WILEY | Hindawi

Research Article

Performance Analysis of Display Field Communication with Advanced Receivers

Pankaj Singh [ID],[1] Byung Wook Kim [ID],[2] and Sung-Yoon Jung [ID][1]

[1]*Department of Electronic Engineering, Yeungnam University, Gyeongsan, Gyeongbuk 38541, Republic of Korea*
[2]*Department of Information and Communication Engineering, Changwon National University, Changwon, Gyeongsangnam 51140, Republic of Korea*

Correspondence should be addressed to Sung-Yoon Jung; syjung@ynu.ac.kr

Display field communication (DFC) is an imperceptible display-to-camera (D2C) communication approach that provides dual-mode, full-frame, visible light communication capabilities. Unlike conventional screen-to-camera communication approaches, DFC embeds data imperceptibly in the spectral domain of individual video frames. This paper analyzes the practical performance of the DFC scheme with advanced receivers, including zero forcing (ZF), minimum mean square error (MMSE), and maximum likelihood (ML). A 249 × 262 color image is used for embedding data consisting of eight individual information vectors with their elements 2-QAM and 4-QAM modulated. The color image is separated into three individual channels, i.e., red (R), green (G), and blue (B). A lossy display-camera channel is considered in the presence of Gaussian noise, blooming, and various geometric distortions. Simulation results show that the ML receiver outperforms MMSE and ZF receivers. In addition, independent RGB data channels are evaluated to compare the symbol error rate of each channel. The proposed color DFC algorithm can be a viable candidate for practical scenarios in applications like smart content transmission and for supporting robust communication performance with advanced receivers, while the data embedded in the images remain unobtrusive to the human eye.

1. Introduction

Optical camera communication (OCC) [1–5] has been rapidly emerging as a compelling technology for wireless communications, generally consisting of light-emitting diodes (LED) as a transmitter and a camera as a receiver. The widespread use of multiple cameras in handheld devices, and the explosion of related smart content-oriented display technology, such as digital signage [6], gives OCC systems extensive applications in information technology. Smart content data can be defined as content that is intelligibly personalized to a specific person, delivering a targeted message for a personal experience. The smart content transmission industry is expected to be approximately $32 billion in the next five years, having applications in the fields of education, hospitality, government, corporate environment, cinema, advertising, and many more [7]. Moreover, smartphone or tablet cameras are used, not only for capturing images but also for communicating information. For example, the pervasive quick response (QR) codes communicate a short code to smartphone cameras, and recent research has explored using screen-to-camera communications for large data transfers. In such a scenario, display-to-camera (D2C) communications will play an important role when the smart content data becomes more personalized, using concepts from artificial intelligence, big data, and augmented reality.

The most popular D2C communication is QR code where information is encoded into a two-dimensional (2D) barcode [8–13]. In other words, QR codes are well-known digital advertising methods that send data to a camera from digital marks in print media. However, QR codes are typically limited by their size and location. This limits the amount of information that can be encoded. On the other hand, as the demand for communications through multimedia services increases, interactive applications in the ubiquitous computing environments require a large amount of

data to be transmitted. In the D2C communications environment, an image on an electronic display (e.g., TVs, monitors, billboards, and projector screens) is a transmitter, and the camera is a receiver [14–17]. In other words, a camera is used as both an image sensor to capture an image and a communication receiver to obtain the information transmitted from the display pixels. In addition, because an ordinary display can embed large amounts of data, it is possible for users to provide a full-frame display while simultaneously transmitting data.

As mentioned above, display- (or screen-) to-camera communications is a technology where an LCD screen and a camera sensor can communicate via device-to-device communications [1, 18, 19]. First inspired by the traditional RF modulation scheme, PIXNET [20] proposed encoding information in 2D spatial frequencies of an image. PIXNET includes a perspective corrective algorithm, blur-adaptive orthogonal frequency-division multiplexing (OFDM) coding, and an ambient light filter. Another approach to screen-camera communications was proposed in color barcode streaming for smartphones (COBRA) [11]. COBRA was designed to achieve one-way communications between small-sized screens and the low-speed cameras in smartphones using 2D color barcodes. The limited available throughput in screen-to-camera links was further enhanced by LightSync [12], which improves frame synchronization between transmitter and receiver and can nearly double the achievable throughput. The creators of HiLight [21] introduced a new scheme for screen-camera communications without any coded images. Leveraging the properties of the orthogonal transparency (alpha) channel, HiLight "hides" the bits by changing the pixel translucence instead of modifying the red-green-blue (RGB) color. Another mechanism for high-rate, flicker-free screen-camera communications was proposed, which is similar to QR coding techniques, but the data are embedded in the image spatially using a content-adaptive method [22]. The proposed scheme considers some blocks of the pixels in an image, embedding data as a texture in each pixel block (but not in the edges of the texture, because the human eye is very sensitive to changes in the texture of an image). Two methods are applied to embed data spatially in the texture of an image. The first is texton analysis, which is a machine learning technique to detect the desired texture in an image. The other is pixel-based texture analysis, which employs detection of the so-called *good region* (in which changes remain imperceptible to the human eye) in an image to embed data.

These kinds of hidden display-camera communications techniques have emerged as a new paradigm that embeds data imperceptibly into regular videos while remaining unobtrusive to human viewers. Various other studies have also been conducted on how to embed data that can remain unobtrusive to the human eye in a displayed image [15–17, 22–27]. In [15], a screen-to-camera communications system was proposed where data embedding is done on the relative brightness of an image block. This was done by increasing (encode bit 1) or decreasing (encode bit 0) the brightness of the block of pixels. Moreover, the messages are embedded into selected video frames using watermarking that is not

perceptible to the human eye and is subsequently played at a high frame rate. However, the details of system implementation and performance evaluation were not presented. Yuan et al. [16] presented a watermarking technique for embedding data into an image frame based on the Laplacian pyramid method. However, fundamental analysis of the error rate performance was not presented. Wang et al. [17] proposed a method called InFrame that uses a *complementary frame* concept and embeds a data frame into a pair of multiplexed video frames. Jo et al. [23] presented DisCo, which enables displays and cameras to communicate with each other while also displaying and capturing images for human consumption. Messages were transmitted by temporally modulating the display brightness at high frequencies so they are imperceptible. Messages were received by a rolling shutter camera that converts the temporally modulated incident light into a spatial flicker pattern. Zhang et al. [25] presented ChromaCode, which introduced a uniform color space for unobtrusive data embedding. The bits are embedded into pixels using the most accurate color difference formula, CIEDE2000, in a perceptually uniform color space, CIELAB. The authors also proposed a novel adaptive embedding scheme in an outcome-based philosophy, which accounts for both pixel lightness and frame texture and ensures flicker invisibility over the full frame. Overall, most of the above methods for D2C communications embed data into the spatial domain of an image (or video), which can directly affect image perception by the human eye. On the other hand, in the display field communication (DFC) scheme [28, 29], the data are embedded in the spectral (or frequency) domain of an image, and they still provide dual-mode, full-frame, visible light communication functionalities.

