https://doi.org/10.2352/ISSN.2470-1173.2019.12.HVEI-213

NARVAL, A No-Reference Video Quality Tool for Real-Time Communications

Augustin Lemesle, [§†] ***Alexis Marion,*** [§†] ***Ludovic Roux*** [§] ***and Alexandre Gouaillard*** [§]
[§] ***CoSMo Software Consulting, Singapore;***
[†] ***CentraleSupélec, France***

Abstract

In this paper we introduce two new no-reference metrics and compare their performance to state-of-the-art metrics on six publicly available datasets having a large variety of distortions and characteristics. Our two metrics, based on neural networks, combine the following features: histogram of oriented gradients, edges detection, fast fourier transform, CPBD, blur and contrast measurement, temporal information, freeze detection, BRISQUE and Video BLIINDS. They perform better than Video BLIINDS and BRISQUE on the six datasets used in this study, including one made up of natural videos that have not been artificially distorted. Our metrics show a good generalization as they achieved high performance on the six datasets.

Introduction

Video traffic over Internet or on mobile networks is experiencing a huge increase in the volume of data transferred. According to the Cisco Visual Networking Index released in June 2017 [1], live video traffic (streaming, video conferencing) should grow dramatically from 3% of Internet video traffic in 2016 to 13% by 2021, which translates as 1.5 exabytes[1] per month in year 2016, growing to 24 exabytes per month in year 2021.

As for any application that deals with video, Quality of Experience (QoE) for the end user is very important. Many tools and metrics have been developed to assess automatically the QoE for video applications. Measurements provided by several metrics on the same video can be combined together to provide a more robust assessment of video. A good example of this approach has been proposed by Netflix with their Video Multimethod Assessment Fusion (VMAF) metric [2]. This metric helps to assess routinely and objectively the quality of thousands of videos encoding with dozens of encoding settings. But it requires the availability of the original reference non distorted video to compute the quality score of the same video distorted by video compression. This method, well adapted to video streaming where original non distorted video is available, cannot be applied to real-time communication where the original video is not available.

In the recent years, several authors have proposed new metrics or a combination of features to assess video quality without needing any reference. These methods are referred to as no-reference (NR) metrics. Usually, these techniques have been evaluated only on a single dataset. Some studies have applied metrics to two or three different datasets, which gives a first hint at their capacity to generalize to a larger variety of distortions than those of the initial dataset on which they were developed.

[1] 1 exabyte = 1 million terabytes

In this paper, we propose to evaluate further the ability to generalize of current state-of-the-art NR video quality assessment metrics by using them on six publicly available datasets, including one made up of natural videos that have not been artificially distorted. These datasets exhibiting very different distortion types.

In addition, we propose NARVAL and BB-NARVAL, two NR video assessment methods designed as a combination of several NR metrics and aiming at being robust and able to generalize to many types of distortions. NARVAL, which stands for Neural network-based Aggregation of no-Reference metrics for Video quAlity evaLuation, relies on a dense neural network to perform the aggregation of the measures provided by eight individual no-reference metrics. BB-NARVAL adds BRISQUE and Video BLIINDS to NARVAL to enhance its performances.

We used BB-NARVAL to assess the quality of videos relayed by several WebRTC media servers during a video conferencing load test comparative study [3]. The test consisted in creating rooms with 7 participants until the goal of 70 rooms (490 participants) was reached. The objective was to compare how each media server were able to handle the load and how the video quality they deliver evolves as the number of video conferences increased. The results showed that BB-NARVAL consistently gave lower scores as the quality of videos worsen towards the end of the test.

The contribution of this paper is threefold. First we introduce two NR video quality tools, then we evaluate the performance of our tools on various video datasets. Finally, we give an insight of the capacity of our tools and of state-of-the-art NR video quality metrics to generalize their measurements to different types of distortions.

Previous Work

A comprehensive and detailed review of NR video quality metrics has been published in 2014 [4]. Video quality assessment techniques can be classified into three categories: full-reference (FR) techniques which require full access to the reference video, reduced-reference (RR) techniques which need a set of coarse features extracted from the reference video, and no-reference (NR) techniques which do not require any reference video.

For a long time, NR pixel-based features have measured degradation of visual quality generated by distortions such as blur or blocking. Blur is a reduction of edge sharpness. Cumulative Probability of Blur Detection (CPBD) [5] is a metric developed for detecting blur in an image. Sharpness can be measured by using variance of Laplacian. This technique has been used in [6] as a measure to control auto-focusing, assuming that a well focused image has sharper edges. In [7], Fast Fourier Transform gives a measure of the image sharpness. Another blur measurement, implemented using the method proposed in [8], first detects the edges in a frame then evaluates whether or not they are blurred.

