Image Broadcasting for Heterogeneous User Devices in MIMO Networks

Soyoung Jang[1], Seok-Ho Chang[1], Minyeong Kim[1], and Sunghyun Cho[2]

[1]Computer Science and Engineering, Dankook University, Yongin 16890, South Korea
[2]Information and Communications Engineering, DGIST, Daegu 42988, South Korea

***Abstract*—This paper considers a multimedia broadcasting scenario in which two types of heterogeneous users with different display resolutions and different numbers of antennas stay in the service area. We propose an image broadcasting scheme that uses the image super-resolution (SR) techniques, spatial diversity, and diversity-multiplexing tradeoff (DMT) achieving codes. The proposed scheme broadcasts a low-resolution (LR) image to two types of users, along with residual pixel-error map containing high-frequency details of high-resolution (HR) image. Then, a user retaining an HR screen employs SR to reconstruct an HR image from the received LR image, and exploits the residual map to further enhance the image quality. Our scheme properly trains the neural network models of the deep learning-based SR by taking into account the source coding rates of the images. Considering the relationship between the number of antennas and screen resolution, based on hardware space of user devices, the proposed scheme encodes an LR image with spatial diversity, and encodes residual map with DMT-achieving codes. Numerical evaluation shows that our scheme significantly outperforms the baseline strategy that broadcasts either HR or LR images.**

I. Introduction

In recent years, there has been significant demand for the wireless multimedia services [1], [2]. Multiple-input multiple-output (MIMO) channels offer large gains in terms of link reliability and data rate. Some space-time codes, such as Golden code [3], achieve the optimal diversity-multiplexing tradeoff (DMT) characteristics. In mobile multimedia broadcasting systems such as digital video broadcasting (DVB), there exist heterogeneous user devices in the same service area, and they in general have different numbers of antennas and different display resolutions. As an example, a tiny mobile phone may have a single antenna and a low-resolution (LR) screen, due to its limited hardware space. On the other hand, a device such as a tablet or notebook computer usually has more than one antenna and a high-resolution (HR) screen.

Let N_{t} and N_{r} denote the numbers of transmit and receive antennas, respectively. Then, we can achieve a spatial multiplexing rate of up to $\min(N_{\mathrm{t}}, N_{\mathrm{r}})$. Note that the base station (BS) typically has more antennas than does the user. Suppose that there are users with a single antenna in the broadcasting service area. Then, from the fact that $\min(N_{\mathrm{t}}, 1)$ is unity, the BS needs to broadcast images using spatial diversity rather than spatial multiplexing, such that the data is decodable by all types of users. This may lead to a significant performance loss for users retaining more than one antenna.

We next describe an issue on the display resolution, when two types of users having HR and LR screens reside in the same service area. If the BS broadcasts an HR image to both users, a user with an LR screen needs to downsample the received HR image to satisfy the display resolution. This results in the quality degradation compared to the case where LR image is sent to a user with an LR screen. On the other hand, if the BS broadcasts an LR image to both users, a user with an HR screen needs to enhance the resolution of the received LR image, which leads to the quality impairment. Namely, forcing the BS to broadcast either HR or LR images for all types of users yields a performance loss, compared to the one-to-one transmission case.

To address the problem, we propose a strategy that efficiently serves heterogeneous users. As a simple example, we consider a broadcasting system where a BS adopts two antennas, and two different types of users stay in the service area: i) a *big user* (i.e., a user of large hardware space) with two antennas and an HR screen, ii) a *small user* (i.e., a user of small hardware space) with a single antenna and an LR screen. For this setup, we propose a broadcasting scheme which employs image super-resolution (SR) techniques [4], spatial diversity, and the optimal DMT-achieving codes.

Specifically, the proposed scheme broadcasts an LR image to two types of users, along with additional information we refer to as *residual pixel-error map* (or, simply, *residual map*) containing high-frequency details of HR image. Then, a big user employs SR algorithms to reconstruct an HR image from the received LR image, and exploits the residual map to further enhance the image quality. In the proposed scheme, the neural network models of the SR are adequately trained, according to the source coding rates of the images. Regarding the space-time codes, the LR image bitstream is encoded by spatial diversity techniques such as the Alamouti code, while the residual map is encoded by the optimal DMT-achieving code such as the Golden code. As a result, LR image is decodable by both the big and small users, while residual map is decodable by only the big user.

The work in [5] addressed the issues considered in this paper. In [5], scalable video coding, unequal error protection, and superposition coding are incorporated, whereas this paper adopts image SR techniques. Note that in [6], the SR algorithm was studied for the transmission of LR videos. The performance of the scheme proposed in [6] is compared to that of the simple LR video transmission scheme, where the

978-1-5386-8088-9/19/$31.00 ©2019 IEEE

bilinear interpolation is used to reconstruct HR video.