DFC embeds data in and extracts from the spectral domain so that the properties associated with the frequency coefficients of an image can be employed. In particular, the data are embedded in the designated spectral subbands (SB) of an image. Kim et al. [28] was the first work on DFC where the practical performance of the DFC scheme was evaluated in the presence of additive white Gaussian noise (AWGN) and various geometric distortions. The data are embedded in a 256×256 grayscale image with 16-QAM modulation to achieve a maximum data rate of 9.5 Kbps. However, the scheme considered data embedding in only one dimension (the width) of the frequency-domain image, resulting in 1D-DFC. The same authors extended the concept of 1D-DFC to 2D-DFC [29] where the data are embedded in two-dimensions (width and height) of a grayscale image. It has been shown that 2D-DFC achieves a higher data rate than 1D-DFC, and hence is more appropriate for practical uses in the D2C environment. However, it also uses a grayscale image as input on the transmitter screen. The work by Kim and Jung [30] proposed color DFC, where the authors used three different-colored RGB images as input on the screen. Each independent RGB data channel was evaluated, showing similar performance for all three input images. In other words, it was shown that, despite the different characteristics of the input images, similar output results are observed in both grayscale and color images, indicating that the type and channels
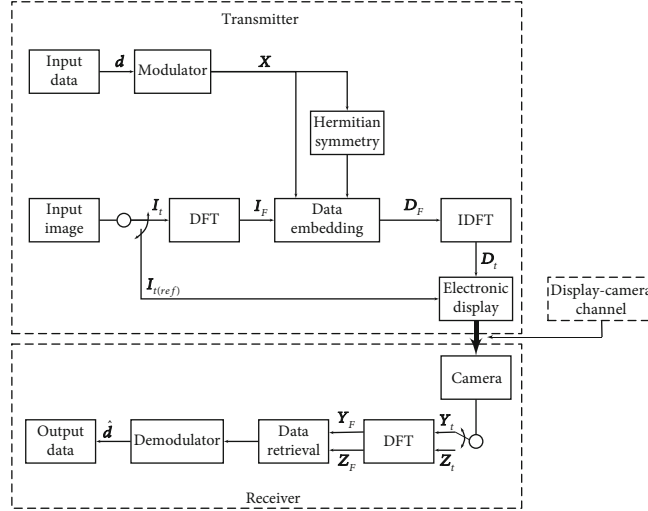
FIGURE 1: System architecture for DFC.

of the input image do not really affect the symbol error rate (SER) performance in screen communications [28, 30].

In this paper, we evaluate the concept of DFC for the color image (as input on the transmitter screen) in order to apply it to a more practical D2C environment. In addition, we mathematically evaluate and analyze the performance of the proposed method for different decoding schemes. Three advanced receivers, including zero forcing (ZF), minimum mean square error (MMSE), and maximum likelihood (ML), are evaluated. Furthermore, the SER in a display-camera channel with various distortions and AWGN noise is evaluated for all the receivers and for various other system parameters, such as modulation order, subbands, and different RGB channels. The rest of this paper unfolds as follows. Section 2 describes the system model and the data embedding process in the proposed color DFC scheme. Section 3 describes the display-camera channel and captures various other channel distortions, such as blooming, that can occur during display-camera communications. Section 4 explains and analyzes the decoding process for all the receivers. Performance of the scheme is assessed and compared in Section 5, and the study concludes in Section 6.

2. Color DFC Scheme

A DFC system is composed of a digital camera pointed at an electronic screen (cf., Figure 1). On the transmitter side, the input spatial-domain images are first converted to the frequency domain by applying discrete Fourier transform (DFT). Because the input image is in color, it should first be separated into individual RGB channels. Then, each channel is converted separately to the frequency domain. At the same time, the modulator modulates the binary input data by mapping bits to binary symbols. Both the data and their Hermitian symmetric equivalents are used in the data embedding process to conserve the spatial property of the transmitted image [28]. Each data channel image is then converted back to the spatial domain using inverse DFT, and all the channels are combined to show (or transmit) the final

image on the screen. On the display device, the data-embedded image and the *reference image* are rendered alternately to minimize the image artifacts that may be visible to the human eye [28].

At the receiver, the frames are received sequentially by the camera, and the images are classified as data-embedded and reference images. The combined spatial-domain image is then separated into its three channels and converted to the frequency domain. Finally, the data are decoded using the various advanced receivers.

2.1. Data Embedding. In the RGB image, the data can be embedded in each of the individual color channels. Therefore, the data rate is tripled in color DFC, compared to gray-scale DFC [28]. The frequency-domain image for a particular channel can be calculated by taking the column-wise one-dimensional discrete Fourier transform (1D-DFT) of the image:

$$\boldsymbol{I}_{F,j} = \left[\boldsymbol{F} \cdot \boldsymbol{i}_{t_1,j}, \boldsymbol{F} \cdot \boldsymbol{i}_{t_2,j}, \dots, \boldsymbol{F} \cdot \boldsymbol{i}_{t_Q,j}\right] = \boldsymbol{F} \cdot \boldsymbol{I}_{t,j}, \quad (1)$$

where $\boldsymbol{F}$ is a $P \times P$ DFT matrix, $\boldsymbol{I}_{t,j}$ is a $P \times Q$ spatial-domain image, $\boldsymbol{i}_{t_q,j}$ is the column vector of $\boldsymbol{I}_{t,j}$ where $q = 1, 2, \cdots, Q$, and j is a particular R, G, or B channel. In the frequency-domain image, each point represents a particular frequency contained in the spatial-domain image. The result of 1D-DFT has low-frequency components on both sides of the frequency-domain image, whereas the high-frequency components lie symmetrically in the central region.

Regarding data embedding, the data are first modulated using quadrature amplitude modulation (QAM); 2-QAM and 4-QAM were used in this study, as shown in the constellation diagram of Figure 2. By using further high-order modulation, i.e., more points on the constellation, although it is possible to transmit more bits per symbol; the points become closer together and are susceptible to noise. Therefore, we will stick to only 2-QAM and 4-QAM in this study. After that, the modulated data and their Hermitian symmetric
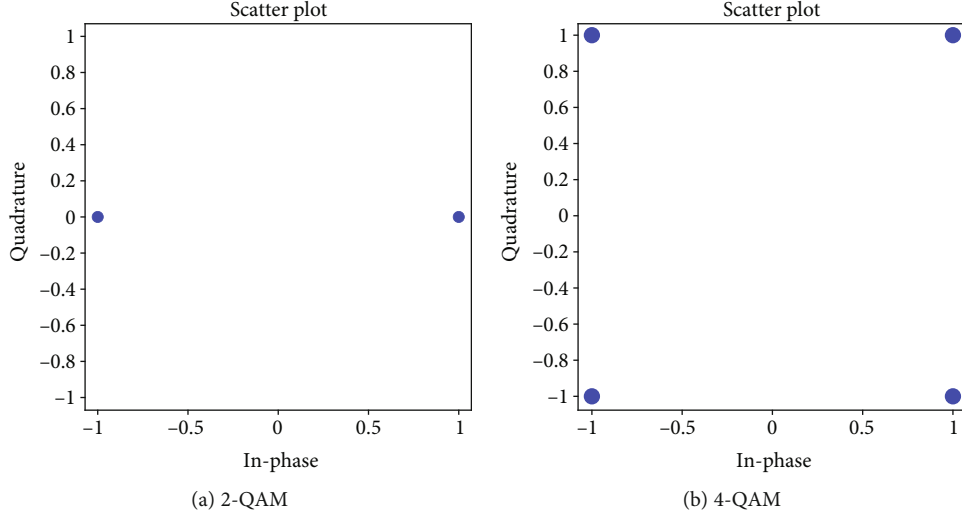
(a) 2-QAM

(b) 4-QAM

Figure 2: Modulated signal constellation.

equivalents are embedded in the frequency-domain image. The Hermitian symmetric equivalents of the data are embedded because, since the spatial-domain image has pixel values of real and positive numbers, the elements of 1D-DFT output will exhibit columnwise conjugate symmetry. Therefore, to embed data in the frequency domain and sustain the real-valued and positive properties for a data-embedded image simultaneously, the data sequence should also have conjugate symmetric properties [28, 29]. Hence, the data matrix, X, can be represented as

$$X_j = \left[X_{1,j}, X_{2,j}, \ldots, X_{Q,j} \right], \qquad (2)$$

with

$$X_{q,j} = \left[\underbrace{\mathbf{1}}_{1\times(s-1)} \left(d_{q,j} \right)^T \underbrace{\mathbf{1}}_{1\times c} \left(\mathrm{flip}\left(d_{q,j}^* \right) \right)^T \underbrace{\mathbf{1}}_{1\times(s-2)} \right]^T, \quad (3)$$

where $d_{q,j} = [d_{q,j}(1), \ldots, d_{q,j}(L)]^T$ is the data vector on the q th column of the data matrix, $j \in \{R, G, B\}$, s is the starting pixel of the data symbol, and "flip" is an operation as defined in [28]. The starting pixel of the data symbol is given as

$$c = P - 2s - 2L + 3, \qquad (4)$$

satisfying

$$s + L < \left\lfloor \frac{P}{2} \right\rfloor + 2, \qquad (5)$$

where L is the number of data symbols per column. In this way, the data structure covers an $L \times Q$ rectangular region, and another $L \times Q$ conjugate symmetric region on the frequency-domain image shown as white bands in Figure 3. These white regions in the frequency-domain image represent the position of the frequency subbands.

