https://doi.org/10.2352/ISSN.2470-1173.2017.17.COIMG-429

Single Image Super-Interpolation using Adjusted Self-Exemplars

Hyun-Ho Kim, Jae-Seok Choi and Munchurl Kim
School of Electrical Engineering, Korea Advanced Institute of Science & Technology (KAIST); Republic of Korea

Abstract

Super-resolution (SR) is an elegant technique that can reconstruct high-resolution (HR) videos/images from their low-resolution (LR) counterparts. Most of the conventional SR methods utilize linear mappings to learn complex LR-to-HR relationships, where these linear mappings are often learned from training. Inspired by our previous linear mapping based SR method [1], we propose a novel super-interpolation based SR method that utilizes adjusted self-exemplars. That is, in order to find sufficient amounts of LR-HR patch pairs in self-exemplars, we iteratively augment self-exemplars from an LR input image to create additional self-exemplars. In doing so, our proposed SR method is able to find well-learned linear mappings on-line from self-exemplars without using external training images, and outperforms other conventional SR methods.

Introduction

Recent mobile phones and TVs come with ultra-high-definition (UHD) displays. Therefore, the HR contents are increasingly demanded. In addition, it often requires HR images for accurate image analysis in many fields such as medical imaging, satellite imaging and military imaging. However, it is difficult to obtain HR contents of good quality from LR ones such as images of full-high-definition (FHD) resolution. Thus, super-resolution, a sophisticated technique that can reconstruct HR videos/images from their LR counterparts, is necessarily required.

To obtain HR images, previous SR methods utilize either external dictionaries or self-exemplars in order to learn LR-to-HR relationship. Among them, some SR methods [1]-[7] used external LR-HR training image patches to learn LR-to-HR mappings. Meanwhile, as for the self-exemplar-based SR methods [8]-[10], they exploited self-similarity by generating self-exemplars, or self-dictionaries from LR inputs.

Inspired by super-interpolation (SI) [1], we take a different approach and propose a novel self-exemplar-based SR method by incorporating a self-similarity technique into SI. Also, additional self-exemplars are generated via iterative augmentation of self-exemplars for learning better LR-to-HR mappings. Experiment results show that the proposed SR methods can reconstruct HR images of higher quality compared to external dictionary-based linear mapping SR methods [1].