Blocking is a discontinuity between adjacent blocks resulting from block-based processing of compression algorithms. Blockiness-blurriness (BB) is a measure for JPEG compressed images as described in [9].

More recently, there have been several new NR metrics developed around the concept of Natural Scene Statistics (NSS). NSS rely on the fact that natural images exhibit a common statistical behaviour which is known to be modified by image and video compression algorithms. Measurement of deviations from regular NSS allows to estimate the perceptual quality of an image. Metrics developed using NSS approach aim at being independent from the type of distortion that impair images, thus being more generic and applicable to a wide range of distorted images and videos. In this category of metrics, we can cite BRISQUE (Blind/Referenceless Image Spatial QUality Evaluator) [10], Natural Image Quality Evaluator (NIQE) [11], Video BLIINDS (BLind Image Integrity Notator using DCT Statistics) [12] and VIIDEO (Video Intrinsic Integrity and Distortion Evaluation Oracle) [13].

BRISQUE [10] operates in the spatial domain. It uses scene statistics of locally normalized luminance coefficients to quantify losses of naturalness in the image due to distortions. It has been evaluated on two datasets, LIVE IQA (Laboratory for Image & Video Engineering — Image Quality Assessment) and TID2008 (Tampere Image Database).

NIQE [11] is based on a collection of NSS features derived from a corpus of natural undistorted images. NIQE is completely blind in the sense that it relies only on measurable deviations from statistical regularities observed in natural images without training on human-rated distorted images. It has been evaluated on LIVE IQA dataset. The authors report NIQE is statistically better than BRISQUE.

Video BLIINDS [12] aims at predicting the quality of videos. It is based both on a spatio-temporal NSS of video scenes in the Discrete Cosine Transform (DCT) domain and on a model that quantifies motion coherency occurring in the scenes. It has been evaluated on two video datasets, LIVE VQA (Video Quality Assessment) and EPFL-PoliMi (École Polytechnique Fédérale de Lausanne — Politecnico di Milano).

VIIDEO [13] is based on NSS regularities that are observed in the space-time domain of natural videos. Inter subband correlations are used to quantify the degree of distortion present in the video. It has been evaluated on LIVE VQA dataset.

Some authors have proposed to use a combination of metrics for VQA. In [14], the authors propose the FC (Feature Combination) model, a combination of five features to perform video quality evaluation, namely contrast, colorfulness, blurriness, spatial information and temporal information. The authors report that FC gives a better video quality assessment than VIIDEO on KoNViD-1k dataset, and has a performance nearly as good as Video BLIINDS.

The same authors propose in [15] an extension of FC: the STFC model (Spatio Temporal Feature Combination). It is the combination of six spatial features: contrast, colorfulness, exposure, sharpness, spatial information and temporal information, and three temporal features: temporal contrast, temporal exposure and temporal sharpness. The STFC model performs better than VIIDEO and Video BLIINDS on KoNViD-1k dataset.

In [16], the authors evaluate eight NR metrics on videos impaired by compression and transmission over lossy networks (0 to 10% packet loss): scene complexity (number of objects or elements present in the frame), video motion (amount of movement), blockiness (discontinuity between adjacent blocks), jerkiness (non-fluent and non-smooth presentation of frames), average blur, blur ratio, average noise and noise ratio. They added two network measurements (bit rate and level of packet loss) for providing an improved NR metric. Values computed by these ten metrics are then combined using several machine learning techniques trained in a supervised way. For this experiment, they used ten videos obtained from the Live VQA dataset. These videos have been compressed at eight different levels using H.264, and impaired by transmission over a network with twelve packet loss rates. They assessed the quality of their results against the scores given by the FR metric Video Quality Metric (VQM) [17], but not against NR metrics.

A study of the performance of Video BLIINDS and VIIDEO on four datasets, namely VQEG Phase I, LIVE VQA, ReTRIeVED and a dataset generated by the authors in a mobile phone network simulator is reported in [18]. They conclude that VIIDEO performed better than Video BLIINDS, however none of them are reliable when applied to the practical application as the scores given to bad quality videos can be higher than the score given to better quality videos.

