

GAN Based Image Deblurring Using Dark Channel Prior

Shuang Zhang*, Ada Zhen, Robert L. Stevenson; University of Notre Dame; Notre Dame, Indiana, 46556

Abstract

A conditional general adversarial network (GAN) is proposed for image deblurring problem. It is tailored for image deblurring instead of just applying GAN on the deblurring problem. Motivated by that, dark channel prior is carefully picked to be incorporated into the loss function for network training. To make it more compatible with neuron networks, its original indifferentiable form is discarded and L_2 norm is adopted instead. On both synthetic datasets and noisy natural images, the proposed network shows improved deblurring performance and robustness to image noise qualitatively and quantitatively. Additionally, compared to the existing end-to-end deblurring networks, our network structure is light-weight, which ensures less training and testing time.

Introduction

Blur is a common artifact for images taken by hand-held cameras. It is mostly caused by object motions, hand shake or out-of-focus. The blurry image is often modeled as convolution of a sharp image and a blur kernel. And the target of deblurring is to restore a latent sharp image from the blurry one. Single image deblurring, however, is a highly ill-posed problem, since it contains insufficient information to recover a unique sharp image.

In the past few years, assorted constraints and regulation schemes have been proposed to exclude implausible solutions. Priors, like total variation prior [1], sparse image prior [2], heavy-tailed gradient prior [3] and dark channel prior [4], are combined with L_1/L_2 norm image regulation term to suppress ringing artifacts and improve quality. Zhen [5] takes advantage of inertia sensor data to gain extra information and estimate spatially varying blur kernels. However, since the blur kernel in reality is more complicated than the model, estimation of blur kernel is inaccurate, which causes ringing artifacts. Furthermore, these methods based on iterative optimization techniques are computationally intensive.

Recently, Convolutional Neural Networks (CNN) and deep learning related techniques have drawn a great attention in computer vision and image processing. Their applications in image deblurring demonstrate promising results. Sun [7] and Schuler [6] use CNN to estimate the spatially-invariant blur kernel and obtain latent image by tradition pipeline. Chakrabarti [13] trained a neural network to predict complex Fourier coefficients of motion kernel. Recently kernel-free end-to-end deblurring methods are proposed by Nah et al. [8] and Kupyn et al. [9]. Nah [8] adopted the multi-scale network to mimic conventional coarse-to-fine optimization methods, and proposed a new realistic blurry image dataset with ground truth sharp images. The work of Kupyn [9] trains the popular Generative Adversarial Network (GAN) on the same dataset with fewer parameters, gains higher PSNR values than that of Nah et al. [8] on the GOPRO dataset, and beats the others on Kohler dataset [10] using SSIM. Although [9] performs well based on metric scores, visually, its deblurred result suffers

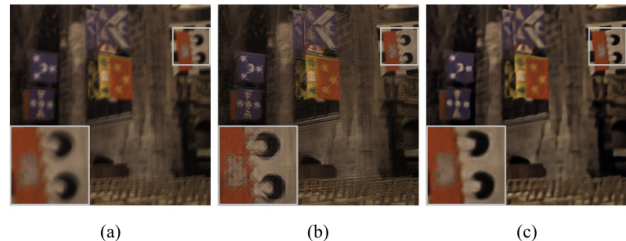


Figure 1. Comparison. (a) Input blurry image. (b) Result of [9]. (c) Our result.

grid artifacts, as illustrated in Fig. 1.

To address this artifact, we utilize the dark channel prior. Dark channel is defined as minimal intensity among three color channels of pixels in a local area. It was first proposed by He et al. [11] for dehazing problem, based on the statistics that haze-free outdoor images have a smaller dark channel than hazy images. Pan et al. [4] applied dark channel prior to image deblurring. They theoretically and empirically proved that comparing with blur images, the dark channel of sharp image is more sparse. And their results demonstrate that dark channel prior contributes to suppressing ringing and other artifacts. In order to enforce the sparsity, they utilize a regulation term of L_0 norm to count the nonzero elements of dark channel maps. Unfortunately, L_0 norm is not differentiable, which makes it hard to utilize in back propagation of neuron networks. Instead of using L_0 norm, we adopt L_2 norm to directly compute difference of the dark channel maps between groundtruth sharp images and deblurred images.