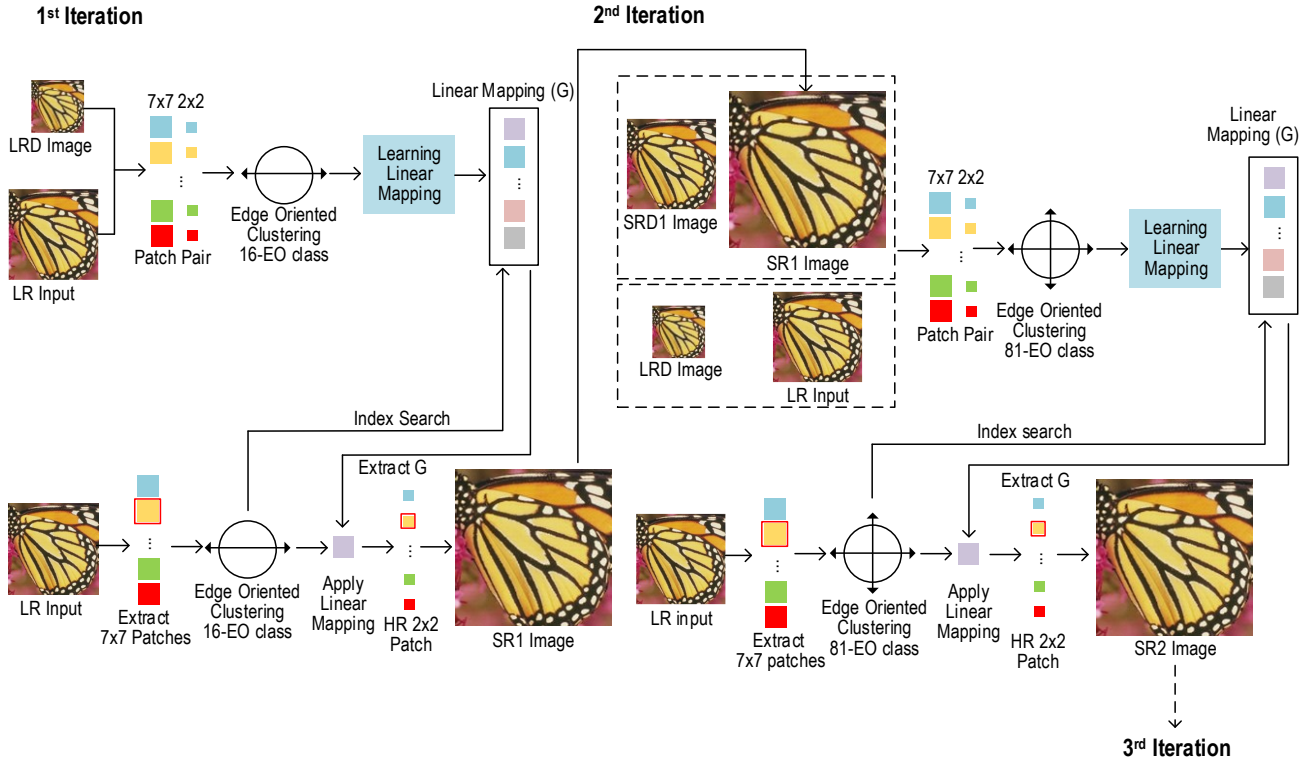

Figure 1. *LR-to-HR conversion of our SR method using on-line learning-based linear mapping with augmentation of self-dictionary exemplars.*

Proposed Method

In order to create self-dictionaries from an LR input image, bicubic interpolation is used to down-scale the LR input image by a scaling ratio of 1/2. This down-scaled image is used as an LR self-dictionary (LRD). Since we often lack of training exemplars in one LRD, we further increase the amount of the training exemplars by applying data augmentation (flipping, rotating and down-sizing) to LR and obtain additional LR-LRD image pairs. Here, 7×7 LRD patches and the corresponding 2×2 LR patches are used to learn LR-to-HR mappings. For each 7×7 LRD patch, we first obtain a 3×3 center patch, which is then divided into four 2×2 sub-patches as

$$Y_{3\times3}=\begin{bmatrix} l_{1,1} & l_{1,2} & l_{1,3} \\ l_{2,1} & l_{2,2} & l_{2,3} \\ l_{3,1} & l_{3,2} & l_{3,3} \end{bmatrix},$$
$$L_1=\begin{bmatrix} l_{1,1} & l_{1,2} \\ l_{2,1} & l_{2,2} \end{bmatrix},\quad L_2=\begin{bmatrix} l_{1,2} & l_{1,3} \\ l_{2,2} & l_{2,3} \end{bmatrix}, \tag{1}$$
$$L_3=\begin{bmatrix} l_{2,1} & l_{2,2} \\ l_{3,1} & l_{3,2} \end{bmatrix} \text{ and } L_4=\begin{bmatrix} l_{2,2} & l_{2,3} \\ l_{3,2} & l_{3,3} \end{bmatrix},$$

where $Y_{3\times3}$ is the 3×3 center patch, and L_{1-4} are the four 2×2 sub-patches. Only 3×3 center patch within each 7×7 patch is used to avoid computation complexity where the 3×3 patch size is empirically found as a good compromise between overall complexity and mapping accuracy.

Next, prior to clustering for learning linear mappings, the gradients of 4 sub-patches for each LRD patch are used to determine a class number for the current LRD patch. Two gradient operators are used and defined as

$$h_x=\begin{bmatrix} 1 & -1 \\ 0 & 0 \end{bmatrix} \text{and } h_y=\begin{bmatrix} 1 & 0 \\ -1 & 0 \end{bmatrix}, \tag{2}$$

where h_x and h_y are a horizontal and vertical gradient operator respectively. For example, for the sub-patch L_1, the gradient operators are applied as

$$\begin{aligned} &g_{x,L1}=<h_x,L_1>=l_{1,1}-l_{1,2} \\ &\text{and } g_{y,L1}=<h_y,L_1>=l_{1,1}-l_{2,1} \end{aligned}, \tag{3}$$

where $g_{x,L1}$ and $g_{y,L1}$ are filtered sub-patches, which are obtained by applying two gradient operators to the sub-patch L_1, and $<,>$ indicates an inner product operator. From this, we can obtain the magnitude and angle of gradients for each sub-patch as

