https://doi.org/10.2352/ISSN.2470-1173.2019.11.IPAS-276

Pixelwise JPEG Compression Detection and Quality Factor Estimation Based on Convolutional Neural Network

Kazutaka Uchida[a,b], Masayuki Tanaka[a,c], and Masatoshi Okutomi[a]
[a] Tokyo Institute of Technology, Tokyo, Japan
[b] Kadinche Corporation, Tokyo, Japan
[c] National Institute of Advanced Industrial Science and Technology, Tokyo, Japan

Abstract
JPEG compression is one of image degradations that often occurs in image storing and retouching process. Estimating JPEG compression degradation property is important for JPEG deblocking algorithm and image forensic analysis. JPEG degradation exists not only in JPEG file format but also in other image formats because JPEG distortion remains after converting to another image format. Moreover, JPEG degradation property is not always uniform within an image in case that the image is collaged from different JPEG-compressed photos. In this paper, pixelwise detection of JPEG-compression degradation and estimation of JPEG quality factor using a convolutional neural network is proposed. The proposed network outputs an estimated JPEG quality factor map and a compression flag map from an input image. Experimental results show that the proposed network successfully infers the quality factors and discriminates between non-JPEG-compressed images and JPEG-compressed images. We also demonstrate that the proposed network can spot a collaged region in a fake image which is comprised of images that have different JPEG compression properties. Additionally, the network reveals that image datasets Set5 and Set14, often used to evaluate super-resolution algorithms, contain JPEG-compressed low quality images, which are inappropriate for such evaluation.

Introduction
Image degradation estimation has been a hot topic for decades since identifying degradation properties plays a key role in image restoration [27] and other applications. Many degradation estimation techniques such as noise level detection [19, 21, 20], blur map estimation [24, 35] and haze map estimation [29] have been researched because they are essential process in image restoration algorithms, i.e. denoising [6], deblurring [4], and dehazing [10, 12]. Pixelwise degradation parameter estimation is also powerful for image forensic analysis like fake image detection because it reveals unnatural degradation patterns in a target image.

JPEG compression is one of the image degradations which occurs in an image capturing pipeline and image retouching processes. Detecting JPEG-compression and estimating JPEG quality factor (Q-factor) give crucial information for JPEG deblocking algorithms. Moreover, detecting JPEG-compression from the pixel data is significant to understand the history of the image as part of image forensic analysis.

There are mainly two approaches for JPEG quality factor estimation of an image. The first approach is exploiting metadata and the DCT coefficients in the JPEG image file. Li *et al*. [18] proposed a method to detect a tampered region of collaged image by analyzing the DCT coefficients. Their method successfully spots the tampered region, however, it is only valid as long as the image is in JPEG format. Once the image is converted to another file format such as BMP and PNG, the method cannot be no longer applied.

The second approach is to use the pixel data of a target image. This is more general way to estimate the JPEG quality factor because it is independent from the file format. Fu *et al*. [11] proposed a model-based estimation method for JPEG quality factor by pixel data. Their method is able to infer the JPEG quality factor of a target image, however, it is not possible to generate pixelwise JPEG quality factor map which is suitable for fake image detection.

The objective of this study is to estimate pixelwise degradation property, focusing on degraded images caused by JPEG compression, from the pixel data of an image. JPEG compression degradation exists not only in JPEG-format image but also PNG, BMP and other file format image because the JPEG degradation remains after the image is converted into another file format. To adapt to general image file formats rather than JPEG format, the estimation algorithm is required to use only the pixel data of an image, i.e. without referring to DCT coefficients and other metadata. Additionally, pixelwise estimation is vital because it enables to detect Photoshopped regions in an image that is collaged from JPEG-compressed and non-compressed images.

Proposed Method
In this section, we propose a neural network for JPEG quality factor estimation and for the detection whether the input image is JPEG-compressed or not.

