Complexity-Reduced Super Resolution for Foveation-based Driving Head Mounted Displays

HYOUNGSIK NAM, (Member, IEEE), AND HANGYEOL KANG
Department of Information Display, Kyung Hee University, Seoul 02447, Republic of Korea
Corresponding author: Hyoungsik Nam (e-mail: hyoungsiknam@khu.ac.kr).
This research was funded by Samsung Display and the Brain Korea 21 Four Program in 2021.

ABSTRACT
In this paper, we propose a foveation-based super resolution (SR) algorithm to create high resolution images from low resolution inputs for virtual reality head mounted displays. Because the proposed SR algorithm is integrated in the previous foveation-based driving technology to cover the small area around the foveation point that requires high rendering quality, the overall computational complexity is substantially reduced, compared to the whole area SR. The target display has 4 times as high resolution as the input image, therefore, the proposed SR algorithm generates 4× as well as 2× SR images at the same time. To support two SR output images, small area SR, and small number of weights, we employ cropping as well as progressive and recursive framework used in the previous MS-LapSRN. We reduce the number of neurons by placing the deconvolutional layer after convolutional layers, compared to the MS-LapSRN. PSNR and SSIM performances of the proposed SR for the 4× scale are estimated as 31.152 dB and 0.935 for Set5, 26.656 dB and 0.858 for Set14, 27.138 dB and 0.830 for BSD100, and 25.078 dB and 0.836 for Urban100. For the target 8K display of 7,680×4,320, the proposed FovSR-integrated driving technology achieves the substantial reductions by 76.7 % and 99.02 % on the number of lines from 7,680 to 1,788 and the number of neurons from 24,518,246,400 to 239,541,184, respectively.

INDEX TERMS Super resolution, low computational complexity, virtual reality, head mounted display, foveation-based driving

I. INTRODUCTION
VIRTUAL reality (VR) has attracted a lot of attention due to its increased potential in various fields such as healthcare, movie, virtual travel, professional sports, education, and gaming [1]–[3]. Because VR devices should provide users with indistinguishable and immersive experiences from the real world by computer-generated images, their displays have to address screen-door effect, motion artifacts, and latency by means of higher resolution than 60 pixels per degree (ppd), wider field of view (FoV) than 160°, and higher frame rate than 90 Hz that surpass human visual system (HVS) [4]–[7].

To reduce the latency to less than 20 ms, foveated rendering and motion predicting algorithms have been introduced. Foveated rendering methods apply different resolutions over the area based on the non-uniform distribution of receptors on the retina, leading to the speed-up of the image generation at the graphics processing unit [8] as well as the reduction of the data bandwidth in the image data transfer [9]. In addition, the deep learning algorithm has been proposed to expedite the motion tracking function by predicting in advance [10], which enables the early preparation of the next image. However, they all need high resolution and high frame rate panels of several thousands of lines, where pixels must be driven within the very short period of time of less than 1 µs. Since it is difficult to charge the pixel in such a narrow time slot, the foveation-based driving scheme has been come up with [11], [12]. Whereas previous foveated rendering technologies require the restoration of the full resolution image right before driving a panel, this foveation-based driving method recovers foveated rendering images directly on the panel by charging multiple lines of pixels at the same time. The resultant charging time has been extended to 4.18 µs, that is 476.2 % of 0.87 µs in the high performance panel of 9,600 lines and 120 Hz frame rate. Consequently, the high resolution (HR) images can be handled through the whole pipeline of VR devices.
On the other hand, there still remains one problem that is a lack of the required HR images for VR devices. In the end, available low resolution (LR) images should be scaled up to the higher resolution one, which is called a super resolution (SR) [13]–[15]. SR technologies have been covering many areas such as medical imaging [16], surveillance [17], and depth maps [18]. Recently, as the resolutions of televisions are increasing to 8K (7680×4320), SR methods are getting more interests [19]. The SR is an ill-posed problem because there are always multiple HR images corresponding to a given LR image [20]. There have been traditional approaches such as interpolation-based [21]–[23], multiple-image-based [24], and example-based methods [25]–[27]. Recently, thanks to the rapid development of deep learning algorithms (DL), many DL-based SR approaches have achieved the state-of-the-art performances in a variety of metrics such as peak signal-to-noise ratio (PSNR), structural similarity (SSIM), mean opinion score (MOS), and so on.