$$\begin{aligned} &M_{L1}=\sqrt{g_{x,L1}^2+g_{y,L1}^2} \\ &\text{and } P_{L1}=\tan^{-1}\left(\frac{g_{x,L1}}{g_{y,L1}}\right) \end{aligned}, \tag{4}$$

where M_{L1} and P_{L1} are a magnitude and an angle of sub-patch L_1, respectively. Similarly, the gradient operators are applied to the other three sub-patches L_{2-4} as well. Next, if the magnitude of the sub-patch gradient is smaller than a predefined threshold, this sub-patch is categorized as a flat region and index 0 is assigned. Otherwise, the sub-patch is regarded as an edge region and index 1 is assigned. Because there are 4 sub-patches for each 7×7 LRD patch, one of 16 (= $(1+1)^4=2^4$) class indexes can be assigned to each 7×7 LRD patch. This process of assigning a class index to each LRD patch will be referred to as classification.

All the LRD-LR patch pairs are grouped into K (=16) clusters, based on the class indexes of the LRD patches. Here, for the kth LRD-LR cluster, one linear mapping can be learnt using ridge regression. First, we vectorize the 7×7 LRD patches of the kth class (L_0^k) and their corresponding 2×2 LR patches, H_0^k . The kth linear mapping, G_0^k , can be obtained as

$$H_0^k=G_0^kL_0^k. \tag{5}$$

Therefore, we need to solve the following optimization problem:

$$G_0^k=\arg\min_{G_0^k}\left\|H_0^k-G_0^kL_0^k\right\|_2^2+\lambda\left\|G_0^k\right\|_F^2, \tag{6}$$

which minimizes the reconstruction error between the original LR patches (H_0^k) and the reconstructed patches ($G_0^kL_0^k$). This L_2 norm minimization problem has a closed form solution as

$$G_0^k=H_0^kL_0^{kT}(L_0^kL_0^{kT}+\lambda I)^{-1}, \tag{7}$$

which becomes the kth linear mapping. By repeating this process for all 16 clusters, we can obtain 16 linear mappings for LR-to-HR conversion.

Next step is to up-scale the LR input image using the learnt linear mappings. We decompose the LR input image into 7×7 patches and perform classification for them. By applying appropriate linear mappings to these 7x7 patches based on their class indexes from classification, we can obtain an up-scaled image SR_1. However, because we utilize only 16 edge directions for classification and an insufficient amount of exemplars, the quality of the SR_1 image may not be satisfactory. To overcome this limitation, we repeat the SR procedure for another 2 more iterations by generating another dictionary pairs.

In the 2nd iteration, we create another self-dictionary pair, SRD_1-SR_1, by applying bicubic interpolation to the SR_1 image with a scaling ratio of 1/2. And now we repeat the SR procedure with LRD-LR and SRD_1-SR_1 pairs. In contrast to the first iteration where the edge magnitude of a sub-patch gradient was only considered, we further utilize edge angles in the 2nd iteration, which are quantized with angle step-size of 90°. Note that the two opposite directions are considered the same direction (e.g. 45° = 225°). Here, index 0 is assigned to sub-patches that are considered a flat region. Index 1 is assigned to sub-patches whose edge angle is between 0° and 90°, and index 2 is assigned to those within 90° to 180°. Thus, each sub-patch can now hold one of 3 indexes, according to its edge direction and magnitude. As a result, one of 81 (= $(2+1)^4=3^4$) class indexes can be assigned to each 7x7 patch, and from this, we can obtain 81 linear mappings G_1. Finally, by applying G_1 to the LR input image, we can obtain an up-scaled image SR_2. The quality of the SR_2 image is higher than that of the previous SR_1 image, but its quality is still not

Table1. Experimental Results (Live Data set)

Images	Bicubic	ANR [10]	SI-3 [1]	Ours	Ours (Ensemble)
12003	31.14	32.69	32.77	32.97	**33.16**
118035	31.89	34.17	35.03	35.50	**35.78**
189080	35.72	38.07	38.36	38.41	**38.69**
253027	25.34	26.92	26.97	27.24	**27.34**
Coinsinfountain	29.90	31.41	31.14	31.22	**31.25**
Lena	35.46	37.37	37.26	37.39	**37.52**
Parrots	36.03	38.22	38.09	38.03	**38.30**
Womanhat	34.10	35.48	35.40	35.56	**35.62**
Average	32.45	34.29	34.38	34.54	**34.71**

sufficient. Therefore, we again generate another self-dictionary pair, SRD_2-SR_2, by applying bicubic interpolation to the SR_2 image with a scaling ratio of 1/2.