II. MIMO Image Broadcasting Schemes

A. Baseline MIMO Image Broadcasting Schemes

The first baseline scheme broadcasts HR images to both the big and small users. The system takes the bitstream from the image source coder, and then the bitstream is converted into a sequence of channel codewords with error detection and correction capability. The coded bitstream is then mapped to constellation symbols. In this baseline scheme, a small user needs to downsample the received HR image to meet the display resolution.

The second baseline scheme broadcasts LR images. In this scheme, a big user needs to enhance the resolution of the received LR image. In the baseline strategies, HR or LR images are encoded by only spatial diversity, which enables both the big and small users to decode the received images.

B. Proposed MIMO Image Broadcasting Scheme

Fig. 1 depicts the proposed scheme, which broadcasts LR images to both users along with residual map. This map describes the difference between the original high-quality HR image and the low-quality HR image reproduced from the LR image using SR techniques. The residual map in general contains high-frequency details of HR image. We compute the map by convex optimization, so as to achieve the best end-to-end performance as measured by the expected distortion (or, equivalently, mean-squared error between the original image and the reproduced one). The LR image is encoded by Alamouti code, and thus is decodable by both the big and small users. On the other hand, the residual map is encoded by Golden code, and is decodable only by a big user. The motivation for this is as follows: i) A big user needs to decode many source bits to meet the quality of its HR screen, but the LR screen of a small user can produce satisfactory quality with a small number of source bits. Thus, for a small user, decoding of only the LR image is sufficient. ii) When the residual map is encoded by Golden code, the image quality for a big user is significantly enhanced.

B.1 Training of Networks Based on Source Coding Rates

For the SR techniques, we employ EDSR [7], which is based on SRResNet architecture [8]. The SRResNet consists of multiple residual blocks, and each residual block consists of convolution, batch normalization, and rectified linear unit (ReLU) layers. EDSR adopts simplified residual blocks without batch normalization layers to achieve better performance and smaller memory footprint. While the original EDSR shows good performance, it is not directly applicable to our task in this paper. The network models of the original EDSR are trained to enhance the resolution of a high-quality LR image, and compressed LR images are not considered. Consequently, the original EDSR fails to restore a high-quality HR image from a compressed LR image.

To adequately reflect the effect of image compression on SR, we train EDSR networks with respect to different source coding rates (or, equivalently, compression rates, or bit rates), instead of using a pre-trained EDSR network. To do so, we utilize DIV2K dataset [9], which is a publicly available dataset designed to train neural networks for image SR, and provides 800 HR images for training. We first generate downsampled LR images from the HR images in DIV2K. Specifically, we apply anti-aliasing lowpass filtering and downsampling to each HR image, and then compress the downsampled image with different bit rates of $0.05, 0.1, 0.15, \ldots, 2.0$ bits per pixel (bpp).

Using the compressed LR images, we train EDSR networks as follows: To begin, we train EDSR with the compression rate of 2.0 bpp. To do so, we use 800 pairs of original HR images and compressed LR images having a bit rate of 2.0 bpp. The SR networks are trained for 300 epochs from scratch. We let $\text{EDSR}_{2.0}$ denote the network models trained for 2.0 bpp. To train EDSR for lower compression rates, we initialize the network with the parameters of $\text{EDSR}_{2.0}$, and run 150 epochs. We use ADAM optimizer [10] with default parameters $\beta_1 = 0.9$ and $\beta_2 = 0.999$. Training each network for 150 epochs takes about 20 hours with an NVIDIA GeForce GTX 1080 Ti graphics card.

B.2 Computation of the Residual Pixel-Error Map

The residual map, denoted by $\mathbf{R}$, contains the difference between the ground-truth HR image, denoted by $\mathbf{U}_0$, and the low-quality HR image, denoted by $\mathbf{U}_1$, which has been reconstructed from the compressed LR image, using the SR techniques (refer to Fig. 1(a)). $\mathbf{R}$, $\mathbf{U}_0$, and $\mathbf{U}_1$ are $N_{\text{HR}} \times N_{\text{HR}}$ matrices (i.e., the resolution of the HR image). In the proposed system, using $\mathbf{U}_1$ and $\mathbf{R}$, the big user reproduces an HR image, $\mathbf{U}_2$, as follows:

$$\mathbf{U}_2 = \mathbf{U}_1 + \mathbf{G} * \mathbf{R}, \tag{1}$$

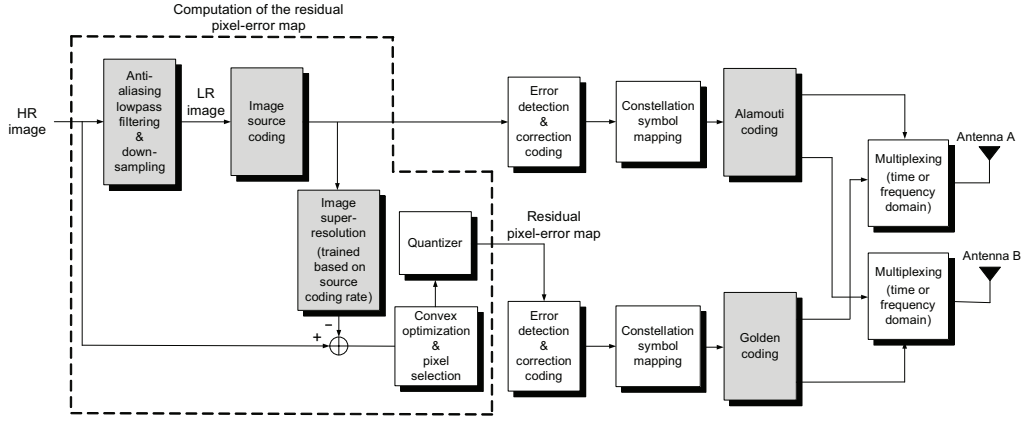
where $*$ indicates a convolution operation and $\mathbf{G}$ is a pre-defined 2-dimensional convolution kernel in the form of $N_{\text{G}} \times N_{\text{G}}$ matrix[1] (N_{G} is an odd number in the range of $3 \le N_{\text{G}} < N_{\text{HR}}$). As an example, we consider the following 3×3 Gaussian convolution kernel:

$$\mathbf{G} = \begin{bmatrix} g_2 & g_1 & g_2 \\ g_1 & 1 & g_1 \\ g_2 & g_1 & g_2 \end{bmatrix}, \tag{2}$$

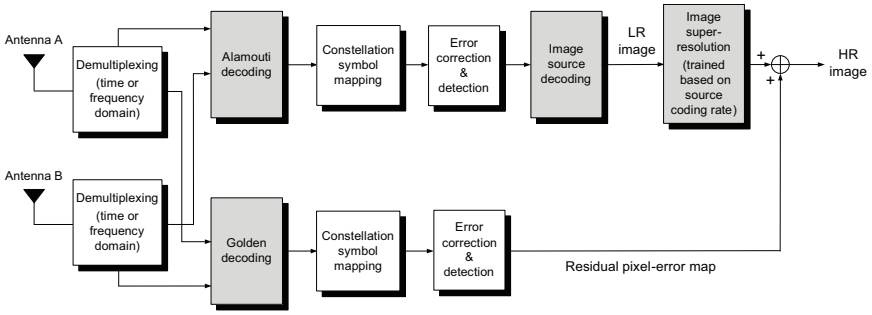
where g_1 and g_2 are real in the range of $0 \le g_2 \le g_1 \le 1$.

In the following, we describe how to compute residual map. Let $\mathbf{R}(i,j)$ denote the element at the ith row and the jth column of $\mathbf{R}$. Also let $\Omega = \{(i_1, j_1), \ldots, (i_L, j_L)\}$ denote a set of indices for L elements selected from N_{HR}^2 elements in $\mathbf{R}$. Subject to the constraint that $\mathbf{R}(x,y) = 0$ for $\forall (x,y) \in \Omega^c$ (i.e., $N_{\text{HR}}^2 - L$ elements in $\mathbf{R}$ are set to zeros), using convex optimization techniques, we find the optimal residual map, $\mathbf{R}^*$, that minimizes the mean squared-error

[1]The matrix resulted from $\mathbf{G} * \mathbf{R}$ has the size of $(N_{\text{HR}} + N_{\text{G}} - 1) \times (N_{\text{HR}} + N_{\text{G}} - 1)$. However, we assume that after convolution, $(N_{\text{G}} - 1)/2$ right-most and left-most columns, and $(N_{\text{G}} - 1)/2$ top-most and bottom-most rows are truncated, such that $\mathbf{G} * \mathbf{R}$ becomes an $N_{\text{HR}} \times N_{\text{HR}}$ matrix.



(a) Transmitter



(b) Receiver of the big user

Fig. 1. The proposed MIMO image broadcasting system.

between $\mathbf{U}_0$ and $\mathbf{U}_2$. The constraint indicates that at most L elements in $\mathbf{R}$ have nonzero values (i.e., L is the maximum number of nonzero-valued elements in $\mathbf{R}$). L should be chosen considering the data rate (or bit budget) assigned to the transmission of residual map.

Convolution kernel, $\mathbf{G}$, in (1) is introduced to better reproduce $\mathbf{U}_2$, in a way that the big user roughly estimates several elements that are located close to a nonzero-valued element $\mathbf{R}(x,y) \neq 0$ for $\forall (x,y) \in \Omega$. As an example, using Gaussian kernel given by (2), the big user estimates up to 8 elements that are spatially correlated with the nonzero-valued element located at the center of $\mathbf{G}$. To compute the optimal residual map, we solve a constrained optimization problem:

$$\begin{aligned} \mathbf{R}^* = \arg\min_{\mathbf{R}} \ & \|(\mathbf{U}_1 + \mathbf{G} * \mathbf{R}) - \mathbf{U}_0\|_{\mathrm{F}}^2 + \alpha \|\mathbf{R}\|_{\mathrm{F}}^2 \\ \text{subject to} \ \ & \mathbf{R}(x,y) = 0 \ \ \text{for} \ \forall (x,y) \in \Omega^c, \end{aligned} \tag{3}$$

