https://doi.org/10.2352/ISSN.2470-1173.2019.5.MWSF-528

How re-training process affect the performance of no-reference image quality metric for face images

Xinwei Liu [1,2], Christophe Charrier [1], Marius Pedersen [2], and Patrick Bours [2]
[1] Normandie Univ, UNICAEN, ENSICAEN, CNRS, GREYC, Caen, France
[2] NTNU - Norwegian University of Science and Technology, Gjøvik, Norway

Abstract

The accuracy of face recognition systems is significantly affected by the quality of face sample images. There are many existing no-reference image quality metrics (IQMs) that are able to assess natural image quality by taking into account similar image-based quality attributes. Previous study showed that IQMs can assess face sample quality according to the biometric system performance. In addition, re-training an IQM can improve its performance for face biometric images. However, only one database was used in the previous study, and it contains only image-based distortions. In this paper, we propose to extend the previous study by use multiple face database including FERET color face database, and apply multiple setups for the re-training process in order to investigate how the re-training process affect the performance of no-reference image quality metric for face biometric images. The experimental results show that the performance of the appropriate IQM can be improved for multiple databases, and different re-training setups can influence the IQM's performance.

Introduction

The face has become one of the most common and successful modalities for biometric recognition in the past decade [1]. However, face recognition is still a challenging issue when degraded face images are acquired. It has been proven that face sample quality has significant impact on accuracy of biometric recognition [2]. Low sample quality is a main reason for matching errors in biometric systems and may be the main weakness of some applications. Biometric image quality assessment approaches are used for measuring image quality and they may help to improve system performance. There are many existing IQMs that have been developed for the evaluation of natural image's quality [3]. According to the properties of face biometric images, only no-reference IQMs might be suitable for the assessment of face image quality. A previous study evaluated 13 no-reference IQMs by using a specific face database which contain only image-based distortions [4, 5]. The results from this study showed that one of the selected IQMs can assess face image quality according to the biometric system performance. In addition, the performance of this IQM can be improved by using high quality face images to re-train it [4, 5]. However, only one database was used in this study and there is only one re-training setup. Therefore, we propose to extend the experiments in [4, 5]: 1) use both high quality face and iris images for the re-training process, 2) use FERET color database [6] to conduct the experiments by using multiple re-training setups. From the results of extended experiments we can know 1) if different biometric modalities can affect the re-trained IQM, 2) whether the no-reference IQM can assess face quality from commonly used database and the performance of the IQM can be improved by re-training it. The structure of the paper is described as follows. We first present related works. Then the experimental setup followed by the experimental results, and then, their analysis are given. At last, the conclusion and future work are presented.

Related work

The existing no-reference IQMs can be grouped into two categories: distortion-specific IQMs and generalized IQMs. Different IQMs can be used for different purposes. Distortion-specific IQMs are designed for specific distortions, so they can performa well on certain distortions. On the other hand, generalized IQMs can assess different types of distortions; however, they may not perform as good as distortion-specific IQMs for a certain distortion. Moreover, some of IQMs are natural scene statistics (NSS)-based metrics which have been trained on image databases. Such IQMs can have better performance on images that similar to trained dataset, and vice versa.

There are several existing studies using no-reference IQMs to assess face sample quality. Abaza *et al.* [7] evaluated no-reference IQMs that can measure image quality factors in the context of face recognition. Then they proposed a face image quality index that combines multiple quality measures. Dutta *et al.* [8] proposed a data-driven model to predict the performance of a face recognition system based on image quality features. They modeled the relationship between image-based quality features and recognition performance measures using a probability density function. Hua *et al.* [9] investigated the impact of out-of-focus blur on face recognition performance. Fiche *et al.* [10] introduced a blurred face recognition algorithm guided by a no-reference blur metric. From these studies we can see that no-reference IQMs can be helpful to assess the quality of face samples. The observed performance is comparable to some metrics proposed in face ISO/IEC standard [11], which are designed specifically for face modality. Based on these studies, Liu *et al.* [4, 5] evaluated the performance of 13 no-reference IQMs on face biometric images. They discovered that the performance of one IQM can be improved by re-training it on face images instead of natural images. However, there was one re-training setup in this study, and only one database was used for evaluation. Thus, in order to better investigate 1) how the re-training setups affect the performance of such IQM, and 2) whether the performance can be improved when using other databases, we propose to extend their experiment in this paper.