The data embedding process is then carried out in the frequency domain using the multiplicative coefficients on the pixel values of an image. A data-embedded image, $D_{F,j}$, in the frequency domain is given as

$$D_{F,j} = I_{F,j} \circ X_j, \qquad (6)$$

where $\circ$ is the Hadamard product operator. The above data-embedded frequency-domain image is then converted to the spatial domain by taking the inverse DFT. Therefore, the data-embedded image in the spatial domain is represented as

$$D_{t,j} = F^H \cdot D_{F,j}. \qquad (7)$$

To be displayed on the screen, each channel of the color image has to be combined as follows:

$$D_t = \sum_j D_{t,j}, \qquad (8)$$

where $j \in \{R, G, B\}$. The above data-embedded image is then displayed on the electronic screen that is captured by the camera. Note that the data-embedded image is placed between the neighboring reference images in the sequence of image frames. This will achieve two important purposes. First, the reference image can be used for the decoding of the embedded data in the camera receiver. Second, it will help an electronic display perform its original purpose, i.e., by rendering images at a high frame rate, the artifacts visible to the human eye can be greatly minimized.

As mentioned above, the frequency-domain image has low-frequency components on both sides of the image, whereas the high-frequency components lie symmetrically in the central region. For data embedding in the frequency domain, we choose several frequency subbands, i.e., the coefficients of several groups of frequency bins. In Figure 3, we can see that subband 3 (SB3) loads data in relatively lower

(a) Data embedding in the high-frequency band (subband 1)

(b) Data embedding in the mid-frequency band (subband 2)

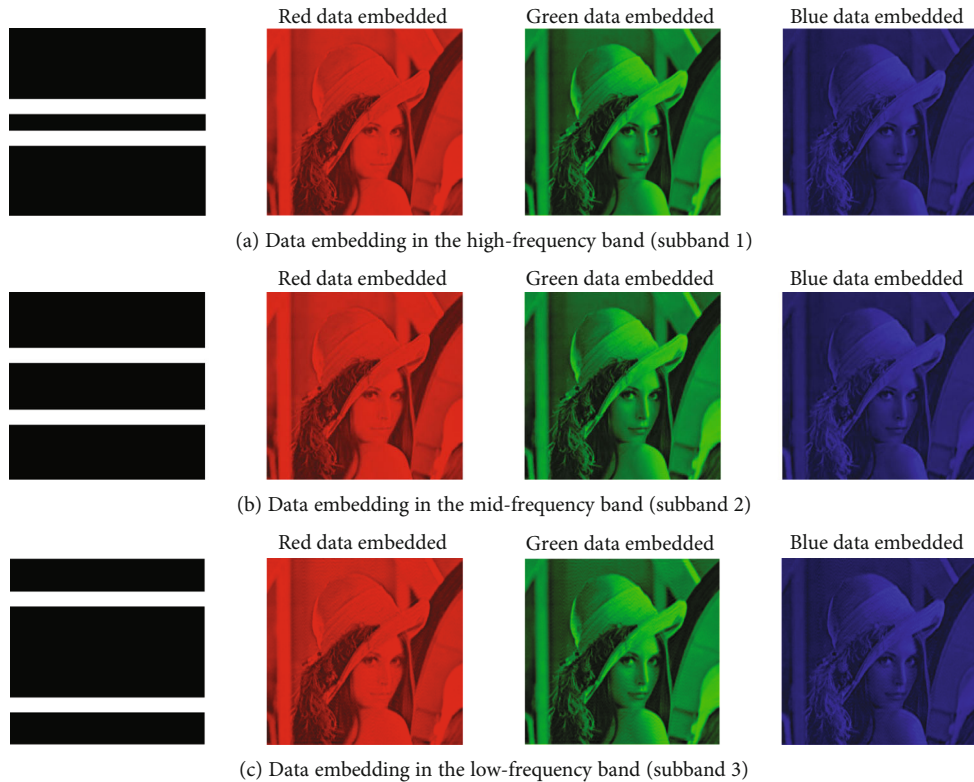(c) Data embedding in the low-frequency band (subband 3)

FIGURE 3: Data-embedded images in the frequency domain with their corresponding RGB components.

frequency bands than subband 2 (SB2), and so on. The corresponding data-embedded images for individual RGB channels are also shown. These individual RGB data-embedded images are combined to make final spatial-domain image to be transmitted on the screen. Figure 4 represents the corresponding combined spatial-domain images. The effect of loading data in various frequency bands can be observed in their spatial-domain counterparts. We can see that fewer detectable artifacts are introduced in SB1 and SB2, while the artifacts become high and easily visible in SB3. In particular, we can observe fewer visible artifacts in Figure 4(c) and clear visible artifacts in Figure 4(d). Therefore, because the low-frequency subbands contain the primary parts perceived in image content by the human eye, mid- or high-frequency subbands are preferred for embedding data.

Note that in the current DFC scenario, a smaller number of data symbols per column (L) are chosen, which results in a small vertical region for data embedding. However, L is a variable, and the value can be increased to embed more data. This may lead to an increased data rate, but a more distorted image at the same time. Recall that the target of the proposed DFC scheme is embedding and extracting data from the spectral domain of an image while simultaneously letting the electronic display perform its original purpose. Hence, the proposed method uses the coefficients in a certain range of frequencies to embed the data in the Fourier domain image in such a way that any artifacts introduced in the corresponding spatial image would be invisible. Moreover, by using reference images and rendering images at a high frame rate, any artifacts still visible to the human eye can be highly minimized.

3. Display-Camera Channel

3.1. Path Loss.
In a DFC system, the pixels of the display are the transmitter, and the camera capturing both the display screen and the background is the receiver. We assume that transmitter and receiver locations are fixed and that the channel characteristics are stationary in time. This assumption is realistic for situations in which the channel varies slowly and can be tracked. Furthermore, it is assumed that the optical axes of the transmitter and the receiver are aligned. When perfect alignment between the data-transmitting screen and the camera are considered, all the light-emitting pixels from the screen are the focus of the camera. In many cases, this assumption is nearly true, or it can be corrected by spatial predistortion techniques. A commercial video-processing equipment exists that removes projective distortion in the case of *off-axis* projection. Display-camera communications involves nonequivalent attenuation between the brightness of different pixels on the transmitter screen through the camera lens onto the image plane. Ideally, for any given aperture of the camera, the attenuation would be constant for every part of the image. However, unavoidable matters of geometric optics result in image illuminance declining as we move outward from the center of the frame. This phenomenon can be approximated by the "cosine fourth" law [31], which can be summarized as follows:
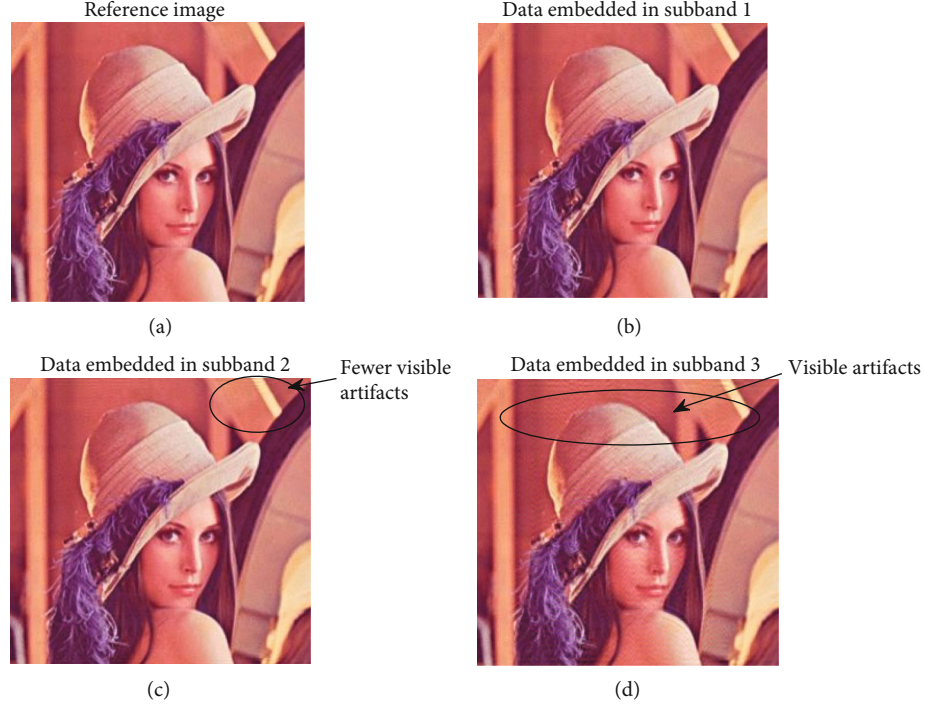
$$\frac{E_\Theta}{E_0} = \cos^4\Theta, \tag{9}$$

