IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received January 9, 2018, accepted February 23, 2018, date of publication March 6, 2018, date of current version May 9, 2018.
Digital Object Identifier 10.1109/ACCESS.2018.2812809

Dual Autoencoder Network for Retinex-Based Low-Light Image Enhancement

SEONHEE PARK, (Student Member, IEEE), SOOHWAN YU, (Student Member, IEEE), MINSEO KIM, KWANWOO PARK, AND JOONKI PAIK, (Senior Member, IEEE)
Department of Image Graduate School of Advanced Imaging Science, Multimedia and Film, Chung-Ang University, Seoul 06974, South Korea
Corresponding author: Joonki Paik (paikj@cau.ac.kr)

This work was supported in part by the Institute for Information & Communications Technology Promotion grant through the Korea government under Grant 2017-0-00250 and in part by the Intelligent Defense Boundary Surveillance Technology Using Collaborative Reinforced Learning of Embedded Edge Camera and Image Analysis.

ABSTRACT This paper presents a dual autoencoder network model based on the retinex theory to perform the low-light enhancement and noise reduction by combining the stacked and convolutional autoencoders. The proposed method first estimates the spatially smooth illumination component which is brighter than an input low-light image using a stacked autoencoder with a small number of hidden units. Next, we use a convolutional autoencoder which deals with 2-D image information to reduce the amplified noise in the brightness enhancement process. We analyzed and compared roles of the stacked and convolutional autoencoders with the constraint terms of the variational retinex model. In the experiments, we demonstrate the performance of the proposed algorithm by comparing with the state-of-the-art existing low-light and contrast enhancement methods.

INDEX TERMS Autoencoder, image processing, image enhancement, neural networks, variational retinex model, unsupervised learning.

I. INTRODUCTION

Digital cameras play a role in sensing information from external world in various applications such as artificial intelligence, remote sensing, surveillance system, and advanced driver assistance system. However, when the amount of incoming light to the sensor is insufficient under poor illumination conditions, the dynamic range of the acquired image is reduced. In addition, the low-light images are corrupted by additive noise because of the limited number of photons received by each pixel. As a result, it is difficult to obtain a high-quality image under the low-light condition and the low-light artifacts may reduce the performance of computer vision applications such as object recognition, detection, and tracking. A theoretically sound approach to solve this problem is image enhancement based on the retinex theory with the understanding of the human visual system (HVS). More specifically, Land *et al.* first proposed the retinex theory to demonstrate the process of the HVS to perceive colors from the retina to visual cortex [1], [2]. They demonstrated that the HVS perceives the colors by the reflected ratios of the light rather than the lightness.

The retinex theory-based image enhancement methods have been further developed to improve the dynamic range of the dark region. These methods enhance the visibility by subtracting the local and global illumination components, but the separation of the illumination and reflectance components is an ill-posed problem. Jobson *et al.* [3] defined the reflectance component as a ratio of the intensity value at the center to the average of the intensity values. They proposed single-scale retinex (SSR) to enhance the dynamic range by eliminating the illumination component, which is estimated by Gaussian low-pass filtering. However, the resulting image shows halo effect near edges because of the continuity of the illumination component. To solve this problem, Rahman *et al.* [4] presented multi-scale retinex (MSR) using multiple Gaussian kernels with different standard deviations. Although the MSR algorithm can suppress the halo effect using the weighted summation of multiple illumination components, the resulting image cannot avoid color distortion. Jobson *et al*. extended their previous work to compensate the color component by applying the color restoration function using the ratio of each color channel [5].

As an alternative approach to enhance the low-light image, variational retinex models were proposed based on priors of illumination and reflectance components [6]–[12]. Kimmel *et al.* [6] estimated the illumination component by minimizing the gradient of illumination component using l_2 regularization. Ma *et al.* [7] estimated the reflectance component which contains high-frequency components using anisotropic total variation (TV)-prior instead of the Gaussian smoothness prior to reduce the amplified noise while preserving the edge. Fu *et al.* [8] demonstrated the relationship between the bright channel and illumination component using the bright channel prior, and presented a variational retinex model which can suppress halo-effect. Park *et al.* [9] penalized the brightness of illumination using quadratic fidelity prior of the illumination component with respect to its enhanced version to suppress over-enhancement of the reflectance component.

Recently, deep learning-based image processing methods have been proposed in the image enhancement field. Lore *et al.* [13] adopted stacked-sparse denoising autoencoder to a low-light image enhancement framework to simultaneously perform the brightness enhancement and noise removal. Shen *et al.* [14] analyzed the property of MSR algorithm in the sense of convolutional neural network (CNN), and proposed MSR-net to enhance the low-light image using the CNN architecture.

In this paper, we present a novel low-light image enhancement framework combining the stacked and convolutional autoencoders based on the retinex theory. In addition, we analyze the relationship between the variational retinex model and the proposed autoencoder-based enhancement methods. Major contribution of the proposed method is twofold: i) Since the stacked autoencoder decomposes the low-light input patch into compact features, it can reconstruct the optimal illumination component with enhanced brightness, and ii) the convolutional autoencoder plays a role in suppressing noise amplification in the reflectance component without degrading sharp edges. In addition, the stacked autoencoder can be regarded as the smoothness term or brightness constraint on the illumination component. In the same manner, the convolutional autoencoder enforces the penalty on the reflectance component to reduce the amplified noise.

The paper is organized as follows. Section II describes the variational retinex model and low-light net as a theoretical background. Section III presents the proposed low-light image enhancement framework using dual autoencoder. Experimental results are shown in Section IV and Section V concludes the paper.

II. RELATED WORKS

A. VARIATIONAL RETINEX MODEL

To enhance the contrast of a low-light image, Kimmel *et al.* [6] estimated the illumination component by minimizing the energy functional in the iterative manner. However, since the illumination and reflectance components are inversely proportional, the low illumination component results in the over-enhancement of the reflectance component. To overcome this problem, they adjusted the amount of gamma correction to enhance the brightness of the estimated illumination component. The resulting image is obtained by multiplying the reflectance and adjusted illumination components.