In this paper, we present a GAN based image deblurring network using dark channel difference as loss function. The proposed technique is not just a straightforward application of GAN. This method focuses on how to combine traditional knowledge with deep learning to make the network achieve better performance. Compared to the previous GAN-based deblurring network, the proposed network has less layers and weights. It leads to less training and testing time, and more importantly achieves favorable results. In addition, the original GOPRO training dataset consists of artificially created blurry images without noise, which are usually different from the real blurry images. To improve the quality of our trained network on more realistic blurry images and increase network robustness, we add random Gaussian noise with variance in a limited range onto the training image patches. The comparison experiments show that our network outperforms Kupyn et al. [9] for both GOPRO test dataset and real noisy blurry images.

Related Work

Conditional General Adversarial Networks

GAN is first proposed by Goodfellow et al. [14] to train a generative network in an adversarial process. It consists of two

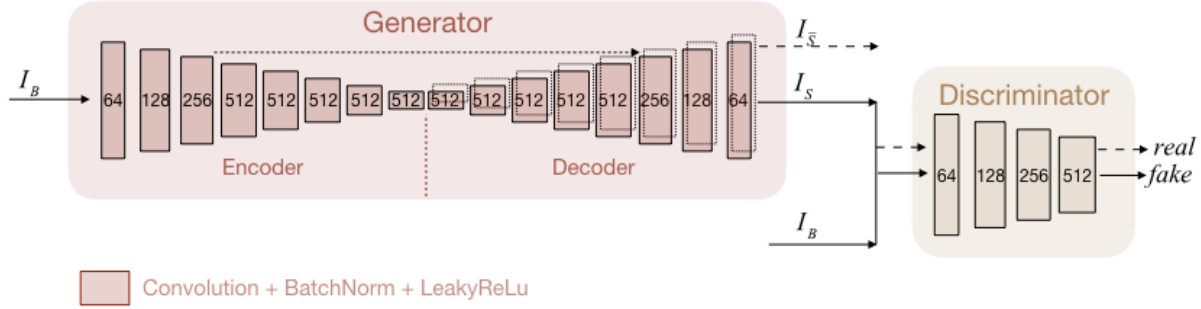


Figure 2. Proposed Network. Our CGAN based network has two sub-networks: generator G and discriminator D . Generator restores sharp image I_S from input blurry image I_B . I_S represents ground truth image. Discriminator can regard input pair (I_S, I_B) as "fake" and (I_S, I_B) as "real". Except for the first layer of discriminator and generator, each block in both generator and discriminator consists of a convolutional layer, batch normalization step [21] and an activation function LeakyReLU [23]. The first layers are not normalized. The digit denotes the number of filters for each block. Dotted lines are skip connection layers in decoder which come from layers with same size in encoder.

networks: a generator G and a discriminator D . Generator generates a fake sample from input noise z , while discriminator estimates the probability that the fake sample is from training data rather than generated by generator. These two networks are simultaneously trained until discriminator cannot tell if the sample is real or fake. This process can be summarized as a two-player min-max game with the following function:

$$\min_G \max_D \mathbb{E}_{\bar{x} \sim P_{\text{data}}(\bar{x})} [\log D(\bar{x})] + \mathbb{E}_{z \sim P_z(z)} [\log(1 - D(G(z)))], \quad (1)$$

where P_{data} denotes distribution over training data \bar{x} and P_z is distribution of input noise z . GAN has been applied to different image restoration problems like super-resolution [16] and texture transfer [17].

Mirza et al. [15] extend GAN into a conditional model (eq. (2)), called Conditional Generative Adversarial Nets (CGAN), so that GAN can make use of auxiliary information to direct both generator and discriminator. Isola et al. [18] adopt CGAN architecture to achieve general image-to-image translation. In [18], more than just random noise z , similar image y is added as input of the generator, where y and \bar{x} share part of features. y and \bar{x} can be pairs of hazing and clear images about same scene, or different color buildings with same structure. Based on network architecture of [18], Kupyń et al. [9] utilize Wasserstein loss [19] and perceptual loss [20] to train a CGAN for deblurring problem.

$$\min_G \max_D \mathbb{E}_{\bar{x} \sim P_{\text{data}}(\bar{x})} [\log D(\bar{x}, y)] + \mathbb{E}_{z \sim P_z(z)} [\log(1 - D(G(z, y, y)))], \quad (2)$$

Dark Channel Prior

For an image I , the dark channel of a pixel p is defined by He et al. [11] as

$$\mathfrak{D}_c(p) = \min_{q \in \mathcal{N}(p)} \left(\min_{c \in \{r, g, b\}} I^c(q) \right), \quad (3)$$

where p and q are pixel locations, $\mathcal{N}(p)$ denotes the image patch centered at p , and I^c is the c -th color channel. As shown in eq. (3), dark channel describes the minimum intensity in an image patch. He et al. [11] observe that dark channel map $D(I)$ in a

haze-free image tends to be zero. Pan et al. [4] use a less restrictive assumption that dark channel map $D(I)$ is sparse rather than zero. Inspired by this, they adopt L_0 -regulation term to enforce less sparse dark channel in a deblurring process, where L_0 norm counts non-zero elements in a dark channel map.