The paper [19] presents a combination of audio and video metrics to assess audio-visual quality. The assessment has been performed on two different datasets. For the NR video metrics, they used a blockiness-blurriness metric [9], BRISQUE [20], the Blind Image Quality Index (BIQI) [21] and NIQE [11]. The best combination for both datasets is the blockiness-blurriness.

Datasets

All videos in the following datasets have been scored by several people using the Absolute Category Rating (ACR) system as defined by International Telecommunication Union (ITU) [22]. From the ACR, the two scores that are most often computed are the Mean Opinion Score and the Differential Mean Opinion Score. In the following datasets, some provide MOS only, some provide DMOS only, and some provide both.

As shown in Table 1 the datasets have a heterogeneous number of reference videos and distortions. A wide range of distortions allows us to test the robustness of metrics regarding distortions but leaves us with few videos per distortion.

Table 1. Number of reference and distorted videos

Distortions	**Datasets**					
	IRCCyN	*KoNViD*	*ReTRiEVED*	*CSIQ*	*LIVE*	*Mobile*
References	24	1200	12	8	10	10
* Distortions	6		18	22	15	20
* Types of distortions	1		6	4	4	5
Videos	192	1200	228	184	160	210

*Per reference video

IRCCyN IVC 1080i

The IRCCyN IVC 1080i video quality dataset contains 24 high quality reference videos sequences of 9 to 12 seconds in 1920×1080 i50 resolution. They have been encoded with H.264 using 7 different bitrates. So, in total, the dataset contains 192 videos. Individual votes and MOS scores obtained by ACR and are provided. [23]

KoNViD-1k

This is a random collection of 1,200 videos selected from the YFCC100m (Yahoo Flickr Creative Commons 100 Million Dataset [24]), a dataset made up of 100 million media objects, of which 793,436 are videos that have been uploaded to Flickr between 2004 and 2014. Unlike traditional datasets, videos collected in KoNViD are natural videos that have not been artificially distorted. Subjective scores of all videos in the KoNViD-1k dataset were obtained by a well-designed crowd-sourcing experiment of ACR judgments. [25]

CSIQ Video Quality Dataset

The CSIQ video dataset consists of 12 reference videos and 216 distorted videos with six different types of distortion (compression with H.264, H.265, Motion JPEG and the Snow codec wavelet-based compression, simulated transmission loss for the H.264 encoded videos, and addition of white noise). All the videos are in the YUV420 format at a resolution of 832×480 and duration of 10 seconds. [26]

ReTRiEVED Video Quality Dataset

The dataset contains 184 test videos obtained from 8 source videos of different content characterized by wide span of spatial and temporal information and different motions. Test video sequences was generated by considering practical transmission scenario by using Network Emulator (NETEM) and Video LAN. Packet loss rate, jitter, delay, and throughput have been considered as possible distortions resulting from video transmission and its value is considered based on ITU and ETSI recommendations. [27] [28]

LIVE VQA (Video Quality Assessment)

The LIVE VQA uses ten uncompressed high-quality videos with a wide variety of content as reference videos. A set of 150 distorted videos were created from these reference videos (15 distorted videos per reference) using four different distortion types: MPEG-2 compression, H.264 compression, simulated transmission of H.264 compressed bitstreams through error-prone IP networks and through error-prone wireless networks. Each video was assessed by 38 human subjects in a single stimulus study with hidden reference removal where the subjects scored the video quality on a continuous quality scale. [29] [30]

LIVE Mobile VQA

The LIVE Mobile VQA dataset consists of 10 RAW HD reference videos and 200 distorted videos (4 compression + 4 wireless packet-loss + 4 frame-freezes + 3 rate-adapted + 5 temporal dynamics per reference), each of resolution 1280 × 720 at 30 fps, and of duration 15 seconds each. The dataset includes the DMOS computed from the ratings that the subjects provided at the end of each video clip. [31] [32] [33]

Metrics

We have evaluated the performance of 12 metrics on the 6 datasets. Nine of these metrics are framewise features, namely:

- *Spatial Information*: As described in ITU standard P.910 [34], spatial information is the standard deviation of an image filtered by a Sobel filter across its two axis.
- *Histogram of Oriented Gradients*: Instead of calculating the standard deviation after a Sobel filter, we can construct an histogram of the gradient magnitude and angle of each pixel using a horizontal and vertical Sobel filter as explained in [35].
- *Edge detection*: We use the Sobel filter in collaboration with a Canny Edge detection to find the edges in a frame and compute the number of edges in a frame.
- *Variance of Laplacian*: Using a Laplacian filter, we compute the variance of the filtered image as a feature for our video frame [6].
- *Fast Fourier Transform*: By performing a Fourier Transform on the video frame, we get a measure of the image sharpness with the method described in paper [7].
- *CPBD*: This is a metric proposed by [5] for detecting blur in an image.
- *Blur measurement*: Another blur measurement, implemented using the method described in [8], finding the edges in a frame and examining whether or not they are blurred.
- *Colorfulness*: The colorfulness of a frame is determined using the method introduced in [36].
- *Contrast*: The contrast of a frame is measured using [37], by calculating the standard deviation of pixel gray-scale intensities.