Experimental setup

In this paper, we use two face databases. The first one is the face image database named 'GC^2–multi-modality biometric image quality database', which was used in [4]. This database has three biometric modalities: contactless fingerprint, VW iris, and face. Three cameras are used for the acquisition: 1) a Lytro first generation Light Field Camera (LFC) (11 Megapixels), 2) a Google Nexus 5 embedded cam- era (8 Megapixels), and 3) a Canon D700 with Canon EF 100mm f/2.8L Macro Lens (18 Megapixels). 50 subjects participated in the acquisition. For the face modality, 2250 raw face images are obtained in the database. In addition, different types of image-based degradations are introduced in the database. Each face image is degraded into five distortion levels (one to five, from little degraded to highly degraded) for eight degradations [4, 5, 12]: high and low contrast distortions, motion blur, Gaussian blur, high and low luminance distortions, Poisson noise, and JPEG compression artifacts. Including the degraded sample images, there are 92250 face images. In addition to the GC^2 database, we use another face database: FERET color face database [13]. It has 269 subjects and there are two acquisition sessions for most of subjects. For each session, 11 different sample images were acquired which contain different face angles and expressions.

One no-reference IQM, ILNIQE2 [14], was evaluated as the best performed metric from 15 selected IQMs in [4, 5]. After re-trained it on high quality face images, its performance can be improved when assessing face quality for GC^2 database. In this paper, we extend this experiment to re-train ILNIQE2 on both high quality face and iris images. Moreover, we evaluate the performance of ILNIQE2 on FERET face database, and then re-train it on 1) high quality face images, 2) both high quality face and iris images. By analyzing the experimental results mentioned above, we can understand how re-training process affect the performance of no-reference image quality metric for face images.

The open source face recognition system use in this paper is 'The PhD (Pretty helpful Development functions for) face recognition toolbox' [15], which is a collection of Matlab functions and scripts for face recognition. The toolbox was produced as a byproduct of Štruc and Pavešić's [16] research work and is freely available for downloading.

An IQM is useful if it can at least give an ordered indication of an eventual performance [17]. Rank-ordered detection error trade-off (DET) characteristics curve is one of the most commonly used and widely understood method used to evaluate the performance of quality assessment approaches. The DET curve used here plots false none match rate (FNMR) versus false match rate (FMR). Grother and Tabassi [17] proposed to use quality-bin-based approaches to evaluate the image quality assessment methods. They believe if a certain percentage of low quality samples are excluded from the dataset, the comparison score would become ?better? (closer to 1 in our case) and the equal error rate (EER) (when FMR and FNMR are equal) would decrease. We use it as one of the methods to represent the performance of ILNIQE2. We omit the percentile low quality samples and keep 80%, 60%, and 40% of highest quality samples from each subject for ILNIQE2 to evaluate its performance [18].

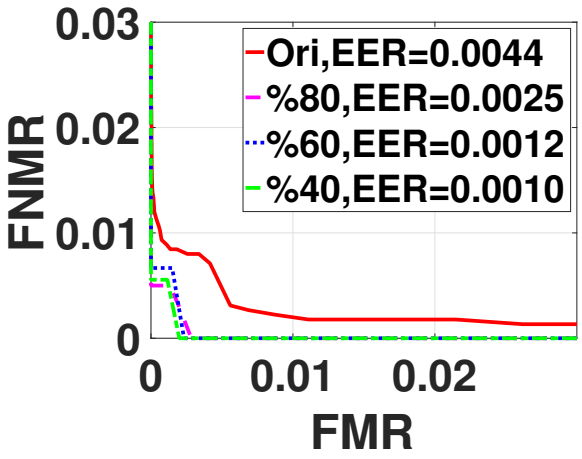


Figure 1. *DET curves with EER for comparison score with and without omitting low quality samples.*