To prevent the over-enhancement of the reflectance component, Park *et al.* [9] presented a modified variational retinex model using an additional data-fidelity term which penalizes the brightness of the illumination component as

$$\arg\min_{f_R, f_L} \lambda_1 \|f_R f_L - g\|_2^2 + \lambda_2 \|\nabla f_L\|_2^2 + \lambda_3 \|\nabla f_R\|_1 + \lambda_4 \|f_L - \hat{f}_L\|_2^2, \quad (1)$$

where $\|f_R f_L - g\|_2^2$ represents the data-fidelity term, $\|\nabla f_L\|_2^2$ and $\|\nabla f_R\|_1$ the smoothness terms on illumination and reflectance components, respectively, $\|f_L - \hat{f}_L\|_2^2$ the data-fidelity term between illumination component and its enhanced version by the gamma correction. λ_1, λ_2, λ_3, and λ_4 represent regularization parameters.

In Park's method, the smoothness term on the illumination component, $\|\nabla f_L\|_2^2$, penalizes the gradient of illumination component using an isotropic TV-prior which is equivalent to the Gaussian smoothness prior. In addition, since the brightness of illumination component is iteratively increased using the fourth quadratic data-fidelity term in (1), this method can estimate the natural reflectance component without gamma correction on the illumination component. Moreover, this method can reduce noise by minimizing the loss of edge information using an anisotropic TV-prior on the reflectance component.

B. LEARNING-BASED IMAGE ENHANCEMENT METHODS

An autoencoder is a feed-forward neural network which aims to extract meaningful features by compressing the input data in an unsupervised manner [15]–[19]. Vincent *et al.* [17] performed noise removal using stacked denoising autoencoder and pairs of original and noise corrupted vectors. This method trains the autoencoder to reconstruct the output vector as close to as the original vector in a self-supervised manner from a compressed representation of noise vector.

Motivated by the denoising autoencoder, Lore *et al.* [13] presented low-light net (LLNet) to enhance the low-light image using the stacked denoising autoencoder with the sparsity prior because the low-dimensional representations provide more compact and meaningful features. To simultaneously perform contrast enhancement and noise removal, Lore *et al.* generated a pair of ground-truth and corrupted images as the training data. The corrupted images were produced by degrading the brightness of ground-truth image using gamma correction and adding white Gaussian noise. The weights and biases of the autoencoder were updated in the back-propagation step to minimize the error between ground-truth and reconstructed images. Moreover, they learned each denoising and contrast enhancement network separately, and presented staged LLNet to independently control the performance of each image enhancement module.

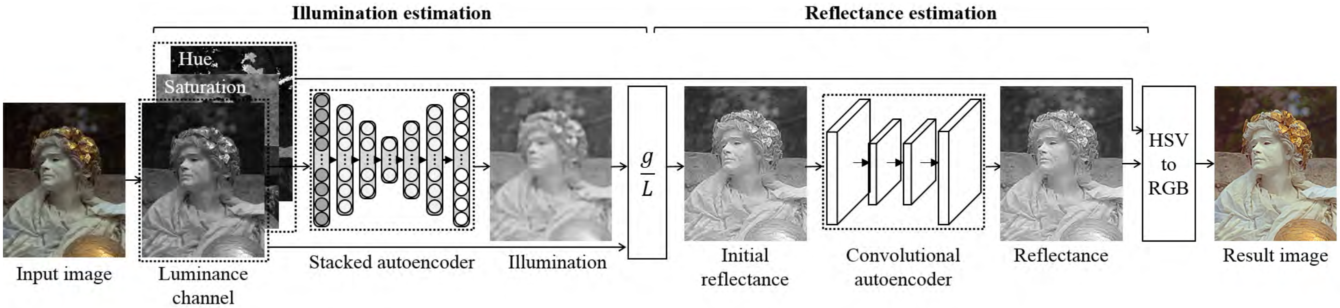

FIGURE 1. The proposed low-light image enhancement framework.

III. DUAL AUTOENCODER NETWORK

This section describes the proposed low-light image enhancement framework using stacked and convolutional autoencoders. The proposed dual autoencoder model estimates the enhanced image in three steps: i) estimation of illumination component using a stacked autoencoder, ii) initial estimation of reflectance component, and iii) refinement of reflectance component using a convolutional autoencoder. In the following subsections, we describe the roles of stacked and convolutional autoencoders.

In addition, we analyze the relationship between the proposed network architecture and each constraint term of modified variational retinex model in [12]. Fig. 1 shows a block-diagram of the proposed low-light image enhancement framework.

A. ILLUMINATION ESTIMATION USING STACKED AUTOENCODER

Based on the retinex decomposition model, the illumination component has a low-frequency characteristic since it smoothly changes in the image space. For that reason, a Gaussian low-pass filter was the most popular in existing illumination estimation methods. On the other hand, the proposed method estimates the enhanced illumination component using the stacked autoencoder which reduces the dimensionality of the input data.

A conventional denoising autoencoder plays a role in estimating the original data given its corrupted version. In the training step of the neural network, if we use a very small number of hidden units, the encoder layer provides more compressed representation of the input data. It implies that the reconstruction of original data cannot be successful. Fig. 2 shows the comparative results of the reconstruction of brightness enhanced illumination given the input low-light image.

As shown in Fig. 2, the neural network with smaller number of hidden units cannot successfully preserve the image structure such as edges and textures. The poor performance in preserving edges implies that the resulting image cannot be consistent with the original image in terms of noise reduction. However, the blurred result of the autoencoder can be used as the illumination component because it has the low-frequency property over the entire image as shown in Fig. 2. In addition, since the proposed method trains the neural network using a pair of low- and high-contrast image patches, the brightness enhanced illumination component can be obtained. In the third, fifth, and seventh columns of Fig. 2, row (d) has better reconstructed edges and details because of more hidden units in the bottleneck layer than row (b) and (c).

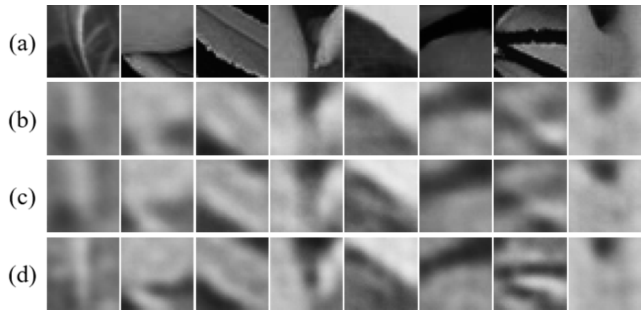

FIGURE 2. Comparative results of the reconstruction of brightness enhanced illumination patches: (a) corrupted input patches, (b) reconstructed patches using 256, 128, and 32 hidden units in each hidden layer, (c) reconstructed patches using 256, 128, and 64 hidden units in each hidden layer, and (d) reconstructed patches using 256, 128, and 96 hidden units in each hidden layer.