Reference image

Data embedded in subband 1

(a)

(b)

Data embedded in subband 2   Fewer visible
                                    artifacts

Data embedded in subband 3   Visible artifacts

(c)

(d)

FIGURE 4: Spatial-domain images with data embedded in different frequency subbands.

where E_Θ and E_0 are the data signal energy on the pixel at the point *off-axis* and point *on-axis*, respectively, and Θ is the angle at which transmitted pixels are off-axis. Consequently, the received pixel intensity can be calculated as

$$\tilde{\boldsymbol{Z}}_t[i] = \boldsymbol{I}_{t(\text{ref})}[i] \cos^4\theta_{t,i},$$
$$\tilde{\boldsymbol{Y}}_t[i] = \boldsymbol{D}_t[i] \cos^4\theta_{t,i},$$
(10)

where i is the index for a given pixel and θ_i is the corresponding off-angle for the respective pixel.

3.2. Blooming. Another kind of distortion that can affect display-camera communications is *blooming*. It occurs due to the charge leakage between neighboring pixels in a charge-coupled device (CCD) sensor. In particular, the charge capacity of a CCD pixel is limited, and when a pixel is full, the charge starts to leak into adjacent pixels. This process is known as blooming. The bloom effect makes the received image look brighter and have a hazy look. In other words, blooming results in an image in which bright light appears to bleed beyond its natural borders. Although under normal circumstances, this imperfection is not noticeable, an intensely bright light could cause the imperfections to be visible. Blooming can be approximated by the blurring effect followed by brightening the blurred image with reduced contrast. The spatial response of an imaging system is described by its point-spread function (PSF) [32]. The blur effect due to imperfect focus has a 2D Gaussian distribution in the spatial domain [32]. Let $\boldsymbol{h}$ denote the PSF characterizing the linear and spatially invariant response of the imaging system. In this

case, the blurred pixels can be modeled as a convolution of received pixels and the PSF [19]:

$$\ddot{\boldsymbol{Z}}_t[i] = \tilde{\boldsymbol{Z}}_t[i] \otimes h_t[i],$$
$$\ddot{\boldsymbol{Y}}_t[i] = \tilde{\boldsymbol{Y}}_t[i] \otimes h_t[i],$$
(11)

where $\otimes$ represents 2D linear convolution and $h_t[i]$ is the attenuation in the ith pixel of the corresponding images, captured by the PSF of the imaging system.

The effect of blooming distortion on an image can be recovered by using a Wiener filter [33]. In particular, the Wiener filter algorithm deconvolves the PSF from the original received image, finally returning the deblurred image. Note that in addition to blurring, there is the presence of noise in an image. In the absence of noise, a Wiener filter is equivalent to an ideal inverse filter. Figure 5 shows the effect of blooming on the transmitted image and the corresponding recovered image. As shown in the figure, blooming causes the received light energy to spread to areas outside the pixel. The amount of spread depends on the type of lens used in the camera. Specifically, blooming can be understood as a low-pass filtering phenomenon that distorts the high-frequency components in the image. Note that blooming occurs only in smartphones having a CCD image sensor. Despite the fact that nearly all smartphones these days use complementary metal-oxide semiconductor (CMOS) image sensors, a few camera phones still have a CCD sensor.

3.3. Noise. The signal quality in the camera receiver is also influenced by noise in the channel. Noise in camera systems manifests as *noise current* on each camera pixel and is

(a) Bloom image　　　　　　(b) Restored image

Figure 5: An image with bloom distortion and the corresponding restored image.

generated due to the photons from environmental lighting. At the output of the camera, the noise current in each camera pixel is a quantized quantity and manifests as fluctuations in the intensity of that pixel. The noise energy accumulated in each pixel can be quantified using the mean value of the variances in pixel intensity. In this paper, we consider noise in a camera pixel to be primarily from the background and that it follows an AWGN characteristic [34, 35] and it is uniform over the image sensor, quantified through AWGN noise variance. Therefore, the captured image can then be represented as

$$Z_t = \ddot{Z}_t + N_{t(\mathrm{ref})},$$
$$Y_t = \ddot{Y}_t + N_t, \qquad (12)$$

where Z_t is the received reference image, Y_t is the received data-embedded image, and N_t is the AWG noise matrix.

A camera is generally composed of an imaging lens, an image sensor (IS), and other image signal processing components. In order to obtain color information, a red, green, or blue filter normally covers the IS in a repeating pattern. This pattern (or sequence) of filters can vary, but one could use a widely adopted Bayer color filter array, which is a repeating 2×2 pattern, for digital acquisition of color images [36]. Therefore, we are able to separate the RGB signal with a color camera. Consequently, the combined image is separated into its individual channels as

$$Z_{t,j} = \ddot{Z}_{t,j} + N_{t(\mathrm{ref}),j},$$
$$Y_{t,j} = \ddot{Y}_{t,j} + N_{t,j}, \qquad (13)$$

where $j \in \{R, G, B\}$. This will help decode each channel's data separately. In addition, because the data are embedded using a multiplicative property on the frequency-domain image, individual images are then transformed into their frequency domain counterparts as follows:

$$Z_{F,j} = F \cdot Z_{t,j},$$
$$Y_{F,j} = F \cdot Y_{t,j}. \qquad (14)$$

3.4. Geometrical Distortion. Because of the nature of the camera imaging mechanism, the electronic screen may not be frontally aligned with the camera. This gives rise to geometric distortion as the screen pixels are captured at a perspective that results in shape distortion. Therefore, to observe practical performance of the proposed color DFC, geometric distortion based on the various vision transformation parameters can be considered. The perspective distortions in the DFC channel can be modeled as a composite effect of video quality reduction due to perspective scaling, rotating, and twisting of the pixel areas from the camera projection. The projection matrix [37] can be expressed as

$$\mathscr{P} = \begin{bmatrix} \varepsilon \cos \phi & -\varepsilon \sin \phi & t \\ \varepsilon \sin \phi & \varepsilon \cos \phi & t \\ 0 & 0 & 1 \end{bmatrix}, \qquad (15)$$

where the scalar, ε, represents scaling factor, ϕ is the rotating angle, and t is the twisting factor. The scale operator performs a geometric transformation which shrinks or zooms the size of an image.

Note that before data retrieval, the boundaries of the electronic display should be accurately detected so that a predefined frequency range for data embedding can be identified. For that, it is assumed that the electronic display is planar. To recognize the borders of the display for precise alignment, Harris corner extraction and the Hough transform, which are widely used geometric correction methods, can be exploited [38]. To resize the distorted image to its original size, one can obtain the missing pixels by interpolation. Because the data are hidden in the intensity values of the spatial-domain image, a high degree of spatial resolution results in high accuracy in data detection. Figure 6 shows the perspective screen alignment for precise data detection when rotating, scaling, and twisting distortions are considered. Furthermore, note that the camera should capture the entire image area because the embedded data are spread over the entire spatial-domain image. Therefore, a large standoff distance, which makes a camera acquire the whole image area, is required. The data are then decoded differently using the three advanced receivers described in the next section.

(a) Distorted image   (b) Reconstructed image

FIGURE 6: Reconstructed image according to the vision parameters of the perspective distortion.

4. Data Decoding

To retrieve data at the receiver, the ability to distinguish embedded data from the original data is required. For this reason, a reference image is inserted between data-embedded frames in the image frame sequence.