Proposed Method

Network Architecture

The proposed network aims at obtaining a generator to restore sharp image I_S from input blurry image I_B . This generator is trained with a discriminator using pairs of blurry image I_B and ground truth sharp image I_S . This structure is shown in fig.2. Except for the first layer of discriminator and generator, each block in both generator and discriminator consists of a convolutional layer, batch normalization step [21] and an activation function LeakyReLU [23] with leaking rate $\alpha = 0.2$. The first layers are not normalized.

Generator The proposed generator adopts an encoder-decoder framework to achieve image-to-image performance. Similar to [18], the encoder consists of a sequence of convolutional layers with stride = 2 and kernel size = 5. And the decoder has a chain of transposed-convolutional layers with same size of stride and kernel. Encoder represents the input image with a bottleneck vector and decoder recovers an image with same size of input from bottleneck vector. A skip architecture is applied by inserting same size of layers from encoder after each layer of decoder. This skip connection refines the details in output image by combining deep, coarse, semantic information and shallow, fine, appearance information [22]. Dropout is also included in decoder to avoid over-fitting.

Discriminator The proposed discriminator contains a series of convolutional layers with stride = 2 and kernel size = 5. The output of discriminator is a scalar, followed by a sigmoid function.

Loss Functions

According to eq. (2), we train discriminator and generator alternatively. The loss function of discriminator is same as adversarial loss:

$$\mathcal{L}_d = \mathbb{E}_{\bar{x}, y} [\log D(\bar{x}, y)] + \mathbb{E}_{y, z} [\log(1 - D(G(z, y, y)))]. \quad (4)$$

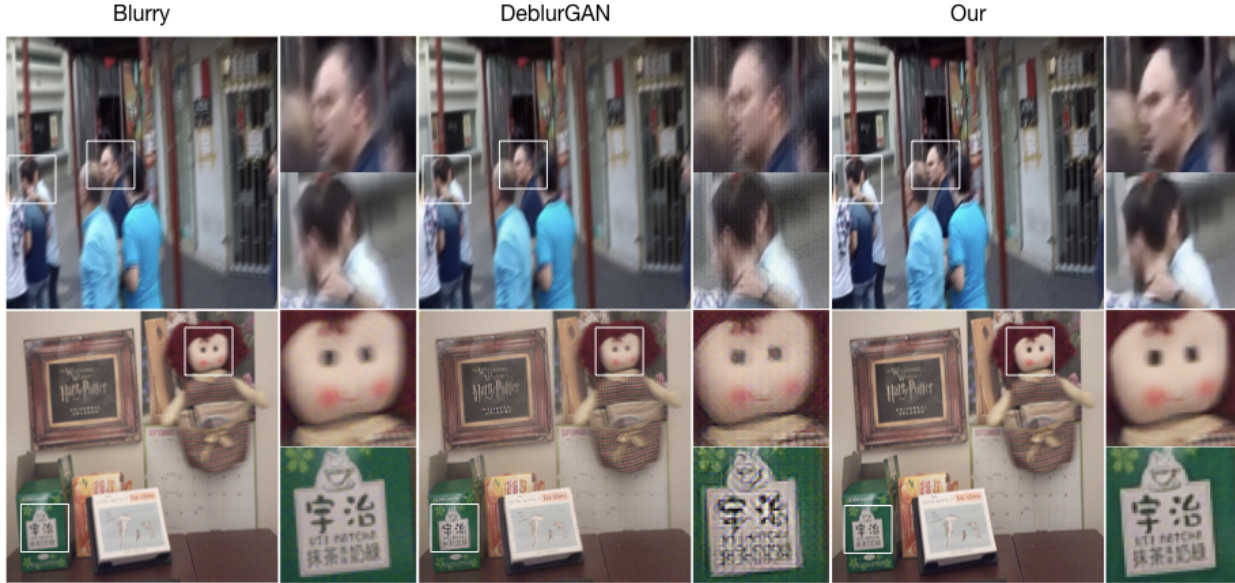


Figure 3. Comparison with DeblurGAN [9]. From top to bottom: image from GOPRO dataset and real nature image. From left to right: blurry images, deblurred results by [9] and our result.