B. REFLECTANCE ESTIMATION USING CONVOLUTIONAL AUTOENCODER

We can estimate the reflectance component using the ratio of the input image to the illumination component estimated in the previous subsection. Given the estimated illumination component in the previous step, the initial reflectance component is estimated as

$$R = \frac{g}{L}, \tag{2}$$

where R represents an initially estimated reflectance component, g the input low-light image, and L the estimated illumination component.

However, since the noise components are amplified during the contrast enhancement process, the noise removal of reflectance component is needed to provide a high-quality image. Although the stacked denoising autoencoder can perform noise removal to a certain degree, it ignores the

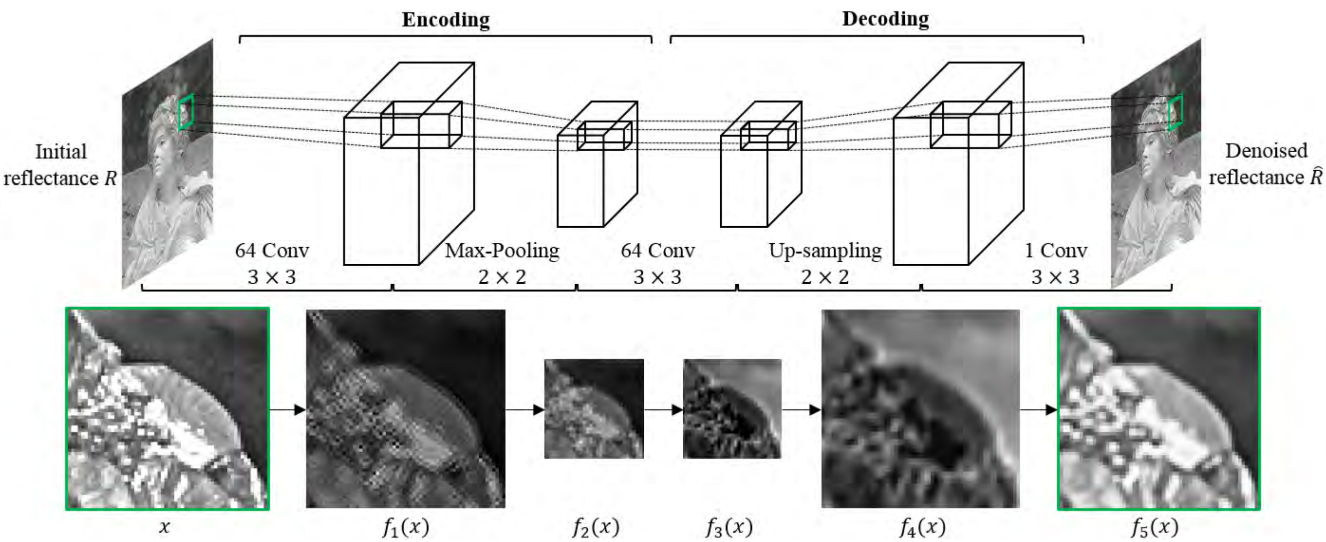

FIGURE 3. **The proposed convolutional autoencoder model.**

two-dimensional structural information of image since the input low-light patch was transformed into one-dimensional vector. On the contrary, the convolutional autoencoder can use the two-dimensional structural information for training. So it results in reducing the loss of details [18], [19]. For that reason, the proposed method can estimate improved reflectance component using convolutional autoencoder to reduce noise amplification while preserving the edge. Fig. 3 shows the proposed fully convolutional autoencoder network model.

The encoding part consists of one convolution layer and one pooling layer, all of which has one activation function. The first convolution layer is represented as

$$f_1(x) = \max(0, w_1 * x + b_1), \tag{3}$$

where x represents the input image and $*$ the convolution operation. w_k and b_k represent the weight and bias of the k-th convolution layer, respectively. The extracted features are compressed by down-sampling the output of convolution layer using the max-pooling as

$$f_2(x) = p(f_1(x)), \tag{4}$$

where $p(\cdot)$ represents the max-pooling operation, which decreases the dimensionality of input data and extracts effective features by taking the maximum activation.

The decoding part consists of one up-sampling and two convolution layers. The second convolution layer is represented as

$$f_3(x) = \max(0, w_2 * f_2(x) + b_2). \tag{5}$$

This layer takes compact and coded features to reconstruct the input image. The up-sampling process can be represented as

$$f_4(x) = u(f_3(x)), \tag{6}$$

where $u(\cdot)$ represents the up-sampling operation. The up-sampling layer plays a role in enlarging the spatial resolution of $f_3(x)$ to have the same dimension of the input image x. Finally, the reconstruction of resulting image is performed in the third convolution layer as

$$f_5(x) = \sigma(w_3 * f_4(x) + b_3), \tag{7}$$

where $\sigma(\cdot)$ represents the sigmoid function. Since the last convolution layer provides the reconstructed resulting image in [0,1], we used the sigmoid function as activation function than rectified linear unit (ReLU).

The weight and bias of each convolution layer are updated to minimize the difference between the reconstructed and its ground-truth images. Mean squared error (MSE) is used to define the loss function as

$$L(\theta) = \frac{1}{n}\sum_{i=1}^{n} \|f(x_i; \theta) - y_i\|^2, \tag{8}$$

where $\theta = \{w_1, w_2, w_3, b_1, b_2, b_3\}$, n represents the number of training data, and y_i the i-th ground-truth image of x_i.

C. MOTIVATION FROM THE VARIATIONAL RETINEX MODEL

In low-light condition, since the reduced amount of incoming light results in the low illumination component, it is highly probable to obtain the over-enhanced reflectance component because the estimated reflectance is inversely proportional to the illumination component. In this subsection, we compare the proposed dual autoencoder network model with the constraint terms of variational retinex model to prevent the over-enhancement of reflectance component [12].

The proposed method estimates the illumination component from the extremely compressed representation of the input low-light image explained in subsection 3.A. Since the proposed stacked autoencoder is trained to provide the blurred illumination component using the small number of hidden units, it can be regarded as the smoothness prior, $\|\nabla f_L\|_2^2$, of the illumination component in the variational retinex model in (1). In addition, the proposed stacked autoencoder learns the mapping between the low- and high-contrast patches to obtain the illumination component of brightness enhanced image. It implies that the proposed stacked autoencoder enforces the data-fidelity between the illumination component of an input low-light image and that of its enhanced version as the fourth term in (1).