Network Structure
Figure 1 shows a convolutional neural network (CNN) structure to estimate JPEG quality factor pixel-by-pixel. The proposed network consists of seven dilated convolutional layers with the dilated rate of one, two, three, four, three, two, and one, respectively. The filter size is 3×3 and the activation function is ReLU for all layers except the last one. The network structore is similar to [33]. The input of the network is an RGB image and the network outputs two maps; a JPEG quality map and a compression flag map. The output maps have the same size as the input image. The JPEG quality map is a pixelwise estimated JPEG quality factor whose value is linearly normalized to $[0,1]$; zero is for the JPEG quality 100% (and for no-compression) and one is for 0%. The compression flag map is a pixelwise estimation whether the

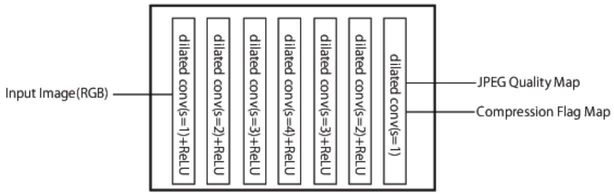


Figure 1: Proposed network structure

pixel is ever JPEG-compressed or not. The value is binary where one means JPEG-compressed and zero means non-compressed. Note that a JPEG compressed image with the quality factor 100% is not equal to the no-compression image because JPEG compression is lossy even for the quality factor of 100%. Thus, the compression flag map is important to discriminate high quality JPEG image and no-compression image.

Training

Training the network is achieved by optimizing parameters within the convolution layers so that the network can predict the true quality factor and compression flag map for a given input image. To ensure that original images for training dataset have never been JPEG compressed, only raw images are used for the dataset.

Training data is generated online from image patches cropped from the raw images. The training data is a mixture of two types of images: JPEG-compressed image and non-compressed image. As for JPEG-compressed images, the quality factors are randomly chosen for each image patch while the values of the compression flag map are set to one. As for non-compressed image (*i.e.* raw image), the values of the JPEG quality map and the compression flag map are set to zero.

The network parameters are optimized by using general optimization algorithm such as stochastic gradient descent (SGD) and Adam [17]. As for the loss function, mean squared errors for the quality factor map and the compression flag map are used.

Inference

Inference of JPEG quality factor and JPEG compression flag are simple feed-foward process. The trained network outputs the estimated quality factor map and compression flag map for a given input image. To normalize the range of output maps, the values are clipped to $[0,1]$.

If the input image is supposed to have a single quality factor (*i.e.* not collaged), the mean value of the quality factor map and the compression flag map are convenient to estimate the JPEG-compression property of the image. If the mean value of the compression flag is greater than 0.5, the input image is considered as ever-JPEG-compressed.

Experiments

To validate the proposed network, the network is trained and some evaluation tasks are executed with the network.

Training

In the training phase, 300 images from a raw image dataset (*i.e.* RAISE-1k dataset [7]) resized to 10% in size (e.g. 493×326) are used. Then, random patches (size of 60×60) are cropped from the resized images. For each patches, data augmentation by flipping and rotating are applied to increase the data by eight times. The proposed network is trained so that it can predict the true JPEG quality map and the true compression flag map from a given JPEG-compressed (or non-compressed) patch generated with a random JPEG quality factor. 10% of the training patches are given as no-compressed and the rest 90% are JPEG-compressed with random JPEG quality factor ranging from 5% to 100%.

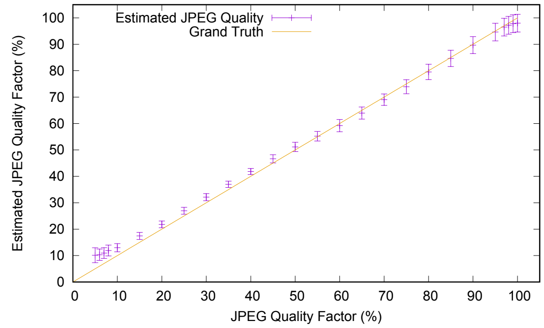


Figure 2: Performance on JPEG quality factor estimation

The optimization of the network parameters are done by Adam. The mini-batch size is 128 and one epoch contains 5.3k mini-batches. The training is finished after 80 epochs of optimization. The implementation is written with Keras framework [5] and run on a PC with a Nvidia TITAN-X GPU.