where $\| \cdot \|_{\mathrm{F}}$ denotes the Frobenius norm and $\alpha \|\mathbf{R}\|_{\mathrm{F}}^2$ is a Tikhonov regularization term. Recall that the constraint in (3) is introduced because of the limited bit budget (i.e., the limited number of bits) assigned to the transmission of residual map. Optimization given by (3) can be expressed as the unconstrained optimization problem:

$$\begin{aligned} \mathbf{R}^* \ = & \arg\min_{\mathbf{R}} \ \|(\mathbf{U}_1 + \mathbf{G} * \mathbf{R}) - \mathbf{U}_0\|_{\mathrm{F}}^2 + \alpha \|\mathbf{R}\|_{\mathrm{F}}^2 \\ & + \beta \sum_{(x,y) \in \Omega^c} |\mathbf{R}(x,y)|^2, \end{aligned} \tag{4}$$

where β is a large constant to force $\mathbf{R}(x,y)$ to be zero for $\forall (x,y) \in \Omega^c$. Eq. (4) can be rewritten as

$$\mathbf{r}^* = \arg\min_{\mathbf{r}} \ \|(\mathbf{u}_1 + \mathbf{B}\mathbf{r}) - \mathbf{u}_0\|^2 + \alpha \|\mathbf{r}\|^2 + \beta \mathbf{r}^T \mathbf{M}_{\Omega^c} \mathbf{r}, \tag{5}$$

where $\mathbf{r}^*$, $\mathbf{r}$, $\mathbf{u}_1$, and $\mathbf{u}_0$ are vectors obtained by row-wise stacking $\mathbf{R}^*$, $\mathbf{R}$, $\mathbf{U}_1$, and $\mathbf{U}_0$, which are $N_{\mathrm{HR}} \times N_{\mathrm{HR}}$ matrices, into $N_{\mathrm{HR}}^2 \times 1$ vectors, respectively; $\mathbf{B}$ is an $N_{\mathrm{HR}}^2 \times N_{\mathrm{HR}}^2$ matrix that represents $\mathbf{G}$, so that $\mathbf{B}\mathbf{r}$ is equal to $\mathbf{G} * \mathbf{R}$; and $\mathbf{M}_{\Omega^c}$ is an $N_{\mathrm{HR}}^2 \times N_{\mathrm{HR}}^2$ diagonal matrix whose diagonal elements are set to either 0 or 1 according to Ω^c (i.e., masking matrix) such that $\mathbf{r}^T \mathbf{M}_{\Omega^c} \mathbf{r}$ is equal to $\sum_{(x,y) \in \Omega^c} |\mathbf{R}(x,y)|^2$. The quadratic optimization problem of (5) can be converted to the linear equation: If the object function in (5) is differentiated with regard to $\mathbf{r}$, and the resulting function is set to zero with $\mathbf{r}^*$ substituted into $\mathbf{r}$, then we have

$$(\mathbf{B}^T\mathbf{B} + \alpha \mathbf{I}_{N_{\mathrm{HR}}^2 \times N_{\mathrm{HR}}^2} + \beta \mathbf{M}_{\Omega^c})\mathbf{r}^* + \mathbf{B}^T(\mathbf{u}_1 - \mathbf{u}_0) = 0, \tag{6}$$

where $(\cdot)^T$ denotes transpose of matrix. From (6), we obtain r^* by using the conjugate gradient descent method.

In the following, based on (6), we present the proposed method of computing residual map, including how to choose the set $\Omega = \{(i_1, j_1), \ldots, (i_L, j_L)\}$ (i.e., how to select L elements that will not be forced to be zeros). The proposed algorithm takes an iterative approach as follows.

Step 1: Let $N_{\mathrm{iter}} (\geq 2)$ denote the number of iterations in this algorithm. For given L $(< N_{\mathrm{HR}}^2)$ and N_{iter}, set a factor $\lambda = (L/N_{\mathrm{HR}}^2)^{\frac{1}{N_{\mathrm{iter}}-1}}$ (< 1).

Step 2: Initialize the cardinality of $\Omega = \{(i_1, j_1), \ldots, (i_{L_{\text{temp}}}, j_{L_{\text{temp}}})\}$ as $L_{\text{temp}} = N_{\text{HR}}^2$ (i.e., all the elements in $\mathbf{R}$ are initially forced to be in Ω).

Step 3: The iteration index, k, is initialized as $k = 1$.

Step 4: Downsample the ground-truth HR image, $\mathbf{U}_0$. Compress the resulting LR image, and from this, reconstruct HR image, $\mathbf{U}_1$, using the SR techniques of which the network models are trained based on the compression rate.