In terms of noise reduction, the proposed method reduces the amplified noise in the enhanced reflectance component by using the convolutional autoencoder. The conventional denoising autoencoder performs the training in the form of the randomly corrupted input and original vectors to reduce the dimensionality. On the other hand, the convolutional autoencoder trains the neural network using the two-dimensional image data. For that reason, the convolutional autoencoder can better preserve the image structure since the weights are shared at all input data. In the proposed method, the convolutional autoencoder can be regarded as the constraint term on the reflectance component using l_1-norm minimization in (1). In addition, the proposed dual autoencoder is trained to provide the output which is equal to input in self-supervised learning manner, the stacked and convolutional autoencoders are regarded as the data-fidelity term $\|f_L - \hat{f}_L\|_2^2$ in the retinex model.

IV. EXPERIMENTAL RESULTS

This section demonstrates the performance of proposed dual autoencoder model by comparing with existing low-light image enhancement methods including Kim's method [20], Jiang's method [21], Jobson's method (MSRCR) [5], Fu's method [8], Guo's method [22], and Park's method [12]. The objective assessments are evaluated using the peak signal-to-noise ratio (PSNR) and structural similarity index measure (SSIM) [23].

In addition, we describe the training condition, to generate proper training data, and analyze the performance of proposed dual autoencoder model. Adam, which is proposed by Kingma and Ba [24], is used for optimization with a learning rate of 0.001. The mini batch size of each autoencoder is respectively set to 512 and 128. The experiments were performed on a personal computer with a CPU of 3.9 GHz, RAM of 24 GByte, and Nvidia GPU Titan Xp. The training of stacked and convolutional autoencoder networks respectively took about 254 and 1,501 seconds using GPU. Given the input image of size 512 × 512, the proposed method provided the enhanced result in 13.3 seconds using CPU.

A. TRAINING DATA GENERATION

The proposed method respectively estimates the illumination and reflectance components in the stacked and convolutional autoencoders as shown in Fig. 1. Since the stacked autoencoder estimates the brightness enhanced illumination component from the input low-light image, we generated a set of training data using a pair of low- and high-contrast patches because it is difficult to naturally obtain low- and high-contrast image pairs [13].

FIGURE 4. Set of ideal images collected from the internet to synthesize the training dataset.

For that reason, we synthesized the low-contrast image patch using about 80,000 patches of size 33 × 33, which were sampled from a set of ideal (high-contrast) images collected from the internet and the dataset used in [25] as shown in Fig. 4. The low-contrast patches were generated using the gamma corrected luminance channel in the HSV color space of ideal image patches as

$$f_p = g_p^{\gamma}, \tag{9}$$

where g_p and f_p represent the low- and high-contrast patches and γ the adjusting parameter which is randomly set to in [1.6, 3.3] for each patch.

In the proposed dual autoencoder model, the convolutional autoencoder performs the noise reduction in the initially estimated reflectance component. To train the convolutional denoising autoencoder, we generated about 190,000 pairs of noisy and noise-free patches size 28 × 28 using the luminance channel in the HSV color space of the ideal image. The noisy patch was synthesized by adding Gaussian noise with zero mean and randomly selected standard deviation $\sigma \in [10, 18]$. Fig. 5 shows a part of synthesized noisy low-contrast patches and their ground-truth.

B. PERFORMANCE OF PROPOSED AUTOENCODER MODELS

In this subsection, we compare the performance of the stacked autoencoder to estimate the illumination component. Since the autoencoder reconstructs the original data from the compressed representation of the input data, the performance depends on the number of hidden units. For that reason, we compared the resulting image with different number of hidden units in the bottleneck layer as shown in Fig. 6.

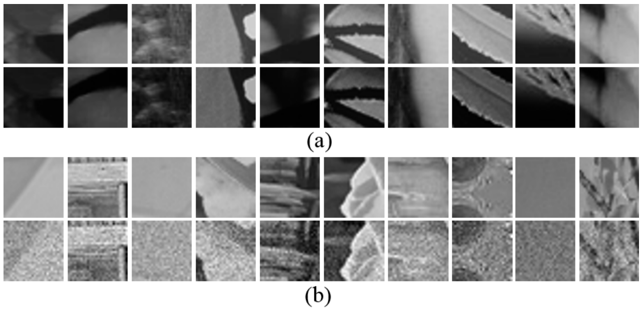

FIGURE 5. **A part of synthesized training dataset for the proposed dual autoencoder network: (a) high- and low-contrast patches for the stacked autoencoder and (b) noise-free and noisy patches for the convolutional autoencoder.**

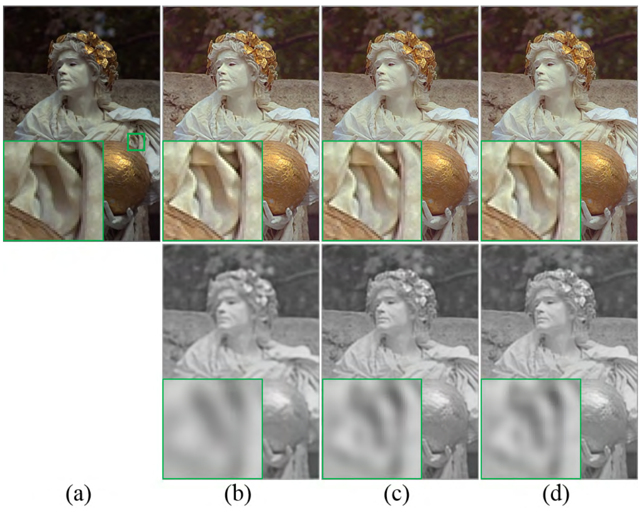

FIGURE 6. **Comparative results using different number hidden units in the bottleneck layer: (a) input image, (b) 32 hidden units (AE: 7.5489), (c) 64 hidden units (AE: 7.5208), and (d) 96 hidden units (AE: 7.5205). The first row shows the enhanced resulting images, and the second row shows the estimated illumination components.**

Since the input patch of size 33 × 33 is transformed to a vector in $\mathbb{R}^{33^2}$, the number of bottleneck layer should be lower than 33^2 to compress the input data.