Then in order to detect jitter, freeze and other temporal distortions, we measured three features based on the differences between frames:

- *Temporal Information*: Defined in the same ITU standard P.910 [34], the temporal perceptual information is the standard deviation of the difference between two consecutive frames of the video.
- *Mean squared difference*: With the difference between two consecutive frames, we can also compute the mean squared difference of those frames.
- *Freeze detection*: As described in [38], we measure whether or not a sequence of frames is frozen when the mean square error of two consecutive frames and with the first frame of the freeze sequence are under a threshold. Empirically we set the threshold to 0.1. Afterwards we return 1 if the frame is found frozen and 0 otherwise.

NARVAL

NARVAL is based on a neural network. We chose each layer activation function empirically to produce the best result in the same way as we chose the number of layers and the neurons in each layer. The final structure is a 4 layers fully connected network. Input layer is connected to a first hidden layer of 1000 neurons with ReLU activation function. A second hidden layer contains 500 neurons and use tanh as activation function. The third hidden layer has 250 neurons and ReLU activation function. At last we reach the output layer which has a single neuron with linear activation function. The output of the network is a score. Between each layer we placed a dropout layer with a dropout probability of 0.2.

To train the network, we used a 5-fold validation method and the mean squared error.

We used as input a function of each feature and metric per video thus eliminating the temporal dimension of the features. For most of the inputs, the mean and the standard deviation of the input were selected, however for some the maximum was a better choice.

After testing exhaustively every combination of features, we retained the following 8 best features for our metric:

- Histogram of Oriented Gradients,
- Edges detection,
- Fast Fourier Transform,
- CPBD,
- Blur measurement,
- Contrast measurement,
- Temporal Information,
- Freeze detection.

BB-NARVAL

BB-NARVAL is a combination of NARVAL and of two NSS metrics calculated on different color-spaces:

- BRISQUE on Y, UV and RGB, with 36 features per frame,
- Video BLIINDS with 46 features per video,
- NARVAL.

BB-NARVAL stands for BRISQUE-BLIINDS-NARVAL.

Experiments

There are two main parts in our work: first, the extraction of features from videos, then the supervised training of a model to predict a score for a given video.

For the feature extraction part, we selected metrics and features published and evaluated on different image quality datasets. After calculating the value given by each metric on the videos of the 6 datasets, we stored the data to be able to reuse them in the training part.

For the second part, we used a fully connected neural network to do a regression. We trained our model on the datasets using a 5-fold fit and then repeating the training multiple times. As each dataset contains multiple distortions, we cannot just split the folds randomly, thus we tried to choose the 5 folds so that all distortions exist in a fold and we kept the same distribution for all tests. Only the mean of the folds will then be taken into account.

All the results displayed are averaged over multiple trainings of the model as well as over the folds used during training. The network has been trained with MOS for IRCCyN, KoNViD and ReTRiEVED and with DMOS for CSIQ, LIVE and Mobile.

Using our model, we first tested our 8 features on each dataset individually. For these features, we found that using the mean and standard deviation values over the video gave the best results. Table 2 shows the results of the training on each feature alone. From this table, we see multiple tendencies, the first being that the blur measurement is the most efficient feature alone for almost all datasets. Then, we see that temporal information and freeze detection are more efficient on KoNViD, ReTRiEVED and also Mobile even if it is less visible. This is expected as ReTRiEVED and Mobile are the only datasets with temporal distortion, and KoNViD is a collection of random videos from the Internet that may present temporal distortions.