4.1. Zero Forcing. In the zero-forcing receiver, we have to find the inverse of the channel matrix. Considering the reference frames' pixel values are equivalent to the channel coefficients, the data can be decoded as

$$\widehat{d}_{q,j,\mathrm{ZF}}(l) = \frac{Y_{F,j}(s+l-1,q)}{Z_{F,j}(s+l-1,q)}, \qquad (16)$$

where $l = 1, 2, \cdots, L$, and $\widehat{\boldsymbol{d}}_{q,j} = [\widehat{d}_{q,j}(1), \dots, \widehat{d}_{q,j}(L)]^{T}$ is the estimated data column vector for the qth column and jth channel of an image. It is clear from the above equation that a ZF receiver does not consider the noise effect. On the other hand, the noise may get enhanced during the decoding process.

4.2. Minimum Mean Square Error. The MMSE receiver tries to minimize the mean square error between the transmitted symbols and the detected symbols, and thus maximizes the signal-to-noise ratio (SNR). Let $\mathscr{W}_{\mathrm{MMSE}}$ denote the MMSE detector. The estimated data column vector, $\widehat{\boldsymbol{d}}_{q,j}$, is then computed as

$$\widehat{d}_{q,j,\mathrm{MMSE}}(l) = \mathscr{W}_{q,j,\mathrm{MMSE}}(l) Y_{F,j}(s+l-1,q), \qquad (17)$$

where

$$\mathscr{W}_{q,j,\mathrm{MMSE}}(l) = \frac{E\left\{Y_{F,j}(s+l-1,q)\widehat{d}_{q,j}(l)\right\}}{E\left\{Y_{F,j}^2(s+l-1,q)\right\}}$$
$$= \frac{Z_{F,j}(s+l-1,q)}{Z_{F,j}^2(s+l-1,q)+\sigma_N^2(s+l-1,q)}. \qquad (18)$$

The final estimated symbol can then be expressed as

$$\widehat{d}_{q,j,\mathrm{MMSE}}(l) = \frac{Z_{F,j}(s+l-1,q)Y_{F,j}(s+l-1,q)}{Z_{F,j}^2(s+l-1,q)+\sigma_N^2(s+l-1,q)}, \qquad (19)$$

where σ_N^2 is the noise power in the received frame. We can see from the above equation that the MMSE receiver attempts to reduce the noise at the receiver based on SNR. Therefore, at a high SNR, MMSE behaves as a ZF receiver.

4.3. Maximum Likelihood. Maximum likelihood finds the minimum distance between the received image frames and the product of all possible transmitted image frames. Let S and M denote the set of transmitted image signal constellation symbol points and the modulation order, respectively. Then, ML detection determines the estimated transmitted data vector as

$$\widehat{d}_{q,j,\mathrm{ML}}(l) = \min_{\boldsymbol{d}_{q,j}\in S^M}\left|Y_{F,j}(s+l-1,q) - Z_{F,j}(s+l-1,q)\cdot d_{q,j}(l)\right|^2, \qquad (20)$$

where $d_{q,j}(l)$ is the transmitted symbol, Y_F is the received data-embedded image, and Z_F is the received reference image. Note that the ML receiver achieves the optimal performance when the transmitted symbol generation probability is equal. This is because it minimizes the error probability at the receiver by comparing the received signal vector with all the possible combinations of transmitted signal vectors to estimate the final symbol.

5. Simulation Results

This section presents the simulation results for the proposed color 1D-DFC scheme from comparing three advanced receivers and evaluating various system parameters, such as RGB channels, subbands, and modulation order. The receiving camera was assumed to be in front of the electronic screen to ensure perfect line-of-sight communication. The color image at 249×262 pixels (cf., Figure 4) was used as the input image, i.e., the image on the electronic display. For the performance evaluation, two different modulation techniques (2-QAM and 4-QAM) were exploited. Frame

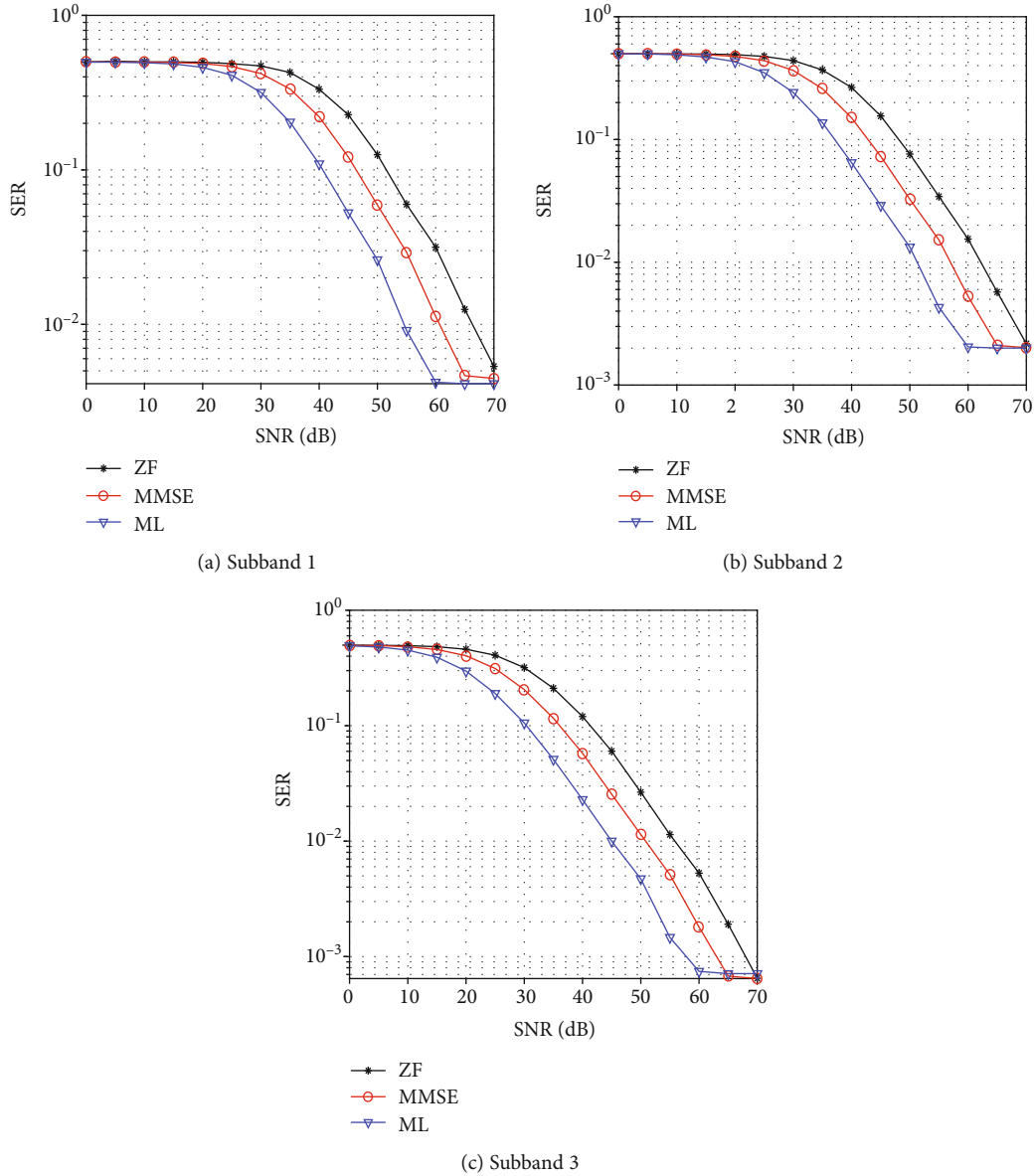(a) Subband 1

(b) Subband 2

(c) Subband 3

FIGURE 7: SER of color DFC for different receivers over the green channel.

synchronization in the D2C link was assumed, and a 30 frame per second off-the-shelf camera for a receiver was considered. The frame rate of the camera is assumed to be greater than the display frame rate. This is because the camera should successfully capture the entire sequence of reference images to decode the transmitted data [28]. In addition, the number of embedded data symbols per image column, i.e., L, was set to 20 vertical pixels. Moreover, the position of the subbands in the frequency-domain image was set by considering the start pixel value, s, equal to 95 for SB1, 75 for SB2, and 45 for SB3. The default modulation in the simulation is 2-QAM unless otherwise specified.