Fig. 6 shows the resulting images using the estimated illumination component by three different stacked autoencoders having 32, 64, and 96 hidden units in the bottleneck layer. Specifically, the resulting image using 32 hidden units provided the best contrast because of more compressed representation of the input low-light image. It implies that the stacked autoencoder can better represent the low-frequency property of the illumination component with the small number of hidden units. Moreover, the higher average entropy (AE) value implies that the resulting image has more image information such as visible edges [26]. Based on the observation, we used 32 hidden units in the bottleneck layer.

In the convolutional autoencoder, we compared the performance of noise reduction using the MSE at different number of convolution filters and different sizes of receptive field of max-pooling layer as summarized in Table 1. As shown in the Table 1, n_1, n_2, and n_3 represent the number of filters used in each convolution layer, and p_1 is the size of receptive field in max-pooling layer. u_1 is a factor of magnification in the decoding part as shown in Fig. 3. As summarized in Table 1, the smaller size of the receptive field in max-pooling layer and more convolution filters provided better denoising performance with the smallest MSE value in the training and validation. Therefore, the proposed method takes the maximum value in the local window of size 2 × 2. The output size of convolution and max-pooling layers are summarized in Table 2.

TABLE 1. **Comparison of the performance of convolution layer with three different filter sizes and receptive fields of the max-pooling layer using the MSE value (10^{-2}).**

Layer	Parameter n_1-p_1-n_2-u_1-n_3	Training loss	Validation loss
1st conv layer	**16**-2-64-2-1	0.1100	0.1800
	32-2-64-2-1	0.1000	0.1700
	64-2-64-2-1	**0.0951**	**0.1500**
Max-pooling	64-**2**-64-2-1	**0.0951**	**0.1500**
	64-**4**-64-4-1	0.2400	0.4000
	64-**7**-64-7-1	0.5100	0.6800
2nd conv layer	64-2-**16**-2-1	0.1100	0.1700
	64-2-**32**-2-1	0.1000	0.1600
	64-2-**64**-2-1	**0.0951**	**0.1500**

TABLE 2. **The output size of each layer in the convolutional autoencoder.**

Layer	Size of output	# of filters	Filter size	Padding
1st conv layer	64 × 28 × 28	64	3 × 3	1
Max-pooling	64 × 14 × 14	-	2 × 2	1
2nd conv layer	64 × 14 × 14	64	3 × 3	1
Up-sampling	64 × 28 × 28	-	2 × 2	1
3rd conv layer	1 × 28 × 28	64	3 × 3	1

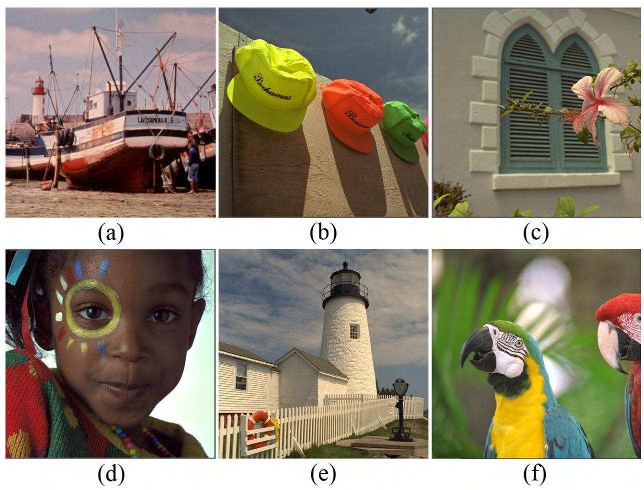

FIGURE 7. **The test images used in the objective assessments.**

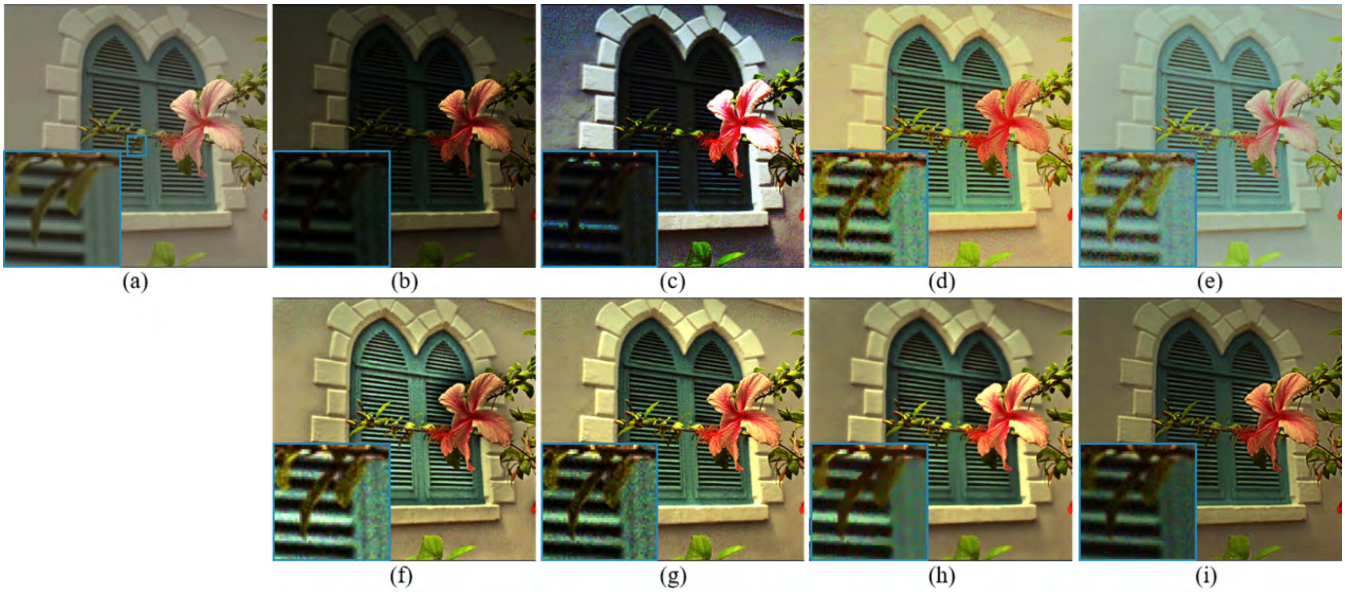

FIGURE 8. The performance comparison of proposed and existing methods using the synthesized low-light image: (a) input image, (b) synthesized low-light image, (c) Kim's method [20], (d) Jiang's method [21], (e) Jobson's method [5], (f) Fu's method [8], (g) Guo's method [22], (h) Park's method [12], and (i) the proposed method.