Table 2. Pearson correlation coefficient of features on the different datasets

Metrics	Datasets					
	MOS			*DMOS*		
	IRCCyN	*KoNViD*	*ReTRiEVED*	*CSIQ*	*LIVE*	*Mobile*
Histogram	0.459	0.240	0.703	0.315	0.218	0.327
FFT	0.094	0.123	0.601	0.123	0.219	0.175
CPBD	0.462	0.399	0.470	0.376	0.128	0.270
Edges	0.313	0.401	0.790	0.326	0.114	0.216
Contrast	0.110	0.451	0.415	0.155	0.181	0.188
Blur	**0.524**	0.473	**0.816**	**0.478**	**0.292**	**0.417**
TI	0.114	**0.530**	0.764	0.126	0.221	0.330
Freeze	0.120	0.491	0.633	0.156	0.235	0.335

Overall, the correlation scores are rather low as some features alone are insufficient for the network to be trained. But when they are combined into NARVAL and BB-NARVAL, they reinforce each other and give much better scores. We will also compute scores given by Video BLIINDS and BRISQUE on the six datasets and compare them to NARVAL and BB-BARVAL.

Table 3. Processing functions for the metrics

Metric	Type and number of features	
	Video-wise	*Frame-wise*
BRISQUE	max	36
Video BLIINDS	46	-
BB	mean,std	3
NIQE	mean, std	1
VIIDEO	1	-

For this network, we had to process the data before the training as described in Table 3 to squeeze the temporal dimension. Each metric was calculated on the gray-scaled video. The results can be found in Table 4 where we also added the best feature as a comparison. The aim with the results in Table 4 is to see whether or not a metric can be generalized to many datasets.

After testing metrics individually and NARVAL, we tried to combine existing metrics and their combination in different colorspaces to improve our results thus creating our second metric: BB-NARVAL. Some results of our tests can be found in Table 5. On IRCCyN the training still fails approximately 1 out of 5 times and in any case it takes more time to train. The results presented here for IRCCyN come only from trainings that did work.

Other combinations of metrics were tested such as adding NIQE or VIIDEO but the results were not satisfying. Indeed, adding NIQE just plainly worsen the results and adding VIIDEO improved LIVE correlation a little while deteriorating KoNViD and CSIQ correlations. We tried to look for more global measurement not only improving the correlation on one dataset so we did

Table 4. Pearson correlation coefficient of metrics on the different datasets

Metrics	Datasets					
	MOS			*DMOS*		
	IRCCyN	*KoNViD*	*ReTRiEVED*	*CSIQ*	*LIVE*	*Mobile*
Blur	0.524	0.473	0.816	0.816	0.292	0.417
NIQE	0.511	0.223	0.777	0.328	0.290	0.314
VIIDEO	0.131	0.208	0.732	0.304	0.622	0.351
BB	0.838	0.595	0.918	0.414	0.236	0.858
BRISQUE	0.808	0.671	0.923	0.674	0.417	0.471
BLIINDS	0.149	**0.742**	0.923	**0.825**	**0.711**	0.885
NARVAL	**0.956**	0.689	**0.924**	0.768	0.601	**0.886**

Table 5. Pearson correlation coefficient of the proposed metric on the different datasets

Metrics	Datasets					
	MOS			*DMOS*		
	IRCCyN	*KoNViD*	*ReTRiEVED*	*CSIQ*	*LIVE*	*Mobile*
BRISQUE	0.808	0.671	0.923	0.674	0.417	0.471
BLIINDS	0.149	0.742	0.923	0.825	0.711	0.885
NARVAL	**0.956**	0.689	0.924	0.768	0.601	0.886
BB-NARVAL	0.929	**0.761**	**0.954**	**0.882**	**0.890**	**0.927**

not include VIIDEO in BB-NARVAL.

Discussion

Firstly, all the datasets differ according to the type of distortions proposed but also to the methods used, the video used or the amplitude of the distortions. For instance, the ReTRIEVED dataset proposes a wider amplitude in the intensity of the distortion thus making some videos so much distorted that it becomes difficult to understand the content while IRCCyN or LIVE datasets show milder distortions. Thus, between two datasets the same DMOS can rate very different impairments.

Thus we can remark that among the six datasets, ReTRiEVED has the highest correlations while LIVE has the lowest. Indeed, ReTRiEVED has a high variance of MOS scores while, on the other hand, LIVE has the second lowest scores variance of all datasets after KoNViD.

In regards to metrics, NIQE is not giving satisfying results on any of the datasets. On ReTRiEVED and IRCCyN, the correlation is slightly better but still inferior to blur feature alone, and it remains very low as compared to other correlations on the ReTRiEVED dataset.

On any other datasets than LIVE for which it has been designed for and ReTRiEVED which is easier to train on, VIIDEO shows a significantly lower performance than other metrics. For other datasets, it presents the lowest results and the network almost does not learn anything from that metric.