5.1. SER for Different Receivers. The symbol error rate performance for all three receivers is compared in Figure 7. The

green data channel was evaluated for comparison, and all the symbols were 2-QAM modulated. We can see that the ZF receiver performs worse in all the subbands, and the ML receiver is the best. This is because ZF is performed by dividing the received frame by the reference frame; it does not care for the noise term and can amplify the noise in the process of estimating the bits. In the MMSE receiver, the coefficients were optimized and take care of the noise term amplification with the factor 1/SNR. Therefore, when the SNR becomes high, MMSE behaves like a ZF receiver. The ML receiver avoids the problem of noise enhancement, since it does not consider equalization. Instead, it estimates the transmitted symbol by choosing the minimum distance between the received image signal vector and all possible combinations of reference image signal vectors. Note that ML is the optimal

(a) ZF receiver

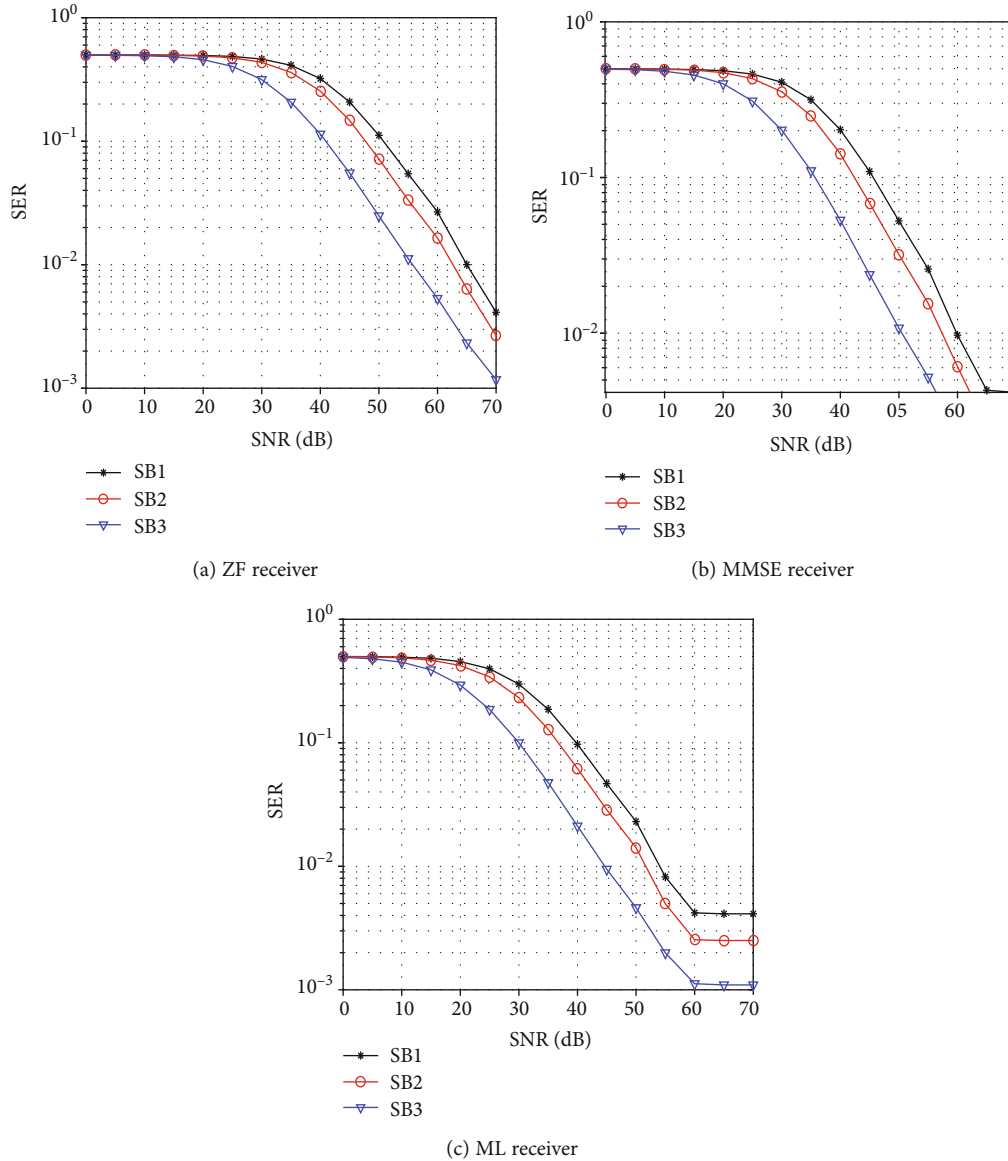(b) MMSE receiver

(c) ML receiver

FIGURE 8: SER of color DFC for different subbands over the red channel.

receiver in our case, because the occurrence probability of all the transmitted symbols is the same. Moreover, we can see that SB1 and SB3 show the worst and best performances, respectively.

5.2. SER for Different Subbands. The various receivers' performance according to the subbands is depicted in Figure 8. For this comparison, we evaluated the red data channel for all the receivers with 2-QAM modulated symbols. For low-SNR values (<15 dB), a similar SER performance can be observed for all the receivers due to poor communications links. However, when the SNR is increased, we can see that the SER performance of low-frequency bands, i.e., SB3, outperforms other bands for all receivers. This is because the energy of the low-frequency band is higher than that of mid- and high-frequency bands. As a result, data embedding

at a low frequency achieves better robustness against noise, compared to the other subbands.

5.3. SER for Different Modulation Order. Figure 9 presents SER as a function of SNR for the different modulation schemes. Here, the blue channel was considered in subband 1. Two modulation schemes were considered, i.e., 2-QAM and 4-QAM, having the modulation order (M) of 2 and 4, respectively. We can see that 2-QAM shows better performance than 4-QAM for all the receivers. This is because if the energy of the constellation plane remains the same, the points on the constellation plane must be closer together with an increasing modulation order (cf., Figure 2). Therefore, as modulation order increases, data transmission becomes more susceptible to noise. The effects of the distance between the adjacent points in the constellation plane become