In the final iteration, we repeat the previous steps using the SRD_2-SR_2 pair and the LRD-LR pair. In this case, edge directions are quantized as an angle step-size of 60°. Here, index 0 is assigned to sub- index 2 is assigned to those within 60° to 120°, and index 3 is assigned to those within 120° to 180°. So each sub-patch can patches that are considered as flat regions. Index 1 is assigned to sub-patches whose edge angles are between 0° and 60°, now hold one of 4 indexes, according to its edge direction and magnitude. Therefore, we can obtain linear mappings G_2, which has 256 (= $(3 + 1)^4$ = 4^4) classes. By applying G_2 to the LR input image, we can reconstruct the final HR image, whose quality is superior to the previous SR_1 image and SR_2 image in terms of PSNR values. Figure 1 shows the overall SR process of the proposed method.

In order to improve PSNR performance, we present another variant of our proposed method that utilizes an ensemble method. Here, we use new gradient operators as

$$h_{x2} = \begin{bmatrix} 2 & -1 \\ -1 & 0 \end{bmatrix} \text{ and } h_{y2} = \begin{bmatrix} -1 & 0 \\ 2 & -1 \end{bmatrix}, \quad (8)$$

where h_{x2} and h_{y2} are a horizontal and vertical gradient operator, respectively, and these operator were obtained empirically. We repeat another SR procedure by using these gradient operators to generate another result image. Next, we average two result images, which are generated by using two different gradient operator sets (2) and (8). Figure 2 shows the comparison between two proposed methods with and without ensemble.

Experimental Results

We conducted various experiments using images in LIVE database [11] for test images. Experiment results show that the proposed method outperforms other conventional SR

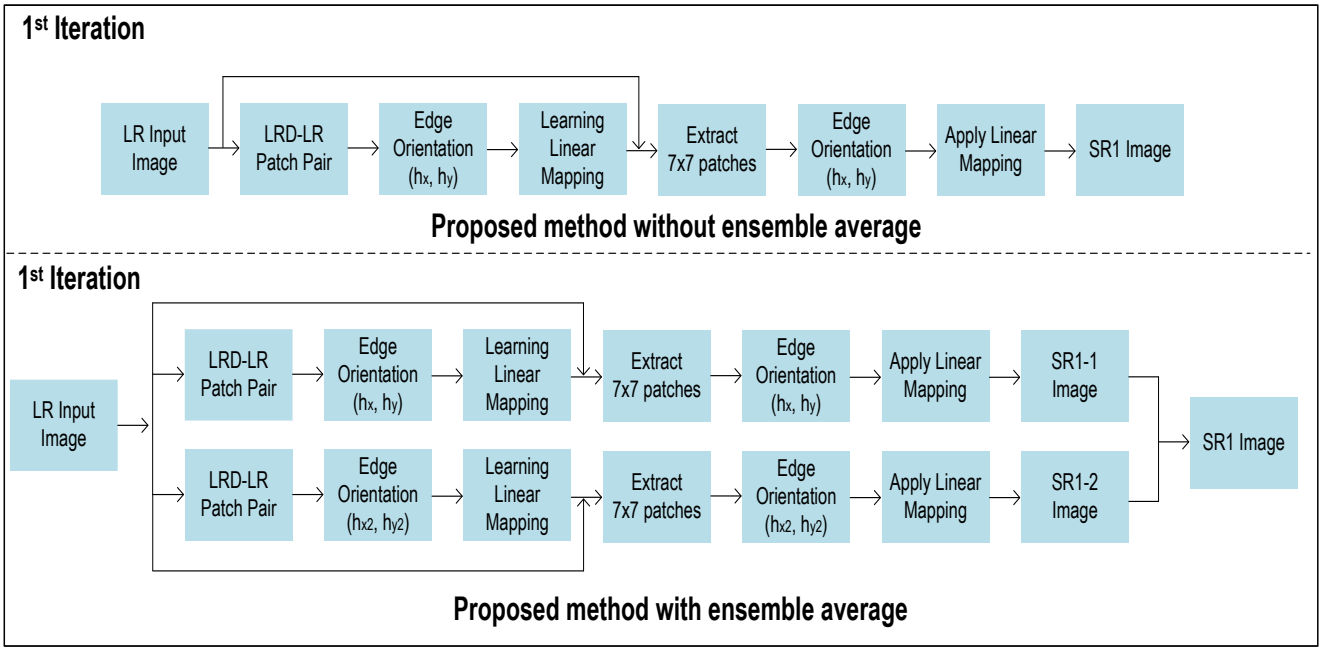

Figure 2. *LR-to-HR conversion of our SR method using ensemble method*

Figure 3. Comparison of HR output images (× 2 on 118035)