In the first place, SRCNN [28] is proposed as a convolutional neural network (CNN) method that consists of optimizing feature extraction, non-linear mapping, and reconstruction steps with bi-cubic interpolated input images. Then, the computational efficiency is improved by the de-convolutional layer at the end of the SR model [29], [30]. Further advancements are accomplished by increasing the depth of the network through skip-connections without gradient vanishing and exploding issues [31]–[33]. In other directions, the different network structures are studied including progressive up-sampling frameworks [34]–[36], recursive learning schemes [36]–[38], as well as generative adversarial networks [39], [40] for more stable and higher performances. However, their performance improvements are obtained by huge computational complexity such as the large numbers of layers, channels, and dimensionalities.

This paper proposes a low complexity SR framework integrated with the previous foveation-based driving scheme for VR head mounted displays (HMDs) [11], [12]. Because the display areas are manipulated with different resolutions according to the distance from the foveation point, the proposed scheme needs only to focus on the small region around the foveation point. As a result, the computational complexity can be substantially reduced, compared to the existing SR networks applied to the whole image area.

**II. PROPOSED FOVEATED SR FRAMEWORK**

When LR images are used as the inputs of the previous foveation-based driving HMD due to the lack of HR images, a separate SR block should be employed to match the image resolution to the target HR display as shown in Fig. 1(a). Since the SR target resolution is such high as several thousands of lines, SR networks require the huge computational power due to the very large number of neurons. Whereas, if the SR block is embedded in the pipeline of the foveated display as depicted in Fig. 1(b), the computational complexity can be substantially reduced because only some small areas in the image need to increase their resolutions. Consequently, the SR integrated foveation-based driving scheme directly gives rise to the vertically reduced image in the resolution by means of the much smaller number of neurons in the network. This integrated SR is named by a foveated SR (FovSR) in this paper.

It is assumed that the resolution of the target HR image is 4 times as high as the LR input image. For example, LR and HR images are 2K and 8K, respectively. Thus, the resultant foveated rendering image can contain 1×, 2×, and 4× resolutions as to the LR image, leading to the requirement of 2× and 4× SR networks separately. To support these two
networks at a time with as the small number of weights and neurons as possible, the proposed SR network adopts the progressive up-sampling framework including recursive blocks that share equivalent weights. It is constructed with the basis of a $2 \times$ SR block as illustrated in Fig. 2(a). It has been also reported that this framework provides more stable training and higher performance for large scale factors [15]. In the first place, the proposed SR model passes the LR image through the convolutional layer (Conv_i) which outputs are used to generate the $2 \times$ resolution image via the convolutional layers (Conv_SR). Then, the final HR image of 4 times as high resolution as the LR image is obtained by conducting one more Conv_SR with the outputs of the first Conv_SR. Because a part of the image needs to be processed, two cropping blocks to extract $2 \times$ and $4 \times$ areas are placed at the front of Conv_i and the second Conv_SR for the $4 \times$ SR. In addition, to transfer the enough amount of information to the SR network, the deconvolutional layer (DeConv) with multiple channels is connected to the following SR network and the final $2 \times$ and $4 \times$ images are obtained through additional convolutional layers (Conv_o2, Conv_o4). The substantial reduction on the number of weights is achieved by sharing all the weights between $2 \times$ and $4 \times$ networks marked by a dot-line box in red in Fig. 2(a). It is also possible to share the weights of Conv_o2 and Conv_o4 when the whole network is trained with respect to $2 \times$ SR and $4 \times$ SR images simultaneously. These up-sampled images are merged into one reduced resolution image for a foveated display panel as presented in Fig. 2(b).

Conv_SR is constructed by cascading $R$ recursive blocks with equivalent weights as shown in Fig. 3, which structure is similar to MS-LapSRN [36]. The repetition of a recursive block enables the reduction on the number of weights regardless of the depth. A recursive block includes $D$ distinct convolutional layers along with activation functions of leaky rectified linear units (LReLUs). For more stable convergence, the proposed Conv_SR model adds residual paths to all ends of recursive blocks as skip-connections. While MS-LapSRN adds a up-sampled residual to a up-sampled image for the generation of the SR image, the proposed one deploys a deconvolutional layer after the addition to reduce the number of neurons in Conv_SR, leading to the smaller number of neurons.