Step 5: From (6), for given $\mathbf{u}_0$, $\mathbf{u}_1$, $\mathbf{B}$, and $\mathbf{M}_{\Omega^c}$ (or, equivalently, $\mathbf{U}_0$, $\mathbf{U}_1$, $\mathbf{G}$, and $\Omega = \{(i_1, j_1), \ldots, (i_{L_{\text{temp}}}, j_{L_{\text{temp}}})\}$, respectively), find the optimal residual map $\mathbf{r}^*$ (or, equivalently, $\mathbf{R}^*$) using the conjugate gradient descent method.

Step 6: If the iteration index is $k = N_{\text{iter}}$ (i.e., $L_{\text{temp}} = L$), we have finally obtained $\mathbf{R}^*$. Otherwise,

1) Update the cardinality of $\Omega = \{(i_1, j_1), \ldots, (i_{L_{\text{temp}}}, j_{L_{\text{temp}}})\}$ as $L_{\text{temp}} = \lambda L_{\text{temp}}$, where λ is a factor defined in Step 1.
2) Select L_{temp} elements from $\mathbf{R}^*$ in such a way that the magnitudes of L_{temp} elements (i.e., $|\mathbf{R}^*(x, y)|$) are the largest in $\mathbf{R}^*$. Accordingly, update the set $\Omega = \{(i_1, j_1), \ldots, (i_{L_{\text{temp}}}, j_{L_{\text{temp}}})\}$ and $\mathbf{M}_{\Omega^c}$.
3) Update $k = k + 1$, and go to Step 5. □

The residual map, $\mathbf{R}^*$, which has been obtained from Steps 1–6, is then quantized as follows: For every element in $\mathbf{R}^*$, we perform the following.

i) $\mathbf{R}^*(x, y) = Q_{\text{level}} \times \text{Round}(\mathbf{R}^*(x, y)/Q_{\text{level}})$, where Q_{level} is a quantization level, and $\text{Round}(\cdot)$ indicates a rounding operation.
ii) $\mathbf{R}^*(x, y)$ is limited to the range of $-Q_{\max}Q_{\text{level}} \le \mathbf{R}^*(x, y) \le Q_{\max}Q_{\text{level}}$.

As a result, $\mathbf{R}^*(x, y)$ is quantized as $n \cdot Q_{\text{level}}$, where $n = -Q_{\max}, -Q_{\max} + 1, \ldots, 0, \ldots, Q_{\max}$.

We next describe the data size of the residual map. Let k_n denote the number of elements in $\mathbf{R}^*$ that have the value of $n \cdot Q_{\text{level}}$ after $\mathbf{R}^*$ has been quantized. Then, the number of ways of assigning N_{HR}^2 elements, each having a value of $n \cdot Q_{\text{level}}$, to $\mathbf{R}$ of $N_{\text{HR}} \times N_{\text{HR}}$ matrix is given by

$$N_{\text{R}} = \binom{N_{\text{HR}}^2}{k_{-Q_{\max}}, k_{-Q_{\max}+1}, \ldots, k_0, \ldots, k_{Q_{\max}}}, \quad (7)$$

where $\binom{n}{l_1, \ldots, l_m} = n!/l_1! \cdots l_m!$ is a multinomial coefficient. From (7), the data size of the residual map is $\log_2 N_{\text{R}}$ (bits).

III. Numerical Evaluation

We evaluate the peak signal-to-noise ratio (PSNR) performances of the proposed and baseline schemes. For the image source coder, we use SPIHT [11], and the performance is evaluated for 100 test images provided by DIV2K [9]. In our evaluation, we employ a resolution of 512×512 for HR images (i.e., $N_{\text{HR}} = 512$), and a resolution of 256×256 for LR images. To compare the image quality, we use the PSNR defined as $10 \log_{10}(255^2/\text{E}[D])$ (dB), where $\text{E}[D]$ is the expected distortion of the image.

For this evaluation, we set the parameters described in Section II.B.2 as follows: $g_1 = 0.2$, $g_2 = 0.02$, $\alpha = 10^{-7}$, $\beta = 10^7$, $N_{\text{iter}} = 15$, $Q_{\text{level}} = 16$, $Q_{\max} = 6$, and L is set depending on the bit budget assigned to the residual map, $\log_2 N_{\text{R}}$ (bits), given by (7). To begin, we observe the PSNR performance of the big user in a SISO system (i.e., the big user has an HR screen and a single antenna), to exclude the effect of space-time codes on the performance, and to more clearly see the effect of two techniques specified in Sections II.B.1 and II.B.2. In Fig. 2, we show the PSNR of the big user for the following cases.