Experimental Results

We examined the trained network by three experiments as follows. First, the accuracy of the JPEG quality and JPEG compression prediction is measured. Second, we apply the network to a photo compositing of two different images that have different JPEG compression histories. Third, we assess image datasets to uncover its JPEG compression history.

JPEG Quality Estimation

To evaluate the trained network, the performance on JPEG quality estimation for JPEG-compressed images is examined. Fifty images from RAISE-1k dataset [7] which did not used in the training phase are selected. The images are resized to 10% in size and JPEG-compressed with the quality factor ranging from 5% to 100%. Estimated JPEG quality factor for a test image is taken as the mean value of the inferred JPEG quality map. The average estimated JPEG quality factors and its standard deviation for various JPEG quality factor on the test dataset is shown in Fig. 2. The network is able to infer the JPEG quality with small errors. Relatively large errors occur for images with high and low quality factor due to flat regions in the images. Since estimating a JPEG quality in a flat region is an ill-posed problem, the network was trained to predict the stochastically optimized value, i.e. around 55%, for the region. Thus, the averaged estimated JPEG quality for images that have flat region tends to drift to mid-quality factor.

Table 1 illustrates the discriminability on JPEG-compressed and non-compressed images. A test image is categorized to non-compressed if the mean value of compression flag map is less than 0.5. The ratio of estimated as non-compressed images in the test dataset are measured. The network exhibits high performance on

Table 1: Performance on JPEG degradation detection

JPEG quality factor	ratio of estimated as non-compressed
97%	0.00
98%	0.00
99%	0.00
100%	0.02
w/o JPEG comp.	1.00

the discriminability.

Fake Image Detection

Since the network can infer pixelwise JPEG quality map, it is useful for fake image detection. We examine the feasibility on a fake image detection task as an application of pixelwise JPEG quality estimation.

A fake image is created as follows. Two images of mountain scenery from the RAISE-1k dataset, that are not used for the training, are selected. They are resized by 10% in size. One image is left uncompressed as shown in Fig.3a, and the other image is JPEG-compressed with the quality factor of 80% as illustrated in Fig.3b. Then, the two images are collaged into an image (Fig.3c) with the layer mask of Fig.3d. The collaged image is a fake image of fictitious mountain view.

To examine if the proposed network can spot the collaged region, the JPEG compression detection is performed for the image. Figure 3e demonstrate the compression flag map, which obviously detects the collaged region. The failure detection in the remaining snow area is caused by the ill-conditioned problem due to the saturated flat region, as mentioned previously.

Image Assessment

Finally, we assess image datasets to reveal its JPEG compression histories.

Image datasets Set5 [1] and Set14 [31] are often used to evaluate super resolution algorithms [15, 9, 16, 32]. Images in the datasets are in BMP file format, however, they might have ever-JPEG-compressed images because the history of the images is unknown. Figure 4 and 5 show the mean estimated JPEG quality and compression flag for each image. As for Set5, 'baby' and 'woman' are predicted as JPEG-compressed because the mean value of the compression flag map is nearly one. As for Set14, 'foreman', 'coastguard', 'ppt3', and 'zebra' are inferred as JPEG-compressed with high values of the compression flag map and relatively low values of the estimated JPEG quality factor. Note that we excluded grayscale images from Set14, namely 'bridge' and 'man' because the prediction for grayscale images is not reliable due to the network training done by only color images.