After $2 \times$ and $4 \times$ SR images are attained, these output images of the proposed FovSR are merged into one reduced resolution image with the LR image. Because two SR output images and a LR input image cannot be directly merged due to their different scaling factors, it is necessary to make their resolutions matching up to the highest one that is 4 times as large as the LR image. First of all, since the foveation-based resolution reduction scheme has the full horizontal resolution, the horizontal scaling factors should be adjusted to the target HR one. Consequently, the horizontal resolutions of LR and $2 \times$ SR images are up-sampled by factors of 4 and 2, respectively. The up-sampling is conducted by simply duplicating pixels. Whereas, the vertical resolutions are dealt
with differently according to whether the image includes higher resolution areas or not. While the 4× SR image has only one resolution, the 2× image contains the area of the 4× SR one and the LR image includes both areas of 4× and 2× SR images. Therefore, the vertical resolutions are up-sampled to the highest one of the given area. This vertical up-sampling is also executed by copying the pixels. As depicted in Fig. 4, when the foveation point is located in the center, the boundaries of 2× and 4× SR regions are calculated over a LR image as marked by red and blue boxes, respectively. Then, the vertical resolutions of the LR image areas including 4× and 2× SR images are up-sampled by factors of 4 and 2. The vertical resolutions of remaining areas are maintained. Because the 2× SR image is scaled up by the factor of 2 compared to the LR image, its corresponding area to the 4× SR image is up-sampled by the factor of 2 in the vertical direction. Finally, the reduced resolution image is directly accomplished by putting them together. If lower resolution regions than a LR image are allowed to reduce the vertical resolution more, those lower resolution areas can be simply made up by averaging the pixel data of multiple lines as proposed in the previous foveation-based driving scheme [11]. For example, the 1/2 reduction on the vertical resolution of a LR image can be accomplished by averaging pixel values of two lines.

III. EVALUATION RESULTS

The proposed FovSR model sets the number of channels in all layers to 64 excluding the convolutional layers for the RGB image reconstruction, Conv_o2 and Conv_o4, that consist of 3 channels. The kernel sizes of all convolutional layers are 3 × 3, and the kernel sizes of the de-convolutional layers are 4 × 4. Also, in order to maintain the spatial dimension over all layers, zero padding is employed. The activation function is a LReLU with a negative slope of 0.2. The mini-batch size is 64 and an Adam optimizer with $\beta_1 = 0.9, \beta_2 = 0.99$ is used. The initial learning rate is $1 \times 10^{-4}$ and reduced by a factor of 2 every 100 epochs. The training is conducted on a Tensorflow framework and a machine with 2.2 GHz Intel Xeon Silver central processing unit (CPU) (128GB RAM) and graphics processing unit (GPU), NVidia TITAN Xp (12GB Memory).

Even though the proposed SR is applied to some partial areas of the image, training and evaluation are fulfilled over the whole area without cropping. A loss function ($L$) is defined as a weighted sum of L1 losses over both 2× and 4× SR images as described in Eq. (1), where $\alpha$ is a weight, $K$ is the number of a training images, $\hat{y}_{2x}$ and $\hat{y}_{4x}$ are 2× and 4× SR images, and $y_{2x}$ and $y_{4x}$ are their ground truth target images. $\alpha$ controls how much the network training focuses on the 4× SR performance. When $\alpha$ is 1.0, the network is updated to optimize only the 4× SR performance without considering the 2× SR performance. But with $\alpha$ of 0.5, it should optimize 4× as well as 2× SR outputs with the same attention. Consequently, the 4× SR performance at $\alpha$ of 0.5 might be lower than that at $\alpha$ of 1.0. When $\alpha$ is less than 1.0, the weights of Conv_o2 and Conv_o4 are shared. However, for $\alpha$ of 1.0, the network without Conv_o2 is trained for 4× HR images and then Conv_o2 is separately optimized only for 2× HR images by freezing weights of remaining layers.

$$
L = \frac{1}{K} \left\{ (1-\alpha) \sum_{i=1}^{K} |y_{2x}^i - \hat{y}_{2x}^i| + \alpha \sum_{i=1}^{K} |y_{4x}^i - \hat{y}_{4x}^i| \right\}
$$