In the deblurring setting, y and \bar{x} denote blurry and sharp image, respectively. The generator loss is defined as combination of adversarial loss, content loss and dark channel loss:

$$\mathcal{L}_g = \mathbb{E}_{y,z} [\log(1 - D(G(z), y))] + \lambda_1 \cdot \mathcal{L}_{content} + \lambda_2 \cdot \mathcal{L}_{darkchannel}, \quad (5)$$

where $\lambda_1 = 100$ and $\lambda_2 = 250$ in our experiments.

Content loss We adopt the traditional content loss to direct the output of generator to ground truth. Although both L_1 and L_2 norm are commonly used, L_1 norm is chosen since it attains less blurring result [18].

$$\mathcal{L}_{content} = \mathbb{E}_{\bar{x},y,z} [|\bar{x} - G(y,z)|_1]. \quad (6)$$

Dark channel loss In order to suppress ringing and grid artifact, dark channel prior is especially chosen. Pan et al. [4] exploit L_0 norm to count non-zero elements in a dark channel map $\mathcal{D}_c(I)$ of an image I . Since L_0 norm is indifferent, L_2 norm is utilized instead which calculates the distance of dark channel map between ground truth and deblurred image.

$$\mathcal{L}_{darkchannel} = \mathbb{E}_{\bar{x},y,z} [|\mathcal{D}_c(\bar{x}) - \mathcal{D}_c(G(y,z))|_2]. \quad (7)$$

Unlike [9], we discard the perceptual loss [20]. Kupyn et al. [9] employ the difference of one feature map in the VGG-19 [24] between ground truth and restored images as perceptual loss. GAN is known for its ability to reserve perceptual feature of an image. Adding an extra perceptual loss seems a noneffective repeat. Our experiment shows that perceptual loss doesn't improve the result, on the contrary, it leads to worse performance.

Experiments

Our network is implemented with Python code based on Tensorflow [25].

Datasets

GOPRO dataset [8] is utilized for training and testing our network. It contains 2103 pairs of blurry and ground truth images in train dataset, and 1111 pairs in test dataset. Resolution of the image are 720p. The blurry image is generated by averaging a sequence (7-15) of continuous sharp images. Sharp image in the middle of sequence is regarded as ground truth. GOPRO dataset is regarded as benchmark by many deblurring algorithms like [8] and [9]. Although GOPRO dataset is widely used, it only employs noise-free images. For natural images, however, noise always accompanies with blur. To test our model on more real images, we add Gaussian noise with *variance* = 0.001 to original GOPRO-Large dataset and create a new GOPRO-noise dataset with 1111 image pairs. A synthetic dataset in [9] is adopted for training. Same as combination version of DeblurGAN in [9], we use both GOPRO train dataset and synthetic dataset to train our network.

Training Process

The proposed network is trained on NVIDIA GeForce GTX 1080 Ti GPU and tested on Mac Pro with 2.7 GHz Intel Core i5 CPU. Similar to [9], the input training pair is randomly cropped as size of 256×256 after downsampled by a factor of two. Weights are initialized to follow Gaussian distribution with zeros mean and standard deviation 0.02. For each iteration of optimization, 1 step is performed on discriminator D , followed by 2 steps on generator G to prevent discriminator loss \mathcal{L}_d from zero. The model is trained for 15 epochs within 2 days, comparing with 200 epochs for 6 days in [9]. Furthermore, despite of instability GAN's training, our method converges to similar result for each and every training task, which demonstrates the robustness of our GAN architecture.

Table. 1 Average PSNR and SSIM.

Dataset	Metrics	[9]	$dc0$	$dc250$	$dc250p$
Original	PSNR	26.63	26.70	27.01	26.45
	SSIM	0.8701	0.8798	0.8813	0.8680
Noisy	PSNR	26.32	26.53	26.83	26.31
	SSIM	0.8524	0.8697	0.8707	0.8604

Result and Comparison

Our test results are mainly compared with state-of-art GAN based deblurring network DeblurGAN [9]. DeblurGAN defeats deep learning networks [7] and [8] on GOPRO dataset. Since the author posted the code online¹, we compare our network with DeblurGAN by directly adopting its uploaded network and latest trained weights. We test our model on GOPRO and GOPRO-noise test datasets.