Datasets BSD100 [22] and Urban100 [13] are also often used to evaluate super-resolution algorithms [14, 28, 16, 26]. Figure 6 and 7 show the histograms of estimated JPEG quality factors for BSD100 and Urban100, respectively. According to the estimation, all BSD100 images are ever-JPEG-compressed with the quality factor of around 95%. As for Urban100, it has 44 images that are estimated as ever-JPEG-compressed out of 100 images. Urban100 contains both high quality images without JPEG-compression and low quality JPEG-compressed images with the quality factor under 80%. Set5, Set14, BSD100, and Urban100 might be inappropriate for evaluation of super-resolution algorithms because they contain low quality JPEG-compressed images.

Dataset Live1 [23] and BSD68 [22] are often used to evaluate JPEG deblocking algorithms [30, 8, 25] and denoising algorithms [34, 2, 3], respectively. Figure 8 shows the histogram of estimated JPEG quality factors of Live1 dataset. It has 12 images that estimated as ever-JPEG-compressed out of 29 images. The estimated JPEG quality factors are relatively high, however, the pre-existing JPEG compression in the dataset may affect the evaluation of the performance of JPEG deblocking algorithms. Thus, the dataset might not be suitable for that purpose. Figure 9 shows the histogram of estimated JPEG quality factors of BSD68 dataset. All BSD68 images are estimated as ever-JPEG-compressed and most images are estimated to have 90% to 95% of the quality factor. With respect to the denoising evaluation, the JPEG noise existing in the dataset may lead to wrong evaluation of the denoising algorithms.

Conclusion

CNN-based pixelwise degradation estimation for JPEG compression is introduced. In previous works, only a single JPEG quality factor is determined per an image or a small block. In contrast, the proposed model infers pixelwise JPEG quality factor, which is useful for fake image detection. Furthermore, unlike previous works, the proposed method can estimate whether the image is ever JPEG-compressed or not from the pixel data. Considering that the JPEG compression degradation remains after being converted to another image format such as PNG and BMP, it is advantageous for investigating the history of an image as part of image forensic analysis. We have shown that the proposed pixelwise JPEG quality factor estimation helps for the fake image detection and image assessment.

References

[1] Marco Bevilacqua, Aline Roumy, Christine Guillemot, and Marie Line Alberi-Morel. Low-complexity single-image super-resolution based on nonnegative neighbor embedding. 2012.

[2] Chang Chen, Zhiwei Xiong, Xinmei Tian, and Feng Wu. Deep boosting for image denoising. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 3–18, 2018.

[3] Jingwen Chen, Jiawei Chen, Hongyang Chao, and Ming Yang. Image blind denoising with generative adversarial network based noise modeling. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3155–3164, 2018.

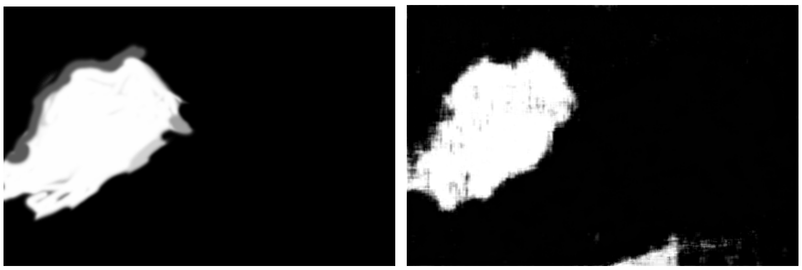
[4] Sunghyun Cho and Seungyong Lee. Fast motion deblurring. In *ACM Transactions on graphics (TOG)*, volume 28, page 145. ACM, 2009.

[5] François Chollet et al. Keras. `https://github.com/keras-team/keras`, 2015.

[6] Kostadin Dabov, Alessandro Foi, Vladimir Katkovnik, and Karen Egiazarian. Image denoising by sparse 3-d transform-domain collaborative filtering. *IEEE Transactions on image processing*, 16(8):2080–2095, 2007.