The ground truth 4× HR resolution images are prepared from a DIV2K dataset including the training dataset of 800 images and the validation dataset of 100 images. LR images and ground truth 2× HR images are generated from ground truth HR images by a bi-cubic interpolation function that is programmed with a resize function of scikit-image at an order of 3, an anti_aliasing of True, and a clip of False [41]. The training stage uses LR, 2× HR, and 4× HR patches cropped by the sizes of 32×32, 64×64, and 128×128, respectively. 32×32 and 64×64 patches are made up by down-sampling 128×128 HR patches. One epoch consists of 32 patches selected from each training image, totally 25,600 patches. Because 32 patches per image are randomly extracted every other epochs by the extract_patches_2d function of scikit-image, the over-fitting problem can be ameliorated due to the
TABLE 1. Performance evaluation of bi-cubic, MS-LapSRN, and FovSR at D of 5, R of 8, and α of 0.5. PSNR and SSIM are estimated for the luminance components of YCbCr-converted images and PSNR$_{RGB}$ is obtained for all RGB components.

| Model    | Scale | Set5 PSNR / SSIM / PSNR$_{RGB}$ | Set14 PSNR / SSIM / PSNR$_{RGB}$ | BSD100 PSNR / SSIM / PSNR$_{RGB}$ | Urban100 PSNR / SSIM / PSNR$_{RGB}$ |
|----------|-------|-----------------------------------|-----------------------------------|-----------------------------------|-----------------------------------|
| Bi-cubic |       | 29.826 / 0.927 / 28.219           | 27.851 / 0.887 / 26.340           | 28.396 / 0.878 / 27.066           | 25.474 / 0.856 / 24.036           |
| MS-LapSRN | 2×   | 33.228 / 0.964 / 31.455           | 30.312 / 0.921 / 28.649           | 30.129 / 0.908 / 28.776           | 27.952 / 0.906 / 26.421           |
| FovSR    |       | 33.332 / 0.966 / 31.557           | 30.447 / 0.922 / 28.768           | 30.188 / 0.908 / 28.833           | 28.140 / 0.908 / 26.597           |
| Bi-cubic |       | 27.907 / 0.887 / 26.178           | 25.475 / 0.817 / 23.866           | 25.729 / 0.797 / 24.400           | 22.879 / 0.775 / 21.445           |
| MS-LapSRN | 4×   | 31.057 / 0.933 / 29.155           | 27.578 / 0.856 / 25.798           | 27.115 / 0.829 / 25.765           | 25.025 / 0.834 / 23.507           |
| FovSR    |       | 31.077 / 0.933 / 29.182           | 27.612 / 0.857 / 25.834           | 27.128 / 0.830 / 25.782           | 25.027 / 0.834 / 23.517           |

TABLE 2. Hardware complexity comparison of MS-LapSRN and FovSR at D of 5 and R of 8.

| Model       | # Weights | # Neurons       |
|-------------|-----------|-----------------|
| MS-LapSRN   | 253,456   | 8,211,456,000   |
| FovSR       | 253,312   | 8,087,040,000   |

effectively increased number of training patches. In addition, more data augmentation schemes are applied randomly to each patch every other epochs by means of the rotations of 0°, 90°, 180°, and 270° as well as horizontal and vertical flips. Therefore, the number of possible patches used in the whole training period is increased by the factor of 8. 64 validation patches per a validation image are randomly selected once in the beginning of the training stage. Finally, the optimized model is tested over 4 datasets of Set5, Set14, BSD100, and Urban100 [15]. PSNR and SSIM values are estimated over the luminance components of YCbCr-converted images. In particular, because the proposed FovSR generates 2× and 4× SR directly from the color images with 3 channels of red, green, and blue components, PSNR performances over all RGB channels (PSNR$_{RGB}$) are also measured. All PSNR values are presented in the unit of dB.

In the FovSR model, there are three hyper-parameters, R, D, and α. R is the number of recursive blocks in Conv _SR, D is the number of convolutional layers in a recursive block, and α is a ratio of 4× SR for a loss function.