(a) ZF receiver
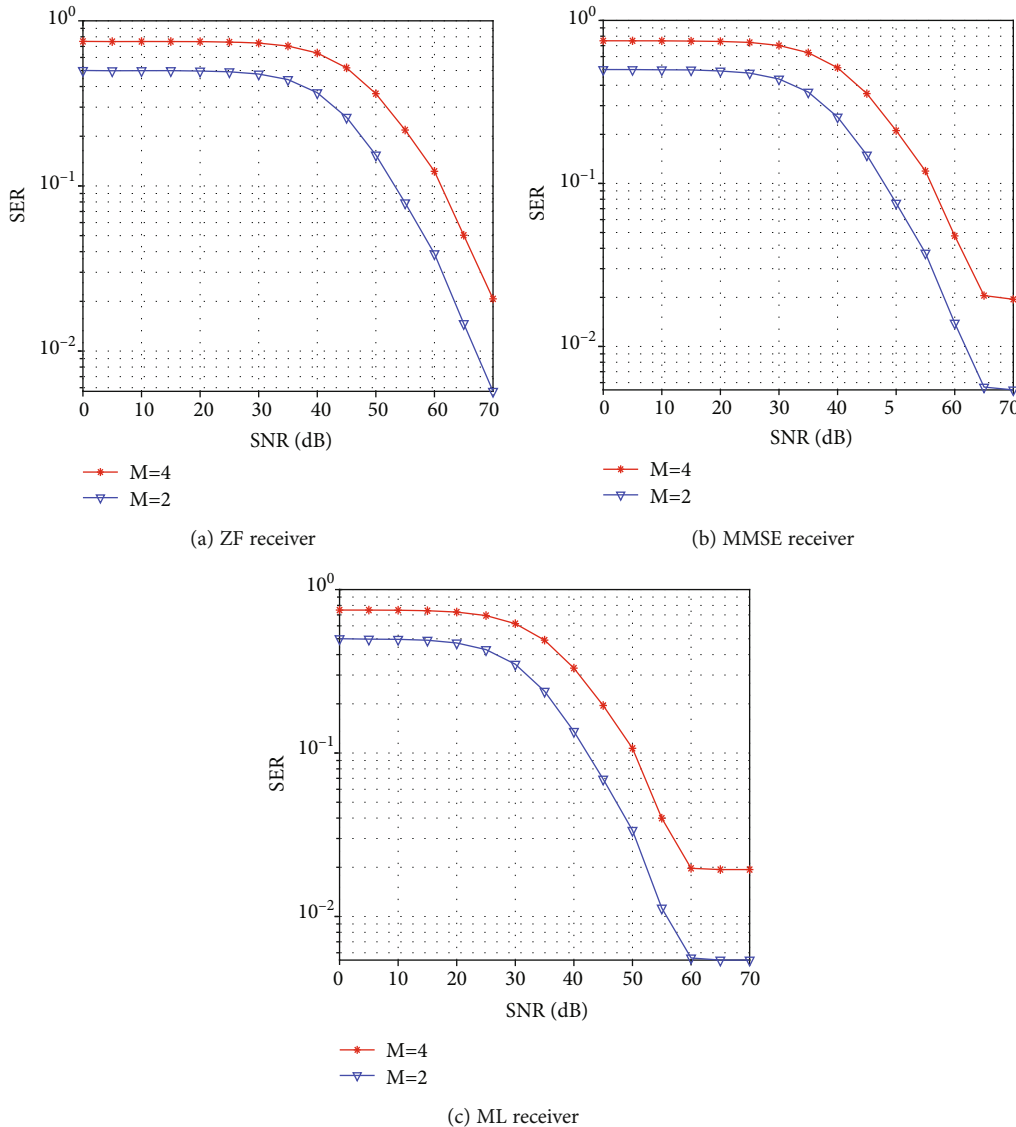
(b) MMSE receiver

(c) ML receiver

FIGURE 9: SER of color DFC for different receivers over the blue channel and subband 1.

significant as SNR increases. In addition, as the modulation order increases, the ML receiver's computational complexity increases exponentially. In particular, MMSE is computationally more complex than ML when the modulation is 2-QAM. However, if the modulation is 4-QAM or higher, ML becomes more complex.

5.4. SER for Different Color Channels. Figure 10 depicts the performance in subband 2 of different RGB channels for the three receivers. We can see that the performance of all the channels is similar, and there is a negligible difference. Therefore, we can say that the type of channel in the input image does not really affect SER performance in DFC. On the other hand, we can deduce that color DFC could provide three times the data rate compared to grayscale DFC, where there is one channel only.

5.5. Peak Signal-to-Noise Ratio. Table 1 shows PSNR values with regard to various subbands. We can see that as the position of the start pixels of the subband increases, the image quality of the data-embedded image is improved. Note that the visual characteristics of the image are located in the low frequencies, while the details and noise are located in higher frequencies. Since SB1 and SB2 occupy medium- to high-frequency regions, when using SB1 and SB2 for data embedding, visual artifacts are hardly ever perceived.

6. Conclusions

This paper evaluates the performance of a colored DFC scheme with three advanced receivers, including zero forcing, minimum mean square error, and maximum likelihood. This approach utilizes an RGB image on an electronic display as a transmitter, and a digital camera is the receiver. Because the RGB image is composed of three channels, the data can be embedded into each individual channel. In addition, we showed that data decoding

(a) ZF receiver

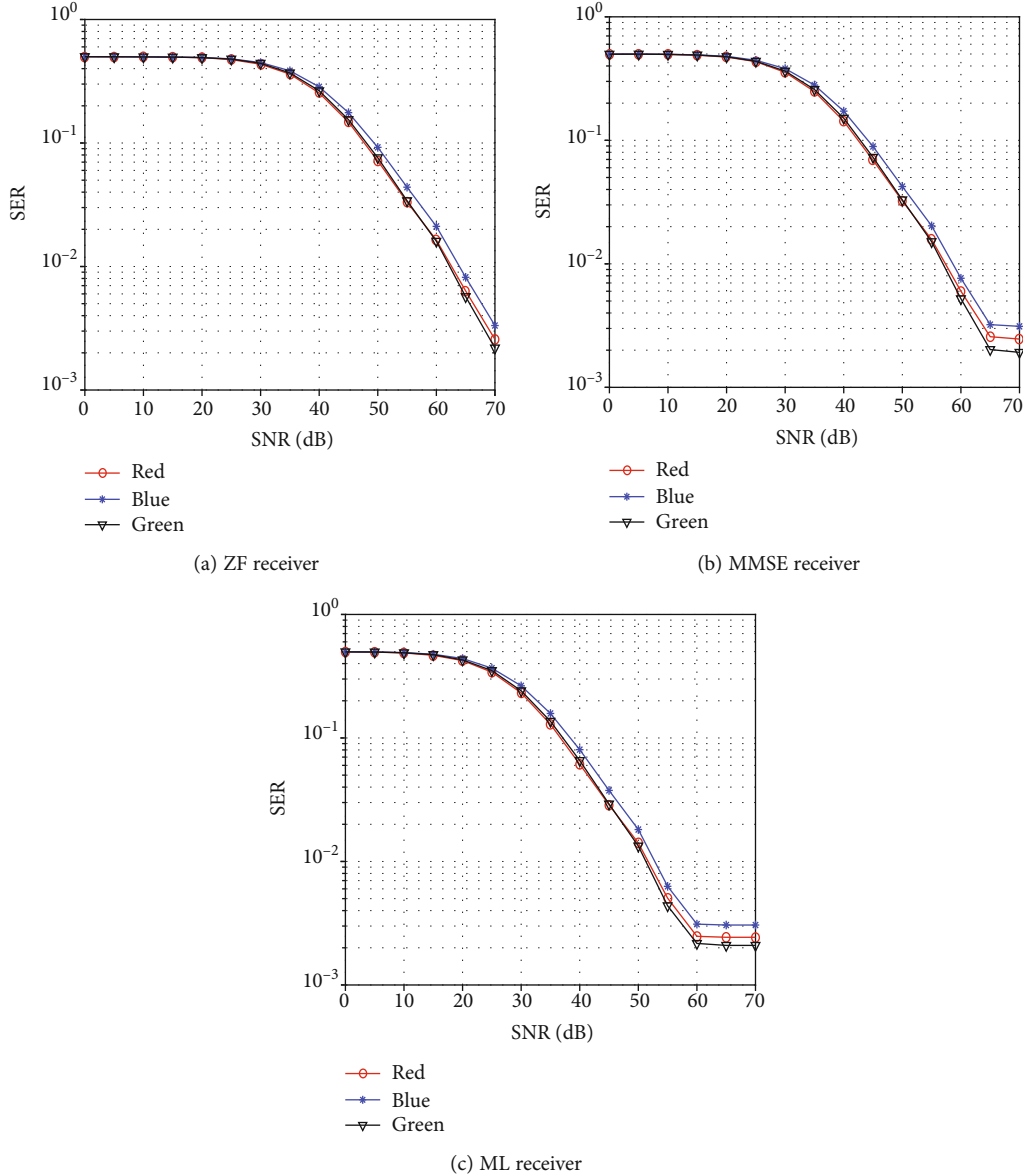(b) MMSE receiver

(c) ML receiver

FIGURE 10: SER of color DFC for the different receivers over subband 2.

TABLE 1: PSNR performance of a data-embedded image for different frequency subbands.

Bands	Values (2-QAM)	Values (4-QAM)
High-frequency band (SB1)	43.96 dB	43.92 dB
Mid-frequency band (SB2)	40.62 dB	40.63 dB
Low-frequency band (SB3)	33.75 dB	33.70 dB

using the ML receiver achieves the best performance. For the application of smart content services, color DFC is an important step towards realizing the potential for robust data communications while supporting the original functionality of displaying image sequences without image artifacts.

Data Availability

The simulation parameter data used to support the findings of this study are included within the article.

Conflicts of Interest

The authors declare that they have no conflicts of interest regarding the publication of this paper.

Acknowledgments

References

[1] N. T. Le, M. A. Hossain, and Y. M. Jang, "A survey of design and implementation for optical camera communication," *Signal Processing: Image Communication*, vol. 53, pp. 95–109, 2017.

[2] N. Saha, N. T. Le, M. S. Ifthekhar, and Y. M. Jang, "Survey on optical camera communications: challenges and opportunities," *IET Optoelectronics*, vol. 9, no. 5, pp. 172–183, 2015.