[7] Duc-Tien Dang-Nguyen, Cecilia Pasquini, Valentina Conot-

(a) Original image A (w/o JPEG comp.) (b) Original image B (JPEG: Q-factor= 80%) (c) Composited fake image



(d) Grand truth (layer mask) (e) Estimated compression flag map

Figure 3: Detection of JPEG-compressed region in a fake image

Filename	baby_GT.bmp	bird_GT.bmp	butterfly_GT.bmp	head_GT.bmp	woman_GT.bmp
Estimated Q-Factor (mean)	95.87%	98.29%	98.61%	99.50%	**78.07%**
Comp. Flag (mean)	**0.98**	0.32	**0.57**	0.01	**1.00**

Figure 4: JPEG compression assessment for Set5

ter, and Giulia Boato. Raise: a raw images dataset for digital image forensics. In *Proceedings of the 6th ACM Multimedia Systems Conference*, pages 219–224. ACM, 2015.

[8] Chao Dong, Yubin Deng, Chen Change Loy, and Xiaoou Tang. Compression artifacts reduction by a deep convolutional network. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 576–584, 2015.

[9] Chao Dong, Chen Change Loy, Kaiming He, and Xiaoou Tang. Image super-resolution using deep convolutional networks. *IEEE transactions on pattern analysis and machine intelligence*, 38(2):295–307, 2016.

[10] Chen Feng, Shaojie Zhuo, Xiaopeng Zhang, Liang Shen, and Sabine Süsstrunk. Near-infrared guided color image dehazing. In *Proc. IEEE 20th International Conference on Image Processing (ICIP)*, number EPFL-CONF-188639, pages 2363–2367, 2013.

[11] Dongdong Fu, Yun Q Shi, and Wei Su. A generalized benford's law for jpeg coefficients and its applications in image forensics. In *Security, Steganography, and Watermarking of Multimedia Contents IX*, volume 6505, page 65051L. International Society for Optics and Photonics, 2007.

[12] Kaiming He, Jian Sun, and Xiaoou Tang. Single image haze removal using dark channel prior. *IEEE transactions on pattern analysis and machine intelligence*, 33(12):2341–2353, 2011.

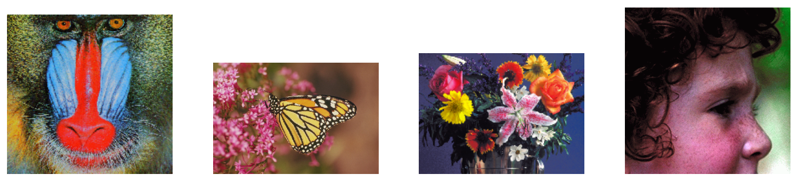
[13] Jia-Bin Huang, Abhishek Singh, and Narendra Ahuja. Single image super-resolution from transformed self-exemplars. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 5197–5206, 2015.

[14] Jia-Bin Huang, Abhishek Singh, and Narendra Ahuja. Single image super-resolution from transformed self-exemplars. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 5197–5206, 2015.

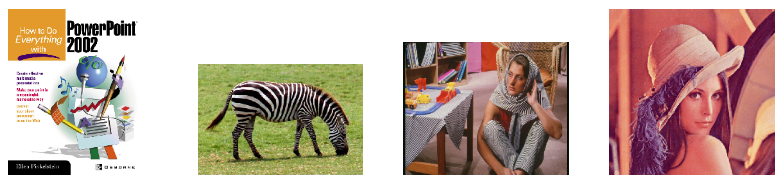
[15] Jiwon Kim, Jung Kwon Lee, and Kyoung Mu Lee. Accurate image super-resolution using very deep convolutional networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1646–1654, 2016.