Bi-cubic interpolation and MS-LapSRN are evaluated with R of 8, D of 5, and α of 0.5 as baseline algorithms. The evaluation is conducted over the whole image area by applying the FovSR without cropping like the training stage. In addition, the MS-LapSRN has been modified to support the foveation area is of the most importance in the foveated display, the FovSR’s performance of 4× SR images has been focused. Consequently, α is set to be 1.0, that is, the loss of 2× SR images is not taken into account during the 4× SR training. Second, the performance is measured as R is reduced from 8 to 3 as presented in TABLE 4. While the best performance is achieved at R of 6, the case of R of 5 is also considered to achieve the minimum number of layers in a Conv _SR for the low computational complexity.

TABLE 3. FovSR performance comparison of 4× SR over α values of 0.1 to 1.0 at D of 5 and R of 8.

| α     | BSD100 PSNR / SSIM / PSNR$_{RGB}$ | Urban100 PSNR / SSIM / PSNR$_{RGB}$ |
|-------|-----------------------------------|-----------------------------------|
| 0.1   | 26.984 / 0.827 / 25.635           | 25.767 / 0.826 / 23.256           |
| 0.5   | 27.128 / 0.830 / 25.782           | 25.027 / 0.834 / 23.517           |
| 0.9   | 27.173 / 0.831 / 25.830           | 25.193 / 0.838 / 23.688           |
| 1.0   | 27.186 / 0.831 / 25.843           | 25.205 / 0.839 / 23.697           |

TABLE 4. FovSR performance comparison of 4× SR over R values of 8 to 3 at D of 5 and α of 1.0.

| R     | BSD100 PSNR / SSIM / PSNR$_{RGB}$ | Urban100 PSNR / SSIM / PSNR$_{RGB}$ |
|-------|-----------------------------------|-----------------------------------|
| 8     | 27.186 / 0.831 / 25.843           | 25.205 / 0.839 / 23.697           |
| 7     | 27.193 / 0.831 / 25.849           | 25.224 / 0.839 / 23.719           |
| 6     | 27.200 / 0.831 / 25.857           | 25.233 / 0.839 / 23.726           |
| 5     | 27.184 / 0.831 / 25.841           | 25.184 / 0.838 / 23.679           |
| 4     | 27.171 / 0.831 / 25.829           | 25.145 / 0.837 / 23.644           |
| 3     | 27.136 / 0.830 / 25.795           | 25.070 / 0.835 / 23.573           |

Lastly, the FovSR is simulated at 6 combinations between R of 5 to 6 and D of 3 to 5 as depicted in Fig 5. Numbers following R and D represent their values, respectively. D and R are determined to be 4 and 5, where the FovSR outperforms the MS-LapSRN of D of 5, R of 8, and α of 0.5. The resultant 2× and 4× SR images are compared in Fig 6 for ground truth, bi-cubic, MS-LapSRN, and proposed FovSR. In addition, PSNR, PSNR$_{RGB}$, and SSIM values are presented in Table 5 for 4 datasets and 2 scales of 2× and 4×. In this evaluation, 2× SR performance values have been estimated for the 1/2 down-sampled images from the original HR images by the bi-cubic interpolation. Therefore, those values of bi-cubic and MS-LapSRN are smaller than them provided in the previous paper [36] where 1/2 down-sampled images and original HR images are used as input and target images, respectively, for the 2× SR evaluation.
After cropping and FovSR model, LR input and $2 \times$ SR images are up-sampled as illustrated in Fig. 4 to be combined with $4 \times$ SR image for the construction of the vertically reduced image. The example of up-sampled LR and $2 \times$ SR images are shown in Fig. 7 along with $2 \times$ and $4 \times$ SR images as well as resultant vertically reduced image and foveated display image. The resolution of the LR image is $2K$ $(1,920 \times 1,080)$ and the FoV is $110^\circ$. $2 \times$ foveated display image. The resolution of the LR image is $2K$ $(1,920 \times 1,080)$ and the FoV is $110^\circ$. $2 \times$ and $4 \times$ regions are calculated as $\pm 592$ pixels and $\pm 252$ pixels from the foveation point at the center of the target $8K$ $(7,680 \times 4,320)$ display. Consequently, the vertical resolution is lowered to 1,788 that is $23.3 \%$ of the target HR image and the number of neurons in the FovSR are substantially reduced to 239,541,184 that is only $0.98 \%$ of 24,518,246,400 neurons required by the separate SR approach processing the whole image area.