Experimental results

Extended experiments using GC^2 database

Instead of using only face image to re-train ILNIQE2, here we use both face and iris images re-training ILNIQE2 to investigate if biometric modality can influence the re-training process. The DET curves with EER for data with and without omitting low quality face samples from GC^2 database for reflex cameras by using ILNIQE2 are given in Fig. 1. Here we only illustrate the results from reflex camera as examples in this paper. The red continuous line represents the original DET curve; the magenta '--' line represents the DET curve when we keep 80% highest quality face samples; the blue ':' line represents the comparison score when we keep 60% highest quality face samples; and the green '-.' line represents the comparison score when we keep only 40% highest quality face samples in the database for the experiment. If a DET curve is closer to the bottom-left point, it means that this set of data lead to a higher face recognition performance. Meanwhile, the lower EER value the better system performance. From Fig. 1 we can see that, DET curves shift closer to bottom-left point when we keep 80%, 60%, and 40% highest quality samples by using the assessment results from re-trained ILNIQE2 to omit low quality samples taken by reflex camera. EER values also decrease when more and more low quality face samples are omitted. It means that by using both face and iris image to re-train ILNIQE2, its performance is improved.

In order to compare the performance of re-trained ILNIQE2 between the setup in [4, 5] and in this paper, we use EER values by omitting lowest quality face sample one by one until only one highest quality face sample is left from each subject as an indicator. The comparison of the change of EER values between two setups is illustrated in Fig. 2. The x-axis in Fig. 2 represents the percentage of kept high quality face samples. The blue line represents the re-trained ILNIQE2 by using only face images, and the red line represents the re-trained ILNIQE2 using both face and iris images. The y-axis represents the EER value. If the EER value has a smooth decreasing tendency when we omit lowest quality samples one by one, it means that the metric used for generating the quality scores can predict the face recognition algorithm well, which represents the high performance of such IQM. From

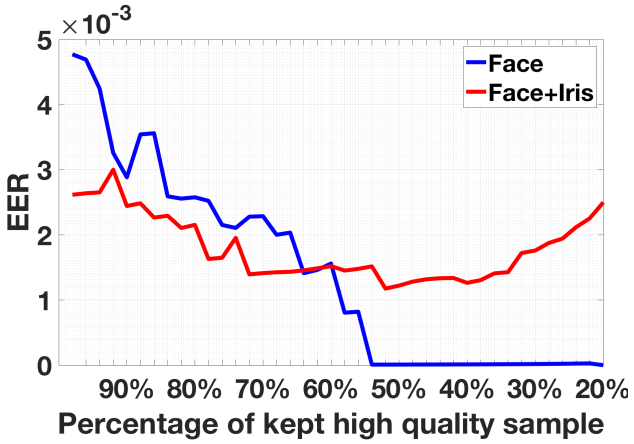


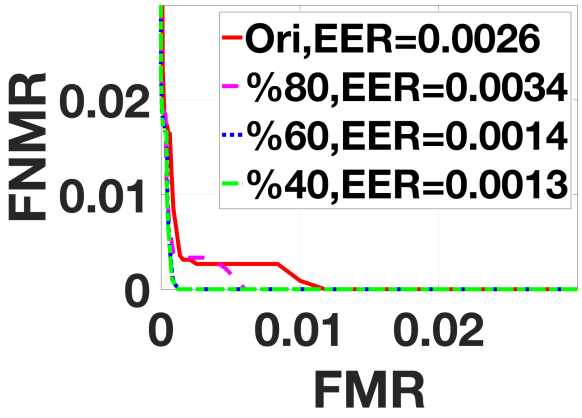
Figure 2. *Comparison of EER by omitting lowest quality sample one by one using re-trained ILNIQE2 for each subject between the two setups.*

Fig. 2 we can see that, the performance of re-trained ILNIQE2 by using face and iris images is better before 60% of high quality samples are still kept in the database. However, After half of the lowest quality face images are omitted, the EER becomes 0 for the blue line. The red line has a increasing at the end. It means that the overall performance of re-trained ILNIQE2 using face and iris is not as good as when only use face images for re-training. In another words, adding iris images for re-training process, the performance of the IQM can be decreased at some point.