[16] Jiwon Kim, Jung Kwon Lee, and Kyoung Mu Lee. Deeply-recursive convolutional network for image super-resolution. In *Proceedings of the IEEE Conference on Computer Vision*

Filename	foreman.bmp	pepper.bmp	comic.bmp	coastguard.bmp
Estimated Q-Factor (mean)	**74.51%**	99.69%	99.02%	**84.17%**
Comp. Flag (mean)	**0.99**	0.00	0.19	**1.00**



Filename	baboon.bmp	monarch.bmp	flowers.bmp	face.bmp
Estimated Q-Factor (mean)	99.69%	99.21%	99.76%	99.45%
Comp. Flag (mean)	0.00	**0.65**	0.00	0.01



Filename	ppt3.bmp	zebra.bmp	barbara.bmp	lenna.bmp
Estimated Q-Factor (mean)	**70.96%**	**83.09%**	99.29%	99.75%
Comp. Flag (mean)	**0.78**	**1.00**	0.03	0.00

Figure 5: JPEG compression assessment for Set14 (except grayscale images)

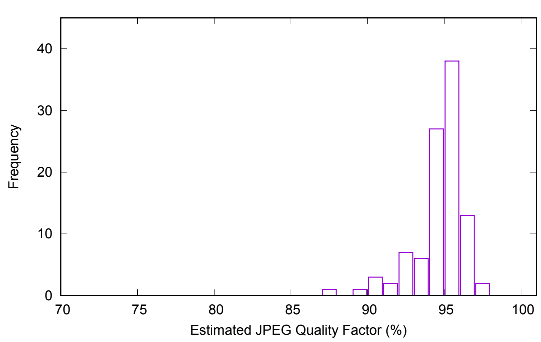


Figure 6: Histogram of estimated JPEG Q-factor for BSD100

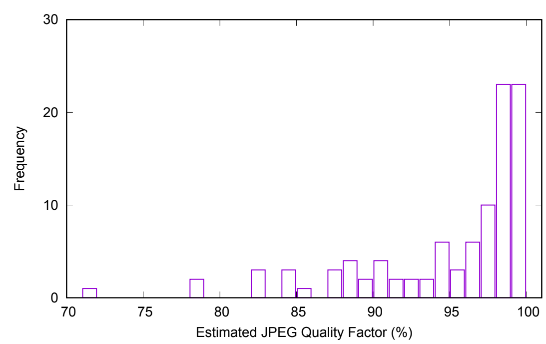


Figure 7: Histogram of estimated JPEG Q-factor for Urban100

and Pattern Recognition, pages 1637–1645, 2016.

[17] Diederik Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

[18] Xiang Hua Li, Yu Qian Zhao, Miao Liao, Frank Y Shih, and Yun Q Shi. Detection of tampered region for jpeg images by using mode-based first digit features. *EURASIP Journal on advances in signal processing*, 2012(1):190, 2012.

[19] Xinhao Liu, Masayuki Tanaka, and Masatoshi Okutomi. Noise level estimation using weak textured patches of a single noisy image. In *Image Processing (ICIP), 2012 19th IEEE International Conference on*, pages 665–668. IEEE, 2012.

[20] Xinhao Liu, Masayuki Tanaka, and Masatoshi Okutomi. Estimation of signal dependent noise parameters from a single image. In *Image Processing (ICIP), 2013 20th IEEE International Conference on*, pages 79–82. IEEE, 2013.

[21] Xinhao Liu, Masayuki Tanaka, and Masatoshi Okutomi. Single-image noise level estimation for blind denoising. *IEEE transactions on image processing*, 22(12):5226–5237, 2013.

[22] D. Martin, C. Fowlkes, D. Tal, and J. Malik. A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecologi-