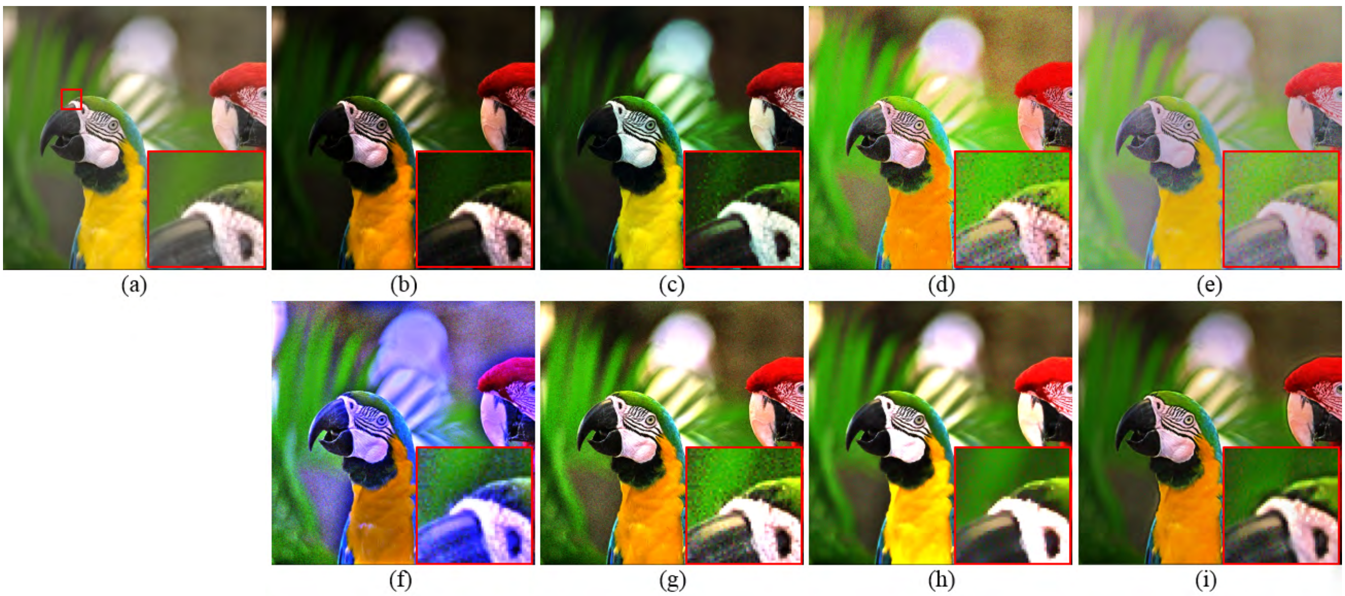

FIGURE 9. The performance comparison of proposed and existing methods using the synthesized low-light image: (a) input image, (b) synthesized low-light image, (c) Kim's method [20], (d) Jiang's method [21], (e) Jobson's method [5], (f) Fu's method [8], (g) Guo's method [22], (h) Park's method [12], and (i) the proposed method.

C. OBJECTIVE ASSESSMENTS

To perform the objective assessments of the proposed method, we synthesized the low-light image by degrading the brightness of the ideal image with an additive Gaussian noise with zero mean and $\sigma = 3$ only in the dark region. Fig. 7 shows a set of ideal images, and the corresponding PSNR and SSIM values are summarized in Table 3.

Figs. 8(a) and 9(a) show the ideal images and Figs. 8(b) and 9(b) show the synthesized low-light images. The histogram-based method cannot provide a successfully enhanced result in both dark and bright regions with an artifact of brightness saturation [20]. As shown in Figs. 8(d) and 9(d), although the transmission map-based method provided the enhanced result using the degradation model of hazy image, it cannot avoid the noise amplification and color distortion [21]. Likewise, although Jobson's method provided the enhanced result using multi-scale retinex with the color correction function to suppress the color

TABLE 3. Objective evaluation of the performance using PSNR and SSIM [23]. The red and blue colors represent the first and second best scores, respectively.

	Synthesized		[20]		[21]		[5]		[8]		[22]		[12]		Proposed	
	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
Fig. 7(a)	11.60	0.5920	14.46	0.4228	16.79	0.6779	16.80	0.5017	8.92	0.4342	17.31	0.6927	14.74	0.6645	17.28	0.7474
Fig. 7(b)	11.79	0.4034	13.75	0.3278	16.07	0.5706	15.02	0.6479	11.64	0.2956	15.66	0.6075	15.70	0.6348	16.91	0.6782
Fig. 7(c)	11.14	0.4574	13.38	0.2995	15.52	0.5955	15.85	0.5719	17.95	0.6873	14.31	0.6267	16.69	0.6646	16.78	0.7200
Fig. 7(d)	14.33	0.3824	15.48	0.4145	18.56	0.6348	13.50	0.5848	11.45	0.4097	16.48	0.6367	15.79	0.5938	17.81	0.6699
Fig. 7(e)	11.18	0.4784	13.02	0.2824	15.42	0.5589	15.32	0.6803	16.49	0.6530	15.60	0.6002	16.98	0.6102	16.51	0.6643
Fig. 7(f)	11.99	0.5034	14.21	0.5408	16.14	0.6519	14.69	0.7058	13.13	0.3167	16.37	0.7149	14.30	0.7020	16.81	0.7469
Avg.	12.01	0.4659	14.05	0.3813	16.42	0.6149	15.20	0.6154	13.26	0.4661	15.95	0.6465	15.70	0.6450	17.02	0.7044