Fig. 3 illustrates the deblurred results of [9] and our model. Blurry image in the first row is picked in GOPRO-noise dataset and the blurry one of second row is real natural image with motion blur taken by camera. According to local patches, although [9] can deal with blur but its results suffer from grid artifacts, while our model with dark channel loss achieves sharper images without grid artifacts. Furthermore, for motion blurry image (second row), the sharp part in input image remains unchanged in our deblurred result, but extra grid artifacts are added to result of [9].

The quantitative performance of the proposed network on two dataset GOPRO and GOPRO-noise is shown in Tab. 1. In our experiment, the coefficient of dark channel loss $\lambda_2 = 250(dc250)$. The results are compared with same network without dark channel loss $dc0$, same network with extra perpetual loss $dc250p$ as well as DeblurGAN [9]. All test images are downsampled by factor of two. The perpetual loss follows what it is in [10]. The proposed model performs best among the comparisons for both noise-free and noisy dataset. DeblurGAN performs less well owing to its grid artifacts. Perceptual loss leads to a worse result. Since GAN is good at preserving perceptual feature already, perceptual loss brings no extra constraints for the network. Comparison with $dc=0$ demonstrates that dark channel loss contributes to better result.

Conclusion

To address deblurring problem using a CGAN based architecture and to tackle the issue with grid artifacts in GAN based deblurring methods, this paper incorporates a dark channel prior. The dark channel prior is employed by L_2 norm rather than L_0 in order to make it more friendly for network training. To validate the deblurring result on more nature images, a noise involved dataset is proposed. The proposed network shows a great deblurring performance for both synthetic and real blurry images.

References

[1] Nicolas Dey, Laure Blanc-Feraud, C Zimmer, Z Kam, J-C Olivo-Marin, and J Zerubia. A deconvolution method for confocal microscopy with total variation regularization. In Biomedical Imaging: Nano to Macro, 2004. IEEE International Symposium on, pp. 1223–1226. (2004).

[2] Anat Levin, Rob Fergus, Fredo Durand, and William T Freeman. Image and depth from a conventional camera with a coded aperture. In ACM Transactions on Graphics (TOG), volume 26, pg. 70. (2007).

[3] Shan Qi, Jiaya Jia, and Aseem Agarwala. High-quality motion deblurring from a single image. In ACM Trans. Graph., volume 27, pg. 73. (2008).

[4] Jinshan Pan, Deqing Sun, Hanspeter Pfister, and Ming-Hsuan Yang. Blind image deblurring using dark channel prior. In CVPR, pg. 1628–1636, (2016).

[5] Ruiwen Zhen and Robert L. Stevenson. Multi-image motion deblurring aided by inertial sensors. Journal of Electronic Imaging, 25(1):013027–013027, (2016).

[6] Christian J. Schuler, Michael Hirsch, Stefan Harmeling, and Bernhard Scholkopf. Learning to deblur. IEEE Trans. Pattern Anal. Mach. Intell., 38(7):1439–1451, (2016).

[7] Jian Sun, Wenfei Cao, Zongben Xu, and Jean Ponce. Learning a convolutional neural network for non-uniform motion blur removal. In CVPR, pg. 769–777, (2015).

[8] Seungjun Nah, Tae Hyun Kim, and Kyoung Mu Lee. Deep multi-scale convolutional neural network for dynamic scene deblurring. pg. 3883–3891, (2017).

[9] Kupyn, Orest Budzan, Volodymyr Mykhailych, Mykola Mishkin, Dmytro Matas, and Jiri Matas. DeblurGAN: Blind Motion Deblurring Using Conditional Adversarial Networks. CVPR camera-ready (2018).

[10] Rolf Khler, Michael Hirsch, Betty Mohler, Bernhard Scholkopf, and Stefan Harmeling. Recording and playback of camera shake: Benchmarking blind deconvolution with a real-world database. In ECCV, pg. 27–40, (2012).

[11] Kaiming He, Jian Sun, and Xiaoou Tang. Single image haze removal using dark channel prior. In CVPR, pg. 1956–1963, (2019).

[12] Tae Hyun Kim, Seungjun Nah, and Kyoung Mu Lee. Dynamic scene deblurring using a locally adaptive linear blur model. arXiv preprint arXiv:1603.04265, (2016).