On the other hand, BRISQUE performs relatively well except on LIVE and Mobile, the common point been these two datasets being the wireless distortions. The blockiness-bluriness metric shows similar performance as BRISQUE on IRCCyN and ReTRiEVED and it has a much better correlation on Mobile. However on LIVE and CSIQ this metric has poor results maybe because of their MPEG-2 distortions.

Video BLIINDS gives the best results overall except on IRCCyN where our network trained with the mean squared error did not manage to learn anything.

Our metric NARVAL also performs well on all datasets giving the same score as BLIINDS on ReTRiEVED and Mobile while also producing a very high correlation on IRCCyN. On the other datasets, NARVAL is behind Video BLIINDS but above BRISQUE or BB.

Table 5 shows that BB-NARVAL is better than BLIINDS and NARVAL on every dataset except IRCCyN where Video BLIINDS has a very negative impact on the training. Overall, BB-NARVAL performs well or very well on all the six datasets.

Finally, for all the metrics, the correlation on the KoNViD dataset is lower than on all or most of the other datasets. We can explain this by the fact that the videos from KoNViD have few similarities of content between them as that they are natural videos randomly selected from the Internet. Moreover, the standard deviation of their scores is very low.

Conclusion

Real time video communication needs robust and widely applicable objective video quality assessment tools. In this paper, we have introduced two new NR metrics, NARVAL and BB-NARVAL, to evaluate the quality of real-time videos. These two metrics exhibit good performance and outperform Video BLIINDS or VIIDEO on six publicly available video datasets. The wide spectrum of distortions and characteristics of the videos of these datasets shows that our metrics are able to generalize.

In future works, we plan to add sound quality assessment to complete our tool. We will also need to apply our metrics to a greater number of challenging larger datasets of naturally distorted videos to be able to generalize our model to real videos without man-made distortions. KoNViD is an example of such a dataset and we can already see some limits to our model as it has quite a low correlation on it.

We used BB-NARVAL to assess the quality of videos relayed by several WebRTC media servers during a video conferencing load test comparative study [3]. We project to further enhance our model to deal with more real videos and to apply our tool to real-time communication applications.

References

[1] Cisco. (2017, June) Visual networking index. Cisco. [Online]. Available: https://www.cisco.com/c/en/us/solutions/collateral/service-provider/visual-networking-index-vni/vni-hyperconnectivity-wp.html

[2] Z. Li, A. Aaron, I. Katsavounidis, A. Moorthy, and M. Manohara. (2016, June) Toward a practical perceptual video quality metric. Netflix. [Online]. Available: https://medium.com/netflix-techblog/toward-a-practical-perceptual-video-quality-metric-653f208b9652

[3] E. André, N. Le Breton, A. Lemesle, L. Roux, and A. Gouaillard, “Comparative study of WebRTC open source SFUs for video conferencing,” in *IPTComm 2018, Principles, Systems and Applications of IP Telecommunications*, Oct. 2018.

[4] M. Shahid, A. Rossholm, B. Lövström, and H.-J. Zepernick, “No-

reference image and video quality assessment: a classification and review of recent approaches," *EURASIP Journal on Image and Video Processing*, vol. 40, Aug. 2014.

[5] N. D. Narvekar and L. J. Karam, "A no-reference image blur metric based on the cumulative probability of blur detection (CPBD)," *IEEE Transactions on Image Processing*, vol. 20, no. 9, pp. 2678–2683, 2011.

[6] J. L. Pech-Pacheco, G. Cristóbal, J. Chamorro-Martínez, and J. Fernández-Valdivia, "Diatom autofocusing in brightfield microscopy: a comparative study," in *15th International Conference on Pattern Recognition (ICPR)*, vol. 3. IEEE, 2000, pp. 314–317.

[7] K. De and V. Masilamani, "Image sharpness measure for blurred images in frequency domain," *Procedia Engineering*, vol. 64, pp. 149–158, 2013.

[8] M. G. Choi, J. H. Jung, and J. W. Jeon, "No-reference image quality assessment using blur and noise," *International Journal of Computer Science and Engineering*, vol. 3, no. 2, pp. 76–80, 2009.

[9] Z. Wang, H. R. Sheikh, and A. C. Bovik, "No-reference perceptual quality assessment of JPEG compressed images," in *ICIP 2002, IEEE International Conference on Image Processing*, Sep. 2002.

[10] A. Mittal, A. K. Moorthy, and A. C. Bovik, "No-reference image quality assessment in the spatial domain," *IEEE Transactions on Image Processing*, vol. 21, no. 12, pp. 4695–4708, Dec. 2012.