IV. CONCLUSION

This paper demonstrates a low complexity super resolution approach focusing on the small area around the foveation point where the high rendering quality is of the importance. In the first place, the proposed FovSR crops the $2 \times$ SR area from the LR input image and increases its resolution by a factor of 2. Then, it crops the $4 \times$ SR area from the $2 \times$ SR image and then repeats the $2 \times$ SR operation to achieve the $4 \times$ SR image in the region around the foveation point. Consequently, the overall foveation-based driving technology with the FovSR has accomplished the dramatic reduction on the number of lines in the vertically reduced image and the number of neurons in the SR network. Therefore, the proposed scheme allows high performance HMD devices to be feasible in the near future by means of small data bandwidth, small latency, enough charging time, as well as high image quality.

REFERENCES

[1] J. Steuer, “Defining virtual reality: Dimensions determining telepresence,” J. Commun., vol. 42, no. 4, pp. 73–93, 1992.
[2] L. P. Berg and J. M. Vance, “Industry use of virtual reality in product design and manufacturing: a survey,” Virtual Real., vol. 21, no. 1, pp. 1–17, 2017.
[3] M. Slater and M. V. Sanchez-Vives, “Enhancing our lives with immersive virtual reality,” Front. Robot. AI, vol. 3, 2016, Art. no. 74.
[4] E. Bastug, M. Bennis, M. Méard, and M. Debbah, “Toward interconnected virtual reality: Opportunities, challenges, and enablers,” IEEE Commun. Mag., vol. 55, no. 6, pp. 110–117, 2017.
[5] B. Bastani, E. Turner, C. Vieri, H. Jiang, B. Funt, and N. Balram, “Foveated Pipeline for AR/VR Head-Mounted Displays,” Inf. Disp., vol. 33, no. 6, pp. 14–35, 2017.
[6] B. Young, “OLED Displays and the Immersive Experience,” Inf. Disp., vol. 34, no. 2, pp. 16–36, 2018.
[7] A. K. Bhowmik, “Advances in Virtual, Augmented, and Mixed Reality Technologies,” Inf. Disp., vol. 34, no. 5, pp. 18–21, 2018.
[8] A. Patney, M. Salvi, J. Kim, A. Kaplanyan, C. Wyman, N. Benty, D. Luebke, and A. Lefohn, “Towards foveated rendering for gaze-tracked virtual reality,” ACM Trans. Graph., vol. 35, no. 6, 2016, Art. no. 179.
[9] C. Vieri, G. Lee, N. Balram, S. H. Jung, J. Y. Yang, S. Y. Yoon, and I. B. Kang, “An 18 megapixel 4.3” 1443 ppi 120 Hz OLED display for wide field of view high acuity head mounted displays,” J. Soc. Inf. Disp., vol. 26, no. 5, pp. 314–324, 2018.
[10] H. Menon, A. Ramachandrapu, and J. Kesinger, “Deep-Learning Based Approaches to Visual-Inertial Odometry for Autonomous Tracking Applications,” in Proc. Soc. Inf. Disp., 2018, pp. 471–474.
[11] S. Park, Y. I. Kim, and H. Nam, “Foveation based reduced resolution driving scheme for immersive virtual reality displays,” Opt. Express, vol. 27, no. 21, pp. 29594–29605, 2019.
[12] J. Bae, J. Lee, and H. Nam, “Variable clock and em signal generation scheme for foveation-based driving oled head-mounted displays,” Electronics, vol. 10, 2021, Art. no. 538.
[13] S. C. Park, M. K. Park, and M. G. Kang, “Super-resolution image reconstruction: a technical overview,” IEEE Signal Process. Mag., vol. 20, no. 3, pp. 21–36, 2003.
[14] W. Yang, X. Zhang, Y. Tian, W. Wang, J.-H. Xue, and Q. Liao, “Deep learning for single image super-resolution: A brief review,” IEEE Trans. Multimedia, vol. 21, no. 12, pp. 3106–3121, 2019.
[15] Z. Wang, J. Chen, and S. C. Hoi, “Deep learning for image super-resolution: A survey,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 43, no. 10, pp. 3365–3387, 2021.
[16] Y. Huang, L. Sha, and A. F. Frangi, “Simultaneous Super-Resolution and Cross-Modality Synthesis in Magnetic Resonance Imaging,” in Deep Learning and Convolutional Neural Networks for Medical Imaging and Clinical Informatics. Springer, Cham, 2019, pp. 437–457.
[17] Z. Wang, J. Chen, and S. C. Hoi, “A robust face super-resolution algorithm and its application in low-resolution face recognition system,” Multimede. Tools Appl., vol. 79, pp. 23909–23934, 2020.
[18] T.-W. Hui, C. C. Loy, and X. Tang, “Depth Map Super-Resolution by Deep Multi-Scale Guidance,” in Proc. ECCV, 2016, pp. 353–369.
[19] Y. Kim, J.-S. Choi, J. Lee, and M. Kim, “A CNN-Based Multi-scale Super-Resolution Architecture on FPGA for 4K/8K UHD Applications,” in Proc. MMM. Springer, 2020, pp. 739–744.
FIGURE 6. SR quality comparison for ground truth, bi-cubic, MS-LapSRN, and proposed FovSR (a) 2 × SR (b) 4 × SR