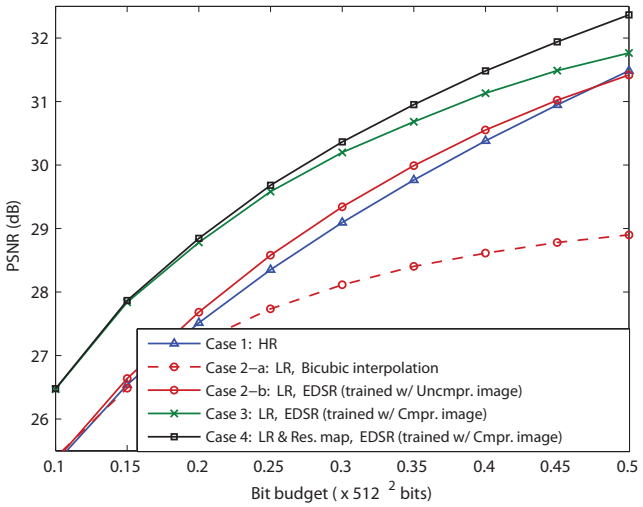


Fig. 2. PSNR performance of a big user in a SISO system with a bandwidth-limited, but error-free channel.

Case 1: HR image is broadcast.
Case 2-a: LR image is broadcast. HR image is reproduced using bicubic interpolation.
Case 2-b: LR image is broadcast. HR image is reproduced using the original SR networks trained with uncompressed LR images.
Case 3: LR image is broadcast. HR image is reproduced using SR networks trained according to the source coding rates of the LR images.
Case 4: Both LR image and residual map are broadcast. HR image is reproduced in the same way as Case 3.

Note that Cases 1 and 2 correspond to the baseline strategies presented in Section II.A, except that no space-time coding is employed. In Fig. 2, x-axis indicates the bit budget (i.e., the number of bits) used to send images and/or residual map. It is assumed that channel is bandwidth-limited, but error-free, such that the bit budget is not used for error-correction coding, but is entirely used to produce compressed images (i.e., source coding) and/or residual map. Note that the bit budget of x bits corresponds to the bit rate of $x/512^2$ bpp for 512×512 HR images, and $x/256^2$ bpp for 256×256 LR images. In Fig. 2, y-axis indicates PSNR value (dB) averaged over 100 test images.

Fig. 2 shows that Case 2-a performs the worst, and Case 2-b performs much better than does Case 2-a, since Case 2-b employs EDSR that is superior to bicubic interpolation. However, it is seen that compared to Case 1, Case 2-b rarely performs

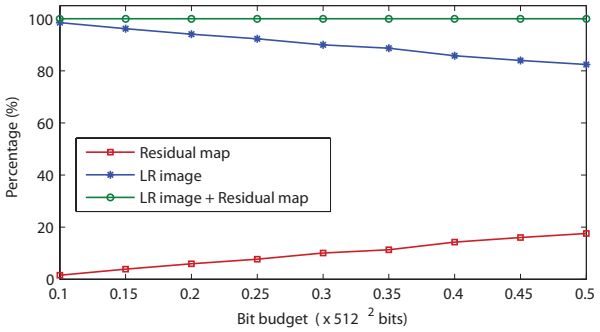


Fig. 3. Percentage of best assigned bit budget.

better, while Case 3 considerably improves the performance of Case 1. As shown, the PSNR gap between Cases 2-b and 3 is large for a small bit budget. This is because source coder compresses images a lot for a small bit budget, and thus the impact of training the network models considering the compression is significant when a bit budget is small. It is also observed that the PSNR gain of Case 3 relative to Case 1 is large for a small bit budget. We conjecture the reason for this as follows: For a small bit budget, images tend to have large quantization errors because of severe compression. Recall that, for a given bit budget of x bits, LR image has a bit rate of $x/256^2$ bpp, whereas HR image has a bit rate of $x/512^2$ bpp. Thus, if we send LR images instead of HR images, we are able to offer higher bit rate for the same bit budget, which results in smaller quantization errors. The effect of compression (i.e., quantization) on the image distortion is as follows: At a high bit rate, the distortion-rate curves of the images show the exponential rate decay of 2, as is stated in [12], whereas the exponential rate decay is larger than 2 at a low or middle bit rate [13]. This indicates that, when a bit rate is low, image distortion caused by quantization decreases rapidly as bit rate increases. As a result, for a small bit budget, it is beneficial to send LR images instead of HR images, so that distortion due to quantization can be reduced significantly thanks to the increased bit rate, though additional errors are introduced when enhancing the resolution of LR images at the receiver (i.e., the additional errors result from SR). On the other hand, for a large bit budget, it is less beneficial to send LR images instead of HR images, because distortion due to quantization does not decrease rapidly as bit rate increases, whereas additional errors result from SR. Thus, as shown in Fig. 2, the PSNR gap between Cases 1 and 3 is small for a large bit budget.

From Fig. 2, it is also observed that Case 4 improves the PSNR performance of Case 3. Note that in Case 4, a given bit budget is divided into two parts, each assigned to the LR image and the residual map. We have tried various ways of dividing bit budget for the LR image and the residual map; for each given bit budget, the best division, which yields the maximum PSNR value, is depicted in Fig. 3. The PSNR curve of Case 4 shown in Fig. 2 and the best assignment curve shown in Fig. 3 are the ones averaged over 100 test images.