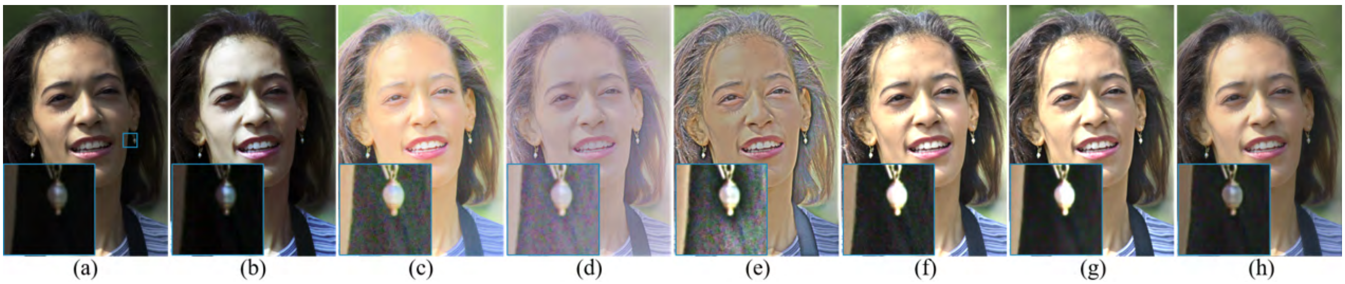

FIGURE 10. The performance comparison of proposed and existing methods using the real low-light image: (a) original image, (b) Kim's method [20], (c) Jiang's method [21], (d) Jobson's method [5], (e) Fu's method [8], (f) Guo's method [22], (g) Park's method [12], and (h) the proposed method.

FIGURE 11. The performance comparison of proposed and existing methods using the real low-light image: (a) original image, (b) Kim's method [20], (c) Jiang's method [21], (d) Jobson's method [5], (e) Fu's method [8], (f) Guo's method [22], (g) Park's method [12], and (h) the proposed method.

distortion, it also shows the noise amplification, which can be easily recognized in the cropped, enlarged image [5].

As shown in Figs. 8(f) and 9(f), the variational retinex model-based method provided better enhanced result using the bright channel prior [8]. However, it loses the edges and textures while suppressing the noise since it estimates the reflectance component using l_2-norm minimization. Although Guo's method provided high-contrast image by refining the illumination map at the low-computational cost, it cannot control the amplified noise as shown in Figs. 8(g) and 9(g) [22]. Park's method provided better performance than Fu's method in preserving the edges by the l_1-norm minimization while suppressing the noise, but the resulting image shows a certain amount of brightness saturation [12]. On the other hand, the proposed method provided better enhanced result in the sense of brightness enhancement and noise reduction without undesired artifacts using retinex-based dual autoencoder network.

D. QUALITATIVE ASSESSMENTS

In this subsection, the performance of the proposed dual autoencoder is evaluated using real low-light images as shown in Figs. 10 and 11. The histogram-based method can provide contrast enhanced image by redistributing the histogram bins in each sub-histogram, but it cannot successfully improve the quality of the dark region of the background [20]. Although Jiang's method provided the significantly enhanced result by using the transmission map estimated from the inverted input low-light image, it cannot avoid the color distortion and noise amplification [21].On the other hand, Jobson's method provided enhanced result by estimating the reflectance component using multiple Gaussian kernels. However, it provides unnaturally enhanced results with a narrow dynamic range [5].

The variational retinex approach provided better enhanced images, but it shows the over-enhancement with amplified noise and black halo near edges [8]. Although Guo's method provided the high-contrast resulting image using refined illumination map, it needs additional post-processing to reduce the noise amplification [22]. In terms of noise removal, Park's method provided better resulting images by using l_1-norm minimization to the reflectance component, but this method cannot avoid the brightness saturation in the bright region [12]. On the other hand, the proposed method can provide significantly enhanced result without saturation and noise amplification using the retinex-based dual autoencoder model than existing image enhancement methods.

V. CONCLUSION

In this paper, we proposed the novel low-light image enhancement framework using the dual autoencoder network model based on the retinex theory. The proposed dual autoencoder model consists of the stacked and convolutional autoencoders to perform both brightness enhancement and noise reduction. In the proposed algorithm, the stacked autoencoder is used to estimate the brightness enhanced and blurred illumination component since the very small number of hidden units generates the very compact features of an input data. The convolutional autoencoder is used to prevent the noise amplification of the estimated reflectance component while preserving the edge. In addition, we observed that the stacked and convolutional autoencoders play a role of the smoothness terms on the illumination and reflectance components in the variational retinex model. Finally, the proposed method can provide the high-quality image in various image processing applications such as robot vision and visual surveillance systems in low-light condition.

REFERENCES

[1] E. H. Land, "The Retinex," *Amer. Sci.*, vol. 52, no. 2, pp. 247–264, 1964.

[2] E. H. Land and J. J. McCann, "Lightness and Retinex theory," *J. Opt. Soc. Amer.*, vol. 61, no. 1, pp. 1–11, 1971.

[3] D. J. Jobson, Z.-U. Rahman, and G. A. Woodell, "Properties and performance of a center/surround Retinex," *IEEE Trans. Image Process.*, vol. 6, no. 3, pp. 451–462, Mar. 1997.

[4] Z. Rahman, D. J. Jobson, and G. A. Woodell, "Multi-scale Retinex for color image enhancement," in *Proc. IEEE Int. Conf. Image Process.*, vol. 3. Sep. 1996, pp. 1003–1006.

[5] D. J. Jobson, Z.-U. Rahman, and G. A. Woodell, "A multiscale Retinex for bridging the gap between color images and the human observation of scenes," *IEEE Trans. Image Process.*, vol. 6, no. 7, pp. 965–976, Jul. 1997.

[6] R. Kimmel, M. Elad, D. Shaked, R. Keshet, and I. Sobel, "A variational framework for Retinex," *Int. J. Comput. Vis.*, vol. 52, no. 1, pp. 7–23, 2003.

[7] W. Ma, J.-M. Morel, S. Osher, and A. Chien, "An L_1-based variational model for Retinex theory and its application to medical images," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2011, pp. 153–160.

[8] X. Fu, D. Zeng, Y. Huang, X. Ding, and X.-P. Zhang, "A variational framework for single low light image enhancement using bright channel prior," in *Proc. IEEE Global Conf. Signal Inf. Process.*, Dec. 2013, pp. 1085–1088.

[9] S. Park, B. Moon, S. Ko, S. Yu, and J. Paik, "Low-light image enhancement using variational optimization-based Retinex model," in *Proc. IEEE Int. Conf. Consum. Electron.*, Jan. 2017, pp. 70–71.

[10] D. Zosso, G. Tran, and S. Osher, "A unifying Retinex model based on non-local differential operators," *Proc. SPIE*, vol. 8657, p. 865702, Feb. 2013.