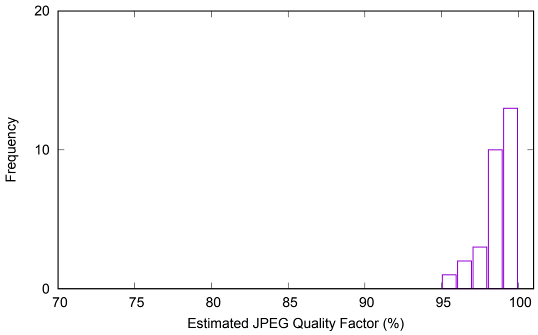


Figure 8: Histogram of estimated JPEG Q-factor for LIVE1

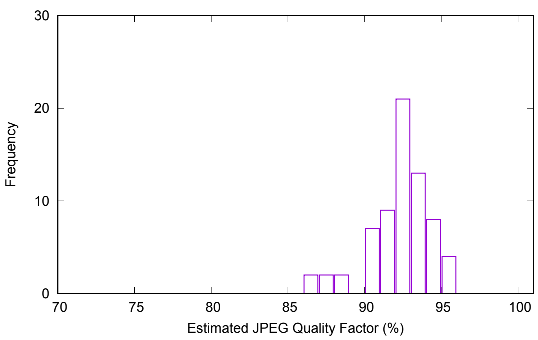


Figure 9: Histogram of estimated JPEG Q-factor for BSD68

cal statistics. In *Proc. 8th Int'l Conf. Computer Vision*, volume 2, pages 416–423, July 2001.

[23] HR Sheikh. Live image quality assessment database release 2. *http://live. ece. utexas. edu/research/quality*, 2005.

[24] Jianping Shi, Li Xu, and Jiaya Jia. Just noticeable defocus blur detection and estimation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 657–665, 2015.

[25] Pavel Svoboda, Michal Hradis, David Barina, and Pavel Zemcik. Compression artifacts removal using convolutional neural networks. *arXiv preprint arXiv:1605.00366*, 2016.

[26] Ying Tai, Jian Yang, and Xiaoming Liu. Image super-resolution via deep recursive residual network. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, volume 1, page 5, 2017.

[27] Kazutaka Uchida, Masayuki Tanaka, and Masatoshi Okutomi. Non-blind image restoration based on convolutional neural network. In *Consumer Electronics (GCCE), 2018 IEEE 7th Global Conference on*. IEEE, 2018.

[28] Zhaowen Wang, Ding Liu, Jianchao Yang, Wei Han, and Thomas Huang. Deep networks for image super-resolution with sparse prior. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 370–378, 2015.

[29] Chia-Hung Yeh, Li-Wei Kang, Ming-Sui Lee, and Cheng-Yang Lin. Haze effect removal from image via haze density estimation in optical model. *Optics express*, 21(22):27127–27141, 2013.

[30] Ke Yu, Chao Dong, Chen Change Loy, and Xiaoou Tang. Deep convolution networks for compression artifacts reduction. *arXiv preprint arXiv:1608.02778*, 2016.

[31] Roman Zeyde, Michael Elad, and Matan Protter. On single image scale-up using sparse-representations. In *International conference on curves and surfaces*, pages 711–730. Springer, 2010.

[32] Kai Zhang, Wangmeng Zuo, Yunjin Chen, Deyu Meng, and Lei Zhang. Beyond a gaussian denoiser: Residual learning of deep cnn for image denoising. *IEEE Transactions on Image Processing*, 2017.

[33] Kai Zhang, Wangmeng Zuo, Shuhang Gu, and Lei Zhang. Learning deep cnn denoiser prior for image restoration. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 3929–3938, 2017.

[34] Kai Zhang, Wangmeng Zuo, and Lei Zhang. Ffdnet: Toward a fast and flexible solution for cnn based image denoising. *IEEE Transactions on Image Processing*, 2018.

[35] Shaojie Zhuo and Terence Sim. Defocus map estimation from a single image. *Pattern Recognition*, 44(9):1852–1858, 2011.