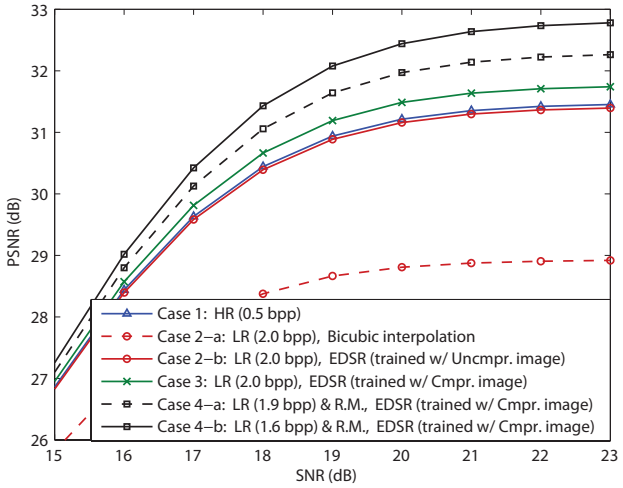
As shown in Fig. 2, when a bit budget is large, the performance gain achieved by residual map is large. The reason for this is conjectured in the following. If we allocate a given bit budget to the LR image, we can reduce quantization errors since the LR image is less compressed thanks to the increased bit rate. On the other hand, if a given bit budget is used for the residual map, we can reduce the errors resulted from SR. As described earlier, at a low bit rate, image distortion caused by quantization rapidly decreases as bit rate increases. Based on this fact, we conjecture that for a small bit budget, it is beneficial to assign most bit budget to the LR image than to the residual map, so that image distortion due to quantization can be reduced significantly because of the increased bit rate. We also conjecture that for a large bit budget, it is less effective to assign most bit budget to the LR image, because distortion caused by quantization does not decrease rapidly as bit rate increases; it is better to assign part of the bit budget to residual map, so that we can reduce the errors resulted from SR. That is, residual map plays a considerable role in the performance improvement when a bit budget is large. This can be observed from Fig. 3. As bit budget increases, the percentage of the bit budget that is best assigned to the residual map increases, while that of the bit budget assigned to the LR image decreases.

We note that, as shown in Fig. 2, even in the case where a big user is the only receiver (i.e., one-to-one transmission instead of broadcasting), our scheme (Case 4) improves the performance of the conventional scheme that transmits an HR image to a big user (Case 1). This result is unexpected in the sense that, our scheme, which downsamples the original image and transmits the resultant lower-resolution one, improves the performance of the conventional scheme that transmits the original image to a user, as long as two techniques specified in Section II.B.1 and II.B.2 are employed.

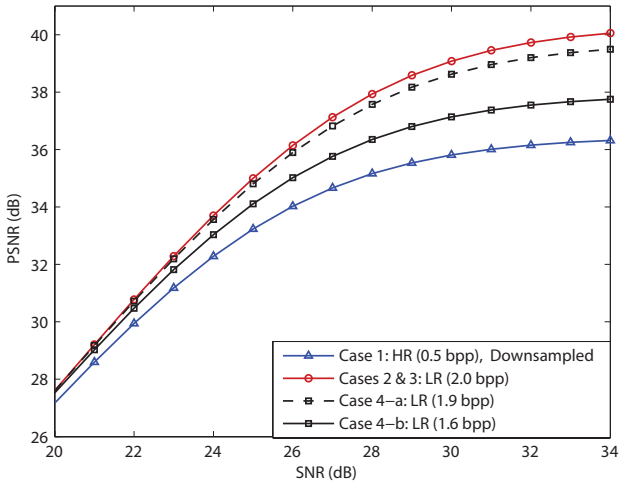
Next, we observe the PSNR qualities of the big and small users in MIMO broadcasting systems. We use QPSK constellation symbols for the proposed and baseline schemes, and the optimal maximum likelihood (ML) decoding is used for the Alamouti and Golden codes. In our evaluation, error correction coding is not considered, and the bit budget is set to 0.5×512^2 bits, which corresponds to the bit rate of 0.5 bpp for HR image, and 2.0 bpp for LR image.

The PSNR performance of the big user is shown in Fig. 4(a), where x-axis indicates the average SNR (dB) in MIMO systems, and y-axis indicates PSNR value (dB) averaged over 100 test images. Two curves of Case 4 shown in Fig. 4(a) (i.e., Cases 4-a and 4-b) are the results obtained by dividing bit budget (i.e., 0.5×512^2 bits) in two different ways: For Case 4-a, the division ratio between the LR image and the residual map is 19:1, and the division ratio for Case 4-b is 4:1. Fig. 4(a) indicates that, for a big user, the proposed scheme (Case 4) outperforms the baseline schemes that broadcast HR image (Case 1) and LR image (Case 2). These results are consistent with those shown in Fig. 2.