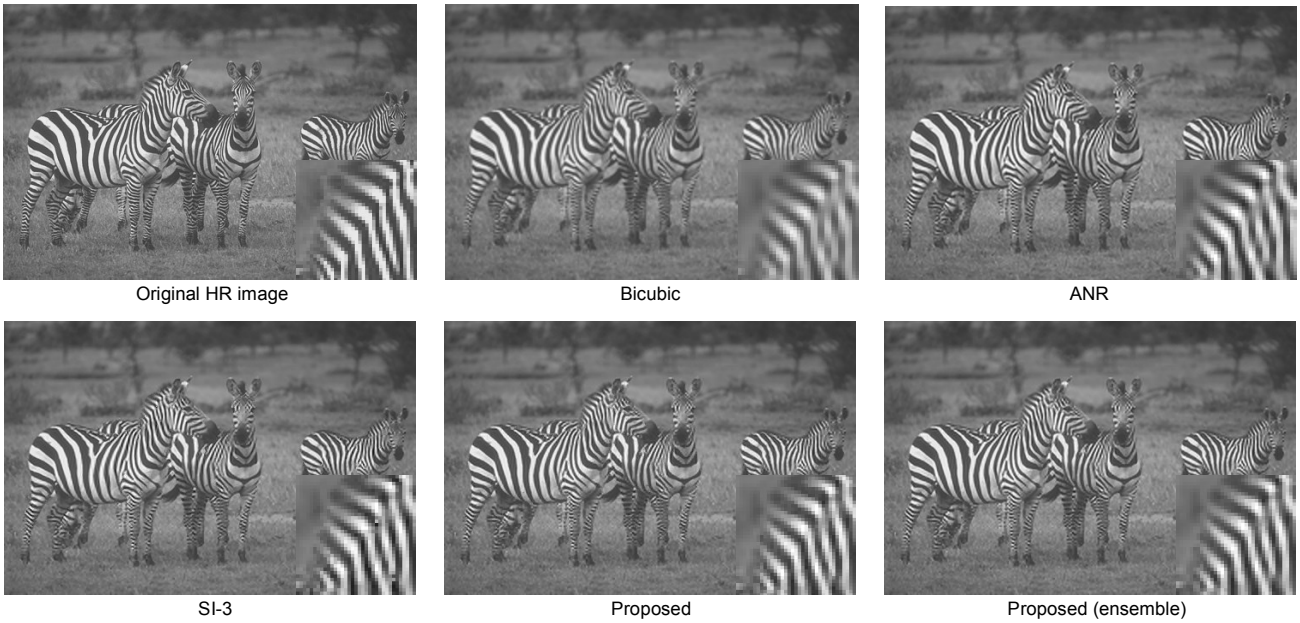

Figure 4. Comparison of HR output images (× 2 on 253027)

methods in terms of PSNR. Table 1 shows PSNR values of the reconstructed HR images using various SR methods for the Live Data set. The average PSNR for our proposed method without ensemble was 34.54 dB, while PSNR for the proposed method with ensemble was 34.71 dB. Compared to bicubic interpolation, our method reconstructs HR images of average 2.26 dB higher PSNR values. Compared to anchored neighborhood regression (ANR) method [13], our method reconstructs HR images of average 0.42 dB higher PSNR values. Compared to SI [1], our method reconstructs HR images of average 0.33 dB higher PSNR values. Figures 3-5 show the visual comparison for the reconstructed images by our method and other SR methods. As shown in Figure 3, it can be observed that the proposed method can reconstruct HR images with sharper edges than the SI-3 method [1]. Also, the ringing artifacts in the HR image of SI-3 are not visible in the HR image of the proposed method. Note that while the conventional SR methods require many external training images to learn linear mappings, our proposed method can learn better linear mappings by only using self-exemplars from LR input images.