TABLE 5. Performance (PSNR/PSNR_RGB /SSIM) comparison over bi-cubic, MS-LapSRN, and proposed FovSR for Set5, Set14, BSD100, and Urban100 datasets

| Dataset  | Scale | Bi-cubic | MS-LapSRN [36] | Proposed FovSR |
|----------|-------|----------|----------------|----------------|
|          |       | PSNR / SSIM / PSNR_RGB | PSNR / SSIM / PSNR_RGB | PSNR / SSIM / PSNR_RGB |
|          | 2x    | 29.826 / 0.927 / 28.219 | 33.228 / 0.964 / 31.455 | 31.395 / 0.946 / 29.733 |
|          | 4x    | 27.907 / 0.887 / 26.178 | 31.057 / 0.933 / 29.155 | 31.152 / 0.935 / 29.265 |
|          | 2x    | 27.851 / 0.887 / 26.340 | 30.312 / 0.921 / 28.649 | 29.024 / 0.908 / 27.474 |
|          | 4x    | 25.475 / 0.817 / 23.866 | 27.578 / 0.856 / 25.798 | 27.656 / 0.858 / 25.883 |
|          | 2x    | 28.396 / 0.878 / 27.066 | 30.129 / 0.908 / 28.776 | 29.297 / 0.899 / 27.959 |
|          | 4x    | 25.729 / 0.797 / 24.400 | 27.115 / 0.829 / 25.765 | 27.138 / 0.830 / 25.795 |
|          | 2x    | 25.474 / 0.856 / 24.036 | 27.952 / 0.906 / 26.421 | 26.432 / 0.881 / 25.980 |
|          | 4x    | 22.879 / 0.775 / 21.445 | 25.025 / 0.834 / 23.507 | 25.078 / 0.836 / 23.573 |

[20] C.-Y. Yang, C. Ma, and M.-H. Yang, “Single-image super-resolution: A benchmark,” in Proc. ECCV. Springer, 2014, pp. 372–386.
[21] R. Keys, “Cubic convolution interpolation for digital image processing,” IEEE Trans. Acoust., Speech, Signal Process., vol. 29, no. 6, pp. 1153–1160, 1981.
[22] C. E. Duchon, “Lanczos filtering in one and two dimensions,” J. Appl. Meteorol. Climatol., vol. 18, no. 8, pp. 1016–1022, 1979.
[23] M. Irani and S. Peleg, “Improving resolution by image registration,” CVGIP-Graph. Model. Image Process., vol. 53, no. 3, pp. 231–239, 1991.
[24] N. P. D. Gallego and J. Ilao, “Multiple-image super-resolution on mobile devices: an image warping approach,” J. Image Video Proc., vol. 2017, 2017, Art. no. 8.
[25] W. T. Freeman, T. R. Jones, and E. C. Pasztor, “Example-based super-resolution,” IEEE Comput. Graph. Appl., vol. 22, no. 2, pp. 56–65, 2002.
[26] J. Yang, J. Wright, T. Huang, and Y. Ma, “Image super-resolution as sparse representation of raw image patches,” in Proc. CVPR. IEEE, 2008, pp. 1–8.
[27] D. Glasner, S. Bagon, and M. Irani, “Super-resolution from a single image,” in Proc. ICCV. IEEE, 2009, pp. 349–356.
[28] C. Dong, C. C. Loy, K. He, and X. Tang, “Image super-resolution using deep convolutional networks,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 38, no. 2, pp. 295–307, 2015.
[29] C. Dong, C. C. Loy, and X. Tang, “Accelerating the super-resolution convolutional neural network,” in Proc. ECCV. Springer, 2016, pp. 391–407.
[30] W. Shi, J. Caballero, F. Huszár, J. Totz, A. P. Aitken, R. Bishop, D. Rueck-
FIGURE 7. Example of SR and merged images generated by the foveation-based driving scheme integrated with the proposed FovSR. The resolutions of images are presented at their bottoms in the form of (height \times width).