[13] Ayan Chakrabarti. A neural approach to blind motion deblurring. In ECCV, pg. 221–235. Springer, (2016).

[14] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, Yoshua Bengio. Generative adversarial nets. In NIPS, pg. 2672-2680, (2014).

[15] Mehdi Mirza, Simon Osindero. Conditional generative adversarial nets. CoRR, abs/1411.1784., (2014).

[16] Christian Ledig, Lucas Theis, Ferenc Huszar, Jose Caballero, Andrew Cunningham, Alejandro Acosta, Andrew Aitken, Alykhan Tejani, Johannes Totz, Zehan Wang, and Wenzhe Shi. Photo-realistic single image super-resolution using a generative adversarial network. arXiv preprint arXiv:1609.04802, (2016).

[17] Chuan Li and Michael Wand. Precomputed real-time texture synthesis with markovian generative adversarial networks. ECCV, (2016).

[18] Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A. Efros. Image-to-image translation with conditional adversarial networks. In CoRR, abs/1611.07004, (2016).

[19] Martin Arjovsky, Soumith Chintala, and Lon Bottou. Wasserstein GAN. ArXiv, arXiv:1701.07875v3, (2017).

[20] Justin Johnson, Alexandre Alahi, and Li Fei-Fei. Perceptual losses for real-time style transfer and super-resolution. In ECCV, (2016).

[21] Sergey Ioffe and Christian Szegedy. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In Proceedings of the 32nd International Conference on International Conference on Machine Learning - Volume 37 (ICML'15), Francis

¹<https://github.com/KupynOrest/DeblurGAN>

- Bach and David Blei (Eds.), Vol. 37. JMLR.org 448-456.(2015).
- [22] E. Shelhamer, J. Long and T. Darrell, Fully Convolutional Networks for Semantic Segmentation, in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 39, no. 4, pg. 640-651, (2017).
- [23] Bing Xu, Naiyan Wang, Tianqi Chen, Mu Li. Empirical evaluation of rectified activations in convolutional network. arXiv preprint arXiv:1505.00853, 2015.
- [24] Karen Simonyan, Andrew Zisserman. Very Deep Convolutional Networks for Large-Scale Image Recognition. ArXiv e-prints, Sept. (2014).
- [25] Abadi, Martn, Barham, Paul, Chen, Jianmin, Chen, Zhifeng, Davis, Andy, Dean, Jeffrey, Devin, Matthieu, Ghemawat, Sanjay, Irving, Geoffrey, Isard, Michael, Kudlur, Manjunath, Levenberg, Josh, Monga, Rajat, Moore, Sherry, Murray, Derek Gordon, Steiner, Benoit, Tucker, Paul A., Vasudevan, Vijay, Warden, Pete, Wicke, Martin, Yu, Yuan and Zhang, Xiaoqiang. TensorFlow: A system for large-scale machine learning. CoRR abs/1605.08695 (2016) .

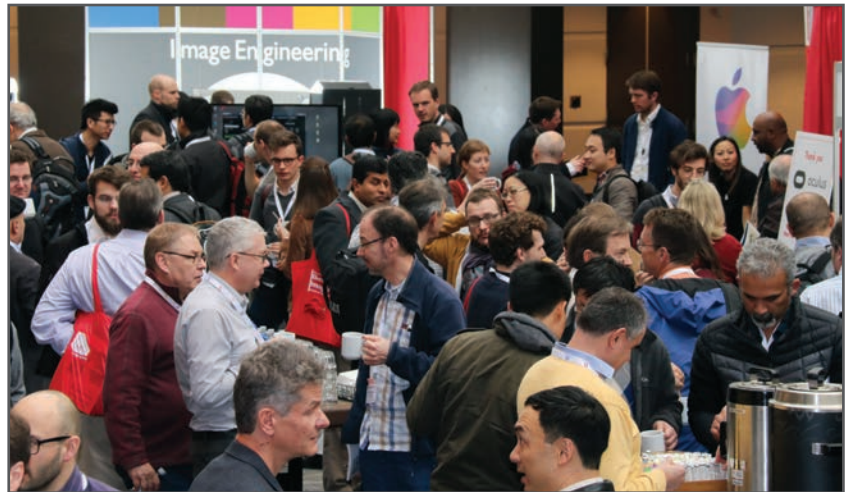
JOIN US AT THE NEXT EI!

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