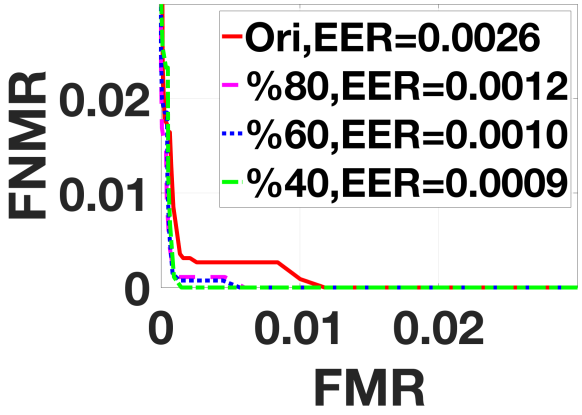
Experiments using FERET database

As introduced previously, we now use FERET color database to evaluate the performance of ILNIQE2 with and without re-training. The training datasets are 1) face images from GC^2 database, and 2) face and iris images from GC^2 database. We present in Fig. 3 the DET curves with EER for data with and without omitting low quality face samples from FERET color face database for reflex camera by using ILNIQE2. Fig. 3 (a) represents the result from original ILNIQE2, (b) represents the result from re-trained ILNIQE2 using face images, and (c) represents the result from re-trained ILNIQE2 using both face and iris images. From Fig. 3 we can see that, the ILNIQE2 under three setups can all assess face images quality because DET curves shift to the bottom-left point and EER values decrease, except omitting 20% lowest quality sample for ILNIQE2 without re-training. Both setups for re-training process can improve the performance of ILNIQE2 because when more and more low quality samples are omitted from the database, EER values in Fig. 3 (b) and (c) are always lower than (a). If we compare EER values in Fig. 3 (b) and (c) we can find out the performance of re-trained ILNIQE2 using face images is better at 80% and 60% percentage of kept high quality face samples.

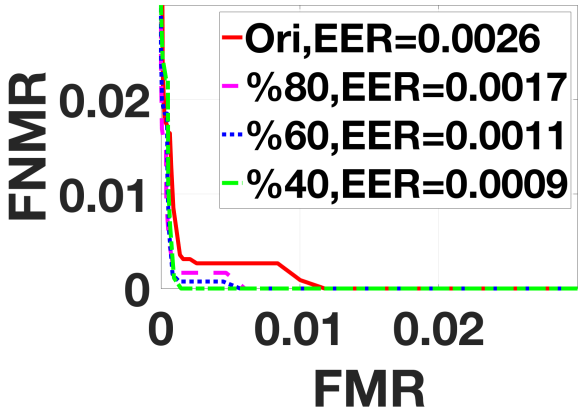
We illustrate EER values by omitting lowest quality face sample one by one until only one highest quality face sample is left from each subject from three setups in Fig. 4. The observation from Fig. 4 is similar to the findings in Fig. 3. The re-trained ILNIQE2 using only face images has the overall better performance. The performance of re-trained ILNIQE2 using both face and iris images has better performance than the original one, but becomes worse when only 40% of high quality samples are left in



(a) Original



(b) Face



(c) Face + iris

Figure 3. *DET curves with EER for comparison score with and without omitting low quality samples for ILNIQE2: (a) original, (b) re-trained using face images, and (c) re-trained using both face and iris images.*

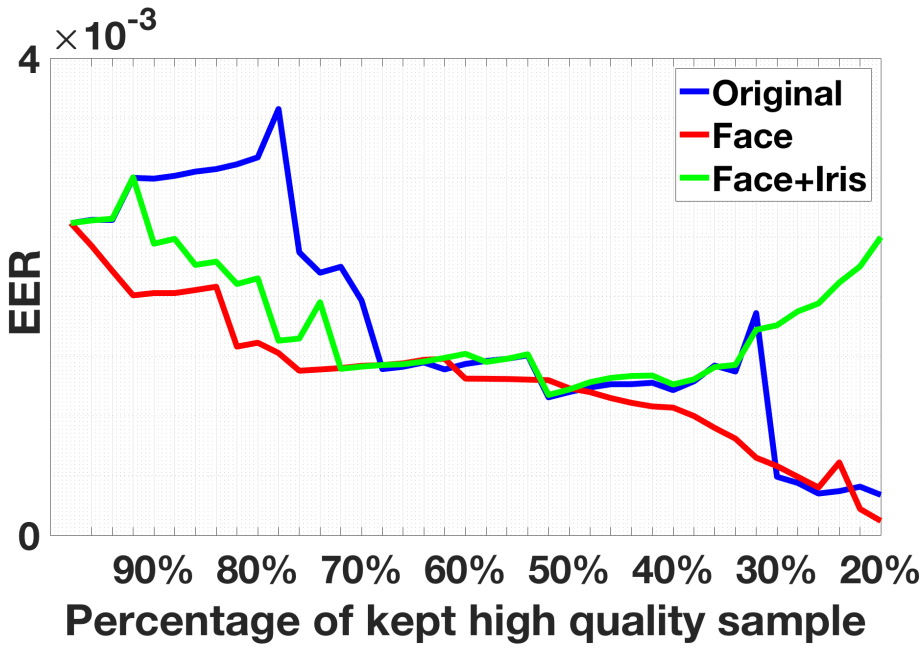


Figure 4. *Comparison of EER by omitting lowest quality sample one by one using re-trained ILNIQE2 for each subject between the three setups.*