Author Biography

Kazutaka Uchida *received bachelor's and master's degrees in control and systems engineering from Tokyo Institute of Technology, Tokyo, Japan, in 2000 and 2003, respectively. He joined Sony Corporation in 2003 and developed next-generation audio and visual applications. In 2008, he cofounded a startup company called Kadinche Corporation, where he currently works on the development of immersive virtual reality applications. He is currently pursuing his PhD.*

Masayuki Tanaka *received his bachelor's and master's degrees in control engineering and Ph.D. degree from Tokyo Institute of Technology in 1998, 2000, and 2003. He joined Agilent Technology in 2003. He was a Research Scientist at Tokyo Institute of Technology since 2004 to 2008. Since 2008, He has been an Associated Professor at the Graduate School of Science and Engineering, Tokyo Institute of Technology. He was a Visiting Scholar with Department of Psychology, Stanford University, CA, USA.*

Masatoshi Okutomi *received a B.Eng. degree from the Department of Mathematical Engineering and Information Physics, the University of Tokyo, Japan, in 1981 and an M.Eng. degree from the Department of Control Engineering, Tokyo Institute of Technology, Japan, in 1983. He joined Canon Research Center, Canon Inc., Tokyo, Japan, in 1983. From 1987 to 1990, he was a visiting research scientist in the School of Computer Science at Carnegie Mellon University, USA. In 1993, he received a D.Eng. degree for his research on stereo vision from Tokyo Institute of Technology. Since 1994, he has been with Tokyo Institute of Technology, where he is currently a professor in the Department of Systems and Control Engineering, the School of Engineering.*

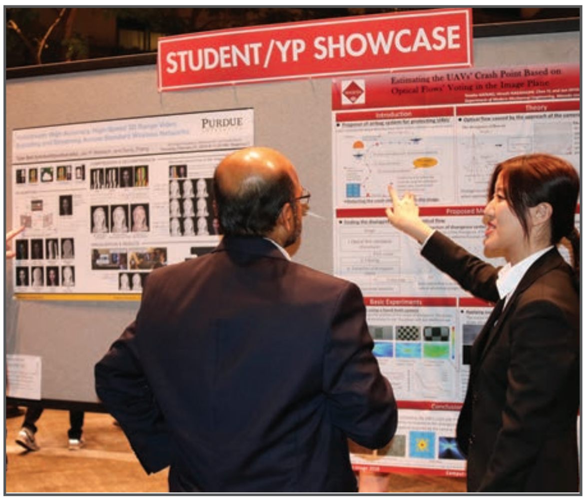
JOIN US AT THE NEXT EI!

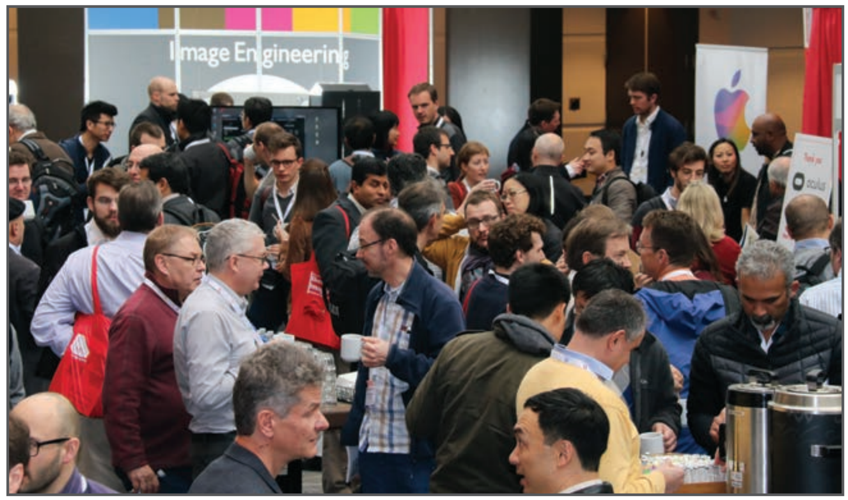
IS&T International Symposium on

Electronic Imaging

SCIENCE AND TECHNOLOGY

Imaging across applications . . . Where industry and academia meet!





• SHORT COURSES • EXHIBITS • DEMONSTRATION SESSION • PLENARY TALKS •
• INTERACTIVE PAPER SESSION • SPECIAL EVENTS • TECHNICAL SESSIONS •

www.electronicimaging.org

IS&T
imaging.org