[11] X. Fu, Y. Liao, D. Zeng, Y. Huang, X. Zhang, and X. Ding, "A probabilistic method for image enhancement with simultaneous illumination and reflectance estimation," *IEEE Trans. Image Process.*, vol. 24, no. 12, pp. 4965–4977, Dec. 2015.

[12] S. Park, S. Yu, B. Moon, S. Ko, and J. Paik, "Low-light image enhancement using variational optimization-based Retinex model," *IEEE Trans. Consum. Electron.*, vol. 63, no. 2, pp. 178–184, May 2017.

[13] K. G. Lore, A. Akintayo, and S. Sarkar, "LLNet: A deep autoencoder approach to natural low-light image enhancement," *Pattern Recognit.*, vol. 61, pp. 650–662, Jan. 2017.

[14] L. Shen, Z. Yue, F. Feng, Q. Chen, S. Liu, and J. Ma. (Nov. 2017). "MSR-net:Low-light image enhancement using deep convolutional network." [Online]. Available: https://arxiv.org/abs/1711.02488

[15] C. Xing, L. Ma, and X. Yang, "Stacked denoise autoencoder based feature extraction and classification for hyperspectral images," *J. Sensors*, vol. 2016, Jun. 2016, Art. no. 3632943.

[16] L. Gondara, "Medical image denoising using convolutional denoising autoencoders," in *Proc. IEEE 16th Int. Conf. Data Mining Workshops (ICDMW)*, Barcelona, Spain, Dec. 2016, pp. 241–246.

[17] P. Vincent, H. Larochelle, Y. Bengio, and P.-A. Manzagol, "Extracting and composing robust features with denoising autoencoders," in *Proc. 25th Int. Conf. Mach. Learn.*, 2008, pp. 1096–1103.

[18] J. Masci, U. Meier, D. Cireşan, and J. Schmidhuber, "Stacked convolutional auto-encoders for hierarchical feature extraction," in *Artificial Neural Networks and Machine Learning*. Springer, vol. 6791. 2011, pp. 52–59.

[19] B. Du, W. Xiong, J. Wu, L. Zhang, L. Zhang, and D. Tao, "Stacked convolutional denoising auto-encoders for feature representation," *IEEE Trans. Cybern.*, vol. 47, no. 4, pp. 1017–1027, Apr. 2017.

[20] Y.-T. Kim, "Contrast enhancement using brightness preserving bi-histogram equalization," *IEEE Trans. Consum. Electron.*, vol. 43, no. 1, pp. 1–8, Feb. 1997.

[21] X. Jiang, H. Yao, S. Zhang, X. Lu, and W. Zeng, "Night video enhancement using improved dark channel prior," in *Proc. IEEE Int. Conf. Image Process.*, Sep. 2013, pp. 553–557.

[22] X. Guo, Y. Li, and H. Ling, "LIME: Low-light image enhancement via illumination map estimation," *IEEE Trans. Image Process.*, vol. 26, no. 2, pp. 982–993, Feb. 2017.

[23] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, Apr. 2004.

[24] D. P. Kingma and J. Ba. (2014). "Adam: A method for stochastic optimization." [Online]. Available: https://arxiv.org/abs/1412.6980

[25] 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, Feb. 2016.

[26] H. Hase, M. Yoneda, M. Sakai, and J. Yoshida, "Evaluation of handprinting variation of characters using variation entropy," *IEICE Trans. D*, vol. 71, no. 6, pp. 1048–1056, 1988.

SEONHEE PARK (S'16) was born in Busan, South Korea, in 1993. She received the B.S. degree in integrative engineering from Chung-Ang University, South Korea, in 2016, where she is currently pursuing the M.S. degree in digital imaging engineering. Her research interests include super-resolution, remote sensing images, denoising, and image enhancement and restoration for display processing.

SOOHWAN YU (S'15) was born in Incheon, South Korea, in 1988. He received the B.S. degree in information and communication engineering from Suwon University, South Korea, in 2013, and the M.S degree in image engineering from Chung-Ang University, South Korea, in 2016, where he is currently pursuing the Ph.D. degree in image engineering. His research interests include image enhancement, super-resolution, and image restoration.

MINSEO KIM was born in Seoul, South Korea, in 1992. She received the B.S. degree in integrative engineering from Chung-Ang University, South Korea, in 2016, where she is currently pursuing the M.S. degree in digital imaging engineering. Her research interests include dehazing, machine learning, and image enhancement and restoration for display processing.

KWANWOO PARK was born in Ulsan, South Korea, in 1994. He received the B.S. degree in integrative engineering from Chung-Ang University, South Korea, in 2017, where he is currently pursuing the M.S. degree in digital imaging engineering. His research interests include denoising, machine learning, and image enhancement and restoration for display processing.

JOONKI PAIK (M'89–SM'12) was born in Seoul, South Korea, in 1960. He received the B.Sc. degree in control and instrumentation engineering from Seoul National University, in 1984, and the M.Sc. and Ph.D. degrees in electrical engineering and computer science from Northwestern University, in 1987 and 1990, respectively. From 1990 to 1993, he joined Samsung Electronics, where he designed the image stabilization chip sets for consumer camcorders. Since 1993, he has been the faculty with Chung-Ang University, Seoul, South Korea, where he is currently a Professor with the Graduate School of Advanced Imaging Science, Multimedia, and Film. From 1999 to 2002, he was a Visiting Professor with the Department of Electrical and Computer Engineering, The University of Tennessee, Knoxville. He has served the Consumer Electronics Society of the IEEE as a member of the Editorial Board. Since 2005, he has been the Head of the National Research Laboratory in the field of image processing and intelligent systems. From 2005 to 2007, he served as the Dean of the Graduate School of Advanced Imaging Science, Multimedia, and Film. From 2005 to 2007, he was the Director of the Seoul Future Contents Convergence Cluster established by the Seoul Research and Business Development Program. In 2008, he was a full-time Technical Consultant for the System LSI Division of Samsung Electronics, where he developed various computational photographic techniques, including an extended depth of field system. He is currently serving as a member of the Presidential Advisory Board for Scientific/Technical Policy with the Korean Government and also a Technical Consultant for the Korean Supreme Prosecutor's Office for computational forensics. He was a recipient of the Chester-Sall Award from the IEEE Consumer Electronics Society, Academic Award from the Institute of Electronic Engineers of Korea, and a Best Research Professor Award from Chung-Ang University.