The PSNR performance of a small user is depicted in Fig. 4(b). As shown, Case 1 exhibits the worst performance. This is because the bit rate of HR image, which is received by the



(a) Big user



(b) Small user

Fig. 4. PSNR performance in a MIMO system.

small user and is then downsampled to yield LR image, is four times lower than that of LR image. That is, the severely compressed HR image leads to large quantization errors in the LR image. Note that Cases 2 and 3, which broadcast LR image, make no difference to the small user, since a small user does not employ SR techniques. From Fig. 4(b), we see that Case 2 (or Case 3) outperforms the proposed scheme (Cases 4-a and 4-b). This is because in Case 2 (or Case 3), bit budget is entirely used for the LR image, whereas the bit budget is divided between LR image and residual map in the proposed scheme. It is also observed that Case 4-a outperforms Case 4-b, because the former assigns more bit budget to the LR image than does the latter. In Fig. 4(b), a small user achieves the maximum PSNR value corresponding to excellent quality (about 40.0 dB) for Case 2 (or Case 3). However, a PSNR of 40.0 dB is not considered as a major performance improvement compared to the PSNR of 39.5 dB for Case 4-a, since such high PSNR values are usually perceived identically by the human visual system [5].

IV. Conclusions

Exploiting the SR techniques, the proposed scheme offers a feasible solution to the image broadcasting scenario where two types of heterogeneous users reside in the service area, with different display resolutions and different numbers of antennas. Numerical evaluation shows that, in the range of PSNRs of our interest, our scheme significantly outperforms the conventional strategy that broadcasts either HR or LR images (Fig. 4). It is also shown that even for one-to-one transmission case instead of broadcasting, our scheme, which downsamples the original image and transmits the resultant lower-resolution one, improves the performance of the conventional scheme that transmits the original image (Fig. 2).

Acknowledgment

This research was supported by the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Science, ICT & Future Planning (2017R1A2B4011148), and by Institute of Information & Communications Technology Planning & Evaluation (IITP) grant funded by the Korea government(MSIT) (No. 2017-0-01973 (Korea-Japan) International collaboration of 5G mmWave based Wireless Channel Characteristic and Performance Evaluation in High Mobility Environments).

References

[1] S.-H. Chang, H.-G. Park, J. W. Choi, and J. P. Choi, "Scalable source transmission with unequal frequency reuses in MIMO cellular networks," *IEEE Trans. Commun.*, vol. 65, pp. 4188–4204, Oct. 2017.

[2] S.-H. Chang, P. C. Cosman and L. B. Milstein, "Iterative channel decoding of FEC-based multiple description codes," *IEEE Trans. Image Process.*, vol. 21, pp. 1138–1152, Mar. 2012.

[3] J.-C. Belfiore, G. Rekaya, and E. Viterbo, "The Golden code: A 2×2 full-rate space-time code with non-vanishing determinants," *IEEE Trans. Info. Theory*, vol. 51, pp. 1432–1436, Apr. 2005.

[4] S. Park, M. Park, and M. Kang, "Super-resolution image reconstruction: A technical overview," *IEEE Signal Process. Mag.*, vol. 20, no. 3, pp. 21–36, May 2003.

[5] A. Vosoughi, S.-H. Chang, S-H. Kim, P. C. Cosman, and L. B. Milstein, "Spatially scalable video broadcasting in multiple antenna systems," in *Proc. Asilomar Conf. Signals, Syst., Comput.*, Nov. 2016.

[6] L. Zhou, B. Zheng, A. Wei, B. Geller, and J. Cui, "A robust resolution-enhancement scheme for video transmission over mobile ad-hoc networks," *IEEE Trans. Broadcast.*, vol. 54, no. 2, pp. 312–321, Jun. 2008.

[7] B. Lim, S. Son, H. Kim, S. Nah, and K. M. Lee, "Enhanced deep residual networks for single image super-resolution," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017.

[8] C. Ledig, L. Theis, F. Huszar, and J. Caballero et. al., "Photo-realistic single image super-resolution using a generative adversarial network," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017.

[9] E. Agustsson and R. Timofte, "NTIRE 2017 challenge on single image super-resolution: dataset and study," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017.

[10] D. Kingma and J. Ba, "Adam: A method for stochastic optimization. In the International Conference on Learning Representations," in *Proc. International Conf. Learning Represent. (ICLR)*, May 2015.

[11] Y. Jin and H.-J. Lee, "A block-based pass-parallel SPIHT algorithm," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 22, no. 7, pp. 1064–1075, Jul. 2012.

[12] S. Lin, A. Stefanov, and Y. Wang, "On the performance of space-time block-coded MIMO video communications," *IEEE Trans. Veh. Technol.*, vol. 56, pp. 1223–1229, May 2007.

[13] M. Shin, S.-H. Kim, L. Toni, and S.-H. Chang "Optimization of joint progressive source and channel coding for MIMO systems," in *Proc. IEEE Global Commun. Conf. (Globecom)*, Dec. 2017.