, Reversible watermarking algorithm using sorting and prediction, IEEE Trans. Circuits Syst. Video Technol. 19 (7) (2009) 989–999.
[23] K. Seshadrinathan, A.C. Bovik, Motion tuned spatio-temporal quality assessment of natural videos, IEEE Trans. Image Process. 19 (2) (2010) 335–350.
[24] Y.Q. Shi, X. Li, X. Zhang, H.T. Wu, B. Ma, Reversible data hiding: advances in the past two decades, IEEE Access 4 (2016) 3210–3237.
[25] T. Sugiyama, Image-Capturing Apparatus, U.S. Patent 0 231 618, 2005.
[26] W.L. Tai, C.M. Yeh, C.C. Chang, Reversible data hiding based on histogram modification of pixel differences, IEEE Trans. Circuits Syst. Video Technol. 19 (6) (2009) 906–910.
[27] D.M. Thodi, J.J. Rodriguez, Expansion embedding techniques for reversible watermarking, IEEE Trans. Image Process. 16 (3) (2007) 721–730.
[28] [online] Available: http://faculty.csie.ntust.edu.tw/~klchung/TY/Perceptual_Effect.rar.
[29] X. Wang, C.Y. Chao, X. Xu, X. Niu, Reversible data-hiding scheme for 2-D vector maps based on difference expansion, IEEE Trans. Inf. Forensic. Secur. 2 (3) (2007) 311–320.
[30] Z. Wang, A.C. Bovik, H.R. Sheikh, E.P. Simoncelli, Image quality assessment: from error visibility to structural similarity, IEEE Trans. Image Process. 13 (4) (2004) 600–612.
[31] X. Wu, L. Zhang, Temporal color video demosaicking via motion estimation and data fusion, IEEE Trans. Circuit Syst. Video Technol. 16 (2) (2006) 231–240.
[32] W.J. Yang, K.L. Chung, H.Y.M. Liao, Efficient reversible data hiding for color filter array images, Inf. Sci. 190 (2012) 208–226.
[33] W.J. Yang, K.L. Chung, L.C. Lin, C.H. Lin, Universal reversible data hiding for arbitrary RGB CFA captured images, in: Proceedings of IEEE International Conference on Information, Communications and Signal Processing (ICICS), Tainan, Taiwan, 2013, pp. 1–5.
[34] W.J. Yang, K.L. Chung, W.N. Yang, 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. 23 (4) (2013) 591–606.
[35] W.J. Yang, K.L. Chung, H.Y.M. Liao, W.K. Yu, Efficient reversible data hiding algorithm based on gradient-based edge direction prediction, J. Syst. Softw. 86 (2) (2013) 567–580.
[36] [online] Available: http://trace.eas.asu.edu/yuv/.
[37] L. Zhang, X. Wu, A. Buades, X. Li, Color demosaicking by local directional interpolation and nonlocal adaptive thresholding, J. Electron. Imaging 20 (2) (2011) 023016.
[38] X. Zhang, Reversible data hiding in encrypted image, IEEE Sig. Process. Lett. 18 (4) (2011) 255–258.
[39] X. Zhang, J. Wang, Z. Wang, H. Cheng, Lossless and reversible data hiding in encrypted images with public key cryptography, IEEE Trans. Circuits Syst. Video Technol. (2016) in early access.
[40] X. Zhang, W. Zhang, Semantic image compression based on data hiding, IET Image Process. 9 (1) (2015) 54–61.