Figure 5. Comparison of HR output images (× 2 on Lena)

Conclusion

In this paper, we proposed a novel self-exemplar-based SR method by incorporating self-similarity technique. In doing so, our proposed SR method is able to learn well-trained linear mappings compared to SI [1]. As a result, the proposed method reconstructed HR images of higher PSNR with reduced ringing artifacts and noise. For future work, our SR method may further be improved by incorporating other sophisticated clustering algorithms to obtain better linear mappings, instead of using EO-based clustering.

Acknowledgment

This work was supported by the National Research Foundation (NRF) of Korea grant funded by the Korea government (No.2014R1A2A2A01006642).

References

[1] J.-S. Choi and M. Kim, "Super-interpolation with edge-orientation-based mapping kernels for low complex upscaling," IEEE Transactions on Image Processing, vol. 25, no. 1, pp. 469–483, 2016.

[2] M. Bevilacqua, A. Roumy, C. Guillemot, and M.-L. A. Morel, "Single-image super-resolution via linear mapping of interpolated self examples," IEEE Trans. Image Process., vol. 23, no. 12, pp. 5334–5347, Dec. 2014.

[3] J. Yang, J. Wright, T. S. Huang, and Y. Ma, "Image super-resolution via sparse representation," IEEE Trans. Image Process., vol. 19, no. 11, pp. 2861–2873, Nov. 2010.

[4] J.-S. Choi, S.-H. Bae, and M. Kim, "A no-reference perceptual blurriness metric based fast super-resolution of still pictures using sparse representation," Proc. SPIE, vol. 9401, p. 94010N, Mar. 2015.

[5] J. Kim, J. K. Lee, and K. M. Lee, "Accurate image super-resolution using very deep convolutional networks," in Proc. IEEE Conf. Comp. Vis. Patt. Recogn., 2016.

[6] C. Dong, C. C. Loy, and X. Tang. Accelerating the super-resolution convolutional neural network. In European Conference on Computer Vision (ECCV), pages 391–407. Springer International Publishing, 2016.

[7] C. Dong, C. C. Loy, K. He, and X. Tang, "Learning a deep convolutional network for image super-resolution," in Proc. 13th Eur. Conf. Comput. Vis., Zurich, Switzerland, Sep. 2014, pp. 184–199.

[8] J. B. Huang, A. Singh, and N. Ahuja, "Single image super-resolution from transformed self-exemplars," in Proc. IEEE Conf. Comput. Vis. Pattern Recog., 2015, pp. 5197–5206.

[9] M. Ebrahimi and E. R. Vrscay. Solving the inverse problem of image zooming using self-examples. In Image analysis and Recognition, 2007.

[10] D. Glasner, S. Bagon, and M. Irani. Super-resolution from a single image. In ICCV, 2009.

[11] H. R. Sheikh et al. (2005). LIVE Image Quality Assessment Database 2. [Online]. Available: http://live.ece.utexas.edu/research/quality

[12] R. Keys, "Cubic convolution interpolation for digital image processing," IEEE Trans. Acoust., Speech, Signal Process., vol. 29, no. 6, pp. 1153–1160, Dec. 1981.

[13] R. Timofte, V. De, and L. Van Gool, "Anchored neighborhood regression for fast example-based super-resolution," in Proc. IEEE Int. Conf. Comput. Vis., Sydney, NSW, Australia, Dec. 2013, pp. 1920–1927.

Author Biography

Hyun-Ho Kim received the B.S. degree in electrical engineering from Kyunghee University, Suwon, South Korea, in 2016. He is currently pursuing the MS degree in electrical engineering at Korea Advanced Institute of Science and Technology (KAIST). His research interests include image processing, pattern recognition and neural networks.

Jae-Seok Choi received the B.S. degree in electrical engineering from Hanyang University, Seoul, South Korea, in 2014, and the M.S. degree in electrical engineering from KAIST, Daejeon, South Korea, in 2016. He is currently pursuing the Ph.D. degree in electrical engineering at KAIST. His research interests include super-resolution, deep learning and image processing.

Munchurl Kim received the B.E. degree in electronics from Kyungpook National University, Daegu, Korea, in 1989, and the M.E. and Ph.D. degrees in Electrical and Computer Engineering from the University of Florida, Gainesville, in 1992 and 1996, respectively. He is a Full Professor of the School of Electrical Engineering, KAIST. His current research interests include high performance video coding, perceptual video coding, visual quality assessments on 3-D/UHD video, computational photography, machine learning and pattern recognition.