[11] A. Mittal, R. Soundararajan, and A. C. Bovik, "Making a "completely blind" image quality analyzer," *IEEE Signal Processing Letters*, vol. 20, no. 3, pp. 209–212, Mar. 2013.

[12] M. A. Saad, A. C. Bovik, and C. Charrier, "Blind prediction of natural video quality," *IEEE Transactions on Image Processing*, vol. 23, no. 3, pp. 1352–1365, Mar. 2014.

[13] A. Mittal, M. A. Saad, and A. C. Bovik, "A completely blind video integrity oracle," *IEEE Transactions on Image Processing*, vol. 25, no. 1, pp. 289–300, Jan. 2016.

[14] H. Men, H. Lin, and D. Saupe, "Empirical evaluation of no-reference VQA methods on a natural video quality database," in *9th International Conference on Quality of Multimedia Experience (QoMEX)*. IEEE, 2017.

[15] H. Men, H. Lin, and D. Saupe, "Spatiotemporal feature combination model for no-reference video quality assessment," in *10th International Conference on Quality of Multimedia Experience (QoMEX)*. IEEE, 2018.

[16] M. Torres Vega, D. C. Mocanu, S. Stavrou, and A. Liotta, "Predictive no-reference assessment of video quality," *Signal Processing: Image Communication*, vol. 52, pp. 20–32, Mar. 2017.

[17] M. H. Pinson and S. Wolf, "A new standardized method for objectively measuring video quality," *IEEE Transactions on Broadcasting*, vol. 50, no. 3, pp. 312–322, Sep. 2004.

[18] C. A. Mello, M. M. .Saraiva, D. P. .Menor, and R. Nishihara, "A comparative study of objective video quality assessment metrics," *Journal of Universal Computer Science*, vol. 23, no. 5, pp. 505–527, May 2017.

[19] H. A. Becerra Martinez and M. C. Q. Farias, "Combining audio and video metrics to assess audio-visual quality," *Multimedia Tools and Applications*, Feb. 2018.

[20] A. Mittal, A. K. Moorthy, and A. C. Bovik, "Blind/referenceless image spatial quality evaluator," in *45th Asilomar Conference on Signals, Systems and Computers (ASILOMAR)*, Nov. 2011.

[21] A. K. Moorthy and A. C. Bovik, "A two-step framework for constructing blind image quality indices," *IEEE Signal Processing Letters*, vol. 17, no. 5, pp. 513–516, May 2010.

[22] ITU, "Methods for the subjective assessment of video quality, audio quality and audiovisual quality of internet video and distribution quality television in any environment," ITU-T, Tech. Rep. P.913, March 2016.

[23] S. Péchard, R. Pépion, and P. Le Callet, "Suitable methodology in subjective video quality assessment: a resolution dependent paradigm," in *International Workshop on Image Media Quality and its Applications*, Sep. 2008.

[24] B. Thomee, D. A. Shamma, G. Friedland, B. Elizalde, K. Ni, D. Poland, D. Borth, and L.-J. Li, "YFCC100m: The new data in multimedia research," *Communications of the ACM*, vol. 59, no. 2, pp. 64–673, 2016.

[25] V. Hosu, F. Hahn, M. Jenadeleh, H. Lin, H. Men, T. Szirányi, S. Li, and D. Saupe, "The Konstanz natural video database (KoNViD-1k)," in *9th International Conference on Quality of Multimedia Experience (QoMEX)*. IEEE, 2017, pp. 1–6.

[26] P. V. Vu and D. M. Chandler, "ViS3: An algorithm for video quality assessment via analysis of spatial and spatiotemporal slices," *Journal of Electronic Imaging*, vol. 23, no. 1, Feb. 2014.

[27] F. Battisti, M. Carli, and P. Paudyal, "QoS to QoE mapping model for wired/wireless video communication," in *Proc. of Euro Med Telco Conference*, Nov. 2014.

[28] P. Paudyal, F. Battisti, and M. Carli, "Study of the effects of video content on quality of experience," in *Proc. of Image Quality and System Performance XII, IS&T/SPIE Electronic Imaging*, vol. 9396, Feb. 2015.

[29] K. Seshadrinathan, R. Soundararajan, A. C. Bovik, and L. K. Cormack, "Study of subjective and objective quality assessment of video," *IEEE Transactions on Image Processing*, vol. 19, no. 6, pp. 1427–1441, 2010.