[31] J. Kim, J. Kwon Lee, and K. Mu Lee, “Accurate image super-resolution using very deep convolutional networks,” in Proc. CVPR, 2016, pp. 1646–1654.

[32] B. Lim, S. Son, H. Kim, S. Nah, and K. Mu Lee, “Enhanced deep residual networks for single image super-resolution,” in Proc. CVPR, 2017, pp. 136–144.

[33] Y. Tai, J. Yang, X. Liu, and C. Xu, “Memnet: A persistent memory network for image restoration,” in Proc. ICCV, 2017, pp. 4539–4547.

[34] Z. Wang, D. Liu, J. Yang, W. Han, and T. Huang, “Deep networks for image super-resolution with sparse prior,” in Proc. ICCV, 2015, pp. 370–378.

[35] W.-S. Lai, J.-B. Huang, N. Ahuja, and M.-H. Yang, “Deep laplacian pyramid networks for fast and accurate super-resolution,” in Proc. CVPR, 2017, pp. 624–632.

[36] ——, “Fast and accurate image super-resolution with deep laplacian pyramid networks,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 41, no. 11, pp. 2599–2613, 2018.

[37] J. Kim, J. Kwon Lee, and K. Mu Lee, “Deeply-recursive convolutional network for image super-resolution,” in Proc. CVPR, 2016, pp. 1637–1645.

[38] Y. Tai, J. Yang, and X. Liu, “Image super-resolution via deep recursive residual network,” in Proc. CVPR, 2017, pp. 3147–3155.

[39] C. Ledig, L. Theis, F. Huszár, J. Caballero, A. Cunningham, A. Acosta, A. Aitken, A. Tejani, J. Totz, Z. Wang et al., “Photo-realistic single image super-resolution using a generative adversarial network,” in Proc. CVPR, 2017, pp. 4681–4690.

[40] X. Wang, K. Yu, S. Wu, J. Gu, Y. Liu, C. Dong, C. C. Loy, Y. Qiao, and X. Tang, “EsrGAN: Enhanced super-resolution generative adversarial networks,” in Proc. ECCV, 2018, pp. 4681–4690.

[41] S. van der Walt, J. L. Schönberger, J. Nunez-Iglesias, F. Boulogne, J. D. Warner, N. Yager, E. Gouillart, and T. Yu, “Single-image super-resolution using sparse regression and natural image prior,” scikit-image: Image processing in Python, vol. 2, 2014, Art. no. e453.
HYOUNGSIK NAM (M’10) received his B.S., M.S. and Ph.D. degrees in EECS from Korea Advance Institute of Science and Technology (KAIST), Daejon Korea, in 1996, 1998, and 2004. He joined Samsung Electronics as a senior engineer in 2005, where he had worked on Active-Matrix Liquid-Crystal Displays. He is currently an associate professor in the department of Information Display at Kyung Hee University, Seoul Korea. His current research interests are low power technologies, integrated circuits, signal/user interfaces for flat panel displays, and machine learning application.

HANGYEOL KANG received the B.S. degree from the Department of Information Display, Kyung Hee University, Seoul, South Korea, in 2020. He is currently pursuing the dual M.S. degree in Kyung Hee University and Institut Polytechnique de Paris. His current research is focused on machine learning applications in various fields.

***