[3] T. Yamazato, I. Takai, H. Okada et al., "Image-sensor-based visible light communication for automotive applications," *IEEE Communications Magazine*, vol. 52, no. 7, pp. 88–97, 2014.

[4] I. Takai, T. Harada, M. Andoh, K. Yasutomi, K. Kagawa, and S. Kawahito, "Optical vehicle-to-vehicle communication system using LED transmitter and camera receiver," *IEEE Photonics Journal*, vol. 6, no. 5, pp. 1–14, 2014.

[5] J. Hu, F. C. P. Shin, K. Y. Sam, and C. Yu, "LED-camera communication system with RGB coding," in *2012 Photonics Global Conference (PGC)*, pp. 1–4, Singapore, Singapore, December 2012.

[6] J. Schaeffler, "What is digital signage?," in *Digital Signage*, pp. 61–92, Focal Press, 2012.

[7] "Global digital signage market statistics, trends, size, 2018 analysis and forecast to 2025," January 2019, https://www.reuters.com/brandfeatures/venturecapital/article?id=34829.

[8] D. Wave, "Information technology-automatic identification and data capture techniques-QR code bar code symbology specification," in *International Organization for Standardization*, vol. 18004, ISO/IEC, 2015.

[9] K. Kamijo, N. Kamijo, and Z. Gang, "Invisible barcode with optimized error correction," in *2008 15th IEEE International Conference on Image Processing*, pp. 2036–2039, San Diego, CA, USA, October 2008.

[10] A. Mohan, G. Woo, S. Hiura, Q. Smithwick, and R. Raskar, "Bokode: imperceptible visual tags for camera based interaction from a distance," *ACM Transactions on Graphics*, vol. 28, no. 3, p. 98, 2009.

[11] T. Hao, R. Zhou, and G. Xing, "COBRA: color barcode streaming for smartphone systems," in *MobiSys '12: Proceedings of the 10th international conference on Mobile systems, applications, and services*, pp. 85–98, New York, NY, USA, June 2012.

[12] W. Hu, H. Gu, and Q. Pu, "LightSync: unsynchronized visual communication over screen-camera links," in *MobiCom '13: Proceedings of the 19th Annual International Conference on Mobile Computing & Networking*, pp. 15–26, New York, NY, USA, September 2013.

[13] A. Wang, S. Ma, C. Hu, J. Huai, C. Peng, and G. Shen, "Enhancing reliability to boost the throughput over screen-camera links," in *MobiCom '14: Proceedings of the 20th annual international conference on Mobile computing and networking*, pp. 41–52, New York, NY, USA, September 2014.

[14] M. R. H. Mondal and J. Armstrong, "Analysis of the effect of vignetting on MIMO optical wireless systems using spatial OFDM," *Journal of Lightwave Technology*, vol. 32, no. 5, pp. 922–929, 2014.

[15] R. Carvalho, C.-H. Chu, and L.-J. Chen, "IVC: imperceptible video communication," in ACM, 2014.

[16] W. Yuan, K. Dana, A. Ashok, M. Gruteser, and N. Mandayam, "Dynamic and invisible messaging for visual MIMO," in *2012 IEEE Workshop on the Applications of Computer Vision (WACV)*, Breckenridge, CO, USA, January 2012.

[17] A. Wang, C. Peng, O. Zhang, G. Shen, and B. Zeng, "InFrame: multiflexing full-frame visible communication channel for humans and devices," in *Proceedings of the 13th ACM Workshop on Hot Topics in Networks - HotNets-XIII*, p. 23, New York, NY, USA, October 2014.

[18] P. H. Pathak, X. Feng, P. Hu, and P. Mohapatra, "Visible light communication, networking, and sensing: a survey, potential and challenges," *IEEE Communications Surveys Tutorials*, vol. 17, no. 4, pp. 2047–2077, 2015.

[19] T. Nguyen, M. D. Thieu, and Y. M. Jang, "2D-OFDM for optical camera communication: principle and implementation," *IEEE Access*, vol. 7, pp. 29405–29424, 2019.

[20] S. D. Perli, N. Ahmed, and D. Katabi, "PixNet: interference-free wireless links using LCD-camera pairs," in *MobiCom '10: Proceedings of the sixteenth annual international conference on Mobile computing and networking*, pp. 137–148, New York, NY, USA, September 2010.

[21] T. Li, C. An, A. Campbell, and X. Zhou, "Hilight: hiding bits in pixel translucency changes," in *VLCS '14: Proceedings of the 1st ACM MobiCom workshop on Visible light communication systems*, pp. 45–50, New York, NY, USA, September 2014.

[22] V. Nguyen, Y. Tang, A. Ashok et al., "High-rate flicker-free screen-camera communication with spatially adaptive embedding," in *IEEE INFOCOM 2016 - The 35th Annual IEEE International Conference on Computer Communications*, pp. 1–9, San Francisco, CA, USA, April 2016.

[23] K. Jo, M. Gupta, and S. K. Nayar, "DisCo: display-camera communication using rolling shutter sensors," *ACM Transactions on Graphics*, vol. 35, no. 5, p. 150, 2016.

[24] T. Li, C. An, X. Xiao, A. T. Campbell, and X. Zhou, "Real-time screen-camera communication behind any scene," in *MobiSys '15: Proceedings of the 13th Annual International Conference on Mobile Systems, Applications, and Services*, pp. 197–211, New York, NY, USA, May 2015.

[25] K. Zhang, Y. Zhao, C. Wu et al., "ChromaCode: a fully imperceptible screen-camera communication system," *IEEE Transactions on Mobile Computing*, p. 1, 2019.

[26] S. Shi, L. Chen, W. Hu, and M. Gruteser, "Reading between lines: high-rate, non-intrusive visual codes within regular videos via implicitcode," in *UbiComp '15: Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, pp. 157–168, New York, NY, USA, September 2015.

[27] A. Wang, Z. Li, C. Peng, G. Shen, G. Fang, and B. Zeng, "Inframe++: achieve simultaneous screen-human viewing and hidden screen camera communication," in *MobiSys '15: Proceedings of the 13th Annual International Conference on Mobile Systems, Applications, and Services*, pp. 181–195, New York, NY, USA, May 2015.

[28] B. W. Kim, H.-C. Kim, and S.-Y. Jung, "Display field communication: fundamental design and performance analysis," *Journal of Lightwave Technology*, vol. 33, no. 24, pp. 5269–5277, 2015.

[29] S.-Y. Jung, H.-C. Kim, and B. W. Kim, "Implementation of two-dimensional display field communications for enhancing the achievable data rate in smart-contents transmission," *Displays*, vol. 55, pp. 31–37, 2018.

[30] S.-A. Kim and S.-Y. Jung, "Screen communication scheme based on display," *The Journal of Korean Institute of*

Communications and Information Sciences, vol. 44, no. 4, pp. 670–673, 2019.

[31] D. A. Kerr, *Derivation of the Cosine Fourth Law for Falloff of Illuminance Across a Camera Image*, Technical Report 4, 2007.

[32] S. Hranilovic and F. R. Kschischang, "A pixelated MIMO wireless optical communication system," *IEEE Journal of Selected Topics in Quantum Electronics*, vol. 12, no. 4, pp. 859–874, 2006.

[33] E. R. Dougherty, *Random Processes for Image and Signal Processing*, SPIE Optical Engineering Press, 1999.

[34] T. Komine and M. Nakagawa, "Fundamental analysis for visible-light communication system using LED lights," *IEEE Transactions on Consumer Electronics*, vol. 50, no. 1, pp. 100–107, 2004.

[35] A. P. Tang, J. M. Kahn, and K.-P. Ho, "Wireless infrared communication links using multi-beam transmitters and imaging receivers," in *Proceedings of ICC/SUPERCOMM '96 - International Conference on Communications*, pp. 180–186, Dallas, TX, USA, June 1996.

[36] P. Cheremkhin, V. Lesnichii, and N. Petrov, "Use of spectral characteristics of DSLR cameras with Bayer filter sensors," *Journal of Physics: Conference Series*, vol. 536, no. 1, article 012021, 2014.

[37] R. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*, Cambridge University Press, 2003.

[38] W. Burger and M. J. Burge, *Principles of Digital Image Processing: Fundamental Techniques*, Springer Science & Business Media, 2010.