[30] K. Seshadrinathan, R. Soundararajan, A. C. Bovik, and L. K. Cormack, "A subjective study to evaluate video quality assessment algorithms," in *Human Vision and Electronic Imaging XV*, SPIE, Ed., vol. 7527, Jan. 2010, pp. 7527–7527.

[31] A. K. Moorthy, L. K. Choi, A. C. Bovik, and G. de Veciana, "Video quality assessment on mobile devices: Subjective, behavioral and objective studies," *IEEE Journal of Selected Topics in Signal Processing*, vol. 6, no. 6, pp. 652–671, Oct. 2012.

[32] A. K. Moorthy, L. K. Choi, G. de Veciana, and A. C. Bovik, "Mobile video quality assessment database," in *IEEE ICC 2012 International Conference on Communications, Workshop on Realizing Advanced Video Optimized Wireless Networks*, June 2012.

[33] A. K. Moorthy, L. K. Choi, G. de Veciana, and A. C. Bovik, "Subjective analysis of video quality on mobile devices," in *Sixth International Workshop on Video Processing and Quality Metrics for Consumer Electronics (VPQM)*, Jan. 2012.

[34] ITU, "Subjective video quality assessment methods for multimedia applications," ITU-T, Tech. Rep. P.910, Apr. 2008.

[35] D. Melcher and S. Wolf, "Objective measures for detecting digital tiling," *NTIA/ITS, T1A1.5*, vol. 5, pp. 95–104, Jan. 1995.

[36] D. Hasler and S. E. Suesstrunk, "Measuring colorfulness in natural images," in *Human vision and electronic imaging VIII*, vol. 5007. SPIE, 2003, pp. 87–96.

[37] E. Peli, "Contrast in complex images," *Journal of the Optical Society of America part A*, vol. 7, no. 10, pp. 2032–2040, 1990.

[38] Q. Huynh-Thu and M. Ghanbari, "No-reference temporal quality metric for video impaired by frame freezing artefacts," in *ICIP 2009, IEEE International Conference on Image Processing*. IEEE, Nov. 2009, pp. 2221–2224.

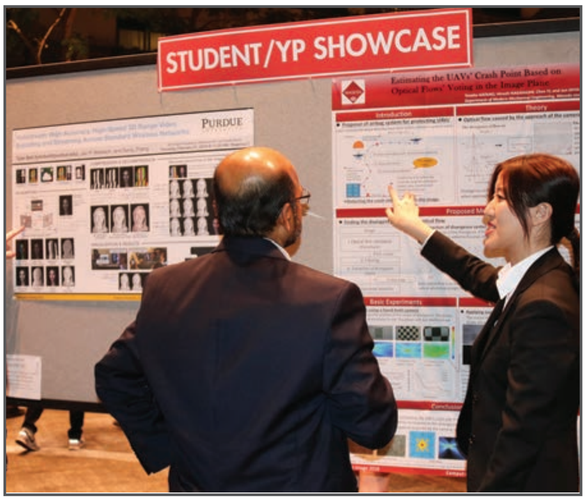
JOIN US AT THE NEXT EI!

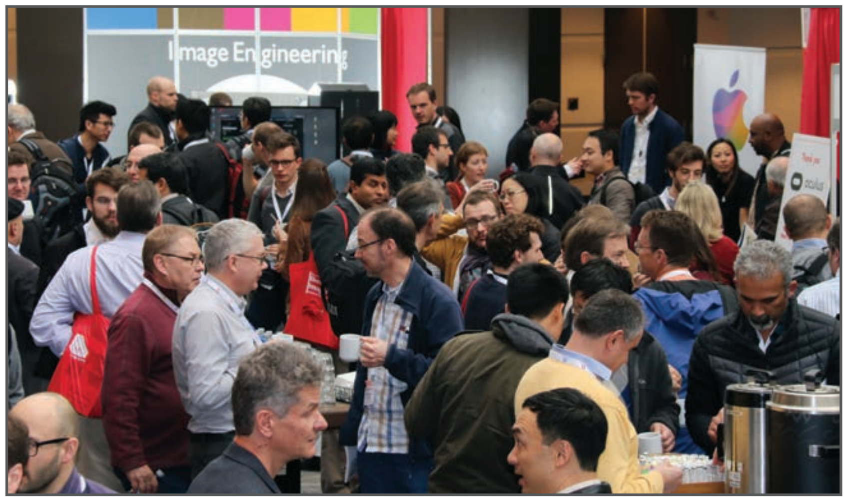
IS&T International Symposium on

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