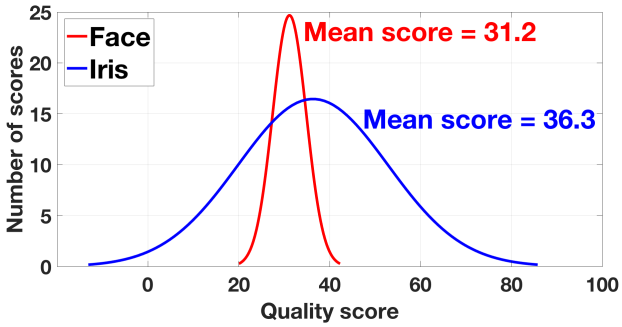


Figure 5. *Quality score.*

the database. In addition, EER values from three setups are very similar between 70% and 45% of high quality samples are left in the database.

In order to discover why the performance of re-trained ILNIQE2 decreased when adding iris images into the training dataset, we use original ILNIQE2 to assess the quality of face and iris images from the training dataset separately. The distribution of the quality scores are given in Fig. 5. The higher quality score represent higher image quality. As we can see from Fig. 5, the mean of the quality score for iris images is 36.3, which is higher than the mean score for face images. The distribution of the quality scores for iris images is also wider than face images. Therefore, when we included iris together with face images in the training dataset, the performance of re-trained ILNIQE2 has been affected. Because there are more high quality images (from iris images) were used for re-training, some of the high quality face images were recognized as low quality samples, so the performance of re-trained ILNIQE2 has been decreased.

Conclusion and future work

In this paper, we extend the experiment from a previous study and conduct new experiments to investigate how different re-training setups affect the performance of no-reference image quality metric for face images. Both face and iris images are used for the re-training process. From the experimental results we can find out that, the quality of training dataset can affect the performance of ILNIQE2 on face biometric images. It could be interested to apply the same protocol for the evaluation of no-reference IQMs on iris biometric images.

References

[1] Jain, Anil K., and Stan Z. Li. Handbook of face recognition. New York: springer, 2011.

[2] Gao, Xiufeng, Stan Z. Li, Rong Liu, and Peiren Zhang. ”Standardization of face image sample quality.” In International Conference on Biometrics, pp. 242-251. Springer, Berlin, Heidelberg, 2007.

[3] Pedersen, Marius, and Jon Yngve Hardeberg. ”Survey of full-reference image quality metrics.” (2009).

[4] Liu, Xinwei, Marius Pedersen, Christophe Charrier, and Patrick Bours. ”Performance evaluation of no-reference image quality metrics for face biometric images.” Journal of Electronic Imaging 27, no. 2 (2018): 023001.

[5] Liu, Xinwei. ”Multi-modality quality assessment for unconstrained biometric samples.” PhD thesis (2018).

[6] Phillips, P. Jonathon, Hyeonjoon Moon, Syed A. Rizvi, and Patrick J. Rauss. ”The FERET evaluation methodology for face-recognition algorithms.” IEEE Transactions on pattern analysis and machine intelligence 22, no. 10 (2000): 1090-1104.

[7] Abaza, Ayman, Mary Ann Harrison, Thirimachos Bourlai, and Arun Ross. ”Design and evaluation of photometric image quality measures for effective face recognition.” IET

Biometrics 3, no. 4 (2014): 314-324.
[8] Dutta, Abhishek, Raymond Veldhuis, and Luuk Spreeuwers. "Predicting face recognition performance using image quality." arXiv preprint arXiv:1510.07119 (2015).
[9] Hua, Fang, Peter Johnson, Nadezhda Sazonova, Paulo Lopez-Meyer, and Stephanie Schuckers. "Impact of out-of-focus blur on face recognition performance based on modular transfer function." In Biometrics (ICB), 2012 5th IAPR International Conference on, pp. 85-90. IEEE, 2012.
[10] Fiche, Ccile, Patricia Ladret, and Ngoc-Son Vu. "Blurred face recognition algorithm guided by a no-reference blur metric." In Image Processing: Machine Vision Applications III, vol. 7538, p. 75380U. International Society for Optics and Photonics, 2010.
[11] International Organization for Standardization (ISO), "Information technology–biometric sample quality–part 5: face image data," ISO/IEC TR 29794-5, International Electrotechnical Commission (IEC) (2010).
[12] Liu, Xinwei, Marius Pedersen, and Jon Yngve Hardeberg. "CID: IQ-a new image quality database." In International Conference on Image and Signal Processing, pp. 193-202. Springer, Cham, 2014.
[13] "FERET color face database," https://www.nist.gov/programs-projects/face-recognition-technology-feret, Accessed: 2017-12-02.
[14] Zhang, Lin, Lei Zhang, and Alan C. Bovik. "A feature-enriched completely blind image quality evaluator." IEEE Transactions on Image Processing 24, no. 8 (2015): 2579-2591.
[15] Štruc Vitomir, "The phd toolbox: Pretty helpful development functions for face recognition," 2012.
[16] Štruc, Vitomir, and Nikola Pavešić. "The complete gabor-fisher classifier for robust face recognition." EURASIP Journal on Advances in Signal Processing 2010, no. 1 (2010): 847680.
[17] Grother, Patrick, and Elham Tabassi. "Performance of biometric quality measures." IEEE transactions on pattern analysis and machine intelligence 29, no. 4 (2007): 531-543.
[18] Poh, Norman, Samy Bengio, and Arun Ross. Revisiting Doddington" s Zoo: A Systematic Method to Assess User-dependent Variabilities. No. EPFL-REPORT-83320. IDIAP, 2006.

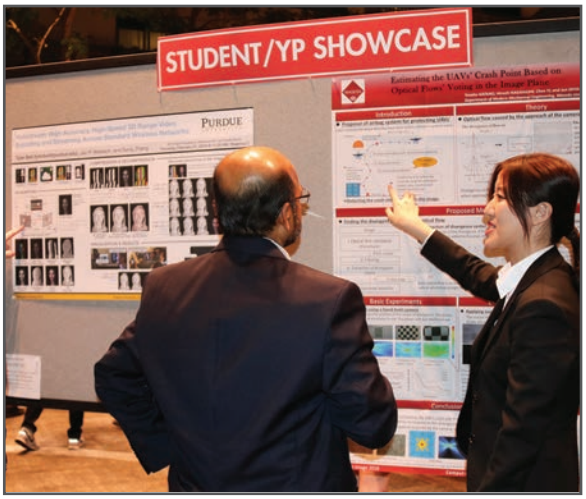
JOIN US AT THE NEXT EI!

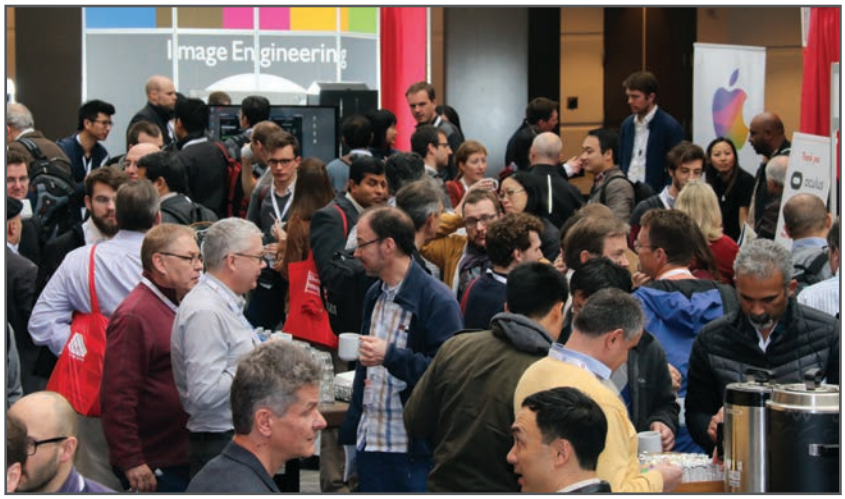
IS&T International Symposium on

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